Rozwiązania technologiczne XXI wieku – skutki i perspektywy rozwoju. Tom 4

Rozwiązania technologiczne XXI wieku – skutki i perspektywy rozwoju. Tom 4

Redakcja:
Izabela Mołdoch-Mendoń
Kamil Maciąg

Lublin 2023

Wydawnictwo Naukowe TYGIEL składa serdeczne podziękowania zespołowi Recenzentów za zaangażowanie w dokonane recenzje oraz merytoryczne wskazówki dla Autorów.

Recenzentami niniejszej monografii byli:

- prof. ndzw. dr hab. inż. Czesław Łukianowicz
- prof. dr hab. inż. Grzegorz Wielgosiński
- dr hab. Agata Opolska-Bielańska
- dr hab. inż. Krzysztof Regulski
- dr hab. inż. Beata Smyrak, prof. AGH
- dr hab. inż. Jarosław Grzegorz Ziółkowski
- dr inż. Martyna Białecka
- dr Renata Brajer-Marczak
- dr inż. Dorota Dardas
- dr inż. Paweł Dymora
- dr n. o zdr. Mariola Janiszewska
- dr Łukasz Karpiński
- dr inż. Marcin Lijewski
- dr inż. Beata Majkowska-Marzec
- dr inż. Łukasz Mrozik
- dr Ewa Pachura
- dr inż. Katarzyna Maria Rutczyńska-Wdowiak
- dr inż. Łukasz Szałata
- dr inż. Agnieszka Szczotok
- dr inż. Artur Wiliński

Wszystkie opublikowane rozdziały otrzymały pozytywne recenzje.

Skład i łamanie:
Monika Maciąg

Projekt okładki:
Marcin Szklarczyk

Korekta:

© Copyright by Wydawnictwo Naukowe TYGIEL sp. z o.o.

ISBN 978-83-67881-09-8

Wydawca:
Wydawnictwo Naukowe TYGIEL sp. z o.o.
ul. Głowackiego 35/341, 20-060 Lublin
www.wydawnictwo-tygiel.pl

Spis treści

Justyna Waśniowska
Pomiar goniofotometryczny rozsyłu światłości w systemie *C-γ* oświetlacza przemysłowego ... 7

Monika Walkowicz, Anna Różańska
Rozwiązania materiałowe i technologiczne w obszarze utleniania powierzchni dotykowych z miedzi i stopów miedzi o właściwościach przeciwdrobnoustrojowych ... 17

Szymon Senyk, Tadeusz Kałdoński
Właściwości tribologiczne smarów plastycznych zawierających heksagonalny azotek boru o różnej granulacji ... 32

Filip Turoboś, Mateusz Krukowski, Nicole Meisner
Przegląd wybranych algorytmów aproksymacyjnych dla problemu komiwojażera ... 55

Michał Bugaj, Krzysztof Wróbel, Katarzyna Szumielewicz
Przegląd wybranych metod wytłumaczalności modeli uczenia maszynowego oraz uczenia głębokiego ... 75

Artur Wilczek, Julia Nowak
Opracowanie wskaźnika inteligencji lingwistycznej na podstawie wybranych cytatów bohaterów literackich ... 85

Magda Konieczna
Metody, techniki i problemy sieci neuronowych ... 95

Filip Turoboś, Mateusz Krukowski
Przybliżone rozwiązania problemu komiwojażera dla grafów semimetrycznych ... 105

Halina Marczak
Wpływ elektromobilności na jakość powietrza ... 130

Monika Walkowicz, Anna Różańska
Problematyka i metodologia badań powierzchni dotykowych z miedzi i stopów miedzi o właściwościach przeciwdrobnoustrojowych ... 141

Andrzej Wartecki, Wiesław J. Jermacz, Adam Lewandowski
Innowacyjno-rozwojowe uwarunkowania działalności spółki wodociągowej w ujęciu ekonomicznym oraz ekologicznym na przykładzie Śremskich Wodociągów Spółka z o.o. ... 152

Agnieszka Kucharska-Jastrząbek, Paulina Dmowska-Jasek
VESTA – damska kamizelka skrytego noszenia – od pomysłu do projektu ... 173

Paulina Dmowska-Jasek, Agnieszka Kucharska-Jastrząbek
Indywidualizacja damskiej kamizelki skrytego noszenia pod kątem ergonomii użytkowania 184

Beata Sokołowska
Przykłady oceny parametrów i/lub markerów w modelowym podejściu z klasyfikacją nadzorowaną w badaniach w realnej i wirtualnej rzeczywistości 198

Patryk Warzybok, Grzegorz Dziatkiewicz
Planowanie ścieżki uczenia się i modelowanie dynamiki zmian kompetencji inżynierskich w zakresie obsługi systemów klasy CAx....230

Paweł Buchwald, Aleksander Dawid
Metoda walidacji aplikacji internetowych przeznaczonych dla osób ze szczególnymi potrzebami....241

Anna Jackiewicz-Zagórska, Jakub Maksymilian Gac, Bartosz Nowak, Łukasz Werner
Modelowanie procesu usuwania z powietrza cząstek aerozolowych w nowoczesnych kompozytowych materiałach polipropylen-ZnO....254

Kamila Steckiewicz
Wyznaczanie temperatury błysku w układach hamulcowych 267

Maciej Słowik, Małgorzata Gulewicz
Symulacyjna i eksperymentalna estymacja pozycji roweru – przykład projektu B+R SmartBrake....286

Grzegorz Drozdowski, Paweł Dziekański
Wpływ zachowań przedsiębiorczych pracowników na funkcjonowanie przedsiębiorstw w dostosowaniu się do otoczenia nowych technologii....302

Bożena Gajdzik
Adaptacyjność przedsiębiorstw w Polsce w ostatnich trzech dekadach. Od restrukturyzacji naprawczej po inteligentne technologie 314

Barbara Wieczorek
Zabezpieczenie połączenia płyta-słup konstrukcji żelbetowych przed katastrofą postępującą w świetle przepisów normowych 328

Barbara Wieczorek
Przegląd i porównanie wyników badań eksperymentalnych połączenia płyta-słup monolitycznych konstrukcji żelbetowych w odniesieniu do wytycznych normowych .340

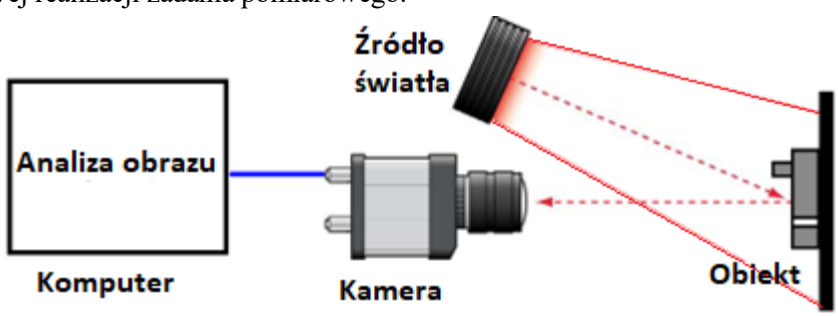
Indeks Autorów....351

Justyna Waśniowska[1]

Pomiar goniofotometryczny rozsyłu światłości w systemie *C-γ* oświetlacza przemysłowego

1. Wstęp

Zadania kontrolne i pomiarowe realizowane z wykorzystaniem analizy obrazu wymagają indywidualnego doboru konfiguracji systemu wizyjnego. W skład każdego systemu wizyjnego wchodzi oświetlacz emitujący promieniowanie elektromagnetyczne w kierunku obiektu. Następnie promieniowanie odbite lub rozproszone na powierzchni obiektu jest rejestrowane na matrycy sensorów zabudowanej w kamerze odpowiedzialnej za akwizycję obrazu sceny. Kolejnym elementem systemu wizyjnego jest oprogramowanie wykorzystywane do analizy obrazu, umożliwiające wyznaczenie parametrów charakterystycznych dla wybranych cech badanego obiektu. Dobór konfiguracji systemu wizyjnego, typu oświetlacza oraz typu kamery rejestrującej obraz jest częścią procesu projektowania wizyjnych systemów pomiarowych. Należy stwierdzić, że prawidłowy dobór oświetlacza do zadania pomiarowego ma kluczowe znaczenie z punktu widzenia prawidłowej realizacji zadania pomiarowego.



Rysunek 1. System wizyjny [1]

Rozwój technik oświetlenia można datować już na 300 000 lat temu w czasach odkrycia ognia przez prehistorycznego człowieka, któremu służył jako źródło ciepła i światła. Ogniska, pochodnie i pierwsze lampy łojowe i olejowe umożliwiły życie w jaskiniach i tworzenie pierwszych malowideł ściennych tj. pierwszych obrazów. Pierwsze rysunki malowane w jaskiniach musiały być przecież wykonane przy użyciu sztucznym oświetleniu. Rozwój techniki oświetleniowej umożliwił budowę pierwszych latarni morskich oraz lamp instalowanych na ulicach. Aktualnie dostępna jest bardzo szeroka gama źródła światła o różnych charakterystykach widmowych i o różnym przeznaczeniu. Wybór oświetlacza do konkretnego zadnia kontrolnego lub pomiarowego jest prowadzony na podstawie analizy charakterystyki widmowej oświetlacza oraz analizy parametrów jego pracy. W artykule przedstawiono pomiary wybranych parametrów pracy oświetlacza, które wpływają na parametry obrazu rejestrowanego przez wizyjny system pomiarowy.

[1] wasniows@agh.edu.pl, AGH Akademia Górniczo-Hutnicza, www.agh.edu.pl.

Pomiar goniofotometryczny jest podstawowym pomiarem jednej z najważniejszych charakterystyk źródeł światła, układów świetlno-optycznych, czyli rozsyłu światłości [2-4]. Realizowany jest z wykorzystaniem zespołu głowicy fotometrycznej i goniometru [4]. Celem wykonywania badań goniofotometrycznych jest przedstawienie rozsyłu światła pochodzącego ze źródła światła, układu świetlno-optycznego, za pomocą bryły fotometrycznej. Bryłę fotometryczną uzyskuje się na podstawie pomiarów natężenia oświetlenia za pomocą luksomierza. Natężenie oświetlenia jest przeliczane na światłość zgodnie z prawem odwrotności kwadratu odległości. Bryła fotometryczna przedstawia rozsył światła w sposób trójwymiarowy. Pomiary natężenia oświetlenia wykonuje się punktowo na wcześniej określonej siatce pomiarowej. Przez to uzyskanie brył fotometrycznych o zadowalającej rozdzielczości kątowej jest czasochłonne. Pomiar może trwać nawet do kilkunastu godzin [2], a niekiedy nawet do kilkudziesięciu godzin.

Można stosować różne konfiguracje goniofotometrów, np.: goniofotometry wykorzystujące ruchome zwierciadło, fotometry ramienne lub inne [2, 5]. Pomiar goniofotometryczny z wykorzystaniem luksomierza jest często stosowany do pomiarów przemysłowych opraw oświetleniowych oraz oświetlaczy. Pomiary wykonywane są w ciemni fotometrycznej – pomieszczeniu o dużych gabarytach, pomalowanym na czarno. Czarny kolor ma na celu pochłaniać światło, aby nie odbijało się od ścian i nie docierało do miernika z innych kierunków.

Pomiar goniofotometryczny można wykonywać w trzech różnych systemach geometrycznych. Mianowicie systemie *A-α, B-β, C-γ*. W niniejszym artykule przestawiono i opisano pomiar goniofotometryczny w systemie *C-γ* z wykorzystaniem goniometru oraz luksomierza [5].

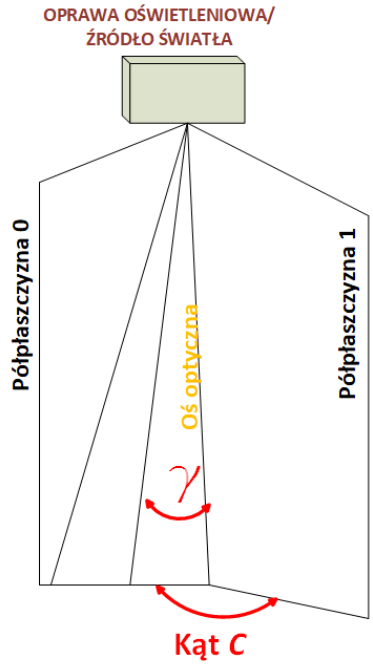
Pomiar goniofotometryczny najczęściej wykonuje się zgodnie z obowiązującymi normami w zależności od badanego układu świetlno-optycznego. Na przykład dla lamp samochodowych odległość fotometrowania wg Regulaminu nr 65 WKG ONZ wynosi 25 m [6]. Według „Dziennika miar i probiernictwa" graniczną odległość fotometrowania definiuje się jako *odległość, od której stosuje się prawo odwrotności kwadratu odległości* [7].

W artykule przedstawiono metodę pomiaru ze szczegółowym opisem urządzeń wykorzystywanych w pomiarach oraz etapów realizacji pomiaru. W podsumowaniu przedstawiono przykładowe wyniki pomiarów oraz kształt wyznaczonej bryły fotometrycznej.

2. Metoda pomiaru z wykorzystaniem goniofotometru

Pomiar goniofotometryczny wykonywany jest w taki sposób, że oprawę oświetleniową mocuje się na uchwycie do goniometru. Goniometr umożliwia obrót oprawy oświetleniowej oświetlacza w dwóch płaszczyznach o kąt *C* i γ.

Pomiar goniofotometryczny pozwala zmierzyć natężenie oświetlenia *E* w danym kierunku w odległości fotometrowania za pomocą luksomierza. Dane pomiaru każdego punktu pomiarowego są zapisywane do pliku, a następnie zgodnie z prawem odwrotności kwadratu odległości wg wzoru (2) obliczana jest światłość i zapisywana do pliku o rozszerzeniu .ldt lub .ies. W pliku .ldt lub .ies znajdują się również takie dane jak liczba półpłaszczyzn, liczba kątów gamma, strumień świetlny, kąt połówkowy rozsyłu światła itd.



Rysunek 2. Schemat przestawiający źródło światła oraz dwie półpłaszczyzny wraz z zaznaczonymi kątami *C* i γ, opracowanie własne

Poniższy wzór (1) pozwala zgodnie z prawem odwrotności kwadratu obliczyć natężanie oświetlenia.

$$E = \frac{I}{R^2} \tag{1}$$

gdzie: *E* – natężenie oświetlenia, *I* – światłość w danym kierunku, *R* – odległość środka świetlnego od głowicy pomiarowej luksomierza.

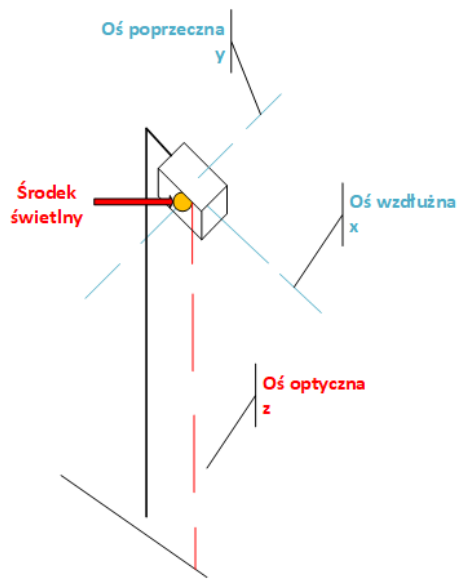
Na podstawie wzoru (1), przekształcając go otrzymujemy wzór umożliwiający obliczenie światłości. Przy czym natężenie oświetlenia uzyskujemy z pomiarów.
A odległość *R* jest to odległość między środkiem świetlnym w tym przypadku głowicą pomiarową luksomierza.

$$I = E \cdot R^2 \tag{2}$$

gdzie: *E* – natężenie oświetlenia, *I* – światłość w danym kierunku, *R* – odległość środka świetlnego od głowicy pomiarowej luksomierza.

Pliki .ies lub .ldt można otworzyć za pomocą specjalistycznych programów np. QLumEdit, Dialux, Speos 2023 R1 bądź innych. W programie QLumEdit można zobaczyć m.in. wykres światłości dla płaszczyzn *C*0-180 oraz *C*90-270. W dalszej części artykułu zostało wyjaśnione w jaki sposób konstruujemy pęk półpłaszczyzn i oznaczamy kąty *C*. Program Dialux umożliwia wykonywanie obliczeń symulacyjnych natężenia oświetlenia. Natomiast oprogramowanie Speos 2023 R1 umożliwia bardziej zaawansowane obliczenia naukowe.

Środek świetlny można opisać jako punkt, który można utożsamić ze środkiem geometrycznym źródła światła. Natomiast oś optyczna jest to prosta, która przechodzi przez środek świetlny oprawy albo źródła światła. Jej kierunek jest kierunkiem światłości o maksymalnej wartości i/lub kierunkiem osi symetrii układu geometrycznego oprawy oświetleniowej. Często jest to kierunek prostopadły do płaszczyzny przechodzącej przez krawędzie (odcinki) otworu wyjściowego oprawy oświetleniowej [5]. Rysunek 3 w sposób uproszczony przedstawia to jak usytuowane w przestrzeni są osie: poprzeczna, wzdłużna, optyczna. Oś wzdłużna i poprzeczna najczęściej zależy od gabarytów oprawy oświetleniowej. Oś wzdłużna jest równoległa do dłuższego boku oprawy, a oś poprzeczna do krótszego.



Rysunek 3. Oprawa oświetleniowa z zaznaczonymi osiami, opracowanie własne na podstawie [5]

System pomiaru określany jako *C-γ* opiera się na spostrzeżeniu, że wiele opraw oświetleniowych posiada oś optyczną skierowaną pionowo w dół. To umożliwia wyznaczenie pęku półpłaszczyzn o wspólnej krawędzi będącej osią optyczną. Następnie należy wyznaczyć kąty gamma. Są to kąty płaskie leżące na każdej półpłaszczyźnie. Kąt gamma to kąt między osią optyczną a półprostą o początku w środku świetlnym oprawy.

System *C-γ* opiera się na wykorzystaniu pęku półpłaszczyzn o wspólnej półprostej, która jest osią optyczną jak pokazano to na rysunku 4.

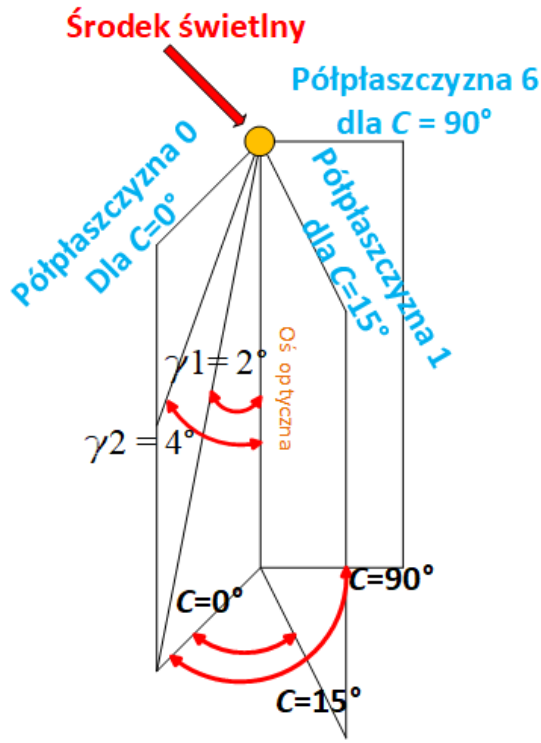
Kąty *C* i *γ* umożliwiają określenie dowolnego kierunku w przestrzeni. Kąt *C* to kąt dwuścienny pomiędzy półpłaszczyzną uznawaną za wyjściową, a zadaną półpłaszczyzną. Na rysunku 4 oznaczono dla uproszczenia tylko półpłaszczyzny 0,1,6. Kąt *C*n to kąt pomiędzy półpłaszczyzną 0 i 1 oraz 0 i 2 itd. aż do n+1 półpłaszczyzn przy czym ostatnia półpłaszczyzna to półpłaszczyzna 0 (ponieważ kąt pełny ma 360°). Upraszczając, kąt 360° dzielimy na liczbę n półpłaszczyzn. Wartość, którą otrzymujemy to wartość kąta dwuściennego pomiędzy półpłaszczyznami. Przy czym kąt *C* to kąt pomiędzy półpłaszczyzną wyjściową a półpłaszczyzną zadaną. Zatem jest wielokrotnością kąta otrzymanego wcześniej kąta dwuściennego.

Kąt γ to kąt płaski zawierający się między prostą o zadanym kierunku i osią optyczną. Kąty g leżą na półpłaszczyźnie *C*(n*wyliczony kąt dwuścienny) (gdzie n – numer kolejnej półpłaszczyzny).

Można wybrać dowolną liczbę półpłaszczyzn. Jednak programy obliczeniowe najczęściej przyjmują, że kąt między dwoma półpłaszczyznami wynosi 15°, a różnica między dwoma kątami $\gamma 1$ i $\gamma 2$ wynosi 2°.

Oczywiście możliwe jest wykonywanie pomiarów z inną liczbą płaszczyzn oraz kątów γ. Takie obliczenia są możliwe np. w oprogramowaniu Speos 2023 R1.

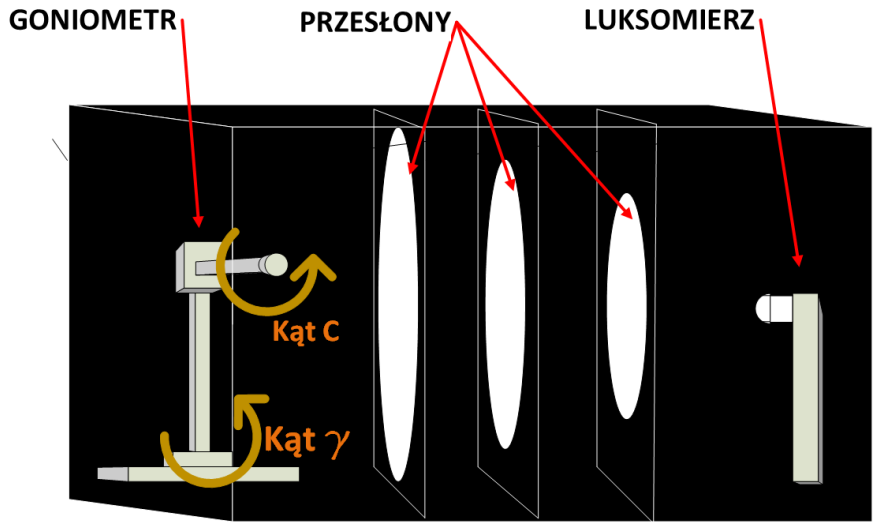
Pomiar w zależności od wyznaczonej siatki pomiarowej (liczby punktów pomiarowych) może trwać od kilku godzin dla mniejszej liczby punktów pomiarowych do kilkunastu godzin dla większej liczby punktów pomiarowych.



Rysunek 4. Uproszczony schemat półpłaszczyzn C oraz zaznaczony jeden z kątów g, opracowanie własne

Na powyższym rysunku przestawiono w sposób schematyczny pęk półpłaszczyzn, kąt $C = 0°$ oraz $C = 15°$ a także zaznaczono kąt gamma.

Ciemnia fotometryczna jest to pomieszczenie o dużych rozmiarach w kształcie prostopadłościanu. Jest pomalowane na czarno tak, aby światło nie odbijało się od ścian. Między goniometrem a luksomierzem najczęściej znajdują się rozmieszczone przysłony tak, aby światło nie docierało z innych kierunków do luksomierza. Z jednej strony pomieszczenia usytuowany jest goniometr. Natomiast z drugiej jest umieszczony w osi optycznej luksomierz. Jak już wspomniano luksomierz stosuje się do pomiaru natężenia promieniowania elektromagnetycznego w zakresie widzialnym, czyli mówiąc prościej natężenia oświetlenia.



Rysunek 5. Ciemnia fotometryczna pokazana na rysunku w sposób schematyczny, opracowanie własne

Goniometr do pomiarów fotometrycznych w systemie *C-γ* jest to urządzenie, na którym zawieszamy oprawę oświetleniową. Umożliwia on, dzięki stosowanym silnikom, obrót oprawy dla dwóch kątów *C* oraz *γ*. Typowa konstrukcja goniometru wygląda w taki sposób, że posiada dwa ramiona. Jedno jest przytwierdzone pionowo do podłoża i obraca się względem podstawy umożliwiając obrót o kąt *γ*. Drugie ramie to pewnego rodzaju wysięgnik, do którego przymocowane są szczęki obracając się o kąt *C*. Szczęki są po to, aby umożliwić przytwierdzenie uchwytu wraz z oprawą. Można stosować inny element umożliwiający mocowanie oprawy. Najczęściej konstrukcja goniometru umożliwia ustalenie zadanej odległości środka świetlnego od głowicy pomiarowej luksomierza z zadaną tolerancją. Oprawa oświetleniowa jest zawieszona na nim w taki sposób, aby środek świetlny pokrywał się ze środkiem głowicy luksomierza oraz oś optyczna pokrywała się z osią głowicy luksomierza.

Luksomierz jest podstawowym i bardzo często stosowanym miernikiem w fotometrii [4]. Luksomierz składa się z głowicy pomiarowej oraz miernika prądu fototelektrycznego, który współpracuje ze wzmacniaczem. Zgodnie z Polskimi Normami zaleca się, aby luksomierze wykorzystywane do pomiarów natężenia oświetlenia wnętrz były wzorcowane i kalibrowane. Na dokładność pomiarową luksomierza wpływa kilka czynników. Jednym z nich jest źródło światła wzorca. Ze względu na to, że jako wzorca używa się najczęściej lampy żarowej mogą pojawić się pewne błędy pomiarowe. Kolejnym czynnikiem wpływającym na błąd pomiarowy luksomierza jest błąd kosinusowy, którego wpływ może ograniczać tzw. korekcja kosinusowa. Jej rolą jest wyeliminowanie błędów pomiarów, dla światła padającego pod innym kątem niż 0°. Innym parametrem charakteryzującym luksomierz jest współczynnik (wskaźnik) korekcji barwowej. Niestety ze względu na to, że istnieje duża liczba różnego rodzaju źródeł światła o różnej charakterystyce widmowej oraz to, że po odbiciu się od ścian, mebli itp. światło pochodzące z różnych źródeł dociera do miernika o jeszcze innej charakterystyce barwowej. To znacznie utrudnia wyeliminowanie pewnych błędów pomiarowych [8].

3. Wyniki badań doświadczalnych

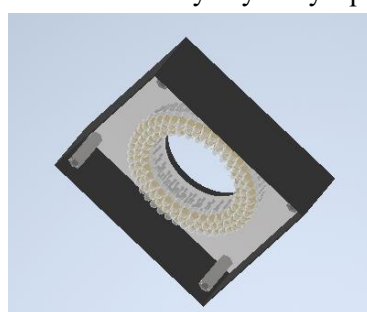
3.1. Opis stanowiska pomiarowego

Stanowisko pomiarowe składało się z goniometru i luksomierza oraz badanego oświetlacza umieszczonych w ciemni fotometrycznej. Goniometr jest skonstruowany w taki sposób, że umożliwia umocowanie oświetlacza tak, aby jego środek świetlny znajdował się w odległości 10 m od głowicy pomiarowej luksomierza.

Badania były wykonywane z siatką pomiarową dla kąta dwuściennego pomiędzy dwoma półpłaszczyznami *C* równemu 15° oraz kąta płaskiego γ zmieniającego się co 2°. Dane w postaci natężenia oświetlenia zostały zebrane i przekształcone na wartość światłości. Wygenerowano później gotowy plik .ldt.

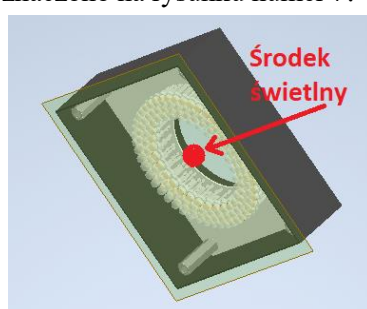
3.2. Badane źródło światła

Badanym źródłem światła jest oświetlacz przemysłowy typu „ring light". Obudowa oświetlacza wykonana jest w kształcie kwadratu o wymiarach ok. 80 x 80 mm z otworem w środku. Diody są rozmieszczone pierścieniowo wokół otworu w sposób prezentowany na rysunku 6. Oś optyczna każdej z diod jest ustawiona równolegle do osi otworu wykonanego w obudowie. Strumień świetlny uzyskany z pomiarów wynosi 338 l m.



Rysunek 6. Badane źródło światła, opracowanie własne

Środek świetlny oraz płaszczyznę przechodzącą przez krawędzie otworu oświetlacza wyjściowego oświetlacza oznaczono na rysunku numer 7.



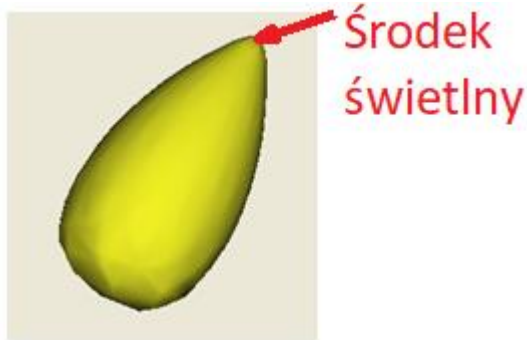
Rysunek 7. Badane źródło światła z zaznaczonym środkiem świetlnym oraz płaszczyzną przechodząca przez krawędzie otworu wyjściowego światła, opracowanie własne

3.3. Bryła fotometryczna badanego źródła światła

Bryła fotometryczna to graficzne trójwymiarowe przestawienie rozsyłu światła w przestrzeni. Bryłę fotometryczną można określić jako zbiór linii o długości przedstawiających

wartość światłości w danym kierunku, wychodzących z jednego punktu tj. środka świetlnego [5].

Na poniższym rysunku 8 pokazana została pokazana bryła fotometryczna badanego źródła światła zwizualizowana z pomocą oprogramowania Speos 2023 R1.



Rysunek 8. Rzeczywista bryła fotometryczna przedstawiona przy pomocy programu Speos 2023 R1, opracowanie własne

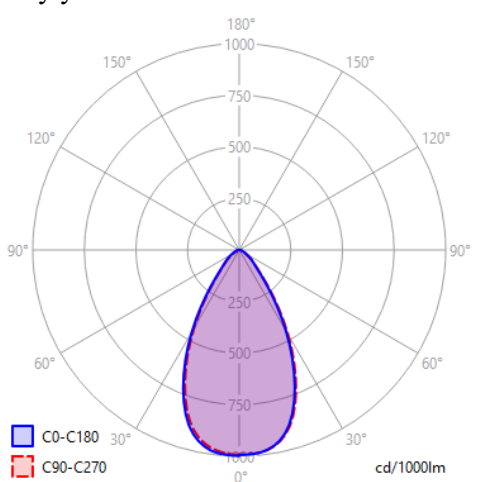
3.4. Wykres światłości

Wykres światłości to przedstawione we współrzędnych biegunowych lub kartezjańskich wartości światłości w zadanych płaszczyznach.

W jaki sposób odczytywać dane z wykresu biegunowego:

Należy wyobrazić sobie, że bierzemy pod uwagę dwie półpłaszczyzny, między którymi kąt wynosi 180°. Na jednej półpłaszczyźnie (od 0° do 180°) są kąty γ co widać na rysunku numer 9. Zatem na wykresie odczytujemy kąt γ w danych dwóch półpłaszczyznach.

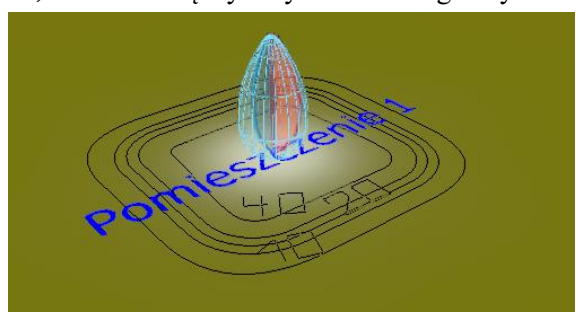
Najczęściej podaje się dwa wykresy dla półpłaszczyzn C0-180° oraz C90-270°. W poniższym przykładzie wykresy nakładają się na siebie ze względu na obrotowo-symetryczny charakter bryły.



Rysunek 9. Wykres światłości w półpłaszczyznach *C*0-180 oraz *C*90-270 wygenerowany w programie QLumEdit we współrzędnych biegunowych, opracowanie własne

3.5. Stosowanie bryły fotometrycznej do obliczeń w Dialux

W tym podrozdziale zostanie przedstawione bardzo proste zastosowanie oprawy oświetleniowej do pomiarów symulacyjnych w środowisku Dialux evo. Jak już wspomniano jest to oprogramowanie służące do badań symulacyjnych natężenia oświetlenia. Jest w nim możliwość wyboru parametrów obliczanych na zadanej powierzchni np. natężenie prostopadłe, natężenie pionowe i poziome itd. Oprogramowanie umożliwia import rysunków .dwg (rysunek CAD). Można też w nim rysować różnego rodzaju elementy. Te rysunki przedstawiają np. budynki, wnętrza budynków. Program umożliwia wybór otoczenia. Oprogramowanie Dialux umożliwia nadawanie powierzchniom koloru i rodzaju materiału, z którego ma być wykonany dany element. Każdy kolor ma inny współczynnik odbicia światła co przekłada się na dalsze obliczenia natężenia oświetlenia w punktach pomiarowych. W zależności od tego czy jest to pomieszczenie, czy zewnątrz budynku lub też ulice, obliczenia są wykonywane według dedykowanych norm.



Rysunek 10. Bryła fotometryczna oraz rozkład natężenia oświetlenia na powierzchni symulowany w oprogramowaniu Dialux evo, opracowanie własne

4. Posumowanie

Pomiar goniofotometryczny służy do wyznaczania bryły fotometrycznej. Jest to możliwe dzięki stosowaniu urządzenia jakim jest goniometr oraz miernikowi natężenia oświetlenia, czyli luksomierzowi. Dane z luksomierza w postaci natężenia oświetlenia w danym punkcie pomiarowym są przeliczane, dzięki prawu odwrotności kwadratu odległości, na światłość. Następnie jest tworzony specjalny plik o rozszerzeniu. .ldt lub .ies. Umieszczane są w nim takie dane jak: typ oświetlacza, strumień światła, liczba półpłaszczyzn oraz liczba kątów gamma, a także światłość w każdym punkcie pomiarowym itp. Pomiar wykonywany jest zgodnie z dedykowanymi normami w ciemni fotometrycznej.

Wykonanie pomiaru światłości w każdym punkcie, a w późniejszym etapie prac umożliwia modelowanie bryły fotometrycznej. Bryła fotometryczna wykorzystywana jest przez projektantów oświetlenia do badań symulacyjnych rozkładu natężenia światła na powierzchni np. w oprogramowaniu Dialux czy Relux. Oprogramowanie Dialux oraz Relux umożliwiają modelowanie struktury stanowiska pomiarowego oraz elementów mających wpływ na odbicia lub rozpraszanie światła z oświetlacza.

Podziękowania

Serdeczne podziękowania za pomoc zarówno merytoryczną, jak i wsparcie w realizacji badań należą się dr hab. inż. Andrzejowi Siomie oraz firmie MILOO-ELECTRONIC Sp. z o.o. za umożliwienie wykonania badań. Dziękuję również rodzinie za wsparcie w dążeniu do osiągania sukcesów.

Literatura

1. Sioma A., *Programowanie systemów wizyjnych w środowisku IVC Studio*, Wydawnictwo Kolumb, Katowice 2010
2. Legierski M., Michałek P., *Zautomatyzowane stanowisko do pomiarów przestrzennych rozsyłów światłości miernikiem matrycowym*, Przegląd Elektrotechniczny, 94, 2018, s. 138-141
3. Legierski M., Michałek P., *Realizacja i zastosowanie szybkiego goniofotometru do pomiarów przestrzennych rozsyłów światłości przednich lamp samochodowych*, Eksploatacja i testy, AUTOBUSY, 2017, s. 1031-1034.
4. Legierski M., Michałek P., *Concept and Implementation of H-V Goniophotometer for Fast Measurement of Luminous Intensity Distribution in C-γ System*, Przegląd Elektrotechniczny, 96, 2020, s. 75-78
5. Żagan W., *Podstawy techniki świetlnej*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2014.
6. Michałek P., Domagała D., *Pomiary fotometryczne błyskowych urządzeń sygnalizacyjnych i ostrzegawczych użyciu szybkiego fotometru*, Eksploatacja i testy, AUTOBUSY, 2017, s. 1042-1045.
7. Dziennik miar i probiernictwa, 6, 1995.
8. Banaszak A., Tabaka P., Wtorkiewicz J., *Analiza wybranych właściwości różnych typów luksomierzy*, Prace Instytutu Elektrotechniki, zeszyt 268, 2015.

Pomiar goniofotometryczny rozsyłu światłości w systemie *C-γ* oświetlacza przemysłowego

Streszczenie

W pracy opisano pomiar goniofotometryczny, który służy do badania parametru pracy źródła światła (układu świetlno-optycznego) jakim jest rozsył światłości w przestrzeni. Jednym z możliwych systemów geometrycznych służących do pomiaru światłości jest system *C-γ*. Został on szczegółowo omówiony w artykule. Rozsył światła w przestrzeni przedstawiamy za pomocą bryły fotometrycznej. Jest ona utworzona na podstawie pomiaru światłości w każdym punkcie pomiarowym. Bryła fotometryczna przestawia rozsył światła w sposób trójwymiarowy. Praca zawiera również opis, bryły fotometrycznej, wykresu światłości oraz innych pojęć związanych z pomiarem goniofotometrycznych. W krótkim wprowadzeniu do metody wykonywania badań przytoczono wzór odnoszący się do prawa odwrotności kwadratu odległości oraz to jak na jego podstawie obliczyć światłość. Wyszczególniono także, które wartości są mierzone. Przestawiono również wyniki badań doświadczalnych oświetlacza przemysłowego. Opracowano jego bryłę fotometryczną, a także wykres światłości.

Słowa kluczowe: pomiar goniofotometryczny, system *C-γ,* bryła fotometryczna

Goniophotometric measurement in the *C-γ* system

Abstract

The article describes the goniophotometric measurement, which is used to study the operating parameter of the light source (light-optical system), which is the distribution of light in space. One of the possible geometrical systems for measuring luminous intensity is the *C-γ* system. It has been discussed in detail in the article. Light distribution in space is represented by a photometric solid. It is created on the basis of luminous intensity measurement at each measurement point. The photometric solid shows the light distribution in a three-dimensional way. The article also presents a description of the photometric solid, the luminous intensity diagram and other concepts related to the goniophotometric measurement. In a short introduction to the test method, the formula referring to the inverse square law and how to calculate luminous intensity on its basis were quoted. It is specified which values are measured. The results of real tests of the industrial illuminator are also presented. Its photometric body and luminous intensity diagram were developed.

Keywords: goniophotometric measurement, *C-γ* system, photometric solid

Monika Walkowicz[1], Anna Różańska[2]

Rozwiązania materiałowe i technologiczne w obszarze utleniania powierzchni dotykowych z miedzi i stopów miedzi o właściwościach przeciwdrobnoustrojowych

1. Wprowadzenie

Zagrożenia epidemiczne spowodowane zakażeniami patogenami alarmowymi, w szczególności drobnoustrojami opornymi na działanie antybiotyków, to jedno z współczesnych wyzwań systemów opieki zdrowotnej [1, 2], które zapoczątkowały rozwój kompleksowych krajowych, europejskich i światowych działań ukierunkowanych na monitorowanie stanu zdrowia społeczeństwa, jego oceny i w przypadku stwierdzenia nieprawidłowości wprowadzania działań prewencyjnych [3, 4]. Takie działania wymagają jednak współpracy i zaangażowania wielu różnych podmiotów w celu dostarczenia wiarygodnych informacji o częstości występowania zakażeń i ich wpływu na zdrowie populacji. Jak podkreślają autorzy różnych prac takich kompleksowych danych jest wciąż niewiele [5, 6]. W literaturze dostępnych jest wiele opracowań dotyczących skali problemu zakażeń wielolekoopornymi bakteriami, głównie z wykorzystaniem standardowych współczynników epidemiologicznych w odniesieniu do poszczególnych form zakażeń, grup drobnoustrojów, czy w ujęciu geograficznym – na różnych poziomach, od pojedynczych oddziałów szpitalnych do danych bardziej zagregowanych, np. na poziomie kraju. Jednak dla celu zobrazowania skutków takich zakażeń w ujęciu mogącym stanowić porównanie z innymi problemami zdrowia publicznego, funkcjonują specyficzne współczynniki, np. wskaźnik DALY (ang. *Disability Adjusted Life Year*) [7]. Wskaźnik ten definiowany jest sumarycznie jako liczba lat życia z niepełnosprawnością (YLD, ang. *Years Lived with Disability*) i liczba utraconych lat życia z powodu przedwczesnego zgodu (YLL, ang. *Years of Life Lost*). Wskaźnik ten może być obliczany łącznie dla całej populacji, ale także dla poszczególnych grup wiekowych, płci, czy z uwzględnieniem innych kryteriów. Metodyka ta wykorzystana została w badaniach międzynarodowego zespołu ds. zakażeń szpitalnych, a wyniki analiz opublikowano w „Lancet Infection Disease" [8]. Dowiedziono, że blisko 70% zakażeń szczepami wieloopornych bakterii nabywanych jest przez ludzi w trakcie hospitalizacji w jednostkach opieki zdrowotnej. Dwoma najbardziej newralgicznymi grupami wiekowymi są pacjenci powyżej 65. roku życia oraz niemowlęta i noworodki wcześniaki z małą masą urodzeniową. Wskaźnik DALY na 100 000 osób przyjmuje dla tych skrajnych grup wiekowych wartości na poziomie odpowiednio 700 i 2000 lat stanowiąc o uszczerbku na zdrowiu dla jednej osoby i dla jednego zakażenia średnio na poziomie 3 i 7 dni. Wydawać by się mogło, że to niewiele, ale należy zaznaczyć, że to wyniki jedynie statystyczne. Jeden pacjent może w ogóle nie doświadczyć

[1] monika.walkowicz@agh.edu.pl, AGH Akademia Górniczo-Hutnicza, Wydział Metali Nieżelaznych, Katedra Przeróbki Plastycznej i Metaloznawstwa Metali Nieżelaznych, al. A. Mickiewicza 30, 30-059 Kraków.
[2] a.rozanska@uj.edu.pl, Zakład Kontroli Zakażeń i Mykologii, Katedra Mikrobiologii Wydziału Lekarskiego CMUJ, ul. Czysta 18, 31-121 Kraków.

zakażeń, inny z kolei w trakcie życia nabędzie je kilkanaście lub kilkadziesiąt razy, co dodatkowo nierzadko zakończy się ciężkim kalectwem.

W Polsce DALY jest wyższy od średniej wartości dla wszystkich krajów europejskich (190 DALY/100 000 osób vs 170 DALY/100 000 osób [8]) świadcząc z jednej strony o większych czynnikach ryzyka tj. większej zapadalności na choroby i jednocześnie gorszych konsekwencjach tj. niskiej jakości usług leczenia, a także samej profilaktyki. Umownie DALY uważany jest za wyznacznik rozwoju cywilizacyjnego społeczeństwa. Wyniki przedstawionych badań są zbieżne z danymi publikowanymi rokrocznie przez Europejskie Centrum Zapobiegania i Kontroli Chorób (ECDC, ang. *European Centre for Disease Prevention and Control*). Przykładowo tylko w odniesieniu do naszego zachodniego sąsiada (Niemiec), w Polsce notuje się więcej inwazyjnych zakażeń wywołanych przez szczepy *Acinetobacter baumannii, Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa* [8]. Części z tych zakażeń oczywiście trudno jest uniknąć, jednak sporemu odsetkowi z nich można zapobiegać prostymi metodami takimi jak higiena szpitalna czyli dezynfekcja środowiska oraz higiena rąk. Jednak wyniki polskich badań, zarówno obserwacyjnych, jak i ankietowych wskazują na poważne rozbieżności praktyki higieny rąk z zaleceniami. Otóż, wg wyników badania obserwacyjnego, jedynie 5% pielęgniarek i 17% lekarzy w naszym kraju myje ręce przed kontaktem z pacjentem [9]. Po kontakcie z chorym wartości te są już na wyższym poziomie, odpowiednio 25% i 53% [10], jednak bezwzględnie powinny wynosić 100%.

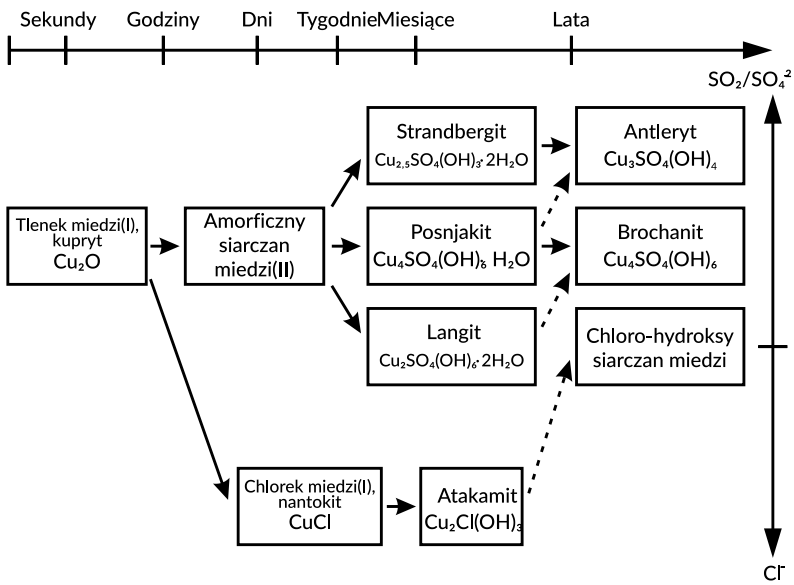
Dlatego też coraz chętniej w krajach na całym świecie sięga się po rozwiązania wspomagające, do których należy m.in. stosowanie na oddziałach szpitalnych wyrobów z powierzchniami przeciwdrobnoustrojowymi z miedzi i jej stopów. Takie remedium stosuje aktualnie około 8% szpitali w Polsce, zwłaszcza w zachodniej części kraju. Zaznaczyć należy jednak, że to temat względnie młody gdyż te instalacje wdrożone zostały do jednostek opieki zdrowotnej w przeciągu ostatnich kilku lat, zwłaszcza po 2015 roku kiedy Ministerstwo Zdrowia wydało standardy akredytacyjne przyznające dodatkowe punkty szpitalom stosującym miedziane wyroby dotykowe [11]. Informacja ta w połączeniu z faktem, że takie instalacje są powoli, ale sukcesywnie wdrażane do krajowej służby zdrowia jest pozytywna, tym bardziej że powszechnie znane są problemy natury organizacyjnej, ekonomicznej i prawnej z którymi borykają się krajowe jednostki opieki medycznej.

Efektywne upowszechnienie miedzianych, dotykowych wyrobów przeciwbakteryjnych wymaga znajomości zagadnień materiałowych, technicznych i technologicznych z zakresu inżynierii materiałowej, zwłaszcza inżynierii powierzchni koncentrujących się na doborze odpowiedniego stopu stanowiącego superpozycję właściwości przeciwdrobnoustrojowych, odporności na korozję i właściwości wytrzymałościowych. W szczególności z wdrożeniem powierzchni z materiałów na bazie miedzi związany jest: problem utleniania czyli korozji miedzi i jej stopów wywołanej kontaktem z potem rąk ludzkich oraz z powietrzem atmosferycznym o zróżnicowanym składzie chemicznym; zagadnienie stosowanych w polskich placówkach opieki zdrowotnej środków dezynfekcyjnych i związany z nimi problem odporności korozyjnej stopów miedzi pozostających w kontakcie z różnymi substancjami czynnymi zawartymi w tych preparatach; temat alergennego uwalniania niklu ze stopów miedzi zawierających ten pierwiastek z uwagi na niebezpieczeństwo wywoływania uczuleń skórnych na skutek kontaktu z powierzchniami dotykowymi; proces obróbki mechanicznej materiałów na osnowie miedzi jako symulację

rysowania powierzchni metalu pod wpływem długoczasowej eksploatacji, głównie pod kątem jej wpływu na skuteczność przeciwbakteryjną oraz problem kształtowania zwilżalności miedzi i jej roli w mikrobiologicznych testach efektywności przeciwbakteryjnej. W niniejszej pracy przedstawiono pierwsze ze wspomnianych zagadnień.

2. Tematyka korozji miedzi w zastosowaniach na przeciwdrobnoustrojowe powierzchnie dotykowe

Zagadnienie utleniania miedzi w zastosowaniach na dotykowe wyroby przeciwbakteryjne można analizować zarówno w aspekcie kontaktu metalu z powietrzem lub atmosferą tlenu, jak i z potem rąk ludzkich, a także mogącymi imitować go mieszaninami wody morskiej. Pierwsze zagadnienie współcześnie dobrze i szeroko opisane w literaturze [12, 13] związane jest z powstawaniem warstewek tlenkowych (Cu_2O, CuO) jedno- lub dwufazowych, zależnych od temperatury i czasu trwanie procesu utleniania, co opisuje tzw. prawo paraboliczne [14]. Jeśli dodatkowo uwzględni się obecność zanieczyszczeń atmosferycznych siarką i chlorem (np. w środowisku oddziału szpitalnego) to sekwencja tworzenia związków chemicznych w produktach korozji miedzi obejmuje dużo bardziej złożone związki siarczanów, chlorków, czy też kwasów karboksylowych, co zilustrowano na rysunku 1.



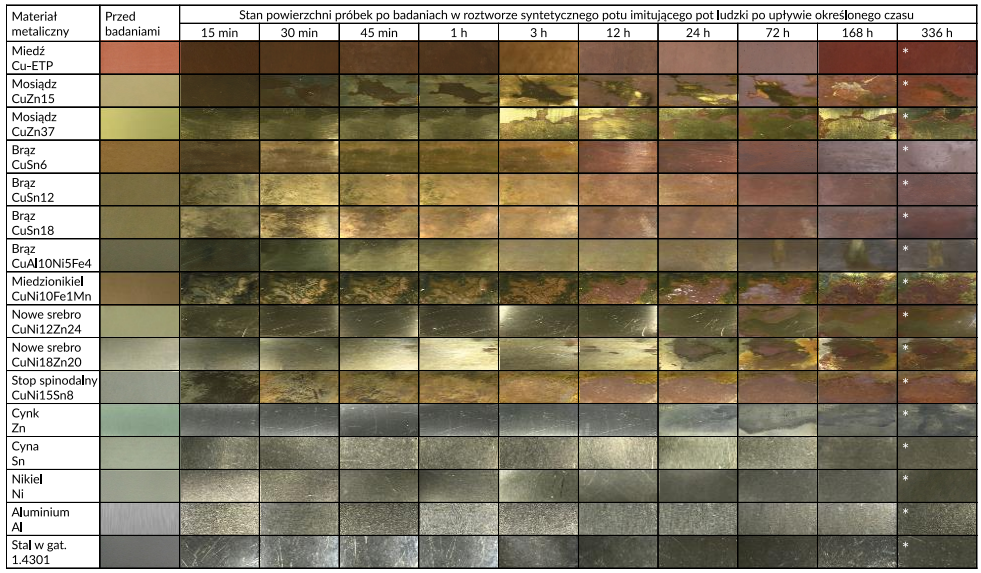
Rysunek 1. Wpływ czasu i obecności w powietrzu siarki i chloru na tworzenie związków chemicznych jako produktów korozji miedzi i jej stopów [15]

Sam mechanizm narastania warstewek korozyjnych w środowisku wilgotnym, mówiąc potocznie patynowanie, obejmuje rozpuszczanie i wytrącanie związków na bazie miedzi lub jej dodatków, o czym szeroko raportują prace naukowe [16, 17]. Z kolei zastosowanie miedzi i jej stopów na elementy maszyn i urządzeń pracujących w kontakcie z wodą morską, w tym m.in. wymienników ciepła może także stanowić w kontekście miedzi przeciwdrobnoustrojowej ciekawy punkt odniesienia, mimo że składem chemicznym i obecnością drobnoustrojów powodujących biokorozję woda morska różni się od potu

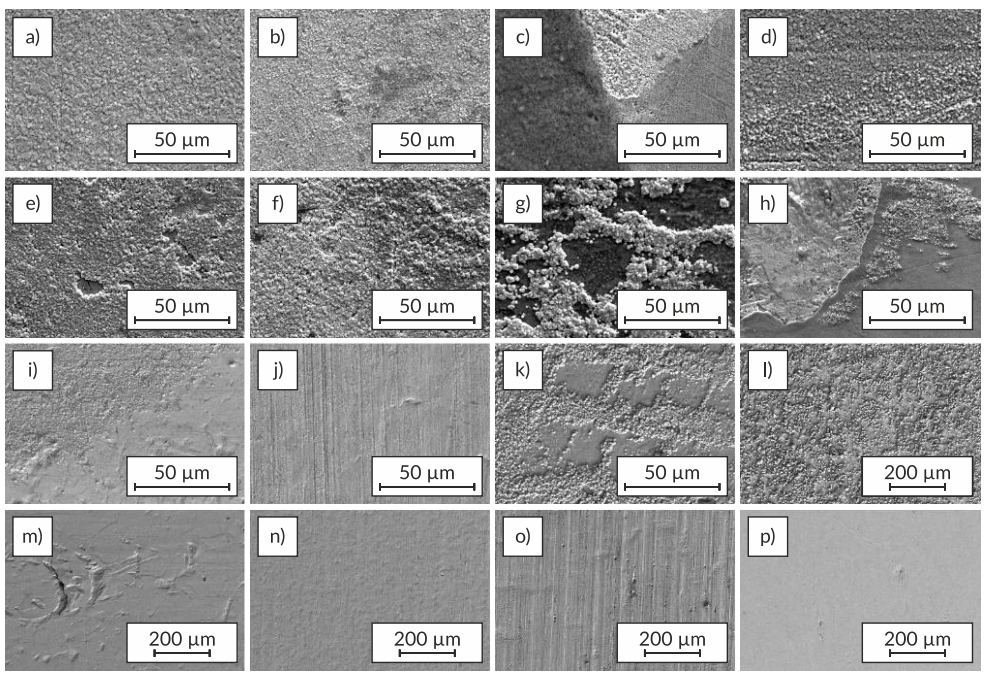
rąk ludzkich. W tego typu instalacjach wodnych warstwy korozyjne przyczyniając się wręcz do perforacji rur utworzone są przez tlenochlorek miedzi $CuCl_2 \cdot 3Cu(OH)_2$, wodorotlenek miedzi $Cu(OH)_2$ lub węglan miedzi(II) $CuCO_3$ [18]. Nasuwa się w związku z tym pytanie: co powstaje w wyniku kontaktu miedzi z potem rąk ludzkich? Odpowiedź w kontekście rozstrzygnięcia czy utlenianie powierzchni miedzi i jej stopów wpływa na skuteczność przeciwbakteryjną powierzchni stanowiła meritum badań własnych nad przedmiotowym zagadnieniem.

Rozważania rozpoczęto od poddania wytworzonych w warunkach przemysłowych lub laboratoryjnych próbek blach o kształcie kwadratu i boku 25 mm z miedzi wysokiej czystości chemicznej (Cu-ETP; 99,9 %) i jej najpopularniejszych tj. handlowych/komercyjnych stopów (mosiądze $CuZn_{15}$ i $CuZn_{37}$; brązy cynowe $CuSn_6$, $CuSn_{12}$ i $CuSn_{18}$; brąz aluminiowy $CuAl_{10}Ni_5Fe_4$; miedzionikiel $CuNi_{10}Fe_1Mn$; mosiądze wysokoniklowe tzw. nowe srebra $CuNi_{12}Zn_{24}$ i $CuNi_{18}Zn_{20}$; stop spinodalny $CuNi_{15}Sn_8$) oraz w celach porównawczych z innych metali (cynk, cyna, nikiel, aluminium, stal) kontaktowi z roztworem syntetycznego potu imitującego pot ludzki. W tym celu wykorzystano metodykę opisaną w normie EN 1811 [19], na podstawie której sporządzono bazowy rozwór składający się z mocznika CH_4N_2O, chlorku sodu NaCl, kwasu mlekowego $C_3H_6O_3$ i wodorotlenku sodu NaOH, w którym zanurzano indywidualnie materiały na czas od 15 minut do 336 godzin tj. 14 dni w temperaturze 30°C. Na rysunku 2 ukazano stan powierzchni próbek po krótko- i długoczasowych testach zanurzeniowych.

Na podstawie analizy wzrokowej powierzchni miedzi i jej stopów stwierdzono, że wszystkie materiały uległy reakcjom ze składnikami sztucznego potu i pokryły się produktami korozji. Największą odporność wykazał mosiądz $CuZn_{37}$, na powierzchni którego ogniska korozji wystąpiły tylko w niektórych miejscach i który niezmiennie zachował połysk metaliczny przez cały czas trwania testu. Powierzchnia stopów wysokomiedziowych stała się wizualnie zbliżona do czystej miedzi. Także stopy o wysokiej zawartości niklu i innych dodatków takich jak m.in. aluminium, czy cynk które uznawane są powszechnie za bardziej odporne na korozję, po teście stały się matowe i całkowicie pokryły osadami. Inne niż na osnowie miedzi materiały metaliczne wykazały w krótkoczasowych testach odporność korozyjną, ale pod wpływem wydłużonego czasu badania uległy zmatowieniu powierzchni. Dodatkowo badania morfologii powierzchni materiałów wykonane na skaningowym mikroskopie elektronowym wskazały na dużą chropowatość warstwy wierzchniej o znacznej porowatości (rys. 3), z kolei mikroanaliza rentgenowska (EDS) na obecność tlenków miedzi, chlorków miedzi i głównych dodatków stopowych w wykształtowanych warstwach. Na podstawie analizy wzrokowej powierzchni miedzi i jej stopów stwierdzono, że wszystkie materiały uległy reakcjom ze składnikami sztucznego potu i pokryły się produktami korozji.



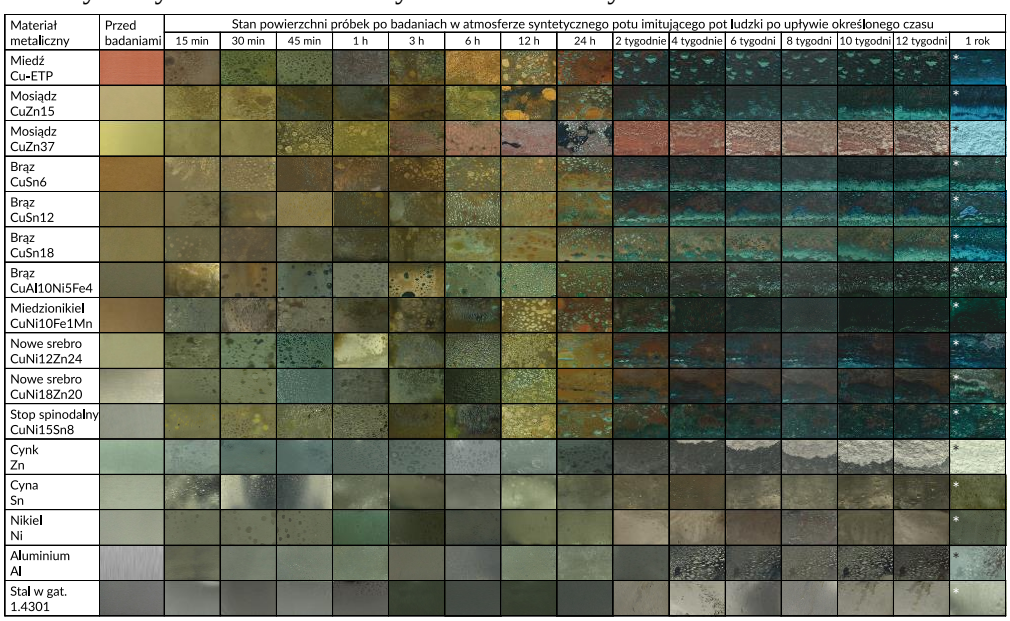
Rysunek 2. Powierzchnia próbek metali po testach zanurzeniowych w roztworze syntetycznego potu sporządzonym według normy EN 1811. Na rysunku symbol „*" oznacza próbki poddane badaniom strukturalnym na mikroskopie skaningowym



Rysunek 3. Mikroanaliza powierzchni próbek blach po trwających 2 tygodnie testach zanurzeniowych w roztworze syntetycznego potu sporządzonym wg normy EN 1811 dla: a) Cu-ETP b) $CuZn_{15}$, c) $CuZn_{37}$, d) $CuSn_6$, e) $CuSn_{12}$, f) $CuSn_{18}$, g) $CuAl_{10}Ni_5Fe_4$, h) $CuNi_{10}Fe_1Mn$, i) $CuNi_{18}Zn_{20}$, j) $CuNi_{12}Zn_{24}$, k) $CuNi_{15}Sn_8$, skaningowy mikroskop elektronowy, powiększenie × 1000; l) Zn, m) Sn, n) Ni, o) Al (gat. 1050A), p) X5CrNi18-10 (gat. 1.4301), skaningowy mikroskop elektronowy, powiększenie × 150

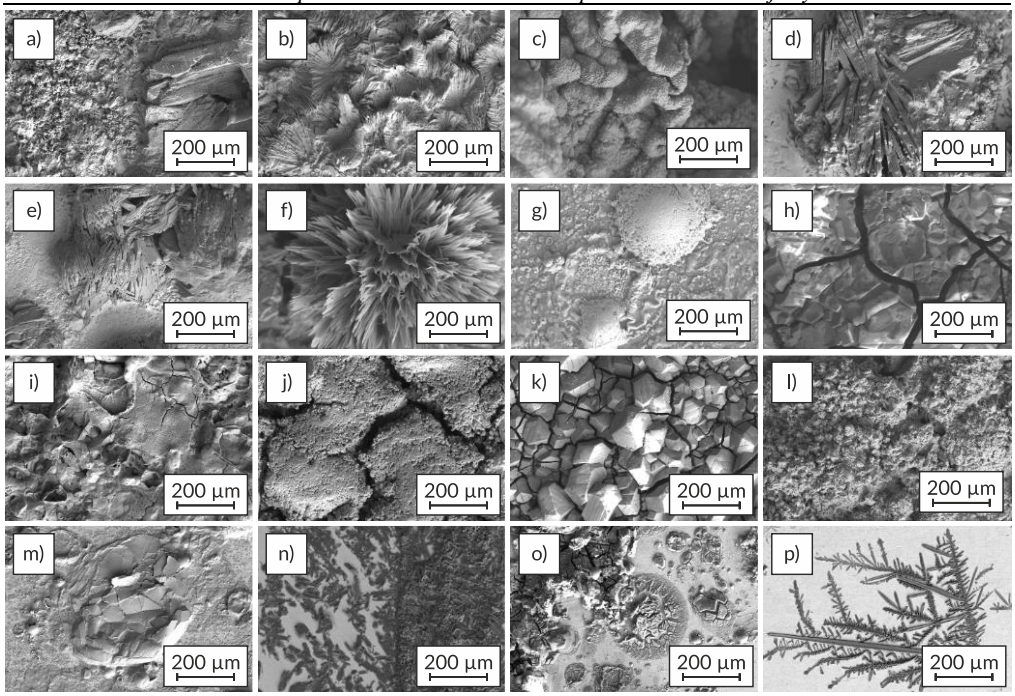
Największą odporność wykazał mosiądz $CuZn_{37}$, na powierzchni którego ogniska korozji wystąpiły tylko w niektórych miejscach i który niezmiennie zachował połysk metaliczny przez cały czas trwania testu. Powierzchnia stopów wysokomiedziowych stała się wizualnie zbliżona do czystej miedzi. Także stopy o wysokiej zawartości niklu i innych dodatków takich jak m.in. aluminium, czy cynk które uznawane są powszechnie za bardziej odporne na korozję, po teście stały się matowe i całkowicie pokryły osadami. Inne niż na osnowie miedzi materiały metaliczne wykazały w krótkoczasowych testach odporność korozyjną, ale pod wpływem wydłużonego czasu badania uległy zmatowieniu powierzchni. Dodatkowo badania morfologii powierzchni materiałów wykonane na skaningowym mikroskopie elektronowym wskazały na dużą chropowatość warstwy wierzchniej o znacznej porowatości (rys. 3), z kolei mikroanaliza rentgenowska (EDS) na obecność tlenków miedzi, chlorków miedzi i głównych dodatków stopowych w wykształtowanych warstwach.

Dla porównania badania korozyjne przeprowadzono także w atmosferze sztucznego potu, przy czym test opierał się na metodyce normy ISO 3160 [20], której bazowy roztwór oprócz wspomnianego wcześniej mocznika i kwasu mlekowego zawierał także chlorek amonu NH_4Cl i kwas octowy CH_3COOH. Temperatura badań wynosiła 40°C a maksymalny czas testu 1 rok. Wyniki ukazano na rysunku 4.



Rysunek 4. Powierzchnia próbek metali po testach korozyjnych w atmosferze syntetycznego potu sporządzonego według normy ISO 3160. Na rysunku symbol „*" oznacza próbki poddane badaniom strukturalnym na mikroskopie skaningowym

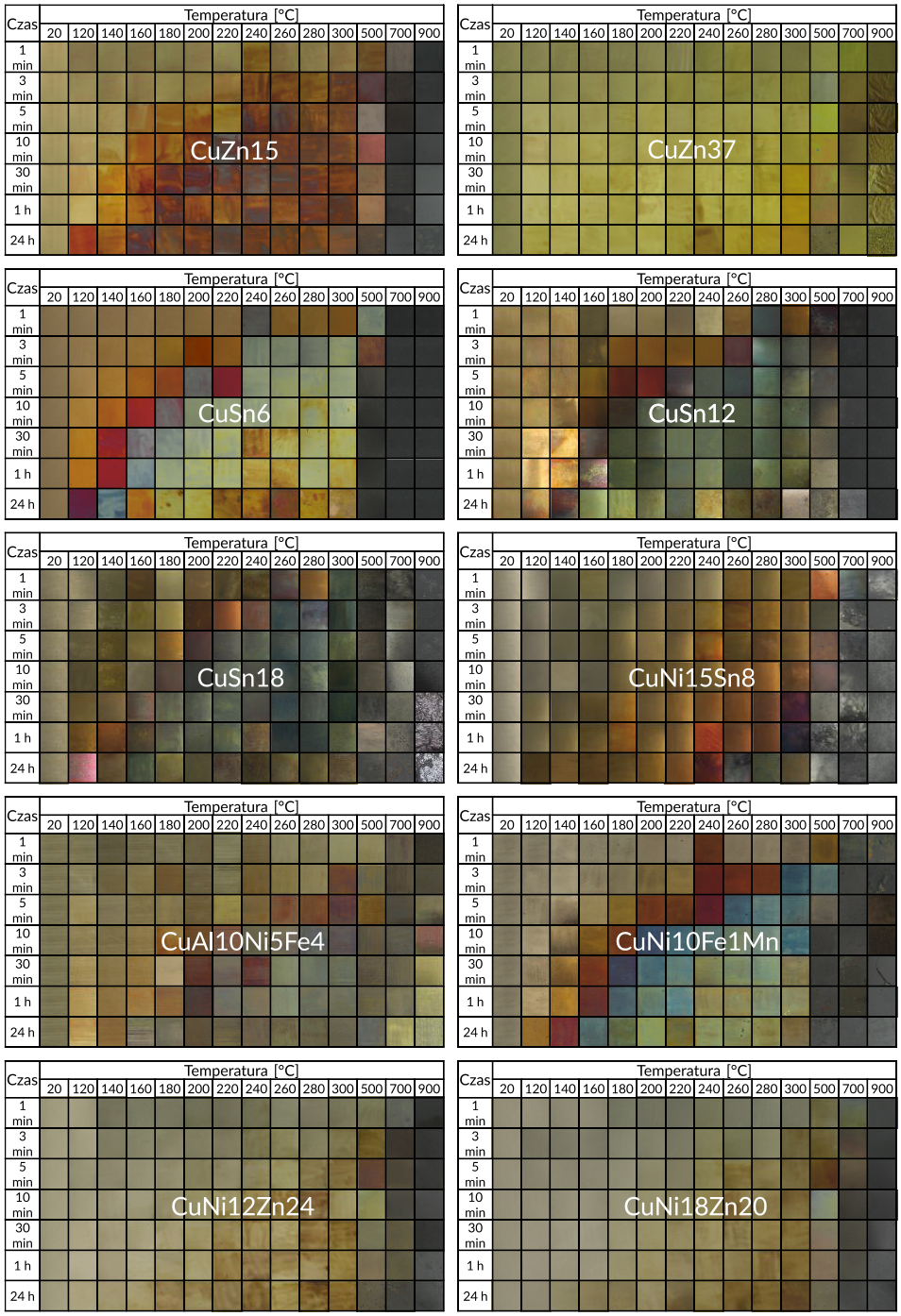
Analizując stan powierzchni materiałów uwagę zwraca dużo większa niż w badaniach zanurzeniowych według normy EN 1811 ilość produktów korozji, świadcząc tym samym o dużo bardziej inwazyjnych warunkach testu prowadzonego w atmosferze sztucznego potu. Potwierdzenie stanowiły także ukazane na rysunku 5 wyniki mikrostruktury warstwy wierzchniej, zgrubiałe i porowate utworzone przez związki na bazie tlenków i chlorków.



Rysunek 5. Mikroanaliza powierzchni próbek blach po trwających 1 rok testach korozyjnych w atmosferze syntetycznego potu sporządzonym wg normy ISO 3160 dla: a) Cu-ETP b) $CuZn_{15}$, c) $CuZn_{37}$, d) $CuSn_6$, e) $CuSn_{12}$, f) $CuSn_{18}$, g) $CuAl_{10}Ni_5Fe_4$, h) $CuNi_{10}Fe_1Mn$, i) $CuNi_{18}Zn_{20}$, j) $CuNi_{12}Zn_{24}$, k) $CuNi_{15}Sn_8$, skaningowy mikroskop elektronowy, powiększenie × 1000; l) Zn, m) Sn, n) Ni, o) Al (gat. 1050A), p) X5CrNi18-10 (gat. 1.4301), skaningowy mikroskop elektronowy, powiększenie × 150

Aby precyzyjnie określić zdolności przeciwbakteryjne utlenionej miedzi i jej stopów konieczne jest dokładne wyznaczenie warunków powstawania warstw korozyjnych i dokonanie analizy ich składu chemicznego. Jak wskazują dotychczasowe badania nad tym zagadnieniem, w wyniku oddziaływania potu rąk ludzkich z wyrobami z miedzi i jej stopów tworzy się na ich powierzchni warstewka tlenku miedzi(I) Cu_2O, która po dwóch latach eksploatacji w rzeczywistych warunkach ma grubość od około 50 nm do 250 nm [21]. Bazując na tej wiedzy w ramach badań własnych próbki wyrobów płaskich walcowanych z miedzi i jej stopów (materiały metaliczne takie jak wymienione wcześniej) zostały poddane procesowi utleniania w atmosferze powietrza z wykorzystaniem podwyższonych temperatur w zakresie od 120°C do 900°C i w czasie od 1 minuty do 24 godzin. Proces utleniania miał na celu uzyskanie w warunkach laboratoryjnych warstw tlenkowych symulujących warstwy powstające w wyniku korozji potu rąk ludzkich w trakcie rzeczywistej eksploatacji tego typu wyrobów.

Na rysunku 6 ukazano wybrane zdjęcia powierzchni blach ze stopów miedzi poddanych obróbce cieplnej w atmosferze powietrza w temperaturach wyszczególnionych w górnej część ilustracji, i w odstępach czasowych ukazanych w kolumnie z lewej strony zdjęcia. Uwagę zwraca kolorystyka utworzonych tlenków, która obejmuje szeroką paletę barw.



Rysunek 6. Kolorystyka korozyjnych warstw tlenkowych na powierzchni stopów miedzi powstałych w procesie obróbki cieplnej w atmosferze powietrza w temperaturze 120-900 °C i czasie 1 min – 24 h

Wytworzone w ten sposób warstwy powierzchniowe zostały następnie poddane analizie jakościowej i ilościowej za pomocą metody katodowej redukcji. W badaniu każdy z utlenionych materiałów stanowił katodę w obwodzie prądowym elektrolizy, anodę stanowił drut platynowy, natomiast elektrolitem był siarczan sodu. Katodową redukcję warstw tlenkowych w temperaturze 25°C prowadzono przy stałej gęstości prądu. Rezultatem badań były krzywe chronopotencjometryczne uzyskane dla próbek miedzi utlenionych w 200°C, 300°C i 600°C w czasie 1 minuty, 30 minut i 45 minut. Stwierdzono, że w zależności od warunków temperaturowo-czasowych obróbki cieplnej na krzywej występują dwa skoki potencjału wynikające z obecności tlenku miedzi(II) CuO oraz tlenku miedzi(I) Cu_2O. Wybrane wyniki pomiarów kulometrycznych przedstawiono w tabeli.

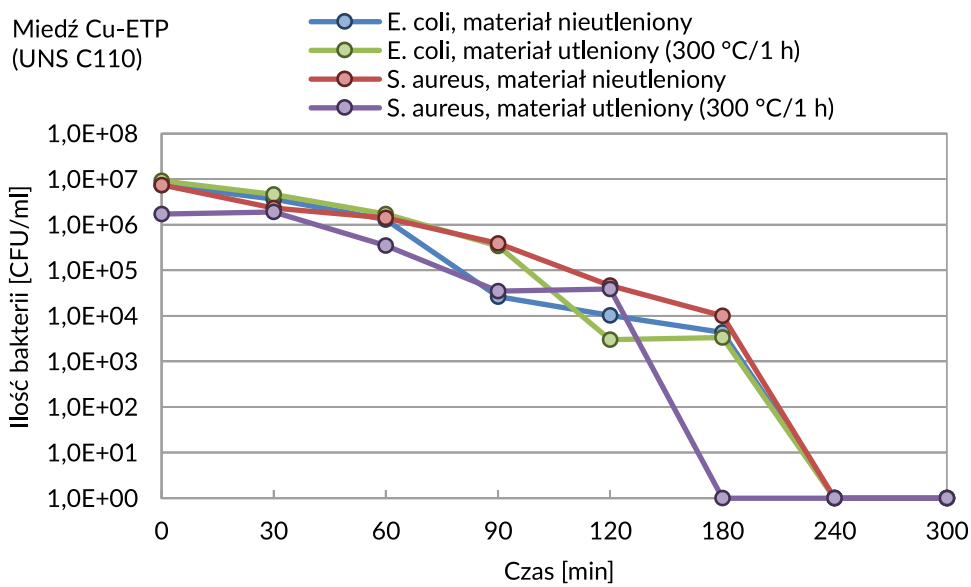
Tabela 1. Wyniki badań chronopotencjometrycznych metodą kulometryczną składu chemicznego i grubości warstw powierzchniowych na miedzi utlenionej w atmosferze powietrza w temperaturze 200-600°C i czasie 1-60 minut

Warunki utleniania		Gęstość prądu,	Czas katodowej redukcji związków		Grubość warstewki związku	
Temperatura	Czas		Cu_2O	CuO	Cu_2O	CuO
°C	min	mA/cm^2	s	s	nm	nm
200	1	0,50	80	-	40	-
		0,50	94	-	50	-
	5	0,50	177	-	110	-
		0,50	257	-	160	-
	30	0,50	376	13	230	4
		0,50	363	13	220	4
	60	1,00	215	3	270	2
250	5	0,50	648	18	400	6
		0,50	755	20	470	6
		0,50	927	28	570	9
	30	0,50	662	138	410	44
		0,50	784	176	480	57
	60	1,00	255	70	320	45
300	1	0,10	265	-	30	-
		0,25	360	-	110	-
	5	0,25	4665	285	1440	46
		0,50	1085	215	670	69
		0,50	1525	425	940	137
	30	0,50	3892	108	2400	35
	45	1,00	810	85	1000	55
		1,00	995	100	1230	64
400	1	0,50	250	-	160	-
		0,50	132	-	80	-
		0,50	192	-	120	-
600	1	1,00	1595	25	1970	16
		1,00	1595	25	1970	16
		1,00	1445	40	1780	16
		1,00	2220	30	2740	19

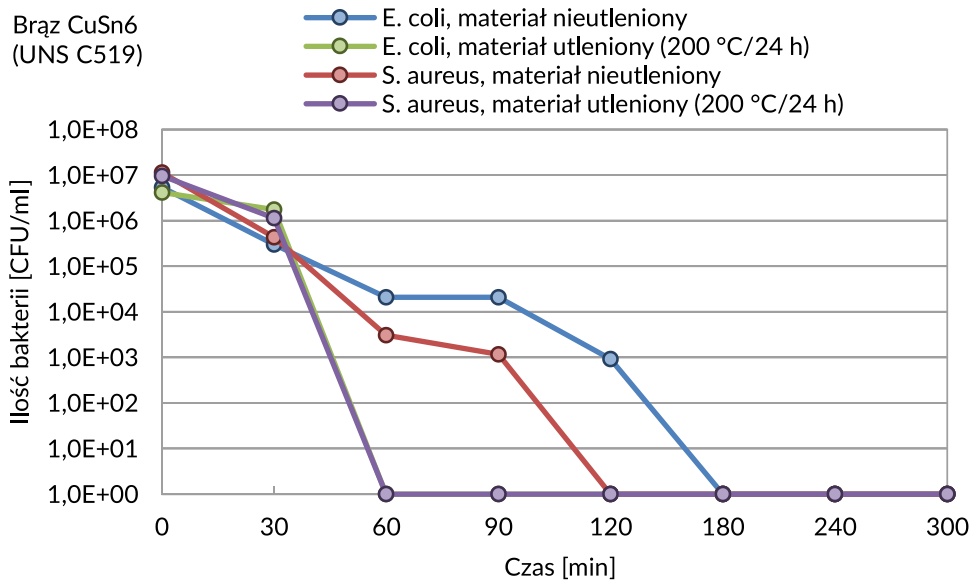
Źródło: [15].

Zaobserwowano wzrost całkowitej grubości warstewek tlenkowych na powierzchni miedzi od około 40 nm dla materiału utlenionego w temperaturze 200°C w krótkim czasie 1 minuty do ponad 2000 nm dla materiału obrobionego cieplnie w temperaturze 300°C w czasie 3 minut. Tym samym nawiązując do danych literaturowych [21] uzyskane zostały w badaniach chronopotencjometrycznych metodą kulometryczną powierzchniowe warstwy o składzie chemicznym i grubości odwzorowującej tlenki, które tworzą się na powierzchni miedzi w rzeczywistych warunkach eksploatacyjnych w wyniku kontaktu materiału metalicznego z potem rąk ludzkich.

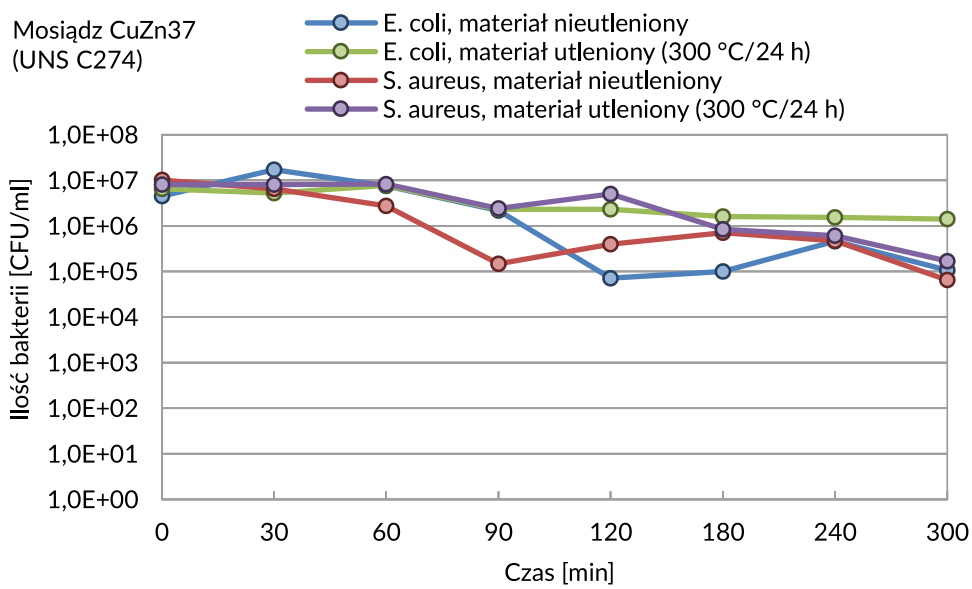
Tak utlenione próbki w kolejnym etapie procedury badawczej poddano mikrobiologicznym badaniom skuteczności przeciwdrobnoustrojowej realizowanym w Katedrze Mikrobiologii Uniwersytetu Jagiellońskiego Collegium Medicum. Rezultaty ukazano na rysunkach 7-10 w postaci wykresów zależności liczby bakterii od czasu, osobno dla miedzi Cu-ETP, brązu cynowego $CuSn_6$, mosiądzu $CuZn_{37}$ i nowego srebra $CuNi_{18}Zn_{20}$ przed i po procesie utleniania. Testy prowadzono z wykorzystaniem dwóch drobnoustrojów tj. pałeczki okrężnicy (łac. *Escherichia coli*) i gronkowca złocistego (łac. *Staphylococcus aureus*). Początkowa ilość kolonii około 10 milionów, która nanoszona została na powierzchnię każdej z próbek została zredukowana zarówno na próbkach utlenionych, jak i nieutlenianych. Zbliżony charakter krzywych dla wszystkich badanych materiałów daje podstawę do stwierdzenia, że utlenianie nie wpływa w istotny sposób na własności przeciwbakteryjne miedzi i jej stopów.



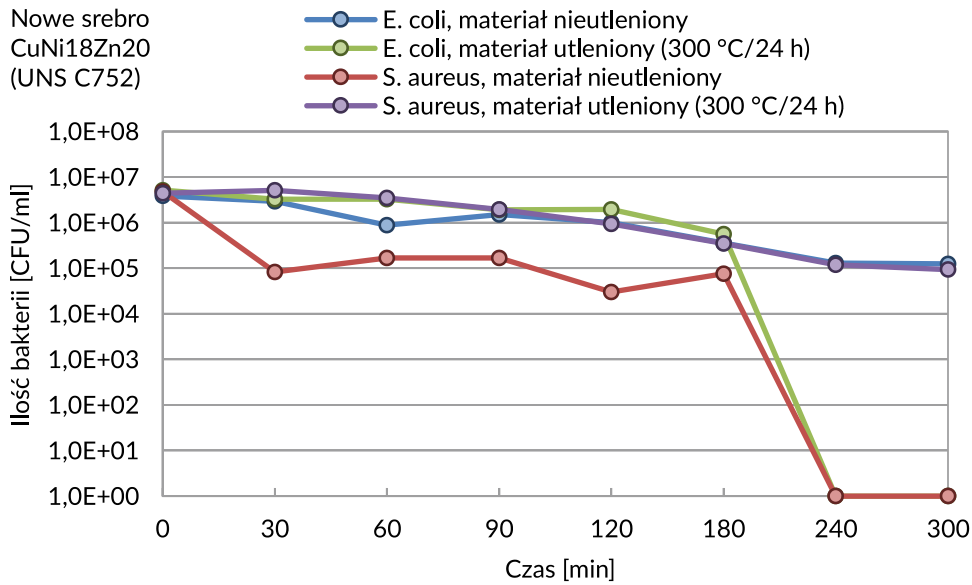
Rysunek 7. Wykresy zależności liczby bakterii pałeczki okrężnicy (łac. *Escherichia coli*) i gronkowca złocistego (łac. *Staphylococcus aureus*) od czasu dla powierzchni miedzi Cu-ETP przed i po procesie utleniania w atmosferze powietrza w podwyższonej temperaturze jako symulacja eksploatacji wyrobów z powierzchniami dotykowymi



Rysunek 8. Wykresy zależności liczby bakterii pałeczki okrężnicy (łac. *Escherichia coli*) i gronkowca złocistego (łac. *Staphylococcus aureus*) od czasu dla powierzchni mosiądzu CuZn37 przed i po procesie utleniania w atmosferze powietrza w podwyższonej temperaturze jako symulacja eksploatacji wyrobów z powierzchniami dotykowymi



Rysunek 9. Wykresy zależności liczby bakterii pałeczki okrężnicy (łac. *Escherichia coli*) i gronkowca złocistego (łac. *Staphylococcus aureus*) od czasu dla powierzchni brązu $CuSn_6$ przed i po procesie utleniania w atmosferze powietrza w podwyższonej temperaturze jako symulacja eksploatacji wyrobów z powierzchniami dotykowymi



Rysunek 10. Wykresy zależności liczby bakterii pałeczki okrężnicy (łac. *Escherichia coli*) i gronkowca złocistego (łac. *Staphylococcus aureus*) od czasu dla powierzchni mosiądzu wysokoniklowego $CuNi_{18}Zn_{20}$ przed i po procesie utleniania w atmosferze powietrza w podwyższonej temperaturze jako symulacja eksploatacji wyrobów z powierzchniami dotykowymi

Dla oceny skuteczności przeciwbakteryjnej przyjęto kryteria zastosowane przez Souli [22], zgodnie z którymi redukcja gęstości zawiesiny w granicach od 2 do 3 logarytmów oznacza właściwości bakteriostatyczne, a redukcja powyżej 3 log – właściwości bakteriobójcze. Uzupełnienie testów mikrobiologicznych stanowiły badania z wykorzystaniem mikroskopu fluorescencyjnego polegające na różnicowym barwieniu komórek bakteryjnych pozwalając tym samym zaobserwować martwe i żywe drobnoustroje. W wyniku tych badań dla materiałów na osnowie miedzi stwierdzono nie stwierdzono obecności żywych drobnoustrojów, z kolei na materiale referencyjnym (porównawczym), który stanowiła stal zaobserwowano koncentrację żywych bakterii.

3. Podsumowanie

Miedź i stopy miedzi swoje właściwości przeciwdrobnoustrojowe zawdzięczają jonom Cu^{+} i Cu^{2+}. Aby jony te wydzieliły się na powierzchni metalu musi zajść proces utleniania, czyli korozji. Jego obecność powoduje zmatowienie warstwy wierzchniej, a tym samym pogorszenie ogólnych walorów wizualnych i estetyki powierzchni miedzi. Jak wykazano w niniejszej pracy wytworzenie w wyniku utleniania warstewki korozyjne nie wpływają negatywnie tj. nie osłabiają aktywności przeciwdrobnoustrojowej metalu. Biorąc pod uwagę powyższe, w kwestii doboru materiału metalicznego na osnowie miedzi na powierzchnie dotykowe rozwiązaniem są stopy z dodatkami pierwiastków podwyższających odporność korozyjną. W tym przypadku mowa o takich dodatkach stopowych jak m.in. aluminium czy cynk. Tworzą one na powierzchni metalu szczelną warstewkę tlenkową, odpowiednio tlenku aluminium Al_2O_3 i tlenku cynku ZnO ograniczając w ten sposób dostęp tlenu do metalu bazowego (miedzi). Na podstawie przedstawionych w pracy wyników badań stwierdzono, że materiały tego typu wykazują w kontakcie z potem rąk

ludzkich wyższą aniżeli czysta miedź odporność korozyjną, zachowując przy tym zadawalający poziom skuteczności przeciwbakteryjnej w kontakcie z drobnoustrojami kolonizującymi powierzchnie dotykowe wyrobów w miejscach użyteczności publicznej. Dodając do tego fakt, iż materiały te muszą być podatne do przetwórstwa plastycznego, formowania i obróbki wykańczającej wyrobu gotowego, a także powszechnie dostępne/komercyjne i względnie atrakcyjnie ekonomicznie, to taką rolę z powodzeniem spełnia mosiądz $CuZn_{37}$. Jest materiałem o wysokim poziomie własności mechanicznych i wykazuje odporność korozyjną w kontakcie ze składnikami syntetycznego potu. Dodatkowo półwyroby hutnicze z tego stopu są wytwarzane w warunkach krajowych zakładów metalurgiczno-przetwórczych z branży metali nieżelaznych.

Uwagi ogólne/Podziękowania

Projekt został sfinansowany ze środków Narodowego Centrum Badań i Rozwoju: umowa nr PBS3/A9/32/2015.

Literatura

1. Chang R.Y.K., Nang S.C., Chan H.K., Li J., *Novel antimicrobial agents for combating antibiotic-resistant bacteria*, Advanced Drug Delivery Reviews, 187, 2022, s. 114378.
2. Monteiro A., Cardoso J., Guerra N., Ribeiro E., Viegas C., Verde S., Sousa-Uva A., *Exposure and health effects of bacteria in healthcare units: an overview*, Applied Sciences, 12, 2022, s. 1958.
3. Jasovský D., Littmann J., Zorzet A., Cars O., *Antimicrobial resistance – a threat to the world's sustainable development*, Upsala Journal of Medical Sciences, 121, 2016, s. 159-164.
4. de Kraker M.E., Stewardson A.J., Harbarth S., *Will 10 million people die a year due to antimicrobial resistance by 2050?* PLOS Medicine, 13, 2016, s. e1002184.
5. Wernli D., Jorgensen P.S., Harbarth S., *Antimicrobial resistance: the complex challenge of measurement to inform policy and the public*, PLOS Medicine, 14, 2017, s. e1002378.
6. Prestinaci F., Pezzotti P., Pantosti A., *Antimicrobial resistance: a global multifaceted phenomenon*, Pathogens and Global Health, 109, 2015, s. 309-318.
7. Grosse S.D., Lollar D.J., Campbell V.A., Chamie M., *Disability and disability-adjusted life years: not the same*, Public Health Reports, 124, 2009, s. 197-202.
8. Cassini A., Högberg L.D., Plachouras D., Quattrocchi A., Hoxha A., Simonsen G.S., Colomb-Cotinat M., Kretzschmar M.E., Devleesschauwer B., Cecchini M., Ouakrim D.A., Oliveira T.C., Struelens M.J., Suetens C., Monnet D.L., *Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis*, The Lancet Infectious Diseases, 19, 2019, s. 56-66.
9. Garus-Pakowska A., Sobala W., Szatko F., *Observance of hand washing procedures performed by the medical personnel before patient contact. Part I*, International Journal of Occupational Medicine and Environmental Health, 26, 2013, s. 113-121.
10. Garus-Pakowska A., Sobala W., Szatko F., *Observance of hand washing procedures performed by the medical personnel after patient contact. Part II*, International Journal of Occupational Medicine and Environmental Health, 26, 2013, s. 257-264.
11. Dziennik Urzędowy Ministra Zdrowia, Obwieszczenie Ministra Zdrowia z dnia 28 października 2015 r., w sprawie standardów akredytacyjnych w zakresie udzielania świadczeń zdrowotnych oraz funkcjonowania podmiotów leczniczych wykonujących inwazyjne procedury zabiegowe i operacyjne.
12. Choudhary S., Sarma J.V.N., Pande S., Ababou-Girard S., Turban P., Lepine B., Gangopadhyay S., *Oxidation mechanism of thin Cu films: A gateway towards the formation of single oxide phase*. AIP Advances, 8, 2018, 8, s. 055114.

13. Aromaa J., Kekkonen M., Mousapour M., Jokilaakso A., Lundström M., *The oxidation of copper in air at temperatures up to 100 °C*, Corrosion and Materials Degradation, 2, 2021, s. 625-640
14. Itoh M., *Time-dependent power laws in the oxidation and corrosion of metals and alloys*, Scientific Reports, 12, 2022, s. 1-9.
15. Walkowicz M., *Miedź przeciwdrobnoustrojowa. Materiały – Powierzchnie dotykowe – Aplikacje*, Oficyna wydawnicza Impuls, Kraków 2018, s. 97.
16. Leygraf C., Chang T., Herting G., Wallinder I.O., *The origin and evolution of copper patina colour*, Corrosion Science, 157, 2019, s. 337-346.
17. FitzGerald K.P., Nairn J., Skennerton G., Atrens A., *Atmospheric corrosion of copper and the colour, structure and composition of natural patinas on copper*, Corrosion Science, 48, 2006, s. 2480-2509.
18. Edwards M., Ferguson J.F., Reiber S.H., *The pitting corrosion of copper*, Journal AWWA, 86, 1994, s. 74-90
19. PN-EN 1811 1811+A1:2015-09, Zalecana metoda badania uwalniania się niklu z części wyrobów umieszczanych w przekłuwanych miejscach ludzkiego ciała oraz produktów przeznaczonych do bezpośredniego i długotrwałego kontaktu ze skórą.
20. PN-ISO 3160-1:2000 Koperty zegarków i akcesoria -- Powłoki ze stopów złota -- Wymagania ogólne.
21. Fredj N., Kolar J., Prichard D., Burleigh T., *Study of relative color stability and corrosion resistance of commercial copper alloys exposed to hand contact and synthetic hand sweat, Corrosion Science*, 2013, 76, s. 415-423.
22. Souli M., Galani I., Plachouras D., Panagea T., Armaganidis A., Petrikkos G., Giamarellou H., *Antimicrobial activity of copper surfaces against carbapen-emase-producing contemporary gram-negative clinical isolates*, Journal of Antimicrobial Chemotherapy, 68, 2013, s. 852–857.

Rozwiązania materiałowe i technologiczne w obszarze utleniania powierzchni dotykowych z miedzi i stopów miedzi o właściwościach przeciwdrobnoustrojowych

Streszczenie

Praca zawiera kompendium wybranych zagadnień materiałowych i technologicznych z obszaru utleniania miedzi i stopów miedzi o immanentnych cechach przeciwdrobnoustrojowych na powierzchnie dotykowe wyrobów przeznaczonych do miejsc użyteczności publicznej, w szczególności jednostek opieki zdrowotnej. Materiały tego typu o dowiedzionej w mikrobiologicznych badaniach laboratoryjnych i klinicznych skuteczności w eliminacji ze swoich powierzchni drobnoustrojów chorobotwórczych, w tym wielolekoopornych patogenów alarmowych stanowią współcześnie wsparcie tradycyjnych metod dezynfekcji, czyli przeciwdziałania rozprzestrzeniania bakterii, wirusów i grzybów w miejscach dużych skupisk ludzkich i/lub środowisku o silnej koncentracji zakażeń (np. szpitalnym). Problematyka ta z uwagi na czas pandemii koronawirusa jest wciąż aktualna i obejmuje szereg aspektów tematycznych bezpośrednio i pośrednio związanych z implementacją do szpitali, przychodni, szkół, lotnisk i innych obiektów o ponadstandardowej przepustowości, częstotliwości i trybie użytkowania materiałów na osnowie miedzi przeciwdrobnoustrojowej w postaci wyrobów z powierzchniami dotykowymi. W szczególności, tak jak przedstawiono w niniejszej publikacji, dotyczą one zagadnień materiałowych, technicznych i technologicznych w aspekcie doboru stopu miedzi o odpowiednich właściwościach mechanicznych i chemicznych (odporność na korozję) oraz podatności do przetwórstwa plastycznego, formowania i obróbki wykańczającej wyrobu gotowego. Nie sposób nie wspomnieć o silnym akcencie mikrobiologicznym przedstawionej tematyki w zakresie doboru metody oceny efektywności metalu w eliminacji drobnoustrojów (kontekst zdrowia publicznego) oraz wydźwięku społecznym tj. możliwości przeciwdziałania zakażeniom rozprzestrzenianym drogą kontaktową w obiektach użyteczności publicznej. W pracy przedstawiono analizę literaturową oraz wyniki badań własnych nad problemem korozji pod wpływem eksploatacji tj. kontaktu z potem rąk ludzkich komercyjnych/handlowych stopów miedzi.

Słowa kluczowe: stopy miedzi, inżynieria powierzchni, korozja, utlenianie, właściwości przeciwdrobnoustrojowe

Material and technological challenges of oxidation of touch surfaces from antimicrobial copper and its alloys

Abstract

The work contains a compendium of selected material and technological issues in the area of oxidation of copper and copper alloys with inherent antimicrobial properties on the touch surfaces of products intended for public places, in particular healthcare units. Materials of this type, proven in microbiological laboratory and clinical tests to be effective in eliminating pathogenic microorganisms from their surfaces, including multi-drug resistant alarm pathogens, nowadays support traditional methods of disinfection, i.e. preventing the spread of bacteria, viruses and fungi in places of large human concentrations and/or in an environment with high concentration of infections (e.g. hospital). Due to the time of the coronavirus pandemic, this issue is still valid and covers a number of thematic aspects directly and indirectly related to the implementation of antimicrobial copper-based materials in the form of touch surfaces. In particular, as presented in this publication, they concern material, technical and technological issues in terms of the selection of a copper alloy with appropriate mechanical and chemical properties (corrosion resistance) and susceptibility to plastic processing, forming and finishing of the finished product. It is impossible not to mention the strong microbiological accent of the presented topic in terms of the selection of the method for assessing the effectiveness of metal in eliminating microorganisms (public health context) and the social overtones, i.e. the possibility of preventing infections spread by contact in public facilities. The paper presents a literature analysis and the results of own research on the problem of corrosion under the influence of exploitation, i.e. contact with the sweat of human hands of commercial copper alloys.

Keywords: copper alloys, surface engineering, corrosion, oxidation, antimicrobial properties

Szymon Senyk[1], Tadeusz Kałdoński[2]

Właściwości tribologiczne smarów plastycznych zawierających heksagonalny azotek boru o różnej granulacji

1. Wprowadzenie

Smary plastyczne są jedną z grup środków smarnych stosowanych w systemach tribologicznych. Wykorzystywane są, gdy substancja smarująca nie może być doprowadzana w sposób ciągły do smarowanego węzła tarcia lub tam, gdzie olej smarny nie może być utrzymywany w wystarczającej ilości. Pozwalają na dokładne uszczelnienie układów trących przed dostępem wody i zanieczyszczeń. Zapewniają także dobrą ochronę przeciwkorozyjną powierzchni metalowych. Stosowane są do smarowania elementów działających w wysokich temperaturach, przy dużych naciskach jak i zmiennych obciążeniach. Wśród głównych obszarów zastosowań smarów plastycznych wymienia się łożyska toczne i ślizgowe. Istnieją także smary plastyczne przeznaczone do smarowania lin, prowadnic, przegubów czy też przekładni otwartych wykorzystywanych w przemyśle motoryzacyjnym, górnictwie i hutnictwie [1-5]. Smary do zastosowań w łożyskach stanowią około 65% całkowitego asortymentu tej grupy środków smarnych [3].

W literaturze i dokumentach normatywnych można odnaleźć wiele definicji smaru plastycznego. Wszystkie z nich są jednak spójne pod względem treści, dlatego przyjąć można, że smary plastyczne to stałe lub półpłynne układy koloidalne składające się z oleju bazowego i środka zagęszczającego. Cząsteczki oleju ulokowane są w przestrzennej siatce stworzonej przez cząsteczki zagęszczacza. W wyniku oddziaływania sił kapilarnych i adsorpcji składniki smaru plastycznego tworzą jednolitą strukturę. Ponadto smary plastyczne zawierają dodatki, kształtujące ich określone właściwości. Udział objętościowy oleju bazowego wynosi zwykle 70÷96%, a środka zagęszczającego 4÷30%. Ilość dodatków w smarze plastycznym wynosi maksymalnie 10% [1]. Ustalenie proporcji składników smaru plastycznego jest koncepcją jego wytwórcy, która powinna być ukierunkowana na planowe zastosowanie produktu i związane z tym niezbędne jego właściwości.

W trudnych warunkach eksploatacji układów tribologicznych (ekstremalnie niska lub wysoka temperatura, ekstremalne naciski kontaktowe, warunki próżniowe, obecność promieniowania) smary plastyczne nie są w stanie wytworzyć odpowiednio trwałego filmu smarnego. Aktualne wyzwania w badaniach nad smarami plastycznymi przedstawione zostały przez autora publikacji [6]. Obejmują one m.in. opracowanie smarów, które zapewnią dłuższą trwałość eksploatacyjną smarowanych elementów i będą pracować w trudniejszych warunkach (ekstremalnie niskie i wysokie temperatury oraz prędkości) oraz optymalizację ich przepływu w węzłach tarcia. Uzasadnione są, wobec tego badania naukowe w zakresie optymalizowania składu smarów plastycznych, a w konsekwencji zmiany ich określonych właściwości.

[1] szymon.senyk@wat.edu.pl, Szkoła Doktorska/Wydział Inżynierii Mechanicznej, Wojskowa Akademia Techniczna, https://www.wojsko-polskie.pl/wat/.
[2] tadeusz.kaldonski@wat.edu.pl, Wydział Inżynierii Mechanicznej, Wojskowa Akademia Techniczna, https://www.wojsko-polskie.pl/wat/.

2. Dodatki kształtujące właściwości tribologiczne smarów plastycznych

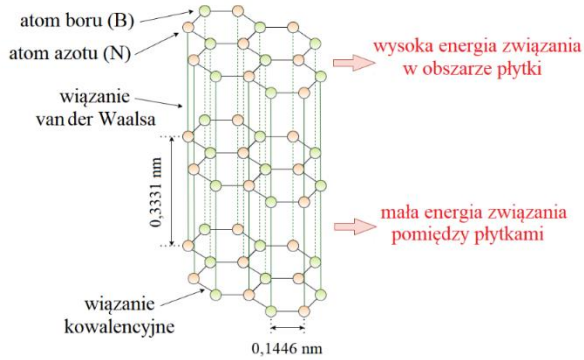
2.1. Heksagonalny azotek boru

Jedną z grup dodatków szeroko stosowanych w praktyce naukowej i produkcyjnej w celu poprawy właściwości smarów plastycznych są smary stałe. Za główny cel ich stosowania Krawiec uznaje *rozszerzanie obszaru parametrów wymuszeń (nacisku, prędkości poślizgu, temperatury) zapewniających węzłowi ślizgowemu pracę przy tarciu normalnym, z równoczesną minimalizacją występujących sił tarcia i zużycia* [2]. W ramach smarów stałych stosowanych w obszarze tribologii wyróżnić należy związki o budowie warstwowej (lamelarnej). Niektóre z nich takie jak dwusiarczek molibdenu (MoS_2), grafit czy kwas borowy (H_3BO_3) pochodzą ze źródeł naturalnych minerałów. Inne powstają natomiast w wyniku przemian chemicznych i są to np. dwusiarczek wolframu (WS_2), diselenki, fluorowany grafit czy będący przedmiotem rozważań w tym artykule heksagonalny azotek boru (h-BN) [7].

Heksagonalny azotek boru charakteryzuje się wysoką stabilnością termochemiczną (temperatura topnienia w pobliżu 2600ºC) oraz oksydacyjną, przez co zachowuje swoje właściwości smarujące w wysokich temperaturach. Utrzymuje on niską wartość współczynnika tarcia w temperaturach do 900ºC, podczas gdy inne podobne związki jak dwusiarczek molibdenu i grafit tracą swoje właściwości smarujące w niższych temperaturach. Ta odmiana azotku boru wykazuje dobrą przewodność cieplną przy braku przewodności elektrycznej oraz niską reaktywność chemiczną, co stanowi jej przewagę nad grafitem. Izolacyjny charakter h-BN wynika z faktu, że w przestrzeniach międzywarstwowych występują pary elektronowe, podczas gdy np. w graficie są to niesparowane elektrony. Heksagonalny azotek boru jest materiałem chemicznie obojętnym wobec większości substancji nawet w wysokich temperaturach. Odznacza się doskonałymi właściwościami elektroizolacyjnymi, także w próżni oraz wysoką wytrzymałością dielektryczną [7-15].

Zastosowanie heksagonalnego azotku boru jako dodatku do olejów smarnych i smarów plastycznych poza wymienionymi właściwościami uwarunkowane jest jego budową strukturalną (rys. 1). W każdej z warstw ułożone naprzemiennie atomy boru i azotu połączone są orbitalem sp^2 i tworzą silne wiązanie kowalencyjne typu s. Długość wiązania pomiędzy atomami obu tych pierwiastków wynosi 0,1446 nm. Odległość między sąsiadującymi warstwami to 0,3331 nm, a wiązanie między nimi ma charakter słabych międzycząsteczkowych sił van der Waalsa o energii około 16,7 kJ/mol. Słabe oddziaływania między poszczególnymi warstwami h-BN świadczą o jego anizotropowości i implikują mały opór tarcia wewnętrznego. W ruchu ślizgowym warstwy atomowe układają się równolegle do jego kierunku, a następnie z łatwością ścinają, zapewniając niskie opory tarcia [7-15]. Heksagonalny azotek boru wykazuje, wobec tego naturalne właściwości smarujące, będące pochodną jego struktury krystalicznej, co Hebda i Wachal [16] ujęli w ramy strukturalnej teorii tarcia smarów stałych.

Właściwości fizykochemiczne tego materiału są podobne do grafitu, przez co może być stosowany w wielu podobnych ślizgowych węzłach tarcia, w warunkach wysokiej temperatury i ekstremalnych nacisków [7]. Potencjał tribologiczny tej odmiany azotku boru w porównaniu do dwusiarczku molibdenu i grafitu związany jest w szczególności z wysoką stabilnością i czystością środowiska pracy (biały kolor proszku h-BN) [10]. Zaletą heksagonalnego azotku boru jest także jego nietoksyczność i mała szkodliwość dla środowiska naturalnego [9].



Rysunek 1. Graficzne zobrazowanie struktury krystalicznej heksagonalnego azotku boru, opracowanie własne na podstawie [17]

2.2. Nanododatki do smarów plastycznych

Nanotechnologia jest jedną z najdynamiczniej rozwijających się dziedzin nauki. Z tego względu w ostatnich latach badacze zajmujący się tribologią zwrócili swoją uwagę ku potencjałowi nanocząstek, możliwych do wykorzystania także jako dodatki do olejów smarnych i smarów plastycznych.

Za najbardziej charakterystyczną cechę nanomateriałów uznaje się dużą powierzchnię granic rozdziału. W przypadku nanoproszków, do których zalicza się heksagonalny azotek boru, jest to powierzchnia zewnętrza. Powoduje ona silną reaktywność chemiczną oraz skłonność do aglomeracji. Skłonność do tworzenia aglomeratów i agregatów oprócz rodzaju materiału i jego morfologii zależy także od ośrodka jego ulokowania [18]. Tendencja związana ze stosowaniem nanododatków do smarów plastycznych, wynika nie tylko z dynamicznego rozwoju nanotechnologii i związanej z tym dobrej koniunktury, ale przede wszystkim z przesłanek, pozwalających sądzić, iż tego typu materiały mogą wpływać pozytywnie na właściwości tribologiczne smarów plastycznych. Właściwości tribologiczne nanomateriałów, a tym samym możliwe mechanizmy ich oddziaływania determinowane są przez ich skład, kształt i wielkość cząstek, krystaliczność (defekty struktury), stężenie oraz stabilność dyspersji w środkach smarnych [19].

Mechanizmy smarowania nanocząstkami, umożliwiające zmniejszenie tarcia i zużycia elementów współpracujących dzielone są na dwie grupy. W pierwszej z nich wyróżnia się procesy związane z bezpośrednim oddziaływaniem nanocząstek w strefie tarcia (efekt łożyska kulkowego oraz tworzenie filmu ochronnego). Druga grupa odnosi się do wpływu nanocząstek na powierzchnie trące (efekt naprawy i polerowania powierzchni) [20, 21]. Działanie nanocząstek w ramach tych mechanizmów jest następujące [19-22]:

- efekt łożyska kulkowego (ang. *rolling effect*) – cząstki o sferycznym lub quasi-sferycznym kształcie zachowują się jak małe łożyska kulkowe, które toczą się pomiędzy powierzchniami trącymi. Jest to związane ze zmianą tarcia ślizgowego na mieszankę tarcia ślizgowego i tocznego. Ten mechanizm smarowania zachodzi w systemach tribologicznych, które pracują w warunkach małego obciążenia, gdyż wówczas możliwe jest zachowanie kształtu i sztywności nanocząstek;
- tworzenie warstwy ochronnej – na smarowanych powierzchniach trących tworzy się warstwa zapewniająca ich ochronę przed tarciem oraz zużyciem. Nanocząstki wyka-

zują łatwość w tworzeniu takich warstw poprzez adsorpcję lub reakcje chemiczne z uwagi na dużą powierzchnię właściwą i dużą energię powierzchniową;

- efekt naprawczy – charakteryzuje się osadzaniem nanocząstek zawartych w środku smarnym w miejscach ubytku materiału z powierzchni trących;
- efekt polerowania/wygładzania (ang. *polishing/smoothing effect*) – występuje wówczas, gdy chropowatość smarowanej powierzchni jest zmniejszana przez ścieranie wspomagane nanocząstkami. W stykach tarciowych nanocząstki mogą wypełniać ubytki powierzchniowe, które mogą pełnić rolę zasobników nanocząstek w strefie styku.

Należy jednak zaznaczyć, że nie każdy z wymienionych efektów może mieć miejsce w przypadku nanocząstek heksagonalnego azotku boru odziaływujących w strefie tarcia. Największe wątpliwości budzi w tym przypadku efekt łożyska kulkowego, głównie z uwagi na fakt, że forma heksagonalna jest miękką odmianą azotku boru, przez co efekt toczenia jest trudny do uzyskania.

W ostatnich latach w literaturze tematu odnaleźć można wiele publikacji z zakresu smarów plastycznych zawierających nanododatki. Podsumowanie tego obszaru badań w monografii z roku 2019 przedstawili Rawat i Harsha [4], przytaczając publikacje, których autorzy stosowali jako nanododatki do smarów plastycznych m.in. nanostruktury węglowe, grafen, zredukowany tlenek grafenu, dwusiarczek molibdenu, grafit, politetrafluoroetylen, węglan i boran wapnia, fosforan cyrkonu, metale (miedź, cyna), tlenki metali (CuO, TiO_2, Al_2O_3, CeO_2), fluorki (CaF_2, CeF_3), nanokompozyty, a także popiół z łusek ryżowych wzbogacony kwarcem. Zastosowanie wymienionych materiałów wpływało na poprawę właściwości tribologicznych bazowych smarów plastycznych. Z tego względu rozpatrywanie nanocząstek heksagonalnego azotku boru jako dodatku do smarów plastycznych również wydaje się uzasadnione.

2.3. Właściwości tribologiczne smarów plastycznych zawierających h-BN

Pojęcie właściwości tribologicznych smarów plastycznych jest nierozerwalnie związane z realizowanym w układzie tribologicznym procesem smarowania, umożliwiającym zaistnienie, w miejsce tarcia zewnętrznego ciał stałych, tarcia wewnętrznego w substancji smarującej. Zachowanie się substancji smarującej w konkretnym węźle tribologicznym, szczególnie w warunkach tarcia granicznego, określa się mianem smarności. Determinuje ona zdolność takiej substancji do wytworzenia na danej powierzchni trwałej warstwy (filmu) granicznej. Taka zdolność zależna jest między innymi od właściwości substancji stosowanej do smarowania określonego węzła tarcia, a tym samym od jej składu.

Wśród wielu publikacji na temat dodatków stosowanych do modyfikowania właściwości tribologicznych smarów plastycznych, zaledwie kilkanaście dotyczy zastosowania w tym celu heksagonalnego azotku boru.

Przedmiotem opublikowanych w tym zakresie rozważań naukowych był najczęściej wpływ różnej zawartości heksagonalnego azotku boru na właściwości tribologiczne smaru plastycznego [8, 23-37]. Stężenia dodatku nie przekraczały zwykle 10%, aczkolwiek wspomnieć należy pracę [34], w której zastosowana była większa ilość h-BN równa 30%. Rzadkością w analizowanym zbiorze literatury światowej były prace skupiające uwagę na wpływie granulacji heksagonalnego azotku boru na charakterystyki tribologiczne smarów plastycznych [31, 38], a jedyna praca uwzględniająca w tym względzie nanoodmianę tego smaru stałego została opublikowana w [38]. Dotyczyła ona jednego

stężenia heksagonalnego azotku boru, podobnie jak inne prace odnoszące się do mikro h-BN [39-41]. W większości publikacji jako bazę stosowano smar litowy [8, 23, 24, 30-32, 35, 38-40], aczkolwiek można napotkać również badania ze smarami wapniowymi [37], polimocznikowymi [33, 42], opracowanymi w oparciu o polialfaolefiny i sebacynian dioktylu [29] oraz kompozycjami cytrynianu acetylotributylu i octanomaślanu celulozy [34]. Opisane w literaturze światowej badania tribologiczne w omawianym zakresie najczęściej realizowane były na aparacie czterokulowym.

W pracach opublikowanych w ostatnich latach przedmiotem badań były nanocząstki heksagonalnego azotku boru. Okazały się one skutecznym dodatkiem poprawiającym właściwości tribologiczne smarów plastycznych, do których je wprowadzono [11, 32, 33, 35, 38]. Dotychczas tylko jedna publikacja [38] dotyczyła porównania wpływu nano- i mikrocząstek heksagonalnego azotku boru na właściwości tribologiczne smaru plastycznego. Jej autorzy dowiedli, że smar z zawartością nanocząstek h-BN wykazywał najlepsze właściwości tribologiczne. Podobne analizy przeprowadzone zostały dla związków pokrewnych względem h-BN, tj. grafitu [43] i dwusiarczku molibdenu [44]. Wnioski płynące z treści obu artykułów były podobne – nanocząstki wprowadzone do smaru plastycznego okazywały się skuteczniejsze niż cząstki o większych rozmiarach.

Autorzy publikacji [34] wykazali, że wszystkie zastosowane przez nich smary stałe (grafit, h-BN, politetrafluoroetylen) miały tendencję do aglomerowania i tworzenia większych struktur przestrzennych. Takie zachowanie cząstek może prowadzić do pogorszenia właściwości tribologicznych modyfikowanych smarów i znacznej zmiany ich reologii. Problem aglomerowania cząstek może szczególnie uwidaczniać się w przypadku nanomateriałów, dlatego też zwykle tego typu smary stałe stosowane są w mniejszych stężeniach, zdaniem autorów publikacji [38], nieprzekraczających 4%.

Na podstawie wyników prac opublikowanych przez różnych autorów należy przyjąć, że heksagonalny azotek boru pozytywnie wpływa na właściwości smarnościowe smarów plastycznych pod warunkiem doboru odpowiednich parametrów tego dodatku. Nieliczne przypadki braku poprawy właściwości tribologicznych smarów plastycznych po dodaniu do nich heksagonalnego azotku mogły być wynikiem złego doboru bazowego środka smarnego, nieodpowiednio dobranego stężenia bądź granulacji dodatku.

Właściwości tribologiczne smarów plastycznych, do których wprowadzono heksagonalny azotek boru zależały od rozmiaru cząstek dodatku, ich stężenia oraz rodzaju bazowego środka smarnego. Warto podkreślić, że rekomendowanym parametrem opisu zbioru cząstek był rozkład granulometryczny, a nie średni rozmiar cząstek, gdyż w przypadku materiałów o dużym rozrzucie rozmiarów pojedynczych elementów materii, wynik uśredniony nie odzwierciedla rzeczywistej granulacji próbki.

Tribochemiczne oddziaływanie heksagonalnego azotku boru wprowadzonego do smarów plastycznych opierało się w głównej mierze na tworzeniu warstw ochronnych na smarowanych powierzchniach współpracujących. Mogły one powstawać na zasadzie osadzania lub adsorbowania cząstek h-BN na smarowanych powierzchniach. Aktywność przeciwtarciowa i przeciwzużyciowa takich warstw związana była z budową strukturalną tego materiału, przekładającą się na łatwość poślizgu międzywarstwowego i w konsekwencji niskie opory tarcia. Łatwość w tworzeniu warstw ochronnych na smarowanych podłożach przez smary zawierające nanocząstki była pochodną ich rozmiarów i dużej powierzchni właściwej.

Po przeprowadzonej analizie literatury tematu dostrzeżono potrzebę rozpatrzenia wpływu różnych odmian heksagonalnego azotku boru na właściwości tribologiczne różnych smarów plastycznych. Tak obszerne badania są rzadko spotykane w literaturze. Zdaniem autorów są one niezbędnym etapem rozważań na temat zastosowania smarów plastycznych z dodatkiem heksagonalnego azotku boru w rzeczywistych układach tribologicznych. Konieczne jest ponadto rozpoznanie właściwości fizyko-chemicznych stosowanego dodatku oraz szczegółowa analiza prowadzonych badań tribologicznych. Identyfikacja morfologii cząstek, ich rozmiaru oraz powierzchni właściwej i struktury porowatej stanowi punkt wyjścia do rozważań na temat ich oddziaływania z substancją smarującą oraz interakcji w systemach tribologicznych.

3. Cel pracy

Celem pracy była ocena wpływu rozkładu granulometrycznego i stężenia heksagonalnego azotku boru na właściwości tribologiczne litowego i wapniowego smaru plastycznego. Badania parametrów smarnościowych zrealizowano na aparacie czterokulowym. Przeprowadzono również pomiary właściwości fizykochemicznych poszczególnych odmian h-BN. Ich rozpoznanie umożliwiło dokładną ocenę wpływu cech tych dodatków na właściwości tribologiczne bazowych smarów plastycznych.

4. Metodyka i obiekty badań

4.1. Bazowe smary plastyczne

Do badań wybrano dwa smary bazowe: z zagęszczaczem litowym oraz wapniowym. Takie rodzaje smarów plastycznych są najczęściej wykorzystywane w technice smarowniczej – ich łączny udział w asortymencie produkcyjnym szacowany jest na 80% [45]. Smary bazowe nie zawierały dodatków uszlachetniających.

Bazowy smar litowy stanowił kompozycję oleju mineralnego zagęszczanego mydłem prostym w postaci kwasu 12-hydroksystearynowego. Bazowy smar wapniowy opracowany był w oparciu o olej mineralny zagęszczany prostym mydłem wapniowym stanowiącym produkt reakcji mieszaniny kwasów tłuszczowych z wodorotlenkiem wapnia.

Bazową próbkę smaru litowego oznaczono w niniejszej pracy jako L0, a bazową próbkę smaru wapniowego jako W0. Próbki z zawartością heksagonalnego azotku boru opracowane na bazie smaru litowego (L) i wapniowego (W) oznaczano według następującego schematu: rodzaj smaru (L lub W) stężenie dodatku – rodzaj dodatku np. L 1-1 oznacza smar litowy z 1% zawartością h-BN 1.

4.2. Próbki heksagonalnego azotku boru i procedura przygotowania kompozycji smarnych

W badaniach wykorzystano cztery rodzaje heksagonalnego azotku boru. Dodatki wprowadzane były do bazowych smarów plastycznych w stężeniach masowych wynoszących 1%, 3%, 5% oraz 10%. Wybór stężeń był podyktowany między innymi wynikami analiz przeprowadzonych dotychczas w Wojskowej Akademii Technicznej w Warszawie [8, 46]. Także w innych pracach przytoczonych w analizie literatury tematu najczęściej stosowane były stężenia h-BN nieprzekraczające 10%. Ponadto w związku z doniesieniami [38] o aglomerowaniu nanocząstek, obserwowanym często po przekroczeniu ich 4% zawartości w smarach plastycznych, postanowiono rozpatrzyć zarówno mniejsze jak i większe od tej wartości stężenia.

Rysunek 2. Schemat procedury przygotowania smarów plastycznych z zawartością heksagonalnego azotku boru, opracowanie własne

Procedura przygotowania próbek smarów plastycznych z zawartością heksagonalnego azotku boru (rys. 2) rozpoczynana była od wstępnego, ręcznego wymieszania odpowiednio odmierzonych składników kompozycji. Następnie smary poddawano mieszaniu z wykorzystaniem miksera z obrotową misą przy stałej prędkości obrotowej (pierwszy stopień mieszania w 5-stopniowej skali urządzenia). Ta część procedury trwała godzinę, a osadzony na ścianach misy smar plastyczny zbierano na bieżąco szpatułką. Po procesie mieszania kompozycje poddawano 30-minutowemu działaniu ultradźwięków. Podobne sposoby przygotowania próbek smarów plastycznych z dodatkiem różnych smarów stałych przy wykorzystaniu miksera przedstawiono w publikacjach [47-52]. Sonikację kompozycji zastosowano z kolei w opracowaniach [51, 53]. Ustalenia czasu trwania poszczególnych etapów przyjętej procedury dokonano empirycznie w oparciu o informacje wskazane w literaturze [47-53], kierując się przy tym kryterium jednorodności uzyskanej mieszaniny.

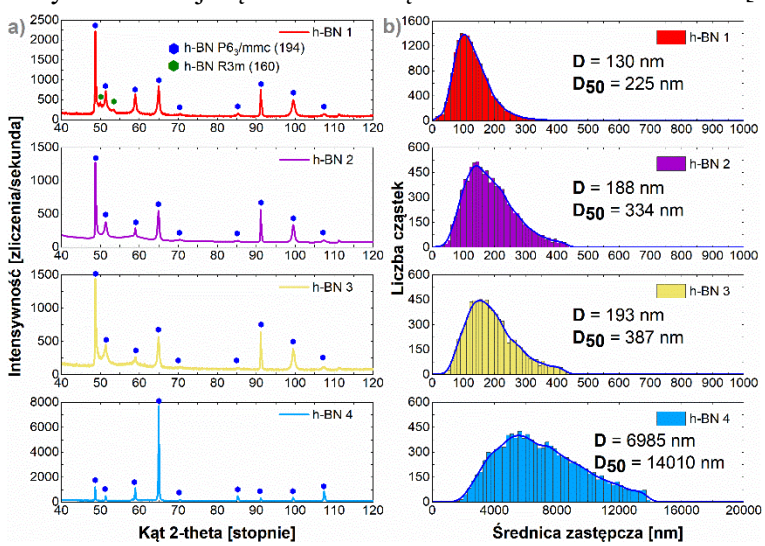
4.3. Identyfikacja właściwości próbek heksagonalnego azotku boru

Wobec chęci porównania wpływu nanocząstek i mikrocząstek heksagonalnego azotku boru na charakterystyki tribologiczne smarów plastycznych niezbędne wydawało się rozpoznanie właściwości fizykochemicznych zgromadzonych materiałów, istotnych z punktu widzenia procesu smarowania. Wśród takich parametrów wyróżnić należy morfologię i rozkład rozmiarów cząstek h-BN oraz strukturę porowatą tych cząstek opisywaną przez powierzchnię właściwą, całkowitą objętość porów i rozmiary tych porów. Takie parametry zostały oznaczone dla czterech próbek heksagonalnego azotku boru stosowanych w zrealizowanych badaniach. Przeprowadzono ponadto badania związane z określeniem struktury krystalicznej próbek.

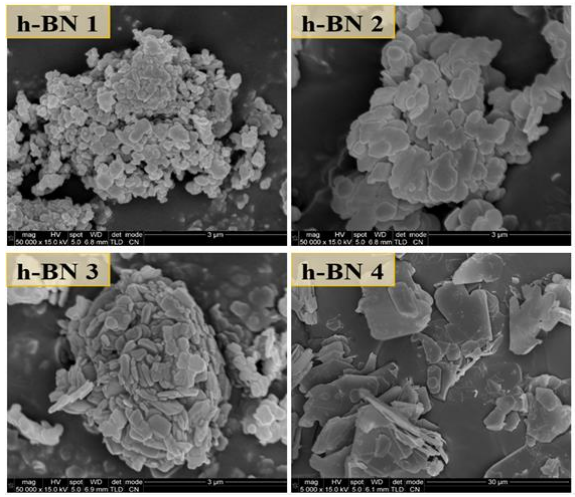
W celu identyfikacji i potwierdzenia heksagonalnej struktury krystalicznej czterech próbek stosowanego materiału przeprowadzono badania metodą dyfrakcji promieniowania rentgenowskiego (XRD, ang. *X-ray diffraction*) na urządzeniu Rigaku Ultima IV. Analizę wyników (rys. 3a) przeprowadzono w oparciu o uzyskane dyfraktogramy rentgenowskie z wykorzystaniem oprogramowania PDXL (Rigaku) oraz bazy danych PDF4+ 2021. Każda z próbek zawierała azotek boru o układzie heksagonalnym. W oparciu o dostępną bazę dyfraktogramów w próbce h-BN 1 zidentyfikowano dwa rodzaje sieci krystalicznej. W pozostałych próbkach wykryto wyłącznie typową dla tego typu materiału grupę $P6_3/mmc$. O różnej morfologii cząstek obecnych w poszczególnych próbkach heksago-

nalnego azotku boru świadczyły różnice w intensywności promieniowania wykrytego przez detektor przy poszczególnych wartościach kąta 2-theta [17].

Dodowem potwierdzającym powyższą tezę o różnej morfologii cząstek były wyniki pomiarów przeprowadzonych za pomocą skaningowego mikroskopu elektronowego (SEM, ang. *scanning electron microscope*) Nova NanoSEM 200. Na rysunku 4 przedstawiono obrazy mikroskopowe cząstek pochodzących z czterech próbek heksagonalnego azotku boru. Kształt tych cząstek uznano za nieizometryczny. Widoczne były także różnice w rozmiarze cząstek poszczególnych próbek h-BN, które szczegółowo określono wykorzystując metodę analizy obrazu. Dokonano w tym celu segmentacji obrazów na obiekty (odpowiadające obrysom ziaren) i wykonano analizę powierzchni tych obiektów [17]. Zgromadzone dane przetworzono w skrypcie [54], gdzie powierzchnię ziaren przeliczono na obiekty o kształcie koła i takiej samej powierzchni jak ówcześnie uzyskane. Dla każdego z takich zastępczych kół wyznaczono średnicę, a uzyskane wyniki przedstawiono w postaci histogramów (rys. 3b). Średnie wartości średnic takich kół dla każdej z próbek podano na rysunku jako D. Przyjmując, że średnica koła D jest średnicą kuli, w kolejnym kroku obliczono objętość każdej z takich kul i wyznaczono średnicę zastępczą D_{50}, dla której mniejsze i większe od niej ziarna stanowiły połowę całkowitej objętości danej próbki heksagonalnego azotku boru. Cząstki o najmniejszym rozmiarze wykryto w próbce h-BN 1, a znaczna część z nich wykazywała rozmiary typowe dla nanocząstek. Parametry opisujące rozmiary cząstek pochodzących z próbek h-BN 2 i h-BN 3 były zbliżone, natomiast zdecydowanie największe mikrocząstki odnotowano dla h-BN 4 [17].



Rysunek 3. Wyniki badań struktury krystalicznej i rozmiarów cząstek próbek heksagonalnego azotku boru: a) dyfraktogramy rentgenowskie, b) rozkłady rozmiarów cząstek [17]

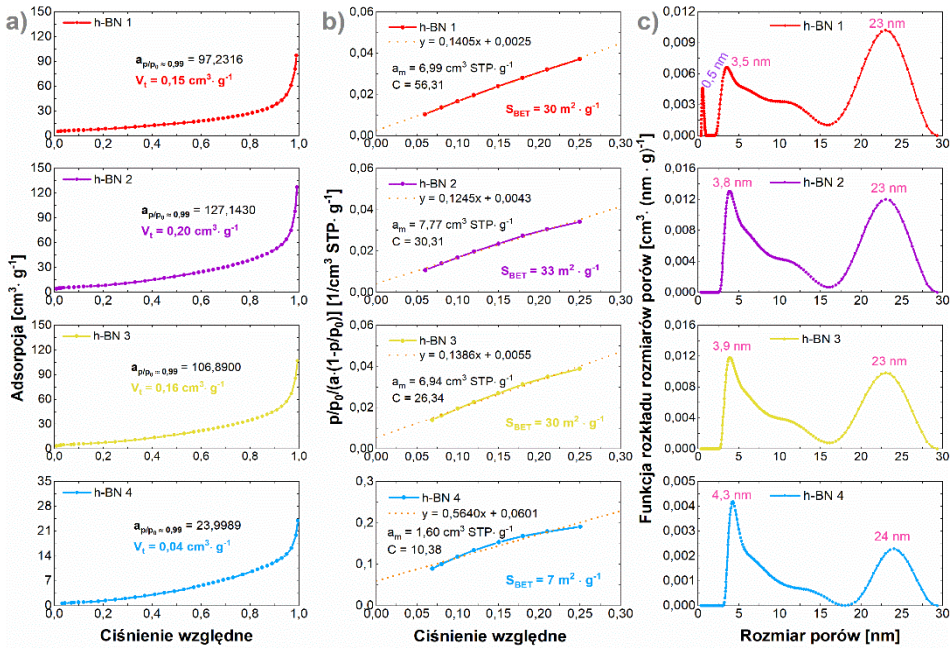


Rysunek 4. Obrazy cząstek z różnych próbek heksagonalnego azotek boru [17]

Właściwości adsorpcyjne materiałów są często charakteryzowane w oparciu o niskotemperaturowe izotermy adsorpcji azotu lub argonu. Uzyskanie takich izoterm pozwala w następstwie na określenie powierzchni właściwej, całkowitej objętości porów oraz rozkładu rozmiarów porów. Takie parametry określono dla próbek heksagonalnego azotku boru, stanowiących obiekty badań. Izotermy adsorpcji azotu wyznaczono w temperaturze -196°C za pomocą analizatora adsorpcyjnego ASAP 2020. Powierzchnię właściwą S_{BET} cząstek heksagonalnego azotku boru określono metodą BET, której nazwa jest akronimem stworzonym od nazwisk jej autorów (Brunauer, Emmett and Teller) [55]. Dla opisu doświadczalnych izoterm adsorpcji h-BN we współrzędnych równania BET przyjęto przedział ciśnień względnych 0÷0,25, w którym doświadczalne izotermy adsorpcji wykazywały niemal liniowy charakter. Całkowitą objętość porów V_t wyznaczono w odniesieniu do jednego punktu izotermy adsorpcji azotu odpowiadającego ciśnieniu względnemu $p \cdot p_o^{-1} \approx 0{,}99$ [56]. Funkcje rozkładu rozmiarów porów (PSD, ang. *pore size distribution*) wyznaczono z izoterm adsorpcji azotu metodą nielokalnej teorii funkcjonału gęstości (2D-NLDFT, ang. *2D – Non-local density functional theory*) dla porów w kształcie szczeliny węglowej przy założeniu energetycznej niejednorodności i geometrycznego pofałdowania ścian porów przedstawioną w pracy [57].

Wyniki pomiarów adsorpcyjnych przedstawiono na rysunku 5. Największą powierzchnię właściwą wynoszącą 33 $m^2 \cdot g^{-1}$ i całkowitą objętość porów wynoszącą 0,20 $cm^3 \cdot g^{-1}$ odnotowano dla próbki h-BN 2. Zbliżone wartości tych parametrów uzyskano dla próbek h-BN 1 i h-BN 3. Próbka h-BN 4 wykazała najmniejszą powierzchnię właściwą równą 7 $m^2 \cdot g^{-1}$, jak i najmniejszą całkowitą objętość porów równą 0,04 $cm^3 \cdot g^{-1}$. Wszystkie próbki heksagonalnego azotku boru wykazywały niejednorodne pory, głównie w zakresie mezoporów, których rozmiary wynosiły około 4 nm i 23 nm. Dodatkowo dla h-BN 1 zidentyfikowano pory o rozmiarach rzędu 0,5 nm, dlatego ten rodzaj heksagonalnego azotku boru uznano za materiał nanoporowaty [17].

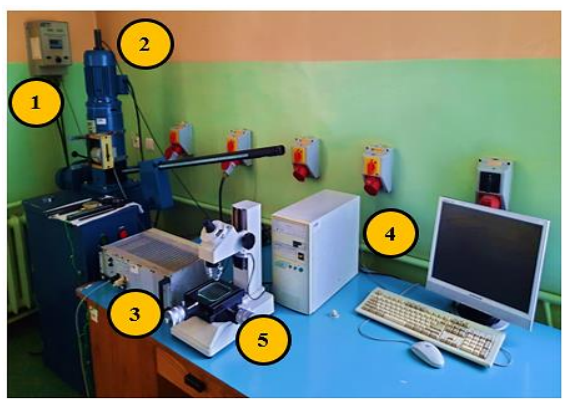
Wyniki badań właściwości fizykochemicznych heksagonalnego azotku boru wykorzystano w ocenie wpływu tego dodatku na właściwości tribologiczne smarów bazowych.



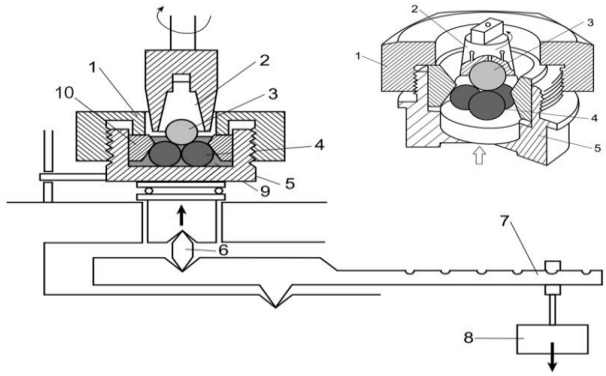
Rysunek 5. Wyniki pomiarów adsorpcyjnych próbek h-BN: a) doświadczalne niskotemperaturowe izotermy adsorpcji, b) izotermy adsorpcji we współrzędnych równania BET, c) rozkłady rozmiarów porów [17]

4.4. Metodyka badania właściwości tribologicznych smarów plastycznych

Urządzeniem najpowszechniej stosowanym do oceny właściwości tribologicznych środków smarnych jest aparat czterokulowy. Badania w ramach obecnej pracy zrealizowano na aparacie czterokulowym T-02 wyprodukowanym przez Instytut Technologii Eksploatacji w Radomiu. Widok stanowiska badawczego przedstawiono na rysunku 6, a schemat skojarzenia badawczego standardowego aparatu czterokulowego na rysunku 7.



Rysunek 6. Stanowisko do badań właściwości smarnościowych substancji smarujących: a) aparat czterokulowy T 02 wraz z oprzyrządowaniem: 1 – mikroprocesorowy regulator obrotów silników asynchronicznych SM – 303, 2 – aparat czterokulowy T-02, 3 – mikroprocesorowy sterownik tribologiczny, 4 – komputer sterujący, 5 – mikroskop pomiarowy [opracowanie własne]



Rysunek 7. Schemat węzła tarcia aparatu czterokulowego: 1 – pokrywa mocująca kule dolne, 2 – uchwyt kuli górnej, 3 – ruchoma kula górna, 4 – nieruchome kule dolne, 5 – naczynie z substancją smarującą, 6 – pryzmat, 7 – dźwignia, 8 – obciążenie, 9 – badana ciecz, 10 – pierścień mocujący [58]

Ocena właściwości smarnościowych środków smarnych na wspomnianym urządzeniu może być prowadzona w oparciu o zapisy wielu dokumentów normatywnych. Metodyka badań przyjęta na potrzeby badań opisanych w bieżącej pracy obejmowała zarówno parametry zaczerpnięte z norm przedmiotowych [59-61] jak i parametry nienormatywne. Dla każdego z oznaczanych parametrów wykonywano po trzy pomiary, przyjmując warunek, aby uzyskane wartości nie różniły się o więcej niż 10% od ich średniej, co stanowi norma [59].

Pierwszym z wybranych do oznaczenia parametrów było graniczne obciążenie zużycia G_{oz}, obliczane z normatywnej zależności [59]:

$$G_{oz} = 0{,}52 \cdot \frac{P}{d^2}, \left[\frac{N}{mm^2}\right], \quad (1)$$

gdzie: 0,52 – współczynnik, wynikający z rozkładu sił w węźle tarcia aparatu czterokulowego (czworościanu foremnego), P – obciążenie zadane [N], d – średnia średnica skaz na kulach [mm].

Parametr ten posłużył scharakteryzowaniu właściwości przeciwzużyciowych analizowanych smarów plastycznych. Wykonanie takiego oznaczenia zgodnie z polskimi normami [59, 60] polega na przeprowadzeniu biegu testowego przy obciążeniu 1471,5 N (lub innym będącym wielokrotnością 490,5 N), w czasie 60 ±1 s. Norma europejska [61] takie warunki badań określa jako próbę zużycia, pozostawiając badaczowi dowolność w doborze czasu trwania testu i obciążenia. Różnice w warunkach oznaczania granicznego obciążenia zużycia dotyczą prędkości obrotowej kuli górnej – według PN/C-04147 jest to 500 obr./min, a według PN/C-04362 jest to 1450÷1500 obr./min. Kierując się wynikami testów próbnych realizowanych przy większej prędkości obrotowej i przy zadeklarowanych obciążeniach, podczas których następowało nadmierne zużywanie kul testowych, a niekiedy nawet ich zespawanie, przyjęto pierwszą z podanych prędkości obrotowych. Wartości granicznego obciążenia zużycia G_{oz} wyznaczono przy trzech obciążeniach zadanych tj. 490,5 N, 981,0 N oraz 1471,5 N. Podczas wszystkich testów obserwowano i rejestrowano przebieg momentów tarcia.

Norma PN-76/C-04146 zawiera jeden unikalny względem pozostałych dokumentów parametr charakteryzujący właściwości przeciwzatarciowe środków smarnych, tj. obciążenie zacierające P_t. Testy wykonywane powinny być przy płynnie rosnącym z prędkością 408,8 N/s obciążeniu, gdy prędkość obrotowa górnej kuli wynosi 500 obr./min. Efektem

tego testu jest charakterystyka momentu tarcia w funkcji czasu, na podstawie której wskazywany jest moment, w którym doszło do przerwania filmu smarnego, skutkującego nagłym wzrostem oporów tarcia w węźle tarcia aparatu czterokulowego. Obciążenie zacierające P_t wyznacza się wykorzystując zależność:

$$P_t = t \cdot 408{,}8\ [N], \tag{2}$$

gdzie: t – czas, w którym nastąpił nagły wzrost oporów w węźle tarcia [s], 408,8 – wartość prędkości przyrastania obciążenia w czasie testu [N/s].

Z obciążeniem zacierającym P_t powiązany jest nienormatywny parametr określany jako graniczny nacisk zatarcia p_{oz}. Metodykę jego oznaczania wzorowano na pracach [62, 63], z tą różnicą, że dla każdego z badań przyjęto maksymalną wartość obciążenia P_{max}, jaką można było uzyskać na aparacie czterokulowym. Graniczny nacisk zatarcia odzwierciedla zatem naciski w węźle tarcia przy maksymalnym, możliwym do uzyskania na aparacie czterokulowym obciążeniu P_{max} (przy którym aparat czterokulowy samoistnie wyłącza się). Wartość p_{oz} wyznaczano z zależności:

$$p_{oz} = 0{,}52 \cdot \frac{P_{max}}{d^2}, \left[\frac{N}{mm^2}\right], \tag{3}$$

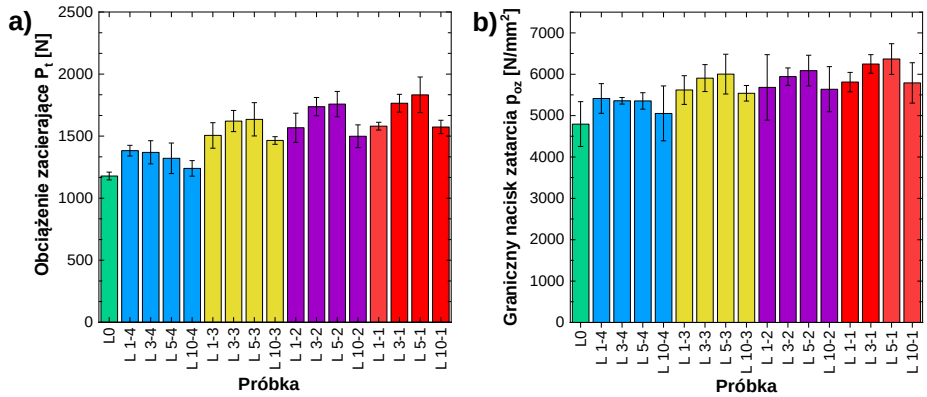
gdzie: 0,52 – współczynnik, wynikający z rozkładu sił w węźle tarcia aparatu czterokulowego (czworościanu foremnego), P_{max} – maksymalna wartość obciążenia zadanego (przy którym aparat samoistnie się wyłącza) [N], d – średnia średnica skaz na kulach [mm].

Zarówno obciążenie zacierające P_t jak i graniczny nacisk zatarcia p_{oz} charakteryzują właściwości przeciwzatarciowe smarów plastycznych i są parametrami wzajemnie się uzupełniającymi. Pierwszy parametr opisuje zachowanie środka smarnego do momentu przerwania filmu smarnego, z kolei drugi charakteryzuje jego właściwości w całym zakresie obciążeń węzła tarcia aparatu czterokulowego przy płynnie wzrastającym obciążeniu.

Ostatnim z oznaczanych parametrów były charakterystyki średniej średnicy skaz w funkcji czasu testu realizowanego na aparacie czterokulowym. Próby wykonywano przy stałym obciążeniu zadanym wynoszącym 300 N. Doboru obciążenia dokonano w oparciu o wskazania normy [61]. Obciążenie dobrano w taki sposób, aby podczas testu nie dochodziło do przerwania filmu smarnego, skutkującego nagłym wzrostem średniej średnicy skaz. Wzorem poprzednio realizowanych prac [8, 46] okres trwania biegów testowych wynosił kolejno: 1 minutę, 5 minut, 15 minut oraz 30 minut, a prędkość obrotowa kuli górnej 1450 obr./min. Takie charakterystyki pozwalają na ocenę właściwości przeciwzużyciowych smarów plastycznych.

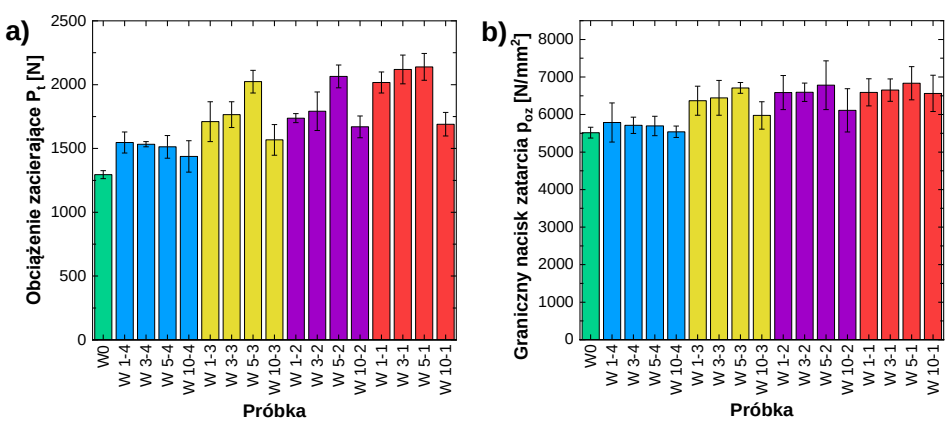
5. Wyniki badań tribologicznych i ich analiza

Na rysunkach 8 i 9 przedstawiono wyniki obciążenia zacierającego P_t i granicznego nacisku zatarcia p_{oz} uzyskane odpowiednio dla kompozycji smaru litowego i wapniowego.



Rysunek 8. Obciążenie zacierające P_t (a) oraz graniczny nacisk zatarcia p_{oz} (b) dla smarów litowych, opracowanie własne

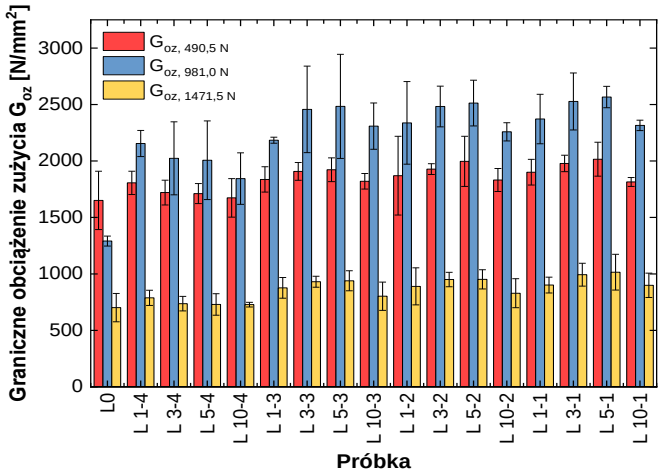
Dla bazowej próbki smaru litowego (L0) wartość obciążenia zacierającego P_t wynosiła 1178,71 N, co oznacza, że zgodnie z zależnością (2) do przerwania warstwy smarującej dochodziło średnio już po 2,88 s. Zastosowanie heksagonalnego azotku boru przyczyniło się do wzrostu odporności warstwy smarującej na zacieranie, co skutkowało zwiększonymi o 5,2÷55,5% względem smaru bazowego wartościami tego parametru uzyskanymi dla próbek z dodatkiem h-BN. Poprawa właściwości przeciwzatarciowych następowała wraz ze zmniejszaniem średniego rozmiaru cząstek heksagonalnego azotku boru. Najskuteczniejszy okazał się smar plastyczny zawierający 5% h-BN 1, a zatem dodatek z największym udziałem frakcji nanocząstek (rys. 3b). W przypadku wartości granicznego naciku zatarcia p_{oz} reprezentującego zachowanie substancji smarującej aż do osiągnięcia maksymalnego obciążenia P_{max} na dźwigni aparatu czterokulowego zaobserwowano podobne zależności. Po wprowadzeniu do smaru litowego heksagonalnego azotku boru odnotowano wzrost wartości tego parametru o 5,4÷32,8% względem próbki bazowej. Zastosowanie smaru stałego przyczyniło się do stabilizowania nacisków w strefach styku. Z tego względu na kulach testowych powstały mniejsze skazy zużyciowe, które stanowiły podstawę do wyznaczenia wartości granicznego nacisku zatarcia p_{oz} zgodnie z zależnością (3).



Rysunek 9. Obciążenie zacierające P_t (a) oraz graniczny nacisk zatarcia p_{oz} (b) dla smarów wapniowych, opracowanie własne

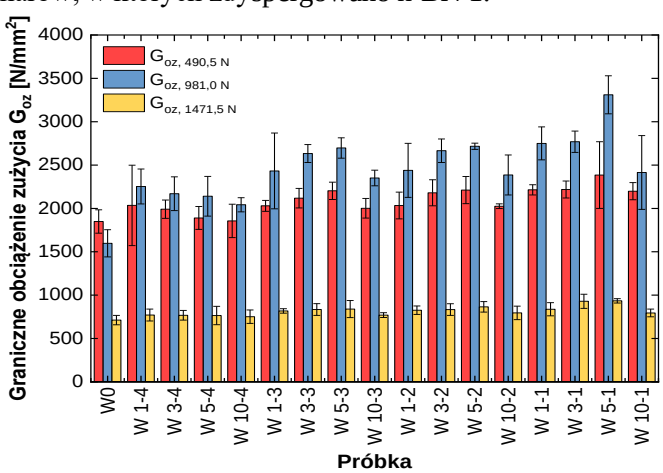
Kompozycja bazowa (W0) wykazała najgorsze właściwości przeciwzatarciowe spośród próbek smarów wapniowych opisywane przez obciążenie zacierające P_t wynoszące 1293,53 N. Wydłużenie czasu odporności warstwy smarującej na zacieranie nastąpiło po wprowadzeniu do smaru bazowego heksagonalnego azotku boru, co skutkowało wzrostem wartości tego obciążenia o 11,1÷65,3%. Największą wartość tego parametru osiągnięto dla próbki z dodatkiem 5% h-BN 1 i była ona równa 2139,39 N. Właściwości przeciwzatarciowe smarów plastycznych w całym cyklu obciążania węzła tarcia aparatu czterokulowego także poprawiły się po wprowadzeniu do smaru bazowego heksagonalnego azotku boru. Wartości granicznego nacisku zatarcia p_{oz} zależne były od średniego rozmiaru cząstek dodatku i wzrastały dla określonych stężeń wraz z jego zmniejszaniem. Istotny wpływ w tym zakresie należy także przypisać wielkości frakcji nanocząstek obecnych w danym materiale. Najwięcej takich cząstek, na podstawie rozkładów granulometrycznych przedstawionych na rysunku 3b, odnotowano w h-BN 1. Skutkowało to najlepszymi właściwościami przeciwzatarciowymi smarów plastycznych, zarówno litowych, jak i wapniowych, zawierających tę odmianę h-BN. Materiał (h-BN 4) zawierający niemal wyłącznie mikrocząstki – o średnim rozmiarze 6985 nm – okazał się najmniej skutecznym dodatkiem przeciwzatarciowym. Przypuszczalnie cząstki o takich rozmiarach mogły w małej liczbie (w porównaniu do próbek h-BN o mniejszej granulacji i większym udziale frakcji nanocząstek) przedostawać się do stref kontaktu powierzchni trących, co skutkowało niewielką poprawą właściwości przeciwzatarciowych względem smaru bazowego.

Na rysunku 10 przedstawiono wartości granicznego obciążenia zużycia G_{oz} uzyskane dla smarów litowych. Niezależnie od obciążenia zadanego wartości tego parametru były najmniejsze dla próbki bazowej, co wiązało się z największą średnią średnicą skaz obserwowaną na kulach testowych przy smarowaniu tą kompozycją. Próbka bazowa (L0) jako jedyna wykazała większe wartości granicznego obciążenia zużycia G_{oz} dla obciążenia 490,5 N w porównaniu z obciążeniem 981,0 N. Związane było to ze zwiększeniem średniej średnicy skaz odpowiednio z 0,395 mm na 0,629 mm. Właściwości przeciwzużyciowe smaru litowego poprawiły się po wprowadzeniu do smaru bazowego heksagonalnego azotku boru, o czym świadczą zwiększone wartości granicznego obciążenia zużycia G_{oz}. Najbardziej efektywne okazało się zastosowanie heksagonalnego azotku boru h-BN 1 w stężeniu 5%, szczególnie dla obciążenia 981,0 N. Wielkość cząstek heksagonalnego azotku boru była zatem, obok ich stężenia, czynnikiem decydującym o właściwościach przeciwzużyciowych smarów plastycznych. W miarę wzrostu średniego rozmiaru cząstek dodatku i zwiększaniu udziału frakcji mikrocząstek w nim zawartych, wartości granicznego obciążenia zużycia G_{oz} – przy określonym stężeniu – zmniejszały się. Były jednak zawsze większe niż te odnotowane dla próki bazowej (L0).



Rysunek 10. Graniczne obciążenie zużycia G_{oz} dla smarów litowych przy różnych obciążeniach (dla każdej próbki słupki ukazują kolejno – od lewej: $G_{oz,\ 490,5\ N}$; $G_{oz,\ 981,0\ N}$; $G_{oz,\ 1471,\ 5\ N}$) [opracowanie własne]

Na rysunku 11 przedstawiono wyniki granicznego obciążenia zużycia G_{oz} odnotowane dla smarów wapniowych. Podobnie jak w przypadku smaru litowego próbka bazowa wykazywała najgorsze właściwości przeciwzużyciowe. Wartości średniej średnicy skaz na kulach testowych smarowanych kompozycją bazową wynosiły 0,372 mm i 0,566 mm odpowiednio dla obciążenia 490,5 N i 981,0 N. Zaobserwowano w związku z tym mniejszą wartość granicznego obciążenia zużycia G_{oz} przy większym z podanych obciążeń. Zastosowanie heksagonalnego azotku boru ponownie przyczyniło się do poprawy właściwości przeciwzużyciowych próbki bazowej, a najlepsze efekty uzyskano dla kompozycji z 5% zwartością h-BN 1 przy obciążeniu 981,0 N. Należy przy tym przypomnieć, że materiał h-BN 1 wykazał najmniejsze cząstki o średnim rozmiarze równym 130 nm i największy udział frakcji nanocząstek (rys. 3b). Właściwości przeciwzużyciowe smarów plastycznych z taką samą procentową zawartością dodatków (h-BN 2, h-BN 3, h-BN 4) o większym średnim rozmiarze cząstek i mniejszym udziale frakcji nanocząstek były gorsze w porównaniu do smarów, w których zdyspergowano h-BN 1.

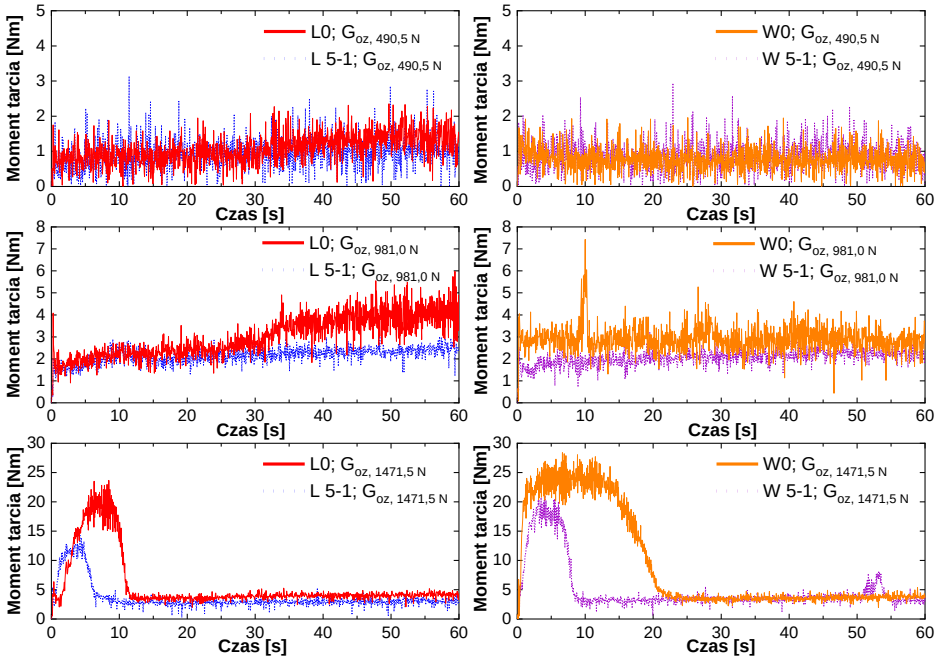


Rysunek 11. Graniczne obciążenie zużycia G_{oz} dla smarów wapniowych przy różnych obciążeniach (dla każdej próbki słupki ukazują kolejno – od lewej: $G_{oz,\ 490,5\ N}$; $G_{oz,\ 981,0\ N}$; $G_{oz\ 1471,\ 5\ N}$) [opracowanie własne]

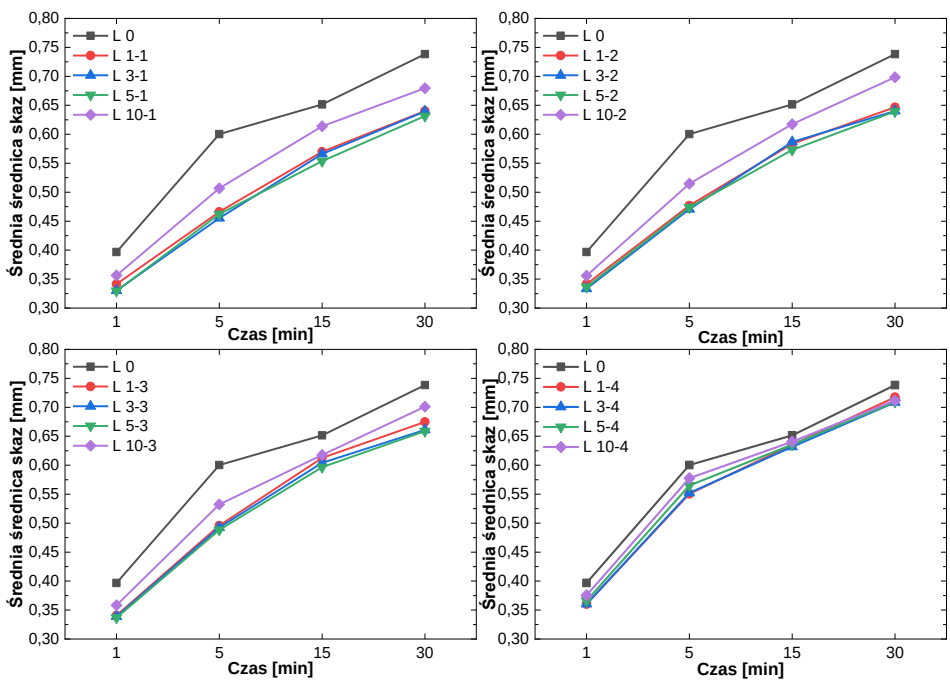
Wartości granicznego obciążenia zużycia G_{oz} uzyskane dla smarów litowych i wapniowych z zawartością h-BN prowadzą do konkluzji, że optymalne warunki działania cząstek smaru stałego odnotowano przy obciążeniu 981,0 N. Szczególnie dobrze spisywał się dodatek h-BN 1 z największym udziałem frakcji nanocząstek, o czym świadczą najmniejsze średnice skaz zużycia odnotowane dla smarów zawierających h-BN 1. Dla każdej z próbek smarów litowych i wapniowych wartości granicznego obciążenia zużycia uzyskane przy obciążeniu 1471,5 N były najmniejsze. Związane było to ze zwiększeniem średniej średnicy skaz w miarę wzrostu obciążenia zadanego. Takie zjawisko jest naturalne dla każdej substancji smarującej, a zgodnie z zależnością (1) determinuje wartości granicznego obciążenia zużycia.

Powyższe spostrzeżenia o działaniu przeciwzużyciowym h-BN potwierdzają przebiegi momentów tarcia przedstawione na rysunku 12, odnotowane przy różnych obciążeniach dla próbek bazowych oraz smarów z zawartością 5% h-BN 1, które wykazały najlepsze właściwości przeciwzużyciowe oceniane w oparciu o graniczne obciążenie zużycia G_{oz}. Przy obciążeniu 490,5 N charakterystyki uzyskane dla porównywanych próbek były zbliżone, zarówno w przypadku smaru litowego, jak i wapniowego. Zwiększenie obciążenia do 981,0 N skutkowało większymi różnicami – wartości momentu tarcia były wyższe dla próbek bazowych. Skutkowało to zwiększonym zużyciem kul badawczych względem próbek z dodatkiem h-BN i tym samym mniejszymi wartościami granicznego obciążenia zużycia G_{oz}. Także w przypadku obciążenia 1471,5 N i smarowaniu kompozycjami bazowymi odnotowano większe wartości momentu tarcia i tym samym słabsze właściwości przeciwzużyciowe. Przy smarowaniu mieszaninami z 5% zawartością h-BN 1 dochodziło do wzrostów momentów tarcia w początkowej fazie testu, które były jednak mniej intensywne i krótsze niż w przypadku smarów bazowych. Ukazane przebiegi wskazują zatem na przeciwzużyciowe działanie cząstek heksagonalnego azotku boru, przyczyniających się do zmniejszenia oporów tarcia w węźle aparatu czterokulowego i tym samym zmniejszania średnic skaz na kulach badawczych w porównaniu do kompozycji bez dodatku.

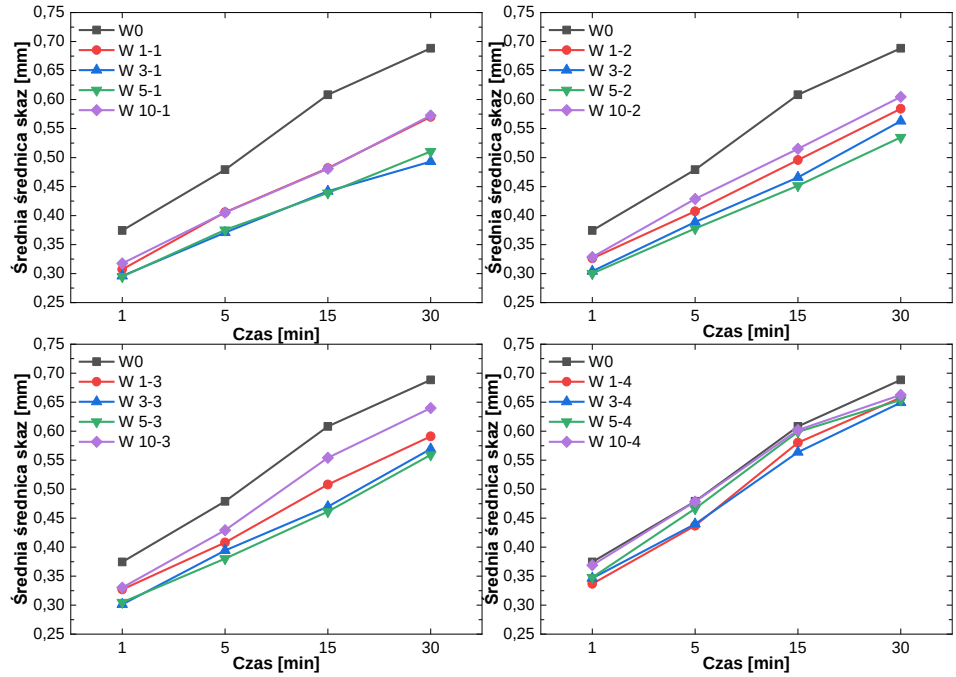
Właściwości przeciwzużyciowe przygotowanych kompozycji w funkcji czasu testu realizowanego na aparacie czterokulowym przy stałym obciążeniu zadanym 300 N oceniono w oparciu o charakterystyki przedstawione na rysunku 13 dotyczące smarów litowych i na rysunku 14 dotyczące smarów wapniowych. Na oddzielnych wykresach zamieszczono wyniki uzyskane dla smarów plastycznych zawierających poszczególne rodzaje heksagonalnego azotku boru w porównaniu do próbki bazowej.



Rysunek 12. Przykładowe przebiegi momentów tarcia dla wybranych próbek smarów litowych oraz wapniowych zarejestrowane podczas oznaczania granicznego obciążenia zużycia G_{oz}, opracowanie własne



Rysunek 13. Charakterystyki średniej średnicy skaz w funkcji czasu biegu testowego uzyskane przy stałym obciążeniu 300 N dla smarów litowych, opracowanie własne



Rysunek 14. Charakterystyki średniej średnicy skaz w funkcji czasu biegu testowego uzyskane przy stałym obciążeniu 300 N dla smarów wapniowych, opracowanie własne

W przypadku obu smarów bazowych średnie średnice skaz uzyskane w każdym punkcie testowym były większe od wartości odnotowanych dla kompozycji zawierających heksagonalny azotek boru. Przeciwzużyciowe działanie smaru stałego uwidaczniało się szczególnie w przypadku dłuższych oznaczeń trwających 5, 15 oraz 30 minut. Zastosowanie dodatku o największej granulacji (h-BN 4) przyczyniło się do najmniejszej zmiany średniej średnicy skaz w funkcji czasu testu względem bazowego smaru litowego oraz wapniowego. Jak już wcześniej stwierdzono, należy przepuszczać, że dopływ tak dużych cząstek do strefy tarcia mógł być utrudniony z uwagi na ich rozmiary, które są dużo większe niż w przypadku pozostałych próbek heksagonalnego azotku boru. Najmniejsze średnie średnice skaz zaobserwowano dla kompozycji zawierających dodatek h-BN 1 w stężeniach 3% i 5%. Dodatek ten charakteryzował się najmniejszym średnim rozmiarem cząstek i największym udziałem frakcji nanocząstek (rys. 3b). Ponownie okazało się, że wartości średniej średnicy skaz zmniejszały się dla określonych stężeń wraz ze zmniejszaniem granulacji heksagonalnego azotku boru.

Podsumowując uzyskane wyniki badań właściwości tribologicznych smarów plastycznych należy bez wątpliwości przyjąć, że zastosowanie heksagonalnego azotku boru przyczyniło się do poprawy charakterystyk smarnościowych bazowego smaru litowego i wapniowego.

Na ogół zmniejszenie średniego rozmiaru cząstek heksagonalnego azotku boru wiązało się z poprawą właściwości smaru plastycznego. Wniosek ten jest zbieżny z obserwacjami przedstawionymi przez innych autorów [38]. Dla smarów z zawartością h-BN 4 odnotowywano najmniejsze wartości obciążenia zacierającego P_t, granicznego nacisku zatarcia p_{oz} i granicznego obciążenia zużycia G_{oz}. W przypadku smarów z tym dodatkiem średnie

średnice skaz odnotowane przy testach realizowanych przy obciążeniu 300 N także były największe. Związane było to z największym rozmiarem cząstek tej próbki materiału i prawdopodobnie ze słabo rozwiniętą strukturą porowatą. Właściwości smarnościowe kompozycji bazowych poprawiały się w większym zakresie po zastosowaniu heksagonalnego azotku boru o mniejszych cząstkach. Dla dodatków o podobnych parametrach dotyczących średniego rozmiaru cząstek i podobnej zawartości nanocząstek tj. h-BN 2 i h-BN 3 uzyskano zbliżone rezultaty badań tribologicznych. Najlepsze efekty odnotowano dla smarów z zawartością dodatku o najmniejszym średnim rozmiarze i największym udziale frakcji nanocząstek (h-BN 1). Należy przypuszczać, że mechanizmy oddziaływania nanocząstek odgrywały znaczącą rolę w kształtowaniu właściwości tribologicznych smarów plastycznych. Nanocząstki heksagonalnego azotu boru są w stanie wytworzyć na smarownych powierzchniach trwałe warstwy ochronne. Uzyskane rezultaty badań zdają się potwierdzać to spostrzeżenie. Słuszne wydaje się także przypuszczenie, że nanocząstki są w stanie działać naprawczo, wypełniając powstałe na powierzchniach ubytki materiału. Taki efekt ich działania prawdopodobnie uwidocznił się podczas dłuższych testów przeciwzużyciowych wykonywanych na potrzeby charakterystyk przedstawionych na rysunkach 13 i 14. Uzyskano wówczas mniejsze średnice skaz niż w przypadku dodatków o mniejszym udziale tej frakcji.

Dla h-BN 4 właściwości tribologiczne smarów plastycznych, do których go wprowadzono były zwykle tym lepsze im mniejsze było jego stężenie. Dla pozostałych próbek najskuteczniejsze okazywało się na ogół zastosowanie 5% dodatku.

Rozwinięta struktura porowata próbek h-BN 1, h-BN 2 oraz h-BN 3 mogła skutkować lepszym związaniem cząstek heksagonalnego azotku boru ze smarem plastycznym, a tym samym ułatwionym ich dopływem do strefy tarcia aparatu czterokulowego. Podobna sugestia została przedstawiona przez autorów publikacji [32]. Potwierdzenie wniosków o możliwych mechanizmach oddziaływania mikrocząstek i nanocząstek heksagonalnego azotku boru wprowadzanych do smarów plastycznych wymaga jednak przeprowadzenia dodatkowych badań mikroskopowych i tribochemicznych.

6. Wnioski

W oparciu o przeprowadzone badania i analizę sformułowano następujące wnioski:

- wprowadzenie do bazowych smarów plastycznych heksagonalnego azotku boru przyczyniło się do wzrostu odporności na zacieranie oraz wzrostu odporności przeciwzużyciowej warstwy smarującej;
- właściwości smarnościowe litowych i wapniowych smarów plastycznych zawierających heksagonalny azotek boru były zależne od zastosowanego stężenia dodatku oraz rozkładu granulometrycznego cząstek h-BN, w tym udziału frakcji nanocząstek;
- właściwości smarnościowe smarów plastycznych poprawiały się wraz ze zmniejszaniem średniego rozmiaru cząstek h-BN i wzrostem udziału frakcji nanocząstek. Najlepsze rezultaty osiągnięto w związku z tym dla smarów plastycznych zawierających h-BN 1. Najskuteczniejsze okazało się na ogół zastosowanie tego dodatku w stężeniu 5%.

Podziękowania

Praca została dofinansowana przez Wojskową Akademię Techniczną w ramach projektu nr UGB/22-834/WAT/2023.

Literatura

1. Czarny R., *Smary plastyczne*, Wydawnictwa Naukowo-Techniczne, Warszawa 2004.
2. Krawiec S., *Kompozycje smarów plastycznych i stałych w procesie tarcia stalowych węzłów maszyn*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2011.
3. Zajezierska A., *Biodegradowalne smary plastyczne*, Instytut Nafty i Gazu, Kraków 2016.
4. Rawat S.S., Harsha A.P., *Current and Future Trends in Grease Lubrication* w Katiyar J.K., Bhattacharya S., Patel V.K., Kumar V. (red.), *Automotive Tribology*, Rozdział 9, Springer Singapore, 2019, s. 147-182.
5. Płaza S., Margielewski L., Celichowski G., *Wstęp do tribologii i tribochemia*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2005.
6. Lugt P.L., *A Review on Grease Lubrication in Rolling Bearings*, Tribology Transactions, 52, 4, 2009, s. 470-480.
7. Reeves C.J., Menezes P.L., Lovell M.R., Jen T-C., *Tribology of Solid Lubricants*, [w:] Menezes P.L., Nosonovsky M., Ingole S.P., Kailas S.V., Lovell M.R. (red.), *Tribology for Scientists and Engineers*, Springer, New York 2013, s. 447-494.
8. Kałdoński T., *Tribologiczne zastosowania azotku boru*, Wojskowa Akademia Techniczna, Warszawa 2006.
9. Wang J., Ma F., Sun M., *Graphene, hexagonal boron nitride, and their heterostructures: properties and applications*, RSC Advances, 27, 2017.
10. Kimura Y., Wakabayashi T., Okada K., Wada T., Nishikawa H., *Boron nitride as a lubricant additive*, Wear, 232, 1999, s. 199-206.
11. Yan C., Zeng Q., Hao Y., Xu Y., Zhou M., *Friction-induced hardening behaviors and tribological properties of 60NiTi alloy lubricated by lithium grease containing nano-BN and MoS2*, Tribology Transactions, 62(5), 2019, s. 812-820.
12. Sahoo R.R., Biswas S.K., *Effect of Layered MoS2 Nanoparticles on the Frictional Behavior and Microstructure of Lubricating Greases*, Tribology Letters, 53, 1, 2014, s. 157-171.
13. Chkhartishvili L., Tabatadze G., Nackebia D., Bzhalava T., Kalandadze I., *Hexagonal Boron Nitride as a Solid Lubricant Additive (An Overview)*, Nano Studies, 14, 2016, s. 91-98.
14. Lipp A., Schwetz A., Hunold K., *Hexagonal Boron Nitride: Fabrication, Properties and Applications*, Journal of European Ceramic Society, 5, 1989, s. 3-9.
15. Radomski J., Rećko W.M., Ketling-Szemley M., *Własności azotku boru i metody jego otrzymywania*, WEMA, Warszawa 1979.
16. Hebda M., Wachal A., *Trybologia*, Wydawnictwa Naukowo-Techniczne, Warszawa 1980.
17. Senyk S., Chodkiewicz A., Gocman K., Szczęśniak B., Kałdoński T., *Hexagonal Nano and Micro Boron Nitride: Properties and Lubrication Applications*, Materials, 15(3), 955, 2022.
18. Kurzydłowski K., Lewandowska M., *Nanomateriały inżynierskie. Konstrukcyjne i funkcjonalne*, Wydawnictwo Naukowe PWN SA, Warszawa 2010.
19. Tang Z., Li S., *A review of recent developments of friction modifiers for liquid lubricants (2007–present)*, Current Opinion in Solid State and Materials Science, 18, 2014, s. 119-139.
20. Gulzar M., Masjuki H.H., Kalam M.A., Varman M., Zulkifli N.W.M., Mufti R.A., Zahid R., *Tribological performance of nanoparticles as lubricating oil additive*, Journal of Nanoparticle Research, 18, 223, 2016.
21. Lee K., Hwang Y., Cheong S., Choi Y., Kwon L., Lee J., Kim S.H., *Understanding the Role of Nanoparticles in Nano-oil Lubrication*, Tribology Letters, 35, 2009, s. 127-131.
22. Zhao J., Huang Y., He Y., Shi Y., *Nanolubricant additives: A review*, Friction, 9, 5, 2021, s. 891-917.
23. Kałdoński T., *Wpływ azotku boru na właściwości użytkowe smarów plastycznych do łożysk ślizgowych*, Materiały VIII Kongresu Eksploatacji Urządzeń Technicznych, Krynica Górska, Polska, 23–26 września 1997.

24. Kałdoński T., *Wpływ rodzaju smaru plastycznego i azotku boru na proces samosmarowania łożyska porowatego*, Materiały VIII Kongresu Eksploatacji Urządzeń Technicznych, Krynica Górska, Polska, 23-26 września 1997.
25. Kałdoński T., Krzemiński, K., *Zastosowanie azotku boru jako dodatku do smaru plastycznego wypełniającego łożysko toczne*, Materiały VIII Kongresu Eksploatacji Urządzeń Technicznych, Krynica Górska, Polska, 23-26 września 1997.
26. Kałdoński T., *Badania i technologia optymalnego stosowania azotku boru α-BN jako dodatku smarnościowego w sprzęcie wojskowym*, Sprawozdanie z realizacji projektu badawczego PB 0T 00A 014 08, WAT, Warszawa 1998.
27. Kałdoński T., Nosal S., Zwierzycki W., *Wpływ zawartości azotku boru produkcji WAT na właściwości antyfrettingowe substancji smarujących*, Materiały IV Konferencji Poltrib'97, Białobrzegi, Polska, 10-12 czerwca 1997.
28. Pawlak Z., Kałdoński T., Lisewski M., Urbaniak W., Oloyede A., *The Effect of Hexagonal Boron Nitride Additive on the Effectiveness of Grease-Based Lubrication of a Steel Surface*, Industrial Lubrication and Tribology, 64(2), 2012, s. 84-89.
29. Bajer J., *Wpływ dodatku smarnego na charakterystyki tribologiczne niskotemperaturowego smaru plastycznego*, Tribologia, 6, 2006, s. 67-80.
30. Krawiec M., Leśniewski T., *Badanie efektywności smarowania litowego smaru plastycznego napełnionego azotkiem boru*, Górnictwo Odkrywkowe, 2013, 54, s. 82-86.
31. Podgornik B., Kosec T., Kocijan A., Donik Č., *Tribological behaviour and lubrication performance of hexagonal boron nitride (h-BN) as a replacement for graphite in aluminium forming*, Tribology International, 81, 2015, 267-275.
32. Wang T., Li Z., Li J., He Q., *Impact of Boron Nitride Nanoparticles on the Wear Property of Lithium Base Grease*, Journal of Materials Engineering and Performance, 29(8), 2020, s. 4991-5000.
33. Wu C., Zie Y., Zhao H., Yang H., Li X., Ni J., *Effects of hBN and $CaCO_3$ Nanoparticles on Tribological and Vibration Properties of Polyurea Grease on Rolling Bearing*, Tribology Letters 70, 95, 2022.
34. Gorbacheva S.N., Yadykova A.Y., Ilyin S.O., *Rheological and tribological properties of low-temperature greases based on cellulose acetate butyrate gel*, Carbohydrate Polymers 272 118509, 2021.
35. Guan Q., Zhang X.W., He J., Li Z., Wei Y.L., Ren F.E., Gong N., Ding H.H., Wang W.J., *Effect of WS2 and h-BN Nano-additive on Extreme Pressure and Anti-wear Properties of Semi-fluid Lithium Base Grease*, Surface Technology, 51, 8, 2022, s. 233-242.
36. Amiruddin H., Abdollah M.F.B., Norani M.N.M., *Lubricity and mechanical stability of bio-grease formulated from non-edible vegetable oil*, Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2022.
37. Wu C., Hong Y., Ni J., Teal P.D., Yao L., Li X., *Investigation of mixed hBN/Al2O3 nanoparticles as additives on greaseperformance in rolling bearing under limited lubricant supply*, Colloids and Surfaces A: Physicochemical and Engineering Aspects, 659, 130811, 2023.
38. Kumar N., Saini V., Bijwe J., *Dependency of Lithium Complex Grease on the Size of hBN Particles for Enhanced Performance*, Tribology Letters, 71, 20, 2023.
39. Józwiak P., Siczek K., *Badania wpływu obecności azotku boru w smarze na opory ruchu sworznia drążka kierowniczego*, Archiwum Motoryzacji, 58, 2012, s. 123-137.
40. Torres H., Podgornik B., Jovicevic-Klug M., Rodríguez Ripoll M., *Compatibility of graphite, hBN and graphene with self-lubricating coatings and tool steel for high temperature aluminium forming*, Wear, 490-491, 204187, 2022.
41. Krawiec S., Krawiec M., Leśniewski T., *Evaluation of lubrication Efficiency based on wear during variable loads and sliding velocities*, Tribologia, 4, 2016, s. 127-136.

42. Wen M.Y., Guo P., Hui J.D., Hui, Z.M., Geng F., *The effect of additives on tribological behaviours in polyurea grease*, Materials Research Innovations, 19, 2015, Issue sup5: Global Conference on Materials Science and Engineering (CMSE 2014).
43. Kumar N., Saini V., Bijwe J., *Tribological Investigations of Nano and Micro-sized Graphite Particles as an Additive in Lithium-Based Grease*, Tribology Letters, 68, 124, 2020.
44. Hu E. Z., Xu Y., Hu K.H., Hu X.G., *Tribological properties of 3 types of MoS2 additives in different base greases*, Lubrication Science, 29, 2017, s. 541-555.
45. Honary L., Richter E., *Biobased Lubricants and Greases: Technology and Products*, Wiley, 2011.
46. Senyk S., Kałdoński T., *Analysis of the influence of hexagonal boron nitride on tribological properties of greases*, Tribologia, 301, 3, 2022, s. 101-112.
47. Shah A., Bagi S., Aswath P., *Wear and Friction of Greases Containing Organic and Inorganic Sulfur Carriers*, Tribology Online, Vol. 12, 4/2017, s. 162-170.
48. Bagi S.., Aswath P., *Mechanism of Friction and Wear in MoS2 and ZDDP/F-PTFE Greases under Spectrum Loading Conditions*, Lubricants, 3, 2015, s. 687-711.
49. Nehme G., *The Importance of Variable Speeds under Extreme Pressure Loading in Molybdenum Disulfide Greases Using Four-Ball Wear Tests*, Tribology Transactions, 56:6, 2013, s. 977-985.
50. Nehme G., *Tribological behavior and wear prediction of molybdenum disulfide grease lubricated rolling bearings under variable loads and speeds via experimental and statistical approach*, Wear, 376-377, 2017, s. 876-884.
51. Vyavhare K., Aswath P., *Tribological Properties of Novel Multi-Walled Carbon Nanotubes and Phosphorus Containing Ionic Liquid Hybrids in Grease*, Frontiers in Mechanical Engineering, 5(15), 2019.
52. Krawiec S., Krawiec M., Leśniewski T., *Wpływ dodatków uszlachetniających w smarze plastycznym na efektywność smarowania jego kompozycjami z PTFE, Cu*, Tribologia, 3, 2014, s. 109-117.
53. Wu C., Zie Y., Zhao H., Yang H., Li X., Ni J., *Effects of hBN and CaCO3 Nanoparticles on Tribological and VibrationProperties of Polyurea Grease on Rolling Bearing*, Tribology Letters, 70, 95, 2022
54. Lopez-Sanchez M.A. *GrainSizeTools: A Python Script for Grain Size Analysis and Paleopiezometry Based on Grain Size*, Journal of Open Source Software, 3, 2018, s. 863.
55. Brunauer S., Emmett P.H., Teller E., *Adsorption of Gases in Multimolecular Layers*, Journal of the American Chemical Society, 60, s. 309-319.
56. Gregg S.J., Sing K.S.W., Salzberg, H.W., *Adsorption Surface Area, and Porosity*, Journal of The Electrochemical Society, 114, 279C, 1967.
57. Jagiello J., Olivier J.P., *2D-NLDFT Adsorption Models for Carbon Slit-Shaped Pores with Surface Energetical Heterogeneity and Geometrical Corrugation*, Carbon, 55, 2013, s. 70-80.
58. Poradnik TOTAL: *Przemysłowe środki smarne*, Warszawa 2003.
59. PN/C-04147:1976: Przetwory naftowe – Badanie własności smarnych olejów i smarów.
60. PN/C-04362:2017-03: Przetwory naftowe – Badanie właściwości smarnych olejów i smarów.
61. PN-EN ISO 20623:2018-02: Przetwory naftowe i produkty podobne. Oznaczenie właściwości przeciwzatarciowych i przeciwzużyciowych środków smarowych. Metoda czterokulowa (warunki europejskie).
62. Szczerek M., Tuszyński W., *Badania tribologiczne. Zacieranie*, Biblioteka Problemów Eksploatacji, Radom 2000.
63. Tuszyński W., *Badanie przeciwzużyciowych oddziaływań środków smarowych (rozprawa doktorska)*, Wojskowy Instytut Techniki Pancernej i Samochodowej, Sulejówek 1999.

Właściwości tribologiczne smarów plastycznych zawierających heksagonalny azotek boru o różnej granulacji

Streszczenie

Smary plastyczne to środki smarne stanowiące kompozycje oleju bazowego, środka zagęszczającego oraz dodatków odpowiedzialnych za kształtowanie ich określonych właściwości. Rozwój i poprawa efektywności nowoczesnych smarów plastycznych związane są często z zastosowaniem innowacyjnych dodatków, takich jak np. nanomateriały. Jednym z nich może być heksagonalny azotek boru (h-BN).

W niniejszej pracy opisano badania nad zastosowaniem nanocząstek i mikrocząstek heksagonalnego azotku boru wprowadzanych w różnych stężeniach do litowego i wapniowego smaru plastycznego. Zastosowano cztery rodzaje h-BN, które scharakteryzowano pod kątem rozpatrywanego zastosowania wykorzystując w tym celu skaningową mikroskopię elektronową (SEM), dyfrakcję rentgenowską (XRD) oraz pomiary adsorpcyjne. Określono w ten sposób morfologię i rozmiar cząstek h-BN, ich budowę krystaliczną oraz strukturę porowatą. Badania tribologiczne wykonano na aparacie czterokulowym, oznaczając parametry charakteryzujące właściwości smarnościowe analizowanych kompozycji.

W wyniku zrealizowanych prac stwierdzono, że właściwości tribologiczne smarów plastycznych poprawiają się wraz ze zmniejszaniem granulacji dodatku. Najlepszą efektywność smarowania osiągnięto dla smarów bazowych zawierających nano h-BN w określonych stężeniach. Identyfikację właściwości dodatku uznano za istotny element takich rozważań uławiający szczegółową analizę wpływu h-BN na właściwości tribologiczne smarów plastycznych.

Słowa kluczowe: tribologia, smary plastyczne, heksagonalny azotek boru, nanododatki

Tribological properties of greases containing hexagonal boron nitride with different granulation

Abstract

Greases are lubricants that are compositions of base oil, thickener and additives responsible for forming their specific properties. The development and improvement of the effectiveness of modern greases are often associated with the use of innovative additives, such as nanomaterials. One of them may be hexagonal boron nitride (h-BN).

This paper describes the research on the use of nanoparticles and microparticles of hexagonal boron nitride introduced in various concentrations into lithium and calcium base grease. Four types of h-BN were used, which were characterized for the considered application using scanning electron microscopy (SEM), X-ray diffraction (XRD) and adsorption measurements. In this way, the morphology and size of h-BN particles, their crystal structure and porous structure were determined. Tribological tests were performed on a four-ball apparatus, determining the parameters characterizing the lubricity properties of the analysed compositions.

As a result of the completed work, it was found that the tribological properties of greases improve with the reduction of additive granulation. The best lubrication efficiency was achieved for greases containing nano h-BN in certain concentrations. Identification of additive properties was considered an important element of such considerations, facilitating a detailed analysis of the influence of h-BN on the tribological properties of greases.

Keywords: tribology, greases, hexagonal boron nitride, nanoadditives

Filip Turoboś[1], Mateusz Krukowski[2], Nicole Meisner[3]

Przegląd wybranych algorytmów aproksymacyjnych dla problemu komiwojażera

1. Wprowadzenie

Mnogość algorytmów przybliżonych pojawiających się w ostatnich dekadach dla najróżniejszych problemów NP-trudnych sprawia, że nawet doświadczonym naukowcom trudno jest nadążyć za takim tempem rozwoju. W tym artykule uwagę zwrócimy na szczególne zagadnienie optymalizacji kombinatorycznej, pojawiające się w całej gamie rozważań teoretycznych i praktycznych – problem komiwojażera (ang. *TSP*). Ze względu na jego trudność, zamiast podejmować prób jego dokładnego rozwiązania, uwagę skupimy na metodach przybliżonych, którym poświęcona jest poniższa praca. W przeciwieństwie do dokładnych algorytmów rozwiązujących TSP, metody przybliżone działają znacznie szybciej (w czasie wielomianowym), co okupione jest pewną nieoptymalnością uzyskiwanych przez nie rozwiązań.

1.1. Cel artykułu

Celem tego artykułu jest przedstawienie możliwie szerokiego przeglądu metod heurystycznych dla problemu komiwojażera w sposób przystępny, bez zaawansowanych detali implementacyjnych – tak, aby nawet początkujący badacz zajmujący się teorią optymalizacji kombinatorycznej był w stanie bez trudu przyswoić sobie podstawowe koncepcje przedstawiane w ramach tej pracy. Ze względu na obszerność tej tematyki, omówienie wszystkich dostępnych na tę chwilę technik przybliżonych nie jest możliwe w praktyce. Czytelnika zainteresowanego semimetryczną wersją problemu komiwojażera odsyłamy do artykułu Krukowskiego i Turobosia [1], gdzie porównano heurystyki dedykowane temu wariantowi. Obiektywne zestawienie ze sobą przedstawianych rodzin metaheurystyk, które można wykorzystywać do rozwiązywania problemu komiwojażera w pełnej ogólności byłoby materiałem na obszerną monografię. Głównym argumentem, który przemawia za tym jest fakt, że niemal każda z zaproponowanych technik ma po kilka parametrów, które mogą przyjmować dziesiątki całkowicie różnych wartości (zazwyczaj owe parametry są funkcjami czy metodami służącymi do przeszukiwania sąsiedztwa rozpatrywanego rozwiązania). Jest to interesujący punkt wyjścia do dalszych badań – z uwagi na rozległość tego tematu, prace nad nim zostały podzielone na konkretne części (analiza teoretyczna, implementacyjna i porównawcza). Kontynuacja naszych badań, w której przedstawimy część implementacyjną i zestawienie wyników, jest zaplanowana na najbliższą przyszłość.

[1] filip.turobos@p.lodz.pl, Instytut Matematyki, Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej, Politechnika Łódzka.
[2] mateusz.krukowski@p.lodz.pl, Instytut Matematyki, Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej, Politechnika Łódzka.
[3] meisnernicole@gmail.com, Wydział Matematyki, Informatyki i Mechaniki, Uniwersytet Warszawski.

1.2. Organizacja tekstu

Artykuł zorganizowany jest w następujący sposób: w pierwszej sekcji wprowadzamy niezbędne pojęcia z teorii grafów oraz optymalizacji kombinatorycznej. Jednocześnie ustalona zostanie notacja, stosowana w dalszych częściach artykułu. W dalszej części przejdziemy do zasadniczego przeglądu wybranych metod uzyskiwania przybliżonych rozwiązań dla problemu komiwojażera, które podzieliliśmy na następujące kategorie:

- inteligentne metaheurystyki ogólnego przeznaczenia;
- heurystyki konstrukcyjne;
- heurystyki optymalizacji lokalnej.

Szczegóły przedstawionego podziału staną się jasne w dalszej części artykułu. Końcową część pracy poświęcono na krótkie zestawienie i wnioski.

2. Preliminaria

W celu ustalenia wspólnego formalizmu, wprowadzone zostaną niezbędne pojęcia z teorii grafów, jak również teorii optymalizacji kombinatorycznej. Umożliwi nam to również wprowadzenie spójnej symboliki

2.1. Niezbędne pojęcia z teorii grafów

Niniejsze wprowadzenie bazuje na monografiach [2,3], stamtąd też pochodzi większość zastosowanych w niniejszym tekście oznaczeń.

Parę $G := (V, E)$ nazywamy **grafem**, jeżeli V jest skończonym, niepustym zbiorem, zaś $E \subset \{\{x, y\} : x, y \in V,\ x \neq y\}$. Elementy V i E nazywamy, odpowiednio, **wierzchołkami** i **krawędziami**. **Graf ważony**, to para postaci (G, ω), gdzie $G = (V, E)$ jest grafem, $E \neq \emptyset$, zaś ω: $E \to (0, +\infty)$ jest nieujemną funkcją, nazywaną wagą. Czytelnik zaznajomiony z podstawowymi pojęciami z teorii grafów jest w stanie natychmiast zauważyć, że w ramach powyższej definicji grafy postrzegamy jako grafy *proste* (bez krawędzi wielokrotnych lub łączących wierzchołek z nim samym) i *nieskierowane,* tj. takie, w których kolejność wierzchołków nie ma znaczenia. Nieformalnie można to wyjaśnić w ten sposób, że krawędzie należy traktować jako drogi *dwukierunkowe*. W dalszej części tekstu mogą pojawić się nawiązania do **grafów skierowanych**. W takich grafach przyjmować będziemy, że zbiór krawędzi składa się z par uporządkowanych. Oznacza to, że zbiór krawędzi należy rozpatrywać jako $E \subset \{(x, y) \in V^2 : x, y \in V, x \neq y\}$. Dla ustalonego grafu G możemy zdefiniować następujące pojęcia pomocnicze [2,4,5]:

- Graf F nazwiemy **podgrafem** G, jeżeli $V(F) \subset V(G)$ i $E(F) \subset E(G)$, co oznaczamy przez $F \subset G$.
- Graf P nazwiemy **ścieżką**, jeżeli jego wierzchołki P można uszeregować w ciąg tak, by dwa wierzchołki były połączone krawędzią wtedy i tylko wtedy, gdy następują po sobie w owym ciągu.
- Graf C nazwiemy **cyklem**, jeżeli usunięcie dowolnej krawędzi z C przekształca go w ścieżkę. Jeżeli G nie zawiera żadnego cyklu jako podgrafu, to określamy go jako **acykliczny**.
- Podgraf $H \subset G$ nazwiemy **ścieżką/cyklem Hamiltona**, jeżeli jest ścieżką/cyklem i odwiedza każdy z wierzchołków G dokładnie raz. Rodzinę wszystkich cykli

Hamiltona w grafie G oznaczamy przez $\mathfrak{h}(G)$. Jeżeli zachodzi $\mathfrak{h}(G) \neq \emptyset$ to G nazwiemy **grafem Hamiltonowskim**.

Wprowadzone pojęcie podgrafu jest niezależne od funkcji wagi w grafie G (o ile taka waga została dla grafu G określona). Niemniej, jeżeli ω jest wagą w G, zaś F jest podgrafem G, to wówczas $\left(F, \omega|_{E(F)}\right)$ jest grafem ważonym, zaś $\omega|_{E(F)}$ nazwiemy **wagą indukowaną**. Dla podgrafu F grafu ważonego (G, ω) możemy zdefiniować jego wagę jako

$$\Omega(F) := \sum_{e \in E(F)} \omega\,(e).$$

W szczególności w roli F można rozpatrywać cały graf G. Dysponując kompletem pojęć z teorii grafów niezbędnym do zdefiniowania problem komiwojażera (skrótowo TSP, ang. *Travelling Salesperson Problem*), możemy sformułować go następująco, korzystając z wprowadzonej uprzednio notacji[4]:

Dla ważonego grafu Hamiltonowskiego (G, ω) ***znaleźć cykl Hamiltona*** H ***o możliwie najmniejszej wadze.***

Mając już skonkretyzowany problem, możemy już nawiązać zarówno do jego natury, jak również sposobu jego rozwiązywania. Potrzebne pojęcia wprowadzimy w kolejnej sekcji.

2.2. Najważniejsze pojęcia związane z optymalizacją kombinatoryczną

W poniższej części artykułu przedstawiamy kilka przydatnych definicji z optymalizacji i teorii złożoności. Naszym punktem wyjścia będzie samo pojęcie problemu optymalizacyjnego (które w różnych ujęciach można znaleźć w [10-12]). Dla niepustego, skończonego zbioru X i dowolnej funkcji $f: X \to (0, +\infty)$, problem znalezienia elementu $x^* \in X$ spełniającego

$$f(x^*) = \min_{x \in X} f(x),$$

nazywamy będziemy **problemem optymalizacyjnym**. Zbiór X i funkcja f w problemie optymalizacyjnym nazywane są często **przestrzenią poszukiwań** i **funkcją celu**. Każdy element spełniający równość w definicji nazywać będziemy **rozwiązaniem optymalnym**. Elementy X nazywać będziemy **rozwiązaniami dopuszczalnymi**. Nie tak rzadką sytuacją jest fenomen, w którym kilka rozwiązań dopuszczalnych jest jednocześnie optymalne.

W przypadku problemu komiwojażera określonego na grafie (G, ω), łatwo zauważyć, że X stanowi zbiór wszystkich cykli Hamiltona, zaś f jest funkcją wagi, przykładaną do stanowiącego rozpatrywany cykl podgrafu. W szczególności w grafie pełnym jest ich $0.5 \cdot (n-1)!$, tak więc przejrzenie wszystkich rozwiązań dopuszczalnych z wykorzystaniem tzw. metody siłowej wydaje się być niemożliwe do przeprowadzenia dla większych grafów. Oczywiście sam rozmiar przestrzeni poszukiwań nie wyklucza możliwości znalezienia rozwiązania optymalnego (wystarczy spojrzeć na problem sortowania, gdzie liczba możliwych ustawień elementów tablicy n-elementowej to $n!$). Znanym jest jednak fakt, że problem komiwojażera nie daje się rozwiązywać szybko, nawet dla stosunkowo niewielkich wartości n. Zamiast skupiać się na próbach rozwiązania tego

[4] W tym miejscu warto podkreślić, że dalece nie jest to jedyny sposób definiowania problemu komiwojażera, przykłady równoważnych sformułowań można znaleźć w [6-9].

problemu, poszukiwać będziemy **rozwiązań przybliżonych**, tj. takich rozwiązań dopuszczalnych, których wartość funkcji celu jest mała – czyli dostatecznie zbliżona do minimum globalnego funkcji f, by spełniać nasze oczekiwania. Oczywiście takie podejście ma sens, gdy jest aplikowalne do całej rodziny problemów, a nie do pojedynczej instancji – mowa wówczas o **metodzie/algorytmie aproksymacyjnym**. W pewnych sytuacjach takie metody mają gwarancję na jakość uzyskiwanego rozwiązania. To oszacowanie dla jakości przybliżenia odpowiada na pytanie „ile razy gorsze od rozwiązania optymalnego może być rozwiązanie przybliżone, uzyskane przez tę technikę?".

Chcąc sformalizować to pojęcie [10-13], ustalmy $s \geq 1$ i niech $\mathcal{P}$ będzie zbiorem indeksów dla pewnej rodziny problemów optymalizacyjnych o funkcjach celu

$$f_p: X_p \to (0, +\infty),\ p \in \mathcal{P}.$$

Odwzorowanie $\mathfrak{a}: \mathcal{P} \to \bigcup_{p\in\mathcal{P}} X_p$ nazwiemy s-optymalną aproksymacją dla tej rodziny, jeżeli dla każdego $p \in \mathcal{P}$ zachodzi

$$\mathfrak{a}(p) \in X_p \text{ i } \frac{f_p\big(\mathfrak{a}(p)\big)}{f_p\big(x_p^*\big)} \leq s,$$

gdzie x_p^* jest rozwiązaniem optymalnym dla p −tego problemu. Element $\mathfrak{a}(p) \in X_p$ nazywamy wówczas $\boldsymbol{s}$ **−optymalną aproksymacją** dla p −tego problemu z tej rodziny.

Oczywiście najważniejszym argumentem przemawiającym za stosowaniem rozwiązań przybliżonych w miejscu rozwiązań dokładnych jest szybkość, z jaką możemy je uzyskać. Chcąc o niej dyskutować, wprowadzimy jeszcze jedną, ostatnią definicję pomocniczą, jaką jest złożoność obliczeniowa algorytmu [14-19]:

Niech $\mathcal{P}$ będzie zbiorem indeksów dla pewnej rodziny problemów optymalizacyjnych, zaś $\mathcal{M}: \mathcal{P} \to \mathbb{N}$ będzie funkcją, która każdemu $p \in \mathcal{P}$ przypisuje liczbę jednostek pamięci w danym modelu obliczeniowym (dla naszych rozważań możemy się ograniczyć do klasycznej maszyny Turinga), użytych do opisu p −tego problemu. Powiemy, że algorytm $\mathfrak{a}: \mathcal{P} \to \mathbb{R}$ ma **złożoność obliczeniową (czasową)** nie większą niż $F_\mathfrak{a}: \mathbb{N} \to \mathbb{N}$, jeżeli istnieje stała $\alpha \geq 0$, taka że dla p-tego problemu z tej rodziny algorytm ten wykonuje co najwyżej $\alpha \cdot F_\mathfrak{a}\big(\mathcal{M}(p)\big)$ operacji. Mówimy wówczas, że $\mathfrak{a}$ można wykonać w czasie $\mathcal{O}\big(F_\mathfrak{a}(n)\big)$.

3. Przegląd wybranych metod uzyskiwania rozwiązań przybliżonych dla problemu komiwojażera

Poniżej prezentujemy zestawienie znanych metod aproksymacji dla TSP. Podział tych metod, zapowiedziany w początkowej części pracy, podyktowany jest specyfiką konkretnych podejść. Dla wygody czytelnika przypominamy go raz jeszcze, wraz z wyjaśnieniem poszczególnych stosowanych przez nas terminów:

- ***Inteligentne metaheurystyki*** – techniki te są aplikowalne do różnorodnych problemów optymalizacyjnych i zostały (do pewnego stopnia) przystosowane do potrzeb problemu komiwojażera. Metody te stanowią potężne i niezwykle wszechstronne narzędzia optymalizacyjne, które można wykorzystać do mnóstwa różnorakich problemów. Niestety, ich ogólność jest też pewnego rodzaju wadą – nie wykorzystują one bowiem specyficznych zjawisk dla zadanego problemu, więc uzyskiwane przez nie rozwiązania mają często tendencję do bycia nieco gorszymi niż w przypadku metod

opracowywanych z myślą o konkretnych problemie. Ponadto dość często są one stosunkowo czasochłonne i wymagają dużych mocy obliczeniowych by uzyskać wartościowe rozwiązanie.

- ***Heurystyki konstrukcyjne*** – te algorytmy dostarczają metodę-wzorzec konstrukcji rozwiązania przybliżonego dla zadanej instancji problemu komiwojażera, wytwarzając rozwiązanie *od podstaw*. Można je efektywnie łączyć z algorytmami z poprzedniej grupy (jako metoda sprytnej inicjalizacji), ale również są dobrym punktem wyjścia dla algorytmów z kolejnego kręgu.
- ***Heurystyki lokalnej optymalizacji*** – specyfiką działania tych algorytmów jest wprowadzanie udoskonaleń do już istniejącego rozwiązania, w celu jego poprawy. Zazwyczaj dokonywane jest to przy pomocy wprowadzania niewielkich zmian do aktualnie opracowywanego rozwiązania dopuszczalnego, które zmniejszają koszt podróży komiwojażera wzdłuż rozpatrywanej trasy. Metody te mogą być używane samodzielnie, jako techniki poszukiwania rozwiązania z losowym punktem wyjścia, lub mogą być dodawane jako końcowe *szlify* do rozwiązania uzyskanego przy pomocy innej techniki.

Można też wyróżnić heurystyki dedykowane dla konkretnych wariantów problemu – są to heurystyki z gwarancją jakości uzyskiwanego rozwiązania, ale jednocześnie wymagają, by graf, na którym rozpatrujemy daną instancję TSP spełniał pewne dodatkowe założenia. Wymagałoby to jednak znacznie głębszego wejścia w szczegóły techniczne, na które nie mamy miejsca w tym artykule.

3.1. Inteligentne metody heurystyczne dla problemu komiwojażera

Techniki omawiane w ramach poniższej podsekcji są zdolne do wyprodukowania rozwiązania przybliżonego dla dowolnej instancji problemu komiwojażera, po uprzednim dostosowaniu. Przyjrzyjmy się teraz algorytmom z tej grupy, nadmieniając jednocześnie, że są one stosowalne do wielu innych problemów optymalizacyjnych.

Ponieważ zarówno złożoność czasowa, jak i obliczeniowa poniższych technik zależna jest od różnorakich aspektów związanych z ich hiperparametryzacją (np. liczba iteracji, liczba agentów itd.), nie będziemy się zagłębiać w ten aspekt, ograniczając się do zaprezentowania samego przeglądu rozważanych heurystyk. Jednocześnie należy podkreślić raz jeszcze, że metody te zazwyczaj nie są tak wydajne czasowo, jak heurystyki dedykowane konkretnie dla problemu komiwojażera. Nie zmienia to faktu, że są one niezwykle wszechstronne i można je zastosować do innych problemów o naturze optymalizacyjnej.

Dodatkowo warto zaznaczyć, że poniższa lista w żadnym wypadku nie jest wyczerpująca, ponieważ omówienie wszystkich metod heurystycznych, które można wykorzystać w problemie komiwojażera jest niemalże niemożliwe z uwagi na ogromną liczbę tych metod.

3.1.1. Algorytm mrówkowy

Pomimo tego jak zaskakujące się to wydaje, jeśli chodzi o problem komiwojażera, zaskakująco dużo jesteśmy w stanie nauczyć się od niepozornych, malutkich stworzeń – mrówek. Podczas swoich wędrówek, insekty te pozostawiają na swoich ścieżkach feromony. Służą one za wskazówki dla innych członków kolonii, którą ścieżką należy podążać (pasjonującą opowieść o eksperymentach przeprowadzanych na mrówkach można znaleźć w rozdziale „The Amateur Scientist" w [20]). W pewnym uproszczeniu, im

więcej feromonów znajduje się na danej ścieżce, tym większe prawdopodobieństwo tego, że kolonia mrówek postanowi sprawdzić daną ścieżkę.

W latach 90. kilku autorów [21-24] uzmysłowiło sobie, że jesteśmy w stanie naśladować feromonowy trik stosowany przez mrówki do rozwiązywania (w sposób przybliżony) pewnych problemów, które w swoim sformułowaniu polegają na pojęciu odległości, którą należy minimalizować. W szczególności jest to aplikowalne do TSP – technika ta została nazywana algorytmem mrówkowym (ang. *ant colony optimization*, w skrócie ACO).

Sama technika polega na rozrzuceniu wirtualnych mrówek po n miastach, które składają się na rozpatrywaną instancję problemu komiwojażera. Każdej mrówce nakazujemy odwiedzić każde z miast dokładnie raz, a następnie wrócić bezpośrednio do punktu wyjścia. Po ukończeniu swojej podróży pamięć mrówki jest czyszczona, a sama mrówka jest gotowa na kolejną podróż.

W trakcie swojej wędrówki, każdy z sztucznych insektów zapamiętuje odwiedzone już miasta – wyklucza to możliwość wielokrotnego odwiedzenia tego samego miasta. Ponadto, mrówki zostawiają ślad feromonowy na każdej drodze, przez którą przechodzą – w matematycznej notacji drogi te odpowiadają pojedynczym krawędziom w grafie pełnym, opisującym rozpatrywaną instancję TSP. Ilość feromonu pozostawiana na każdej przebytej ścieżce między dwoma miastami jest taka sama – ale z uwagi na to, że krawędzie różnią się między sobą długością, to stężenie feromonu na poszczególnych trasach będzie różne (im dłuższa trasa, tym mniejsze zagęszczenie feromonu).

Po dotarciu do miasta, każda z mrówek jest w stanie wybrać jedną ze ścieżek, prowadzących do kolejnych, nieodwiedzonych dotychczas miast. Ów insekt z dużym prawdopodobieństwem wybierze ścieżkę o największym zagęszczeniu feromonów. Jednocześnie najmniej prawdopodobnym jest wybór miasta, do którego ścieżka ma najmniejsze ich zagęszczenie.

Okazuje się, że wraz z upływem czasu (mierzonego liczbą kolejnych iteracji) nasze sztuczne mrówki będą w znakomitej większości poruszać się po cyklu Hamiltona, który jest stosunkowo dobrym przybliżeniem rozwiązania TSP.

Szczegółowy opis algorytmu mrówkowego w kontekście nieco innego problemu można znaleźć m.in. w [25].

3.1.2. Algorytm świetlikowy

Mrówki nie są jedynym gatunkiem insektów, od którego matematycy i informatycy są w stanie wiele się nauczyć. Równie interesujące pod pewnymi względami okazują się chrząszcze z rodziny *Lampyridae*, znane pod nazwą *robaczki świętojańskie* lub, bardziej kolokwialnie, *świetliki*. W toku ewolucji te intrygujące owady wykształciły w swoich odwłokach organy zdolne do emitowania światła (ciekawym faktem jest to, że pigmenty umożliwiające tym stworzeniom bioluminescencję noszą dość niepokojącą nazwę: *lucyferyny*).

Jednym z głównych powodów, dla których świetliki są takie *błyskotliwe* jest możliwość przyciągnięcia potencjalnego partnera[5]. W 2008 Xin-She Yang [26] zaobserwował, że zachowania tych insektów może być kamieniem węgielnym dla algorytmu optymalizacyjnego, który rozwijał w kolejnych swoich pracach [27, 28] i który można znaleźć również w książce [29].

[5] W tym miejscu warto wspomnieć, że istnieją gatunki robaczków świętojańskich takie jak Photuris versicolor, które używają sygnałów świetlnych do wabienia swoich ofiar (zwykle innych świetlików).

Każdy z wirtualnych świetlików reprezentować będzie pewien cykl Hamiltona w rozpatrywanej przez nas instancji problemu komiwojażera na grafie (K_n, ω). Wygodnie jest rozważać reprezentowane przez nie trasy jako permutacje zbioru miast (odpowiadające kolejności ich odwiedzania).

Dystans pomiędzy świetlikami A i B definiujemy jako najmniejszą liczbę transpozycji niezbędną do przetransformowania permutacji przechowywanej przez robaczka A w permutację reprezentowaną przez B. Dodatkowo, każdy z insektów emituje *sztuczne światło*, którego natężenie zależy od wagi cyklu Hamiltona przechowywanego przez tego świetlika. Innymi słowy, intensywność wirtualnej bioluminescencji jest bezpośrednio powiązana z jakością reprezentowanego przez dany świetlik rozwiązania.

Założenie, które przyświeca interakcjom między świetlikami, stanowi o tym, że słabiej świecące świetliki są przyciągane do tych, które świecą jaśniej (artykuły [30, 31] zawierają obszerny opis tego, na czym polega ruch świetlików). Technicznie rzecz ujmując, mniej *błyskotliwe* świetliki zmieniają przechowywane przez siebie permutacje w taki sposób, by zbliżać się do permutacji reprezentowanych przez silnie świecące robaczki (vide [32]).

Podobnie jak w przypadku wirtualnych mrówek, sztuczne świetliki również stanowią ciekawy i skuteczny przykład wykorzystania inteligencji roju do przybliżania rozwiązań problemu komiwojażera.

Kończąc tę sekcję, pozwolimy sobie odnotować jeszcze jeden warty uwagi fakt, jakim jest możliwość wykorzystania algorytmu świetlikowego do dopasowywania hiperparametrów innych metaheurystyk – przykład takiego wykorzystania można znaleźć w artykule Ariyaratne, Fernando i Weerakoona [33].

3.1.3. Algorytm genetyczny

Algorytm ten, bazujący na procesie ewolucji naturalnej został zaprojektowany i spopularyzowany przez Johna Hollanda [34]. Z uwagi na ich wszechstronność, algorytmy genetyczne (GA) zapewniają stosunkowo dobrą jakość rozwiązań dla wielu rodzin problemów optymalizacyjnych [35-37]. Jest to jedna z najpopularniejszych metaheurystyk populacyjnych, która stała się punktem wyjścia i inspiracją dla wielu współczesnych metod ewolucyjnych (w tym również tych omawianych w dalszej części tego artykułu). Pomimo swoich licznych rozszerzeń i wariantów, algorytmy genetyczne pozostają efektywną metodą rozwiązywania wielu niezwykle złożonych problemów po dziś dzień.

Aby poprawnie wykorzystać potęgę algorytmu genetycznego, użytkownik musi zdefiniować pięć kluczowych dla działania komponentów:

- Schemat kodowania, tj. metoda reprezentowania rozwiązań dopuszczalnych dla rozpatrywanego problemu. W terminologii używanej w GA zakodowane rozwiązanie dopuszczalne nazywane jest zwykle *osobnikiem*.
- Funkcja dopasowania – mierzy jakość uzyskiwanego osobnika, zwykle jest to po prostu funkcja celu, którą optymalizujemy w ramach działania algorytmu.
- Operator krzyżowania – funkcja generująca nowego osobnika na podstawie dwóch innych. Wynikowy osobnik dziedziczy pewne charakterystyki od każdego z *rodziców*.
- Operator mutacji – funkcja wprowadzająca niewielką, losową zmianę w osobniku celem zdywersyfikowania zbioru rozpatrywanych rozwiązań dopuszczalnych. Jest też mechanizmem umożliwiającym wydostanie się z lokalnych optimów, które nie są jednocześnie globalnymi.

- Operator selekcji – jego rola zależy w istotny sposób od zastosowanej implementacji. Co do reguły, funkcja ta stanowi odpowiednik selekcji naturalnej, wykluczającej z populacji najmniej dostosowane osobniki (tj. te o najmniejszej wartości funkcji celu).

Dokładny opis algorytmu genetycznego wraz z dyskusją na temat roli każdego z opisanych wyżej komponentów można znaleźć m.in. w [29].

Sam algorytm zaczyna od utworzenia populacji losowo wygenerowanych rozwiązań dopuszczalnych. Następnie poddana ona zostaje serii *cykli genetycznych*. Pojedynczy cykl rozpoczyna się przez utworzenie nowego pokolenia jako wynik krzyżowań osobników z poprzedniej generacji. Z pewnym prawdopodobieństwem nowo uzyskane osobniki są poddawane mutacjom. Następnie każdy osobnik z nowej populacji podlega ocenie przez oszacowanie funkcji dopasowania. Na koniec operator selekcji pozbywa się najgorszych (pod względem funkcji dopasowania) osobników, zmniejszając z powrotem rozmiar populacji do odgórnie zamierzonej liczności. Oczywiście zaprezentowany powyżej schemat postępowania nie jest jedynym możliwym – wykorzystanie operatora selekcji w sposób prowadzący do innych rozwiązań niż pokoleniowe można znaleźć np. w [38].

Jeżeli zaś chodzi o wykorzystanie algorytmu genetycznego w rozwiązywaniu problemu komiwojażera, zasadniczym problemem, z którym należy zmierzyć się w pierwszej kolejności jest sposób kodowania osobników. Wyniki bowiem mogą drastycznie zależeć od wybranej przez nas reprezentacji cyklów Hamiltona. Badania nad efektywnością różnorodnych operatorów genetycznych oraz powiązanych z nimi schematów kodowania jest kontynuowane od kilku dekad i wiele wskazuje na to, że temat ten jeszcze dalece nie został wyczerpany – zaciekawionego Czytelnika odsyłamy do prac [39-43]. Warto też nadmienić, że GA może być również używany w połączeniu z innymi metodami heurystycznymi. Przykładowo, kombinacja algorytmu genetycznego i optymalizacji rojem cząsteczek do rozwiązywania TSP została opisana w [44].

3.1.4. Algorytm memetyczny

Algorytm memetyczny jest ulepszoną wersją algorytmu genetycznego. Został zaproponowany we wczesnych latach 90’ przez Moscato w celu zwalczenia pewnych niedoskonałości oryginalnej wersji algorytmu genetycznego [45-49].

Moscato zauważył, że algorytm genetyczny doskonale radzi sobie ze znajdywaniem dość dobrych rozwiązań dopuszczalnych, ale zazwyczaj przeciętnie radzi sobie z wprowadzaniem ostatecznych szlifów do uzyskanego rozwiązania (gdy najlepsze znalezione dotychczas rozwiązanie dopuszczalne jest zbliżone do optymalnego). Nazwa opracowanej metody nawiązuje do konceptu memów – wyjaśnienie pochodzenia tego terminu można znaleźć w [50] (mimo pewnych podobieństw do koncepcji internetowych memów, jest to pojęcie naukowe). Mem można postrzegać jako pewien informacyjny/kulturowy odpowiednik pojęcia genu. W przeciwieństwie do genów, memy ewoluują w trakcie życia posiadającego je osobnika. Ponadto, każdy osobnik dodaje swojego osobistego charakteru do memu – można to rozumieć przez własną interpretację pojęcia, własne idee powiązane z konkretnym memem itd. Podobnie jak w przypadku genów, idee też ulegają procesowi selekcji naturalnej i są podatne na fenomeny takie jak mutacja i rozprzestrzenianie się w populacji (obszerną dyskusję na ten temat można znaleźć m.in. w pracach [51,52]).

Algorytm memetyczny może być zatem postrzegany jako dodatkowa warstwa zachowań poszukiwań lokalnych wokół najlepszych uzyskanych dotychczas rozwiązań, stanowiąc interesującą kombinację tych dwóch podejść. Precyzyjniej, różnica pomiędzy GA i MA leży w przebiegu pojedynczego cyklu genetycznego. Standardowy cykl genetyczny składa się z krzyżowania, mutacji i selekcji. Algorytm memetyczny dokłada rozwój personalny każdego z osobników poprzez metody przeszukiwania lokalnego. Dobór konkretnej techniki przeszukiwań lokalnych zależy przede wszystkim od specyfiki rozwiązywanego problemu, ale zazwyczaj jest pewnego rodzaju podejściem wspinaczkowym. Algorytmów memetycznych można używać nie tylko do symetrycznych instancji problemu komiwojażera – studium zastosowania MA do asymetrycznego TSP można znaleźć w [45], jak również w zawartych tam referencjach.

3.1.5. Sieci neuronowe

Sieci neuronowe stanowią potężne i wszechstronne narzędzie, zdolne do rozwiązywania wielu problemów optymalizacyjnych w całej rzeszy różnorodnych gałęzi nauki. Pomimo tego, że ich główne zastosowanie leży w analizie danych i szeroko pojmowanej sztucznej inteligencji, w 1985 roku Hopfield i Tank z powodzeniem wykorzystali sieć neuronową do rozwiązywania problemu komiwojażera [53]. Ich pierwsze podejście nie było zwieńczone wybitnymi sukcesami, co potwierdza reprodukcja eksperymentu przez Wilsona i Pawley'a, wykazująca, że zaproponowana w oryginalnej pracy architektura sieci często nie potrafiła wygenerować dopuszczalnego rozwiązania dla zadanego problemu – szczegóły tej analizy można znaleźć w pracy [54]. Nawet gdy sieć Hopfielda i Tanka uzyskiwała dopuszczalne rozwiązanie, jego jakość była porównywalna z uzyskiwanym w kilku losowych próbach rozwiązaniem [54]. Po części wynikało to z faktu, że sieci zaimplementowane według tej pierwszej koncepcji w dużej mierze polegały na funkcjach kary, które miały na celu znaczące obniżanie jakości niepoprawnych rozwiązań. Obszerną dyskusję na ten temat można znaleźć w artykule Potvina [55], jak również w referencjach zawartych w owej pracy.

Inne podejście, dużo bardziej intuicyjne z ludzkiej perspektywy zasugerowane zostało przez Durbina i Willshawa w [56]. Ich próba bazowała na samoorganizującej się sieci Kohonena (więcej o tych sieciach można znaleźć w oryginalnej pracy Kohonena [57]). Obrazowo można opisać to podejście jako metodę *gumki recepturki*. Okrągłą opaskę rozpiętą na $M > n$ punktach (M jest pewną liczbą większą od liczby n miast rozpatrywanych w konkretnej instancji problemu komiwojażera) umieszczamy na centralnej części mapy opisującej położenia poszczególnych miast

Pozycje owych M punktów rozpinających naszą opaskę są modyfikowane pod wpływem dwóch, do pewnego stopnia, przeciwstawnych sił. Pierwsza z nich przesuwa punkty w kierunku miast (na zewnątrz) celem objęcia całokształtu trasy, druga zaś odpowiada za ściskanie naszej opaski, sprawiając, że tworzona ścieżka pozostaje możliwie krótka. Metoda ta jest niestety dość ograniczona do dwuwymiarowego problemu komiwojażera w wersji Euklidesowej [55]. Inne udoskonalenia tej metody dotyczące sposobów aktualizacji pozycji punktów rozpinających *opaskę* można znaleźć w [58-61]. Przykłady, porównania i testy wydajnościowe można znaleźć również w [62, 63].

3.1.6. Symulowane wyżarzanie

„Wyżarzanie" jest terminem wywodzącym się z metalurgii i materiałoznawstwa. Opisuje on proces zastosowania obróbki cieplnej do konkretnego materiału, polegający na jego rozgrzaniu, podtrzymywaniu wysokiej temperatury, a następnie powolnym schłodzeniu materiału do wyjściowej ciepłoty. Proces ten służy eliminacji wszelkiego rodzaju skaz i nieczystości, które mogą występować w wyżarzanym materiale.

Symulowane wyżarzanie (SA, ang. *Simulated Annealing*) jest algorytmem naśladującym ostatni etap tego procesu. Szeroki opis badań nad zastosowaniami tej metody można znaleźć w pracy [64], gdzie autorzy używają jej również do wytwarzania przybliżonych rozwiązań problemu komiwojażera. Procedura ta zaczyna się od wygenerowania pewnego rozwiązania dopuszczalnego x_c. Początkowa i końcowa *temperatura* w owym procesie musi zostać ustalona na samym początku działania algorytmu. Pierwsza z tych wartości zazwyczaj jest ustalana jako 1, a końcowa ciepłota zazwyczaj jest pewną ujemną potęgą liczby 10. Algorytm ten stopniowo przechodzi w kolejnych iteracjach od początkowej wartości temperatury do końcowej (przejście to rzadko kiedy jest liniowe i mnóstwo tzw. schematów wychładzania jest możliwe do wykorzystania w tym celu – jako punkt odniesienia może posłużyć praca [65]). Na każdym etapie tego procesu pewna liczba technik optymalizacyjnych jest wykonywana. Każda taka podprocedura składa się ze znalezienia pobliskiego rozwiązania x_{nc} (utworzonego poprzez wprowadzenie niewielkiej perturbacji w aktualnie rozważanym rozwiązaniu dopuszczalnym). Następnie, zastępujemy dotychczasowe rozwiązanie z prawdopodobieństwem

$$p_{x_{nc}} := \min\left\{\exp\left(\frac{f(x_c) - f(x_{nc})}{t}\right), 1\right\}$$

gdzie: f jest funkcją kosztu pojedynczego rozwiązania, zaś t jest temperaturą. W niektórych przypadkach ułamek pojawiający się w tym wyrażeniu obejmuje również w mianowniku stałą Boltzmanna. Równanie to, stanowiące kluczowy element procedury symulowanego wyżarzania pochodzi z artykułu [66].

Warto nadmienić, że początkowe rozwiązanie nie musi być losowe. Może ono zostać utworzone z wykorzystaniem jakiejś formy heurystyki konstrukcyjnej, o ile takie podejście jest możliwe w rozważanym problemie. Pomysł ten jest omawiany w kontekście TSP w dalszej części tego artykułu, jak również w sekcji 6.2.1 pracy [67].

Zazwyczaj najtrudniejszy element układanki, jaką stanowi algorytm symulowanego wyżarzania stanowi zdefiniowanie metody perturbacji. Zazwyczaj badacze [68, 69] używają jednej z wielu uznanych heurystyk optymalizacyjnych, typu k-OPT lub metody Lin-Kernighana, które będą omawiane w dalszej części tego artykułu. Tym, co odróżnia SA od standardowego zachowania metod optymalizacji lokalnej jest fakt, iż modyfikacje wprowadzane do aktualnie rozpatrywanego rozwiązania dopuszczalnego mogą je pogarszać, zamiast polepszać. Wydawać się to może dość nieintuicyjne, by akceptować zmiany pogarszające jakość rozwiązania, ale pozwala to na wydostanie się z minimów lokalnych.

3.1.7. Przeszukiwanie tabu

Przeszukiwanie tabu[6] to powszechnie znana metaheurystyka, która z zasady ma przezwyciężać słabe strony metody iterowanego wyszukiwania lokalnego. Panuje konsensus [70], że wyszukiwanie tabu zostało wprowadzone przez Freda Glovera [71] i spopulary-

[6] Za Encyklopedią Britannica: słowo tabu", czasem pisane jako taboo", to słowo pochodzenia polinezyjskiego oznaczające zachowanie, które uznawane jest za zbyt święte albo zbyt wyklęte by zwykły człowiek mógł je pojąć.

zowane w jego dalszych pracach [72-74]. O ile iterowane przeszukiwanie lokalne zatrzymuje się po znalezieniu lokalnego optimum, natomiast TS w pewnym sensie *zawraca*, próbując wybrać inną kombinację ruchów, która potencjalnie może doprowadzić do lepszego rozwiązania niż dotychczas uzyskane.

Procedura zaczyna się od wyboru początkowego rozwiązania. Podobnie jak w poprzednich metodach, może ono – choć nie musi – być losowe, ale może być też wynikiem działania innej metody heurystycznej. Inicjalizujemy tzw. listę tabu jako pustą listę. Następnie wykonywana jest pewna liczba iteracji – każda z nich polega na wyborze najlepszego *sąsiadującego* rozwiązania dopuszczalnego, które nie znajduje się na liście tabu. Podobnie jak w poprzednich technikach pojęcie sąsiedztwa rozwiązania dopuszczalnego jest definiowane przez użytkownika i zazwyczaj polega na możliwości uzyskania innego rozwiązania przy pomocy pojedynczego przyłożenia operacji mutującej [75]. Dotychczasowe rozwiązanie jest umieszczane na liście tabu. W ten sposób algorytm unika utknięcia w optimach lokalnych, jak również jednocześnie uniemożliwia mu to *krążenie* wokół kilku wielokrotnie odwiedzonych rozwiązań. Aby heurystyka ta pozostawała efektywna czasowo, lista tabu powinna mieć pewną odgórnie ustaloną pojemność $\mathcal{C}$. Po jej przekroczeniu, tj. w sytuacji, gdy $\mathcal{C}$ rozwiązań znajduje się na tej liście, kolejny dodany obiekt nadpisze najstarszą pozycję listy. Nie należy jednak zbytnio ograniczać wartości $\mathcal{C}$ – zbyt krótka lista tabu może sprawić, że algorytm utraci swoją odporność na powracanie do pułapki pojedynczego optimum lokalnego. Dyskusję na temat samego rozmiaru listy tabu można znaleźć np. w pracy [76].

Bardziej wysublimowaną techniką niż operowanie rozmiarem listy, którą możemy zastosować w celu usprawnienia działania TS są tzw. kryteria aspirujące. Kryteria te mają na celu oszacowanie czy *złamanie tabu*, tj. zignorowanie obecności danego rozwiązania na liście tabu może być korzystne dla działania algorytmu. Może mieć to miejsce w sytuacji, gdy dane rozwiązanie dopuszczalne można poprawić na kilka różnych sposobów, co otwiera nam drogę do nowych, nieeksploatowanych dotychczas fragmentów przestrzeni poszukiwań. Gdy kryteria te są spełnione, pozwalają ominąć ograniczenia narzucane przez listę. Czytelnika zainteresowanego bardziej szczegółowym opisem odsyłamy do prac Glovera, jak również do artykułu Maleka i innych [77].

3.1.8. Przeszukiwanie zmiennego sąsiedztwa

Hansen i Mladenovic̀ [78, 79] w swoich pracach zaproponowali interesującą technikę metaheurystyczną (o nazwie przeszukiwanie zmiennego sąsiedztwa, ang. *Variable Neighborhood Search*, VNS) w swoich pracach, która bazowała na wprowadzeniu pewnych poprawek do metody iterowanego przeszukiwania lokalnego [80]. W pracy [81] autorzy podejmują się wykorzystania jej w problemie komiwojażera, zarówno w wersji symetrycznej, jak też asymetrycznej. W obydwu przypadkach zastosowana metoda okazała się satysfakcjonująco skuteczna. Co do zasady, technika ta opiera się o eksplorację otoczenia najlepszego dotychczas znalezionego rozwiązania. Otoczenie to jest wyznaczane przez jedną z kilku technik modyfikujących nieznacznie rozpatrywane rozwiązanie – poprawna implementacja VNS dysponuje kilkoma takimi technikami. Po znalezieniu minimum lokalnego z wykorzystaniem jednej techniki, algorytm zmienia metodę modyfikacji rozwiązania i z jej pomocą stara się znaleźć minimum lokalne, eksplorując otoczenie najlepszego dotychczas znalezionego rozwiązania.

Technika ta jest w dużej mierze elastyczna i może być jeszcze bardziej efektywna w połączeniu z algorytmem, który jest w stanie wygenerować zadowalający punkt początkowy.

3.2. Heurystyki konstrukcyjne

Poniższa sekcja poświęcona będzie uzyskiwaniu przybliżonych rozwiązań problemu komiwojażera konstruowanych *od podstaw*. W niniejszym artykule ograniczymy się do krótkiego ich zestawienia, bazującego w dużym stopniu na [82], gdzie można znaleźć bardziej obszerne omówienie większości poniższych metod. Niemniej, przy pomocy algorytmów z tej części naszej listy można uzyskać dość dobre rozwiązania dopuszczalne, które mogą posłużyć za punkt wyjścia dla innych technik udoskonalających – choćby tych opisanych w pozostałych częściach poniższej listy.

3.2.1. Metoda najbliższego sąsiada

Metody najbliższego sąsiada bazują na prostym podejściu zachłannym, w którym komiwojażer będzie za każdym razem kierował się do najbliższego nieodwiedzonego dotychczas miasta. Założenie to dość szybko jednak prowadzi nas do coraz bardziej ograniczonego wyboru kolejnych lokacji. To postępujące zawężenie zbioru możliwości nierzadko zmusza konstrukcję metodą najbliższego sąsiada do uwzględnienia krawędzi o bardzo wysokim koszcie.

Ta niedoskonałość równoważona jest dużą szybkością działania rozpatrywanej metody. Przeciętnie czas jej wykonania wynosi $O(n^2)$, aczkolwiek przeprowadzenie wstępnych obliczeń może zmniejszyć tę wartość. Zaciekawionego czytelnika odsyłamy do stosownej literatury [82].

W przypadku metrycznych instancji problemu komiwojażera, istnieje logarytmiczne ograniczenie na stosunek wagi uzyskanego przybliżenia do wagi optymalnego rozwiązania naszego problemu, o czym mówi pierwsze twierdzenie w pracy Rosenkrantza i in. [83]. Precyzyjniej, stosunek ten nie może przekroczyć wartości $0.5 + \lceil \log_2(n) \rceil$, gdzie n to liczba miast w rozważanym problemie komiwojażera, zaś $\lceil a \rceil$ jest tzw. *sufitem* z liczby, tj. najmniejszą liczbą całkowitą, nie mniejszą niż a. W tym samym artykule autorzy wykazują, że dla zadanego $R \geq 1$ i odpowiednio dużej wartości n, istnieje instancja TSP dla której stosunek wagi przybliżenia do rozwiązania optymalnego przekracza R. Prostsze konstrukcje analogicznych przykładów można znaleźć w pracy Hurkensa i Woegingera [84].

3.2.2. Techniki wstawiające

Poprzednio omawiana metoda może być zobrazowana jako rysowanie łamanej krok po kroku. Metody omawiane w tej podsekcji stanowią interesującą alternatywę dla takiego podejścia. Rozpoczynając od wyboru pojedynczej krawędzi, heurystyki wstawiające produkują sekwencję cykli, z których każdy zawiera wszystkie wierzchołki występujące w poprzedzającym go cyklu i pojedynczy dodatkowy wierzchołek. Innymi słowy, heurystyki wstawiające można zilustrować jako poszerzanie dotychczas uzyskanego cyklu o jedno *ogniwo* na raz. Ideą przyświecającą temu podejściu jest wstawianie nowego wierzchołka do istniejącego cyklu w taki sposób, by powiększał wagę nowo utworzonego cyklu o możliwie najmniejszą wartość. Proces ten kończy się w momencie gdy żaden wierzchołek nie jest już dostępny, tj. uzyskany końcowy cykl jest cyklem Hamiltona w wyjściowym grafie. W artykule [83] autorzy dowodzą (Thm. 3), że w przypadku metrycznym heurystyki wstawiające dla problemu komiwojażera również mają logarytmiczne ograniczenie na stosunek błędu aproksymacji (i podobnie jak poprzednio, ograniczenie to zależne jest od rozmiaru grafu).

Schemat dodawania wierzchołków do istniejącego cyklu jest dość prosty. Zdecydowanie bardziej intrygującą kwestią jest sposób, w jaki heurystyka ustalać będzie kolejność, w której wierzchołki dodawane będą do rozrastającego się cyklu. W poniższym zestawieniu prezentujemy pięć najważniejszych podtypów heurystyk wstawiających:

1. metoda najbliższego sąsiada;
2. metoda najtańszej inkluzji;
3. metoda losowa;
4. metoda najdalszego sąsiada;
5. metoda otoczki wypukłej.

Dyskusję na temat ich sposobu działania rozpoczynamy od omówienia metody *najbliższego sąsiada*. W tym podejściu na każdym etapie konstrukcji kolejny dobierany wierzchołek minimalizuje odległość od najbliższego wierzchołka już istniejącego cyklu.

Druga metoda na tej liście, tj. metoda *najtańszej inkluzji* nieznacznie różni się od poprzedniej techniki. W tym podejściu kolejne wierzchołki dobierane są w taki sposób, by wstawienie go do obecnie rozpatrywanej konstrukcji było możliwie najtańsze. Po więcej szczegółów odsyłamy zaciekawionego czytelnika do artykułu Rosenkrantza [83], jednocześnie podkreślając, że udoskonalona wersja tej techniki wraz z opisem znajduje się również w [85].

Technika *losowa* wydaje się być dość wymowną nazwą, nie wymagającą obszernych wyjaśnień. W każdym kroku wybieramy losowy wierzchołek, który w optymalny sposób dodajemy do aktualnie konstruowanego cyklu Hamiltona. Pomimo bycia dość losową, chaotyczną metodą, rezultaty jej działania są bardzo często porównywalne lub lepsze od dwóch uprzednio sformułowanych rozwiązań heurystycznych. Jednocześnie zachęcamy do zapoznania się z artykułem [86] porównującym różnorakie heurystyki wstawiające, jak również kilka innych technik uzyskiwania rozwiązań przybliżonych.

Przedostatnia heurystyka uwzględniona na tej liście jest dość sprzeczna z intuicją na pierwszy rzut oka. W metodzie *najdalszego sąsiada* zaczynamy od wyboru losowej, zazwyczaj jednej z najdłuższych krawędzi jako punktu wyjścia. W pojedynczym etapie konstrukcji dla każdego nieuwzględnionego jeszcze w cyklu wierzchołka obliczamy jego odległość od dotychczasowej konstrukcji (rozumianej jako jego odległość od najbliższego wierzchołka znajdującego się już w cyklu). Następnie wybieramy wierzchołek, który minimalizuje ową odległość i dodajemy go do tworzonego rozwiązania. Pomimo dość nieoczywistego kryterium, które zostało zastosowane w tym podejściu, metoda ta daje jedne z najlepszych rezultatów wśród wszystkich heurystyk wstawiających. Badacze spekulują, że dobra jakość uzyskiwanych rozwiązań dopuszczalnych jest konsekwencją faktu, że w owej metodzie już w początkowych krokach tworzony jest ogólny, surowy zarys ogólnego kształtu trasy, która będzie następnie dopracowywana.

Idea ta jest też punktem wyjścia dla ostatniej z omawianych w tej części technik, jaką jest *heurystyka otoczki wypukłej*. W odróżnieniu od poprzednich heurystyk, w tej metodzie rozpoczynamy od wygenerowania otoczki wypukłej na zbiorze wszystkich wierzchołków. Tak uzyskana figura służy nam za punkt wyjścia w tej technice. Kolejne wierzchołki dodawane są do tak utworzonego cyklu zgodnie z odgórnie ustalonym schematem (zwykle jedną z czterech omówionych uprzednio metod). Do aktualnych badań nad tą i innymi heurystykami wstawiającymi, czytelnika odsyłamy do pracy [87].

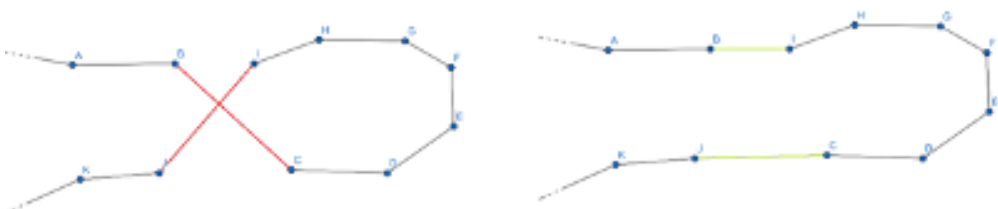
3.3. Techniki optymalizacyjne dla rozwiązań dopuszczalnych (heurystyki udoskonalające)

Dysponując rozwiązaniem dopuszczalnym rozpatrywanego problemu, które jest już dość dobrej jakości, można sobie zadać pytanie czy możliwym jest dalsze jego poprawienie. Heurystyki udoskonalające udzielają na to pytanie pozytywnej (zazwyczaj) odpowiedzi, wprowadzając małe zmiany w strukturze rozwiązania, które poprawiają jego jakość. W ramach tego zestawienia przedstawione zostaną dwie takie metody:

3.3.1. Metody k-OPT

Według naszej najlepszej wiedzy, heurystyki k-OPT stanowią jedne z najwcześniejszych dokonań w dziedzinie optymalizacji lokalnej. Ta konkretna rodzina heurystyk została zaproponowana przez Shena Lina w artykule [88]. Autor ten w swojej publikacji badał metody uzyskiwania 2- i 3-optymalnych tras komiwojażera (warto podkreślić, że Lin w swojej pracy [88] rozpatrywał też problem uzyskiwania tras 4-optymalnych, co okazało się być zatrważająco czasochłonne, a do tego przynosiło stosunkowo niewielkie zyski). W tym miejscu należałoby wyjaśnić, jak rozumieć trasę k-optymalną – do tego będzie potrzebna definicja ruchu k-OPT. Dla prostoty omówimy tutaj jedynie przykład ruchu 2-OPT, ponieważ dla wyższych wartości k drastycznie rośnie liczba możliwych rezultatów takich ruchów, a co za tym idzie – konieczność rozpatrzenia większej liczby możliwości. Ciekawego Czytelnika odsyłamy do pracy [89], gdzie wszystkie możliwości ruchów 3-OPT można zobaczyć rozrysowane na rysunku 1.

Skrótowo, ruch 2-OPT dla rozpatrywanej trasy komiwojażera polega na wzięciu pary należących do niej krawędzi i zamiany ich końców, jeżeli poprawia to jakość uzyskiwanego cyklu Hamiltona. W przypadku Euklidesowym na płaszczyźnie można pomyśleć, że ruch ten czyni trasę wolną od 'węzłów' lub 'pętelek', co dość dobrze ilustruje poniższy rysunek 1.



Rysunek 1. Przykładowa ścieżka przed i po wykonaniu ruchu 2-OPT. Po zastąpieniu czerwonych krawędzi przez zielone, trasa uległa widocznemu skróceniu [opracowanie własne]

3.3.2. Heurystyka Lin-Kernighana

Heurystyka Lin-Kernighana [90] jest pewnym krokiem naprzód względem metod k-OPT. Metody k-OPT mają złożoność rzędu $\mathcal{O}(n^k)$ – dla dużych wartości n nasz dobór k zaczyna być mocno ograniczany przez czas wykonania. Heurystyka omawiana w tej części artykułu stanowi swoiste remedium dla tego problemu – elastyczne modyfikowanie wartości k dla techniki k-OPT wraz z innymi udoskonaleniami nie tylko gwarantują wysoką jakość uzyskanej trasy (przeprowadzenie pełnej procedury opisanej w oryginalnej pracy Lina i Kernighana gwarantuje, że trasa jest co najmniej tak dobra, jak po przeprowadzeniu pełnej procedury 3-OPT), ale również jest stosunkowo szybka.

Dla niektórych języków programowania (np. Python) istnieją biblioteki open-source, implementujące tę technikę.

Podobnie jak z większością poprzednich technik, istnieją jej liczne rozszerzenia i usprawnienia. Serię bardzo interesujących artykułów na ten temat napisał K. Helsgaun, [91-93] do lektury których odsyłamy zaciekawionego Czytelnika.

4. Wnioski i podsumowanie

W ramach poniższej pracy dokonano przeglądu i pogrupowania metod aproksymacyjnych dla problemu komiwojażera. W części poświęconej inteligentnym heurystykom przedstawiono 8 metod uzyskiwania przybliżonych rozwiązań problemu komiwojażera. Pierwsze cztery (algorytm mrówkowy, genetyczny, memetyczny i świetlikowy) stanowią tzw. heurystyki populacyjne. W celu uzyskania zbieżności zachodzi konieczność generowania dużej liczby permutacji, co sprawia, że techniki te są kosztowne czasowo, gdy w rozpatrywanej instancji problemu liczba punktów zaczyna przekraczać dziesiątki tysięcy punktów. Jeśli nie mamy możliwości zrównoleglania swoich obliczeń staje się to tym bardziej dotkliwym problemem. Dla większych instancji lepszym wyborem stają się omawiane w trzech ostatnich podsekcjach pierwszej części heurystyczne metody wykorzystujące mechanizm przeszukiwania lokalnego. Z kolei w opinii znaczącej większości autorów, sieci neuronowe nie są do końca odpowiednim narzędziem do rozwiązywania TSP, pomimo możliwości ich dostosowania do problemów optymalizacji kombinatorycznej. Jest to wypadkowa jakości uzyskiwanych rozwiązań (gorszej lub porównywalnej względem pozostałych inteligentnych metod) oraz przeciętnej skalowalności.

Druga część poświęcona została heurystykom konstrukcyjnym, gdzie przedstawiono dwie ogólne metody wraz z kilkoma ich wariantami. O ile metody zachłanne, prezentowane w pierwszej kolejności, są bardzo wydajne czasowo nawet dla bardzo dużych instancji, o tyle metody wstawiające dają zdecydowanie lepsze rezultaty, pomimo możliwego dłuższego czasu wykonania. Jedne i drugie mogą służyć za doskonałe narzędzia do tworzenia rozwiązań, będących punktami wyjścia dla metod bazujących na lokalnym przeszukiwaniu czy udoskonalaniu posiadanego już rozwiązania. Same w sobie mają często dość ograniczoną przestrzeń poszukiwań, co istotnie ogranicza ich uniwersalność jako samodzielnych metod.

Trzecia część naszych rozważań przybliżała wybrane metody udoskonalające już istniejące rozwiązanie dopuszczalne. Podobnie jak techniki przeszukiwania lokalnego, ich punktem wyjścia mogło być rozwiązanie losowe, ale znacznie lepsze rezultaty uzyskać można poprzez połączenie przedstawionych w tej sekcji metod z dowolnymi heurystykami przedstawianymi w ramach pierwszych dwóch grup. Metody 2-OPT i 3-OPT stanowią niejako pewien standard obecnie stosowanych metod heurystycznych i są zwykle aplikowane domyślnie do rozwiązań generowanych innymi technikami.

Przygotowany przegląd nie tylko przybliża sposoby działania omawianych metod, ale również oferuje znaczącą liczbę odnośników do literatury, która umożliwia Czytelnikowi dalszą eksplorację wybranych fragmentów tej tematyki. Warto podkreślić, że powyższe zestawienie stanowi jedynie ułamek z setek znanych i sprawdzonych metod aproksymacyjnych dla problemu komiwojażera. Pomimo tak znaczącego wyeksploatowania tematu, TSP i przybliżone metody jego rozwiązywania nadal stanowią wartościowy i ciekawy materiał badawczy. Wynika to nie tylko z istotności samej tematyki, ale przede wszystkim z powodu uniwersalności wypracowywanych podejść i możliwości zastosowania proponowanych metod również do innych problemów.

Uwagi ogólne/Podziękowania

Autorzy pragną wyrazić wdzięczność pracownikom Fundacji Tygiel za cierpliwość i wsparcie w procesie składania stosownej dokumentacji. Ponadto, składają też wyrazy szacunku dla p. Jana Filipowicza za wiele interesujących i głębokich dyskusji, które pozytywnie wpłynęły na jakość wypracowanego artykułu.

Filip Turoboś jest finansowany z 2% subsydium dla szkolnictwa wyższego dla uczelni biorących udział w konkursie IDUB ('Inicjatywa Doskonałości – Uczelnia Badawcza'), jak również jest finansowany w ramach grantu „FU^{2}N – Fundusz Udoskonalania Umiejętności Młodych Naukowców" wspierającego doskonałość naukową pracowników Politechniki Łódzkiej, nr grantu: 4/2023.

Literatura

1. Krukowski M., Turoboś F., *Comparison of approximation algorithms for the travelling salesperson problem on semimetric graphs*, arXiv:2108.13070 [math.CO].
2. Diestel R., *Graph theory*, Springer-Verlag, New York 2000.
3. Fletcher P., Hoyle H., Patty C.W., *Foundations of Discrete Mathematics*, PWS-Kent Pub. Co., Boston, 1991.
4. Bondy A., Murty M.R., *Graph Theory, Graduate Texts in Mathematics 244*, Springer-Verlag, London 2008.
5. Gross J.L., Yellen J., Anderson M., *Graph Theory and Its Applications, Third Edition*, Chapman and Hall/CRC, 2018.
6. Dantzig G.B., Fulkerson D.R., Johnson S.M., O*n a Linear-Programming, Combinatorial Approach to the Traveling-Salesman Problem*, Operations Research, 7(1), 1959, s. 58-66.
7. Laporte G., *The traveling salesman problem: An overview of exact and approximate algorithms*, European Journal of Operational Research, 59(2), 1992, s. 231-247.
8. Matai R., Singh S., Mittal M.L., *Traveling Salesman Problem: an Overview of Applications, Formulations, and Solution Approaches*, [w:] Davendra D. (red.), *Traveling Salesman Problem, Theory and Applications*, IntechOpen, 2010.
9. Miller C.E., Tucker A.W., Zemlin R.A., *Integer Programming Formulation of Traveling Salesman Problems*, J. ACM, 7(4), 1960, 326-329.
10. Ausiello G., Marchetti-Spaccamela A., Crescenzi P., Gambosi G., Protasi M., Kann V., *Complexity and Approximation*, Springer, Berlin, Heidelberg, 1999.
11. Anddreasson N., Evgrafov A., Patriksson M., *An Introduction to Continuous Optimization: Foundations and Fundamental Algorithms. Revised Third edition*, Dover Publications, 2020.
12. Weise T., *Global Optimization Algorithm: Theory and Application*, https://archive.org/details/Thomas_Weise__Global_Optimization_Algorithms_Theory_and_Application, [data dostępu: 22.03.2023].
13. Crescenzi P., Kann V., Halldórsson M. Karpinski M., Woeginger G., *A compendium of NP optimization problems*, http://www.csc.kth.se/ viggo/problemlist/ [data dostępu: 21.03.2023].
14. Goldreich O., *Computational Complexity: A Conceptual Perspective*, Cambridge University Press, New York 2008.
15. Goldreich O.P, *NP, and NP-Completeness: The Basics of Computational Complexity*, Cambridge University Press, New York 2010.
16. Arora S., Barak B., *Computational complexity: a modern approach*, Cambridge University Press, 2009.
17. Du D.Z., Ko K.I., *Theory of computational complexity (Vol. 58)*, John Wiley & Sons, 2011.
18. Papadimitriou C.H., Steiglitz K., *Combinatorial optimization: algorithms and complexity*, Courier Corporation, 1998.

19. Cormen T.H., Leiserson C.E., Rivest R.L., Stein C., *Introduction to Algorithms (3rd ed.)*, London, England: MIT Press, 2009.
20. Feynman R.P., Leighton R., Hutchings E., Gates B., *"Surely you're joking, Mr. Feynman!": adventures of a curious character*, New York: W.W. Norton & Company, 2018.
21. Dorigo M., Maniezzo V., Colorni A., *Ant system: optimization by a colony of cooperating agents. IEEE Transactions on Systems*, Man and Cybernetics, Part B (Cybernetics), 26(1), 1996, s. 29-41.
22. Dorigo M, Ganbardella L., *Ant colony system: a cooperative learning approach to the traveling salesman problem*, IEEE Trans Evol Comput, 1(1), 1997, s. 53-66.
23. Aziz Z.A., *Ant Colony Hyper-heuristics for Travelling Salesman Problem*, Procedia Computer Science, 76, 2015, s. 534-538.
24. Yang J., Shi X., Marchese M., Liang Y., *An ant colony optimization method for generalized TSP problem*, Progress in Natural Science, 18(11), 2008, s. 1417-1422.
25. Shmygelska A., Hoos H.H., *An ant colony optimisation algorithm for the 2D and 3D hydrophobic polar protein folding problem*, BMC Bioinformatics, 6(1), 30, 2005, s. 1-22.
26. Yang X.S., *Nature-Inspired Metaheuristic Algorithms*, Bristol, UK, Luniver Press, 2008.
27. Yang X.S., *Firefly Algorithms for Multimodal Optimization*, [w:] Watanabe O., Zeugmann T. (red.), *Stochastic Algorithms: Foundations and Applications*, SAGA 2009. Lecture Notes in Computer Science, vol 5792. Springer, Berlin, Heidelberg, 2009, s. 169-178.
28. Yang X.S., *Firefly algorithm, stochastic test functions and design optimization*, International Journal of Bio-Inspired Computation, 2(2), 2010, s. 78-84.
29. Yang X.S., *Nature-Inspired Metaheuristic Algorithms*, Elsevier, London 2014.
30. Jati G.K., *Suyanto, Evolutionary Discrete Firefly Algorithm for Travelling Salesman Problem*, [w:] Bouchachia A. (red.), *Adaptive and Intelligent Systems*, ICAIS 2011. Lecture Notes in Computer Science, vol 6943. Springer, Berlin, Heidelberg, 2011.
31. Jati G.K., Manurung R., *Discrete Firefly Algorithm for Traveling Salesman Problem*, Swarm Intelligence and Bio-Inspired Computation, 2013, s. 295-312.
32. Kumbharana S.N., Pandey G.M., *Solving travelling salesman problem using firefly algorithm*, International Journal for Research in science & advanced Technologies, 2(2), 2013, s. 53-57.
33. Ariyaratne M.K.A., Fernando T.G.I., Weerakoon S., *A self-tuning firefly algorithm to tune the parameters of ant colony system*, International Journal of Swarm Intelligence, 3(4), 2018, s. 309-331.
34. Holland J.H., *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, MIT press, Cambridge 1992.
35. Canyurt O.E., Öztürk H.K., *Three different applications of genetic algorithm (GA) search techniques on oil demand estimation*, Energy Conversion and Management, 47(18-19), 2006, s. 3138-3148.
36. Ghaheri A., Shoar S., Naderan M., Hoseini S.S., *The applications of genetic algorithms in medicine*, Oman Medical Journal, 30(6), 2015, s. 406.
37. Ross P., Corne D., *Applications of genetic algorithms*, AISB Quaterly on Evolutionary Computation, 89, 1994, s. 23-30.
38. Goldberg D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Company, Reading, MA, 1989.
39. Abdoun O., Abouchabaka J., Tajani C., *Analyzing the Performance of Mutation Operators to Solve the Travelling Salesman Problem* International Journal of Emerging Sciences, 2, 2012, s. 61-77.
40. Grefenstette J., Gopal R., Rosmaita B., Van Gucht D., *Genetic Algorithms for the TSP*, Proceedings of the 1st International Conference on Genetic Algorithms and Their Applications, Lawrence Erlbaum Associates, Mahwah NJ, 1985, s. 160-168.

41. He R., Ma C., Li Y., Diao A., *Study on multi-objective travelling salesman problem for hazardous materials transportation based on improved genetic algorithm*, International Journal of Computing Science and Mathematics, 4(2), 2013, s. 131-139.
42. Larrañaga P., Kuijpers C.M.H., Murga R.H., Inza I., Dizdarevic S., *Genetic Algorithms for the Travelling Salesman Problem: A Review of Representations and Operators*, Artificial Intelligence Review, 13(2), 1999, s. 129-170.
43. Umbarkar A.J., Sheth P.D., *Crossover operators in Genetic Algorithm: a review*, ICTACT Journal of Soft Computing, 6(1), 2015, s. 1083-1092.
44. Liao Y.F., Yau D.H., Chen C.L., *Evolutionary algorithm to traveling salesman problems*, Computers & Mathematics with Applications, 64(5), 2012, s. 788-797.
45. Buriol L., França P.M., Moscato P., *A New Memetic Algorithm for the Asymmetric Traveling Salesman Problem*, Journal of Heuristics, 10(5), 2004, s. 483-506.
46. Moscato P., Norman M.G., *A Memetic Approach for the Traveling Salesman Problem. Implementation of a Computational Ecology for Combinatorial Optimization on Message-Passing Systems*, [w:] Valero M., Onate E., Jane M., Larriba J.L., Suarez B. (red.), Parallel Computing and Transputer Applications. IOS Press, 1992, s. 187-194.
47. Moscato P., Tinetti F., *Blending Heuristics with a Population-Based Approach: A Memetic Algorithm for the Traveling Salesman Problem*, CeTAD, Report 92-12. Universidad Nacional de La Plata, Argentina, 1992, s. 1-43.
48. Moscato P., *Memetic Algorithms: A Short Introduction*, [w:] Corne D., Dorigo M., Glover F. (red.), *New Ideas in Optimization*, McGraw-Hill, 1999, s. 219-235.
49. Moscato P., Cotta C., *A Gentle Introduction to Memetic Algorithms*, [w:] Glover F., Kochenberger G.A. (red.), *Handbook of Metaheuristics*, International Series in Operations Research & Management Science, Springer, Boston, MA, vol 57, 2003, s. 105-144.
50. Laurent J., *A Note on the Origin of 'Memes'/'Mnemes'*, Journal of Memetics, 3 (1), 1999, s. 14-19.
51. Zou P., Zhou Z., Chen G., Yao X., *A novel memetic algorithm with random multi-local-search: a case study of TSP*, Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753). 2004 Congress on Evolutionary Computation, 2004, s. 2335-2340.
52. Heylighen F., Chielens K., *Evolution of Culture, Memetics*, Encyclopedia of Complexity and Systems Science, 2009, s. 3205-3220.
53. Hopfield J.J., Tank D.W., *Neural computation of decisions in optimization problem*, Biological Cybernetics, 52, 1985, s. 141-152.
54. Wilson G.V., Pawley G.S., *On the Stability of the Travelling Salesman Problem Algorithm of Hopfield and Tank*, Biological Cybernetics, 58, 1988, s. 63-70.
55. Potvin J.Y., *State-of-the-art survey - the traveling salesman problem: A neural network perspective*, ORSA Journal on Computing, 5(4), 1993, s. 328-348.
56. Durbin R. Willshaw D., *An Analogue Approach to the Traveling Salesman Problem using an Elastic Net Method*, Nature 326, 1987, s. 689-691.
57. Kohonen T., *Self-Organized Formation of Topologically Correct Feature Maps*, Biological Cybernetics, 43(1), 1982, s. 59-69.
58. Boeres M.C.S., de Carvalho L.A.V., *A Faster Elastic Net Algorithm for the Traveling Salesman Problem, Proceedings of the Int. Joint Conf. on Neural Networks, Baltimore*, MD, 1992, s. 215-220.
59. Burr D.J., *An Improved Elastic Net Method for the Traveling Salesman Problem*, Proceedings of the IEEE Int. Conf. on Neural Networks, San Diego, CA, 1988, s. 69-76.
60. Chen J., Chen J., Zhang X., *Solving the Traveling Salesman Problem Using Elastic Net Integrate with SOM*, IEEE International Conference on Automation and Logistics, Jinan, 2007, s. 2234-2237.

61. Cochrane E.M., Beasley J.E., *The co-adaptive neural network approach to the Euclidean travelling salesman problem*, Neural Networks, 16(10), 2003, s. 1499-1525.
62. La Maire B.F.J., Mladenov V.M., *Comparison of neural networks for solving the travelling salesman problem*, IEEE Proceedings of the NEUREL 2012, Serbia, 2012, 1-5.
63. Skubalska-Rafajłowicz E., *Exploring the Solution Space of the Euclidean Traveling Salesman Problem Using a Kohonen SOM Neural Network*, Artificial Intelligence and Soft Computing. ICAISC 2017. Lecture Notes in Computer Science, Springer, Cham, vol 10245, 2017, s. 165-174.
64. Kirkpatrick S., Gelatt C., Vecchi M., *Optimization by Simulated Annealing*, Science, 220(4598), 1983, s. 671-680.
65. Collins N.E., Eglese R.W., Golden B.L., *Simulated Annealing – An Annotated Bibliography*, American Journal of Mathematical and Management Sciences, 8:3-4, 1988, s. 209-307.
66. Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., *Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics, 21(6), 1953, s. 1087-1092.
67. Eglese R.W., *Simulated annealing: A tool for operational research*, European Journal of Operational Research, 46(3), 1990, s. 271-281.
68. Aarts E.H., Korst J.H., van Laarhoven P.J., *A quantitative analysis of the simulated annealing algorithm: A case study for the traveling salesman problem*, Journal of Statistical Physics, 50(1-2), 1988, s. 187-206.
69. Tian P., Ma J., Zhang D.M., *Application of the simulated annealing algorithm to the combinatorial optimisation problem with permutation property: An investigation of generation mechanism*, European Journal of Operational Research, 118(1), (1999), s. 81-94.
70. Gendreau M., *An Introduction to Tabu Search*, [w:] Glover F., Kochenberger G.A. (red.), *Handbook of Metaheuristics*, International Series in Operations Research & Management Science, vol 57. Springer, Boston, MA, 2003, s. 37-54.
71. Glover F., *Future paths for integer programming and links to artificial intelligence*, Computers & Operations Research, 13(5), 1986, s. 533-549.
72. Glover F., *Tabu search – part I*, ORSA Journal on computing, 1(3), 1989, s. 190-206.
73. Glover F., *Tabu search – part II*, ORSA Journal on computing, 2(1), 1990, s. 4-32.
74. Glover F., Laguna M. *Tabu Search*, [w:] Du D.Z., Pardalos P.M. (red.), *Handbook of Combinatorial Optimization*, Springer, Boston, MA, 1998, s. 621-757
75. Knox J., *Tabu search performance on the symmetric traveling salesman problem*, Computers & Operations Research, 21(8), 1994, s. 867-876.
76. Tsubakitani S., Evans J.R., *Optimizing tabu list size for the traveling salesman problem*, Computers & Operations Research, 25(2), 1998, s. 91-97.
77. Malek M., Guruswamy M., Pandya M., Owens H., *Serial and parallel simulated annealing and tabu search algorithms for the traveling salesman problem*, Annals of Operations Research, 21, 1989, s. 59-84.
78. Hansen P., Mladenovic̀ N., *An introduction to variable neighborhood search*, [w:] *Meta-heuristics: Advances and Trends in Local Search Paradigms for Optimization*, Kluwer Academic Publishers, MA, USA, 1999, s. 433-458.
79. Hansen P., Mladenovic̀ N., *Variable neighborhood Search: Principles and Applications*, European Journal of Operational Research, 130, 2001, s. 449-467.
80. Gendreau M., Potvin J.Y. (red.), *Handbook of Metaheuristics*, International Series in Operations Research & Management Science, Springer, Cham, 2019.
81. Hore S., Chatterjee A., Dewanji A., *Improving variable neighborhood search to solve the traveling salesman problem*, Applied Soft Computing, 68, 2018, s. 83-91.
82. Reinelt G. *The Traveling Salesman – Computational Solutions for TSP Applications*, Springer-Verlag Berlin Heidelberg, 1994.

83. Rosenkrantz D.J., Stearns R.E., Lewis II, P.M., *An Analysis of Several Heuristics for the Traveling Salesman Problem*, SIAM Journal on Computing, 6(3), 1977, s. 563-581.
84. Hurkens C.A.J., Woeginger G.J., *On the nearest neighbor rule for the traveling salesman problem*, Operations Research Letters, 32(1), 2004, s. 1-4.
85. Hassin R., Keinan A., *Greedy heuristics with regret, with application to the cheapest insertion algorithm for the TSP*, Operations Research Letters, 36(2), 2008, s. 243-246.
86. Golden B., Bodin L., Doyle T., Stewart W., *Approximate Traveling Salesman Algorithms*, Operations Research, 28 (3, ii), 1980, s. 694-711.
87. Huang W., Yu J.X., *Investigating TSP Heuristics for Location-Based Services*, Data Sci. Eng., 2, 2017, s. 71-93.
88. Lin S., *Computer Solutions of the Traveling Salesman Problem*, Bell System Technical Journal, 44(10), 1965, s. 2245-2269.
89. Blazinskas A., Misevicius A., *Combining 2-opt, 3-opt and 4-opt with k-swap-kick perturbations for the traveling salesman problem*, Proceedings of the 17th International Conference on Information and Software Technologies, 2011, s. 45-50.
90. Lin S., Kernighan B.W., *An Effective Heuristic Algorithm for the Traveling-Salesman Problem*, Operations Research, 21, 1973, s. 498-516.
91. Helsgaun K., *An effective implementation of the Lin–Kernighan traveling salesman heuristic*, European Journal of Operational Research, 126 (1), 2000, s. 106-130.
92. Helsgaun K., *General k-opt submoves for the Lin–Kernighan tsp heuristic*, Mathematical Programming Computation, 1(2-3), 2009, s. 119-163.
93. Helsgaun K., *An extension of the Lin-Kernighan-Helsgaun tsp solver for constrained traveling salesman and vehicle routing problems*, Roskilde: Roskilde University, 2017.

Przegląd wybranych algorytmów aproksymacyjnych dla problemu komiwojażera

Streszczenie
Problem komiwojażera jest jednym z najważniejszych zagadnień optymalizacji kombinatorycznej w historii matematyki i nadal pozostaje jednym z największych jej wyzwań. Celem poniższego artykułu jest przedstawienie krótkiego i przejrzystego zestawienia najpopularniejszych narzędzi stosowanych w badaniach operacyjnych do uzyskiwania przybliżonych rozwiązań dla tego problemu – zarówno w pełnej ogólności, jak również w pewnych specyficznych scenariuszach.
Słowa kluczowe: problem komiwojażera, optymalizacja kombinatoryczna, badania operacyjne, teoria grafów

A survey of approximate solution techniques for the Traveling Salesperson Problem

Abstract
The Traveling Salesperson problem is one of the most important challenges ever formulated. Arising from the graph theory and combinatorial optimization, it still bothers the minds of great mathematicians and algorithmics alike, posing indomitable challenge. The purpose of this paper is to present a short and easily accessible comparison of most common tools used in operational research to obtain approximate solutions to TSP – in its full generality, as well as in more specific scenarios.
Keywords: Traveling Salesperson Problem, combinatorial optimization, operations research, graph theory

Michał Bugaj[1], Krzysztof Wróbel[2], Katarzyna Szumielewicz[3]

Przegląd wybranych metod wytłumaczalności modeli uczenia maszynowego oraz uczenia głębokiego

1. Wprowadzenie

W dzisiejszych czasach uczenie maszynowe oraz uczenie głębokie jest niezwykle popularnym i istotnym tematem. Obserwując zmiany w podejściu do wykorzystywania algorytmów w kwestii ingerowania w codzienne życie, łatwość wprowadzania zmian wykorzystując nowoczesną technologię i mnogość zastosowania w najróżniejszych dziedzinach można stwierdzić, że stał się to temat bardzo powszechny i wszechobecny. Wszechstronność zastosowania algorytmów uczenia maszynowego w codziennym życiu nie jest już niczym zaskakującym i nie budzi już zdziwienia, wręcz przeciwnie – jest czymś oczekiwanym, a nawet wymaganym aby dany produkt lepiej się sprzedawał. Począwszy od rynku profilowanych reklam internetowych i reklam mobilnych gdzie wykorzystywane są dane użytkowników, ich zainteresowania i trendy, przechodząc płynnie w sektor bankowy do modelowania ryzyka kredytowego oraz obserwując najnowsze zmiany wprowadzane chociażby w strukturze analizowania danych medycznych pod kątem doradczym. Duża łatwość w implementacji algorytmów uczenia maszynowego oraz uczenia głębokiego wymaga również bardzo dużej uwagi i świadomości. Trzeba się zastanowić nie tyle nad ogromnymi możliwościami najnowszych algorytmów, co nad ich realną oceną sytuacji, czynników które biorą udział w procesie oszacowania zagadnienia. Każdy algorytm, nawet najlepszy pod kątem dobranych metryk, może okazać się kompletnie bezużyteczny w kontekście nowego zagadnienia, jeśli nie będziemy w stanie zrozumieć na jakiej podstawie podejmuje on kluczowe decyzje. W tym temacie z pomocą przychodzi zagadnienie zwane wytłumaczalnością i interpretowalnością modeli uczenia maszynowego oraz uczenia głębokiego [1, 2]. W tej pracy został przedstawiony przegląd wybranych metod opisujących to zagadnienie wraz z przykładami obrazującymi istotność tego narzędzia do oceny jakości przeprowadzonych predykcji. Wytłumaczalność modeli uczenia maszynowego skupia się na opisie czynników decydujących o tym dlaczego dana predykcja została oceniona dokładnie w taki sposób oraz o metodzie dotarcia do danego wyniku. Niezależnie od tego czy jest to model drzewa decyzyjnego czy też sieć neuronowa do oceny trudnych danych medycznych istotne jest to aby zrozumieć które elementy miały kluczowy wpływ na podjęcie takiej decyzji, co dzięki wprowadzeniu szeregu algorytmów ułatwiających zrozumienie procesu decyzyjnego jest wyjątkowo przystępne. Dzięki temu można w łatwy sposób przedstawić danej osobie dlaczego jej wniosek kredytowy został odrzucony, dowiedzieć się co ma wpływ na ocenę klasyfikacji gatunku ptaków czy też znaleźć błędy w przygotowanym zbiorze uczącym, na którym model osiąga bardzo wysokie metryki w trakcie treningu.

[1] meyk@agh.edu.pl, Katedra Robotyki i Mechatroniki, Wydział Inżynierii Mechanicznej i Robotyki, Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, http://krim.agh.edu.pl/pl/.
[2] krwrobel@agh.edu.pl, Katedra Robotyki i Mechatroniki, Wydział Inżynierii Mechanicznej i Robotyki, Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, http://krim.agh.edu.pl/pl/.
[3] szumielewiczkatarzyna@gmail.com.

2. LIME – opis działania algorytmu

Pierwszą z opisanych metod wytłumaczalności modeli jest metoda LIME [3] – biorąca swoją nazwę z angielskiej nazwy local interpretable model-agnostic explanations. Opiera się ona o interpretację lokalnych modeli zastępczych, które są wykorzystane do interpretacji indywidualnych predykcji modeli uczenia maszynowego oraz uczenia głębokiego. Metoda polega na aproksymacji zjawiska przy pomocy modelu zastępczego aby wyjaśnić indywidualne predykcję, a nie całe zagadnienie. Model zastępczy przy treningu używa jedynie danych wyjściowych – predykcji badanego modelu – oraz generowanych przez siebie danych wejściowych pomijając kompletnie zbiór danych uczących. Takie podejście pozwala wyuczyć model zastępczy pewnych patternów zachowań bazujących na niewielkich zmianach wprowadzanych do oryginalnego zbioru ocenianego. Badając zmiany wyników poprzez wprowadzanie nieznacznych zmian w parametrach indywidualnych instancji pozwala na przybliżenie lokalnych wyników. Na tej podstawie algorytm LIME trenuje przybliżony, interpretowalny model do lokalnej generalizacji zjawiska. W ten sposób przygotowany algorytm jest dobrą aproksymacją predykcji lokalnych, jednakże niekoniecznie musi odzwierciedlać predykcje globalne zjawiska.

3. SHAP – opis działania algorytmu

Drugą z wybranych metod wytłumaczalności modeli jest koncepcja wartości Shapleya [4], która jest oparta na teorii gier kooperacyjnych, w których przypisuję się unikalny podział wpływu parametrów na końcową ocenę modelu. Cechy modelu współpracują ze sobą – opisują pełniej zagadnienie zdefiniowane jako cel predykcji modelu – i osiągają pewien efekt [5]. Niektóre wagi cech mają większy udział w finalnej predykcji inne znikomy. Za pomocą wartości Shapleya jest możliwy opis dystrybucji istotności wpływu każdej z badanych cech modelu na końcową predykcję algorytmu [6]. W ten sposób otrzymujemy optymalny podział wag każdej z cech oraz rozkład wartości cechy wraz z jej zakresem istotności. Ważną obserwacją jest to, że każda cecha w rozumieniu globalnej interpretowalności ma pewien uśredniony wpływ na finalną predykcję modelu, jednakże zakresy wartości wewnątrz pojedynczej cechy mogą mieć pozytywną korelację z celem algorytmu jak i negatywną korelację, wskazującą na zjawisko przeciwne. Ocena wpływu wartości każdej cechy na finalną ocenę zagadnienia w ramach lokalnej interpretowalności modelu dla pojedynczej predykcji została pokazana na figurze 1.

Jedną z kluczowych własności wartości SHAP jest to, że wartości są sumowane zawsze do różnicy pomiędzy wynikiem gdy wszystkie cechy są brane pod uwagę – biorą udział w predykcji, a sytuacją w której żadna z cech nie jest wzięta pod uwagę. Oznacza to, że wszystkie wartości cech będą sumowane do różnicy między oczekiwanymi wartościami predykcji, a aktualnymi wynikami dla konkretnego przypadku analizowanej predykcji, stąd algorytm jest w stanie ocenić wpływ każdej wartości cechy dla indywidualnej predykcji w odniesieniu do globalnej wyjaśnialności modelu bazującej na uśrednionych wartościach wpływu wszystkich cech. Rozkład każdej cechy wraz z indywidualnymi wartościami oraz ich wpływem pokazuje wykres dla globalnej interpretowalności. Przykład takiego rozkładu wartości cech wraz z ich wpływem zaprezentowano na figurze 2.

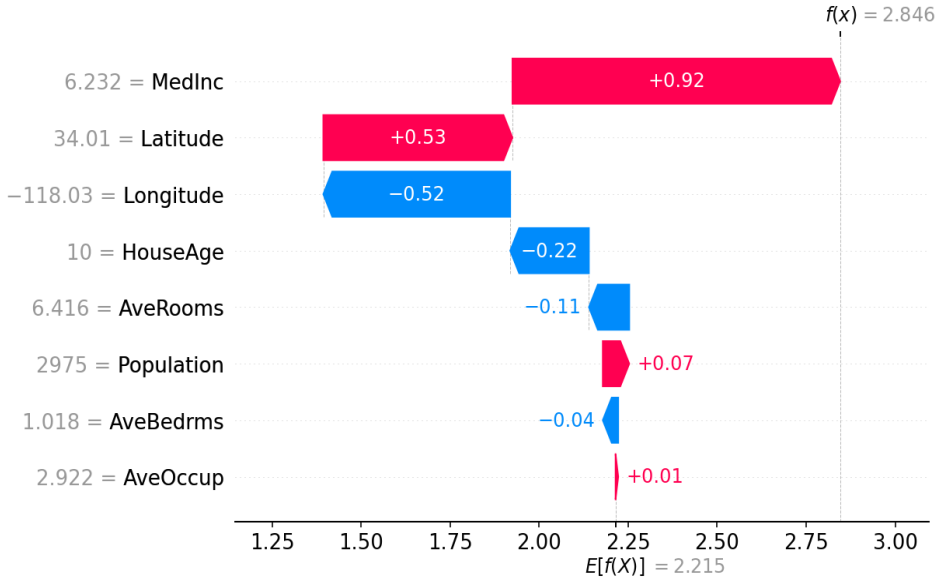


Figura 1. Ocena wpływu wartości każdej cechy na finalną ocenę zagadnienia w ramach lokalnej interpretowalności modelu dla pojedynczej predykcji [7]. Czerwonym kolorem zaznaczono pozytywny wpływ danej cechy na finalną ocenę modelu, na niebiesko negatywny wpływ cechy na opis zagadnienia

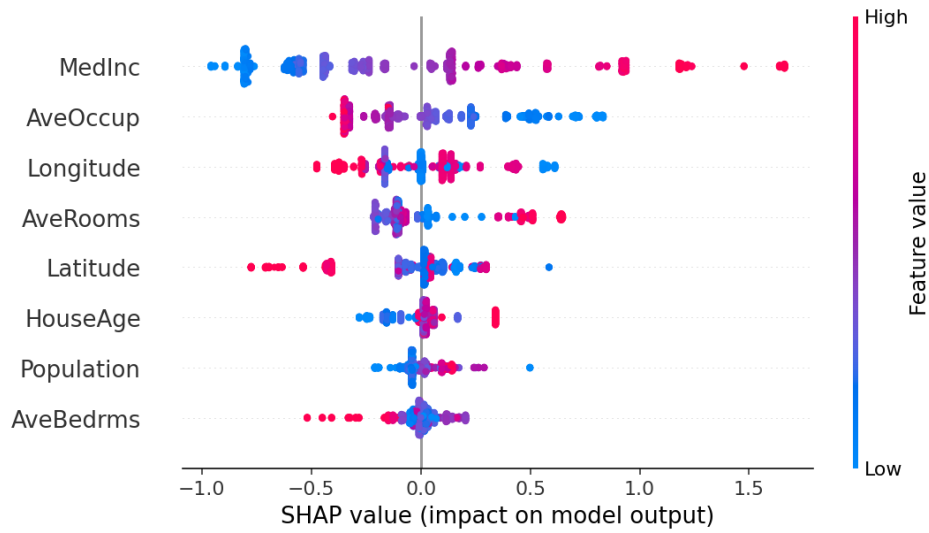


Figura 2. Wykres przedstawiający rozkład wartości każdej z cech wraz z ich wpływem na finalną ocenę modelu [7]

Kolejnym z wielu atutów podejścia metodologii SHAP jest możliwość wyjaśnienia danych graficznych korzystając z metody Deep SHAP. Jest to bardzo szybka implementacja oparta na zasadzie aproksymacji wartości Shapleya w modelach uczenia głębokiego. Deep SHAP używa rozkładu próbek oraz równań Shapleya do procesu linearyzacji składników algorytmu. Przykład na podstawie rozpoznawania cyfr zaprezentowano na figurze 3.

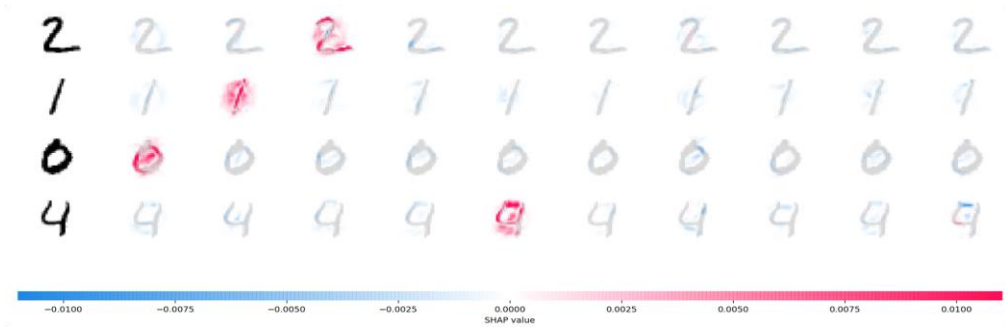


Figura 3. Wytłumaczalność rozpoznawania cyfr za pomocą algorytmu Deep SHAP [7]

Metodę SHAP można również wykorzystywać w procesie interpretacji i wyjaśnialności danych graficznych, konkretnie w klasyfikacji obrazów oraz wykazywaniu, które cechy odpowiadają za podjętą decyzję.

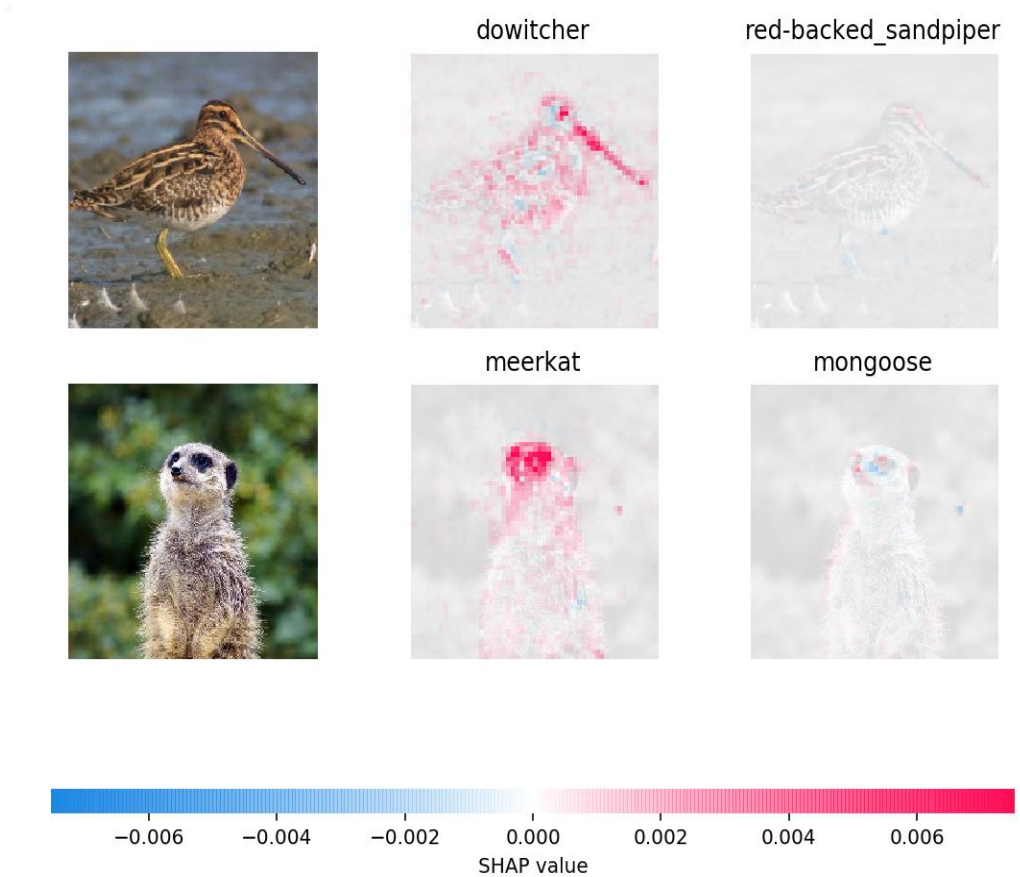


Figura 4. Wyjaśnialność algorytmu SHAP na podstawie zdjęć Szlamca i Surykatki [7]

Na figurze 4 przedstawiono porównanie klasyfikacji dwóch zwierząt wraz z obrazami przedstawiającymi, które cechy decydują o podjętej przez algorytm decyzji. Na pierwszym zdjęciu przedstawiony jest szlamiec oraz dwa obrazy, na których zaznaczono

kolorem czerwonym pixele wskazują na pozytywne wartości SHAP określające korelację klasyfikacji z cechą. Natomiast niebieskie pixele odpowiadają za negatywne wartości SHAP, które zmniejszają prawdopodobieństwo przynależności obrazu do danej klasy zagadnienia. Na lewym obrazie widać Szlamca, o którego klasyfikacji decyduje głównie jego długi dziób co wyraźnie zostało zaznaczone. Po prawej stronie widać parę niebieskich pixeli wskazujących na brak przynależności do tej klasy oraz brak czerwonych pixeli co dodatkowo pokazuje czynniki decyzyjne przy ocenie tego zagadnienia.

Drugim z przykładów jest obrazek surykatki. Ocenę tego zwierzęcia możemy zobaczyć na dwóch powyższych obrazkach odpowiadających za wyjaśnialność modelu. Na pierwszym z nich możemy zauważyć nagromadzenie pozytywnych wartości SHAP w okolicach oczu surykatki, czego zdecydowanie brakuje gdy patrzymy na klasę mangusta widoczną po prawej stronie.

4. Porównanie metody globalnej i lokalnej

Bardzo ważnym elementem zrozumienia modelu uczenia maszynowego lub modelu głębokiego jest to co tak naprawdę użytkownik końcowy chce zrozumieć. W uproszczeniu można przyjąć, że istnieją dwa rodzaje wyjaśnialności modeli – wyjaśnialność globalna oraz wyjaśnialność lokalna. Każda z nich skupia się na różnym podejściu do tematu i wyniki oparte o obie metody będą inne dlatego nie należy stosować ich wymiennie do opisywania zjawisk. Celem wyjaśnialności modelu jest przypisanie wag cech parametrów, które pozwalają modelowi podejmować decyzje. Globalna wyjaśnialność jest zbiorem średnich wszystkich predykcji dotyczących badanego zagadnienia i skupia się przede wszystkim na wyjaśnieniu w jakim stopniu każda cecha ze zbioru uczącego wpływa na ocenę modelu gdy bierzemy pod uwagę wszystkie dane. Jest to bardzo dobry generalizator modelu za pomocą którego będziemy w stanie pobieżnie zrozumieć zjawisko i sprawnie ocenić, który zestaw cech jest pozytywnie skorelowany z obranym przez nas celem, a który wskazuje na przeciwne współzależność. Globalna wyjaśnialność modelu pozwala również na ograniczenie mniej istotnych cech branych do procesu uczenia co wpływa pozytywnie na prędkość uczenia znacznie skracając czas oraz w wielu przypadkach redukując szum informacyjny wpadający do modelu co skutkuje stworzeniem lepiej generalizującego algorytmu.

Model skupiający się na lokalnej interpretacji zagadnienia opiera się o wyjaśnienie jak poszczególne cechy wpływają na ocenę pojedynczej predykcji. Jest to niezwykle istotne, ponieważ globalna wyjaśnialność daje nam uśrednioną wartość jak cecha wpływa na ocenę zagadnienia, w przypadku lokalnej interpretowalności możemy ocenić jak konkretna wartość cechy wpływa na pojedynczą predykcję. Taki mechanizm jest niezwykle istotny w przypadku oceny przypadków, które wykazują sprzeczną korelację niż generalizujący model.

5. Przykłady zastosowania

W celu wyjaśnienia istotności wytłumaczalności oraz interpretowalności modeli uczenia maszynowego i uczenia głębokiego przeprowadzono eksperyment mający na celu pokazać dlaczego nawet satysfakcjonująco wytrenowane modele, które osiągają bardzo wysokie wyniki na zbiorach walidacyjnych bez odpowiednich narzędzi służących do wyjaśnienia, o co oparte są predykcje, mogą okazać się mocno niewiarygodne, błędne a w konsekwencji niebezpieczne. Przygotowany zbiór danych, na którym został przeprowadzony

eksperyment, zawiera dwie klasy: klasa pierwsza – Wilk i klasa druga – Husky. Zadaniem modelu jest nauczenie się rozpoznawania jakie zwierze jest widoczne na obrazie. W tym celu wytrenowano model typu ResNet do klasyfikacji obrazów. Osiągnął on 100% skuteczności w nauce na zbiorze treningowym oraz ponad 90% skuteczności na zbiorze walidacyjnym. Następnie przedstawiono wyniki takiego algorytmu osób z grupy badawczej do oceny jakości modelu. Większość z nich, 68%, nie dostrzegła problemu tkwiącego w złym przygotowaniu zbioru uczącego. W kolejnym kroku zastosowano metodologię LIME aby pokazać która część obrazu decydowała o klasyfikacji zwierzęcia jako husky albo wilka. Przykład błędnej klasyfikacji przedstawiono na figurze 5.

(a) Husky classified as wolf (b) Explanation

Figura 5. Przykład błędnej klasyfikacji wraz z przedstawionym wyjaśnieniem modelu [8]

Jak można łatwo zauważyć za klasyfikację odpowiada śnieg będący tłem na obrazie – na szaro są zaznaczone obszary nie biorące udziału w klasyfikacji, pozostałe obszary są obszarami decydującymi. W tym eksperymencie chodziło o pokazanie istotności zrozumienia modeli pod względem czynników decyzyjnych – jeśli chodzi o tematykę wrażliwą, gdzie mamy do czynienia ze zdrowiem lub życiem – np. diagnostyka medyczna – to trudno sobie wyobrazić podobną sytuację nawet na najlepiej działających modelach.

Kolejnym przykładem wykorzystania metodologii wyjaśnialności lokalnej jest sektor bankowy [8], a konkretnie system przyznawania kredytów. Każda klient składający wniosek o pożyczkę jest poddawany szczegółowemu sprawdzeniu wielu czynników. W ramach konkursu przygotowanego na platformie Kaggle bank Home Credit przygotował zbiór danych, dla których analitycy mieli za zadanie przygotować modele uczenia maszynowego do automatycznego oceniania klientów pod kątem ryzyka niespłacenia przyznanego kredytu. Zbiór danych uczących zawierał wiele cech, które brały udział w procesie uczenia oraz oceniania indywidualnych predykcji, jak i wyznaczeniu globalnego wpływu cechy na finalną ocenę modelu. Bazując na podstawowym modelu LightGBM i biorąc wszystkie cechy ze zbioru uczącego do procesu uczenia modelu po dodaniu części algorytmu SHAP odpowiedzialnego za interpretowalność i wyjaśnialność wyników zbiór tych parametrów można ograniczyć z kilkuset do kilkudziesięciu z tą są dokładnością modelu. Taka operacja pozwala na lepszą generalizację modelu, znacznie szybszy proces uczenia modelu oraz optymalizację zużycia zasobów potrzebnych do eksploatacji modelu. W odróżnieniu od modeli typu black-box każda z predykcji tego modelu może zostać indywidualnie wyjaśniona, co pozwala na uargumentowanie podjęcia decyzji w stosunku do klienta, któremu wniosek został odrzucony. Takie podejście pozwala na poprawę jakości oceny zagadnienia oraz umożliwia jej zrozumienie w oparciu o cechy modeli, które zostały wprowadzone do modelu podczas procesu uczenia.

Wyjaśnialność modeli na danych tabelarycznych oraz danych tekstowych jest dość intuicyjną kwestią, natomiast niezwykle ciekawym tematem jest umiejętność interpretowalności wyników dostarczonych przez algorytmy uczenia głębokiego z wykorzystaniem sieci neuronowych do klasyfikacji obrazów. W jednym z przykładów zastosowania metodologii LIME [3] został przygotowany eksperyment mający na celu pokazanie możliwości wyjaśnialności decyzji modelu. Ze zbioru obrazków zostały wybrane dwa podobne zdjęcia głowy sowy oraz przeciętego jabłka zaprezentowane na figurze 6 i 7.

Figura 6. Zdjęcie przedstawiające głowę sowy [9]

Figura 7. Zdjęcie przedstawiające przecięte jabłko [9]

Model InceptionV3 został wytrenowany na zbiorze danych ImageNet. Dodatkowo w celu uzyskania wyników dotyczących wytłumaczalności modelu została użyta biblioteka LimeImageExplainer wchodząca w skład algorytmu LIME. Na każdym z obrazów została przeprowadzona inferencja z użyciem skryptu interpretującego pokazującego, które obszary obrazu odpowiadają za dokonanie predykcji modelu oraz w jakim stopniu dany obszar decyduje o wyniku klasyfikacji danego modelu. Wyniki wyjaśnialności modelu LIME zostały przedstawione na figurze 8.

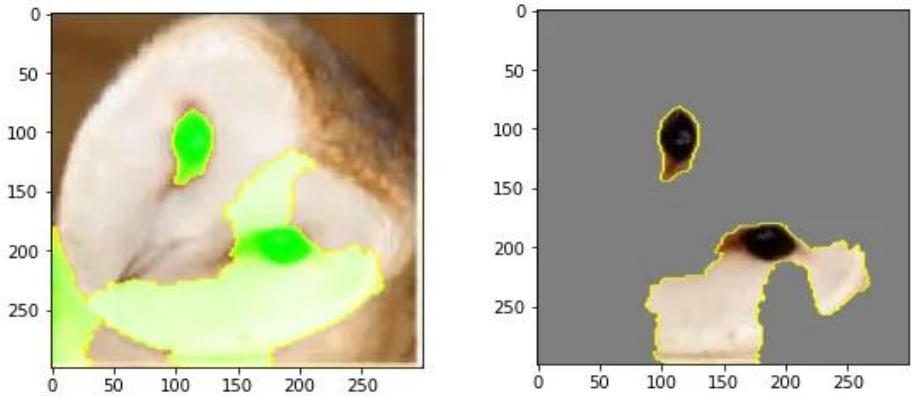


Figura 8. Zdjęcie przedstawiające wytłumaczalność modelu opisujące predykcję dla głowy sowy [9]

Na obrazie widzimy zaznaczone oczy, oraz część głowy które decydują o sklasyfikowaniu tego obrazu jako głowa sowy. Gradient koloru jest informacją pokazującą jaki wpływ ma dany obszar na końcową ocenę modelu – im intensywniejszy kolor tym większe ma znaczenie podczas podejmowania decyzji przez model. Ten sam algorytm interpretacyjny został użyty do oceny drugiego zdjęcia aby sprawdzić czy sieć ma zdolność do odróżnienia dość podobnych obrazów oraz które grupy pikseli decydują o podjęciu takiej decyzji.

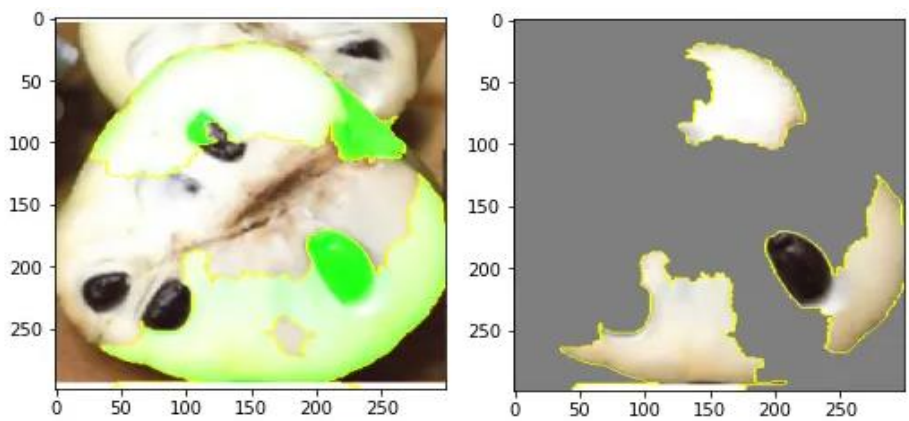


Figura 9. Zdjęcie przedstawiające fragmenty jabłka decydujące o ocenie algorytmu klasyfikacyjnego [9]

Obszary decydujące o predykcji głowy sowy oraz przekrojonego jabłka, widoczne na figurze 9, są bardzo zbliżone do siebie. W tym wypadku algorytm oparł swoją decyzję o widoczność jednej z pestek oraz kształt fragmentów pozostałego obrazu widocznych na zdjęciu. Najistotniejszym parametrem okazała się pestka oraz obszar pod szypułką przekrojonego jabłka.

6. Wnioski

Wyjaśnialność modeli jest kolejnym krokiem do wprowadzania narzędzi doradczych oraz budowania systemów inteligentnych w naszym otoczeniu. Codzienna praca wielu specjalistów może zostać ułatwiona poprzez dodanie do niej warstwy doradczej opartej

na wytłumaczalności modeli. Wiele sektorów już korzysta z algorytmów uczenia maszynowego i algorytmów uczenia głębokiego w procesach produkcyjnych jednakże znacząca większość z nich nie posiada warstw wytłumaczalności, które byłyby niezbędne do realnej oceny opisywanego przez nie zagadnienia. Przedstawione przykłady jasno uwidaczniają problematykę braku zrozumienia czynników decydujących o predykcji modeli oraz jej wpływ na podejmowane decyzje. Kwestia wyjaśnialności modeli uczenia maszynowego oraz uczenia głębokiego staje się niezwykle istotną częścią tworzenia algorytmów sztucznej inteligencji. Aby być przekonanym o skuteczności, ale też i o procesie decyzyjnym algorytmu do podjęcia kluczowej rozstrzygnięcia należy wiedzieć w jaki sposób algorytm analizuje dane, następnie przetwarza informacje z nich uzyskane oraz które cechy decydują o finalnym wyniku jego działania.

Literatura

1. Ribeiro M.T., Singh S., Guestrin C., *Why should I trust you?: Explaining the predictions of any classifier*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.
2. Shrikumar A., Greenside P., Anshul K., *Learning important features through propagating activation differences*, arXiv preprint arXiv:1704.02685, 2017.
3. https://lime.readthedocs.io/en/latest/ [data dostępu 22.03.2023].
4. Shapley L.S., *A value for n-person games*, 1953.
5. Bugaj M., Wrobel K., Iwaniec J., *Model Explainability using SHAP Values for LightGBM Predictions*, IEEE XVIIth International Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH), Polyana (Zakarpattya), Ukraine, 2021, s. 102-106
6. https://github.com/slundberg/shap [data dostępu 21.03.2023].
7. https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html#The-additive-nature-of-Shapley-values [data dostępu: 20.03.2023].
8. https://hackernoon.com/dogs-wolves-data-science-and-why-machines-must-learn-like-humans-do-41c43bc7f982 [data dostępu 22.03.2023].
9. https://medium.com/publicis-sapient-france/explain-your-ml-model-predictions-with-local-interpretable-model-agnostic-explanations-lime-82343c5689db [data dostępu: 20.03.2023].

Przegląd wybranych metod wytłumaczalności modeli uczenia maszynowego oraz uczenia głębokiego

Streszczenie

W pracy przedstawiono przegląd wybranych metod wytłumaczalności modeli uczenia maszynowego oraz uczenia głębokiego wraz z przykładami zastosowania. Opisano koncepcję wytłumaczalności modeli uczenia głębokiego oraz uzasadniono dlaczego jest to istotna kwestia przy tworzeniu algorytmów uczenia maszynowego. Następnie, dokonano analizy obecnego stanu wiedzy dotyczącej zastosowania metod wytłumaczalności modeli oraz przytoczono przykłady błędnej interpretacji danych w oparciu o celowe złe praktyki popełnione przy tworzeniu zbiorów uczących dla metodologii LIME. Przygotowano analizę porównawczą wybranych metod wytłumaczalności modeli wykorzystując metodologię LIME oraz SHAP. Opisano dwie metody podejścia do wytłumaczalności, globalnej skupiającej się na przedstawieniu związków wszystkich obserwacji na wynik oraz lokalnej odzwierciedlającej wpływ cech pojedynczych przypadków na końcowy rezultat algorytmu. Pokazano jaką przewagę mają podejścia do modelowania danych z zastosowaniem metod wytłumaczalności modeli nad modelami typu black-box.

Słowa kluczowe: wytłumaczalność modeli, SHAP, LIME, uczenie maszynowe, uczenie głębokie

Review of selected methods of explainability of machine learning and deep learning models

Abstract

The paper presents a review of the selected methods of machine learning and the deep learning model's explainability with examples of use. The concept of explainability of deep learning models is described and justification of the importance of the issue for developing machine learning algorithms was presented. Then, the current state of knowledge of the model's explainability methods was analyzed, and examples of misinterpretation of data based on intentionally bad practices committed when creating training sets for the LIME methodology were shown. A comparative analysis of selected methods of the model's explainability was prepared using the LIME and SHAP methodology. Two methods of approach to the model's explainability have been described: global, focusing on presenting the relationships between all observations and the result, and local, reflecting the influence of individual case characteristics on the final result of the algorithm. It has been shown the advantage of approaches to data modeling with the use of model explainability methods over black-box models.

Keywords: model interpretability, SHAP, LIME, machine learning, deep learning

Artur Wilczek[1], Julia Nowak[2]

Opracowanie wskaźnika inteligencji lingwistycznej na podstawie wybranych cytatów bohaterów literackich

1. Wprowadzenie

Przez wiele lat panowało przekonanie, że inteligencja jest jednostkowa i można ją mierzyć przy pomocy testów IQ. Jednak w latach 80. XX wieku Howard Gardner, amerykański psycholog i neurobiolog, zaproponował teorię inteligencji wielorakiej (ang. *multiple intelligences*), która zakłada, że inteligencja jest wieloraka i składa się z co najmniej ośmiu różnych składowych. Teoria inteligencji wielorakiej została bardzo dobrze przyjęta przez naukowców i przyczyniła się do zmiany sposobu postrzegania inteligencji, a także wpłynęła na koncepcje edukacyjne, zwracając uwagę na istotę podejścia indywidualnego do każdego dziecka [1].

Jedną ze składowych inteligencji wielorakiej jest inteligencja lingwistyczna.

Elokwencja, bogate słownictwo, poprawność interpunkcyjna czy też ortograficzna – te kilka głównych cech osób o wyższym ilorazie inteligencji lingwistycznej stanowi bardzo dobry przykład tego, jak często składowa ta utożsamiana jest z ogólną inteligencją w rozumieniu odbiorców, gdyż umiejętność wyrażania myśli oraz idei poprzez słowo kreuje obraz nadawcy jako osoby inteligentnej. Dodatkowo inteligencja lingwistyczna, dzięki skupianiu się wokół słów, stwarza możliwości analizy nie tylko komunikacji ustnej, ale również pisemnej.

Dzieła literackie, ze względu na liczne dialogi oraz wartość kulturową, stanowią bardzo dobre źródło danych do różnego typu analiz z wykorzystaniem technik przetwarzania języka naturalnego (ang. *natural language processing*). Dzięki zaawansowanym technikom sztucznej inteligencji, takim jak drzewa decyzyjne (ang. *decision trees*), możliwa jest analiza korelacji inteligencji ogólnej z jej składową, inteligencją lingwistyczną, w celu opracowania wskaźnika inteligencji lingwistycznej na podstawie uzyskanych ważności cech wybranych parametrów.

Opracowanie wskaźnika inteligencji lingwistycznej stanowi nie tylko ciekawe wyzwanie, ale również ma liczne praktyczne zastosowania. Przede wszystkim stwarza możliwości rozwojowe programów do automatycznej klasyfikacji osób na osoby o wyższym, jak i niższym ilorazie inteligencji, wyłącznie w oparciu o językowy aspekt ich wypowiedzi. W przypadku bohaterów literackich, opracowanie takiego programu pozwoliłoby nie tylko na klasyfikacje fikcyjnych bohaterów, ale w przypadku literatury, z której z narracji nie wynika, kto według autora należy do której z tych grup, na poznanie potencjalnych sympatii bądź też uprzedzeń, które podświadomie przypisywał danym bohaterom. Takie informacje pozwoliłyby na uzupełnienie profilu danego autora, a także na poznanie procesu zmian poglądów na przestrzeni wieków.

[1] artur_wilczek@outlook.com, Zakład Nauki o Danych, Centrum Badań i Rozwoju, NASK – Państwowy Instytut Badawczy.
[2] nowak_julia@outlook.com, Instytut Filologii Angielskiej, Uniwersytet Pedagogiczny im. KEN w Krakowie.

1.1. Cel pracy

Celem pracy było opracowanie wskaźnika inteligencji lingwistycznej na podstawie cytatów dwóch grup bohaterów literackich powieści Jane Austen pt. „Duma i uprzedzenie". Pierwszy etap prac dotyczył podziału bohaterów literackich ze względu na iloraz inteligencji ogólnej, wynikający z narracji, a następnie cytaty bohaterów z obu grup zostały poddane procesowi przetwarzania języka naturalnego, w celu wyodrębnienia istotnych parametrów z punktu widzenia analizy. Kolejnym etapem było dobranie odpowiedniej techniki sztucznej inteligencji umożliwiającej poznanie ważności cech, aby na ich podstawie określić wskaźnik inteligencji lingwistycznej. Ostatni etap prac dotyczył odpowiedniego wytrenowania modelu.

Zestawienie danych, w postaci procentowego udziału wytypowanych parametrów w każdym zdaniu, wraz z odpowiednią techniką sztucznej inteligencji, w tym przypadku drzew decyzyjnych, pozwoliło na opracowanie wskaźnika inteligencji lingwistycznej.

1.2. Teoria inteligencji wielorakiej

W 1983 roku Howard Gardner przedstawił teorię inteligencji wielorakiej. Według niego inteligencja jest wieloraka i składa się z co najmniej ośmiu różnych składowych, które działają niezależnie od siebie i mogą działać w różnych kombinacjach. Teoria ta zyskała dużą popularność i wpłynęła na sposób postrzegania zdolności ludzkich. Na inteligencję wieloraką, według Gardnera, składa się osiem jej rodzajów [1]:

1. Inteligencja lingwistyczna – zdolność do używania słów i języka, zarówno w piśmie, jak i w mowie. Osoby z tą inteligencją mają zwykle dobrze rozwinięte umiejętności czytania, pisania i komunikacji werbalnej.
2. Inteligencja logiczno-matematyczna – zdolność do rozwiązywania problemów matematycznych i logicznych. Osoby z tą inteligencją są zwykle dobre w przedmiotach ścisłych, rozumowaniu abstrakcyjnym i analizie danych.
3. Inteligencja przestrzenna – zdolność do myślenia przestrzennego i orientacji w trójwymiarowej przestrzeni. Osoby z tą inteligencją mają zwykle dobrą wyobraźnię przestrzenną, umiejętności rysunkowe, architektoniczne lub w zakresie projektowania.
4. Inteligencja ruchowa – zdolność do kontroli ruchu ciała i manipulowania przedmiotami. Osoby z tą inteligencją mają zwykle dobrą koordynację ruchową i umiejętności sportowe.
5. Inteligencja muzyczna – zdolność do rozumienia i tworzenia muzyki. Osoby z tą inteligencją mają zwykle dobrze rozwinięty słuch absolutny, umiejętności wokalne lub instrumentalne.
6. Inteligencja interpersonalna – zdolność do rozumienia i komunikowania się z innymi ludźmi. Osoby z tą inteligencją są zwykle empatyczne, mają dobre umiejętności społeczne i łatwo nawiązują kontakty.
7. Inteligencja intrapersonalna – zdolność do rozumienia samego siebie i swoich emocji. Osoby z tą inteligencją są zwykle introspektywne i mają dobrze rozwinięte zrozumienie swoich potrzeb i motywacji.
8. Inteligencja przyrodnicza – zdolność do rozumienia świata naturalnego i jego elementów. Osoby z tą inteligencją mają zwykle dobrze rozwinięte umiejętności obserwacyjne i rozumieją prawa przyrody.

1.3. Dotychczasowe badania

Teoria inteligencji wielorakich Howarda Gardnera jest tematem licznych badań naukowych. Wielu autorów opisało osiem rodzajów inteligencji zidentyfikowanych przez Gardnera, a także związki między nimi. Badania dotyczące inteligencji lingwistycznej obejmują różne jej aspekty, takie jak rozumienie tekstu, umiejętności pisania, język mówiony i słownictwo. W swoich pracach naukowych autorzy przedstawiają wyniki badań nad inteligencją lingwistyczną, które potwierdzają, że jest to ważny czynnik wpływający na ogólną inteligencję człowieka oraz na sukcesy w różnych dziedzinach życia, w tym w nauce, pracy i życiu społecznym. Znacznie przeważająca część badań na temat inteligencji lingwistycznej dotyczy jednak zastosowania teorii inteligencji wielorakich Howarda Gardnera w celu analizy oraz poprawy jakości nauczania w szkołach. Prace takie jak „Theory of Multiple Intelligences in Classrooms: A Case Study of Karachi" [2] czy „Wykorzystanie teorii inteligencji wielorakich Gardnera w programach nauczania zintegrowanego" [3] posługują się teorią Gardnera w celu rozwoju systemu nauczania.

Zdecydowaną mniejszością okazują się badania inteligencji lingwistycznej na podstawie tekstu pisanego. Artykuł pt. „Learning and Evaluating General Linguistic Intelligence" [4] stara się opracować ocenę inteligencji, jednak bazą badań nie jest teoria Howarda Gardnera, tylko ogólna definicja inteligencji lingwistycznej, a przedmiotem badań nie są bohaterowie literaccy, a zbiór danych SQuAD (ang. *Stanford Question Answering Dataset*).

2. Przedmiot badań

Inteligencja lingwistyczna cechuje się przede wszystkim dobrze rozwiniętymi umiejętnościami komunikacji werbalnej. Znalezienie korelacji między inteligencją ogólną a inteligencją lingwistyczną związane było przede wszystkim z klasyfikacją bohaterów na tych o wyższym jak i niższym ilorazie inteligencji oraz zestawieniem tego z ich umiejętnościami komunikacji werbalnej. Biorąc pod uwagę, że w przypadku fikcyjnych bohaterów literackich, bohaterowie jak i ich wszystkie wypowiedzi stworzone są przez konkretnego autora, ocena poziomu inteligencji, na podstawie której dokonano podziału bohaterów na dwie grupy, została przeprowadzona w oparciu o narrację oraz zestawiona z cytatami bohaterów. Przedmiotem badań była powieść pt. „Duma i uprzedzenie" Jane Austen, z której do grupy o ponadprzeciętnym ilorazie inteligencji sklasyfikowany został Fitzwilliam Darcy, natomiast do drugiej grupy sklasyfikowana została Lidia Bennet oraz Pani Bennet. Aby zmaksymalizować szansę na wyodrębnienie cech różnicujących, najbardziej rozsądnym rozwiązaniem było skupienie się na konkretnej pozycji literackiej, ponieważ każdy autor w danym momencie życia może inaczej interpretować inteligencję lingwistyczną.

2.1. Jane Austen

Jane Austen (1775-1817) była brytyjską powieściopisarką, uważaną za jedną z najważniejszych postaci literatury angielskiej XVIII i XIX wieku. Austen jest znana przede wszystkim z pisanych przez siebie powieści, które poruszają tematykę życia arystokracji angielskiej z przełomu XVIII i XIX wieku, a jej styl literacki charakteryzuje się błyskotliwym humorem, umiejętnością obserwacji i trafnymi spostrzeżeniami [5].

Austen urodziła się w Steventon, w hrabstwie Hampshire jako siódme dziecko pastora kościoła anglikańskiego, George'a Austena, oraz jego żony, Cassandry. Wraz z rodzeńst-

wem, Jane otrzymała staranne wykształcenie, którego trzon stanowiła nauka języków obcych, matematyki oraz literatury klasycznej [5].

Kontekst społeczno-kulturowy, w jakim dorastała Austen, miał wpływ na tematykę i styl jej powieści. Anglia przeżywała wtedy dynamiczne zmiany polityczne i społeczne, związane m.in. z rewolucją przemysłową oraz początkami rozwoju ruchów emancypacyjnych [5].

W swoich powieściach Austen często poruszała tematykę związaną z życiem arystokracji, ukazując jej zwyczaje, konwenanse oraz problemy związane z dziedziczeniem majątków i małżeństwami aranżowanymi. W wielu z jej dzieł pojawia się także wątek emancypacyjny, dotyczący sytuacji kobiet w społeczeństwie angielskim tamtej epoki. Austen podejmowała także tematykę związana z miłością, często ukazując ją w kontraście do wyborów materialnych i konwenansów społecznych [5].

Twórczość Austen cieszyła się w swoim czasie umiarkowanym zainteresowaniem, lecz po jej śmierci jej powieści stały się jednymi z najczęściej czytanych w literaturze angielskiej [5].

2.2. „Duma i uprzedzenie”

Jednym z kluczowych elementów, które czynią tę książkę tak wyjątkową, jest różnorodność postaci, zarówno pod względem osobowości, jak i statusu społecznego. Na początku powieści poznajemy rodzinę Bennetów, której pięć córek odgrywa kluczową rolę w fabule. Elizabeth Bennet, główna bohaterka, jest sprytną i niezależną kobietą, która odrzuca konwencjonalne poglądy na małżeństwo i poszukuje miłości opartej na wzajemnym szacunku i zrozumieniu. Jej siostry, w tym Jane, są bardziej tradycyjne i skupiają się na znalezieniu odpowiedniego małżonka, co prowadzi do wielu zabawnych sytuacji i nieporozumień. Obok rodziny Bennetów, pojawia się szereg innych postaci, które wpływają na losy bohaterów. Pan Darcy, drugi główny bohater, początkowo wydaje się arogancki i zarozumiały, ale w miarę rozwoju fabuły poznajemy jego prawdziwe uczucia i motywacje oraz wysoką inteligencję [6].

Różnorodność postaci w „Dumie i Uprzedzeniu” pozwala na przedstawienie szerokiej gamy poglądów, postaw i wartości. Austen krytycznie odnosi się do pewnych elementów społeczeństwa, takich jak konwenanse i uprzedzenia, a jednocześnie ukazuje ich złożoność i subtelność. Przez to, czytelnik jest zachęcany do samodzielnego myślenia i oceny postaci i ich zachowań. Jednakże sam fakt przedstawienia szerokiej gamy postaci o zupełnie różnych, barwnych osobowościach jest kluczowym elementem powieści oraz powodem wybrania jej do analizy dialogów [6].

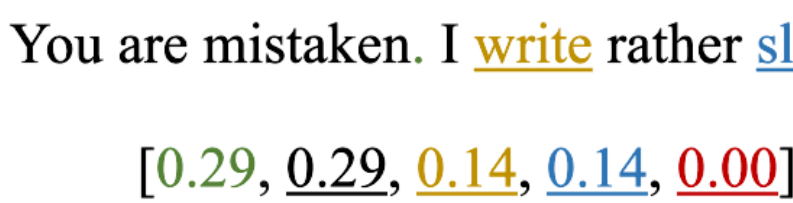
3. Metodyka

Opracowanie wskaźnika inteligencji lingwistycznej wymagało pozyskania informacji o bohaterach literackich, aby sklasyfikować ich na podstawie narracji do danej grupy. W grupie o ponadprzeciętnym ilorazie inteligencji znalazł się Fitzwilliam Darcy (20 cytatów), natomiast w grupie o przeciętnym ilorazie inteligencji znalazła się Lidia Bennet (10 cytatów) oraz Pani Bennet (10 cytatów). Następnym niezbędnym krokiem było dobranie odpowiedniej techniki z zakresu sztucznej inteligencji, umożliwiającej interpretację ważności cech analizowanych danych – czyli informacji, które z analizowanych parametrów są istotne w tego typu analizie i na których warto się skupić. W związku z tym w analizie wykorzystano las losowy składający się z 15 drzew decyzyjnych. W ramach

sprawdzenia poprawnego działania modelu, został on wytrenowany 100 razy dla różnych konfiguracji losowania danych początkowych, a następnie z wykorzystaniem średniej arytmetycznej wyznaczono średnie wartości badanych parametrów.

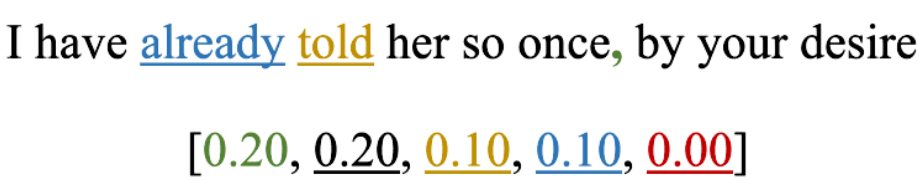
3.1. Zbiór danych

Podział bohaterów literackich na tych o wyższym jak i niższym ilorazie inteligencji w oparciu o narrację został zestawiony z cytatami tychże bohaterów, dzięki czemu otrzymano zbiór danych składający się z 40 cytatów. Każdy cytat został poddany przetwarzaniu języka naturalnego [7], dzięki czemu z każdego cytatu wyznaczono takie parametry jak liczba znaków interpunkcyjnych, liczba ważnych słów, czyli słów niepospolitych (przeciwieństwo stop listy), liczba czasowników, liczba przysłówków oraz liczba przymiotników. W celu uniknięcia zależności tychże parametrów od długości wypowiedzi, każdy z nich został podzielony przez sumę słów danego cytatu, dzięki czemu otrzymane parametry posiadają wartości z przedziału od 0 do 1, co odpowiada ich procentowemu udziałowi w danym cytacie.



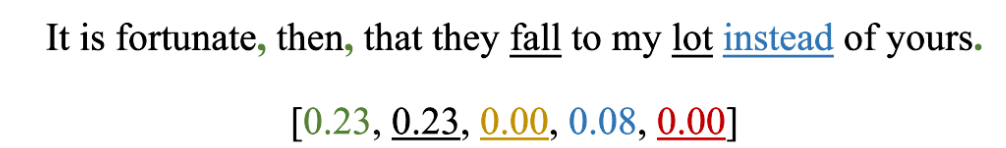
Rysunek 1. Procentowy udział parametrów w cytacie – przykład nr 1. Źródło: opracowanie własne

Na pierwszym miejscu znajduje się parametr będący stosunkiem liczby znaków interpunkcyjnych do sumy wszystkich słów (29%), na drugim miejscu parametr będący stosunkiem liczby ważnych słów do sumy wszystkich słów (29%), na trzecim miejscu parametr będący stosunkiem liczby czasowników do sumy wszystkich słów (14%), na czwartym miejscu parametr będący stosunkiem liczby przysłówków do sumy wszystkich słów (14%) oraz na piątym miejscu parametr będący stosunkiem liczby przymiotników do sumy wszystkich słów (0%).



Rysunek 2. Procentowy udział parametrów w cytacie – przykład nr 2. Źródło: opracowanie własne.

W kolejnym przykładzie parametr będący stosunkiem liczby znaków interpunkcyjnych do sumy wszystkich słów wynosi (20%), parametr będący stosunkiem liczby ważnych słów do sumy wszystkich słów wynosi (20%), parametr będący stosunkiem liczby czasowników do sumy wszystkich słów wynosi (10%), parametr będący stosunkiem liczby przysłówków do sumy wszystkich słów wynosi (10%) oraz parametr będący stosunkiem liczby przymiotników do sumy wszystkich słów wynosi (0%).



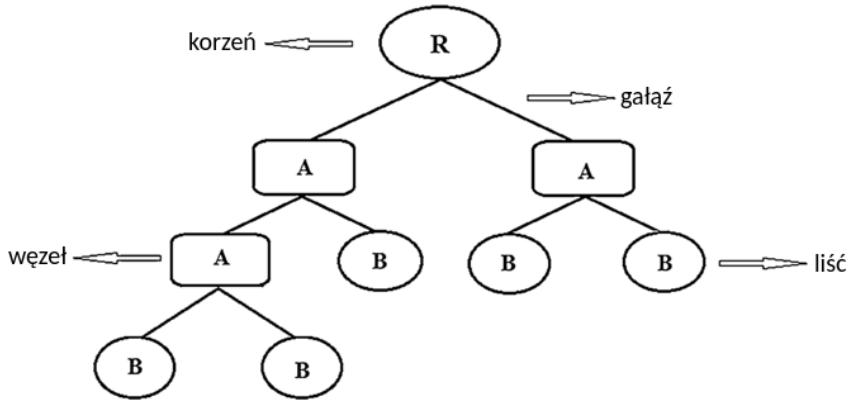
Rysunek 3. Procentowy udział parametrów w cytacie – przykład nr 3. Źródło: opracowanie własne

W ostatnim przykładzie parametr będący stosunkiem liczby znaków interpunkcyjnych do sumy wszystkich słów wynosi (23%), parametr będący stosunkiem liczby ważnych słów do sumy wszystkich słów wynosi (23%), parametr będący stosunkiem liczby czasowników do sumy wszystkich słów wynosi (0%), parametr będący stosunkiem liczby przysłówków do sumy wszystkich słów wynosi (8%) oraz parametr będący stosunkiem liczby przymiotników do sumy wszystkich słów wynosi (0%). To, co wyróżnia ten przykład, to widoczny na rysunku błąd analizy maszynowej. O ile słowo lot możemy pominąć, ponieważ analiza ta nie obejmuje rzeczowników, tak słowo fall, będące czasownikiem, nie zostało poprawnie sklasyfikowane, co wskazuje na to, że algorytm wyznaczania ważnych słów (czyli przeciwieństwa słów należących do stop listy języka angielskiego) oraz ich klasyfikacji jest bardzo restrykcyjny i w niektórych przypadkach nie działa poprawnie.

Zbiór danych został podzielony na zbiór treningowy (85% całego zbioru), czyli zbiór danych, na którym model uczy się klasyfikacji znając poprawną odpowiedź, oraz zbiór testowy (pozostałe 15% całego zbioru), czyli zbiór danych, na którym sprawdzana jest efektywność nauczenia się modelu. Zabieg podzielenia zbioru danych na dwie części przed uczeniem modelu jest konieczny, aby dane testowe nie były znane modelowi, co wyklucza możliwość zapamiętania klasyfikacji konkretnego przypadku. W każdej próbie wytrenowania modelu zbiór danych został wymieszany losowo przed podzieleniem, aby uzyskać niezależność wyników.

3.2. Drzewo decyzyjne

Drzewo decyzyjne (ang. *Decision Tree)* to technika uczenia maszynowego w dziedzinie klasyfikacji i regresji. Jego zadaniem jest stworzenie modelu decyzyjnego, który przewiduje wartość na podstawie zbioru reguł. Drzewem decyzyjnym jest graf-drzewo, które składa się z korzenia, gałęzi, węzłów oraz liści. Korzeń drzewa tworzony jest przez wybrany atrybut, gałęzie łączą korzeń z wierzchołkami, wierzchołki, z których wychodzi co najmniej jedna krawędź (węzły), sprawdzają warunek dotyczący danej obserwacji i wybierają jedną z gałęzi, natomiast pozostałe wierzchołki (liście) odpowiadają za decyzje o klasyfikacji obiektów [8].

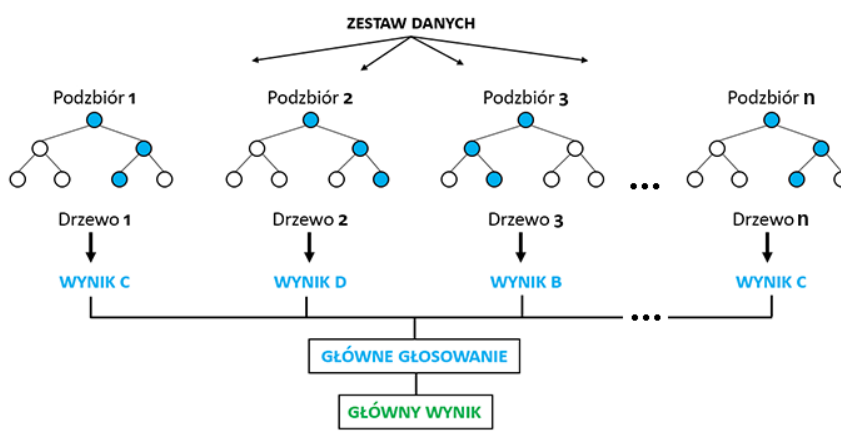


Rysunek 4. Przykład algorytmu drzewa decyzyjnego. Źródło: opracowanie własne na podstawie [9]

Oprócz wielu zalet, takich jak łatwość interpretacji wyników, możliwość obsługi zarówno danych dyskretnych, jak i ciągłych, a także szybkość działania dla dużych zbiorów danych, należy zwrócić uwagę na możliwe problemy związane z przetrenowaniem modelu, które mogą prowadzić do niskiej skuteczności w przypadku nowych danych. Aby temu zapobiec można na przykład ograniczyć głębokość drzewa, ustawić maksymalną liczbę liści drzewa lub usunąć gałęzie, które nie przynoszą istotnej korzyści w zakresie klasyfikacji nowych danych. Niestety techniki te nie zawsze przynoszą oczekiwany rezultat, w związku z czym należy zastosować bardziej rozbudowane modele uczenia maszynowego, jak na przykład las losowy [8].

3.3. Las losowy

Las losowy (ang. *Random Forest*) to technika uczenia maszynowego, która polega na łączeniu wielu drzew decyzyjnych w jednym modelu. Las losowy jest mniej podatny na przetrenowanie niż pojedyncze drzewa decyzyjne, ponieważ wykorzystuje zasadę ensemblingu (ang. *ensemble*), czyli łączenia wielu słabszych modeli w jeden silny. Zmniejsza to ryzyko dopasowania do szumu w danych uczących.



Rysunek 5. Przykład algorytmu lasu losowego. Źródło: opracowanie własne na podstawie [10]

Las losowy wykazuje większą zdolność do radzenia sobie z problemami dużego wymiaru danych, wysokiej wariancji i niedoboru danych, ponieważ każde drzewo decyzyjne wykorzystuje losowe podzbiory danych uczących i atrybutów do budowy modelu. Do tego tak jak drzewa decyzyjne, pozwala poznać ważność cech, co jest niezwykle ważne przy interpretacji wyników i rozbudowaniu modelu [8].

4. Wyniki

Kilkukrotne wytrenowanie opracowanego modelu uczenia maszynowego z losowym podziałem zbioru danych pozwoliło na osiągnięcie wyników w postaci dokładności dopasowania zbioru treningowego na poziomie 96,97%, natomiast zbioru testowego na poziomie 74,83%. Takie wyniki przekładają się następująco na ważność cech wybranych parametrów cytatów:

Tabela 1. Ważność cech rozpatrywanych parametrów

Parametr	Ważność cechy
Procentowy udział znaków interpunkcyjnych	44,57%
Procentowy udział ważnych słów	15,41%
Procentowy udział czasowników	6,73%
Procentowy udział przysłówków	29,61%
Procentowy udział przymiotników	3,68%

Źródło: opracowanie własne.

Z analizy ważności cech wynika, że najważniejszymi parametrami w klasyfikacji bohaterów literackich na tych o wyższym jak i niższym poziomie inteligencji jest procentowy udział znaków interpunkcyjnych (44,57%), procentowy udział przysłówków (29,61%) oraz procentowy udział ważnych słów (15,41%). Interpunkcja odgrywa kluczową rolę w języku pisanym, ponieważ pozwala na wyrażanie precyzyjnych znaczeń i intencji. Poprawne użycie znaków interpunkcyjnych umożliwia czytelnikom właściwe zrozumienie tekstu oraz ułatwia jego interpretację, bez tego zdania mogą stać się niejasne, nieprecyzyjne i trudne do zrozumienia. Poprawne stosowanie interpunkcji wymaga od piszącego znajomości gramatyki i składni, a także zdolności do logicznego myślenia. Wprowadzenie nawet najmniejszej zmiany w użyciu interpunkcji może zmienić znaczenie zdania i wywołać różne skojarzenia w umyśle czytelnika. Dlatego też umiejętność stosowania interpunkcji jest niezwykle ważna dla każdego, kto chce jasno i skutecznie przekazać swoje myśli za pomocą języka pisanego [11]. Drugim kluczowym parametrem są przysłówki. Duża waga przysłówków może wynikać z ich skuteczności w przekazywaniu komunikatów oraz dużej pomocy w przedstawianiu otaczającej nas rzeczywistości. Przysłówek opisuje procesy, czynności, stany, wykazuje więc dużą przydatność w wielu uniwersalnych sytuacjach i nadaje wagę innym użytym słowom. Trzecim kluczowym parametrem są ważne słowa. Są to przeciwieństwa stop listy (ang. *stop words*), czyli słów, które są najczęściej używane w każdym języku (takie jak np. przyimki, zaimki, spójniki) i nie dodają zbyt wielu informacji do tekstu. Zatem słowa istotne to takie słowa, które niosą ważne informacje i stanowią klucz do zrozumienia treści zdania lub wypowiedzi. Słowa te wyrażają najważniejsze pojęcia, idee, atrybuty lub cechy, które odnoszą się do tematu poruszanego w tekście lub wypowiedzi. Właściwe użycie słów istotnych pozwala na przekazanie konkretnej i jasnej informacji, a jednocześnie pozwala na wyrażenie intencji autora lub mówcy. Dlatego też ich stosowanie jest kluczowe w jasnym i precyzyjnym przekazywaniu treści [12].

5. Podsumowanie

Spośród stu przeprowadzonych testów, zauważono bardzo dużą rozbieżność wyników, spowodowaną losowym podziałem zbioru danych na dane treningowe oraz dane testowe a także błędami analizy maszynowej. Rozbieżność ta spowodowana jest przede wszystkim ograniczeniami związanymi z pozyskaniem i rozbudową zbioru danych, przez co opracowany model uczenia maszynowego w niektórych scenariuszach podziału danych miał problem poprawną z klasyfikacją danych testowych. Dodatkowo zastosowanie parametrów cytatów w postaci procentowego udziału w każdym cytacie z jednej strony pozwoliło na zmniejszenie możliwości manipulacji danymi wejściowymi w postaci wyboru konkretnych cytatów bardziej lub mniej złożonych, natomiast z drugiej strony wykluczyło użycie dodatkowego parametru, jakim jest całkowita liczba słów w danym cytacie. Należy również pamiętać o błędach analizy maszynowej wykorzystywanych bibliotek. Biorąc zatem pod uwagę, że parametry stanowiły jedynie procentowy udział części mowy takich jak czasowniki, przysłówki, przymiotniki oraz procentowy udział znaków interpunkcyjnych i ważnych słów, wynik klasyfikacji równy 74,83% jest wynikiem jak najbardziej akceptowalnym.

Z opracowanego wskaźnika inteligencji lingwistycznej wynika, że w przypadku klasyfikacji bohaterów na tych o wyższym jak i niższym ilorazie inteligencji, najważniejsze parametry to procentowy udział znaków interpunkcyjnych (44,57%), procentowy udział przysłówków (29,61%) oraz procentowy udział ważnych słów (15,41%). Należy jednak pamiętać, aby nie przypisywać ważności parametru do którejkolwiek z analizowanych grup – ważność cech to informacja w jakim stopniu dany parametr przyczynił się do uzyskania mechanizmu klasyfikującego, a nie wyniku.

Uzyskany wynik klasyfikacji nie jest wystarczający do opracowania zaawansowanego programu do klasyfikacji bohaterów literackich, jednak opracowanie wskaźnika inteligencji lingwistycznej poprzez wykazanie ważności cech analizowanych przykładów stanowi bardzo dobrą podstawę do opracowania takiego programu, po wcześniejszym opracowaniu algorytmu do automatycznego pozyskiwania cytatów bohaterów literackich oraz zwiększeniu liczby analizowanych parametrów, przez co możliwe będzie wykorzystanie dużo bardziej złożonych modeli uczenia głębokiego.

Literatura

1. Howard G., *Frames of Mind: The Theory of Multiple Intelligences*, Basic Books, 2011.
2. Nasim F., Khan F., *Theory of Multiple Intelligences in Classrooms: A Case Study of Karachi*, Journal of Independent Studies and Research-Management, Social Sciences and Economics, 2008, s. 58-62.
3. Wojnar-Płeszka A., *Wykorzystanie teorii inteligencji wielorakich Gardnera w programach nauczania zintegrowanego*, Oficyna Wydawnicza AFM, Kraków 2021, s. 95-110.
4. Yogatama D., De Masson C., Connor J., Kocisky, T., Chrzanowski, M., Kong, L., Lazaridou A., Ling W., Yu L., Dyer C., Blumsom P., *Learning and Evaluating General Linguistic Intelligence*, DeepMind, Londyn 2019.
5. Mitton G.E., *Jane Austen and Her Times*, Blackmask, Londyn 2016.
6. Austen J., *Pride and Prejudice*, Wordsworth Classics, 2007.
7. Bird S., Klein E., Loper E., *Natural Language Processing with Python*, O'Reilly Media, 2009.
8. Muller A.C., Guido S., *Introduction to Machine Learning with Python*, O'Reilly Media, 2016.
9. *An Introduction to Decision Tree* https://datascience.eu/mathematics-statistics/decision-tree/ [data dostępu: 01.04.2023].

10. Breiman L., Friedman J.H., Olshen R.A. and Stone C.J., *Classificationand and regression trees*, Routledge, 2017.
11. McLeod D., *The Importance of Punctuation*, https://grammarist.com/punctuation/importance-of-punctuation/ [data dostępu: 01.04.2023].
12. Ganesan K., *What are Stop Words?*, https://www.opinosis-analytics.com/knowledge-base/stop-words-explained/ [data dostępu: 01.04.2023].

Opracowanie wskaźnika inteligencji lingwistycznej na podstawie wybranych cytatów bohaterów literackich

Streszczenie
Teoria inteligencji wielorakich, opracowana przez Howarda Gardnera, wyróżnia osiem typów inteligencji. Każdy z nich odpowiada za umiejętności w innym obszarze. Inteligencja lingwistyczna, która stanowi przedmiot badań niniejszej pracy, przejawia się przede wszystkim w ponadprzeciętnych umiejętnościach komunikacji interpersonalnej. Opracowanie wskaźnika inteligencji lingwistycznej pozwala na lepsze zrozumienie związku między językiem a inteligencją oraz na rozwijanie narzędzi do oceny kompetencji językowych w różnych kontekstach. Badania te są istotne nie tylko dla literatury, ale także dla innych dziedzin, takich jak nauki o języku, socjolingwistyka oraz psychologia. W niniejszej pracy opracowany został wskaźnik inteligencji lingwistycznej na podstawie wybranych cytatów bohaterów literackich. Z dzieł literackich wybrani zostali bohaterowie cechujący się wysokim jak i niskim poziomem inteligencji, wynikającym z narracji. Następnie cytaty bohaterów z wyróżnionych grup zostały zestawione ze sobą w celu analizy różnic stylu ich wypowiedzi. Pod uwagę zostały wzięte takie aspekty jak częstotliwość występowania interpunkcji, słów o małym znaczeniu oraz różnych części mowy. Analiza tekstu została przeprowadzona z wykorzystaniem biblioteki NLTK środowiska programistycznego Python, natomiast wskaźnik inteligencji lingwistycznej został opracowany z wykorzystaniem algorytmu sztucznej inteligencji, jakim jest drzewo decyzyjne.
Słowa kluczowe: teoria inteligencji wielorakich, inteligencja lingwistyczna, literatura, sztuczna inteligencja, uczenie maszynowe

Development of an index of linguistic intelligence based on selected quotes of literary characters

Abstract
The theory of multiple intelligences, developed by Howard Gardner, distinguishes eight types of intelligence. Each of these is responsible for skills in a different area. Linguistic intelligence, which is the subject of this study, manifests itself primarily in above-average interpersonal communication skills. The development of an index of linguistic intelligence allows a better understanding of the relationship between language and intelligence and the development of tools to assess linguistic competence in different contexts. This research is relevant not only for literature, but also for other fields such as language science, sociolinguistics and psychology. In the present study, an index of linguistic intelligence was developed on the basis of selected quotations of literary characters. Characters characterised by high as well as low levels of intelligence, derived from the narrative, were selected from literary works. The quotations of the protagonists from the distinguished groups were then juxtaposed to analyse the differences in the style of their speech. Aspects such as the frequency of punctuation, words with little meaning and different parts of speech were taken into account. Text analysis was carried out using the NLTK library of the Python programming environment, while the linguistic intelligence index was developed using the artificial intelligence algorithm decision tree.
Keywords: theory of multiple intelligences, linguistic intelligence, literature, artificial intelligence, machine learning

Magda Konieczna[1]

Metody, techniki i problemy sieci neuronowych

1. Wprowadzenie

Sieci neuronowe są jedną z dziedzin nauczania maszynowego, która znajduje coraz szersze zastosowania w różnych dziedzinach życia człowieka. Zainteresowanie wykorzystywaniem sieci nauronowych do rozwiązywania problemów związanych z przetwarzaniem danych spowodowało rozwój wielu modeli i technik z nimi związanych. W artykule omówiono różne typy parametrów sieci neuronowej takie, jak funkcje aktywacji, funkcje straty, metody regularyzacji, metody optymalizacji modeli i wskazano dalsze możliwe kierunki badań.

Sztuczna inteligencja to dziedzina nauki, która naśladuje zachowania ludzi w rozwiązywaniu problemów i uczy się na podstawie zebranych danych zgodnie z określonymi zasadami. Uczenie maszynowe jest częścią sztucznej inteligencji, która posługuje się algorytmami do znajdowania wzorców z danych i przewidywania wyników na podstawie przetworzonych danych [1]. Nauczanie maszynowe można podzielić na trzy kluczowe kategorie: nauczanie nadzorowowane, nienadzorowane i uczenie ze wzmocnieniem. W tabeli zostały zaprezentowane przykładowe metody każdego z typów nauczania maszynowego.

Tabela 1. Typy nauczania maszynowego

Uczenie nadzorowane	Uczenie nienadzorowane	Uczenie ze wzmocnieniem
Regresja liniowa Metoda k-najbliższych wektorów Maszyna wektorów nośnych Drzewo decyzyjne Lasy losowe Regresja liniowa Regresja logistyczna Sieci neuronowe Algorytm Naive Bayes Algorytm AdaBoost	Klastrowanie K-means Klastrowanie hierarchiczne Algorytmy globalnej optymalizacji Propagacja afiniczności Model mieszaniny gaussowskiej Rozmyte k-średnie Klastry K-średnich Samo-organizujące się mapy	Metody Monte Carlo Programowanie dynamiczne Temporal difference

Źródło: [2, 3].

Sieci neuronowe są nadzorowaną metodą nauczania maszynowego. Dzieli się je na jednokierunkowe i rekurencyjne sieci neuronowe. Informacje przechodzą z warstwy wejściowej do wyjściowej przez wiele warstw ukrytych [4]. Sieci neuronowe z więcej niż trzema warstwami nazywam się głębokimi sieciami neuronowymi.

Sieci neuronowe w ostatnim czasie znalazły zastosowanie w różnych dziedzinach, takich jak:

- Medycyna – obrazowanie medyczne z wykorzystaniem obrazów MRI (ang. *Magnetic Resonance Imaging*) i CT (ang. *Computed Tomography*);

[1] m.konieczna@dyd.akademiabialska.pl, Wydział Nauk Technicznych, Akademia Bialska im. Jana Pawła II.

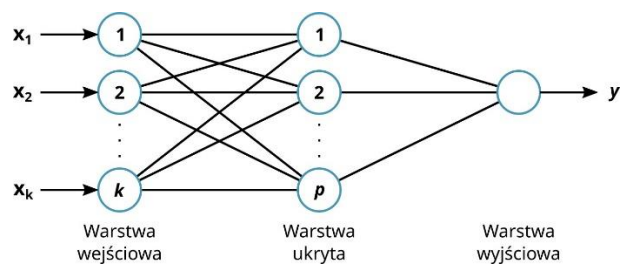
- widzenie komputerowe – rozpoznawanie twarzy, opisywanie zdjęć;
- przetwarzanie języka naturalnego – odpowiadanie na pytania, rozpoznawanie mowy, klasyfikacja dokumentów;
- biologia – klasyfikowanie komórek, analizowanie interakcji między białkami;
- generowanie obrazów – kolorowanie obrazów, zwiększanie rozdzielczości obrazów usuwanie szumów;
- systemy rekomendacji – wyszukiwanie w sieci, rekomendacja produktów;
- robotyka – przenoszenie obiektów, które łatwo zlokalizować lub unieść;
- prognozowanie finansowe i logistyczne;
- zamiana tekstu na mowę;
- i wiele innych zastosowań.

Sieci neuronowe mogą przetwarzać różne typy danych i różne rozmiary danych, w tym dane binarne i wielokrotne liczby klas, zaszumione dane z wartościami odstającymi, dane surowe i wstępnie przetworzone.

Sieci neuronowe obejmują wiele metod i technik. Umożliwiają one uczenie się wzorców z danych przy mniejszym udziale człowieka. Jednak efektywność uczenia się zależy od wiedzy człowieka [1].

2. Struktura sieci neuronowej

Najprostsza sieć neuronowa składa się z trzech warstw: wejściowej, ukrytej i wyjściowej, co zostało przedstawione na rysunku 1. Warstwa wejściowa ma kilka neuronów wejściowych $(x_1, x_2, x_3 \dots x_n)$, których liczba zależy od zestawu danych użytych w procesie uczenia się. Dane te mogą być w różnych formatach: obraz, dźwięk, film, tekst, dane liczbowe. Wszystkie neurony warstwy wejściowej mają przypisane wagi $(w_1, w_2, \dots w_n)$. Warstwa wyjściowa zawiera kilka neuronów lub jeden reprezentujący klasę wyjściowego zestawu danych. Warstwa ukryta jest odpowiedzialna za naukę sieci, wykonywanie obliczeń. Na poziomie tej warstwy dochodzi do modyfikacji parametrów sieci.



Rysunek 1. Budowa sieci neuronowej. Źródło: [5]

Sieć neuronowa oblicza wynik *y* sumując iloczyn wag wektorów i wartości wektorów wejściowych, wartość *t* określa wyjście lub przewidywaną wartość *y* [6].

$$y_i = \begin{cases} 1, jeśli \sum_i w_i x_i > t \\ 0, jeśli \sum_i w_i x_i \leq t \end{cases} \quad (1)$$

gdzie: x_i – dane wejściowe, w_i – wagi wektora wejścia, y_i – dane wyjściowe, t – wyraz wolny

Jeżeli warunek t zastąpimy stałą b równanie ma postać:

$$y_i = \begin{cases} 1, jeśli \sum_i w_i x_i - b > 0 \\ 0, jeśli \ \sum_i w_i x_i - b \leq 0 \end{cases} \quad (2)$$

gdzie: b – wyraz wolny

Jeżeli wartość $b > 0$ to wynikiem funkcji jest 1, a jeżeli $b < 0$ to y przyjmuje wartość 0. Każda zmiana wagi lub odchylenia zmieni wynik i spowoduje, że będzie on bliższy pożądanej wartości [7].

3. Funkcje aktywacji

Funkcje aktywacji to funkcje używane w sieciach neuronowych do obliczania ważonej sumy danych wejściowych i parametrów [8]. Wynik działania funkcji decyduje czy dany neuron jest wartościowy dla sieci czy nie. Jej celem jest wygenerowanie danych wyjściowych sieci neuronowej, które zawierają nowe parametry. Funkcja aktywacji może również zaktualizować dane używając algorytm gradientu prostego lub stochastycznego gradientu prostego. Wyróżniamy dwa rodzaje funkcji aktywacji: liniowe i nieliniowe.

3.1. Liniowe i nieliniowe funkcje aktywacji

W tym rozdziale zostały omówione nieliniowe funkcje aktywacji, które mają właściwości liniowości, takich jak funkcje ReLu (ang. *Rectified Linear Activation Function*). ReLu jest jedną z najpopularniejszych funkcji aktywacji w sieciach neuronowych. Głównymi zaletami tej funkcji są: prostota obliczeń, rozproszenie reprezentacji, czyli wejściowe wartości ujemne na wyjściu dają wartość zero, natomiast pozostałe zachowują wartości dodatnie. Kolejną zaletą jest zachowanie funkcji podobne do funkcji liniowej. Ma ona tą samą wydajność i działanie co funkcja liniowa. Nie ma problemu z zanikaniem gradientu, jak w przypadku innych architektur sieci głębokiego uczenia. Stąd funkcja ReLu jest częściej wykorzystywana niż funkcja sigmoidalna. Wzór funkcji podany jest poniżej [9].

$$f(x) = \max(0, x_i) = \begin{cases} x_i, x_i > 0 \\ 0, \ x_i < 0 \end{cases} \quad (3)$$

gdzie: x_i – dane wejściowe

Funkcja ReLu nie ma górnej granicy. Problemem są również neurony, które w wyniku braku aktualizacji podczas opadania gradientu, przestają reagować na zmiany danych wejściowych i zostają w martwym punkcie (martwy neuron). Z tego względu powstało wiele udoskonaleń funkcji ReLu [10].

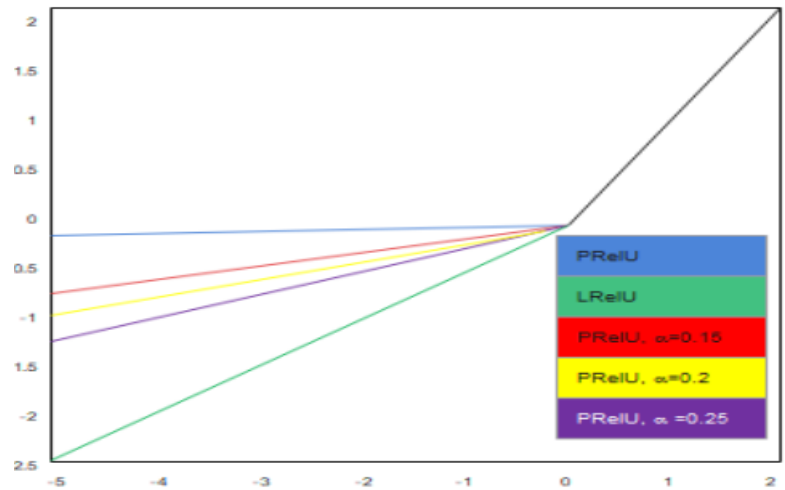
Jedną z tych modyfikacji jest funkcja LReLu (ang. *Leaky Rectified Linear Activation Function*). Jej zaletą jest zmniejszenie znikającego gradientu poprzez wprowadzenie parametru alfa. LReLu oblicza gradient z bardzo małą stałą wartością dla gradientu ujemnego alfa w zakresie 0,01 [8]. Wzór funkcji LReLu jest przedstawiony poniżej [9].

$$f(x) = \begin{cases} x_i, x_i > 0 \\ \alpha_i x_i, x_i < 0 \end{cases} \quad (4)$$

gdzie: $\alpha_i - stała$

Współczynnik α_i jest stałą, która rozwiązuje problem martwych neuronów powstały w wyniku mapowania ujemnego wejścia ReLu do zera [11].

Ponadto kolejnymi modyfikacjami funkcji ReLu są funkcje PReLu (ang. *Parametric Rectified Linear Activation Function*) i ELU (ang. *Exponential Linear Unit*), które podobnie jak LReLu proponują rozwiązanie problemu martwych neuronów. Zależności między tymi funkcjami przedstawione zostały na rysunku 2.



Rysunek 2. Funkcje LRelu, PRelu z różnymi wartościami α. Źródło: [11]

Inną funkcją aktywacji jest funkcja SoftMax (znormalizowana funkcja wykładnicza). Jest ona uogólnieniem funkcji logistycznej [12]. Funkcja normalizuje wynik wyjściowy i podaje go jako prawdopodobieństwo, dlatego jest często wybierana jako funkcja aktywacji ostatniej warstwy sieci neuronowej. Poniżej wzór tej funkcji [8].

$$softmax(x_i) = \frac{e^{x_i}}{\sum_j^n e^{x_j}} \tag{5}$$

gdzie: x_i – dane wejściowe

Funkcja Softmax w przeciwieństwie do funkcji sigmoidalnych, które są używane do klasyfikacji binarnej, mogą być używane do problemów klasyfikacji wieloklasowej [11].

Wybór funkcji aktywacji zależy od danego problemu, który jest do rozwiązania i użytych danych.

3.2. Przykłady nieliniowych funkcji aktywacji

Przykładowe nieliniowe funkcje aktywacji to [13-14]:

1. Funkcja sigmoidalna

$$f(x) = \frac{1}{1+e^{-x}} \tag{6}$$

- funkcja najczęściej używana ze względu na nieliniowość;
- ciągła i różniczkowalna;
- nie jest symetryczna względem zera, zatem znaki wszystkich wartości wejściowych neuronów będą takie same.

2. Tangens hiperboliczny

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{7}$$

- ciągła i różniczkowalna;
- przyjmuje wartości od -1 do 1;
- jest równoważna pod względem wydajności z funkcją sigmoidalną.

3. SiLu

$$a_k(s) = z_k \alpha(z_k) \tag{8}$$

$$z_k = \sum w_{ik} s_i + b_k \tag{9}$$

gdzie: s – wktor wejścia, w_{ik} – wagi, b_k – wyraz wolny

- używana w głębokich sieciach neuronowych;
- używana tylko w uczeniu przez wzmacnianie.

4. dSiLu

$$a_k(s) = \alpha(z_k) \cdot (1 + z_k(1 - \alpha(z_k))) \tag{10}$$

gdzie: s – wktor wejścia, w_{ik} – wagi, b_k – wyraz wolny

- używana do aktualizacji uczenia ze spadkiem gradientu;
- wydajność znacznie lepsza od funkcji sigmoidalnej.

5. Funkcja Gaussa

$$y_i(x) = \sum_{j=1}^{k} w_{ij} \phi(||x - c_j||) \tag{11}$$

gdzie: x – wktor wejścia, k-liczna neuronów w warstwie ukrytej, c_j – środek j – tej warstwy ukrytej neuronu w_{ij} – wagi

- opiera się na krzywej Gaussa;
- bada odległość od określonego punktu za pomocą miary euklidesowej.

3.3. Funkcje straty

Wyróżniamy dwie główne techniki uczenia nadzorowanego:

1. Regresję – jest to problem uczenia nadzorowanego, w którym dane są wykorzystywane do przewidywania wartości rzeczywistych etykiet niewidocznych danych. Algorytmy regresji obejmują regresję liniową, regresję liniową jądra, maszyna wektorów nośnych, regresja metodą LASSO (operator najmniejszej bezwzględnej redukcji i wyboru).
2. Klasyfikację – celem tej metody jest etykietowanie danych ze skończonego zestawu klas. Klasyfikacje binarne odnoszą się do klasyfikacji opartej na zestawie dwóch klas, a klasyfikacja wieloklasowa odnosi się do klasyfikacji opartej na zestawie trzech lub więcej klas [2, 16].

Głównym zadaniem trenowania modelu jest, aby przewidywał on wyniki wektora docelowego na podstawie nowego wektora wejściowego [17]. W regresji przewidywana wartość docelowa wektora wejściowego jest ciągła, a w klasyfikacji dyskretna. Wszystkie funkcje strat służą do opisu jakości funkcji uczenia zarówno w klasyfikacji, jak i w regresji. Głównym zadaniem przy obliczaniu straty jest obliczenie różnicy pomiędzy prawdziwą etykietą a przewidywaną. Jeśli wartość różnicy jest bliska 0, to funkcja uczenia

jest wartościowa, w innym przypadku należy zmodyfikować model [1]. Wzór (12) przedstawia funkcję w modelu regresji [18].

$$f(x) = min \sum_{i=1}^{n} l(f(x_i) - y_i) + R_\lambda(f) \tag{12}$$

gdzie: $f(x_i) - y_i$różnica między funkcją uczenia. w modelu a wynikiem y_i, l-funkcja straty, $R_\lambda(f) -$ regularyzacja modelu

Wzór (13) przedstawia funkcję straty w modelu klasyfikacji [18].

$$f(x) = min \sum_{i=1}^{n} l\big(y_i f(x_i)\big) + R_\lambda(f) \tag{13}$$

gdzie: $y_i -$ etykieta dla x_i $y_i f(x_i)$różnica między funkcją uczącą w modelu a hiperpłaszczyzną

Bez rozróżniania na rodzaje modeli funkcję straty można zapisać wzorem (7) [18].

$$min_L \|P(L) - X\|_l + R_\lambda(L) \tag{14}$$

gdzie: P – operator, l – funkcja straty, $R_\lambda(L) -$ czynnik regularyzacji

4. Optymalizacja sieci neuronowych

Algorytmy optymalizacyjne to jedne z ważniejszych składowych sieci neuronowych. Ich zadaniem jest zmiana atrybutów sieci neuronowej, takich jak wagi i szybkość uczenia, w celu zmniejszenia strat. Algorytmy lub strategie optymalizacyjne odpowiadają za zmniejszenie strat i zapewnienie jak najdokładniejszych wyników. Skuteczne algorytmy optymalizacji wpływają na wydajność modelu sieci neuronowych. Techniki optymalizacji możemy podzielić na:

1. Metoda optymalizacji pierwszego rzędu, taka jak metoda gradientu stochastycznego, która służy do aktualizacji wag modelu sieci neuronowych. Zmiana i aktualizacja parametrów modelu prowadzi do zminimalizowania strat.
2. Metody optymalizacji wysokiego rzędu, takie jak metoda *Newtona*.
3. Optymalizacja heurystyczna bez pochodnych, np. metoda zejścia współrzędnych, metody optymalizacji globalnej i wyszukiwanie w celu stworzenia sieci o mniejszej złożoności i wysokim współczynniku klasyfikacji [19].

Gradient prosty to algorytm optymalizacji pierwszego rzędu, który jest zależny od pochodnej pierwszego rzędu funkcji straty. Oblicza, w jaki sposób należy zmienić wagi, aby funkcja mogła osiągnąć minima. Dzięki wstecznej propagacji straty są przenoszone z jednej warstwy do drugiej, a parametry modelu, są modyfikowane w zależności od strat, tak aby można było zminimalizować straty.

Wadą tej metody jest zmiana wag na całym zbiorze po obliczeniu gradientu. Zatem w przypadku dużego zbioru danych, osiągnięcie minimum może zająć bardzo dużo czasu oraz wymaga dużej pamięci do obliczenia gradientu na całym zbiorze danych.

Stochastyczny spadek wzdłuż gradientu jest algorytmem wyliczającym gradienty na podstawie losowej próby. W tym przypadku parametry modelu są zmieniane po obliczeniu strat na każdym przykładzie treningowym. Ze względu na większą szybkość można go używać wobec dużych zestawów danych [3,20].

Parametry modelu są często aktualizowane oraz charakteryzują się dużą wariancją i fluktuacjami funkcji strat przy różnych intensywnościach. Wadą modelu jest duża zmienność parametrów modelu oraz brak przerwania pracy algorytmu po osiągnięciu minimum.

5. Regularyzacja

Regularyzacja sieci neuronowych to ograniczenie modelu w celu jego uproszczenia i zmniejszenia ryzyka przetrenowania [21].

Na etapie nauki modelu stopień regularyzacji jest kontrolowany przy pomocy hiperparametrów. Są to parametry algorytmu uczącego, które nie są przez ten algorytm modyfikowane. Należy je wyznaczyć przed procesem ucznia się.

Głównym zadaniem w klasyfikacji danych jest znalezienie dobrej funkcji uczącej lub modelu uczenia się, który może dokładnie przybliżyć docelowy wynik na podstawie próbek wejściowych. Funkcja uczenia sieci neuronowej ma na celu dopasowanie wagi, która minimalizuje funkcję straty.

Istnieją różne techniki regularyzacji, takie jak [21]:

1. Regularyzacja danych: przekształcenie uczącego zbioru danych w nowy zbiór danych. Przekształcanie danych może być wykonane przez wykonywanie ekstrakcji cech, wstępne przetwarzanie, redystrybucję zbioru danych lub generowanie nowej próbki, np. zmiana rozmiaru, obracanie, przesuwanie boczne, które nazywane są powiększaniem danych.
2. Regularyzacja architektury pracy sieci: Ten typ rozważa udoskonalenie funkcji mapowania wejścia-wyjścia właściwości dopasowane do danych wejściowych, takie jak podział wagi, wybór funkcji aktywacji, modele szumów, wielozadaniowość uczenie się, odpadanie i wybór modelu.
3. Regularyzacja funkcji strat: obejmuje m.in. entropię krzyżową i błąd średniokwadratowy.

Wśród różnych metod regularyzacji sieci neuronowych, najczęściej stosowane są metody regularyzacji L1 i L2. Metody te zasadniczo regulują wartości bezwzględne parametrów wagowych tak, aby wariancja w każdej warstwie nie stała się znacznie większa niż w poprzedniej warstwie [22].

6. Przetrenowanie i niedotrenowanie sieci neueronowych

Celem dobrego modelu uczenia maszynowego jest dobre uogólnienie danych treningowych na dowolne dane z dziedziny problemu. Pozwala to sieci neuronowej wyznaczać dane na podstawie danych, których model nie używał do nauki. Nadmierne dopasowanie i niedopasowanie to dwie kluczowe koncepcje uczenia maszynowego, które są najczęstszymi przyczynami słabej wydajności modelu. Przyczyny nadmiernego dopasowania można podzielić na trzy rodzaje:

1. Uczenie się szumu na zbiorze treningowym. Występuje, gdy zbiór treningowy jest zbyt mały lub znajdują się w nim niereprezentatywne dane, lub jest zbyt wiele szumów.
2. Duża ilość danych wejściowych. Modele mogą się znacznie różnić w przypadku różnych zestawów danych.
3. Wielokrotność porównywania procedur, które są wczechobecne w algorytmach indukcyjnych, a także innych algorytmach uczących.

6.1. Metoda wczesnego zatrzymania

Strategia ta stosowana jest w celu uniknięcia zjawiska spowolnienia szybkości uczenia się. Ta kwestia oznacza, że dokładność algorytmów po pewnym czasie przestaje się poprawiać, a nawet pogarsza się z powodu nauki hałasu. Zatrzymanie algorytmu powinno następić w optymalnym punkcie. Jeśli model kontynuuje naukę po punkcie, błąd

walidacji wzrośnie podczas uczenia. Jeśli przestanie się uczyć przed punktem to powstaje zjawisko niedopasowania modelu. Jeśli zatrzymamy się po punkcie, model staje się nadmiernie dopasowany.

Tak więc celem jest znalezienie dokładnego punktu, w którym należy przerwać trening. W ten sposób uzyskuje się idealne dopasowanie modelu pomiędzy niedotrenowaniem i przetrenowaniem. W przypadku sztucznych sieci neuronowych proces uczenia polega na znalezieniu idealnego zestawu wagi i wyrazu wolnego.[23]

6.2. Redukcja sieci

Uczenie się hałasu jest jedną z ważnych przyczyn nadmiernego dopasowania. Zatem redukcja szumów staje się jednym z kierunków badań nad nadmiernym dopasowaniem. Jedną z metod jest zmniejszenie rozmiaru końcowych klasyfikatorów w uczeniu relacyjnym, zwłaszcza w drzewie decyzyjnym. Redukcja danych stosowana jest w celu zmniejszenia złożoności klasyfikacji poprzez wyeliminowanie mniej znaczących lub nieistotnych danych, a wreszcie w celu zapobieżenia nadmiernemu dopasowaniu i poprawienia dokładności klasyfikacji. Redukcja danych przed i po to dwa standardowe podejścia do radzenia sobie z hałasem.

Algorytmy wstępnego przycinania działają podczas procesu uczenia się. Zwykle używają kryteriów zatrzymania, aby określić, kiedy przestać dodawać warunki do reguły lub dodawać regułę do opisu modelu. Post-przycinanie dzieli zestaw treningowy na dwa podzbiory: zestaw rosnący i zestaw przycinający. W porównaniu do algorytmu wstępnego uczenia te algorytmy ignorują problemy z nadmiernym dopasowaniem podczas procesu uczenia się. Zamiast tego zapobiegają nadmiernemu dopasowaniu poprzez usuwanie warunków i reguł z modelu generowanych podczas nauki. Podejście to jest znacznie dokładniejsze, ale mniej efektywne [23].

W uczeniu maszynowym algorytm nie jest jedyną składową wpływającą na ostateczną dokładność klasyfikacji. W wielu przypadkach istotny wpływ na wydajność może mieć ilość i jakość zbioru danych treningowych, zwłaszcza w obszarze nadzorowanego uczenia się.

Trening modelu jest w rzeczywistości procesem dostrajania jego hiperparametru. Dobrze dobrane parametry sprawiają dobrą równowagę między dokładnością i regularnością treningu, a następnie hamują efekt przetrenowania, jak również niedopasowanie.

7. Podsumowanie

W ostatnim czasie sieci neuronowe stały się doskonałym wyborem do rozwiązywania różnych problemów z dużą ilością danych. Oprócz ogromnych zalet korzystania z sieci neuronowych istnieje szereg wad. Uczenie zajmuje dużo czasu, zwłaszcza w przypadku zbioru danych o dużym rozmiarze, a czas trwania jest trudny do oszacowania. Wymagana jest duża liczba próbek uczących do trenowania zaprojektowanego modelu. Dodatkowo nie ma warunku ani reguły, aby wybrać odpowiedni model do rozwiązania danego problemu. Sieci neuronowe oparte są na wartościach liczbowych, więc wszystkie dane wejściowe sieci neuronowej powinny być w postaci numerycznej. Optymalizacja parametrów, takich jak szybkość uczenia, liczba węzłów w warstwach ukrytych i liczba ukrytych warstw, może być trudnym zadaniem i wymaga eksperymentowania oraz doświadczenia. Sposób, w jaki sztuczna sieć neuronowa przetwarza dane wejściowe i generuje żądane wyniki, jest nadal niejednoznaczny [24].

Znaczący postęp w zastosowaniu sieci neuronowych z zastosowaniem technik głębokiego nauczania prowadzi jednak do znacznej poprawy parametrów sieci neuronowej. Pomimo tego pozostaje wiele do zrobienia. W celu poprawy wydajności sieci neuronowych doskonali się i opracowuje nowe funkcje aktywacyjne, które dobrze współpracują z danym zbiorem danych oraz wzmacnia się i proponuje nowe funkcje straty, które poprawiają wydajność modelu. Wreszcie, stosuje się różne techniki powiększania danych do manipulacji i zwiększenia liczby próbek treningowych oraz techniki regularyzacji, takie jak porzucanie i zwiększanie liczby warstw sieci neuronowej w treningu głębokiego uczenia się.

Literatura

1. Nayef B.H., Sheikh Abdullah S.N.H., Sulaiman R., Alyasseri Z.A.A.K., *Varians of neural networks:a review*, Malaysian Jurnal of Computer Science, 35(2), 2022, s. 158-178.
2. Erickson B.J., Korfiatis P., Akkus Z., Kline T.L. *Machine learning for medical imaging*, Radiographics, 37(2), 2017, s. 505-515.
3. Geron A., *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow*, Helion SA, Gliwice 2020.
4. Mijwil M., Alsaadi, A., *Overview of Neural Networks. Computer Engineering Techniques Department*, 1, 2, 2019.
5. Fath A.H., Madanifar F., Abbasi M., *Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems*, Petroleum 6(1), 2020, s. 80-91.
6. Zuo Y., Li B., Zhao Y., Jiang Y., Chen Y. C., Chen P., Du S., *All-optical neural network with nonlinear activation functions*, Optica, 6(9), 2019, s. 1132-1137.
7. Nielsen M.A., *Neural networks and deep learning*, San Francisco, CA, USA: Determination press, 2015.
8. Nwankpa C., Ijomah W., Gachagan A., Marshall S., *Activation functions: Comparison of trends in practice and research for deep learning*, arXiv preprint arXiv:1811.03378. 2018.
9. Pachouly J., Ahirrao S., Kotecha K., Selvachandran G., Abraham, A., *A systematic literature review on software defect prediction using artificial intelligence: Datasets, Data Validation Methods, Approaches, and Tools*, Engineering Applications of Artificial Intelligence, 111, 104773, 2022.
10. Varshney M., Singh P., *Optimizing nonlinear activation function for convolutional neural networks. Signal, Image and Video Processing*, 15(6), 2021, s. 1323-1330.
11. Ding B., Qian H., Zhou J., *Activation functions and their characteristics in deep neural networks*, In 2018 Chinese control and decision conference (CCDC), IEEE, 2018, s. 1836-1841.
12. Sadi M.H., Mahani A., *Accelerating deep convolutional neural network base on stochastic computing*, Integration, 76, 2021, s.113-121.
13. Sharma S., Sharma S., Athaiya A., *Activation functions in neural networks. towards data science*, 6(12), 2017, s. 310-316.
14. Karlik B., Olgac A.V., *Performance analysis of various activation functions in generalized MLP architectures of neural networks*, International Journal of Artificial Intelligence and Expert Systems, 1(4), 2011, s. 111-122.
15. Aigner S., Körner M., *The Importance of Loss Functions for Increasing the Generalization Abilities of a Deep Learning-Based Next Frame Prediction Model for Traffic Scenes*, Machine Learning and Knowledge Extraction, 2(2), 2020, s. 78-98.
16. Cai W., Chen J., Li M., *Exploring the encoding layer and loss function in end-to-end speaker and language recognition system*, arXiv preprint arXiv:1804.05160, 2018.

17. Nie F., Hu Z., Li X., *An investigation for loss functions widely used in machine learning*, Communications in Information and Systems, 18(1), 2018, s. 37-52.
18. Sadi M. H., Mahani A., *Accelerating deep convolutional neural network base on stochastic computing*, Integration, 76, 2021, s. 113-121.
19. Alistarh D., Hoefler T., Johansson M., Konstantinov N., Khirirat S., Renggli C. *The convergence of sparsified gradient methods,* Advances in Neural Information Processing Systems, 31, 2018.
20. Kukačka J., Golkov V., Cremers D., *Regularization for deep learning: A taxonomy*, arXiv preprint arXiv:1710.10686, 2017.
21. Varshney M., Singh P., *Optimizing nonlinear activation function for convolutional neural networks*, Signal, Image and Video Processing, 15(6), 2021, s. 1323-1330.
22. Kim B., Ryu K. H., Heo S., *A neural network regularization method to address variance inflation in autoencoders*, IFAC-PapersOnLine, 55(7), 2022, s. 744-749.
23. Ying X., *An overview of overfitting and its solutions*, Journal of physics: Conference series, Vol. 1168, No. 2, IOP Publishing, 2019.
24. Cilimkovic M., *Neural networks and backpropagation algorithm*, Institute of Technology Blanchardstown, Blanchardstown Road North Dublin, Vol. 15, 2015.

Metody, techniki i problemy sieci neuronowych

Streszczenie
Sieci neuronowe stają się coraz ważniejszą częścią sztucznej inteligencji. W wyniku rozwoju badań nad nimi powstały nowe lub zostały zmodyfikowane modele ich optymalizacji. Artykuł przedstawia najważniejsze metody i techniki wykorzystywane w sieciach neuronowych oraz wskazuje pole do przeprowadzenia dalszych badań nad wydajnością działania algorytmów z ich wykorzystaniem. Omówione zostały zagadnienia związane z funkcją aktywacji, straty, optymalizacją, regularyzacją oraz przetrenowaniem i niedotrenowaniem sieci neuronowych. Artykuł wskazuje również najważniejsze problemy, które uniemożliwiają wydajne zastosowanie tej metody do pracy z różnymi zestawami danych.
Słowa kluczowe: sieci neuronowe, sztuczna inteligencja, uczenie maszynowe, funkcje aktywacji

Methods, techniques and problems of neural networks

Abstract
Neural networks are becoming an increasingly important part of artificial intelligence. As a result of the development of research on them, new or modified models and techniques for their optimization have been created. The article presents the most important methods used in neural networks and indicates the field for further research on the performance of the algorithm using them. I have discussed issues related to the activation function, loss function, optimization, regularization as well as overfitting and underfitting of a neural network. The article also points out the most important problems that prevent the efficient use of this method to work with different datasets.
Keywords: neural networks, artificial intelligence, machine learning, activation functions

Filip Turoboś[1], Mateusz Krukowski[2]

Przybliżone rozwiązania problemu komiwojażera dla grafów semimetrycznych

1. Wprowadzenie

Ostatnie dekady to czas zacieśniania więzów między matematyką dyskretną i teorią przestrzeni metrycznych (szczegółowy opis związku tych dziedzin można znaleźć np. w [1, Preface]). Zasadniczym powodem tego zjawiska jest fakt, iż przestrzenie metryczne dają abstrakcyjne, choć niezwykle użyteczne, narzędzia do pracy nad wieloma zagadnieniami matematyki dyskretnej. Zjawisko to odnotowywano już na przełomie XX i XXI wieku w pracach m.in. Piotra Indyka [2] i Nathana Liniala [3].

Na przestrzeni ostatnich lat społeczność matematyczna zaczęła przywiązywać uwagę nie tylko do przestrzeni metrycznych, ale również do ich potencjalnych uogólnień i możliwych zastosowań w obszarze matematyki dyskretnej. Wśród najpopularniejszych uogólnień wyróżnić należy przestrzenie hemimetryczne, gdzie pomijany jest aksjomat symetrii z klasycznej definicji metryki, tj. „$d(x,y) = d(y,x)$ dla każdej pary punktów $x, y \in X$" [4-6]. Przestrzenie hemimetryczne pojawiają się naturalnie w miejscach, gdzie czas służy do pomiaru odległości między. Innym często pojawiającym się konceptem jest przestrzeń pseudometryczna, gdzie niewymagany jest pierwszy aksjomat metryki:

$$\forall_{x,y \in X}\, d(x,y) = 0 \;\Leftrightarrow x = y,$$

przy jednoczesnym zachowaniu pozostałych dwóch [7]. Okazuje się, że wspomniane aksjomaty pojawiają się naturalnie przy wielu rozważaniach w logice rozmytej [8,9].

Rozwój zastosowań hemimetryk i pseudometryk w różnorodnych gałęziach matematyki jest zjawiskiem niezwykle obiecującym. Podążając za aktualnym trendem w dziedzinie, niniejszy artykuł skupia się na rodzinie uogólnień przestrzeni metrycznych noszących nazwę przestrzeni semimetrycznych. Są to przestrzenie, w których obowiązują jedynie dwa pierwsze akjomaty metryki, a nierówność trójkąta nie jest wymagana.

Rozdział 2 podaje podstawowe własności przestrzeni semimetrycznych uwzględniając fakt, że każda przestrzeń semimetryczna spełnia słabsze warianty nierówności trójkąta, tj. γ −nierówność wielokąta oraz β −nierówność trójkąta. Zgodnie z wiedzą autorów, pierwsze zastosowanie β − nierówności trójkąta w optymalizacji kombinatorycznej datowane jest na rok 1994. Autorzy Bandelt, Crama i Spieksma w pracy [10] wykorzystali tę własność do konstrukcji algorytmu aproksymacyjnego dla wielowymiarowego problemu przydziału z dekomponowalnymi kosztami (bardziej aktualne badania w tym nurcie można znaleźć m.in. w [11]). Idea wykorzystania struktur β − metrycznych szybko zyskała uznanie, a niedługo później wykorzystano je w problemie komiwojażera, co stanowi punkt wyjściowy niniejszego artykułu. Jednocześnie, rozdział 2 opisuje powiązania

[1] filip.turobos@p.lodz.pl, Instytut Matematyki, Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej, Politechnika Łódzka.
[2] mateusz.krukowski@p.lodz.pl, Instytut Matematyki, Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej, Politechnika Łódzka.

przestrzeni semimetrycznych z teorią grafów, wprowadzając jednocześnie podstawowe pojęcia wierzchołka, krawędzi, grafu nieskierowanego i prostegu, wagi, cyklu Hamiltona itd.

Rozdział 3 opisuje metodę uzyskiwania minimalnego drzewa spinającego na grafie. Metoda ta pozwala w Rozdziale 4 na uzyskanie cyklu Hamiltona w grafie spełniającym γ −nierówność wielokąta (Twierdzenie 5). Zgodnie z wiedzą autorów, jest to pierwszy znany przykład w literaturze zastosowania γ −nierówności wielokąta w kontekście problemu komiwojażera.

Uzyskane w Rozdziale 4 przybliżone rozwiązanie jest 2γ −optymalne i rodzi pytanie o możliwość poprawy tego wyniku. Rozdział 5 odpowiada na to pytanie twierdząco i opisuje podejście, będące wariacją na temat algorytmu Christofidesa. Metoda nie tylko wykorzystuje minimalne drzewo spinające (opisane w Rozdziałach 3 i 4), ale również doskonałe skojarzenie o minimalnym koszcie (uzyskiwane za pomocą algorytmu Edmondsa). Końcowe Twierdzenie 8 wykazuje, że γ −nierówność wielokąta pozwala na konstrukcję $\frac{3\gamma}{2}$ −optymalnego rozwiązania przybliżonego problemu komiwojażera. Ponownie jest to obszar dotychczas niezbadany w literaturze naukowej.

2. Wokół przestrzeni semimetrycznych

Definicja. Niech X będzie niepustym zbiorem. Funkcja $d: X \times X \to [0, +\infty)$ nazywana jest semimetryką, jeżeli spełnia następujące warunki:

$$\forall_{x,y \in X}\ d(x,y) = 0 \ \Leftrightarrow\ x = y,$$

$$\forall_{x,y \in X}\ d(x,y) = d(y,x).$$

Parę (X, d) określa się wówczas mianem przestrzeni semimetrycznej.

Literatura omawiająca przestrzenie semimetryczne z czysto teoretycznej perspektywy jest dość bogata, a następujące pozycje można uznać za stosunkowo dobre wprowadzenie do ich teorii [12-19]. Niektórzy autorzy [20-22] dodatkowo wymagają spełnienia przez taką przestrzeń pierwszego aksjomatu przeliczalności. Pomimo, że nie jest to zakładane bezpośrednio w definicji, przestrzenie wykorzystywane w tym artykule są skończone, więc jest on automatycznie spełniony.

Wśród przestrzeni semimetrycznych wygodnie jest wyróżnić te, które wyposażone są w pewien odpowiednik nierówności trójkąta. Dla potrzeb przedstawianej argumentacji będą to przestrzenie β −metryczne i γ −wielokątowe, które definiuje się w sposób następujący:

Definicja. Przestrzeń semimetryczna (X, d) nazywana jest:

- β −metryczną, jeżeli $\beta \geq 1$ jest najmniejszą wartością, dla której d spełnia β −nierówność trójkąta:

$$\forall_{x,y,z \in X}\ d(x,z) \leq \beta\big(d(x,y) + d(y,z)\big), \tag{1}$$

- γ −wielokątową, jeżeli $\gamma \geq 1$ jest najmniejsza wartością, dla której d spełnia γ −nierówność wielokąta:

$$\forall_{n \in \mathbb{N}}\ \forall_{x_1,\ldots,x_n \in X}\ d(x_1,x_n) \leq \gamma \cdot \sum_{k=1}^{n-1} d\,(x_k, x_{k+1}). \tag{2}$$

Obydwa powyższe rodzaje przestrzeni zostały solidnie przebadane od strony teoretycznej w literaturze [14, 15, 17, 19, 23-32]. Zaobserwować można również, że każda przestrzeń γ −wielokątowa jest również przestrzenią β −metryczną dla pewnego $\beta \leq \gamma$ (w dalszej części przedstawiony zostanie przykład, gdy wartości te się nie pokrywają). Dla przestrzeni skończonych prawdziwa jest też implikacja przeciwna, tj. każda przestrzeń β −metryczna spełnia też γ −nierówność wielokąta dla pewnego $\gamma \leq \beta^{|X|-1}$. W artykule [31] można znaleźć lepsze oszacowanie, tj. $\gamma \leq \beta^{\lceil \log_2(|X|-1) \rceil}$ (zakładając, że X ma więcej niż jeden punkt).

Dość łatwą i jednocześnie wciąż użyteczną obserwacją – odnotowaną po raz pierwszy w [15] bez dowodu – jest fakt, że każda skończona przestrzeń semimetryczna spełnia nierówności zawarte w twierdzeniu 1.

Twierdzenie 1. Każda skończona przestrzeń semimetryczna (X, d) spełnia γ −nierówność wielokąta dla

$$\gamma = \sup\left\{\frac{d(x_1,x_n)}{\sum_{k=1}^{n-1} d(x_k,x_{k+1})} : n \in \mathbb{N},\ x_1, \dots, x_n \in X,\ x_1 \neq x_n\right\}, \tag{3}$$

Jak również β −nierówność trójkąta dla

$$\beta = \sup\left\{\frac{d(x,z)}{d(x,y)+d(y,z)} :\ x, y, z \in X,\ x \neq z\right\}. \tag{4}$$

Dowód: Bez straty ogólności zakłada się, że X ma przynajmniej 2 elementy. Niech:

$$s := \sup\left\{\frac{d(x_1,x_n)}{\sum_{k=1}^{n-1} d(x_k,x_{k+1})} : n \in \mathbb{N},\ x_1, \dots, x_n \in X,\ x_1 \neq x_n\right\}. \tag{5}$$

Jest to dobrze określona, skończona wartość spełniająca dla dowolnych $x_1, x_2 \in X$, $x_1 \neq x_2$ nierówność:

$$s \geq \frac{d(x_1,x_2)}{d(x_1,x_2)} = 1. \tag{6}$$

Niech $n \in \mathbb{N}$ i n-elementowy ciąg $x_1, \dots, x_n \in X$ będą ustalone. Rozważyć należy dwa przypadki:

- jeżeli $x_1 = x_n$, to:

$$d(x_1, x_n) = 0 \leq s \cdot \sum_{k=1}^{n-1}\ d(x_k, x_{k+1}); \tag{7}$$

- jeżeli $x_1 \neq x_n$, to:

$$d(x_1, d_n) \quad = \quad \frac{d(x_1, x_n)}{\sum_{k=1}^{n-1}\ d(x_k, x_{k+1})} \cdot \sum_{k=1}^{n-1}\ d(x_k, x_{k+1})$$

$$\leq \quad \sup\left\{\frac{d(y_1,y_n)}{\sum_{k=1}^{n-1} d(y_k,y_{k+1})} : y_1,\dots,y_n \in X,\ y_1 \neq y_n\right\} \cdot \sum_{k=1}^{n-1} d(x_k,x_{k+1})$$

$$= \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad s \cdot \sum_{k=1}^{n-1} d(x_k,x_{k+1}). \quad (8)$$

Zatem $\gamma \leq s$. W drugą stronę, nie wprost można przypuszczać, że istnieje $\varepsilon > 0$ taki, że $\gamma = s - \varepsilon$. Istnieje zatem $n \in \mathbb{N}$ i elementy $x_1, \dots, x_n \in X$ takie, że:

$$\frac{d(x_1,x_n)}{\sum_{k=1}^{n-1} d(x_k,x_{k+1})} > s - \varepsilon. \quad (9)$$

Mnożąc obydwie strony nierówności przez jej mianownik uzyskuje się:

$$d(x_1,x_n) > (s-\varepsilon) \cdot \sum_{k=1}^{n-1} d\,(x_k,x_{k+1}), \quad (10)$$

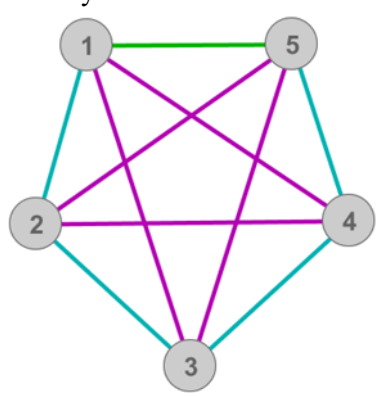
sprzeczność z $\gamma = s - \varepsilon$. Zatem (X,d) spełnia γ −nierówność wielokąta. Dla drugiej części twierdzenia wystarczy przeprowadzić identyczne rozumowanie dla ustalonego $n = 3$.

W tym momencie widocznym jest, że skończona przestrzeń semimetryczna spełnia obydwie te nierówności, przy czym zachodzi $\beta \leq \gamma$. O tym, że stałe te mogą być różne świadczyć może poniższy przykład.

Przykład. Niech $X := \{1,2,3,4,5\}$ i niech $d: X \times X \to [0,+\infty)$ będzie semimetryką zadaną wzorem:

$$d(x,y) := \begin{cases} 0, & x = y, \\ 1, & |x-y| = 1, \\ 20, & |x-y| = 4, \\ 4, & \text{w p.p.} \end{cases} \quad (11)$$

Powyższa przestrzeń semimetryczna zwizualizowana została na rysunku 1.



Rysunek 1. Ilustracja przestrzeni (X,d), niebieskie krawędzie odpowiadają odległości 1, fioletowe odpowiadają odległości 4, zaś zielona – 20, opracowanie własne

Powyższe przekształcenia pozwalają zauważyć, że (X, d) jest przestrzenią $4-$ metryczną, ale nie jest $4-$ wielokątowa, ponieważ dla ciągu skończonego $(1,2,3,4,5)$ można zaobserwować, że:

$$d(x_1, x_5) = 20 \leq \gamma \cdot \sum_{k=1}^{4} d\,(x_k, x_{k+1}) = 4\gamma \qquad (1)$$

To dowodzi, że $\gamma \geq 5$. Nietrudno jest wykazać, że 5 jest optymalną wartością tej stałej.

Z teoretycznego punktu widzenia sama świadomość istnienia stałych β i γ w zupełności wystarcza do pracy nad *matematyczną abstrakcją*. W praktyce jednak znajomość dokładnych wartości tychże stałych jest niezwykle przydatna. W dalszej części przedstawiony zostanie algorytm, który pozwala obliczać wartości tych stałych w czasie wielomianowym.

2.1. Teoria grafów z semimetrycznego punktu widzenia

Mając do dyspozycji warsztat semimetryczny w niniejszej części pracy przedstawione zostaną niezbędne pojęcia z teorii grafów, jednocześnie wprowadzając stosowną konwencję notacyjną [33, 34].

Definicja. Graf definiowany jest jako para $G := (V, E)$, gdzie V jest skończonym, niepustym zbiorem, zaś $E \subset \{\{x, y\} : x, y \in V,\ x \neq y\}$. Elementy V i E nazywane są, odpowiednio, wierzchołkami i krawędziami. Graf ważony, to para postaci (G, ω), gdzie $G = (V, E)$ jest grafem, $E \neq \emptyset$, zaś $\omega: E \to (0, +\infty)$ jest nieujemną funkcją, nazywaną wagą.

Dla grafu G, jego zbiór wierzchołków i krawędzi bywają również oznaczane symbolami $V(G)$ i $E(G)$, odpowiednio [33]. Jest to istotne w sytuacji, gdy rozpatrywane jest kilka grafów jednocześnie. Czytelnik zapoznany z terminologią teorii grafów z pewnością zauważy też, że rozważane struktury są *nieskierowane* i *proste*. „*Nieskierowanie*" oznacza brak orientacji krawędzi grafu, tj. krawędź rozumiana jest jako zbiór $\{x, y\}$, a nie para uporządkowana (x, y). Z kolei grafy „*proste*" to takie, w których nie występują „*pętle*" (czyli krawędzie od wierzchołka do niego samego) lub „*wielokrotnych krawędzi*" (czyli E jest zwykłym zbiorem, a nie multizbiorem). Zapotrzebowanie na tego typu obiekty pojawi się w końcowej części artykułu, natomiast poniżej wprowadzona zostanie odpowiednia notacja.

Definicja. Para $\mathbb{G} := (V, \mathbb{E})$ nazywana jest multigrafem, jeżeli V jest skończonym, niepustym zbiorem, zaś $\mathbb{E} \subset \{(k, \{x, y\}) : x, y \in V,\ x \neq y,\ k \in \mathbb{N}_0\}$. Elementy $\mathbb{E}$ nazywane są multikrawędziami (V pozostaje zbiorem wierzchołków). Multigraf ważony to para $(\mathbb{G}, \omega)$, gdzie $\mathbb{G} = (V, \mathbb{E})$ jest multigrafem, $\mathbb{E} \neq \emptyset$, a $\omega: \mathbb{E} \to (0, +\infty)$ jest wagą.

Indeks k w powyższej definicji służy do odróżnienia poszczególnych krawędzi łączących tę samą parę wierzchołków. Co ważniejsze, proponowana definicja wagi w obydwu definicjach może odbiegać nieco od tej z literatury z powodu założenia, że ω jest funkcją dodatnią. Przykładowo, Bondy i Murty w [35] dopuszczają ω o wartościach rzeczywistych, zaś Bläser [36] zezwala jedynie na wagi o nieujemnych wartościach wymiernych.

Przechodząc dalej, wprowadzona zostanie seria powiązanych pojęć [33, 35, 37].

Definicja. Niech G będzie grafem.

- Graf F jest podgrafem G, jeżeli $V(F) \subset V(G)$ i $E(F) \subset E(G)$, co oznaczane jest przez $F \subset G$.

- Graf P nazywany jest ścieżką, jeżeli jego wierzchołki P można uszeregować w ciąg tak, aby dwa wierzchołki były połączone krawędzią wtedy i tylko wtedy, gdy następują po sobie w owym ciągu.
- Graf C jest cyklem, jeżeli usunięcie dowolnej krawędzi z C przekształca go w ścieżkę. Jeżeli G nie zawiera żadnego cyklu jako podgrafu, to nazywany jest grafem acyklicznym.
- Podgraf $H \subset G$ nazywany jest ścieżką/cyklem Hamiltona, jeżeli jest ścieżką/cyklem i odwiedza każdy z wierzchołków G dokładnie raz. Rodzina wszystkich cykli Hamiltona w grafie G oznaczana będzie przez $\mathfrak{h}(G)$. Jeżeli zachodzi $\mathfrak{h}(G) \neq \emptyset$ to G określany jest mianem grafu Hamiltonowskiego.

Poniżej przytoczona definicja ścieżki pozwala wprowadzić jeszcze jedno istotne pojęcie.

Definicja. Graf G nazywany jest spójnym, jeżeli dla dowolnych $x, y \in V(G)$ istnieje ścieżka $P \subset G$ taka, że $x, y \in V(P)$.

Wprowadzone pojęcie podgrafu jest inherentnie niezależne od funkcji wagi w grafie G (o ile została zdefiniowana). Niemniej, jeżeli ω jest wagą w G i F jest podgrafem G, to wówczas $\left(F, \omega|_{E(F)}\right)$ jest grafem ważonym, zaś $\omega|_{E(F)}$ określić należy mianem *wagi indukowanej*. Ponadto, wygodną notacją dla dalszych potrzeb będzie waga podgrafu (lub całego grafu), którą definiuje się jako:

$$\omega(F) := \sum_{e \in E(F)} \omega\,(e). \tag{13}$$

Jest to rzecz jasna pewne nadużycie notacyjne, bowiem ten sam symbol „ω" posłuży do oznaczania zarówno wag krawędzi, jak podgrafów. Formalnie może być to postrzegane za niewłaściwe, bowiem w roli argumentów ω pojawiają się różne typy obiektów. Konwencja ta nie powinna jednak prowadzić do nieporozumień, zwłaszcza że krawędzie oznaczane są przez małe, a grafy przez wielkie litery alfabetu łacińskiego.

Powyższe pojęcia pozwalają na zdefiniowanie problemu komiwojażera (skrótowo oznaczanego TSP, ang. *Travelling Salesperson Problem*), stanowiącego centralne zagadnienie w niniejszym artykule:[3]

Dla ważonego grafu Hamiltonowskiego (G, ω) ***znaleźć cykl Hamiltona*** H ***o możliwie najmniejszej wadze.***

TSP stanowi jedną z najsłynniejszych zagadek matematycznych, której historia dalece wykracza poza spektrum tego artykułu. Zainteresowanego Czytelnika mogą zaciekawić odnośniki do literatury, gdzie omawiana jest zarówno historia samego problemu, jak również próby jego rozwiązania [39, 42-49].

Przedstawione sformułowanie TSP zakłada odgórnie $\mathfrak{h}(G) \neq \emptyset$. Nie ma wszak sensu szukać cyklu Hamiltona o minimalnej wadze, jeżeli w grafie nie ma żadnego takiego cyklu. Niemniej, w dalszej części artykułu analizowane będą jedynie grafy pełne, które w trywialny sposób spełniają $\mathfrak{h}(G) \neq \emptyset$. Pomimo pozornej restrykcyjności tego podejścia, okazuje się, że nadal stanowi ono równoważny problem do wyjściowego. Powodem, dla którego można dokonać takiego ograniczenia jest fakt, że zawsze istnieje możliwość zdefiniowania $\omega(\{x, y\})$ jako „*odpowiednio dużej wartości*" dla krawędzi $\{x, y\}$, które

[3] W tym miejscu warto podkreślić, że dalece nie jest to jedyny sposób definiowania problemu komiwojażera, przykłady równoważnych sformułowań można znaleźć w pracach [37-40].

nie występują w oryginalnym grafie G. Formalne ujęcie tej powszechnie znanej intuicji można przedstawić za pomocą twierdzenia 2.

Twierdzenie 2. Niech (G, ω_G) będzie grafem o $n := |V(G)| \geq 3$ wierzchołkach. Wówczas istnieje graf pełen (K_n, ω) taki, że:

1. G jest podgrafem K_n i $\omega|_G = \omega_G$.
2. Jeżeli H jest cyklem Hamiltona o minimalnej wadze w (K_n, ω), to zachodzi dokładnie jedna z poniższych sytuacji:
   a. $\omega(H) \leq \omega(G)$ i H jest też cyklem Hamiltona o minimalnej wadze w G;
   b. $\omega(H) > \omega(G)$ i G nie jest Hamiltonowski.

Dowód. Niech K_n będzie grafem pełnym na wierzchołkach grafu G i niech ω będzie zdefiniowana na K_n w następujący sposób:

$$\omega(\{x, y\}) := \begin{cases} \omega_G(\{x, y\}), & \text{jeżeli } \{x, y\} \in E(G), \\ \sum_{e \in E(G)} \omega_G(e), & \text{w p.p.} \end{cases} \tag{14}$$

Widać z powyższego, że G jest podgrafem K_n i $\omega|_G = \omega_G$, tym samym kończąc dowód pierwszej części twierdzenia.

Na potrzeby drugiej części, niech H będzie cyklem Hamiltona o minimalnej wadze w (K_n, ω). Rozpatrzyć należy następujące przypadki:

- Jeżeli $\omega(H) \leq \omega(G)$, to H nie może zawierać krawędzi niewystępujących w $E(G)$, ponieważ w przeciwnym wypadku:

  $$\omega(H) > \omega(G) = \sum_{e \in E(G)} \omega(e). \tag{15}$$

  Co więcej, nie istnieje cykl Hamiltona H' w (G, ω_G), taki że $\omega(H') < \omega_G(H)$, ponieważ taki cykl byłby również cyklem Hamiltona o tej samej wadze w (K_n, ω). Zatem H jest cyklem Hamiltona o minimalnej wadze w (G, ω_G).
- Jeżeli $\omega(H) > \omega(G)$, to istnienie cyklu Hamiltona H' w G implikowałoby:

  $$\omega(H') \leq \omega(G) < \omega(H). \tag{16}$$

  To jednak przeczy minimalności H w K_n (ponieważ H' również byłby cyklem Hamiltona w K_n), co kończy dowód.

W świetle powyższego, możliwym jest ograniczenie się do sformułowania TSP na pełnych grafach ważonych, tj. $G = (K_n, \omega)$. Na takich grafach naturalnie wprowadzić można strukturę semimetryczną w intuicyjny sposób:

$$d(x, y) := \begin{cases} \omega(\{x, y\}), & \text{where } x \neq y, \\ 0, & \text{where } x = y. \end{cases} \tag{17}$$

Konkludując tę część, wprowadzona zostanie metryczna terminologia do grafów ważonych, mianowicie graf $G = (K_n, \omega)$ nazywany będzie:

- metrycznym, jeżeli d jest metryką na $V(G)$,
- β −metrycznym, jeżeli d spełnia β −nierówność trójkąta;
- γ −metrycznym, jeżeli d spełnia γ −nierówność wielokąta.

2.2. Optymalizacja i teoria złożoności

Mając na uwadze to, że niniejszy artykuł może być obiektem zainteresowania zarówno Czytelników zaznajomionych z algorytmiką, jak również teoretyków przestrzeni

metrycznych, poniżej przedstawione zostało krótkie streszczenie niezbędnych w dalszej części pojęć z optymalizacji i teorii złożoności. Punktem wyjścia będzie definicja problemu optymalizacyjnego (w tym miejscu warto nadmienić, że w niektórych monografiach poniższa definicja ma zgoła inną formę, *vide* [50]. Szczegółowe omówienie problemów optymalizacyjnych można znaleźć m.in. w [51, 52]).

Definicja. Niech X będzie niepustym, skończonym zbiorem i niech $f: X \to (0, +\infty)$ będzie dowolną funkcją. Problem znalezienia elementu $x^* \in X$ spełniającego

$$f(x^*) = \min_{x \in X} f(x), \tag{18}$$

nazywany jest problemem optymalizacyjnym.

Zbiór X i funkcja f w problemie optymalizacyjnym nazywane są odpowiednio *przestrzenią poszukiwań* i *funkcją celu*. Każdy element spełniający równość w definicji nosi miano *rozwiązania optymalnego* (może istnieć więcej niż jeden taki element). Elementy X nazywane są *rozwiązaniami dopuszczalnymi*.

Skończoność zbioru X sugerować może, że problemy optymalizacyjne w myśl wprowadzonej definicji są trywialne. Czy istnieją realne przeszkody w tym, żeby znaleźć minimalną wartość $f(x)$, sprawdzając kolejno wszystkie elementy X? O ile w teorii jest to całkowicie poprawne podejście, o tyle w praktyce jest to zbyt kosztowne i czasochłonne do przeprowadzenia. Żeby przedstawić skalę problemu, warto zastanowić się nad poniższym przykładem. TSP jest w rozumieniu powyższej definicji problemem optymalizacyjnym – wystarczy dla zadanego grafu (G, ω) położyć $X := \mathfrak{h}(G)$ i $f := \omega$. Już dla $G = K_{61}$ istnieje $\frac{1}{2} \cdot 60!$ różnych cykli Hamiltona, (dla K_n liczba ta wynosi $\frac{(n-1)!}{2}$), co jest liczbą większą niż liczba cząsteczek w obserwowalnym wszechświecie – wynoszącą około 10^{80})! Przeszukanie wszystkich rozwiązań dopuszczalnych wydaje się być nierealne.

Zamiast poszukiwać rozwiązania optymalnego, możliwym jest ograniczenie się jedynie do znajdowania *rozwiązań przybliżonych*, które (w pewnym sensie) będą wystarczająco dobre. Zazwyczaj jednak obiektem poszukiwań jest *metoda aproksymacyjna*, tj. algorytm, który można zastosować do całej rodziny takich problemów optymalizacyjnych.

Definicja. Niech $s \geq 1$ i niech $\mathcal{P}$ będzie zbiorem indeksów dla pewnej rodziny problemów optymalizacyjnych o funkcjach celu $f_p: X_p \to (0, +\infty)$, $p \in \mathcal{P}$. Odwzorowanie $\mathfrak{a}: \mathcal{P} \to \bigcup_{p \in \mathcal{P}} X_p$ nazywane jest s-optymalną aproksymacją dla tej rodziny, jeżeli dla każdego $p \in \mathcal{P}$ zachodzi:

$$\mathfrak{a}(p) \in X_p \text{ i } \frac{f_p(\mathfrak{a}(p))}{f_p(x_p^*)} \leq s, \tag{19}$$

gdzie x_p^* jest rozwiązaniem optymalnym dla p −tego problemu. Element $\mathfrak{a}(p) \in X_p$ nazywany jest wówczas s −optymalną aproksymacją dla p −tego problemu z tej rodziny.

Zainteresowany Czytelnik jednocześnie zaproszony jest do porównania powyższej definicji z podejściem opisanym w monografiach [50] oraz [53].

Dysponując powyższą terminologią, możliwym jest przejście do właściwej części artykułu, poświęconej eksploracji dwóch s −optymalnych metod aproksymacyjnych dla TSP, tj.

- metody minimalnego drzewa spinającego dla metrycznych i γ −metrycznych grafów;

- modyfikacji algorytmu Christofidesa dla grafów γ −metrycznych.

Zanim jednak przeprowadzone zostaną analizy obu technik, poniżej krótko zaprezentowane zostały podstawy złożoności algorytmicznej. Mając na uwadze ograniczoność tego wprowadzenia, dociekliwy Czytelnik może czuć się zachęcony do eksploracji szerokiej literatury dotyczącej tej tematyki, zawierają ją prace [54-58]. Poniższa definicja jest pewną modyfikacją wersji podanej w [59].

Definicja. Niech $\mathcal{P}$ będzie zbiorem indeksów dla pewnej rodziny problemów optymalizacyjnych, zaś $\mathcal{M}: \mathcal{P} \to \mathbb{N}$ będzie funkcją, która każdemu $p \in \mathcal{P}$ przypisuje liczbę jednostek pamięci (w danym modelu obliczeniowym, np. klasycznej maszynie Turinga) użytych do opisu p −tego problemu. Algorytm $\mathfrak{a}: \mathcal{P} \to \mathbb{R}$ ma złożoność obliczeniową (czasową) nie większą niż $F_{\mathfrak{a}}: \mathbb{N} \to \mathbb{N}$, jeżeli istnieje stała $\alpha \geq 0$, taka że dla p −tego problemu z tej rodziny algorytm ten wykonuje nie więcej niż $\alpha \cdot F_{\mathfrak{a}}\big(\mathcal{M}(p)\big)$ operacji. Wówczas o $\mathfrak{a}$ można powiedzieć, że jest wykonywalny w czasie $\mathcal{O}\big(F_{\mathfrak{a}}(n)\big)$.

Czas powrócić do poruszonego wcześniej problemu obliczania stałych β i γ dla skończonej przestrzeni semimetrycznej. Do znalezienia β wystarcza obliczenie wszystkich ilorazów postaci $\frac{d(x,z)}{d(x,y)+d(y,z)}$ i wybór największego z nich. To rozwiązanie metodą siłową ma złożoność $\mathcal{O}(n^3)$, gdzie n jest kardynalnością przestrzeni X.

Problem znajdowania stałej γ jest nieco bardziej złożony, ale daje się go rozwiązać w tym samym czasie, co jest tezą poniższego twierdzenia 3.

Twierdzenie 3. Niech (X, d) będzie skończoną przestrzenią semimetryczną o n punktach. Stała γ, dla której (X, d) spełnia γ – nierówność wielokąta może być znaleziona w czasie $\mathcal{O}(n^3)$.

Dowód: Niech $G := (K_n, \omega)$ będzie pełnym grafem ważonym, gdzie waga ω zdefiniowana jest jako $\omega(\{x, y\}) := d(x, y)$ dla wszystkich $x, y \in V(K_n) = X$. Wykonanie algorytmu znajdowania najkrótszych ścieżek techniką Floyda-Warshalla (której obszerny opis można znaleźć w [59], zaś sam algorytm pochodzi z artykułu [60]) zwraca funkcję $D: X \times X \to [0, \infty)$, gdzie $D(x, y)$ jest długością najkrótszej ścieżki łączącej punkty x i y, tj.

$$D(x, y) = \min\Big\{\sum_{k=1}^{n-1} d\,(x_k, x_{k+1}) : n \in \mathbb{N},\ x_1 = x,\ x_n = y,\ x_k \in X \text{ for } k \leq n\Big\}. \quad (20)$$

Ponadto, złożoność obliczeniowa tego algorytmu wynosi $\mathcal{O}(n^3)$. Niech funkcja $\Gamma: \{(x, y) \in X \times X : x \neq y\} \to [0, \infty)$ będzie zadana wzorem:

$$\Gamma(x, y) := \frac{d(x,y)}{D(x,y)}. \quad (21)$$

Każda krawędź G ma dodatnią wagę, więc Γ jest dobrze określona. Ponadto $\Gamma(x, y) \geq 1$ bo najkrótsza ścieżka łącząca parę różnych elementów x i y jest nie dłuższa niż $\{x, y\}$, innymi słowy $D(x, y) \leq d(x, y)$. Co więcej:

$$\gamma = \max_{(x,y)} \Gamma(x, y). \quad (22)$$

Znalezienie tego maksimum wymaga przejrzenia $\frac{n(n-1)}{2}$ wartości funkcji Γ (z uwagi na symetrię $\Gamma(x, y) = \Gamma(y, x)$). Złożoność czasowa całej procedury (z uwzględnieniem algorytmu Floyda-Warshalla i bez wykorzystania obliczeń równoległych) wynosi więc $\mathcal{O}(n^3)$.

Powyższe twierdzenie uzasadnia, dlaczego grafy γ – wielokątne są obiektem przedstawianych badań – każdy pełny graf skończony spełnia bowiem postulowany warunek dla pewnego γ! Oznacza to, że poprawiona wersja metody Christofidesa będzie działać dla dowolnego grafu pełnego, nawet bez spełniania warunku trójkąta. Ponadto, grafy które nie są pełne mogą zostać stosownie uzupełnione zgodnie z Twierdzeniem 2, tak więc przedstawiona w dalszej części pracy technika może działać na dowolnych grafach skończonych.

3. Minimalne drzewo spinające w grafach metrycznych, $\beta -$ metrycznych i $\gamma -$metrycznych

Kluczowym pojęciem, które będzie wykorzystywane w tej części artykułu jest definicja minimalnego drzewa spinającego:

Definicja. Niech G będzie grafem.

- Drzewo definiowane jest jako spójny, acykliczny podgraf $T \subset G$.
- Drzewo T spełniające $V(T) = V(G)$ będzie nazywane drzewem spinającym dla grafu G. Rodzina wszystkich drzew spinających dla G oznaczana jest przez $\mathcal{S}(G)$.

Jeżeli (G, ω) jest grafem ważonym, to dowolne $T \in \mathcal{S}(G)$ spełniające

$$\omega(T) = \min_{T' \in \mathcal{S}(G)} \omega(T') \tag{23}$$

nazywane jest minimalnym drzewem spinającym (w skrócie MST, ang. *Minimal Spanning Tree*).

Przed wprowadzeniem heurystyki minimalnego drzewa spinającego, przydatnym pojęciem będzie koncept przejścia po drzewie, który znacząco ułatwi dalsze rozważania (definicja bazuje na [33]).

Definicja. Niech G będzie grafem. n-elementowy ciąg wierzchołków $(x_1, \dots, x_n)$ taki, że $\{x_k, x_{k+1}\} \in E(G)$ dla wszystkich $k < n$ nazywany będzie spacerem (w grafie G). Spacer ten jest zamknięty, jeżeli $x_1 = x_n$. Jeżeli T jest drzewem w G, to dowolny spacer w T, który wykorzystuje każdą z krawędzi dokładnie dwukrotnie stanowić będzie *przejście* po drzewie T.

Powyższe pojęcia umożliwiają sformułowanie ogólnego *framework* dla metody aproksymacyjnej minimalnego drzewa spinającego (skrótowo metody MST), która zaprezentowano w poniżej omówionych krokach.

Krok 1. Dla pełnego grafu ważonego $G := (K_n, \omega)$ znaleźć dowolne minimalne drzewo spinające T.

Krok 2. Wykorzystując metodę przeszukiwania wgłąb (skrótowo DFS, ang. *depth-first search*) skonstruować przejście po drzewie T (zależne od wybranego początkowo wierzchołka x_0).

Krok 3. Wykonać procedurę skracania na uzyskanym przejściu, uzyskując cykl Hamiltona na grafie G.

Pierwsze dwa kroki tego podejścia są szeroko znane, a ich opis można znaleźć m.in. w [58, 59, 61, 62]. Punktem wyjścia dla dalszych rozważań będzie zatem krok 3, tj. procedura skracania.

Definicja. Niech $G := (K_n, \omega)$ będzie pełnym grafem ważonym o $n \in \mathbb{N}, n \geq 3$ wierzchołkach i niech T będzie minimalnym drzewem spinającym w G. Dla wybranego wierzchołka $x_1 \in V(T)$, który określony będzie mianem *korzenia*, wykonanie metody

DFS skutkuje uzyskaniem przejścia T postaci $(x_1, \dots, x_{2n-1})$. Rezultatem procedury skracania (skrótem) jest ciąg wierzchołków $(y_1, \dots, y_n, y_{n+1})$ powstały z $(x_1, \dots, x_{2n-1})$, zgodnie z następującym wzorem:

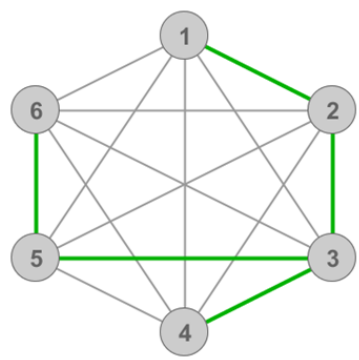
$$y_i := \begin{cases} x_1, & i = 1, n+1, \\ x_{m(i)}, & i = 2, \dots, n, \end{cases} \tag{24}$$

gdzie $m(i) := \min\{j: i \leq j \leq 2n-1 \text{ and } x_j \notin \{x_1, \dots, x_{j-1}\}\}$.

Zapisany powyżej skrót $(y_1, \dots, y_n, y_{n+1})$ można uzyskać wykreślając powtarzające się wierzchołki z ciągu $(x_1, \dots, x_{2n-1})$. Jedynym wyjątkiem jaki należy uwzględnić jest wierzchołek x_{2n-1}, który jest taki sam jak x_1. W skróconym ciągu $S := (y_i)_{i=1}^{n+1}$ każdy wierzchołek poza korzeniem występuje dokładnie raz. Zatem S jest w rzeczywistości cyklem Hamiltona w grafie G, który nazywany będzie w dalszej części artykułu (T, x_1) − skrótem G.

Warto w tym miejscu podkreślić, że zarówno dobór drzewa spinającego, jak również jego korzenia mają wpływ na ostateczny kształt wynikowego cyklu Hamiltona. Aby zilustrować ten fakt, warto przeanalizować poniższy przykład.

Przykład: Niech dany będzie graf pełny K_6 o wierzchołkach $V := \{1, \dots, 6\}$. Niech T będzie drzewem spinającym o krawędziach $E(T) = \{\{1,2\}, \{2,3\}, \{3,4\}, \{3,5\}, \{5,6\}\}$.



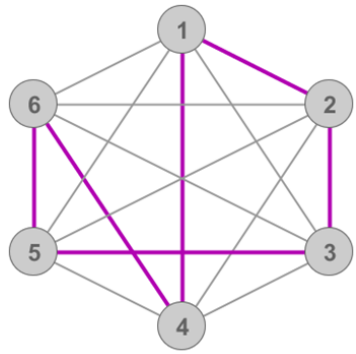
Rysunek 2. Ilustracja minimalnego drzewa spinającego T, opracowanie własne

W pierwszej kolejności rozważane będzie przejście zaczynające się od wierzchołka 1. Kolejne odwiedzane wierzchołki to 2 i 3, aby w kolejnym kroku dokonać wyboru między wierzchołkiem 5 lub 4. Wybierając wierzchołek 5, następny w kolejności pozostaje wierzchołek 6. Po powrocie do wierzchołka 3, odwiedzany jest wierzchołek 4. Następnie wracając aż do 1, uzyskiwane jest przejście R_1:

$$R_1 := (1,2,3,5,6,5,3,4,3,2,1). \tag{25}$$

Wykonując procedurę skracania, usuwane są powtórzenia wierzchołków (oprócz ostatniej jedynki). Rezultatem końcowym tego procesu jest cykl Hamiltona S_1:

$$S_1 := (1,2,3,5,6,4,1). \tag{26}$$



Rysunek 3. Ilustracja cyklu Hamiltona S_1 [opracowanie własne]

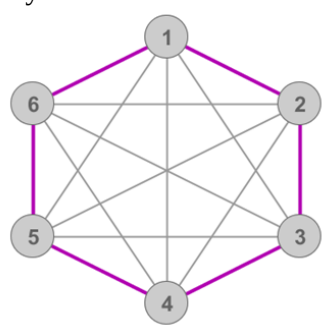
Ciąg S_1 nie jest jedynym $(T, 1)$ −skrótem, bowiem wybór wierzchołka 5 nie był jedynym możliwym. Wybierając w jego miejsce wierzchołek 4, cofając się do 3 i dopiero wtedy odwiedzając 5 i 6, uzyskiwane jest nieco inne przejście R_2, tj.

$$R_2 := (1,2,3,4,3,5,6,5,3,2,1). \tag{27}$$

Po skróceniu tego przejścia otrzymywany jest poniżej wskazany cykl Hamiltona:

$$S_2 := (1,2,3,4,5,6,1), \tag{28}$$

który został przedstawiony na rysunku 4.



Rysunek 4. Ilustracja cyklu Hamiltona S_2 [opracowanie własne]

4. Efektywność heurystyki minimalnego drzewa spinającego

Na podstawie powyższej sekcji wiadomym staje się, że metoda MST zwraca cykl Hamiltona dla zadanego grafu G. Czas zatem porównać jakość uzyskiwanego w ten sposób rozwiązania TSP z wynikiem optymalnym. Dobrze znaną obserwację pozwalającą zestawić ze sobą wagi rozwiązania optymalnego problemu komiwojażera C^* i wagi minimalnego drzewa spinającego T można zwięźle zapisać w postaci poniższej nierówności:

$$\omega(T) \leq \omega(C^*). \tag{29}$$

Dowód tego faktu można skrócić do zauważenia, że usunięcie pojedynczej krawędzi z cyklu C^* daje pewne drzewo spinające T^*, którego koszt (z definicji MST) jest niemniejszy niż $\omega(T)$. W dalszej części zostanie wprowadzony klasyczny rezultat dla heurystyki MST w kontekście grafów metrycznych.

Twierdzenie 4. Niech (K_n, ω) będzie pełnym grafem metrycznym, zaś T jego minimalnym drzewem spinającym, z wyróżnionym wierzchołkiem $x \in V(T)$. Dowolny $(T, x) -$ skrót S spełnia wówczas nierówność:

$$\omega(S) \leq 2\omega(T) \leq 2\omega(C^*). \tag{30}$$

Dowód: Niech $(x_k)_{k=1}^{2n-1}$ będzie przejściem po drzewie T i niech $(y_k)_{k=1}^{n+1}$ będzie ciągiem wierzchołków tworzących (T, x) −skrót S. Dla każdego $i \leq n$ zachodzi:

$$\begin{aligned} d(y_i, y_{i+1}) \quad &= d\big(x_{m(i)}, x_{m(i+1)}\big) \leq d\big(x_{m(i)}, x_{m(i)+1}\big) + d\big(x_{m(i)+1}, x_{m(i+1)}\big) \\ &\leq \cdots \leq \textstyle\sum_{j=m(i)}^{m(i+1)-1} d\left(x_j, x_{j+1}\right), \end{aligned} \tag{31}$$

co implikuje:

$$\begin{aligned} \omega(S) = \sum_{i=1}^{n} d\left(y_i, y_{i+1}\right) \leq \sum_{i=1}^{n} \sum_{j=m(i)}^{m(i+1)-1} d\left(x_j, x_{j+1}\right) \leq \sum_{k=1}^{2n-2} d\left(x_k, x_{k+1}\right) \\ = 2\omega(T). \end{aligned} \tag{2}$$

W połączeniu z obserwacją z początku tej sekcji artykułu kończy to dowód.

W języku teorii optymalizacji, Twierdzenie 4 świadczy o tym, że metoda MST stanowi $2 -$ optymalną aproksymację dla TSP na grafach metrycznych. Założenie metryczności grafu jest kluczowe, ponieważ w oszacowaniu (31) wykorzystywana jest wielokrotnie nierówność trójkąta.

Celem autorów w tej sekcji pracy będzie więc wykazanie odpowiednika powyższego twierdzenia dla grafów $\gamma -$ wielokątnych. Jego dowód przebiega w bardzo zbliżony sposób.

Twierdzenie 5. Niech (K_n, ω) będzie pełnym grafem spełniającym $\gamma -$ nierówność wielokąta, T jego minimalnym drzewem spinającym o wyróżnionym wierzchołku $x_1 \in V(T)$. Dowolny $(T, x_1) -$ skrót S spełnia oszacowanie:

$$\omega(S) \leq 2\gamma\omega(T) \leq 2\gamma\omega(C^*). \tag{33}$$

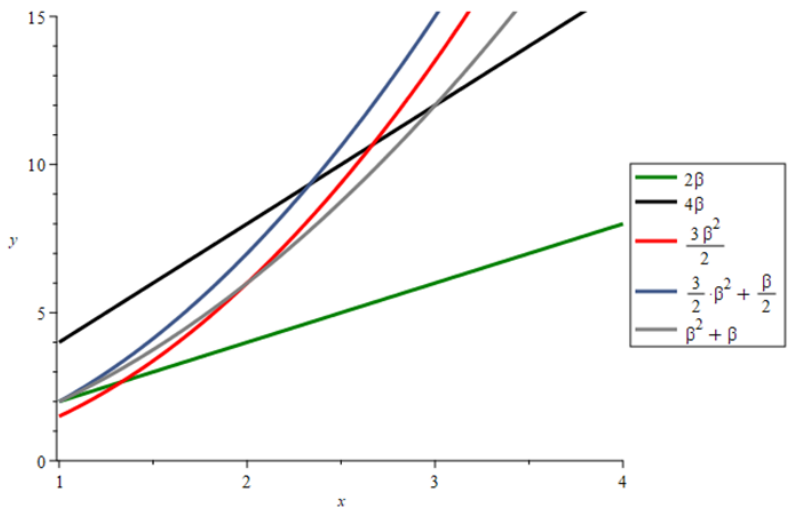
Dowód. Ponownie niech $(x_k)_{k=1}^{2n-1}$ będzie przejściem po drzewie T, z kolei niech $S := (y_k)_{k=1}^{n+1}$ będzie jego $(T, x_1) -$ skrótem. Wykorzystując założenie o γ – metryczności grafu G uzyskiwana jest nierówność:

$$\forall_{i=1,\ldots,n}\ d(y_i, y_{i+1}) = d\big(x_{m(i)}, x_{m(i+1)}\big) \leq \gamma \textstyle\sum_{j=m(i)}^{m(i+1)-1} d\big(x_j, x_{j+1}\big). \tag{34}$$

W konsekwencji dostajemy kończącą dowód nierówność:

$$\begin{aligned} \omega(S) = \sum_{i=1}^{n} d(y_i, y_{i+1}) \leq \gamma \sum_{i=1}^{n} \sum_{j=m(i)}^{m(i+1)-1} d\big(x_j, x_{j+1}\big) \\ \leq \gamma \textstyle\sum_{k=1}^{2n-2} d(x_k, x_{k+1}) \leq 2\gamma\omega(T) \end{aligned} \tag{35}$$

Z perspektywy teorii optymalizacji, powyższe twierdzenie określa, że heurystyka MST stanowi $2\gamma -$ optymalną aproksymację dla TSP. Naturalnym wydawać się może w tym miejscu pytanie, jak owo oszacowanie plasuje się na tle innych estymat znanych z literatury. Dotychczas stosowania idea struktur $\gamma -$ wielokątnych w grafie nie była jeszcze (wg wiedzy autorów) eksplorowana w kontekście TSP przez społeczność matematyczną. Zamiast tego badacze skupiali się raczej na $\beta -$ metrycznych grafach, gdzie uzyskano cztery metody aproksymacyjne, które zdobyły największe uznanie na tym polu. Pierwsza z nich, stanowiąca $\frac{3\beta^2+\beta}{2} -$ optymalną metodę aproksymacyjną została zaproponowana w 1995 przez Andreae i Bandelta [24]. Dzisiaj ma jednak głównie wartość historyczną, ponieważ w 2001 roku Andreae samodzielnie uzyskał $(\beta^2 + \beta)$ aproksymację [63]. W 2000 Bender i Chekuri [64] zaproponowali 4β-optymalne przybliżenie, zaś Böckenhauer, Hromkovič, Klasing, Seibert i Unge wprowadzili $\frac{3\beta^2}{2} -$ optymalną metodę aproksymacji w 2002 (chociaż oryginalny opis algorytmu z pracy [65] zawierał błąd, 11 lat później został poprawiony w artykule poświęconym analizie tej heurystyki [66]). Dyskusję „która technika aproksymacyjna jest najlepsza?" prezentują krzywe na rysunku 5.



Rysunek 5. Zestawienie oszacowań dla grafów $\beta -$ metrycznych [opracowanie własne]

Przykładowo, $\frac{3\beta^2}{2} -$ optymalna metoda aproksymacyjna (czerwona krzywa) jest gorsza niż $(\beta^2 + \beta) -$ aproksymacja (kolor jasnoszary) dla $\beta > 2$, ale w sytuacji gdy $\beta < 2$ okazuje się być lepsza. Gdzie oszacowanie postaci 2γ plasuje się na tym wykresie dla struktur β-metrycznych?

Odpowiedź na to pytanie wymusza konieczność założenia pewnego rodzaju zależności między β i γ. Jak zauważono już wcześniej, prawdziwa jest nierówność $\gamma \geq \beta$, ale tym razem potrzebne będzie oszacowanie z góry na γ jako funkcję zmiennej β. Ponieważ w pełnej ogólności takie oszacowanie nie istnieje, rozpatrywany będzie szczególny przypadek, gdy $\gamma \in [\beta, 2\beta)$ i $\beta \geq 3$. Te warunki wydają się być stosunkowo mało restrykcyjne, co zostało zilustrowane w poprzedniej części artykułu przykładem grafu,

dla którego są one spełnione. Istotną obserwacją jest, że gdy $\gamma \in [\beta, 2\beta)$ i $\beta \geq 3$ to prawdziwy jest ciąg nierówności następującej postaci:

$$2\beta \leq 2\gamma < 4\beta \leq \beta^2 + \beta < \frac{3}{2}\beta^2 < \frac{3\beta^2+\beta}{2}. \tag{36}$$

Oznacza to, że istnieje szeroka klasa grafów, dla których metoda aproksymacyjna bazująca na MST jest najlepszą dotychczas techniką znaną w literaturze. Co więcej, weryfikacja, czy zadany graf spełnia te warunki jest relatywnie łatwa i szybka – wystarczy bowiem wyznaczyć stałe β i γ dla tego grafu (zgodnie z Twierdzeniem 3.), po czym sprawdzić czy warunki $\beta > 3$ i $\gamma < 2\beta$ zostały spełnione. Jeżeli tak, zastosowanie metody MST wydaje się być najlepszym z dotychczas zaproponowanych rozwiązań.

Tę część artykułu konkluduje dowód stwierdzenia, że 2γ jest najlepszym oszacowaniem górnym dla metody MST. Dla dowolnego $\varepsilon > 0$ istnieje pełny graf ważony (K_n, ω_n) taki, że:

$$\omega_n(S) > (2\gamma - \varepsilon)\omega_n(T), \tag{37}$$

gdzie: S jest skrótem dla pewnego minimalnego drzewa spinającego T.

Przykład: Niech $\gamma \geq 1$ i niech $G := (K_n, \omega_n)$ będzie pełnym grafem ważonym o n wierzchołkach $x_1, \dots, x_n$, gdzie:

$$\omega_n\big(\{x_i, x_j\}\big) := \begin{cases} 1, & i = 1,\ j > 1 \\ 2\gamma, & i, j > 1,\ i \neq j. \end{cases} \tag{38}$$

Tak zadany graf G spełnia $\gamma -$ nierówność wielokąta. Ponadto w grafie tym istnieje dokładnie jedno minimalne drzewo spinające T, którego krawędzie są postaci:

$$E(T) = \{(x_1, x_k) :\ k \in \{2, \dots, n\}\}. \tag{39}$$

Ponieważ każda krawędź T ma wagę 1, to zachodzi $\omega_n(T) = n - 1$. Przeszukiwanie T wgłąb (zaczynając od x_1) zwraca przejście:

$$W := (x_1, x_2, x_1, x_3, x_1, \dots, x_n, x_1). \tag{40}$$

Uzyskany w wyniku procedury (T, x_1) –skrót $S := (x_1, x_2, x_3, \dots, x_n, x_1)$ ma zatem wagę:

$$\omega_n(S) = 1 + (n-2) \cdot 2\gamma + 1 = 2 \cdot \big(1 + \gamma(n-2)\big). \tag{41}$$

Proste rachunki pozwalają stwierdzić, że:

$$\lim_{n\to\infty} \frac{\omega_n(S)}{\omega_n(T)} = \lim_{n\to\infty} \frac{2\cdot\big(1+\gamma(n-2)\big)}{(n-1)} = 2\gamma. \tag{42}$$

5. Algorytm Christofidesa dla grafów γ-metrycznych

Skonstruowawszy $2\gamma -$ optymalną metodę aproksymacji dla TSP, można się zastanowić, czy możliwe jest poprawienie tego wyniku w klasie grafów $\gamma -$ metrycznych? Okazuje się, że odpowiedź na to pytanie jest pozytywna, a rozważania na ten temat stanowić będą ostatnią część przedstawianego artykułu.

Centralna idea prezentowanej metody pochodzi z pracy Christofidesa, który zaprojektował $\frac{3}{2}-$ optymalną aproksymację dla TSP zdefiniowanego dla klasy grafów

metrycznych [67]. Celem obranym przez autorów jest rozszerzenie jego metody na klasę grafów $\gamma -$ metrycznych, co stanowić będzie przełomowy wynik w tej gałęzi optymalizacji kombinatorycznej.

Niezbędne w dalszej pracy okaże się przytoczenie kilku definicji [33].

Definicja 1. Niech (V, E) będzie grafem. Liczba krawędzi incydentnych z $x \in V$, tj.

$$\deg(x) := |\{\{x, y\} \in E : y \in V\}|, \tag{43}$$

nazywana jest stopniem wierzchołka $x \in V$.

Definicja 2. Niech (V, E) będzie grafem. Zbiór krawędzi $E_M \subset E$ nazywany jest skojarzeniem, jeżeli żadna para krawędzi E_M nie posiada wspólnego wierzchołka, tzn. dla dowolnych krawędzi $\{x_1, y_1\}, \{x_2, y_2\} \in E_M$ zachodzi $\{x_1, y_1\} \cap \{x_2, y_2\} = \emptyset$. Jeżeli każdy wierzchołek jest końcem którejś krawędzi z E_M, tj.:

$$\forall_{x \in V}\ \exists_{y \in V}\ \{x, y\} \in E_M, \tag{44}$$

to skojarzenie E_M określane jest jako doskonałe.

Każde skojarzenie może być postrzegane jako podgraf (V, E), w którym E_M jest zbiorem krawędzi, zaś V jest zbiorem wierzchołków. W grafach ważonych możliwym jest dalsze uszczegółowienie tej definicji w następujący sposób:

Definicja 3. Niech (G, ω) będzie grafem ważonym i niech $M \subset G$ będzie skojarzeniem doskonałym w tym grafie. M nazywane jest skojarzeniem doskonałym o minimalnym koszcie, jeżeli dla dowolnego innego skojarzenia doskonałego $M' \subset G$ zachodzi $\omega(M) \leq \omega(M')$.

W tym momencie zdefiniowane zostały wszystkie narzędzia niezbędne do przybliżenia konstrukcji rozwiązań przybliżonych dla TSP zaprojektowanej przez Christofidesa. Dla zadanego ważonego grafu pełnego $G := (K_n, \omega)$ należy wykonać pięć, poniżej omówionych kroków.

Krok 1. Znaleźć w G minimalne drzewo spinające T.

Krok 2. Niech $V_{odd}(T)$ będzie zbiorem wierzchołków o nieparzystych stopniach w T. Niech F będzie zdefiniowane jako indukowany podgraf G rozpięty na wierzchołkach $V_{odd}(T)$.

Krok 3. Znaleźć doskonałe skojarzenie o minimalnym koszcie w F (takie skojarzenie istnieje na mocy lematu o ściskaniu dłoni [33] – F ma bowiem parzystą liczbę wierzchołków). Można tego dokonać z wykorzystaniem algorytmu Edmondsa [68] lub jego rozszerzeń [69].

Krok 4. Znaleźć spacer Eulera W (czyli sekwencję wierzchołków, która przechodzi przez każdą krawędź grafu dokładnie raz) w multigrafie $\mathbb{G} = (V(T), \mathbb{E})$, w którym zbiór krawędzi zdefiniowany jest jako:

$$\mathbb{E} := \{(0, e) : e \in E(T)\} \cup \{(1, e) : e \in E(M)\}. \tag{45}$$

Krok 5. Wykonać procedurę skracania na W, uzyskując cykl Hamiltona C w wyjściowym grafie G.

Na pierwszy rzut oka może nie być oczywistym, dlaczego w multigrafie $\mathbb{G}$ skonstruowanym w **Kroku 4.** powyższego algorytmu musi istnieć spacer Eulera. Wątpliwości te rozwiewa poniższy lemat, który można znaleźć w [70], [33] lub [71].

Lemat. Jeżeli każdy wierzchołek spójnego (multi-)grafu $\mathbb{G}$ jest parzystego stopnia, to istnieje spacer Eulera w $\mathbb{G}$.

Dyskusję na temat złożoności czasowej algorytmu Christofidesa można zacząć od obserwacji, że w najgorszym scenariuszu w **Kroku 3.** może zajść konieczność znalezienia doskonałego skojarzenia o minimalnym koszcie w całym grafie G. Jest to zatem wąskie gardło całej tej procedury, należy więc zadbać o sprawne wykonanie tego kroku. Algorytm Gabowa [72-73] do znajdowania doskonałego skojarzenia o minimalnym koszcie w grafie ważonym działa w złożoności czasowej $\mathcal{O}(|V(G)|^3)$, aczkolwiek warto nadmienić, że istnieją bardziej efektywne rozwiązania tego problemu przy założeniu, że wszystkie wagi są liczbami naturalnymi. Nie zakładając tego w powyższych rozważaniach, pozostaje jedynie zaakceptować fakt, że złożoność algorytmu Gabowa determinuje jednocześnie złożoność całej powyższej procedury.

W klasie grafów metrycznych Christofides wykazał [67], że cykl Hamiltona C uzyskany w wyniku działania przedstawionego algorytmu spełnia nierówność $\omega(C) \leq \omega(T) + \omega(M)$. Istotnie, w grafach metrycznych prawdziwa jest nierówność:

$$\omega(M) \leq \frac{1}{2}\omega(C^*), \tag{46}$$

gdzie: C^* stanowi rozwiązanie optymalne dla TSP zdefiniowanego na (G, ω). Ostatetecznie, aplikując nierówność $\omega(T) \leq \omega(C^*)$, Christofides doszedł do wniosku, iż C jest $\frac{3}{2}$ – optymalnym przybliżeniem dla TSP, tzn.

$$\omega(C) \leq \frac{3}{2}\omega(C^*). \tag{47}$$

oryginalne rozumowanie Christofidesa (a co za tym idzie, uzyskana przez niego stała $\frac{3}{2}$) opiera się na nierówności trójkąta. Celem niniejszego artykułu jest przezwyciężenie tego ograniczenia i rozszerzenie tej techniki na klasę grafów $\gamma -$ wielokątnych.

Przedstawiane w pracy podejście powtarza pierwsze 3 kroki oryginalnej metody, tak jak przy rozumowaniu w przypadku grafów metrycznych. Na koniec **Kroku 3.** M jest nadal doskonałym skojarzeniem o minimalnym koszcie, ale oszacowanie (46) może nie być już prawdziwe, ponieważ rozpatrywany graf może nie spełniać nierówności trójkąta! Jaki więc odpowiednik tego oszacowania można podać w przypadku grafów $\gamma -$ metrycznych? Odpowiedzi udziela nam poniższy, uzyskany przez autorów wynik.

Twierdzenie 6. Niech C^* będzie optymalnym rozwiązaniem problemu komiwojażera w pełnym grafie $\gamma -$ metrycznym $G := (K_n, \omega)$. Niech F będzie dowolnym podgrafem indukowanym na G o parzystej liczbie wierzchołków. Jeżeli M jest doskonałym skojarzeniem o minimalnym koszcie na F, to zachodzi nierówność:

$$\omega(M) \leq \frac{\gamma}{2}\omega(C^*). \tag{48}$$

Dowód. Niech $C^* := (y_1, \dots, y_n, y_1)$, gdzie $y_1 \in F$ jest dowolnie wybranym wierzchołkiem. Niech $z_1 := y_1$ i $C_F := (z_1, z_2, \dots, z_m, z_1)$ będzie takim ponumerowaniem wierzchołków grafu F, że z_i poprzedza z_j w C^*, o ile $i < j$ (nieformalnie, wierzchołki F numerowane są w kolejności ich odwiedzania przez komiwojażera, zakładając że zaczyna on podróż w y_1). Następnie należy zaobserwować, że zachodzi:

$$\forall_{i=1,\dots,m}\ \omega(\{z_i, z_{i+1}\}) \leq \gamma \sum_{j=k_i}^{k_{i+1}-1} \omega\left(\{y_j, y_{j+1}\}\right), \tag{49}$$

gdzie $y_{k_i} = z_i$ i $y_{k_{i+1}} = z_{i+1}$ (oraz $z_{m+1} := z_1$). Stąd wywnioskować można, że

$$\omega(C_F) = \sum_{i=1}^{m} \omega(\{z_i, z_{i+1}\}) \leq \gamma \sum_{i=1}^{m} \sum_{j=k_i}^{k_{i+1}-1} \omega\big(\{y_j, y_{j+1}\}\big) = \gamma\omega(C^*). \quad (50)$$

Następnie dokonywany jest podział cyklu C_F na dwa skojarzenia doskonałe poprzez naprzemienny dobór krawędzi:

$$\begin{aligned} M_1 &:= (V(F), \{\{z_1, z_2\}, \{z_3, z_4\}, \dots, \{z_{m-1}, z_m\}\}), \\ M_2 &:= (V(F), \{\{z_2, z_3\}, \{z_4, z_5\}, \dots, \{z_m, z_1\}\}). \end{aligned} \quad (51)$$

Zaobserwować można, że $\omega(M_1) + \omega(M_2) = \omega(C_F)$, co z kolei oznacza, że $\omega(M_1) \leq \frac{\omega(C_F)}{2}$ lub $\omega(M_2) \leq \frac{\omega(C_F)}{2}$. Niemniej, jeżeli M jest doskonałym skojarzeniem o minimalnym koszcie, to zachodzi:

$$\omega(M) \leq \min\{\omega(M_1), \omega(M_2)\} \leq \frac{\omega(C_F)}{2} \leq \frac{\gamma}{2}\omega(C^*), \quad (51)$$

co kończy dowód.

Szczególnym przypadkiem w powyższym twierdzeniu jest sytuacja, gdy $\gamma = 1$, stanowiąca wynik uzyskany przez Christofidesa.

Kolejnym wyzwaniem jest przeprojektowanie **Kroku 4.** przez modyfikację algorytmu Hierholzera tak, by zagwarantować prawidłowe wykorzystanie krawędzi z uzyskanego skojarzenia M. Niech $\mathbb{G} := \big((V, \mathbb{E}), \omega\big)$ będzie multigrafem ważonym, w którym:

- $V(\mathbb{G}) := V(G) = V(T)$;
- $\mathbb{E}(G) := \{(0, e) : e \in E(T)\} \cup \{(1, e) : e \in E(M)\}$.

Zmodyfikowany algorytm Hierholzera można zapisać w poniższy sposób:

1. Należy wybrać wierzchołek początkowy $x_1 \in V(\mathbb{G})$, incydentny z krawędzią z $E(M)$. Niech $W := (y_1)$, gdzie $y_1 := x_1$, gdzie y_1 zdefiniowany będzie jako „*ostatnio odwiedzony wierzchołek*".
2. W jest iteracyjnie przedłużany zgodnie z następującymi regułami:
   a. Należy dokonać wyboru sąsiada y „*ostatnio odwiedzonego wierzchołka*" połączonego z nim niewykorzystaną dotychczas krawędzią z $E(M)$ – jeżeli to możliwe. W przeciwnym razie, należy wybrać krawędź z $E(\mathbb{G})$ – to jest zawsze możliwe, ponieważ wierzchołki $V(\mathbb{G})$ mają parzysty stopień. Warto podkreślić, że dzięki doborowi wierzchołka y_1 pewnym jest, że krawędź $\{y_1, y\}$ pochodzi z $E(M)$.
   b. y jest dodawany na koniec ciągu W. Następnie y jest oznaczany jako „*ostatnio odwiedzony wierzchołek*", zaś krawędź, której użyto, by dotrzeć do tego wierzchołka, jako wykorzystaną. Jeżeli istnieje niewykorzystana krawędź incydentna do y, należy powrócić do punktu a). Zaprojektowana w ten sposób procedura zatrzymać może się jedynie po powrocie do wierzchołka y_1.
3. Jeżeli w grafie istnieją jakieś niewykorzystane krawędzie, dokonać należy wyboru wierzchołka $x \in W$, który posiada sąsiada poza W. Ponawiając procedure opisaną w poprzednim kroku, uzyskuje się zamknięty spacer W_x. Ostatnie wystąpienie x w wyjściowym spacerze W należy zastąpić przez ciąg wierzchołków W_x.

Mając do dyspozycji spacer Eulera $W := (x_1, \dots, x_{k+1}), k = |E(M)| + |E(T)|$, możliwym staje się przejście do **Kroku 5.** prezentowanej metody, tj. do procedury

skracania. W porównaniu z klasycznym podejściem Christofidesa, należy zachować dużo większą ostrożność – naiwne skrócenie spaceru W mogłoby pozbyć się bowiem krawędzi ze skojarzenia M. W takim wypadku końcowym oszacowaniem na wagę uzyskiwanego cyklu Hamiltona C staje się:

$$\omega(C) = \sum_{k=1}^{n} d(y_k, y_{k+1}) \leq \gamma\omega(W) = \gamma\big(\omega(M) + \omega(T)\big) \leq \gamma\left(\omega(C^*) + \frac{\gamma}{2}\omega(C^*)\right). \tag{3}$$

Ponieważ γ zazwyczaj przekracza 1, należy podjąć wszelkie kroki, by uniknąć składnika $\frac{\gamma^2}{2}\omega(C^*)$ w ostatecznym oszacowaniu. W tym celu należy zmodyfikować procedurę skracania spaceru $W = (x_1, \ldots, x_{k+1})$ tak, by zachować w wynikowym cyklu Hamiltona $\hat{C} := (y_1, \ldots, y_n, y_{n+1})$ wszystkie krawędzie z doskonałego skojarzenia:

1. Niech $y_1 := x_1$ i $y_2 := x_2$. Z konstrukcji dokonanej w **Kroku 4.** spaceru Eulera W wynika, że $\{x_1, x_2\} \in E(M)$. Niech $i := 3$ i $j := 3$.
2. Dopóki $i \leq n$ wykonywane są następujące kroki:
   a. Jeżeli $x_j \notin V(M)$ wystąpił już w $\hat{C}$, zwiększyć j o 1.
   b. Jeżeli $x_j \notin V(M)$ nie pojawił się jeszcze w $\hat{C}$, niech $y_i := x_j$, następnie dokonać inkrementacji zarówno i, jak również j.
   c. Jeżeli $x_j \in V(M)$, $\{x_j, x_{j+1}\} \notin E(M)$ i $\{x_{j-1}, x_j\} \notin E(M)$, zwiększyć j.
   d. W przeciwnym wypadku, tj. $x_j \in V(M)$ i zachodzi $\{x_j, x_{j+1}\} \in E(M)$ lub $\{x_{j-1}, x_j\} \in E(M)$, przyjąć $y_i := x_j$, po czym dodać 1 do obydwu zmiennych i oraz j.

Warunki (a)–(d) nie tylko są rozłączne, ale również gwarantują, że każdy wierzchołek wystąpi w wynikowym cyklu Hamiltona $\hat{C}$ dokładnie jeden raz. Jeżeli $x_j \notin V(M)$ to jego włączenie do $\hat{C}$ ma miejsce przy pierwszym napotkaniu tego wierzchołka w W. Co więcej, wierzchołki $x_j \in V(M)$ również zostają dodane do $\hat{C}$ dokładnie raz. Wynika to z faktu, że $\deg_M(x_j) = 1$, a co za tym idzie warunek (d) nie może być spełniony więcej niż raz. Wynikowy cykl Hamiltona $\hat{C}$ zawiera więc wszystkie krawędzie z $E(M)$ lub, innymi słowy, żadna krawędź $E(M)$ nie została usunięta w trakcie procedury skracania.

Przedstawiony poniżej wynik jest kluczowy dla dowodu, że opisana autorska metoda stanowi $\frac{3\gamma}{2}-$ optymalną aproksymację dla TSP w grafach $\gamma-$ metrycznych. Sedno prezentowanego rozumowania leży w oszacowaniu łącznej wagi tych krawędzi $\hat{C}$, które nie należą do skojarzenia doskonałego M:

Twierdzenie 7. Niech $G := (K_n, \omega)$ będzie pełnym grafem $\gamma-$ wielokątnym, dla którego:

- T jest minimalnym drzewem spinającym (patrz **Krok 1.**);
- M jest doskonałym skojarzeniem o minimalnym koszcie na podgrafie indukowanym w G przez wierzchołki $V_{odd}(T)$ (patrz **Kroki 2.** i **3.**);
- W jest spacerem Eulera wygenerowanym przez zmodyfikowany algorytm Hierholzera (patrz **Krok 4**);

- $\hat{C}$ będzie cyklem Hamiltona skonstruowanym przez zaproponowaną wyżej procedurę skracania.

Wówczas zachodzi:

$$\sum_{e\in E(\hat{C})\setminus E(M)} \omega(e) \leq \gamma\omega(T). \tag{54}$$

Dowód: Jeżeli $e \in E(\hat{C}) \setminus E(M)$ wówczas $e = \{y_k, y_{k+1}\}$ dla pewnych $y_k, y_{k+1} \in V(G)$. Ponadto, istnieje taki ciąg wierzchołków

$$x_{k_i} := y_k,\ x_{k_i+1},\ \ldots,\ x_{k_j} := y_{k+1}, \tag{55}$$

że

$$\{x_{k_i}, x_{k_i+1}\}, \ldots, \{x_{k_j-1}, x_{k_j}\} \in E(M) \cup E(T), \tag{55}$$

ponieważ każdy z wierzchołków x_k pochodzi ze spaceru Eulera W. Ponadto zachodzi

$$\forall_{l=0,\ldots,k_j-k_i-1}\ \{x_{k_i+l}, x_{k_i+l+1}\} \notin E(M), \tag{56}$$

ponieważ w przeciwnym razie oznaczałoby to, że w procedurze skracania usunięto krawędź ze skojarzenia M, co nie jest możliwe. Zatem z $\gamma-$ nierówności wielokąta

$$\forall_{e\in E(\hat{C})\setminus E(M)}\ \omega(e) \leq \gamma \sum_{l=0}^{k_j-k_i-1} \omega\left(\{x_{k_i+l}, x_{k_i+l+1}\}\right). \tag{57}$$

Co więcej, dla różnych i, j ciągi wierzchołków (55) są rozłączne, co wynika z faktu, że W przechodzi przez każdą krawędź $\mathbb{G}$ dokładnie raz. W konsekwencji udaje się uzyskać oszacowanie

$$\sum_{\{y_k,y_{k+1}\}\in E(\hat{C})\setminus E(M)} \omega(\{y_k, y_{k+1}\}) \leq \gamma \sum_{\{k_i:\, y_k=x_{k_i}\}} \sum_{l=0}^{k_j-k_i-1} \omega\left(\{x_{k_i+l}, x_{k_i+l+1}\}\right) \leq \gamma E(T), \tag{58}$$

które kończy dowód. □

Mając do dyspozycji Twierdzenie 7, dalsze rozumowanie jest stosunkowo proste: jeżeli $E(M) \subset E(\hat{C})$ to wówczas

$$\omega(\hat{C}) = \sum_{e\in E(\hat{C})\setminus E(M)} \omega(e) + \sum_{e\in E(M)} \omega(e) \leq \gamma\omega(T) + \frac{\gamma}{2}\omega(C^*) \leq \frac{3\gamma}{2}\omega(C^*). \tag{59}$$

Oszacowanie to pozwala sformułować końcowy i najważniejszy rezultat przedstawiany w niniejszej pracy:

Twierdzenie 8. Przy założeniach Twierdzenia 7 istnieje $\frac{3\gamma}{2}$ – optymalne rozwiązanie przybliżone dla problemu komiwojażera określonego na grafie (K_n, ω), które może zostać skonstruowane w czasie $\mathcal{O}(n^3)$.

6. Wnioski/Podsumowanie

W pracy opracowano dwa algorytmy generujące przybliżone rozwiązania dla semimetrycznej wersji problemu komiwojażera. Pierwszy z algorytmów, oparty w głównej mierze na konstrukcji minimalnego drzewa spinającego, pozwala na uzyskanie $2\gamma-$ optymalnego rozwiązania. Algorytm drugi stanowi wariację na temat algorytmu Christofidesa i wykorzystuje zarówno minimalne drzewo spinające, jak i doskonałe

skojarzenie o minimalnym koszcie. Dzięki tej poprawce uzyskany cykl Hamiltona okazuje się być $\frac{3\gamma}{2}$ – optymalnym rozwiązaniem. Oba wyniki stanowią nowatorskie ujęcie tematu, wykorzystując narzędzia przestrzeni semimetrycznych do zagadnienia problemu komiwojażera.

Naturalnym wydaje się pytanie, jak zaproponowana w tej pracy metoda prezentuje się na tle wyników o zbliżonym charakterze użytkowym. Przedstawione w pracy wyprowadzenie, z uwagi na swoją obszerność, ma jedynie charakter teoretyczny – praktyczne porównanie jakości rozwiązań uzyskiwanych przez poszczególne metody aproksymacyjne można znaleźć w pracy [74].

Uwagi ogólne i podziękowania

Autorzy pragną wyrazić wdzięczność pracownikom Fundacji Tygiel za cierpliwość i wsparcie w procesie składania stosownej dokumentacji. Przede wszystkim jednak składają wyrazy szacunku dla anonimowego recenzenta, którego cenne wskazówki i krytyczne uwagi pozytywnie wpłynęły na jakość wypracowanego artykułu.

Filip Turoboś jest finansowany z 2% subsydium dla szkolnictwa wyższego dla uczelni biorących udział w konkursie IDUB (‘Inicjatywa Doskonałości – Uczelnia Badawcza’), jak również jest finansowany w ramach grantu „FU2N – Fundusz Udoskonalania Umiejętności Młodych Naukowców” wspierającego doskonałość naukową pracowników Politechniki Łódzkiej, nr grantu: 4/2023.

Literatura

1. Blumenthal L., *Theory and Applications of Distance Geometry*, Oxford University Press, Oxford 1953.
2. Indyk P., *Sublinear time algorithms for metric space problems*, Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing – STOC, 99, 1999.
3. Linial N., *Finite metric spaces: combinatorics, geometry and algorithms*, Proceedings of the Eighteenth Annual Symposium on Computational Geometry – SCG, 02, 2002.
4. Brattka V., *Recursive quasi-metric spaces*, Theoretical Computer Science, 305(1-3), 2003, s. 17-42.
5. Gustavsson J., *Metrization of quasi-metric spaces*, Mathematica Scandinavica, 35, 1974, s. 56-60.
6. Smyth M.B., *Topology*, [w:] Abramsky S., Maibaum T.S.E. (red.), *Handbook of Logic in Computer Science (Vol. 1): Background: Mathematical Structures*, Oxford University Press, Inc., New York 1993, s. 641-761.
7. Howes N.R., *Modern Analysis and Topology*, Universitext, Springer, New York 1995.
8. Goetschel R., Voxman W., *A pseudometric for fuzzy sets and certain related results*, Journal of Mathematical Analysis and Applications, 81 (2), 1981, s. 507-523.
9. Lowen R.: *On the relation between fuzzy equalities pseudometrics and uniformities*, Quaestiones Mathematicae, 7 (4), 1984, s. 407-419.
10. Bandelt H.J., Crama Y., Spieksma F.C.R., *Approximation algorithms for multi-dimensional assignment problems with decomposable costs*, Discrete Applied Mathematics. The Journal of Combinatorial Algorithms, Informatics and Computational Sciences, 49 (1-3), 1994, s. 25-50.
11. Chen L.H., Hsieh S.Y., Hung, L.J., Klasing R., *Approximation algorithms for the p-hub center routing problem in parameterized metric graphs*, Theoretical Computer Science, 806, 2020, s. 271-280.

12. Bessenyei M., Pàles Zs., *A contraction principle in semimetric spaces*, Journal of Nonlinear and Convex Analysis, 18 (3), 2017, s. 515-524.
13. Ceder J.G., *Some generalizations of metric spaces*, Pacific Journal of Mathematics, 11 (1), 1961, s. 105-125.
14. Chrząszcz K., Jachymski J., Turoboś F., *On characterizations and topology of regular semimetric spaces*, Publicationes Mathematicae Debrecen, 93, 2018, s. 87-105.
15. Chrząszcz K., Jachymski J., Turoboś F., *Two refinements of Frink's metrization theorem and fixed point results for Lipschitzian mappings on quasimetric spaces*, Aequationes Mathematicae, Springer International Publishing, 2018, s. 1-21.
16. Gottlieb L.A., Kontorovich A., Nisnevitch P., *Nearly optimal classification for semimetrics*, Journal of Machine Learning Research, 18 (37), 2017, s. 1-22.
17. Jachymski J., Turoboś F., *On functions preserving regular semimetrics and quasimetrics satisfying the relaxed polygonal inequality*, Revista de la Real Academia de Ciencias Exactas, Fısicas y Naturales. Serie A. Matematicas. RACSAM, 114, 159, 2020, s. 1-11.
18. Pareek C., *Moore Spaces, Semi-Metric Spaces and Continuous Mappings Connected with Them*, Canadian Journal of Mathematics. Journal Canadien de Mathematique, 24 (6), 1972, s. 1033-1042.
19. Wilson W.A., *On semi-metric spaces*, American Journal of Mathematics, 53, 1931, s. 361-373.
20. Cook H., *A characterization of continuously semimetrizable spaces*, [w:] Reed, G.M. (red.), *Set-Theoretic Topology*, Academic Press, Inc., New York, San Francisco, London 1977, s. 91-95.
21. Lutzer D.J., *Semimetrizable and stratifiable spaces*, General Topology and its Applications, 1 (1), 1971, s. 43-48.
22. Kunen K., Vaughan J.E., *Handbook of Set-Theoretic Topology*, North-Holland Publishing Co., Amsterdam 1984.
23. An T.V., Tuyen L.Q., Dung N.V., *Stone-type theorem on b-metric spaces and applications*, Topology and its Applications ,185/186, 2015, s. 50-64.
24. Andreae T., Bandelt H.J., *Performance guarantees for approximation algorithms depending on parametrized triangle inequalities*, SIAM Journal on Discrete Mathematics, 8 (1), 1995, s. 1-16.
25. Bakthin I.A., *The contraction mapping principle in almost metric spaces*, Functional Analysis, Ul'yanovsk. Gos. Ped. Inst., 30, 1989, s. 26-37.
26. Bourbaki N., *Elements of Mathematics. General Topology Part 2*, Hermann, Paris; Addison-Wesley Publishing Co., 1966.
27. Dung, N.V., Hang, V.T.L., *On relaxations of contraction constants and Caristi's theorem in b-metric spaces*, Journal of Fixed Point Theory and Applications, 18, 2016, s. 267-284.
28. Fagin R., Kumar R., Sivakuma D., *Comparing top k lists*, SIAM Journal on Discrete Mathematics, 17, 2003, s. 134-160.
29. Paluszyński M., Stempak K., *On quasi-metric and metric spaces*, Proceedings of the American Mathematical Society, 137, 2009, s. 4307-4312.
30. Schroeder V., *Quasi-metric and metric spaces*, Conformal Geometry and Dynamics. An Electronic Journal of the American Mathematical Society, 10, 2006, s. 355-360.
31. Suzuki T., *Basic inequality on b −metric space and its applications*, Journal of Inequalities and Applications, 256, 2017, s. 1-11
32. Xia Q., *The Geodesic Problem in Quasimetric Spaces*, Journal of Geometric Analysis, 19, 2009, s. 452-479
33. Diestel R., *Graph theory*, Springer-Verlag, New York 2000.
34. Fletcher P., Hoyle H., Patty C.W., *Foundations of Discrete Mathematics*, PWS-Kent Pub. Co., Boston 1991.
35. Bondy A., Murty M.R., *Graph Theory*, Graduate Texts in Mathematics 244, Springer-Verlag, London 2008.

36. Bläser M., *Metric TSP*, [w:] Kao, M.Y. (red.), *Encyclopedia of Algorithms*, Springer, Boston, MA, 2008.
37. Gross J.L., Yellen, J., Anderson, M.: *Graph Theory and Its Applications, Third Edition*, Chapman and Hall/CRC, 2018.
38. Dantzig, G.B., Fulkerson D.R., Johnson S.M., *On a Linear-Programming, Combinatorial Approach to the Traveling-Salesman Problem*, Operations Research, 7(1), 1959, s. 58-66.
39. Laporte G., *The traveling salesman problem: An overview of exact and approximate algorithms*, European Journal of Operational Research, 59(2), 1992, s. 231-247.
40. Matai R., Singh S., Mittal M.L., *Traveling Salesman Problem: an Overview of Applications, Formulations, and Solution Approaches, Traveling Salesman Problem, Theory and Applications*, Donald Davendra, IntechOpen, 2010.
41. Miller C.E., Tucker A.W., Zemlin R.A., *Integer Programming Formulation of Traveling Salesman Problems*, Journal of the ACM, 7 (4), 1960, s. 326-329
42. Applegate D.L., Bixby R.E., Chvatal V., Cook W.J., *The traveling salesman problem: a computational study*, Princeton University Press, 2007.
43. Cook W., *In Pursuit of the Traveling Salesman: Mathematics at the Limits of Computation*, Princeton; Oxford: Princeton University Press, 2012.
44. Fleischmann B., *A new class of cutting planes for the symmetric travelling salesman problem*, Mathematical Programming, 40, 1988, s. 225-246.
45. Fonlupt J., Naddef D., *The traveling salesman problem in graphs with some excluded minors*. Mathematical Programming, 53, 1992, s. 147-172.
46. Gutin G., Yeo A., *TSP tour domination and Hamilton cycle decompositions of regular digraphs*, Operations Research Letters, 28 (3), 2001, s. 107-111.
47. Reinelt G., *The Traveling Salesman – Computational Solutions for TSP Applications*, Lecture Notes in Computer Science 840, Springer-Verlag Berlin Heidelberg, 1994.
48. Snyder L.V., Shen Z.J.M., *Traveling Salesman Problem*, [w:] *Fundamentals of Supply Chain Theory*, John Wiley & Sons, 2019, s. 403-461.
49. Van Bevern R., Slugina V.A., *A historical note on the* $\frac{3}{2}$*-approximation algorithm for the metric traveling salesman problem*, History of Mathematics, 53, 2020, s. 118-127.
50. Ausiello G., Marchetti-Spaccamela A., Crescenzi P., Gambosi G., Protasi M., Kann V., *Complexity and Approximation*, Springer, Berlin, Heidelberg, 1999.
51. Anddreasson N., Evgrafov A., Patriksson M., *An Introduction to Continuous Optimization: Foundations and Fundamental Algorithms*, Revised Third edition, Dover Publications, 2020.
52. Weise T., *Global Optimization Algorithm: Theory and Application*, 2008, https://archive.org/details/Thomas_Weise__Global_Optimization_Algorithms_Theory_and_Application [data dostępu: 3.03.2023].
53. Crescenzi P., Kann V., Halldórsson M., Karpinski M., Woeginger G., *A compendium of NP optimization problems*, 2005, http://www.csc.kth.se/ viggo/problemlist/ [data dostępu: 5.03.2023].
54. Goldreich O., *Computational Complexity: A Conceptual Perspective*, Cambridge University Press, New York 2008.
55. Goldreich O., *P, NP, and NP-Completeness: The Basics of Computational Complexity*, Cambridge University Press, New York 2010.
56. Arora S., Barak B., *Computational complexity: a modern approach*, Cambridge University Press, 2009.
57. Du D.Z., Ko K.I., *Theory of computational complexity*, John Wiley & Sons, 58, 2011.
58. Papadimitriou C.H., Steiglitz K., *Combinatorial optimization: algorithms and complexity*, Courier Corporation, 1998.

59. Cormen T.H., Leiserson C.E., Rivest R.L., Stein C., *Introduction to Algorithms (3rd ed.)*, London, England: MIT Press, 2009.
60. Floyd R.W., *Algorithm 97: Shortest path*, Communications of the ACM, 5 (6), 1962, s. 345.
61. Narsingh D., *Graph theory with applications to engineering and computer science*, New Delhi, Prentice-Hall, 92, 1974.
62. Nešetřil J., Milková E., Nešetřilová H., *Otakar Borůvka on minimum spanning tree problem Translation of both the 1926 papers, comments, history*, Discrete Mathematics, 233(1-3), 2001, 3-36.
63. Andreae T., *On the traveling salesman problem restricted to inputs satisfying a relaxed triangle inequality*, Networks, 38(2), 2001, s. 59-67.
64. Bender M.A., Chekuri C., *Performance guarantees for the TSP with a parameterized triangle inequality*, Information Processing Letters, 73(1-2), 2000, s. 17-21.
65. Böckenhauer H.J., Hromkovič J., Klasing R., Seibert S., Unger W., *Towards the notion of stability of approximation for hard optimization tasks and the traveling salesman problem*, Theoretical Computer Science, 285, 2002, s. 3-24.
66. Krug S., *Analysis of a near-metric TSP approximation algorithm*, RAIRO Theoretical Informatics and Applications, 47 (3), 2013, s. 293-314.
67. Christofides N., *Worst-case analysis of a new heuristic for the traveling salesman problem*, Technical Report 388, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh 1976.
68. Edmonds J., *Paths, trees, and flowers*, Canadian Journal of Mathematics. Journal Canadien de Mathematiques 17, 1965, s. 449-467.
69. Cook W., Rohe A., *Computing minimum-weight perfect matchings*, INFORMS Journal on Computing, 11(2), 1999, s. 138-148.
70. Balakrishnan V.K., *Schaum's Outline of Theory and Problems of Graph Theory*, Schaum's Outline Series, McGrav-Hill, 1997.
71. Levin O., *Discrete Mathematics: An Open Introduction*, CreateSpace Independent Publishing Platform, 2015.
72. Gabow H.N., *Implementation of Algorithms for Maximum Matching on Nonbipartite Graphs*, Ph.D. Thesis, Stanford University, 1974.
73. Gabow H.N., *Data Structures for Weighted Matching and Nearest Common Ancestors with Linking*, [w:] Johnson D.S. (red.), *Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms*, Association for Computing Machinery, New York 1990, s. 434-443.
74. Krukowski M., Turoboś F., *Comparison of approximation algorithms for the travelling salesperson problem on semimetric graphs*, arXiv, 2108.13070 [math.CO].

Przybliżone rozwiązania problemu komiwojażera dla grafów semimetrycznych

Streszczenie

Celem postawionym w tym artykule jest rozszerzenie stosowalności metody aproksymacyjnej Christofidesa na klasę wszystkich pełnych grafów ważonych o dodatnich wagach. W tym celu wprowadzona zostaje klasa grafów semimetrycznych i wykazujemy, że każdy graf skończony jest w istocie $\gamma -$ metryczny. Następnie przytaczamy pojęcia z teorii grafów i teorii optymalizacji, które są niezbędne do analizowania problemu komiwojażera. Przedstawiona zostaje heurystyka minimalnego drzewa spinającego, której efektywność badamy w kontekście grafów spełniających $\gamma -$nierówność wielokąta. Ostatnia część artykułu poświęcona jest kluczowemu wynikowi, dostosowującego technikę Christofidesa do grafów $\gamma -$ metrycznych i pozwalającemu na uzyskanie $\frac{3\gamma}{2} -$ optymalnego przybliżenia dla problemu komiwojażera.

Słowa kluczowe: przestrzenie b-metryczne, przestrzenie semimetryczne, problem komiwojażera, minimalne drzewo spinaję, optymalizacja kombinatoryczna

Approximate solutions to the Traveling Salesperson Problem on semimetric graphs

Abstract

The primary aim of the paper is to extend the applicability of Christofides approximation technique to the scope of all complete (finite) weighted graphs with positive weights. To this end, we acquaint the Reader with the class of semimetric graphs and prove that every finite graph admits a γ-polygon structure. Subsequently, we revisit those concepts from graph and optimization theory, which are necessary for the investigation of the Travelling Salesperson Problem (TSP for short). We move on to the minimal spanning tree heuristic, studying its effectiveness through the lens of γ-polygon graphs. The final part of the paper adjusts the technique of Christofides and culminates in obtaining $\frac{3\gamma}{2}$ −approximate solution to the TSP.

Keywords: b-metric spaces, semimetric spaces, MST, TSP, combinatorial optimization

Halina Marczak[1]

Wpływ elektromobilności na jakość powietrza

1. Wprowadzenie

Termin elektromobilność oznacza ogół zagadnień związanych z użytkowaniem pojazdów elektrycznych. Zgodnie z ustawą o elektromobilności i paliwach alternatywnych [1] pojazd w pełni elektryczny to pojazd samochodowy napędzany energią elektryczną magazynowaną w baterii trakcyjnej przez podłączenie pojazdu do zewnętrznego źródła zasilania. Pojazd częściowo elektryczny to pojazd samochodowy hybrydowy (typu *plug-in*), posiadający napęd spalinowo-elektryczny, w którym energia elektryczna jest gromadzona przez dołączenie do zewnętrznego źródła prądu. Według danych Polskiego Stowarzyszenia Paliw Alternatywnych (PSPA) [2] liczba elektrycznych samochodów osobowych w Polsce wynosi 61570 sztuk (stan na koniec grudnia 2022 r.), przy czym w pełni elektryczne samochody osobowe stanowią 51% ogółu tych samochodów. Liczba elektrycznych pojazdów dostawczych i ciężarowych w Polsce to 3125 sztuk, a autobusów elektrycznych 821 sztuk (stan na koniec grudnia 2022 roku). Rosnący udział bezemisyjnych pojazdów w ogólnej liczbie pojazdów wzmacnia przekonanie, że będzie stopniowo malał wpływ transportu na zanieczyszczenie powietrza. Obecnie w Polsce transport odpowiada za prawie jedną czwartą (22,3%) emisji dwutlenku węgla [3]. Ponadto udział emisji tlenków azotu i tlenku węgla z transportu w ogólnej ich emisji w Polsce kształtuje się na poziomie 35,5% dla tlenków azotu i 13,2% dla tlenku węgla [3]. Brak emisji do powietrza z pojazdów elektrycznych w miejscach ich użytkowania to ważny argument uzasadniający celowość działań promujących rozwój rynku tych pojazdów.

Kompleksowa ocena wpływu na środowisko pojazdu elektrycznego jest wypadkową ocen wpływu na poszczególnych etapach cyklu życia (istnienia) pojazdu [4]. Cztery zasadnicze etapy cyklu życia pojazdu to kolejno: wydobycie surowców potrzebnych do produkcji zespołów i części składowych pojazdu, produkcja materiałów konstrukcyjnych, części składowych pojazdu i ich transport do miejsc montażu, eksploatacja pojazdu, zagospodarowanie pojazdów wycofanych z eksploatacji.

Etap eksploatacji obejmuje użytkowanie pojazdu, zabiegi obsługowe, w tym ładowanie baterii, naprawy i konserwacje oraz przerwy w korzystaniu z pojazdu. Emisja zanieczyszczeń na tym etapie w literaturze [5] określana jest mianem emisji eksploatacyjnej. Na etapie eksploatacji pojazdu ważnym kryterium oceny wpływu pojazdu na środowisko jest emisja do powietrza w procesie wytwarzania i dystrybucji energii elektrycznej potrzebnej do ładowania baterii pojazdu elektrycznego. Tematem badawczym w pracy jest określenie wielkości emisji dwutlenku węgla z produkcji energii elektrycznej na bazie paliw kopalnych (węgiel, gaz ziemny). Wybranie paliw kopalnych do analizy ma uzasadnienie w tym, że aktualnie w Polsce węgiel ma największy udział w strukturze paliw wykorzystywanych w produkcji energii elektrycznej. Wyniki badań można zastosować do porównania wielkości emisji dwutlenku węgla z produkcji energii elektrycznej w oparciu o inne niż węgiel nośniki energii. Główną myślą przewodnią pracy jest stwierdzenie, że

[1] h.marczak@pollub.pl Katedra Zrównoważonego Transportu i Źródeł Napędu, Wydział Mechaniczny, Politechnika Lubelska.

emisyjność pojazdów elektrycznych na etapie ich eksploatacji będzie tym mniejsza, im bardziej ekologiczne źródła energii będą wykorzystywane do wytwarzania energii elektrycznej. Poprawę jakości powietrza w Polsce można również uzyskać poprzez zwiększenie sprawności wytwarzania energii elektrycznej w energetycznych blokach węglowych i gazowych.

2. Metoda badań i założenia

W pracy wykorzystano metodę analizy danych, które są niezbędne do realizacji celu badań. Przedmiotem rozważań objęto dane techniczne pojazdów elektrycznych, dane o strukturze paliw wykorzystywanych do produkcji energii elektrycznej w elektrowniach cieplnych zawodowych w Polsce oraz wartości wskaźników emisji dwutlenku węgla z produkcji energii elektrycznej. Źródłami informacji są przedmiotowe publikacje naukowe, publikacje ośrodków badawczych, raporty przygotowane przez producentów pojazdów elektrycznych, opracowania i raporty organizacji i stowarzyszeń branży elektromobilności oraz opracowania urzędów statystycznych.

Opierając się na zebranych danych określono wielkość emisji dwutlenku węgla towarzyszącej wytwarzaniu i dystrybucji energii elektrycznej w ilości potrzebnej do ładowania baterii pojazdu elektrycznego. W rozważaniach uwzględniono paliwa kopalne, jako surowiec energetyczny do przetwarzania w elektrowni. W obliczeniach wielkości emisji dwutlenku węgla wzięto pod uwagę jedynie emisję na etapie produkcji i dystrybucji energii elektrycznej. W badaniu pominięto emisję dwutlenku węgla na etapie „od pozyskania surowca do bramy elektrowni" (wydobycie surowca (węgla, gazu ziemnego), przeróbka, transport surowca do elektrowni).

Z przedstawionych rozważań wynika, że pojazdy elektryczne na etapie ładowania baterii mają udział w emisji do powietrza zanieczyszczeń. Wielkość tej emisji zależy od rodzaju źródła energii wykorzystywanego w produkcji energii elektrycznej [6, 7]. Im większy jest udział w wytwarzaniu energii elektrycznej czystych źródeł energii (energia odnawialna, energia jądrowa), tym mniejszy jest wpływ na zanieczyszczenie powietrza pojazdu elektrycznego na etapie ładowania.

Zgromadzone dane wykorzystano ponadto do wyznaczenia wielkości emisji dwutlenku węgla z pojazdu samochodowego wyposażonego w silnik spalinowy benzynowy. Uwzględniono emisję na etapie „od źródła do zbiornika" (od pozyskania ropy naftowej do dostawy paliwa na stację paliw i dalej do zbiornika w samochodzie) oraz na etapie użytkowania pojazdu.

3. Wielkość emisji dwutlenku węgla na etapie eksploatacji pojazdu elektrycznego

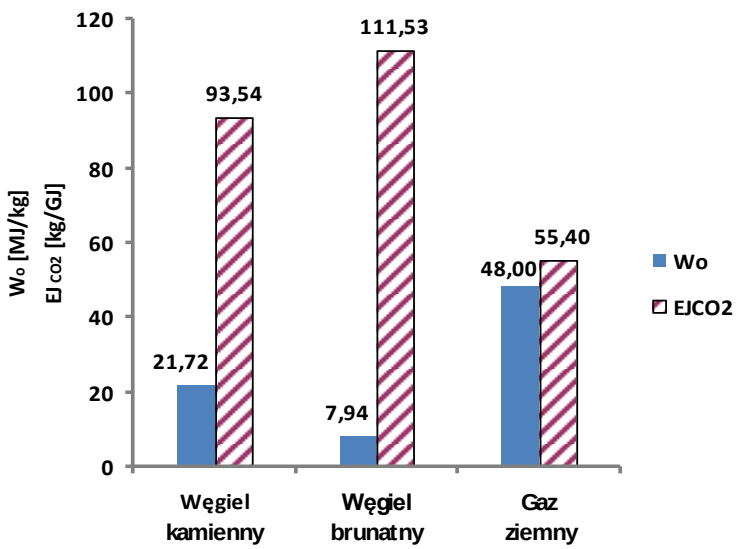
Energia elektryczna wytwarzana jest w oparciu o zasoby środowiska – surowce energetyczne nieodnawialne oraz źródła energii odnawialnej. W Polsce głównym paliwem dla energetyki jest węgiel. Energia elektryczna wytworzona z węgla kamiennego miała ponad 50,1% udziału w całkowitej produkcji energii elektrycznej w Polsce w roku 2022, podczas gdy z węgla brunatnego ponad 26,8% (tab. 1).

Tabela 1. Produkcja energii elektrycznej w Polsce w roku 2022

Wyszczególnienie	Produkcja energii elektrycznej [GWh]
Elektrownie zawodowe konwencjonalne w tym: węgiel kamienny węgiel brunatny gaz ziemny	144740 87761 46978 10002
Elektrownie zawodowe wodne	2815
Instalacje odnawialnych źródeł energii w tym: elektrownie wiatrowe pozostałe instalacje	27602 18305 9297
Ogółem	175157

Źródło: opracowanie własne na podstawie [8].

Na rysunku 1 przedstawiono wartości opałowe węgla i gazu ziemnego wykorzystywanych w polskich elektrowniach i elektrociepłowniach zawodowych oraz jednostkowe wskaźniki emisji dwutlenku węgla ze spalania węgla i gazu ziemnego. Wskaźniki emisji wyrażone są w kilogramach dwutlenku węgla (kg CO_2) na gigadżul (GJ) ciepła pozyskanego odpowiednio z węgla i z gazu ziemnego. Wartości opałowe i wartości wskaźników emisji zaczerpnięto z literatury [9].



Rysunek 1. Wartość opałowa paliw (W_o) i jednostkowy wskaźnik emisji dwutlenku węgla (EJ_{CO2}) ze spalania paliw w polskich elektrowniach i elektrociepłowniach. Źródło: opracowanie własne na podstawie [9]

Sprawność wytwarzania energii elektrycznej w elektrowniach zawodowych cieplnych w Polsce wyniosła w roku 2021 42,5% [10]. W roku 2019 wartość tej sprawności kształtowała się na poziomie 38,1% [11]. Wyraża ona iloraz ilości wytworzonej energii elektrycznej i ilości energii chemicznej paliwa zamienionej na ciepło pary wodnej. Jednostkowy wskaźnik emisji dwutlenku węgla dla procesu wytwarzania energii elektrycznej w elektrowniach cieplnych można wyznaczyć ze wzoru:

$$JWE_{CO2} = \frac{EJ_{CO2}}{\eta} \tag{1}$$

gdzie: JWE_{CO2} – jednostkowy wskaźnik emisji dwutlenku węgla (CO_2) dla procesu wytwarzania energii elektrycznej, wyrażony w kg CO_2/MWh, EJ_{CO2} – jednostkowy wskaźnik emisji dwutlenku węgla ze spalania paliwa, wyrażony w kg CO_2/MWh, η – sprawność wytwarzania energii elektrycznej.

Obliczony według wzoru (1) w oparciu o dane: EJ_{CO2} = 93,54 kg/GJ = 336,74 kg/MWh, η = 42,5% wskaźnik JWE_{CO2} dla produkcji energii elektrycznej z węgla kamiennego w elektrowniach cieplnych wynosi 792,33 kg CO_2/MWh. Wartość wskaźnika JWE_{CO2} dla produkcji energii elektrycznej z węgla brunatnego wynosi 944,72 kg CO_2/MWh (wynik otrzymano dla danych: EJ_{CO2} = 111,53 kg/GJ = 401,51 kg/MWh, η = 42,5%). Wartość wskaźnika JWE_{CO2} dla produkcji energii elektrycznej z gazu ziemnego wynosi 469,27 kg CO_2/MWh (dla EJ_{CO2} = 55,48 kg/GJ = 199,44 kg/MWh, η = 42,5%). W tabeli 2 przedstawiono wpływ sprawności wytwarzania energii elektrycznej na jednostkowy wskaźnik emisji dwutlenku węgla.

Odnosząc się tylko do produkcji energii elektrycznej z węgla (134739 GWh, tab. 1) można stwierdzić, że w Polsce w roku 2022 węgiel kamienny odpowiadał za 65,13% tej produkcji, a węgiel brunatny za 34,87%. Na tej podstawie można wyznaczyć, że średnia ważona wartość wskaźnika JWE_{CO2} dla procesu wytwarzania energii elektrycznej z węgla w Polsce wynosi 845,47 kg CO_2/MWh (dla sprawności η = 42,5%).

Tabela 2. Wpływ sprawności procesu wytwarzania energii elektrycznej na jednostkowy wskaźnik emisji dwutlenku węgla (JWE_{CO2}) dla tego procesu

Paliwo	Sprawność [%]	Wskaźnik JWE_{CO2} [kg CO_2/MWh]
Węgiel kamienny	42,50	792,33
	44,00	765,33
	46,00	735,02
Węgiel brunatny	42,50	944,72
	44,00	912,52
	46,00	872,84
Gaz ziemny	42,50	469,27
	44,00	453,27
	46,00	433,56

Źródło: opracowanie własne.

Z analizy struktury paliw i innych nośników energii pierwotnej wykorzystywanych do wytworzenia energii elektrycznej w Polsce w roku 2022 (tab. 1) wynika, że udział węgla w całkowitej produkcji energii elektrycznej wynosi 76,9%. Mnożąc ten udział przez podaną wyżej wartość 845,47 kg CO_2/MWh można określić, że na każdą megawatogodzinę energii elektrycznej wytworzonej w Polsce przypada emisja 650,17 kg CO_2 spowodowana zużyciem węgla. Analogicznie obliczona wartość emisji dwutlenku węgla ze spalania gazu ziemnego wynosi 26,75 kg CO_2/MWh. Całkowita emisja dwutlenku węgla, jako suma emisji spowodowanej zużyciem węgla i emisji wynikającej ze zużycia paliwa gazowego w elektrowniach cieplnych osiąga wartość 676,92 kg dwutlenku węgla na megawatogodzinę wytworzonej energii elektrycznej.

Zgodnie z raportem Polskiego Towarzystwa Przesyłu i Rozdziału Energii Elektrycznej (PTPiREE) [12] przesyłowi energii elektrycznej przez sieć elektroenergetyczną towarzyszą straty energii. W roku 2021 straty w przesyle energii w Polsce wyniosły 5,5%. Wynika z tego, że pobranie z sieci elektroenergetycznej energii elektrycznej o wartości 1 MWh będzie możliwe pod warunkiem wyprodukowania i oddania do sieci 1,055 MWh energii. W efekcie uwzględnienia strat w przesyle energii zmieni się wartość jednostkowego wskaźnika emisji dwutlenku węgla z 676,92 kg/MWh na 714,15 kg/MWh. Wartość ta informuje, że pobraniu z sieci jednej megawatogodziny energii elektrycznej towarzyszy emisja 714,15 kg dwutlenku węgla spowodowana zużyciem paliw kopalnych (węgiel, gaz ziemny) na wytworzenie energii elektrycznej oraz na pokrycie strat w przesyle energii.

Zasięg i zużycie energii to typowe parametry techniczne samochodów elektrycznych. Wartości tych parametrów dla samochodów elektrycznych wprowadzanych na rynek Unii Europejskiej (UE) wyznaczane są w warunkach laboratoryjnych, obecnie według obowiązującego testu WLTP (ang. *Worldwide Harmonized Light Vehicles Test Procedure*) [13]. Wartości zasięgu pojazdu na jednym ładowaniu wyznaczone w teście WLTP są wyższe od ich wartości osiąganych w warunkach rzeczywistych użytkowania pojazdu. Z kolei wartości zużycia energii w warunkach rzeczywistych użytkowania pojazdu są wyższe od wyznaczonych w teście WLTP. Zmienna sytuacja komunikacyjna na drogach, zmienne warunki pogodowe, a zwłaszcza temperatura, zmiana stylu jazdy kierowców, w tym prędkości jazdy, korzystanie z dodatkowych funkcji pojazdu powoduje wzrost zużycia energii przez pojazd oraz obniżenie zasięgu pojazdu.

Obecnie na rynku dostępne są samochody osobowe elektryczne o zasięgu ponad 600 km. W tabeli 3 przedstawiono wartości zasięgu, pojemności baterii i średniego zużycia energii elektrycznej dla wybranych modeli samochodów. Są to wartości deklarowane przez producentów pojazdów, otrzymane w teście WLTP, który odzwierciedla warunki drogowe miejskie i poza miastem oraz średnią prędkość jazdy podczas testu 46,5 km/h.

Na podstawie analizy wyników (według testu WLTP) przedstawionych w tabeli 3 można stwierdzić, że średnie zużycie energii elektrycznej przyjmuje wartości od 11,9 kWh/100 km do 17,6 kWh/100 km. Wartości te w warunkach rzeczywistych będą bez wątpienia wyższe. Biorąc pod uwagę otrzymane wyniki obliczeń jednostkowego wskaźnika emisji dwutlenku węgla JWE_{CO2} = 714,15 g CO_2/kWh można stwierdzić, że średniemu zużyciu energii elektrycznej przez samochód elektryczny o wartości (11,9-17,6) kWh/100 km odpowiada emisja dwutlenku węgla na poziomie (8498,40-12569,00) g/100 km. Wynika to z faktu, że zużywaną energię elektryczną podczas ładowania pojazdu należy wcześniej wytworzyć.

Tabela 3. Wybrane parametry techniczne przykładowych samochodów osobowych elektrycznych

Marka, model pojazdu	Zasięg [km]	Pojemność akumulatorów [kWh]	Średnie zużycie energii elektrycznej [kWh/100 km]
Smart EQ Fortwo (82 KM), cabrio, 2-drzwiowy	155,0	17,6	13,1
Dacia Spring Cargo (45 KM), furgon, 5-drzwiowy	305,0	26,8	11,9
Fiat 500 (118 KM), hatchback, 3-drzwiowy (Red)	321,0	42,0	14,0
Hyundai Kona (204 KM), SUV, 5-drzwiowy (Platinum)	484,0	64,0	15,4

Tesla Model S Long Range (680 KM), hatchback, 5-drzwiowy	663,0	100,0	17,6
Mercedes EQS (333 KM), hatchback, 5-drzwiowy (Electric Art)	732,0	107,8	16,9

Źródło: opracowanie własne na podstawie [14].

Produkcja energii elektrycznej w Polsce w oparciu głównie o węgiel skutkuje tym, że pojazdy elektryczne podczas ładowania baterii trakcyjnych przyczyniają się pośrednio do emisji zanieczyszczeń do środowiska. W wyniku produkcji energii elektrycznej w elektrowniach cieplnych na bazie paliw kopalnych powstają spaliny, które zawierają, poza dwutlenkiem węgla, również tlenek węgla, tlenki siarki, tlenki azotu, lotne związki organiczne i pyły. Wielkość emisji zanieczyszczeń do powietrza jest w relacji do sprawności technologicznej oczyszczania spalin z procesu wytwarzania energii elektrycznej przed odprowadzeniem tych spalin do środowiska.

Pojazdy elektryczne na etapie ładowania będą w Polsce tym bardziej ekologiczne, im większy udział w polskiej energetyce będą miały źródła energii odnawialnej.

Zwiększenie sprawności wytwarzania energii elektrycznej z paliw kopalnych w elektrowniach cieplnych zawodowych może w pewnym stopniu wpłynąć na obniżenie emisji zanieczyszczeń powodowanej produkcją energii elektrycznej na potrzeby napędu samochodów. W tabeli 4 przedstawiono wyniki badania wpływu sprawności wytwarzania energii elektrycznej w elektrowniach cieplnych na wielkość emisji dwutlenku węgla dla samochodu elektrycznego na etapie eksploatacji. W badaniu uwzględniono, że udział paliw (węgiel, gaz ziemny) w całkowitej produkcji energii elektrycznej w Polsce wynosi 82,64% (stan na rok 2022) (tab. 1). Jednostkowy wskaźnik emisji dwutlenku węgla dla procesu wytwarzania energii elektrycznej z paliw (tab. 4) wyraża emisję wynikającą ze zużycia paliw (węgiel, gaz ziemny) do wytworzenia energii elektrycznej i na pokrycie strat przesyłowych w sieci elektroenergetycznej. Jednostkowy wskaźnik emisji dwutlenku węgla z samochodu elektrycznego (tab. 4) wyraża emisję spowodowaną produkcją energii elektrycznej na potrzeby napędu samochodu o zużyciu energii elektrycznej 17,6 kWh/100 km.

Tabela 4. Wpływ sprawności wytwarzania energii elektrycznej w elektrowniach cieplnych zawodowych na wielkość emisji dwutlenku węgla z samochodu elektrycznego na etapie eksploatacji

Parametr	Jednostka	Sprawność [%]		
		42,50	44,00	46,00
Jednostkowy wskaźnik emisji dwutlenku węgla dla procesu wytwarzania energii elektrycznej z paliw	g/kWh	714,15	689,90	659,81
Jednostkowy wskaźnik emisji dwutlenku węgla z samochodu elektrycznego przy średnim zużyciu energii elektrycznej 17,6 kWh/100 km	g/ km	125,69	121,40	116,13

Źródło: opracowanie własne.

Wzrost sprawności produkcji energii elektrycznej w elektrowniach cieplnych z 42,5% do 46% może przyczynić się do obniżenia jednostkowego wskaźnika emisji dwutlenku węgla z samochodu elektrycznego na etapie eksploatacji o ok. 8,2%.

4. Wielkość emisji dwutlenku węgla z samochodu osobowego z silnikiem benzynowym

Rozważania będą dotyczyć wielkości emisji dwutlenku węgla na etapie eksploatacji samochodu osobowego z silnikiem spalinowym benzynowym. Otrzymane wyniki będzie można porównać z wynikami badania emisji z samochodu elektrycznego na etapie eksploatacji pojazdu. Etap eksploatacji obejmuje użytkowanie pojazdu, czynności obsługowe oraz postoje pojazdu. Dokładna ocena emisji dwutlenku węgla w fazie użytkowania pojazdu spalinowego powinna uwzględniać emisję spowodowaną spalaniem paliwa w celu napędu pojazdu oraz emisję spowodowaną pozyskiwaniem tego paliwa tj. jego produkcją i dystrybucją [5]. Wielkość emisji dwutlenku węgla przez pojazd i zużycie energii należą do ważnych kryteriów oceny oddziaływania pojazdu na środowisko [15].

Na wielkość emisji dwutlenku węgla spowodowanej spalaniem paliwa przez pojazd istotny wpływ ma ilość zużytego paliwa na przejechanym dystansie [16]. Średnie zużycie paliwa przez pojazd na danej odległości warunkuje szereg czynników, m.in. parametry techniczne układu napędowego samochodu, w tym typ układu silnikowego, pojemność i moc silnika, rodzaj skrzyni biegów, masa pojazdu [17] warunki komunikacyjne, warunki pogodowe, prędkość i styl jazdy, rodzaj infrastruktury drogowej oraz wartość opałowa paliwa.

W dokumentacji homologacyjnej każdego pojazdu wyszczególnione jest m.in. średnie jednostkowe zużycie paliwa badane według testu laboratoryjnego WLTP [18] (test obowiązuje od września 2018 roku w odniesieniu do nowych samochodów osobowych wprowadzanych do obrotu w UE). Przykładowo dla nowego konstrukcyjne pojazdu Volkswagen Passat (sedan 4-drzwiowy) wersja Elegance z silnikiem spalinowym benzynowym o pojemności 1,5 dm^3, mocy 110 kW (150 KM), 6-biegową manualną skrzynią biegów, zużycie średnie benzyny według testu WLTP wynosi 5,8 dm^3/100 km [14].

Wielkość emisji dwutlenku węgla ze spalania danej objętości paliwa (benzyny silnikowej) można obliczyć według zależności:

$$E_{CO2} = WE_{CO2} \cdot Z \tag{2}$$

gdzie: E_{CO2} – wielkość emisji dwutlenku węgla spowodowanej spalaniem paliwa, wyrażona w kg/km, WE_{CO2} – wskaźnik emisji dwutlenku węgla pochodzącej ze spalania 1 dm^3 paliwa, wyrażony w kg/dm^3, Z – zużycie paliwa, wyrażone w dm^3/km.

Wskaźnik emisji WE_{CO2} można wyznaczyć ze wzoru:

$$WE_{CO2} = U_C \cdot WE_{CO2/C} \cdot \rho \tag{3}$$

gdzie: WE_{CO2} – wskaźnik emisji dwutlenku węgla ze spalania 1 dm^3 paliwa, wyrażony w kg/dm^3, U_C – udział węgla w paliwie, wyrażony w kg węgla/kg paliwa, $WE_{CO2/C}$ – wskaźnik emisji dwutlenku węgla ze spalania całkowitego i zupełnego 1 kg węgla, wyrażony w kg CO_2/kg ($WE_{CO2/C}$ = 44/12 kg CO_2/kg), ρ – gęstość paliwa, wyrażona w kg/dm^3.

Wyniki obliczeń wielkości charakteryzujących emisję dwutlenku węgla ze spalania paliwa na etapie użytkowania wybranych pojazdów przedstawiono w tabeli 5. Do obliczeń przyjęto, że udział masowy węgla w benzynie wynosi 0,85 kg węgla/kg paliwa, a gęstość benzyny 0,755 kg/dm^3.

Tabela 5. Wielkość emisji dwutlenku węgla spowodowanej spalaniem paliwa (benzyny) na etapie użytkowania pojazdu pasażerskiego

Konfiguracja pojazdu	Zużycie średnie benzyny [dm^3/100 km]	Wskaźnik emisji $WE_{CO2/}$ [kg/dm^3]	Wielkość emisji E_{CO2} [g/km]
Volkswagen Passat Elegance 1.5 (150 KM), sedan, 4-drzwiowy	5,80	2,35	136,30
Volkswagen Passat Elegance 2.0 (280 KM), kombi, 5-drzwiowy	9,00		211,50
Audi A3 1.5 (150KM), sedan, 4-drzwiowy	5,60		131,60
Audi A3 RS3 Quattro 2.5 (400 KM), hatchback, 5-drzwiowy	9,00		211,50
Kia Sportage 1.6 (150 KM), nadwozie SUV, 5-drzwiowy	6,30		148,05
Kia Niro 1.6 (141 KM), nadwozie SUV, 5-drzwiowy, układ silnikowy hybrydowy (HEV)	4,50		105,75
Toyota Yaris Cross 1.5 (125 KM), nadwozie SUV, 5-drzwiowy	5,90		138,65
Toyota Prius 2.0 (223 KM), hatchback, 5-drzwiowy, układ silnikowy plug-in (PHEV)	0,50		11,75

Źródło: opracowanie własne.

Wskaźnik emisji gazów cieplarnianych w cyklu życia benzyny (bez dodatku biokomponentów) wyrażony w gramach ekwiwalentu dwutlenku węgla (g CO_{2eq}) na megadżul energii dostarczonej z benzyną (MJ) wynosi 93,3 g CO_{2eq}/MJ [19]. Dla wyrażenia tego wskaźnika w g CO_{2eq}/dm^3 paliwa uwzględniono wartość opałową benzyny 32 MJ/dm^3 [19]. Wskaźnik emisji gazów cieplarnianych w cyklu życia benzyny w przeliczeniu na decymetr sześcienny paliwa ma wartość 2,99 kg/dm^3. Według [20] wskaźnik emisji dwutlenku węgla w fazie użytkowania pojazdów zasilanych benzyną wynosi 2,83 kg/dm^3.

Na potrzeby dalszych rozważań przyjęto, że w cyklu życia benzyny wskaźnik emisji dwutlenku węgla jest równy wskaźnikowi emisji gazów cieplarnianych (2,99 kg/dm^3). Emisja dwutlenku węgla w gramach na kilometr pochodząca z samochodu Volkswagen Passat Elegance 1.5 (tab. 5) z uwzględnieniem cyklu życia paliwa wynosi 173,42 g/km. Stanowi ona sumę emisji dwutlenku węgla spowodowanej spalaniem paliwa podczas jazdy (136,30 g/km) oraz powstającej w procesie wytwarzania paliwa w ilości umożliwiającej jazdę w danych warunkach (37,12 g/km). Spalanie paliwa generuje 78,6% całkowitej emisji dwutlenku węgla (suma emisji ze spalania i z wytwarzania paliwa).

5. Podsumowanie

Elektromobilność oparta na pojazdach elektrycznych to ważny kierunek zmian w transporcie w dążeniu do jego zeroemisyjności. Mowa tu o pojazdów elektrycznych, które wykorzystują do napędu energię elektryczną gromadzoną w akumulatorach przez podłączenie do zewnętrznego źródła zasilania. Pojazdy elektryczne podczas jazdy nie emitują do powietrza zanieczyszczeń charakterystycznych dla pojazdów spalinowych. Ich rozpowszechnienie przyczyni się do poprawy jakości powietrza i do obniżenia udziału sektora transportu drogowego w zanieczyszczeniu atmosfery.

W kompleksowej ocenie oddziaływania pojazdów elektrycznych na jakość powietrza trzeba uwzględnić emisję do powietrza w całym cyklu istnienia pojazdów. Na etapie eksploatacji pojazdów źródłem zanieczyszczeń jest emisja spowodowana produkcją i dystrybucją energii elektrycznej potrzebnej do ładowania baterii pojazdów elektrycznych. Największy wkład w tę emisję wnoszą technologie wytwarzania energii elektrycznej w oparciu o spalanie paliw kopalnych.

Z analizy struktury paliw i innych nośników energii pierwotnej zużytych do wytworzenia energii elektrycznej w Polsce w roku 2022 wynika, że udział węgla w produkcji energii elektrycznej wynosi 76,9%, a gazu ziemnego 5,7%. Przeprowadzone obliczenia wskazują, że w Polsce na jedną megawatogodzinę pobranej z sieci energii elektrycznej przypada emisja 714,15 kg dwutlenku węgla spowodowana zużyciem paliw kopalnych (węgla i gazu ziemnego) na wytworzenie energii elektrycznej i na pokrycie strat przesyłowych energii.

Analiza parametrów technicznych rozważanych samochodów osobowych elektrycznych wskazuje, że średnie zużycie energii elektrycznej przez te samochody wynosi od 11,9 kWh/100 km do 17,6 kWh/100 km (dla testu laboratoryjnego WLTP). Na podstawie uzyskanych wyników obliczeń jednostkowego wskaźnika emisji dwutlenku węgla wyrażonego w gramach na kilowatogodzinę energii elektrycznej o wartości 714,15 g CO_2/kWh można stwierdzić, że zużyciu energii elektrycznej przez samochód elektryczny w zakresie (11,9-17,6) kWh/100 km odpowiada emisja dwutlenku węgla na poziomie (84,98-125,69) g/km. Wartość ta obrazuje wpływ produkcji energii elektrycznej magazynowanej w akumulatorach pojazdów elektrycznych podczas ich ładowania na zanieczyszczenie powietrza dwutlenkiem węgla, który będąc gazem cieplarnianym przyczynia się do intensyfikacji naturalnego efektu cieplarnianego.

Wyznaczona wielkość emisji dwutlenku węgla w gramach na kilometr pochodząca z samochodu osobowego z silnikiem spalinowym benzynowym (Volkswagen Passat Elegance 1.5) (tab. 5) z uwzględnieniem cyklu życia benzyny wynosi 173,42 g/km. Na wynik ten składa się emisja dwutlenku węgla powstającego w wyniku spalania paliwa w silniku pojazdu oraz generowana w procesie wytwarzania paliwa w ilości potrzebnej do napędu pojazdu. Spalanie paliwa w silniku pojazdu odpowiada za 78,6% całkowitej emisji dwutlenku węgla (suma emisji ze spalania i z wytwarzania paliwa).

Pojazdy elektryczne będą tym bardziej przyjazne dla środowiska na etapie ich eksploatacji im bardziej niskoemisyjne będą technologie wytwarzania energii elektrycznej. Wyzwaniem dla branży energetycznej jest wzrost produkcji energii elektrycznej ze źródeł energii odnawialnej. Wpłynie to na poprawę jakości powietrza i klimatu.

Literatura

1. Obwieszczenie Marszałka Sejmu Rzeczypospolitej Polskiej z dnia 7 kwietnia 2022 r. *w sprawie ogłoszenia jednolitego tekstu ustawy o elektromobilności i paliwach alternatywnych* (Dz. U. z 2022, poz. 1083).
2. Dane PSPA, *Licznik elektromobilności: kolejny rekordowy rok na polskim rynku e-mobility*, pspa.com.pl/aktualności [data dostępu: 19.01.2023].
3. Rocznik statystyczny, *Ochrona środowiska 2022*, Główny Urząd Statystyczny, Warszawa 2022.
4. Kawamoto R., Mochizuki H., Moriguchi Y., Nakano T., Montohashi M., Sakai Y., Inaba A., *Estimation of CO2 emissions of internal combustion engine vehicle and battery electric vehicle using LCA*, Sustainability, MDPI, 11(9), 2019, s. 1-15.

5. Nordelof A., Messagie M., Tillman AM., Ljunggren Soderman M., Van Mierlo J., *Environmental impacts of hybrid, plug-in hybrid, and battery electric vehicles-what can we learn from life cycle assessment*?, The International Journal of Life Cycle Assessment, 19, 2014, s. 1866-1890.
6. Buberger J., Kersten A., Kuder M., Eckerle R., Weyh T., Thiringer T., *Total CO_2-equivalent life-cycle emission from commercially available passenger cars*, Ranewable and Sustainable Energy Reviews, 159, 2022, s. 1-12.
7. Yazdanie M., Noembrini F., Heinen S., Espinel A., Boulouchos K., *Well-to-whell costs, primary energy demand, and greenhouse gas emissions for the production and operation of conventional and alternative vehicles*, Transportation Research Part D: Transport and Environment, 48, 2016, s. 63-84.
8. *Struktura produkcji energii elektrycznej w elektrowniach krajowych, wielkości wymiany energii elektrycznej z zagranicą i krajowe zużycie energii*, Polskie Sieci Elektroenergetyczne, www.pse.pl/dane-systemowe/funkcjonowanie-kse/raporty-miesieczne-z-funkcjonowania-rb/raporty-miesieczne [data dostępu: 13.02.2023].
9. *Wartości opałowe (WO) i wskaźniki emisji CO_2 w roku 2020 do raportowania w ramach systemu handlu uprawnieniami do emisji za rok 2023*, Krajowy Ośrodek Bilansowania i Zarządzania Emisjami, Warszawa 2022.
10. Rocznik statystyczny *Gospodarka paliwowo-energetyczna w latach 2020 i 2021*. Główny Urząd Statystyczny, Warszawa 2022.
11. Marczak H., Droździel P., *Analysis of pollutants emission into the air at the stage of an electric vehicle operation*, Journal of Ecological Engineering, 22(8), 2021, s. 182-188.
12. Raport branżowy PTPiREE, *Energetyka. Dystrybucja. Przesył*, Poznań 2022.
13. Rozporządzenie Komisji (UE) 2017/1151 z 1 czerwca 2017 uzupełniające rozporządzenie (WE) nr 715/2007 Parlamentu Europejskiego i Rady *w sprawie homologacji typu pojazdów silnikowych w odniesieniu do emisji zanieczyszczeń pochodzących lekkich pojazdów pasażerskich i użytkowych* (*Euro 5 i Euro 6*) (Dz. Urz. L 175/1 z 7.7.2017).
14. Autokatalog Instytutu Badań Rynku Motoryzacyjnego SAMAR, www.autokatalog.pl [data dostępu: 08.02.2023].
15. Marczak H., *Evaluation of ecological efficiency resulting from the application of modern buses in urban communications*, Journal of Ecological Engineering, 18(5), 2017, s. 110-117.
16. Siedlecki M., Galant M., Rymaniak M., Ziółkowski A., *Badania emisji zanieczyszczeń pojazdu wyposażonego w silnik z bezpośrednim wtryskiem benzyny w rzeczywistych warunkach eksploatacji*, Autobusy, 12, 2017, s. 404-409.
17. Drabik D., *Analiza możliwości obniżenia zużycia paliwa w samochodach osobowych, Autobusy*, 6, 2017, s. 612-615.
18. Dimaratos A., Tsokolis D., Fontaras G., Tsiakmakis S., Ciuffo B., Samaras Z., *Comparative evaluation of the effect of various technologies on light-duty vehicle CO_2 emision over NEDC and WLTP*, Transportation Research Procedia, 14, 2016, s. 3169-3178.
19. Rozporządzenie Ministra Klimatu i Środowiska z dnia 4 lipca 2022 r. *w sprawie metodyki obliczania emisji gazów cieplarnianych, określenia wskaźników ich emisji oraz wartości opałowej dla poszczególnych paliw i wartości energetycznej energii elektrycznej* (Dz. U. 2022, poz. 1494).
20. Edwards R., Larive J., Beziat J., *Well-to-wheels analysis of future automotive fuels and power trains in the European context*, Report version 3d, EUR 24952 EN, Luxembourg, Publications Office of the European Union, 2011.

Wpływ elektromobilności na jakość powietrza

Streszczenie
Elektromobilność oparta na pojazdach elektrycznych to potrzebny kierunek rozwoju sektora transportu. Może przynieść wiele korzyści dla środowiska i gospodarki. Brak emisji do powietrza z pojazdów elektrycznych w miejscach ich użytkowania to jedna z zalet elektromobilności. W kompleksowej ocenie wpływu na środowisko pojazdów elektrycznych na etapie ich eksploatacji nie można pominąć emisji do powietrza w procesie wytwarzania i dystrybucji energii elektrycznej potrzebnej do ładowania baterii pojazdów elektrycznych. Tematem badawczym w rozdziale monografii jest określenie wielkości emisji dwutlenku węgla z produkcji energii elektrycznej na bazie paliw kopalnych: węgla i gazu ziemnego. Wybór paliw kopalnych do analizy ma uzasadnienie w tym, że aktualnie w Polsce węgiel ma największy udział w strukturze paliw wykorzystywanych w produkcji energii elektrycznej. Do realizacji tematu badań wykorzystano metodę analizy danych technicznych pojazdów elektrycznych, danych o strukturze paliw wykorzystywanych do produkcji energii elektrycznej w elektrowniach cieplnych zawodowych w Polsce oraz wartości wskaźników emisji dwutlenku węgla z produkcji energii elektrycznej. Na podstawie uzyskanych wyników można stwierdzić, że zużyciu energii elektrycznej przez samochód elektryczny w zakresie 11,9–17,6 kWh/100 km odpowiada emisja dwutlenku węgla na poziomie 84,98-125,69 g/km. Wartość ta obrazuje wpływ produkcji energii elektrycznej magazynowanej w akumulatorach pojazdów elektrycznych podczas ich ładowania na zanieczyszczenie powietrza dwutlenkiem węgla. Wyznaczono wielkość emisji dwutlenku węgla pochodzącą z analizowanego samochodu osobowego z silnikiem spalinowym benzynowym z uwzględnieniem cyklu życia benzyny, wynosi ona 173,42 g/km. Emisyjność pojazdów elektrycznych na etapie ich eksploatacji będzie tym mniejsza, im bardziej ekologiczne źródła energii będą wykorzystywane do wytwarzania energii elektrycznej. Poprawę jakości powietrza w Polsce można również uzyskać poprzez zwiększenie sprawności wytwarzania energii elektrycznej w energetycznych blokach węglowych i gazowych.
Słowa kluczowe: elektromobilność, pojazdy elektryczne, etap eksploatacji pojazdu elektrycznego, emisja dwutlenku węgla

Impact of electromobility on air quality

Abstract
Electromobility based on electric vehicles is a necessary direction for the development of the transport sector. It can bring many benefits for the environment and for the economy. The lack of air emissions from electric vehicles in the places of their use is one of the advantages of electromobility. In a comprehensive assessment of the environmental impact of electric vehicles at the stage of their operation, emissions into the air in the process of generating and distributing electricity needed to charge the batteries of electric vehicles cannot be omitted. The research topic in the chapter is to determine the amount of carbon dioxide emissions from the production of electricity based on fossil fuels: coal and natural gas. The choice of fossil fuels for the analysis is justified by the fact that currently in Poland coal has the largest share in the structure of fuels used in the production of electricity. To implement the research topic, the method of analyzing technical data of electric vehicles, data on the structure of fuels used to produce electricity in commercial thermal power plants in Poland and the values of carbon dioxide emission indicators from electricity production were used. Based on the results obtained, it can be concluded that the consumption of electricity by an electric car in the range of 11.9-17.6 kWh/100 km corresponds to the emission of carbon dioxide at the level of 84.98-125.69 g/km. This value illustrates the impact of the production of electricity stored in the batteries of electric vehicles during their charging on carbon dioxide air pollution. The amount of carbon dioxide emissions from the analyzed passenger car with a petrol engine was determined, taking into account the life cycle of petrol, it is 173.42 g/km. The emissivity of electric vehicles at the stage of their operation will be the lower, the more ecological energy sources are used to generate electricity. Air quality in Poland can also be improved by increasing the efficiency of electricity generation in coal and gas power units.
Keywords: electromobility, electric vehicles, electric vehicle operation phase, carbon dioxide emission

Monika Walkowicz[1], Anna Różańska[2]

Problematyka i metodologia badań powierzchni dotykowych z miedzi i stopów miedzi o właściwościach przeciwdrobnoustrojowych

1. Wprowadzenie

Tematyka pracy dotyczy miedzi i stopów miedzi o właściwościach przeciwdrobnoustrojowych, które współcześnie są najskuteczniejszymi materiałami metalicznymi w eliminacji drobnoustrojów chorobotwórczych na powierzchniach dotykowych, przeznaczonych w szczególności do zastosowań w jednostkach opieki zdrowotnej i miejscach użyteczności publicznej takich jak m.in. dworce, lotniska, placówki handlowe, szkoły, teatry, środki transportu zbiorowego, w których występują największe skupiska ludzi będących źródłem zakażeń. Od czasu przeprowadzenia pierwszych niezależnych badań laboratoryjnych i klinicznych w szpitalach w Wielkiej Brytanii i Stanach Zjednoczonych, na podstawie których doświadczalnie dowiedziono skuteczności miedzi w eliminowaniu drobnoustrojów na tle innych materiałów metalicznych takich jak m.in. stal nierdzewna, aluminium czy srebro, środowisko medyczne przyznało, że wprowadzenie powierzchni dotykowych wykonanych miedzi i jej stopów może stanowić istotne wsparcie w zwalczaniu zakażeń związanych z opieką zdrowotną [1]. Rocznie, w całej Unii Europejskiej występuje ponad 5 milionów przypadków zakażeń szpitalnych, z których ponad 50 000 jest śmiertelnych [2]. To przyczyna wysokich kosztów ponoszonych rokrocznie przez systemy opieki zdrowotnej, ale przede wszystkim to problem o ważnym znaczeniu społecznym. Jednym z głównych problemów współczesnego zdrowia publicznego jest narastanie lekooporności bakterii, zwłaszcza w zakażeniach u pacjentów hospitalizowanych i przebywających w jednostkach opieki długoterminowej, ale także epidemie zakażeń wirusowych. Przykładem może tu być epidemia grypy w Polsce w 2009 roku, kiedy to odnotowano ponad 9000 przypadków zakażenia wirusem grypy A/H1N1, a liczba ofiar śmiertelnych wyniosła około 200 osób, w tym wiele małych dzieci i niemowląt. Odnotować należy także światową pandemię wywoływaną przez koronawirusa SARS-CoV-2, w wyniku której według danych Światowej Organizacji Zdrowia (WHO, ang. *World Health Organization*) z dnia 17 kwietnia 2023 roku, liczba zgłoszonych zgonów związanych z COVID-19 na całym świecie wyniosła ponad 11,5 miliona. Należy jednak pamiętać, że liczba ta może być nieco niższa od rzeczywistej liczby zgonów, ponieważ nie wszystkie kraje są w stanie dokładnie monitorować i raportować o wszystkich zgonach z powodu pandemii koronawirusa [3].

Wyniki przeprowadzonych na całym świecie testów mikrobiologicznych pozwoliły na zapoczątkowanie procesu wprowadzania powierzchni dotykowych wykonanych ze stopów miedzi do jednostek opieki zdrowotnej m.in. w Japonii, Niemczech, Francji,

[1] monika.walkowicz@agh.edu.pl, AGH Akademia Górniczo-Hutnicza, Wydział Metali Nieżelaznych, Katedra Przeróbki Plastycznej i Metaloznawstwa Metali Nieżelaznych, al. A. Mickiewicza 30, 30-059 Kraków.
[2] a.rozanska@uj.edu.pl, Zakład Kontroli Zakażeń i Mykologii, Katedra Mikrobiologii Wydziału Lekarskiego CMUJ, ul. Czysta 18, 31-121 Kraków.

Finlandii, Grecji, Wielkiej Brytanii jako skutecznego sposobu wspomagającego eliminację drobnoustrojów ze środowiska, a przez to zapobiegania infekcjom. W związku z powyższym przebadanych pod kątem skuteczności przeciwbakteryjnej 350 różnych gatunków miedzi i jej stopów zostało oficjalnie zatwierdzonych i zarejestrowanych przez Amerykańską Agencję ds. Ochrony Środowiska (EPA, ang. *Environmental Protection Agency*), a także nadano im państwowe oświadczenia zdrowotne. Od tego czasu określany mianem Miedzi Przeciwdrobnoustrojowej (ang. *Antimicrobial Copper*) o nazwie handlowej „Cu+" materiał jest oznaczeniem używanym przez wiodących światowych wytwórców sprzętu oraz wyposażenia medycznego i rehabilitacyjnego. Znak ten niesie ze sobą informację, że produkt wykonany został z miedzi lub jej stopu o właściwościach przeciwdrobnoustrojowych, a tym samym jego powierzchnia dotykowa zapewnia skuteczność eliminacji wirusów oraz mikroorganizmów chorobotwórczych, takich jak bakterie i grzyby [4].

W związku z powyższym od kilku lat produkty Cu+ zaczęto sukcesywnie wprowadzać do placówek medycznych na całym świecie (m.in. kliniki Asklepiosa w Hamburgu, szpitala Selly Oak w Birmingham, szpitala Kitasato w Tokyo, szpitala Del Cobre w Chile) w postaci elementów armatury sanitarnej (m.in. baterii łazienkowych, uchwytów prysznicowych dla niepełnosprawnych, desek sedesowych, przycisków spłukujących do WC, podajników na mydło itp.), włączników oświetlenia, klamek drzwiowych, uchwytów meblowych, stojaków na kroplówki, poręczy łóżek i krzeseł, blatów stolików zabiegowych itp. w miejsce dotychczas stosowanych elementów wykonanych ze stali nierdzewnej, czy tworzyw sztucznych.

Także w Polsce wyroby z powierzchniami dotykowymi z miedzi przeciwdrobnoustrojowej są wdrażane do jednostek opieki zdrowotnej. Przykładem jest Wojewódzki Szpital Specjalistyczny we Wrocławiu, który od 2012 roku bierze udział w polskim programie pilotażowym wprowadzania miedzi do krajowych miejsc użyteczności publicznej, w ramach którego Oddział Nefrologiczny tamtejszej jednostki medycznej wyposażony został w elementy wykonane z przeciwbakteryjnej miedzi i jej stopów, takie jak m.in. klamki i uchwyty meblowe, panele drzwiowe, ławki łazienkowe przyścienne, stoliki przyłóżkowe, kratki łazienkowe podłogowe, okładziny na łóżka i na wózki do przewożenia chorych, dozowniki mydła w łazienkach pacjentów. Wybór oddziału podyktowany był dużą liczbą obecnych na nim zakażeń wieloopornymi szczepami bakterii, szczególnie niebezpiecznymi dla chorych z obniżoną odpornością z powodu dializowania i po transplantacji. Oprócz wrocławskiego szpitala implementacją pełnej gamy asortymentowej miedzianych wyrobów z powierzchniami dotykowymi na swoich oddziałach zainteresowane są kolejne krajowe placówki opieki zdrowotnej.

Warto zaznaczyć, że wyroby z miedzi przeciwdrobnoustrojowej oprócz jednostek opieki zdrowotnej wdrażane są na całym świecie coraz to powszechniej w miejscach użyteczności publicznej w postaci m.in. systemów poręczy schodowych (np. stacja metra Bueras w Santiago w Chile, port lotniczy Cangonhas w São Paulo w Brazylii), okładzin drzwiowych (np. szkoły w Atenach w Grecji) czy kontuarów (np. klinika Ochiai w Chibe w Japonii).

W tym miejscu należy pokreślić, iż zastosowanie wyrobów wykonanych z miedzi przeciwdrobnoustrojowej w polskich jednostkach opieki zdrowotnej zależy od trzech czynników: dostępności na rynku stopów, półwyrobów i wyrobów finalnych, obecności w kraju producentów sprzętu i wyposażenia medycznego oraz podaży na wyroby z miedzi

przeciwbakteryjnej. Wszystkie te warunki w kraju są spełnione. W Polsce istnieje zarówno silnie rozwinięty przemysł miedziowy zakładów metalurgiczno-przetwórczych, jak i rynek producentów wyrobów przeznaczonych do jednostek opieki zdrowotnej. Trzeci czynnik, czyli podaż i dostępność krajowych produktów wykonanych z miedzi o właściwościach przeciwdrobnoustrojowych wynika z dostępności krajowej wiedzy know-how o relacji pomiędzy różnymi gatunkami miedzi i jej stopami a ich właściwościami przeciwdrobnoustrojowymi, a tym samym informacji technicznych o przeciwbakteryjnej miedzi i jej stopach zarówno z zakresu materiałoznawstwa oraz technologii wytwarzania i przetwarzania, jak i informacji z zakresu mikrobiologii. W szczególności w przeciągu ostatnich 10 lat w Polsce w wyniku realizacji różnych projektów badawczo-rozwojowych upowszechnione zostały wyniki badań wpływu m.in. różnej zawartości miedzi i innych pierwiastków oraz różnych stanów umocnienia stopów na ich skuteczność przeciwbakteryjną. Ponadto znany jest stopień wpływu stanu powierzchni miedzi i jej stopów (m.in. brązy, mosiądze, miedzionikle) po procesie chemicznej, termicznej i mechanicznej obróbki odwzorowującej rzeczywiste warunki eksploatacji powierzchni dotykowych na właściwości przeciwdrobnoustrojowe w kontakcie z drobnoustrojami kolonizującymi w krajowych miejscach użyteczności publicznej. W konsekwencji, wytypowane zostały grupy gatunków miedzi i jej stopów niezbędnych do zastosowania na wybrane wyroby o właściwościach przeciwdrobnoustrojowych i na tej podstawie opracowano założenia technologii wytwarzania nowych stopów miedzi oraz produktów z miedzi przeciwbakteryjnej z uwzględnieniem środowiska techniczno-technologicznego krajowych zakładów metalurgicznych oraz uwarunkowań polskiego przemysłu przetwórstwa miedzi. Należy także dodać, że w kraju w dalszym ciągu prowadzone są badania i testy mikrobiologicznych nad skutecznością przeciwdrobnoustrojową materiałów, zarówno metalicznych, jak i tworzyw sztucznych. W konsekwencji opracowana jest polska metodologia przeprowadzania testów mikrobiologicznych pod kątem badań właściwości przeciwdrobnoustrojowych miedzi i jej stopów przeznaczonych na wyroby o powierzchniach dotykowych. Wspomniane informacje dostępne są w monografii pt. *Miedź przeciwdrobnoustrojowa. Materiały – Powierzchnie dotykowe – Aplikacje* [5]. Dodatkowo w obliczu występującego w kraju dużego zapotrzebowania na wyroby o powierzchniach dotykowych wykonanych z miedzi i jej stopów podejmowane stale przez krajowe ośrodki naukowe i przemysłowe interdyscyplinarnych badań nad rozwiązaniami z zakresu przedstawionej problematyki właściwości materiałowych, technologicznych i mikrobiologicznych miedzi Cu+ jest jak najbardziej zasadne.

2. Problematyka krajowych badań nad miedzią przeciwdrobnoustrojową

Spośród wielu naukowych i praktycznych zagadnień istotnych dla rozwoju prekursorskiej w Polsce tematyki miedzi i stopów miedzi o właściwościach przeciwbakteryjnych należy wymienić m.in.:

- identyfikację i klasyfikację stopów miedzi o właściwościach przeciwdrobnoustrojowych przy wykorzystaniu aktualnego stanu wiedzy i danych pochodzących z Amerykańskiej Agencji Ochrony Środowiska;
- metodologię przeprowadzania mikrobiologicznych testów laboratoryjnych i klinicznych oceny skuteczności przeciwdrobnoustrojowej miedzi, stopów miedzi i innych materiałów metalicznych, a na tym tle poznanie mechanizmów eliminacji poszczególnych szczepów bakteryjnych i wirusowych przez jony miedzi;

- technologię wytwarzania półwyrobów hutniczych i wyrobów gotowych z miedzi i stopów miedzi o różnym składzie chemicznym, stanie umocnienia oraz powierzchni, a następnie parametryzację ich właściwości materiałowych, technologicznych i przeciwdrobnoustrojowych;
- charakteryzację skuteczności materiałów na osnowie miedzi pod kątem eliminacji drobnoustrojów w stosunku do innych materiałów metalicznych, i na tej podstawie opracowanie wytycznych w zakresie wyboru konkretnych gatunków miedzi lub stopów miedzi charakteryzujących się najkorzystniejszym zespołem właściwości mechanicznych i użytkowych, ze szczególnym uwzględnieniem odporności przeciwkorozyjnej oraz walorów estetycznych, ceny i skuteczności przeciwdrobnoustrojowej;
- badania nad technologiami przetwórstwa wybranych stopów miedzi na konkretne półwyroby i wyroby finalne przeznaczone do jednostek opieki zdrowotnej i wytypowanych pod kątem wysokich zagrożeń chorobotwórczych innych miejsc użyteczności publicznej.

Wiedza z zakresu wspomnianych tematów umożliwiała opracowanie założeń technologii produkcji nowych wyrobów i na tej podstawie uruchomienie produkcji w Polsce m.in. przez krajowe przedsiębiorstwa, które z uwagi na niszowe produkty Cu+ są zainteresowane rozwojem tematyki miedzi przeciwdrobnoustrojowej w kraju i jej kontynuacją w oparciu o własne zaplecze badawczo-rozwojowe. Zatem oprócz wysokiej wartości naukowej i poznawczej (m.in. testy właściwości przeciwdrobnoustrojowych stopów miedzi, prekursorskie badania wpływu chemicznej, termicznej i mechanicznej obróbki powierzchni na właściwości mechaniczne i użytkowe materiałów metalicznych oraz na zdolność do eliminacji drobnoustrojów) tematyka badań dotyczących miedzi przeciwdrobnoustrojowej ma silny wydźwięk aplikacyjny [6, 7]. Charakter realizowanych w tym obszarze badań materiałowych i mikrobiologicznych nad przeciwdrobnoustrojowymi gatunkami miedzi i jej stopów znajduje swój wyraz w aspekcie społecznym, którym są zaplanowane działania zmierzające do krajowej certyfikacji wyrobów, a także procedury związane z przeprowadzeniem procesu legislacyjnego [8]. Główną ideą podjęcia niniejszych prac jest zapełnienie niszy na polskim rynku i zastosowanie w jednostkach opieki zdrowotnej wyrobów z miedzi Cu+, co bezpośrednio znajdzie swój wyraz w krajowej produkcji stopów miedzi, półwyrobów i wyrobów.

3. Metodologia krajowych badań nad miedzią przeciwdrobnoustrojową

Przedstawiona tematyka badawcza dotyczy przeciwdrobnoustrojowych właściwości miedzi i jej stopów w zastosowaniach na wyroby z powierzchniami dotykowymi, dedykowane do polskich jednostek opieki zdrowotnej oraz innych miejsc użyteczności publicznej. Tematyka prac wpisuje się w szeroki kontekst zagadnień związanych ze skutecznością mikrobiologiczną gatunków miedzi i jej stopów do eliminacji drobnoustrojów chorobotwórczych, ich parametryzacją oraz implementacją na często dotykanych powierzchniach w jednostkach opieki zdrowotnej oraz innych miejscach użyteczności publicznej. Potwierdzona w badaniach klinicznych i laboratoryjnych na całym świecie zdolność do eliminacji mikroorganizmów, takich jak bakterie, grzyby i wirusy przez materiały na bazie miedzi ma charakter interdyscyplinarny, gdyż z jednej strony dotyczy aspektów natury materiałowej i techniczno-technologicznej, z drugiej zaś problematyki mikrobiologicznej oraz zakażeń szpitalnych i związanych z nimi powikłań [9, 10].

W szczególności pierwsza grupa zagadnień obejmuje obszar nauk technicznych zestawionych pod kątem wymagań dotyczących:

- metalurgii w zakresie terminologii, klasyfikacji, składu chemicznego, charakterystyk materiałowych i własności, tolerancji wymiarów i kształtu, oznaczeń, procedur pobierania próbek i metod badań miedzi i stopów miedzi (w formie półwyrobów i wyrobów takich jak m.in. stopy odlewnicze i stopy do przeróbki plastycznej, stopy wstępne, gąski, odlewy, wyroby odlewane nie przerobione plastycznie, materiały wyjściowe do kucia przerobione i nie przerobione plastycznie, blachy, taśmy, płyty, krążki, druty, pręty, rury, kształtowniki, odkuwki),
- inżynierii materiałowej w zakresie: własności powierzchni, wymiarów, odchyłek kształtu, składu chemicznego, własności mechanicznych, wielkości ziarna i metod badań materiałów (m.in. blach, pasów) platerowanych miedzią i stopami miedzi oraz w zakresie charakterystyk materiałowych, ocen ryzyka i przeglądu czynników w przypadku ochrony materiałów metalowych przed korozją.

Druga grupa zagadnień obejmuje natomiast tematykę z obszaru nauk medycznych i nauk o zdrowiu zestawione pod kątem wymagań dotyczących:

- technologii medycznej w zakresie:
  - terminologii, procesu analizy, określania, projektowania, sprawdzania i walidacji użyteczności oraz bezpieczeństwa urządzeń i sprzętu medycznego (m.in. stojaki do kroplówek, uchwyty łazienkowe, meble oraz urządzenia zabiegowe i pomocnicze);
  - klasyfikacji, oznaczeń, metod badań, zasad oceny bezpieczeństwa biologicznego i systemów zarządzania jakością wyrobów medycznych;
  - powołań normatywnych, terminologii, oznaczeń i dokumentacji dostarczanych przez wytwórców wyrobów medycznych;
  - projektowania ergonomicznego sprzętu i urządzeń medycznych;
  - koncepcji, zasad i praktyk odnoszących się do społecznej odpowiedzialności;
- mikrobiologii w zakresie terminologii, klasyfikacji i metodyki badań oceny skuteczności przeciwdrobnoustrojowej materiałów (m.in. metali i tworzyw sztucznych).

Ponadto przedstawiona tematyka badawcza wpisuje się w dużej mierze w problemy społeczne w aspekcie nabytych przez pacjentów w szpitalach infekcji i zakażeń oraz związanych z nimi powikłań i zgonów [11, 12].

W związku z powyższym tematyka badawcza koncentruje się zarówno wokół: metalu czyli miedzi i jej stopów oraz krajowego rynku na wyroby z powierzchniami przeciwdrobnoustrojowymi, jak również drobnoustrojów chorobotwórczych (szczepów bakteryjnych i wirusowych) i testów mikrobiologicznych. Pierwszy obszar badawczy w kraju realizowany jest przez zespół naukowy z Wydziału Metali Nieżelaznych AGH w Krakowie, drugi natomiast przez Katedrę Mikrobiologii Uniwersytetu Jagiellońskiego – Collegium Medicum.

4. Charakter aplikacyjny wyników badań nad miedzią przeciwdrobnoustrojową

Pierwszą przesłanką stanowiącą o zapotrzebowaniu na wyniki badań nad miedzią przeciwdrobnoustrojową, które mają duży potencjał aplikacyjny, jest wyraźna przewaga miedzi i stopów miedzi (jako materiałów inżynierskich do inaktywacji drobnoustrojów) nad innymi materiałami inżynierskimi, które aktualnie wykorzystywane są w polskich

jednostkach opieki zdrowotnej [13]. W obecnych czasach coraz większe znaczenie odgrywa zapobieganie infekcjom, nie zaś ich leczenie, jako tańsza i bardziej skuteczna metoda walki z nimi, a przy tym mniej dotkliwa dla pacjenta [14]. Takie możliwości stwarza zastąpienie tradycyjnych materiałów takich jak m.in. tworzywa sztuczne i stal nierdzewna stopami miedzi. Tym samym zastosowanie nowych, przeciwdrobnoustrojowych elementów wyposażenia jednostek opieki zdrowotnej utrudnia przeżycie, wzrost i wytworzenie biofilmu patogenom, a tym samym ich przetrwanie w środowisku szpitalnym i potencjalne zakażenie kolejnych pacjentów. Jeżeli w ten sposób można zapobiec chociażby jednemu zgonowi pacjenta z powodu zakażenia szpitalnego, to jest to wystarczający powód do tego, aby koszt nowej inwestycji był opłacalny i szybko się zwrócił. Należy podkreślić, iż wyroby na bazie miedzi i jej stopów są w całości przeciwdrobnoustrojowe tzn. nawet kiedy ich powierzchnie zostaną zarysowane, efektywność bójcza pozostaje nadal wysoka, produkty nie ulegają ścieraniu tak jak powłoki lub inne pokrycia [15]. Ponadto stopy miedzi są jedynymi materiałami z powierzchniami przeciwdrobnoustrojowymi zarejestrowanymi przez Amerykańską Agencję ds. Ochrony Środowiska jako oficjalne produkty stosowane w dziedzinie zdrowia publicznego [16].

Interesująca jest także kwestia skuteczności innych materiałów metalicznych oraz tworzyw sztucznych. Otóż wyniki przeprowadzonych na świecie porównawczych badań skuteczności przeciwdrobnoustrojowej miedzi, aluminium, stali nierdzewnej, PVC i polietylenu wyraźnie wskazały skuteczność przeciwbakteryjną miedzi oraz jej brak w przypadku pozostałych ww. materiałów [17, 18]. Warto także poruszyć kwestię przeciwbakteryjnych właściwości srebra z uwagi na dominującą współcześnie w społeczeństwie opinię, jakoby srebro charakteryzowało się największą skutecznością w eliminacji drobnoustrojów chorobotwórczych. W testach laboratoryjnych i klinicznych opisanych w światowym piśmiennictwie i realizowanych w warunkach temperatury pokojowej i wilgotności otoczenia powłoki polimerowe pokryte cząsteczkami srebra zachowują się w ten sam sposób jak kontrolna stal nierdzewna tzn. nie wykazują właściwości przeciwdrobnoustrojowych [19]. Należy jednak zaznaczyć, iż często do oceny efektywności antybakteryjnej powłok na bazie srebra wykorzystuje się Japońską Normę JIS:Z 2801 [20], której wytyczne odnośnie sposobu przeprowadzania testów mikrobiologicznych nie są zbliżone do rzeczywistych warunków panujących w szpitalach i innych miejscach użyteczności publicznej. Otóż japoński standard zaleca wykonywanie testów trwających 24 h w temperaturze 37°C i ponad 90% wilgotności względnej. Ponadto próbka jest owinięta plastikową folią w celu zachowania wilgotności i wyeliminowania strat ciepła. W takich warunkach testowych powłoki zawierające srebro przejawiają właściwości przeciwdrobnoustrojowe, czego przyczyną jest nadmierna wilgoć biorąca udział w reakcjach wymiany jonów srebra. Jednak w przypadku, gdy temperatura i wilgotność spadają do wartości panujących w pomieszczeniach szpitalnych, omawiane powłoki tracą swoje właściwości przeciwdrobnoustrojowe, a ich działanie jest analogiczne do stali nierdzewnej. Dodatkowo srebro o wysokiej zawartości kruszcu, z uwagi na wysoką cenę i niezadawalający poziom właściwości fizycznych nie może być materiałem przeznaczonym na powierzchnie dotykowe do zastosowań w jednostkach opieki zdrowotnej. W odróżnieniu od wszystkie przetestowanych w Wielkiej Brytanii stopów miedzi, które wykazały skuteczność przeciwbakteryjną w różnych warunkach – m.in. w typowych dla pomieszczeń (20°C i 20-24% wilgotności względnej) w czasie od 15 minut do 2 godzin ujawniły redukcję 99,9999% bakterii (m.in. *Acinetobacter baumannii, Pseudomanas aeruginosa,*

Escherichia coli O157:H7, Enterobacter aerogenes, Salmonella enteritidis, Helicobacter pylori, Campylobacter jejuni, Legionella pneumophilia, Listeria monocytogenes, Staphylococcus aureus – również *MRSA, Enterococcus* spp. – w tym szczepy wankomycynooporne, *Clostridioides difficile*, prątki gruźlicy), wirusów (polio, adeno, grypy A/H1N1) oraz grzybów (*Aspergillus niger, Candida albicans*), których ilość obejmowała zakres 250 000-1 000 000 komórek/ml. Ponadto oprócz naturalnych właściwości przeciwdrobnoustrojowych praktycznym aspektem zastosowania miedzi w jednostkach opieki zdrowotnej jest łatwość utrzymania czystości powierzchni dotykowych [21]. Produkty ze stopów miedzi należy czyścić tak samo jak inne powierzchnie dotykowe. Ponadto, wiedząc, iż podstawową obroną przed mikrobami są ogólnie zalecane praktyki higieniczne stosowane wobec powierzchni dotykowych, to powierzchnie z miedzi przeciwdrobnoustrojowej poprzez swoją ciągłą aktywność (24 godz. dziennie, 7 dni w tygodniu) pozwalają na zmniejszenie zanieczyszczenia i kontaminacji drobnoustrojami pomiędzy poszczególnymi okresami dekontaminacji czyszczenia w szpitalach [22].

W związku z powyższym należy podkreślić, iż miedź przeciwdrobnoustrojowa jest materiałem o szerokim spektrum zastosowania. Przeznaczeniem są powierzchnie kontaktu z ciałem człowieka, powierzchnie kontaktu z żywnością oraz powierzchnie wentylacyjno-klimatyzacyjne. W pierwszym przypadku bezpośrednimi obszarami zastosowania są:

- przede wszystkim jednostki opieki zdrowotnej: szpitale, przychodnie, prywatne gabinety lekarskie, stacje sanitarno-epidemiologiczne, zakłady rehabilitacyjne itp.;
- transport publiczny: pojazdy komunikacji zbiorowej, dworce, lotniska, metra itp.;
- budynki użyteczności publicznej: szkoły, biblioteki, urzędy, muzea i miejsca kultury itp.;
- hotelarstwo i gastronomia: restauracje, bary, stołówki, hotele itp.;
- miejsca aktywności sportowa: kluby fitness, baseny itp.

Miedź przeciwdrobnoustrojowa może być także wykorzystywana do poprawy higieny żywności [23]. Higieniczne powierzchnie kontaktowe wykonane z miedzi i jej stopów mogą zmniejszyć liczbę przypadków zakażeń krzyżowych groźnymi patogenami pokarmowymi, takimi jak: *E. coli O157:H7, Campylobacter jejuni, Listeria monocytogenes, Salmonella enteritidis* w procesach produkcji żywności. Co ważne, miedź ma naturalną zdolność eliminacji tych niebezpiecznych drobnoustrojów w szybkim tempie zarówno w temperaturze chłodniczej (4°C), jak i w temperaturze pokojowej (20°C). Ponadto przeciwbakteryjne właściwości miedzi znajdują swoje zastosowanie do poprawy jakości powietrza w pomieszczeniach [24]. W nowoczesnym budownictwie obawa przed działaniem chorobotwórczych drobnoustrojów doprowadziła do priorytetowej potrzeby usprawnienia warunków higienicznych systemów inżynierii sanitarnej (HVAC – skrót obejmujący systemy ogrzewania, wentylacji i klimatyzacji), które są przyczyną ponad 60% wszystkich infekcji powstałych u ludzi spędzających dużo czasu w budynkach (wykazano, że płytki aluminiowe w systemach HVAC były źródłem licznych populacji drobnoustrojów). Szczególnie u ludzi z niską odpornością, kontakt z mikroorganizmami pochodzącymi z systemów inżynierii sanitarnej HVAC może skutkować poważnymi infekcjami prowadzącymi do śmierci. Istnieją liczne publikacje wskazujące na fakt, że miedź eliminuje wiele patogenów powszechnie występujących w systemach HVAC, a na terenie USA prowadzone są intensywne badania kliniczne w celu potwierdzenia hipotezy, że komponenty systemów inżynierii sanitarnej HVAC zawierające miedź istotnie poprawiają jakość powietrza wewnątrz budynków [25-27].

5. Podsumowanie

Podsumowując rozważania przedstawione w pracy i podkreślając na podstawie powyższych informacji wszechstronność zastosowania miedzi i jej stopów jako powierzchni dotykowych należy zaznaczyć, że wyniki badań w postaci wytycznych dotyczących przeciwdrobnoustrojowych gatunków miedzi i jej stopów, półproduktów oraz produktów gotowych, a w konsekwencji technologii ich produkcji są przedmiotem zapotrzebowania i bezpośredniego wykorzystania w praktyce. Grupami docelowi są krajowi producenci:

- materiałów wsadowych w postaci stopów miedzi do przeróbki plastycznej i odlewniczych;
- półproduktów, które stanowią. miedziane wyroby płaskie uzyskane w wyniku procesu walcowania (na zimno i na gorąco) tj. blachy, taśmy, pasy; wyroby ciągnione w postaci rur, prętów i drutów oraz wyroby wyciskane w postaci płaskowników i kształtowników z miedzi przeciwdrobnoustrojowej;
- wyrobów gotowych w postaci: asortymentu i sprzętu medycznego tj. systemów zabudowy bloków operacyjnych i sal szpitalnych, wózków anestezjologicznych, stolików opatrunkowych, sterylizatorów, autoklawów, puszek i pojemników medycznych, elementów elektrostymulatorów itp.; dekoracyjnych elementów meblowych czyli uchwytów szuflad, klamek, wieszaków, prowadnic itp.; systemów i instalacji zasilania gazami medycznymi; innego wyposażenia medycznego np. klamek, okuć drzwi, armatury sanitarnej, poręczy itp.

ale także polskie społeczeństwo i środowisko medyczne w zakresie działań promocyjnych.

Realizacja badań z obszaru miedzi przeciwdrobnoustrojowej umożliwia zdobycie wiedzy know-how dotyczącej wymagań dla wyrobów z miedzi i jej stopów pod kątem właściwości przeciwdrobnoustrojowych, które są odpowiedzią na:

- wymagania technologa i konstruktora tj. właściwości mechaniczne (m.in. wytrzymałość na rozciąganie, granica plastyczności), wielkość ziarna, równomierna struktura, wydzielenia na granicy ziaren itp.;
- wymagania producenta wyrobów gotowych tj. podatność do skrawania, dobre właściwości antykorozyjne, w tym odporność na pot rąk ludzkich, podatność do dalszej przeróbki plastycznej;
- wymagania rynku i konsumenta tj. właściwości bakteriostatyczne, funkcjonalność, cena, estetyka.

Uwagi ogólne/Podziękowania

Projekt został sfinansowany ze środków Narodowego Centrum Badań i Rozwoju: umowa nr PBS3/A9/32/2015.

Literatura

1. Arendsen L.P., Thakar R., Sultan A.H., *The use of copper as an antimicrobial agent in health care, including obstetrics and gynecology*, Clinical Microbiology Reviews, 32, 2019, s. e00125-18.
2. Suetens C., Latour K., Kärki T., Ricchizzi E., Kinross P., Moro M.L., Jans B., Hopkins S., Hansen S., Lyytikäinen O., Reilly J., Deptula A., Zingg W., Plachouras D., Monnet D.L., *Prevalence of healthcare-associated infections, estimated incidence and composite*

antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017, Eurosurveillance, 23, 2018, s. 1800516.
3. Rupani P.F., Nilashi M., Abumalloh R.A., Asadi S., Samad S., Wang S., *Coronavirus pandemic (Covid-19) and its natural environmental impacts*, International Journal of Environmental Science and Technology, 17, 2020, s. 4655-4666.
4. Aillon-Garcia P., Parga-Landa B., Guillen-Grima F., *Effectiveness of copper as a preventive tool in health care facilities. A systematic review*, American Journal of Infection Control, 2023, s. 1-11.
5. Walkowicz M., *Miedź przeciwdrobnoustrojowa. Materiały-Powierzchnie dotykowe-Aplikacje*, Impuls, 2018.
6. Arendsen L.P., Thakar R., Sultan A.H., *The use of copper as an antimicrobial agent in health care, including obstetrics and gynecology*, Clinical Microbiology Reviews, 32, 2019, s. e00125-18.
7. Mitra D., Kang E., Noeh K.G., *Antimicrobial copper-based materials and coatings: potential multifaceted biomedical applications*, ACS Applied Materials & Interfaces, 12, 2020, s. 21259-21282.
8. Patil V., Singh R., Kanade K., Yi G., *Copper nanomaterials derived preventive technologies for Covid-19 pandemic: a review*, Advanced Journal of Chemistry-Section A., 5, 2022, s. 1-9.
9. Bharadishettar N., Bhat U., Udaya B., Panemangalore D., *Coating technologies for copper based antimicrobial active surfaces: a perspective review*, Metals, 11, 2021, s. 711.
10. Zerbib S., Vallet L., Muggeo A., Champs C., Lefebvre A., Jolly D., Kanagaratnam L., *Copper for the prevention of outbreaks of health care–associated infections in a long-term care facility for older adults*, Journal of the American Medical Directors Association, 21, 2020, s. 68-71.
11. Querido M., Aguiar L., Neves P., Pereira C., Teixeira J., *Self-disinfecting surfaces and infection control*, Colloids and Surfaces B: Biointerfaces, 178, 2019, s. 8-21.
12. Champagne V., Sundberg K., Helfritch D., *Kinetically deposited copper antimicrobial surfaces*, Coatings, 9, 2019, s. 257.
13. Scully J., *The Covid-19 pandemic, part 1: can antimicrobial copper-based alloys help suppress infectious transmission of viruses originating from human contact with high-touch surfaces?*, Corrosion, 76, 2020, s. 523-527.
14. Montero D., Arellano C., Pardo M., Vera R., Galvez R., Cifuentes M., Berasain M., Gomes M., Ramirez C., Vidal R., *Antimicrobial properties of a novel copper-based composite coating with potential for use in healthcare facilities*, Antimicrobial Resistance & Infection Control, 8, 2019, s. 16.
15. Tripathy A., Kumar A., Sreedharan S., Muralidharan G., Pramanik A., Nandi D., Sen P., *Fabrication of low-cost flexible superhydrophobic antibacterial surface with dual-scale roughness*, ACS Biomaterials Science & Engineering, 4, 2018, s. 2213-2223.
16. Aillón-García P., Parga-Landa B., Guillén-Grima F., *Effectiveness of copper as a preventive tool in health care facilities. A systematic review*, American Journal of Infection Control, 2023, s. 1-11.
17. Schmidt M., Attaway H., Fairey S., Steed L., Michels H., Salgado C., *Copper continuously limits the concentration of bacteria resident on bed rails within the intensive care unit*, Infection Control & Hospital Epidemiology, 34, 2013, s. 530-533.
18. Michels H., Keevil C., Salgado C., Schmidt M., *From laboratory research to a clinical trial: copper alloy surfaces kill bacteria and reduce hospital-acquired infections*, Health Environments Research & Design Journal, 9, 2015, s. 64-79.

19. Walji S., Aucoin M., *A critical evaluation of current protocols for self-sterilizing surfaces designed to reduce viral nosocomial infections*, American Journal of Infection Control, 48, 2020, s. 1255-1260.
20. Vincent M., Hartemann P., Engels-Deutsch M., *Antimicrobial applications of copper*, International Journal of Hygiene and Environmental Health, 219, 2016, s. 585-591.
21. Schmidt M., Attaway H., Fairey S., Howard J., Mohr D., Craig S., *Self-disinfecting copper beds sustain terminal cleaning and disinfection effects throughout patient care*, Applied and Environmental Microbiology, 86, 2020, s. e.01886-19.
22. Salah I., Parkin I., Allan E., *Copper as an antimicrobial agent: recent advances*, RSC Advances, 11, 2021, s. 18179-18186.
23. Popescu V., Prodan D., Cuc S., Saroşi C., Furtos G., Moldovan A., Carpa R., Bomboş D., *Antimicrobial poly(lactic acid)/copper nanocomposites for food packaging materials*, Materials, 16, 2023, s. 1415.
24. Navaratnam S., Nguyen K., Selvaranjan K., Zhang G., Mendis P., Aye L., *Designing post Covid-19 buildings: approaches for achieving healthy buildings*, Buildings, 12, 2022, s. 74.
25. Gomes I., Simões M., Simões L.C., *Copper surfaces in biofilm control*, Nanomaterials, 10, 2020, s. 2491.
26. Schmidt M., Attaway H., Terzieva S., Marshall A., Steed L., Salzberg D., Hamoodi H., Khan J., Feigley C., Michels H., *Characterization and control of the microbial community affiliated with copper or aluminum heat exchangers of HVAC systems*, Current Microbiology, 65, 2012, s. 141-149.
27. Kauffman R., *Determine the mechanism for copper plating and methods for its elimination from HVAC systems*, ASHRAE Transactions, 114, 2008, s. 360-374.

Problematyka i metodologia badań powierzchni dotykowych z miedzi i stopów miedzi o właściwościach przeciwdrobnoustrojowych

Streszczenie

Praca dotyczy przeciwdrobnoustrojowych właściwości miedzi i jej stopów w zastosowaniach na wyroby z powierzchniach dotykowych przeznaczonymi do użytkowania w jednostkach opieki zdrowotnej oraz innych miejscach użyteczności publicznej. Problematyka ta wpisuje się w szeroki kontekst zagadnień związanych ze skutecznością miedzi i jej stopów w eliminacji drobnoustrojów chorobotwórczych, parametryzacją właściwości powierzchni dotykowych wykonanych z tych materiałów oraz wdrożeniem często dotykanych wyrobów miedzianych do jednostek opieki zdrowotnej oraz innych miejscach użyteczności publicznej. Tematyka potwierdzonej w badaniach klinicznych i laboratoryjnych na całym świecie zdolności materiałów na osnowie miedzi do eliminacji mikroorganizmów ma charakter interdyscyplinarny, gdyż z jednej strony dotyczy aspektów natury materiałowej i techniczno-technologicznej metalu, z drugiej zaś problematyki mikrobiologicznej oraz społecznej z uwagi na problem zakażeń (w tym szpitalnych) nabywanych przez ludzi w miejscach użyteczności publicznej i ich rozprzestrzeniania. Tym tematom poświęcona jest niniejsza praca. Ukazuje problematykę i metodologię badań z obszar nauk technicznych dotyczących inżynierii materiałowej na tle zagadnień z obszaru nauk medycznych i społecznych. Aktualność tematyki w dobie pandemii koronawirusa wynika z konieczności poszukiwania i wdrażania do miejsc użyteczności publicznej materiałów wspomagających przeciwdziałanie rozprzestrzeniania mikroorganizmów chorobotwórczych (wirusów, bakterii, grzybów).
Słowa kluczowe: miedź, stopy miedzi, właściwości przeciwdrobnoustrojowe, inżynieria powierzchni

Issues and methodology of testing touch surfaces made of copper and copper alloys with antimicrobial properties

Abstract

The work concerns the antimicrobial properties of copper and its alloys in applications for products with touch surfaces intended for healthcare units and other public places. This issue fits into the broad context of issues related to the effectiveness of copper and its alloys in the elimination of pathogenic microorganisms, parameterization of the properties of touch surfaces made of these materials and the implementation of frequently touched copper products in healthcare units and other public places. The topic of the ability of

copper-based materials to eliminate microorganisms, confirmed in clinical and laboratory tests all over the world, is interdisciplinary, as on the one hand it concerns the material and technical and technological aspects of the metal, and on the other hand, microbiological and social issues due to the problem of infections (in including hospital) acquired by people in public places and their spread. This work is devoted to these topics. It shows the issues and methodology of research in the field of technical sciences regarding materials engineering against the background of issues in the field of medical and social sciences. The topicality of the subject in the era of the coronavirus pandemic results from the need to search for and implement in public places materials supporting the prevention of the spread of pathogenic microorganisms (viruses, bacteria, fungi).
Keywords: copper, copper alloys, antimicrobial properties, surface engineering

Andrzej Wartecki[1], Wiesław J. Jermacz[2], Adam Lewandowski[3]

Innowacyjno-rozwojowe uwarunkowania działalności spółki wodociągowej w ujęciu ekonomicznym oraz ekologicznym na przykładzie Śremskich Wodociągów Spółka z o.o.

1. Wprowadzenie

W ostatnich latach można zaobserwować przyśpieszenie tempa wdrażania innowacyjnych metod zarządzania i zwiększenie efektywności gospodarowania przedsiębiorstw. Innowacje mają istotny wpływ i wspomagają tempo rozwoju przedsiębiorstw oraz obniżają ceny produktów, zaspokajających ciągle rosnące i zmieniające się ludzkie potrzeby. Wiele badań z tego zakresu pokazuje, iż innowacje są mocno powiązane z podniesieniem wydajności, czy pozycji konkurencyjnej przedsiębiorstwa na rynku. Wydaje się wiec, że zdolność do kreowania i absorpcji innowacji jest jednym z fundamentalnych wyzwań XXI wieku. Analiza i ocena innowacyjności przedsiębiorstw należy do stosunkowo trudnych problemów, które są wynikiem szeregu czynników zarówno wewnątrz przedsiębiorstwa, jak i na zewnątrz, w jego otoczeniu. Duże znaczenie odgrywa tutaj: mikrootoczenie i zróżnicowanie środowiskowe, skala, struktura, strategie, przedsiębiorstw, aspiracje przedsiębiorców, ich kreatywność, twórczość i skłonność do podejmowania ryzyka związanego z działalnością innowacyjną.

Zainteresowanie innowacyjnością można zaobserwować również w działalności jednostek samorządu terytorialnego (j.s.t.). Wynika to częściowo z przenikania pewnych negatywnych zjawisk wynikających z dotykających nas wielu kryzysów o różnym obliczu: migracyjnym, gospodarczym, społecznym, ekologicznym i politycznym. Globalizacja w działalności gospodarczej jest tak widoczna, iż kryzys w jednym regionie może mieć znaczny wpływ na strukturę i rozwój (j.s.t.) w innych częściach Polski, Unii Europejskiej i świata. Różny poziom rozwoju j.s.t. na świecie powoduje przenikanie pewnych sprawdzonych wzorców działań innowacyjnych, które wymuszają sytuacje kryzysowe. Zarządzanie w sytuacjach kryzysowych wymaga przedsiębiorczego i innowacyjnego podejścia, prowadzenie działań trudnych do naśladowania. Zarządzający przedsiębiorstwami stają współcześnie wobec nowych zadań, które narzucają im sytuacje kryzysowe, a ich cechą charakterystyczną są liczne zmiany w branży, sektorze i najbliższym otoczeniu. *W czasach burzliwych* – jak zauważa Peter F. Dracker – *trzeba tak zarządzać przedsiębiorstwem, by mogło wytrzymać nagłe ciosy i wykorzystywać nagłe nieoczekiwane szanse. Co oznacza, że w czasach burzliwych trzeba radzić sobie z problemami o zasadniczym znaczeniu i to radzić sobie dobrze. Zarządzanie w czasach burzliwych musi*

[1] andrzej.wartecki@interia.pl, Zakład Ekonomiki przedsiębiorstw i zarządzania kadrami, Wydział Ekonomiczny, WAS-E w Środzie Wlkp., www.http//wase.pl/.

[2] jermacz.w@interia.pl, www.http//wase.pl//, Zakład Pedagogiki Opiekuńczo-Wychowawczej z Terapią Pedagogiczną, Wydział Nauk o Wychowaniu, WAS-E w Środzie Wlkp.

[3] adam.lewandowski@urzad.srem.pl, doktorant Uniwersytetu Ekonomicznego w Poznaniu, wykładowca Wyższej Szkoły Bankowej w Poznaniu.

się zaczynać od dyskusji nad nowymi, odmiennymi wymaganiami w sferze podstawowych problemów przetrwania i sukcesów w istniejącym biznesie (…). *W czasach burzliwych menadżerowie nie mogą zakładać, że dzień jutrzejszy będzie przedłużeniem dzisiejszego. Przeciwnie, muszą zarządzać z myślą o zmianach jako szansie i jako zagrożeniu* [1].

Spółki komunalne, w których organy samorządu terytorialnego są akcjonariuszami lub udziałowcami muszą zarządzać w sytuacjach kryzysowych z myślą o zmianach jako szansie i zagrożeniu, bo mają w założeniu pełnić rolę służebną wobec społeczności lokalnych, co czyni je z założenia firmami „misyjnymi". Z drugiej jednak strony są one graczami na wolnym rynku i muszą działać w sposób najbardziej zbliżony do firm czysto prywatnych.

Przedmiotem naszych rozważań jest próba identyfikacji wybranych aspektów funkcjonowania samorządowych spółek kapitałowych w perspektywie innowacyjno-rozwojowego uwarunkowania działalności spółki wodociągowej w ujęciu ekonomicznym oraz ekologicznym na przykładzie Śremskich Wodociągów Spółka z o.o.

2. Pojęcie innowacji i ich znaczenie w zakresie rozwoju działalności spółek samorządowych w ujęciu teoretycznym

Punktem wyjścia do naszych rozważań o problematyce dotyczących innowacyjno-rozwojowych uwarunkowań działalności spółki wodociągowej w ujęciu ekonomicznym i ekologicznym jest wyjaśnienie etymologii znaczenia pojęcia innowacje. Innowacje jak podaje słownik języka polskiego, to wprowadzenie czegoś nowego; też: rzecz nowo wprowadzona: nowość, nowatorstwo, innowacja technologiczna i rynkowa [2].

Pierwszy raz pojęcie innowacji do nauk ekonomicznych wprowadził J. A. Schupeter, który skupiał się przede wszystkim na innowacjach technicznych i ich znaczeniu dla gospodarki. Pod względem przedmiotowym określał on innowację niezwykle szeroko.

Według niego obejmuje ona:

- prowadzenie nowego towaru, z jakim konsumenci nie mieli jeszcze do czynienia lub nowego gatunku jakiegoś towaru;
- wprowadzenie nowej metody produkcji jeszcze praktycznie niewypróbowanej w danej dziedzinie przemysłu;
- otwarcie nowego rynku, czyli takiego, na którym dany rodzaj krajowego przemysłu uprzednio nie działał i to bez względu, czy rynek ten istniał wcześniej, czy też nie;
- zdobycie nowego źródła surowców lub półfabrykatów i to niezależnie od tego, czy źródło już istniało, czy też musiało być dopiero stworzone;
- wprowadzenie nowej organizacji jakiegoś przemysłu, np. stworzenie monopolu bądź jego złamanie [3].

Inną i nieco szerszej rozumianą definicją innowacji i bardziej odpowiednia do naszych dalszych rozważań jest ujęcie według Z. Pietrusińskiego. W opinii tego autora innowacje to zmiany celowo wprowadzane przez człowieka lub zaprojektowane przez układ cybernetyczny, które polegają na zastępowaniu dotychczasowych stanów rzeczy innymi, ocenianymi dodatnio w świetle określonych kryteriów składających się w sumie na postęp [4].

Poza Z. Pietrasińskim zagadnieniem innowacji w takim rozumieniu zajmowali się: Cz. Barański, R. Schulz, czy W. Okoń. Ostatni z wymienionych definiuje innowację jako zmianę struktury systemu szkolnego, jako całości lub struktury ważnych jego

składników, mającą na celu wprowadzenie ulepszeń o charakterze wymiernym. Takie podejście do innowacji można było zaobserwować w latach siedemdziesiątych w Ameryce Północnej i Europie Zachodniej. Do przedstawicieli tego nurtu zaliczyć można: P.F. Druckera, Ph. Kotlera, R.W. Griffina, M. E. Porter oraz polskich przedstawicieli zajmującymi się innowacjami, a między innymi S. Marciniaka, I. K. Hejduka, W. M. Grudzewskiego, A. Pomykalskiego i S. Gomułkę. Z kolei Innowacyjność w opinii P. Niedzielskiego ściśle wiąże się z posiadanymi zasobami, ale także umiejętnością ich wykorzystania, czyli dojrzałością innowacyjną (odpowiednim poziomem kultury organizacyjnej, warunkującym wykorzystanie posiadanych zasobów) [5]. Z kolei E. Mazur-Wierzbicka stwierdza, że spożytkowanie nowych pomysłów, dzięki którym powstają korzyści ekonomiczne, ulepszanie technologiczne oraz zmodyfikowane metody produkcji, to interpretacja pojęcia innowacji według M.E. Portera, który innowację utożsamiał z ulepszeniem produktu, procesu, nowych form dystrybucji i koncepcji zarządzania [6]. Najważniejsze w tym ujęciu jest to, że „innowacja jest celowym i zorganizowanym działaniem przedsiębiorców w celu osiągnięcia konkretnego celu ekonomicznego, bardziej precyzyjnego zaspokojenia potrzeb rynkowych, a także rozsądnego wykorzystania posiadanych lub możliwych do pozyskania zasobów". W tej interpretacji innowacja stanowi rezultat postępu wiedzy i wynalazku, czyli jest nowatorskim odkryciem, które zostaje wykorzystane w dziedzinie produkcji i prowadzi do powstawania nowych produktów [7]. W tym nurcie znajduje się także określenie A. Sopińskiej i W. Mierzejewskiej (2017), które stwierdzają, że innowacja to zmiany w metodach wytwarzania, produktach i w organizacji procesu produkcyjnego, opierające się na nowej lub dotychczas niewykorzystanej wiedzy. W procesie tworzenia innowacji istotną rolę odgrywają uwarunkowania wewnętrzne. Wiążą się one z posiadanymi przez przedsiębiorstwo zasobami oraz umiejętnościami ich pozyskania [8].

W szerokim znaczeniu zaprezentowanym przez W. Janasza innowacja oznacza twórczą zmianę [9]. W.M. Grudzewski oraz I.K. Hejduk utożsamiają ją z wiedzą i pomysłem na konkretny produkt lub usługę zaspokajającą potrzeby zgłoszone na rynku, co w dalszej konsekwencji generuje zysk dla przedsiębiorstwa i warunkuje dalszy jego rozwój [10]. Natomiast A. Wziatek i E. Balcerowicz utożsamiają innowację ze zmianą konkurencyjności produktu firmy: cenowej lub/i poza cenowej [11]. Z kolei B. Kowalak podkreśla, że wynalazek staje się innowacją dopiero wtedy, gdy można go kopiować na skalę przemysłową, przy akceptowanym poziomie kosztów [12].

W literaturze przedmiotu interpretacja pojęcia innowacyjności opiera się na czterech podejściach nastawionych na:

1. jednostkę i jej cechy wpływające na tworzenie innowacji (wykształcenie, reatywność, płeć itp.),
2. strukturę, gdzie główny akcent kładzie się na związek pomiędzy strukturą organizacyjną i innowacyjnością oraz pomiędzy przedsiębiorstwem i jego otoczeniem, powiązania interakcyjne, odpowiadające na pytanie: jak wyznaczone działanie wpływa na proces innowacyjny i odwrotnie,
3. systemowe tworzenie innowacji, którego celem jest rozstrzygnięcie jak krajowe i regionalne systemy innowacji mogą wpływać na działalność innowacyjną Przedsiębiorstw [13].

Pojęcie innowacyjności przedsiębiorstwa jest rozważane, jako jednowymiarowe zjawisko, oznaczające zdolność osoby/grupy do wytworzenia i skutecznego zastosowania

nowych pomysłów [14], czy też, jako zdolność przedsiębiorstwa do ciągłego poszukiwania, zastosowania w praktyce i rozpowszechniania innowacji, prowadząca do wzrostu jego nowoczesności oraz umocnienia pozycji konkurencyjnej na rynku [15] oraz jako pomyślne wprowadzenie do praktyki nowej rzeczy albo metody [16].

W świetle przytoczonych definicji można stwierdzić, że jest to bardzo trudny problem z uwagi na wielość definicji oraz brak ogólnie akceptowanych teorii na ten temat. Powoduje to, iż istnieje zapotrzebowanie społeczne, organizacyjne, ekonomiczne i ekologiczne na zajmowanie się tą problematyką, gdyż innowacyjność często mierzy się ilością wprowadzonych innowacji oraz wielkością nakładów przeznaczonych na te działania. Zmiany w sposobie widzenia innowacji i działalności innowacyjnej przedsiębiorstw wprowadziła w Polsce transformacja systemowa.

3. Teoretyczne uwarunkowanie analizy w ujęciu ekonomicznym oraz ekologicznym, działalności spółek samorządowych

Jednym z podstawowych czynników mających wpływ na konkurencyjność przedsiębiorstw jest poziom ich innowacyjności. Zdolność kreowania i wykorzystania innowacji staje się istotnym elementem, determinującym sprawność działania przedsiębiorstwa oraz jego rozwój na konkurencyjnym rynku. Aktywność innowacyjna przedsiębiorstw zależy od wielu czynników zarówno o charakterze zewnętrznym, jak i wewnętrznym, z punktu widzenia przedsiębiorstw. Na podstawie analizy literatury [17] poświęconej innowacyjnym metodom zarządzania przedsiębiorstwem na pierwszym miejscu pojawia się aspekt ekonomiczny. Mając na uwadze akcent ekonomiczny innowacje można analizować według miejsca, w którym rodzi się nowy pomysł lub idea bądź pewne zjawisko, które przyczynia się do powstania innowacji. Źródłem innowacji jest zarówno środowisko zewnętrzne, którego ciągła analiza ekonomiczna pozwala na wprowadzanie zmian, dostosowywanie ich do potrzeb otoczenia, jak i samo przedsiębiorstwo, które analizując własne zasoby i potrzeby działań w obecnie bardzo burzliwym i zmieniającym się otoczeniu. Wszyscy uczestnicy gry rynkowej funkcjonują w środowisku, które obejmuje czynniki trendy makrootoczenia (otoczenie dalsze, otoczenie ogólne, otoczenie globalne). Czynniki te mogą wpływać pozytywnie, ale również mogą hamować innowacyjność i rozwój przedsiębiorstwa. Zdaniem J. Brzóska [18] makrootoczenie tworzy różne zestawy szans i zagrożeń. Jego wpływ jest zróżnicowany pod względem:

- globalnego lub krajowego miejsca lokalizacji przedsiębiorstwa;
- kraju lub regionu funkcjonowania/działania przedsiębiorstwa (np. zróżnicowanie klimatyczne, kulturowe);
- zasięgu działania przedsiębiorstwa i jego ekspansji rynkowej;
- wielkości przedsiębiorstwa, na co może mieć wpływ np. proinnowacyjna polityka państwa.

Aktywności innowacyjna przedsiębiorstw kształtowana jest również pod wpływem uwarunkowań lokalnych i regionalnych, w którym działają. Region może być tutaj rozpatrywany jako organizacja ograniczająca ryzyko i niepewność, jako źródło innowacji oraz jako źródło pozyskiwania różnych umiejętności. Procesy innowacyjne, stanowiące rdzeń nowoczesnych metod zarządzania, zachodzą w specyficznym układzie powiązań przedsiębiorstw, instytucji naukowo-badawczych, organizacji pozarządowych, inicjatyw obywatelskich i sektora publicznego. Rośnie rola środowiska regionalnego i lokalnego, w jakim przychodzi funkcjonować przedsiębiorstwom. Akcentuje się przy tym wagę

polityki władz j.s.t., które mogą i powinny tworzyć korzystne warunki dla kształtowania klimatu sprzyjającego innowacyjności tworzonych samorządowych spółek komunalnych. Jednak jako główne czynniki innowacyjności można tutaj wymienić: typ sektora i rodzaj rynku, w którym funkcjonuje przedsiębiorstwo oraz warunki lokalne i regionalne. Działalność innowacyjna poszczególnych podmiotów może być również zdeterminowana istnieniem w regionie klastrów i regionalnych systemów innowacji oraz wymiana wiedzy, informacji i doświadczeń pomiędzy poszczególnymi podmiotami.

W opinii K. Poznańska [19] za jeden z charakterystycznych atrybutów innowacji uważa inercję innowacyjną, czyli opór przed zmianami jakie może ona powodować poprzez obniżenie poczucia bezpieczeństwa i wygodę pracowników i która pojawia się na każdym szczeblu zarządzania przedsiębiorstwem. Osoby, które projektują innowacje powinny nadawać innowacjom większość z cech, które mogą ułatwić ich planowanie, realizację, a następnie określenie przewidywanych efektów. W rezultacie, dla sprawnej realizacji takiej polityki gospodarczej, wysiłki władzy publicznej wszystkich szczebli skierowane powinny być na organizację procesu powstawania innowacji i tworzenie infrastruktury instytucjonalnej, ułatwiającej ten proces. Uwarunkowania procesów innowacyjnych mogą mieć wartość socjologiczną, psychologiczną, filozoficzną, historyczną, organizacyjną, ekonomiczną i ekologiczną. Uwarunkowania zmieniają się w czasie, mają charakter dynamiczny i zależą od wielu czynników.

W Polsce w okresie transformacji warunki ekonomiczno-ekologiczne i występujące w Unii Europejskiej w tym samym okresie są różne, ale są takie same czynniki, ale mają charakter dynamiczny i w danym momencie wpływające adekwatnie do aktualnej sytuacji politycznej i gospodarczej w danym kraju. Determinanty mają wpływ zarówno na przedsiębiorstwa, jak i na j.s.t. oraz nadzorowane przez nich spółki komunalne. W procesie analizy innowacyjności spółek komunalnych ważną rolę spełniają obok determinantów ekonomicznych czynniki ekologiczne, postulowane globalnie w formie zasad zrównoważonego rozwoju, które obligatoryjnie zobowiązują poszczególne rządy państw do tworzenia przepisów, ustaw, rozporządzeń jako filarów do zrównoważonego rozwoju każdego państwa. Działalność Polski, w zakresie ochrony środowiska, idzie w dobrym kierunku i nie można stwierdzić, że te problemy przez rząd są zaniedbywane.

Zgodnie z wymaganiami polityki Unii Europejskiej nasz kraj jako jej członek jest zobligowany do tworzenia przepisów prawnych dotyczących ochrony środowiska i kształtowania kultury ekologicznej wśród społeczeństwa. Każdy obywatel Polski musi wziąć udział choćby poprzez sortowanie śmieci w ramach gospodarstw domowych. Charakteryzując politykę Polski w zakresie wypełnienia zasad zrównoważonego rozwoju można stwierdzić, że jest ona dobrze ukierunkowana. Nasz kraj obecnie respektuje jako członek Unii Europejskiej wszystkie Dyrektywy Parlamentu Europejskiego oraz decyzje Komisji Europejskiej [20]. W świetle decyzji KE z dnia 23 lutego 2022 roku w kwestii ochrony środowiska został sformułowany wniosek dotyczący zapobieganiu negatywnemu oddziaływaniu na środowisko, niezgodnemu z najważniejszymi konwencjami środowiskowymi. Przedsiębiorstwa objęte zakresem wniosku będą musiały wprowadzić odpowiednie środki („zobowiązanie do dołożenia starań") zależnie od powagi i prawdopodobieństwa różnych skutków. Będą to środki dostępne dla danego przedsiębiorstwa w konkretnych okolicznościach. Obowiązkowe będzie również określenie priorytetów. Przykładem takiej działalności Rządu Polski jest regulacja prawna w formie ustawy „Prawo ochrony środowiska" i zgodnie z nią jest zobligowany do wypracowania

dokumentu strategicznego „Polityka ekologiczna państwa" ze wskazaniem określonych terminów realizacji. Ostatni taki dokument obowiązuje do 2030 roku. Celem głównym Polityki Ekologicznej Państwa 2030 jest rozwój potencjału środowiska na rzecz obywateli i przedsiębiorców, który odpowiada wprost celowi z obszaru „Środowisko" w Strategii na rzecz Odpowiedzialnego Rozwoju i został określony w następujący sposób:

- poprawa jakości środowiska i bezpieczeństwa ekologicznego;
- zrównoważone gospodarowanie zasobami środowiska;
- łagodzenie zmian klimatu i adaptacja do nich oraz zarządzanie ryzykiem klęsk żywiołowych;
- rozwijanie kompetencji (wiedzy, umiejętności i postaw) ekologicznych społeczeństwa;
- środowisko i administracja. Poprawa efektywności funkcjonowania instrumentów ochrony środowiska.

Zawarte w tej strategii wytyczne okażą się skuteczne pod względem ekonomicznym i ekologicznym jeżeli będą monitorowane i kontrolowane przez powołane do tego organy nadzoru. Różnorodność zachodzących zmian w otoczeniu i wewnątrz przedsiębiorstw, będąca często efektem procesów legislacyjnych, wymusza stosowanie unikatowych rozwiązań organizacyjnych w obszarze systemów i metod zarządzania. Znaczącą rolę w tym procesie powinien odgrywać nadzór korporacyjny. Stworzenie systemu pobudzającego potencjał innowacyjny przedsiębiorstwa wydaje się zatem najważniejszym zagadnieniem dotyczącym współczesnej problematyki nadzoru. Zagadnienia te nabierają szczególnego znaczenia w sektorze samorządowym w obszarze sprawowania nadzoru właścicielskiego nad spółkami komunalnymi.

Reasumując, na podstawie analizy źródłowej można stwierdzić, że wyzwanie, jakim jest stwarzanie perspektywy rozwoju i budowy gospodarki opartej na wiedzy, wymaga horyzontalnych, długofalowych oraz kompleksowych działań prowadzonych na różnych płaszczyznach.

4. Założenia metodologiczne przeprowadzonych badań i charakterystyka badanego przedsiębiorstwa

Przedmiotem badań jest wybrany podmiot komunalny – Śremskie Wodociągi Spółka z o.o. specjalizujący się w usługach komunalnych. Sferę podmiotową badań określały zarówno procesy tworzenia w tym przekształcenia oraz funkcjonowania samorządowej spółki kapitałowej z wykorzystaniem czynników innowacyjnych w wymiarze ekonomicznym i ekologicznym.

Celem badań jest analiza i ocena wykorzystania czynników kształtujących działalność innowacyjną skuteczności i ekonomiczności na podstawie sprawozdań finansowych oraz ekonomiczno-ekologicznych aspektów funkcjonowania tego rodzaju spółki kapitałowej w ujęciu praktycznym i funkcjonalnym. Realizacja w/w celu winno istotnie wpłynąć na podwyższenie poziomu świadomości innowacyjnej wśród przedsiębiorców oraz umożliwi wskazanie barier oraz czynników wpływających na działalność innowacyjną funkcjonowania (j.s.t.), a także nadzorowanych przez nich samorządowych spółek komunalnych.

Z przyjętych celów badawczych wyłonił się problem główny w formie pytania: „Jaki jest stan zawansowania i poziom rozwoju innowacyjnych działań funkcjonowania samorządowych spółek kapitałowych w ujęciu praktycznym i funkcjonalnym oraz jaki model

rozwiązań umożliwiłby korelacje funkcji ekonomicznych i ekologicznych dla tego rodzaju podmiotów?".

Skuteczność podejmowanych przez samorządowe spółki kapitałowe działań w aspekcie społeczno-ekonomicznym jest warunkowana wykorzystaniem rozwiązań z zakresu innowacyjnych uwarunkowań prawnych w sferze ekonomiczno-ekologicznej odnoszących się do strategii i zarządzania w różnych sytuacjach kryzysowych spółkami komunalnymi w ujęciu systemowym, co oznacza opracowanie praktycznych wniosków, w których uwzględnione zostanie podejście procesowe oraz analityczne.

Po wstępnej analizie problemu badawczego oraz literatury przedmiotu i analiz zgromadzonych już obserwacji i doświadczeń autorów dotyczących funkcjonowania tego typu podmiotów, sformułowano hipotezę główną, według której przyjęto następujące założenie, „że działalność innowacyjna samorządowych spółek komunalnych zależy od czynników pochodzących z makro i mikrootoczenia oraz korzystania z wiedzy i doświadczeń, które wymuszają sytuacje kryzysowe w zakresie innowacyjnego podejścia, prowadzenia działań trudnych wynikających z licznych zmian w otoczeniu tworzących szanse, a także zagrożenie dla ich rozwoju". W celu zweryfikowania tak przyjętego założenia zastosowana zostanie porównawczo prospektywna analiza funkcjonalna działań Śremskich Wodociągów Spółka z o.o., w oparciu o przeprowadzone badania i analizy zebranych materiałów oraz rocznych sprawozdań finansowych sporządzanych w przyjętym okresie badawczym.

5. Charakterystyka badanego przedsiębiorstwa

Początek historii Śremskich Wodociągów sięga 1907 roku, czyli momentu powstania sieci wodociągowej i Wodociągów Miejskich. Wówczas przedsiębiorstwo połączone było z Gazownią, tworząc jeden Zakład Komunalny. Nadzór pełnili tam urzędnicy miejscy oraz specjalnie powołana komisja. Wodociągi i Gazownia, mimo że działały jako jedna jednostka organizacyjna, posiadały odrębne regulaminy i statuty.

Od 1 lipca 1976 roku powstało Wojewódzkie Przedsiębiorstwo Wodociągów i Kanalizacji w Poznaniu Zakład w Śremie (oprócz Śremu do Poznania należał Nowy Tomyśl, Gniezno i Szamotuły). W tej strukturze Przedsiębiorstwo działało do 1991 roku, wówczas to WPWiK w Poznaniu uległo likwidacji. Śremskie Wodociągi stały się samodzielnym przedsiębiorstwem państwowym, podlegającym strukturalnie Wojewodzie Poznańskiemu.

Z dniem 10 lipca 1997 roku powstało Przedsiębiorstwo Wodociągów i Kanalizacji w Śremie Sp. z o.o., którego właścicielem jest gmina Śrem. Z dniem 18 marca 2015 roku nazwa zmieniona została na: Śremskie Wodociągi spółka z ograniczoną odpowiedzialnością. Zgodnie z zapisem aktu założycielskiego Spółki, przedmiotem jej działalności jest:

- pobór, uzdatnianie i dostarczanie wody (36.00.Z);
- odprowadzanie i oczyszczanie ścieków (37.00.Z);
- roboty związane z budową rurociągów przesyłowych i sieci rozdzielczych (42.21.Z);
- pozostała sprzedaż detaliczna prowadzona poza siecią sklepową, straganami i targowiskami (47.99.Z);
- pozostałe badania i analizy techniczne (71.20.B);
- wytwarzanie paliw gazowych (35.21.Z);
- wytwarzanie energii elektrycznej (35.11.Z);
- produkcja nawozów i związków azotowych (20.15.Z);
- sprzedaż hurtowa wyrobów chemicznych (46.75.Z);
- sprzedaż hurtowa pozostałych półproduktów (46.76.Z).

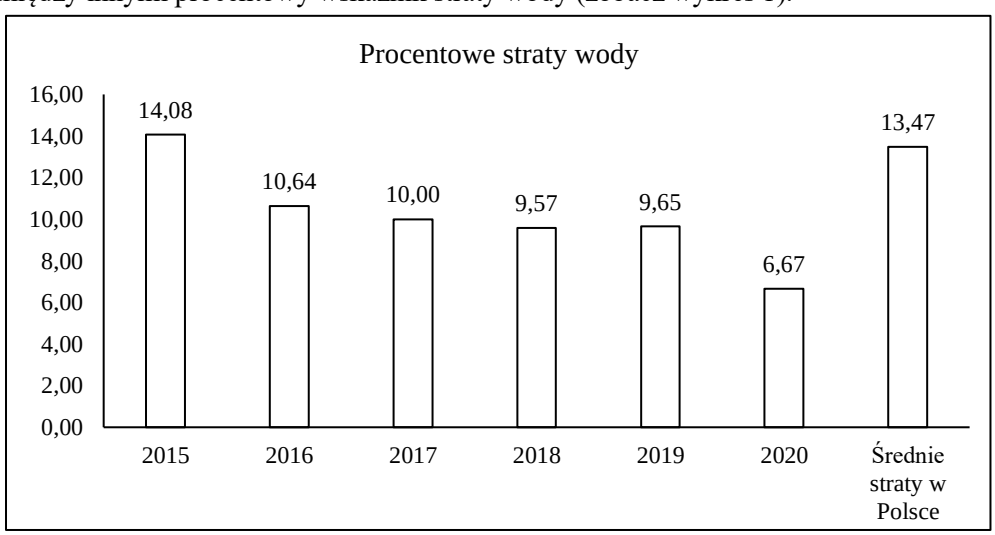
Analiza wybranych informacji ekonomicznych i innowacyjnych Śremskich Wodociągów Spółka z o.o. w świetle przeprowadzonych badań.

W tej części opracowania w ramach przyjętych założeń badawczych, przeprowadzona zostanie analiza kluczowych informacji, które umożliwią diagnozę prowadzonej działalności gospodarczej przedsiębiorstwa Śremskie Wodociągi pod względem ekonomicznych oraz ekologicznym ze szczególnym zwróceniem uwagi na identyfikacje aspektów innowacyjnych.

Realizacja czynności diagnostyczno-badawczych obejmować będzie następujące zagadnienia:

- analiza działalności wodociągowej (procentowe straty wody) obejmujący okres 2015 do 2020 r., działalność inwestycyjna (2010 do 2022 r.);
- analiza wybranych informacji ekonomicznych takich, jak: przychody, koszty, wynik finansowy obejmujący okres 2018 do 2021 r.;
- analiza kosztów rodzajowych, kosztów jednostkowych oczyszczania ścieków, obejmujący okres 2014 do 2020 r.;
- produkcja biogazu oraz polepszacza glebowego.

W obszarze działalności Śremskich Wodociągów funkcjonuje sześć podziemnych ujęć wody, na których znajdują się dwadzieścia trzy studnie głębinowe zaopatrujące będące w zasobach Śremskich Wodociągów pięć stacji uzdatniania wody: Śrem, Nochowo, Gaj, Dąbrowa, Orkowo. Spółka użytkuje także, na podstawie umowy dzierżawy, stację uzdatniania wody w Piotrowie (gmina Brodnica). Obiekty te produkują łącznie w ciągu roku około 2 800 000 m^3 wody o jakości zgodnej z obowiązującymi przepisami sanitarnymi. Efektywność sytemu dostarczania wody dla mieszkańców w/w gmin wyraża między innymi procentowy wskaźnik straty wody (zobacz wykres 1).

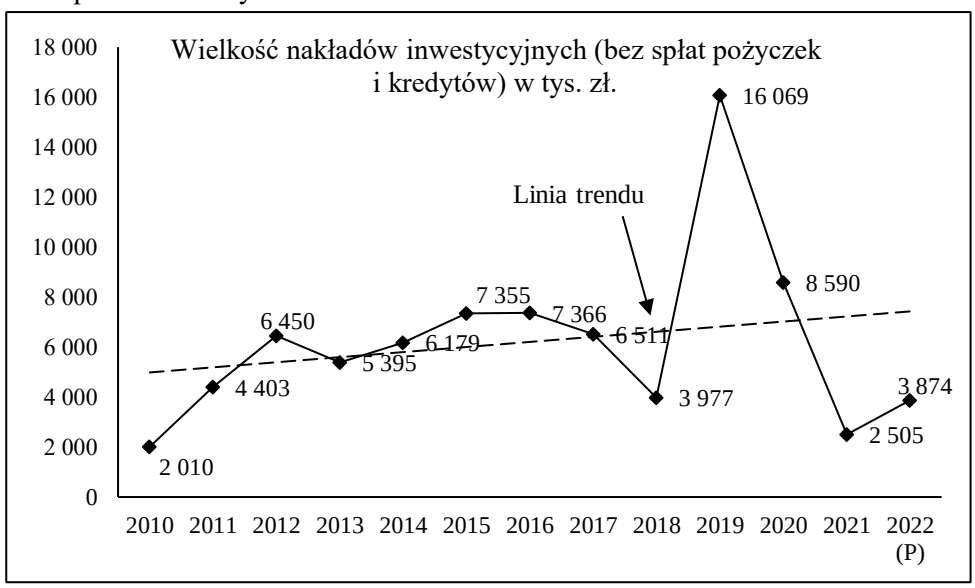


Wykres 1. Procentowe straty wody w zakresie działalności wodociągowej Śremskich Wodociągów w okresie 2015-2020 roku. Źródło: [opracowanie własne]

W okresie od 2016 roku do 2020 roku widać zmniejszające się straty wody, co świadczy o sprawności systemu wodociągowego i jest to rezultat podejmowanych w tym czasie inwestycji w sieć wodociągową. Odnosząc prezentowane na wykresie, procentowe straty

wody do średniej straty w Polsce (13,47%), można stwierdzić, że wskaźnik ten wypada bardzo korzystnie dla Śremskich Wodociągów. Przykładowo w roku 2020 ww. wskaźnik ten wyniósł 6,67%, co oznacza, że w stosunku do średniej krajowej był niższy o 6,68% od średniej krajowej, czyli niższy 58,48%.

Skalę podejmowanych inwestycji przez Śremskie Wodociągi, wyrażonych wartościowo przedstawia wykres 2.



Wykres 2. Wielkość nakładów inwestycyjnych poniesionych przez Śremskie Wodociągi (bez spłat pożyczek i kredytów) w okresie 2010-2022 roku. Źródło: [opracowanie własne]

Łącznie w okresie od 2010 do 2022 roku, Śremskie Wodociągi zainwestowały ponad 82 miliony złotych (82 373 000 zł). Największe nakłady na inwestycje spółka poczyniła w 2019 roku (16 069 tysięcy złotych). Średnio rok rocznie spółka przeznaczał na inwestycje 6 206 tysiąca złotych (najmniej w 2010 roku: 2 010 tysięcy złotych, najwięcej w 2019 roku: 16 069 tysięcy złotych. Z analizy linii trendu (nachylenie dodatnie) wynika systematyczny wzrost nakładów inwestycyjnych z okresami zmian o charakterze rosnącym i malejącym, jako efekt realizacji poszczególnych ale zakończonych etapów.

Zestawienie oraz analiza wybranych informacji ekonomicznych (przychody, koszty, wynik finansowy) obejmujący okres 2018 do 2021 roku (tab. 1).

Tabela 1. Zestawienie przychodów oraz kosztów, WF, inne wskaźniki – Śremskie Wodociągi sp. z o.o. (2018-2021)

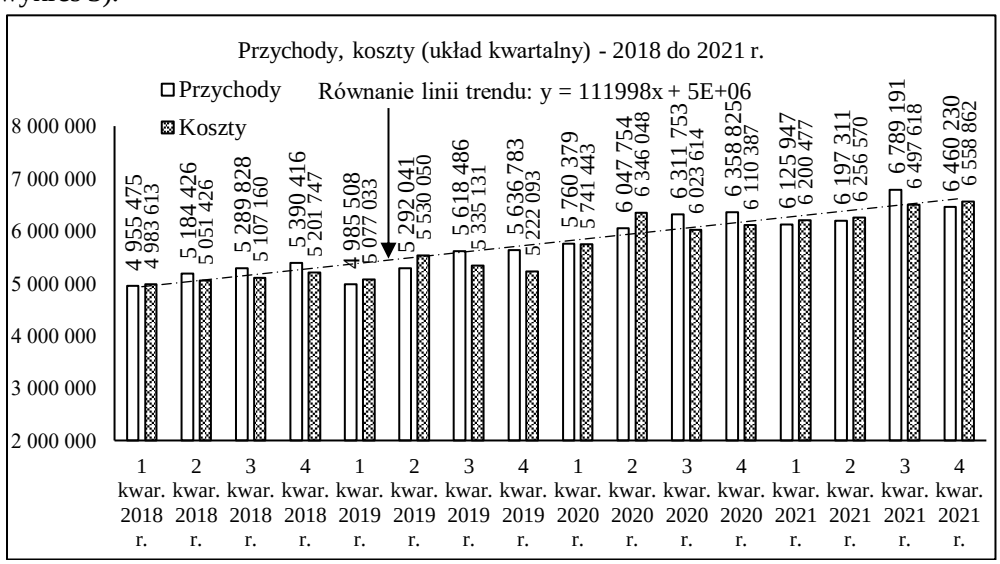
Rok	Wyszczególnienie	I kwartał	II kwartał	III kwartał	IV kwartał	Razem
2018	Przychody	4 955 475	5 184 426	5 289 828	5 390 416	20 820 145
	Koszty	4 983 613	5 051 426	5 107 160	5 201 747	20 343 947
	WF	-28 138	133 000	182 668	188 669	476 199
	Wskaźnik koszt. ogółemem	101	97	97	96	98
	Wskaźnik GMP	-0,57	2,57	3,45	3,50	2,29

2019	Przychody	4 985 508	5 292 041	5 618 486	5 636 783	21 532 818
	Koszty	5 077 033	5 530 050	5 335 131	5 222 093	21 164 308
	WF	-91 525	-238 010	283 355	414 690	368 510
	Wskaźnik koszt. ogółem	102	104	95	93	98
	Wskaźnik GMP	-1,84	-4,50	5,04	7,36	1,71
2020	Przychody	5 760 379	6 047 754	6 311 753	6 358 825	24 478 711
	Koszty	5 741 443	6 346 048	6 023 614	6 110 387	24 221 492
	WF	18 936	-298 294	288 139	248 439	257 219
	Wskaźnik koszt. ogółem	100	105	95	96	99
	Wskaźnik GMP	0,33	-4,93	4,57	3,91	1,05
2021	Przychody	6 125 947	6 197 311	6 789 191	6 460 230	25 572 679
	Koszty	6 200 477	6 256 570	6 497 618	6 558 862	25 513 528
	WF	-74 530	-59 260	291 573	-98 632	59 151
	Wskaźnik koszt. ogółem	101	101	96	102	100
	Wskaźnik GMP	-1,22	-0,96	4,29	-1,53	0,23

Źródło: [opracowanie wasne na podstawie dokumentów finansowych].

Interpretację graficzną wybranych danych, zawartych w tabeli 1, przedstawiają wykresy 3-5.

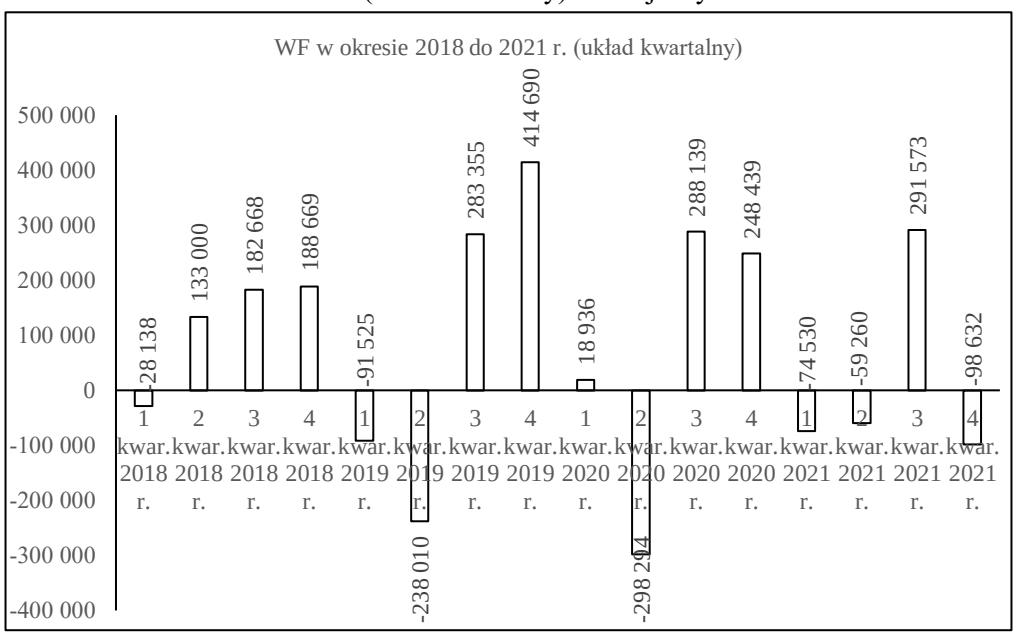
Wielkość przychodów, kosztów w latach 2018 do 2021 w układzie kwartalnym (zob. wykres 3).



Wykres 3. Zmiany wielkość przychodów, kosztów w latach 2018 do 2021 w układzie kwartalnym.
Źródło: tabela 1

W przypadku przychodów oraz kosztów widoczna jest tendencja rosnąca, co potwierdza równanie oraz nachylenie linii trendu. Rozstęp między skrajnymi wielkościami (przychody) wynosi: 1 504 755 zł. Rozstęp między skrajnymi wielkościami (koszty) wynosi: 1 575 249 zł. Ostatecznie koszty były wyższe od przychodów o 70 494 zł. Łącznie w prezentowanym okresie czasu (2018 do 2021 roku) przychody były wyższe od kosztów o 1 161 081 zł.

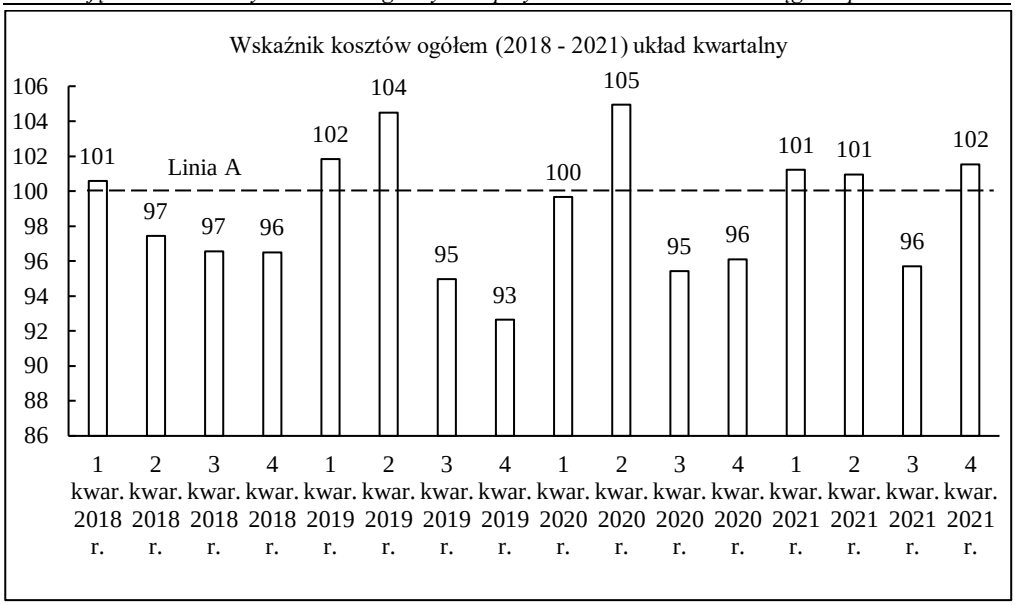
Opłacalności aktywności gospodarczej spółki(wynik finansowy oraz wskaźnik kosztów ogółem) w okresie 2018 do 2021 roku przedstawia wykres 4 oraz 5. Zmiany WF w okresie 2018 do 2021 r. (układ kwartalny) ilustruje wykres 4.



Wykres 4. Zmiany wielkość WF w latach 2018 do 2021 w układzie kwartalnym. Źródło: tabela 1

W okresie prowadzonych badań spółka generowała zarówno dodatni, jak i ujemny wynik finansowy. Najwyższy WF odnotowano w 4 kwartale 2019 roku (414 690 zł), natomiast najniższy w 2 kwartale 2020 roku (-298 294 zł). Łącznie od 2018 do 2021 roku spółka wygenerował dodatni wynik finansowy (1 161 080 zł). Oznacza to również, że średnio w przytoczonym okresie czasu spółka uzyskała WF na poziomie 232 116 zł. Dodać trzeba, że w związku z prowadzoną działalnością, łącznie spółka zrealizowała przychody na poziomie 92 404 353 zł oraz poniosła koszty: 91 243 274 zł.

Zmiany wskaźnika kosztów ogółem w okresie 2018 do 2021 roku (układ kwartalny) przedstawia wykres 5.



Wykres 5. Zmiany wielkość wskaźnika kosztów ogółem w latach 2018 do 2021 w układzie kwartalnym.
Źródło: tabela 1

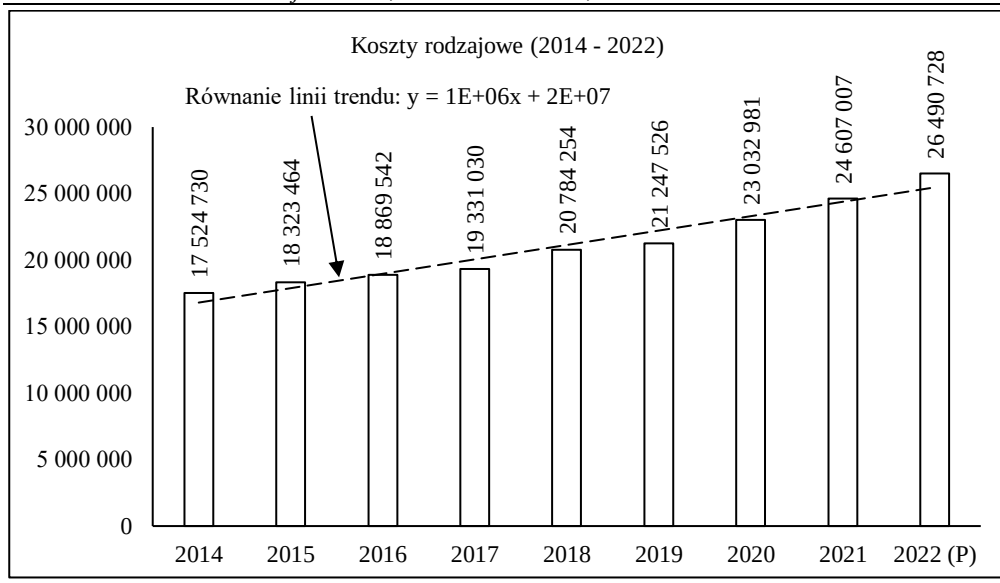
Założenie: wartości wskaźnika kosztów ogółem winien być niższy niż 100% w relacji kosztów do przychodów. Biorąc pod uwagę wartości wskaźnika powyżej 100%, można stwierdzić, że w 6 kwartałach koszty były wyższe od przychodów, co oznacza ujemny wynik finansowy. W 1 kwartale 2020 r. koszty były równe przychodom, oznacza to, że WF = 0. W 8 kwartałach wartość wskaźnika była niższa niż 100%, czyli firma w tych okresach osiągała dodatni WF. Najniższy wskaźników kosztów ogółem można odnotować w 4 kwartale 2019 r.: 93%. Pożądany poziom wskaźnika winien się lokować zdecydowane poniżej 90%.

Na podstawie powyższych interpretacji graficznych (wykresy oraz opisowych) widać, że tego rodzaju aktywność gospodarcza jest niezwykle kosztochłonna i obarczona stosunkowo wysokim ryzykiem. Tym co szczególnie ogranicza, sferę dochodową przedsiębiorstw są stosowane regulacje prawne dotyczące wysokości cen i stawek obowiązujące w danym roku taryfowym. W uzupełnieniu podkreślić jednak należy, że analizowana firma ma charakter samorządowy, której właścicielem są mieszkańcy danej gminy i byłaby nawet dopuszczalna działalność na poziomie koszty są równe przychodom, a w sytuacjach wyjątkowych z uwzględnieniem straty.

Analiza kosztów rodzajowych, kosztów jednostkowych oczyszczania ścieków, obejmujący okres 2014 do 2022 roku.

a) analiza kosztów rodzajowych;

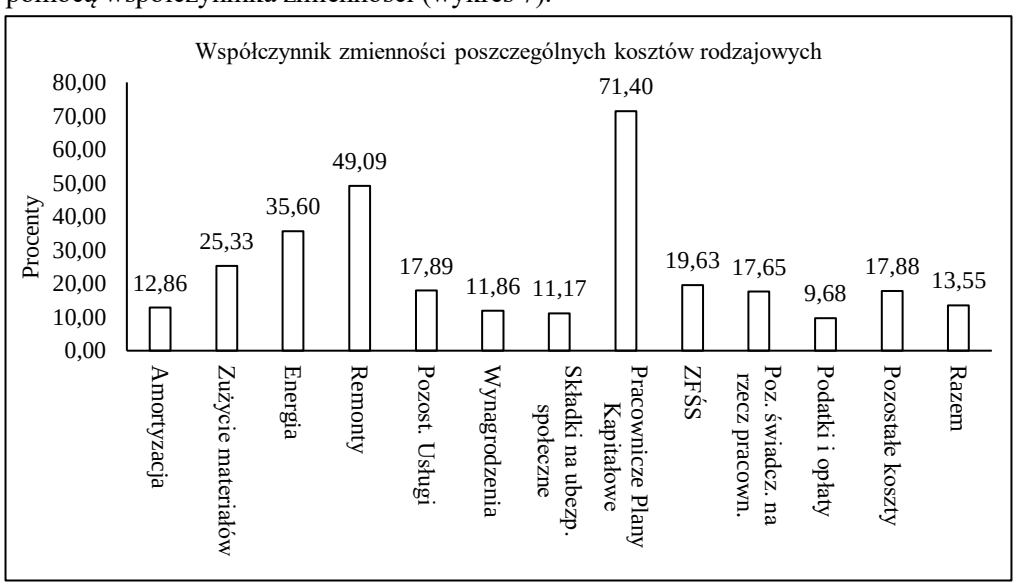
Łączne koszty rodzajowe w okresie 2014 do 2022 roku przedstawia wykres 6.



Wykres 6. Zmiany wielkość kosztów rodzajowych w latach 2014 do 2022. Źródło: opracowanie własne na podstawie danych uzupełniających spółki

Koszty rodzajowe w analizowanym okresie czasu wykazują tendencję rosnącą. Rozstęp kosztów rodzajowych między 2022 a 2014 roku wynosi: 8 965 998. Łącznie koszty rodzajowe w okresie 2014 do 2022 roku wyniosły: 190 211 262 zł, co oznacza, że jest to działalność bardzo kosztochłonna.

Biorąc pod uwagę widoczne zmiany wysokości kosztów rodzajowych w poszczególnych latach tj. od 2014 do 2022 roku, to istotne jest określenie ich zmienności, za pomocą współczynnika zmienności (wykres 7).



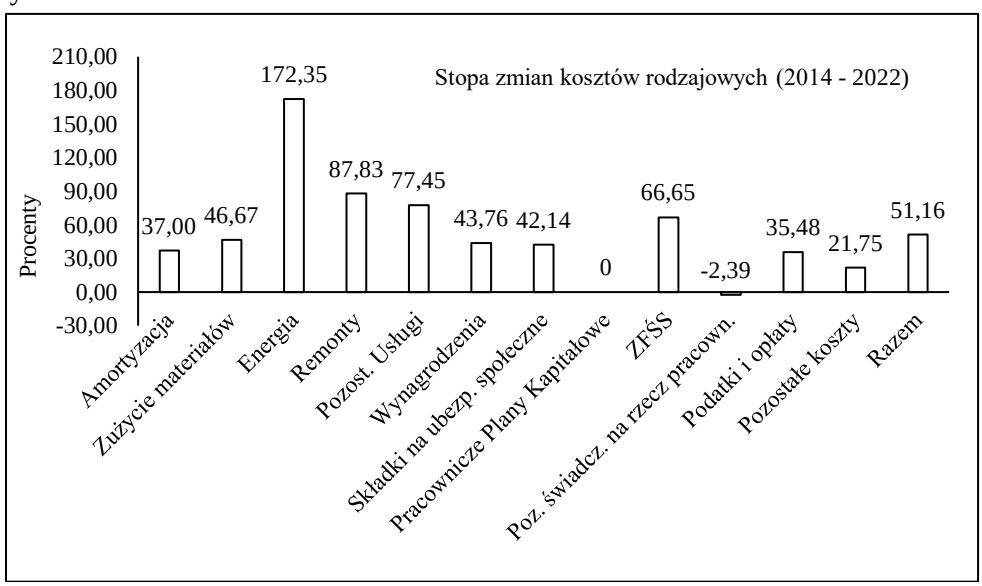
Wykres 7. Współczynnik zmienności kosztów rodzajowych. Źródło: [opracowanie własne]

Interpretacja współczynnika zmienności:

- do 25 % – mała zmienność;
- od 25% do 45% – przeciętna zmienność;
- od 45% do 100% – silna zmienność;
- powyżej 100%- bardzo silna zmienność.

Najwyższą wartość współczynnika zmienności można odnotować w przypadku kosztu „Pracowniczych planów kapitałowych": 71,40% – silna zmienność. Remonty – współczynnik zmienności: 49,09 – silna zmienność. Energia – współczynnik zmienności: 35,60 – przeciętna zmienność. Pozostałe koszty wskazane na wykresie 7 mieszczą się w przedziale małej zmienności. Średnio dla kosztów rodzajowych zawiera się w przedziale malej zmienności: 13,55%.

W kolejnej fazie oceny dynamiki kosztów rodzajowych, stanowiło określenie o ile procent przeciętnie wartości badanej cechy zmieniały się z okresu na okres (przyrost przeciętny) w tym przypadku od 2014 do 2022 r. W tym celu obliczono zarówno średnie tempo zmian oraz stopę zmian. Rozkład stopy zmian w ujęciu graficznym przedstawia wykres 8.



Wykres 8. Stopa zmian kosztów rodzajowych. Źródło:opracowanie własne

Z wykresu 8 wynika, że przeciętnie koszt energii w analizowanym okresie czasu (2014 do 2022 roku) zmieniał się przeciętnie 172,35%. W przypadku remontów stopa zmian wyniosła 87,83%, pozostałe usługi 77,45%. Stopa zmian w odniesieniu do zestawionych kosztów rodzajowych wyniosła 51,16%. Z analizy wskaźnika stopy zmian wynika, że koszty rodzajowe w przedmiotowym okresie czasu wykazywały bardzo znaczące wzrosty, które doprecyzują wnioski zawarte w ramach wykresu 6. W praktyce oznacza to, że analizowana spółka nie stosuje kontrolingu kosztów, stąd takie znaczące odchylenia. Ponadto podkreślić trzeba, że istnieje możliwość wykorzystania owego miernika w prognozowaniu zmian, np. kosztów, WF oraz innych parametrów ekonomicznych chociażby przy tworzeniu planów rzeczowo finansowych.

b) analiza kosztów jednostkowych oczyszczania ścieków.

Wymagania ekologiczne: ścieki dopływające są oczyszczone w procesie technologicznym w stopniu pozwalającym na ich zrzut do odbiornika (rzeka Warta), tj. gdy ich parametry spełniają wymagania określone w pozwoleniu wodnoprawnym, tj.

- wskaźnik określający biochemiczne zapotrzebowanie tlenu (BZT5): 15 g O_2/m^3;
- wskaźnik określający chemiczne zapotrzebowanie tlenu (ChZT): 125 g O_2/m^3;
- zawiesina ogólna: 35 g/m^3;
- azot ogólny: 15 g N/m^3;
- fosfor ogólny: 2 g P/m^3.

Technologia oczyszczania ścieków (ujęcie syntetyczne) – etapy:

Etap 1: oddzielenie stałych zanieczyszczeń obecnych w ściekach za pomocą krat, ich higienizacja, wywóz na gminne składowisko odpadów komunalnych,

Etap 2: usuwanie piasku, wywóz na gminne składowisko odpadów komunalnych,

Etap 3: oczyszczanie wstępne w osadnikach, przepompowywanie do zagęszczaczo-fermentera, transport razem z osadem nadmiernym z zagęszczaczy osadu do komór fermentacyjnych,

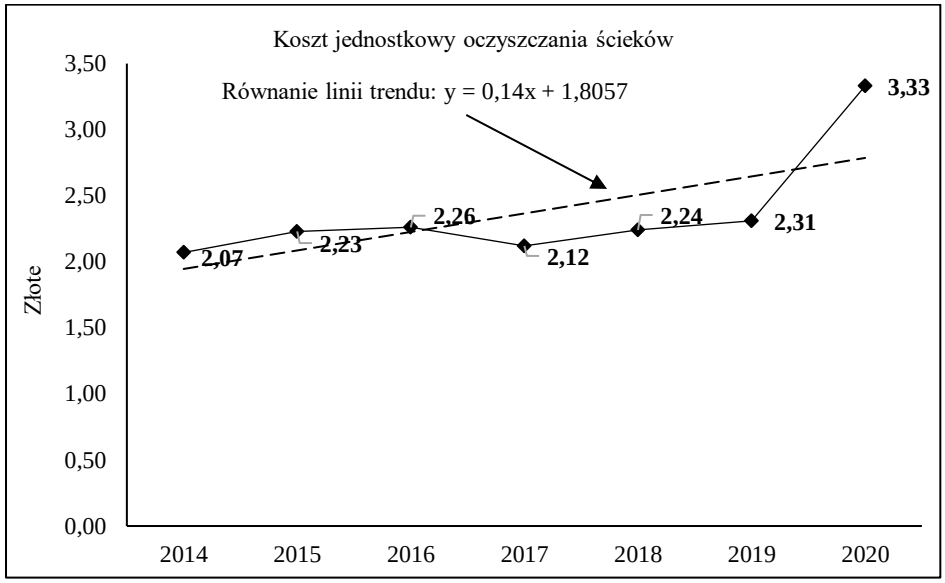
Etap 4: oczyszczanie ścieków w reaktorach biologicznych, które prowadzone jest w trójstopniowym systemie Bardenpho (tworzą go komora defosfatacji, denitryfikacji i nitryfikacji), W procesie sterowania reaktorami biologicznymi wykorzystuje się sondy tlenowe, potencjału redox, azotanów i amoniaku, umożliwiające łatwiejsze sterowanie procesem:

- właściwego utrzymywania tlenu (sterowanie dmuchawami – napowietrzanie) w komorze nitryfikacji reaktora;
- właściwej redukcji amoniaku;
- usuwania azotanów przez kontrolowane zwiększanie recyrkulacji wewnętrznej.

Etap 5: usuwania fosforu, odgazowanie, sklarowanie, skierowanie ścieków przelewami pilastymi przez komorę pomiarową do rzeki Warty,

Etap 6: laboratorium chemiczne – badanie jakości ścieków, bieżąca kontrola poprawności działania oczyszczalni ścieków.

Koszt jednostkowy oczyszczalni ścieków w latach 2014 do 2020 roku przedstawia wykres 9.



Wykres 9. Zmiany kosztów jednostkowych oczyszczania ścieków w latach 2014 do 2020 roku.
Źródło: opracowanie własne

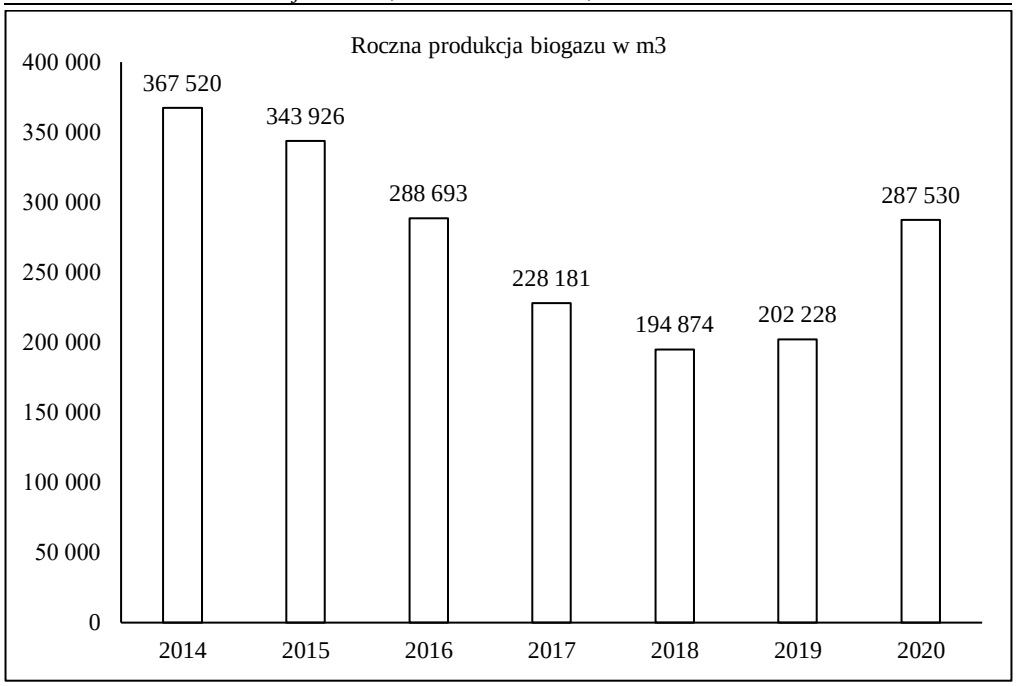
Koszty jednostkowe w latach 2014 do 2020 r. wzrosły z poziomu 2,07 zł (2014 rok) do 3,33 zł (2020 rok). Rozstęp kosztów jednostkowych (2020 do 2014 r.), oczyszczania ścieków we wspomnianym okresie czasu, wynosi 1,26 zł, co oznacza ich wzrost o 160%. Z nachylenia linii trendu wynika, że w prognozie na kolejne lata należy się liczyć z radykalnym wzrostem owych kosztów.

Produkcja biogazu oraz polepszacza glebowego w aspekcie ekologicznym oraz kosztowym.

Analiza produkcji biogazu: technologia, wielkość produkcji w okresie 2014-2020 roku.

Biogaz powstaje w procesie fermentacji metanowej biomasy. Pod względem chemicznym jest mieszaniną głównie metanu, tlenku węgla, siarkowodoru oraz wody w postaci pary wodnej. Biogaz gromadzony jest w zbiorniku, po czym kierowany jest do agregatu. Powstające w agregacie ciepło wykorzystywane jest do utrzymywania wymaganej temperatury fermentującego osadu. Wyprodukowany w czasie fermentacji biogaz jest wykorzystywany w produkcji energii elektrycznej i energii cieplnej jako paliwo gazowe w agregacie kogeneracyjnym o mocy 123 k.w. Ciepło powstałe w procesie jest wykorzystywane do ogrzewania osadu w komorach fermentacyjnych a w okresie zimowym także do ogrzewania budynków technologicznych oczyszczalni ścieków.

Wielkość produkcji biogazu w okresie 2014-2020 roku przedstawia wykres 10.



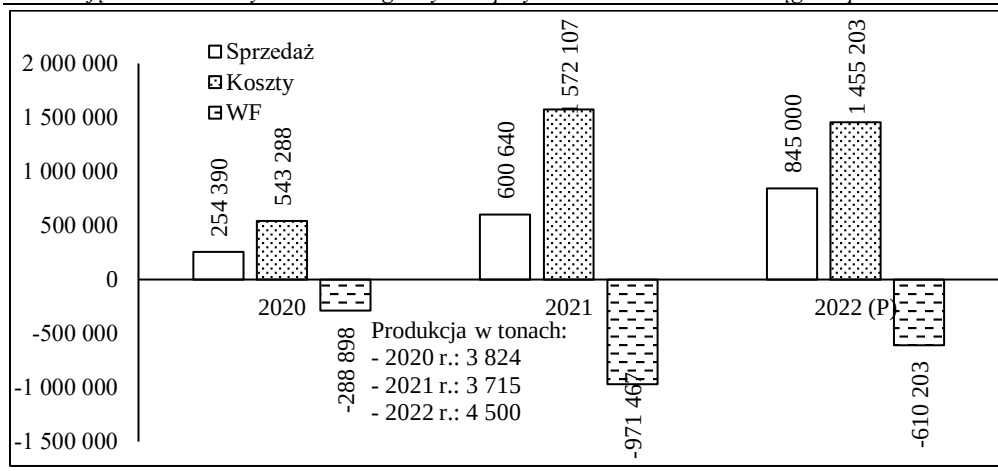
Wykres 10. Produkcji biogazu w okresie 2014-2020 roku. Źródło: dokumenty sprawozdawcze firmy Śremskie Wodociągi Spółka z o.o.

W okresie 2014 do 2020 roku Śremskie Wodociągi wyprodukowały 1 912 952 m^3 biogazu. Od 2014 do 2018 roku produkcja biogazu systematycznie się zmniejszała: spadek o 172 646 m^3 (redukcja o 189%). W okresie 2019 do 2020 roku produkcja biogazu wzrosła o 85 302 m^3. Tendencja wzrostowa produkcji biogazu po 2018 roku jest wyrazem rosnącej roli i znaczenia biogazu, który może być wykorzystany między innymi, jako: paliwo, do ogrzewania, także do gotowania oraz w generatorach prądu. Zaznaczyć trzeba, że w ciągu 2021 roku w zarejestrowanych instalacjach wytworzono ponad 513 mln m^3 biogazu. Prognozuje się, że kolejnych latach, biogaz może zapewnić ok 30% zapotrzebowania na moc. Jak widać gwałtownie rośnie rola i znaczenie OZE.

Produkcja polepszacza glebowego – środek poprawiający właściwości gleby.

W procesie modernizacji oczyszczalni ścieków, skonstruowana została instalacja przetwarzania osadów, dzięki której uzyskano produkt o nazwie handlowej OrCal®, który może być wykorzystywany do: odkwaszania gleby, poprawa jej żyzności, odbudowy w glebie materii organicznej zniszczonej stosowaniem nawozów sztucznych, wzmacnia odporność roślin uprawnych na czynniki chorobotwórcze i stres, np. susze, upały, burze, gradobicia, podtopienia, przymrozki i mrozy, przyduchy śnieżne oraz skutki uboczne stosowania środków ochrony roślin. Dodać należy, ze nawóz pod nazwą OrCal®, decyzją Ministra Rolnictwa i Rozwoju Wsi nr G-936/20 z dnia 26 maja 2020 roku, uzyskał certyfikat (pozwolenie) na podstawie, którego Śremskie Wodociągi mogą polepszacz glebowy wprowadzić do obrotu.

Przychody, koszty wytwarzania oraz wielkość produkcji polepszacza glebowego (OrCal®) przedstawia wykres 11.



Wykres 11. Sprzedaż, koszty produkcji polepszacza glebowego, WF, w okresie 2020-2022.
Źródło: opracowanie własne

W okresie 2020 do 2022 roku łącznie Śremskie Wodociągi wyprodukowały 12 039 ton polepszacza glebowego. Produkcja ww. nawozu na podstawie (ujemny WF) jest kosztowna i niedochodowa. Podkreślić jednak trzeba, że wartością dodaną stanowi fakt, że nawóz ten został wyprodukowany w wyniku przetworzenia osadów, których pełna utylizacja zdecydowanie podniosłaby koszt oczyszczania ścieków (aspekt kosztów alternatywnych).

Pomimo ujemnego WF, należy w tym przypadku uwzględnić tzw. wartość dodaną, wynikającą z bardzo wysokich kosztów utylizacji, które przekraczają koszt produkcji polepszacza glebowego (OrCal®). Można przyjąć, że produkcja biogazu, jak i produkcja w/w nawozu przez Śremskie Wodociągi jest wymiernym przykładem, innowacyjnych rozwiązań w sferze ochrony środowiska naturalnego, OZE. Dodać trzeba, że w dobie kryzysu energetycznego – szczególnie w przedsiębiorstwach, których właścicielem lub udziałowcem jest j.s.t. – należy poszukiwać odmiennych, innowacyjnych sposobów obniżenia kosztów eksploatacyjnych. Dodatkowo można stwierdzić, że obecny kryzys stanowi bardzo ważną inspirację do podejmować różnych decyzji o charakterze innowacyjnym, które z czasem będą przynosić efekty ekonomiczne oraz te dalej idące ekologiczne.

6. Podsumowanie

W niniejszym opracowaniu podjęta została próba identyfikacji i wykorzystania innowacyjno-rozwojowych uwarunkowań działalności spółki wodociągowej w ujęciu ekonomicznym oraz ekologicznym na przykładzie Śremskich Wodociągów Spółka z o.o.

Przeprowadzone rozważania w ramach podjętej problematyki pozwalają na sformułowanie wniosków końcowych.

Wnioski z przeprowadzonych badań:

1. Władze samorządowe muszą samodzielnie podjąć decyzje, w jaki sposób zorganizować rynek usług, by najlepiej zaspokoić potrzeby mieszkańców w aspekcie dopuszczalnych sposobów organizacyjno-prawnych z zastosowaniem innowacyjno-rozwojowych uwarunkowań gospodarki komunalnej przedsiębiorstw samorządowych.

2. Uzyskane wyniki z przeprowadzonych badań na przykładzie Śremskich Wodociągów Spółka z o.o. w pełni potwierdziły założoną hipotezę główną, że działalność innowacyjna samorządowych spółek komunalnych zależy od czynników pochodzących z makro i mikrootoczenia oraz korzystania z wiedzy i doświadczeń, które wymuszają sytuacje kryzysowe w zakresie innowacyjnego podejścia, prowadzenia działań trudnych wynikających z licznych zmian w otoczeniu tworzących szanse, a także zagrożenie dla ich rozwoju.
3. W ramach przyjętych założeń badawczych, przeprowadzona zostanie analiza kluczowych informacji pozwalających na sformułowanie diagnozy prowadzonej działalności gospodarczej przedsiębiorstwa Śremskie Wodociągi pod względem ekonomicznym oraz ekologicznym ze szczególnym zwróceniem uwagi na identyfikacje aspektów innowacyjno-rozwojowych. Analiza wybranych informacji ekonomicznych w wyniku przeprowadzonych badań takich, jak: przychody, koszty, wynik finansowy obejmujący okres 2018 do 2021 r. zostały przedstawia w formie tabel (tab. 1 i 2) i wykresów (1-11) oraz szczegółowych wniosków w toku przeprowadzonych analiz ilościowo-jakościowych, które pozwoliły na następujące podsumowanie wyników z przeprowadzonych badań dotyczących takich zagadnień jak:
    a. Analiza wybranych informacji ekonomicznych w zakresie kształtowania się w analizowanym okresie procentowego wskaźnika straty wody w działalności wodociągowej, który kształtował się średnio na poziomie 10,2% w porównaniu do średniego wskaźnika w Polsce (13,47%). Z kolei przykładowo w roku 2020 roku wskaźnik ten wyniósł 6,67%, co oznacza, że w stosunku do średniej krajowej był niższy o 6,68% od średniej krajowej, czyli niższy 58,48%.
    b. Wielkość nakładów inwestycyjnych w badanym okresie miał charakter rosnący i malejący z wyraźną linią trendu (nachylenie dodatnie) potwierdzającą systematyczny wzrost nakładów inwestycyjnych w ramach zakończonych etapów (wykres 2).
    c. Przychody i koszty w badanym okresie czasowym wykazywały tendencje rosnącą (tab. 2). Natomiast zmiany wielkości wyniku finansowego wykazywały tendencje dodatnio/ujemną, a łącznie dodatni wynik finansowy kształtował się na poziomie (1 161 080 zł), przy uzyskanym przychodzie: 82 404 353 zł I poniesionych kosztach: 91 243 274 zł.
    d. Wskaźnik zmian kosztów rodzajowych w analizowanym okresie charakteryzował się dużo zmiennością kosztów wynikającą między innymi z faktu, że badana spółka nie wykorzystywała dostatecznie rozwiązań kontrolingowych, umożliwiających monitorowanie i kształtowania poziomu kosztów.
    e. Gospodarowanie ściekami w aspekcie wymagań ekologicznych ścieków dopływających i oczyszczonych w procesie technologicznym są realizowane przez spółkę zgodnie ze wskaźnikami określonymi w koncesji wodnoprawnej oraz umożliwiającej na ich zrzut do odbiornika (rzeki Warty) z zachowaniem technologicznych etapów oczyszczania ścieków. Natomiast koszty jednostkowe tego procesu wykazywały trend wzrostowy. Na podstawie nachylenia linii trendu można prognozować na kolejne lata radykalny wzrost kosztów jednostkowych (wykres 9).
    f. W ramach wdrożonych działań innowacyjno-rozwojowych – wielkość produkcji biogazu w realizowanym po 2018 roku przedstawia tendencje wzrostową

z wykorzystaniem jako paliwa do ogrzewania, gotowania oraz wytwarzania energii elektrycznej. W okresie trwającego kryzysu energetycznego zauważalny jest wzrost roli i znaczenia biogazu, czego wyrazem jest 30% pokrycie ogólnego zapotrzebowania na moc (OZE).

Innym potwierdzeniem innowacyjnych działań badanej spółki jest wielkość wdrożonej produkcji polepszacza glebowego rozpatrywana w kategoriach: przychodów, kosztów i poziomu sprzedaży (wykres 11). Działania spółki w tym zakresie jest dobrym przykładem innowacyjnych działań związanych z ochroną środowiska naturalnego (OZE) i godnym naśladowania przez wszystkie j.s.t. w Polsce oraz należących do dobrych praktyk w poszukiwaniu odmiennych innowacyjnych poszukiwań związanych z planowaną w niedalekiej przyszłości cyfryzacją usług i ich cyfrowej dostępności, a także rozliczeń.

Przeprowadzone badania i analizy uzyskanych wyników w ramach podjętej problematyki zostały ograniczona do Śremskich Wodociągów Spółka z o.o. w Województwie Wielkopolskim, co niestety osłabia uniwersalność wniosków płynących z dokonanych badań. Korzystne i wartościowe byłaby zatem przeprowadzenie podobnych badań w innych samorządowych spółkach kapitałowych o tym samym profilu w innych regionach Polski, ponieważ dałoby to możliwość cennych porównań. Niemniej jednak zaobserwowane i opracowane zjawiska mogą stanowić asumpt do szerszej obserwacji i dalszych poszukiwania innych zmiennych, które mogą mieć wpływ na działalność innowacyjną przedsiębiorstw samorządowych.

Literatura

1. Drucker P.F., *Zarządzanie w czasach burzliwych*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 1995, s. 15-47.
2. *Słownik języka polskiego, tom pierwszy*, Warszawa 1978, s. 792.
3. Schumpeter J.A., *Teoria rozwoju gospodarczego*, Warszawa 1960, s. 104.
4. Pietrasiński Z., *Ogólne i psychologiczne zagadnienia innowacji*, Warszawa 1970, s. 16.
5. Niedzielski P., *Rodzaje innowacji*, [w:] Matusiak K.B. (red.), *Innowacje i transfer technologii – Słownik pojęć*, PARP, Warszawa 2005, s. 75.
6. Mazur-Wierzbicka E., *Działalność innowacyjna przedsiębiorstwa w Polsce*, Wydawnictwo Ekonomiczne Uniwersytetu Szczecińskiego, Tarnów, 2015, s. 99.
7. Szczepańska-Woszczyna K., *Kompetencje menedżerskie w kontekście innowacyjności przedsiębiorstwa*. PWN, Druk. Warszawa 2021, s. 43.
8. Sopińska A., Mierzejewska W., *Zasobowe uwarunkowania otwartych innowacji*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 444/2016, Wrocław 2016, s. 446.
9. Janasz W., *Innowacje i ich miejsce w działalności przedsiębiorstw*, [w:] Janasz W. (red.): *Innowacje w modelach działalności przedsiębiorstw*, Uniwersytet Szczeciński, Rozprawy i Studia DXX, Szczecin 2003, s. 49.
10. Wziątek-Kubiak A., Balcerowicz E., *Determinanty rozwoju innowacyjności firmy w kontekście poziomu wykształcenia pracowników*, Ekspertyza przygotowana dla PARP, Wydawnictwo CASE, Warszawa 2009, s. 13.
11. Grudzewski W.M., Hejduk I.K., *Zarządzanie wiedzą w przedsiębiorstwie*, Difin, Warszawa 2004, s. 16.
12. Kowalak B., *Konkurencyjna gospodarka. Innowacje – infrastruktura – mechanizmy rozwoju*, Państwowy Instytut Badawczy, Radom 2006, s. 17.
13. Baruk O.J., *Zarządzanie wiedzą i innowacjami*, A. Marszałek, Toruń 2006, s. 93.
14. Bouncken R.B., *Cultural diversity in innovation teams: surface and deep level effects*, International Journal of Business Research no. 4/2009, s. 17-26.

15. Pomykalski A., *Innowacyjność organizacji*, Wyd. Wyższej Szkoły Kupieckiej, Łódź 2009, s. 8.
16. Stawasz E., *Główne obszary sił motorycznych i napięć w systemie transferu i komercjalizacji technologii w Polsce,* [w:] *Uwarunkowania rozwoju przedsiębiorczości opartej na wiedzy*, Zeszyty naukowe nr 642, Ekonomiczne problemy usług nr 64, Uniwersytet Szczeciński, SOOIPP Annual, Szczecin 2011, s. 13.
17. Richard R., Katz R., *Managing Creativity and Innovation*, Harvard Business School Press, Boston 2003, s. 2.
18. Mruk H., Nestorowicz R., *Uwarunkowania sprawności innowacyjnej przedsiębiorstw*, Wyd. UE w Poznaniu, Poznań 2011, s. 63-171.
19. Brózda J., Marek S., *Otoczenie przedsiębiorstwa*, [w:] Marek S., Białasiewicz M. (red.), *Podstawy nauki o organizacji*, PWE, Warszawa 2008, s. 97.
20. Poznańska K., *Uwarunkowania innowacji w małych i średnich przedsiębiorstwach*, Dom Wydawniczy ABC, Warszawa 1998, s. 41.
21. https://poland.representation.ec.europa.eu/news/sprawiedliwa-i-zrownowazona-gospodarka-2022-02-23_pl. [data dostępu: 28.09.2022].

Innowacyjno-rozwojowe uwarunkowania działalności spółki wodociągowej w ujęciu ekonomicznym oraz ekologicznym na przykładzie Śremskich Wodociągów Spółka z o.o.

Streszczenie

W opracowaniu podjęta została próba przybliżenie innowacyjno-rozwojowych uwarunkowań działalności spółki wodociągowej w ujęciu ekonomicznym oraz ekologicznym na przykładzie Śremskich Wodociągów Spółka z o.o. Z praktyki funkcjonowania tego typu podmiotów gospodarczych wynika, że między funkcjami społecznymi i ekonomicznymi istnieje przestrzeń charakteryzująca się przeciwstawnymi wektorami opisującymi tendencje rozwojowe badanych zjawisk oraz czynników ekonomicznych, szczególnie na poziomie wyników finansowych. Uwzględniając powyższe założenie, cel, przedmiot badań przeprowadzono analizę innowacyjno-rozwojowych uwarunkowań w ujęciu ekonomicznym oraz ekologicznym w zakresie gospodarki ściekami zgodnie z wymaganiami ochrony środowiska naturalnego z zastosowaniem innowacyjnych technologii w produkcji biogazów i polepszacza glebowego. Jest to przykład dobrych praktyk w poszukiwaniu odmiennych i innowacyjnych sposobów obniżania kosztów eksploatacyjnych w samorządowej spółce kapitałowej w obliczu trwającego kryzysu energetycznego oraz planowanej w najbliższej przyszłości cyfryzacji usług i ich dostępności oraz cyfrowych rozliczeń.

Słowa kluczowe: spółka wodociągowa, gospodarka ściekami, innowacje, produkcja biogazów, ekologia

Innovative and developmental conditions of the water supply company's operations in economic and ecological terms on the example of Śrem Waterworks Spółka z o.o.

Abstract

The article attempts to approximate the innovative and developmental conditions of the operation of a water supply company in economic and ecological terms on the example of Śremskie Wodociągi Spółka z o.o. The practice of functioning of this type of economic entities shows that there is a space between social and economic functions characterized by opposing vectors describing the development trends of the studied phenomena and economic factors, especially at the level of financial results. Taking into account the above assumption, purpose and subject of research, an analysis of innovative and developmental conditions in economic and ecological terms was carried out in the field of wastewater management in accordance with the requirements of environmental protection, using innovative technologies in the production of biogas and soil improver. This is an example of good practice in the search for different and innovative ways to reduce operating costs in a local government company in the face of the ongoing energy crisis and the digitization of services and their availability and digital settlements planned for the near future.

Keywords: water supply company, sewage management, innovations, biogas production, ecology.

Agnieszka Kucharska-Jastrząbek[1], Paulina Dmowska-Jasek[2]

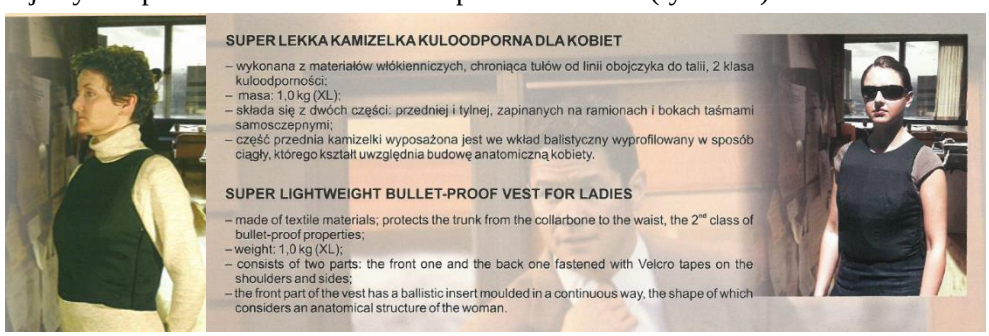
VESTA – damska kamizelka skrytego noszenia – od pomysłu do projektu

1. Wprowadzenie

Koncepcja stworzenia kamizelek balistycznych dla kobiet zrodziła się z dwóch kluczowych powodów.

Pierwszym powodem był fakt, iż obecnie w służbach mundurowych pełni służbę coraz więcej kobiet. W służbach mundurowych podlegających MSWiA służy około 21 000 kobiet, co stanowi około 15% całkowitej liczby funkcjonariuszy (w tym około 16 000 w Policji, około 4 000 w Straży Granicznej i około 200 w Służbie Ochrony Państwa). W Służbie Więziennej służy około 5 000 kobiet [1], w Siłach Zbrojnych RP blisko 6 000 kobiet [2], a w Wojskach Obrony Terytorialnej – blisko 450 żołnierzy to kobiety [3].

Drugi wynikał z doświadczenia Instytutu Technologii Bezpieczeństwa „MORATEX" w opracowywaniu nowoczesnych konstrukcji i technologii włókienniczych wyrobów technicznych oraz ich wdrażania do praktyki przemysłowej. Wśród tych wyrobów wyróżniały się osłony osobiste – kamizelki balistyczne, w tym również pierwsze na rynku krajowym superlekkie kamizelki kuloodporne dla kobiet (rys. 1 i 2).



Rysunek 1. Superlekka kamizelka kuloodporna dla kobiet ITWW „MORATEX" [foto: ITB „MORATEX"]

Produkcja pierwszych konstrukcji kamizelek kuloodpornych dedykowanych dla kobiet została zapoczątkowana w latach 90. XX wieku. Z uwagi na fakt, iż wykorzystywane rozwiązania indywidualizujące kształty osłon tułowia, bezpośrednio dla kobiet, związane były ze znaczną ilością zdefiniowanych kształtów/rozmiarów wyrobu, przez co nie było możliwe osiągnięcie ekonomiki produkcji oraz realnej logistyki zamówień.

[1] ajastrzabek@moratex.eu, Zakład Projektowania Systemów Informatycznych, Instytut Technologii Bezpieczeństwa „MORATEX", http://moratex.eu/.

[2] pjasek@moratex.eu, Zakład Projektowy, Instytut Technologii Bezpieczeństwa „MORATEX", http://moratex.eu/.

Rysunek 2. Kamizelka damska dla kobiet wzór MORATEX XXI (ITWW „MORATEX”)
[foto: ITB „MORATEX”]

1.1. Problematyka badawcza

Funkcjonariuszki służb mundurowych, na chwilę obecną korzystają z osłon tułowia dedykowanych i projektowanych na sylwetkę męską lub tzw. unisex, co odzwierciedla się w specyfikacjach technicznych wyrobów (załącznik do dokumentacji zamówień publicznych [4, 5]). Ze względu na brak idealnego dopasowania, kamizelki do kształtu ciała funkcjonariuszki, istnieje pewność, że nie spełniają one w pełni swoich funkcji ochronnych, zwłaszcza w obszarze pomiędzy biustem a podbrzuszem lub w okolicy pach (rys. 3 i 4).

Rysunek 3. Przykłady niewłaściwego wyprofilowania osłony tułowia do kobiecej sylwetki
[foto: ITB „MORATEX”]

Rysunek 4. Przykłady niewłaściwego dopasowania osłony tułowia do kobiecej sylwetki
[foto: ITB „MORATEX”]

Obszar ten jest newralgiczny z punktu widzenia ochrony narządów wewnętrznych, tj. wątroby, żołądka, nerek. Biorąc pod uwagę, że pod wpływem uderzenia penetratora (pocisku i/lub odłamka) w kamizelkę balistyczną następuje jej gwałtowna, chwilowa deformacja, która w konsekwencji może doprowadzić do tępego, podpancerzowego obrażenia ciała jej użytkownika, czego skutkiem może być ciężkie uszkodzenie ciała, a nawet śmierć, niezmiernie ważne jest dopasowanie kamizelki balistycznej do ciała użytkownika. Należy również dodać, że energia uderzenia pocisku/odłamka jest tylko częściowo absorbowana przez osłonę balistyczną a jej pozostała część jest przekazywana na ciało użytkownika, stąd niezwykle istotne jest, aby osłona była jak najdokładniej dopasowana do ciała użytkowania (zwłaszcza w przypadku kobiet) w celu zminimalizowania ryzyka wspomnianych obrażeń wewnętrznych (efekt BABT).

Jeśli osłona tułowia nie jest prawidłowo dopasowana do ciała (zarówno dla mężczyzn jak i kobiet) i nie zapewnia wymaganej powierzchni ochrony, to negatywnie wpływa na bezpieczeństwo i skuteczność. Możliwość adaptacji rozwiązań powstałych dla mężczyzn wydaje się z punktu bezpieczeństwa niemożliwa i nieefektywna stąd pojawiła się potrzeba projektowania i produkcji osłon balistycznych dla funkcjonariuszek [6].

Osłony ciała projektowane dla kobiet różnią się od męskich lub unifikowanych, głównie ze względu na konieczność wprowadzenia w konstrukcji krzywizn klatki piersiowej wraz z ich różnorodnością rozmiarową. Kamizelki balistyczne projektowane dla mężczyzn lub zunifikowane mogą krańcowo być dedykowane dla kobiet o małym rozmiarowo biuście – z pewnością nie zapewnią jednak wymaganej ergonomii. W zależności od zastosowanych materiałów oraz konstrukcji osłony te nie będą odpowiednie dla kobiet o średnim lub dużym biuście, ze względu na pojawiający się efekt wypchnięcia osłony do przodu, co powoduje zwiększenie luki powstałej pod pachami oraz redukcję powierzchni ochronnej (rys. 3 i 4). Powstałe lokalne uciski osłony na ciało kobiety wpływają (znacząco) negatywnie na dyskomfort, a w sytuacjach krańcowych (ucisk na nerw, tętnicę) powodować mogą dysfunkcję lub przewlekłe bóle.

Znane są rozwiązania indywidualizujące kształty osłon tułowia dedykowane bezpośrednio dla kobiet [7], jednakże, ze względu na fizjologie ciała kobiety, ilość zdefiniowanych kształtów/rozmiarów jest znaczna, przez co nie pozwala na osiągnięcie ekonomiki produkcji oraz realnej logistyki zamówień. Koszty wytworzenia takich osłon są 3-6 razy większe od standardowej, seryjnej produkcji. Konieczność personalizacji wyrobu sprowadza to rozwiązanie do produktu niszowego, ogranicza dostępność dla coraz rosnącej grupy użytkowniczek, a końcowo powoduje częściowe ich wyłączenie.

1.2. Cel projektu

Naprzeciw zdefiniowanemu problemowi pojawiła się koncepcja projektu „Indywidualizacja systemu osłony osobistej" akronim VESTA (POIR 04.01.04-00-0085/18). Projekt wyszedł naprzeciw opisanemu w rozdziale 1.1 problemowi dzięki idei unifikacji grup rozmiarowych poprzez system korygujący sylwetkę, co powinno wpłynąć pozytywnie zarówno na ekonomikę produkcji, jak i dostępność tego rozwiązania dla użytkowniczek, które na dzień dzisiejszy są ograniczane ergonomią oraz bezpieczeństwem stosowanego sprzętu ochronnego.

Biorąc pod uwagę analizę potrzeby rynkowej, jak i stale rosnący udział kobiet w służbach mundurowych [1] zrodziła się idea celowego zaprojektowania systemu Damskiej Wielofunkcyjnej Kamizelki Balistycznej skrytego noszenia (DWKB), przeznaczonej do

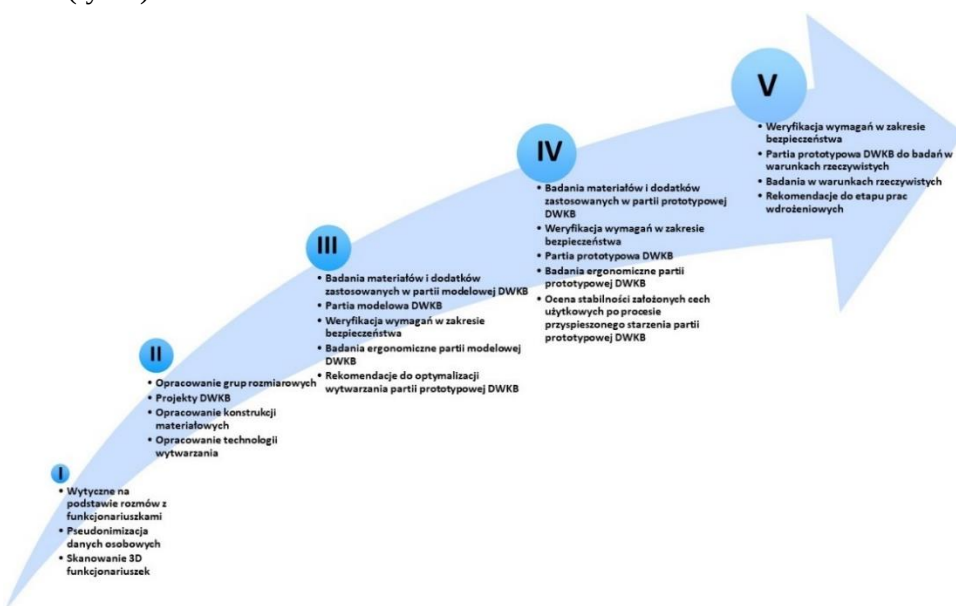
całorocznego użytkowania, a dedykowanej funkcjonariuszkom służb mundurowych. Założeniem niniejszego projektu było uzyskanie systemu o ograniczonej grupie rozmiarowej, co pozwoli na zwiększenie ekonomicznej i logistycznej dostępności oraz efektywnej możliwości realizacji procesu produkcyjnego. Założono, że opracowany w ramach projektu zindywidualizowany system DWKB będzie stanowił nowość w skali krajowej.

2. Realizacja projektu

Prace w ramach realizacji projektu VESTA realizowało Konsorcjum naukowe skupiające:

- Instytut Technologii Bezpieczeństwa MORATEX (Lider projektu);
- Akademię Sztuk Pięknych im. Władysława Strzemińskiego w Łodzi;
- Centralny Instytut Ochrony Pracy – PIB;
- Przedsiębiorstwo Sprzętu Ochronnego MASKPOL S.A.;
- Motex I. Kobylarczyk i wspólnicy spółka jawna.

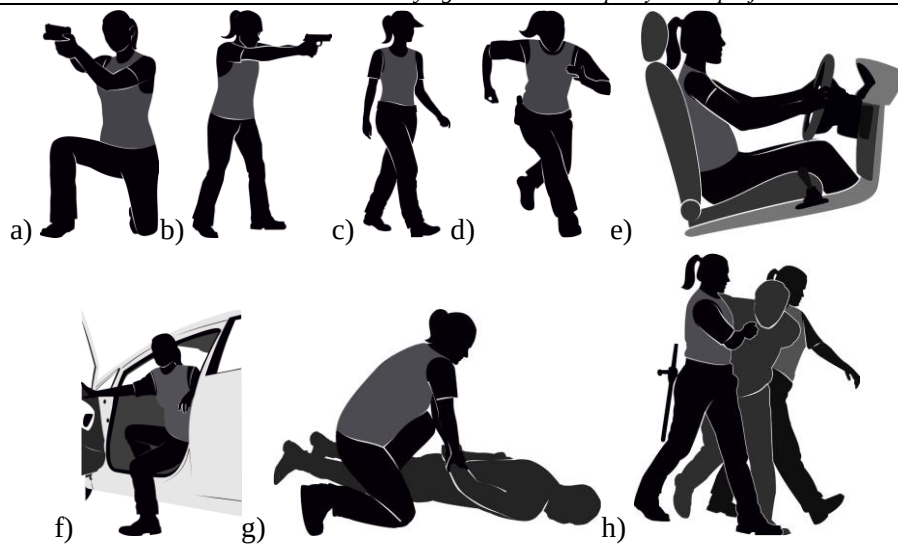
Prace badawcze oraz rozwojowe zostały podzielone na pięć faz, zgodnie ze schematem (rys. 5).



Rysunek 5. Schemat faz realizacji projektu VESTA [opracowanie własne ITB „MORATEX"]

2.1. Projektowanie

Proces projektowania poprzedzony został procesem analitycznym, w trakcie którego wygenerowane zostały wszystkie problemy projektowe dotyczące kwestii użytkowych i funkcjonalnych opracowywanego systemu DWKB. W tym celu przeprowadzono wywiad z funkcjonariuszkami Policji, które były zróżnicowane pod względem budowy anatomicznej. Celem prowadzonych rozmów było określenie warunków motoryki ruchowej funkcjonariuszek, wyselekcjonowanie najważniejszych postaw jakie są przyjmowane podczas pełnienia służby (rys. 6), określenie ograniczeń ruchowych w danych pozycjach wynikających z noszenia kamizelki balistycznej oraz określenie estetycznych i technicznych założeń dla DWKB.



Rysunek 6. Wybrane pozycje przyjmowane przez funkcjonariuszki podczas pełnienia służby: a) klęcząca, b) stojąca, c) podczas pieszego patrolowania, d) w trakcie biegu, e) podczas prowadzenia pojazdu służbowego, f) podczas opuszczania pojazdu służbowego, g) w trakcie zatrzymania podejrzanego, h) podczas prowadzenia podejrzanego [grafiki: ASP w Łodzi]

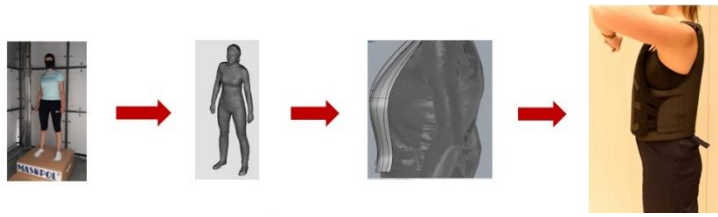
Analizując otrzymane w wyniku przeprowadzonego wywiadu informacje stwierdzono, że konstrukcja kamizelki musi być zaprojektowana w taki sposób, aby:

- uniknąć ucisku w części brzusznej tułowia podczas przyjmowania pozycji siedzącej i pochylonej (kiedy górna przednia część tułowia ulega skuleniu);
- użytkowniczka przyjmując pozycję strzelecką (szczególnie podczas oburęcznego chwytu broni) nie odczuwała dyskomfortu w okolicach pach;
- uzyskać maksymalne dopasowanie systemu DWKB do ciała użytkowniczki;
- otrzymany system DWKB był kompatybilny ze stosowaną przez funkcjonariuszki odzieżą służbową (umundurowaniem).

Projekt rozpoczęto od wytypowania funkcjonariuszek służb mundurowych z terenu całej Polski i zeskanowania ich ciał, ze szczególnym uwzględnieniem wymiarów klatki piersiowej oraz biustu.

Zastosowanie technologii skanowania 3D, w celu jak najdokładniejszego odwzorowania kobiecej sylwetki, miało na celu umożliwienie opracowania dopasowanego systemu DWKB. Odpowiednie dopasowanie kamizelki balistycznej do ciała użytkowniczki pozwala bowiem, co należy wyraźnie podkreślić, na zapewnienie pełniejszej funkcjonalności, szczególnie w obrębie wkładów ochronnych.

Następnie na podstawie wytycznych do konstrukcji nowoopracowywanego wyrobu oraz danych ze skanowania sylwetek, przystąpiono do prac nad projektowaniem systemu DWKB.



Rysunek 7. Etapy prac: skanowanie 3D, projektowanie, wyrób końcowy
[foto: ITB „MORATEX", PSO MASKPOL, ASP w Łodzi]

2.2. Wytworzenie i przebadanie partii modelowej

Proces projektowania obejmował wszystkie elementy opracowywanego systemu:

- wyroby tekstylne (podsystem korekcyjny oraz poszycie kamizelki);
- wyroby kompozytowe (wkłady ochronne zapewniające odporność balistyczną oraz na uderzenie broni białej).

Projektowanie wyrobów tekstylnych, jak i kompozytowych wiązało się z wytypowaniem, a następnie weryfikacją kluczowych parametrów technicznych materiałów i surowców. W przypadku poszycia kamizelki tkaniny, przeznaczone na zewnętrzną stronę wyrobu, zostały dobrane po potwierdzeniu otrzymania wyniku stwierdzającego ograniczone rozprzestrzenianie płomienia ≥ poziom 3 wg PN-EN ISO 14116:2015 [8]. Dzianiny, przeznaczone na wewnętrzną stronę poszycia – bezpośrednio od ciała, zostały dobrane na podstawie parametru przepuszczalności powietrza ≥ 1500 mm/s oraz oporu pary wodnej Ret < 4 m^2Pa/W. W ten sposób poszycie kamizelki, przystosowane do noszenia wewnątrz wkładów ochronnych, miało umożliwiać aktywną regulację różnicy temperatur mikroklimatu przyskórnego użytkowniczki i otoczenia, obniżając potliwość ciała, polepszając komfort termiczny, co miało sprzyjać polepszeniu właściwości psychomotorycznych użytkowniczki.

Projektowanie wyrobów kompozytowych wiązało się z w pierwszej kolejności z dokonaniem oceny wybranych właściwości mechanicznych, fizycznych i strukturalnych wytypowanych materiałów balistycznych i/lub odpornych na przebicie, istotnych dla uzyskania wkładów ochronnych o wymaganych parametrach ochronnych (kulo-, odłamkoodporność oraz opcjonalnie igło-, szpikulco- i ostrzoodporność), na bazie:

- tkanin z włókien para-aramidowych;
- tkanin z włókien UHMWPE;
- nietkanych wyrobów z włókien UHMWPE jednokierunkowych zabezpieczonych filmem ochronnym;
- kompozytów z włókien UHMWPE zabezpieczonych filmem ochronnym;
- kompozytów z włókien para-aramidowych zabezpieczonych filmem ochronnym;
- tkanin z włókien para-aramidowych pokrytych warstwą ochronną.

Następnie na podstawie zweryfikowanych materiałów balistycznych i/lub odpornych na przebicie zostały zaprojektowane i wykonane wkłady w dwóch wersjach:

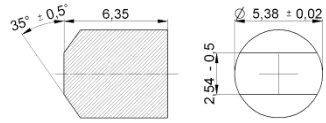
1. podstawowej – wkład balistyczny kuloodporny oraz odłamkoodporny,
2. rozszerzonej – wkład balistyczny kuloodporny oraz odłamkoodporny oraz wkładka odporna na przebicie bronią białą.

Wersja podstawowa miała zapewniać ochronę w obrębie:

- kuloodporności w klasie K2 zgodnie z PN-V-87000:2011 [9] (rys. 8);

Rysunek 8. Pocisk 7,62×25 TT FMJS wykorzystywany do badań kuloodporności w klasie K2 zgodnie z PN-V-87000:2011 [9] [foto: ITB „MORATEX"]

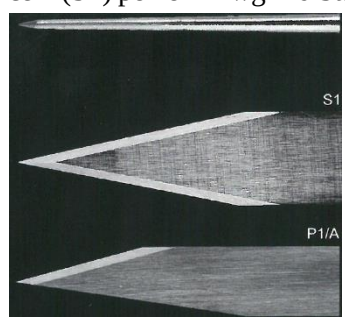
- odłamkoodporności O3 zgodnie z PN-V-87000:2011 [9] (rys. 9)



Rysunek 9. Odłamek standardowy TYP-1 (wg PN-V-87000:2011) wykorzystywany do badań odłamkoodporności zgodnie z PN-V87000:2011 [9]

Wersja rozszerzona miała zapewniać ochronę w obrębie uderzenia:

- ostrzem (S1 i P1), szpikulcem (SP) poziom 2 wg NIJ Standard 0115.00 [10] (rys. 10);



Rysunek 10. Ostrza badawcze, od góry: szpikulec SP, ostrze S1 oraz ostrze P1 wykorzystywane do badań odporności na przebicie zgodnie z NIJ Standard 0115.00 [folder reklamowy firmy TEIJIN]

- igłą typ 18G (1,2 mm x 40 mm) i typ 21G (0,8 mm x 40 mm) z energią 25,0 ±0,5 J – poziom 2 (w oparciu o VPAM KDIW 2004 [11]) wg procedury PBB-12:2008 [12].

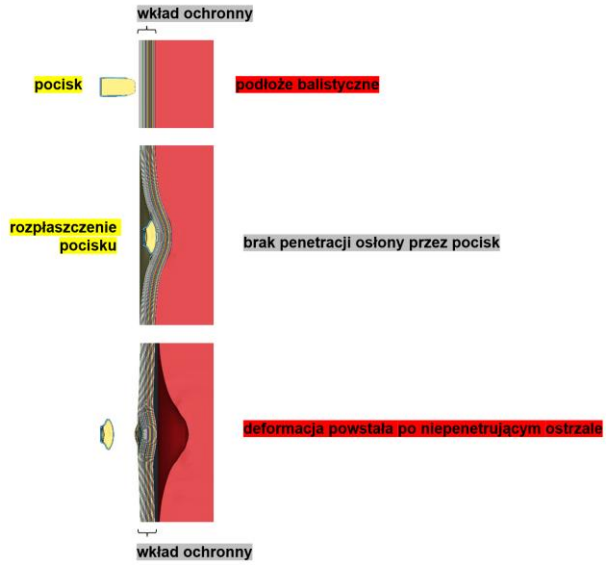
Głównym założeniem projektowanych wkładów ochronnych było ograniczenie ryzyka wystąpienia obrażeń wewnętrznych w przypadku niepenetrującego ostrzału (rys. 11) poprzez obniżenie dynamicznego odkształcenia określonego parametrami:

- głębokości deformacji z 40 mm do <35 mm;
- objętości deformacji <30 cm^3.

Na rysunku 11 przedstawiono schemat ostrzału wkładu ochronnego prowadzony zgodnie z PN-V-87000:2011 [9]. Wkład ochronny umieszczony jest na podłożu balistycznym (plastelinie balistycznej). Po oddaniu strzału zdefiniowanym wcześniej pociskiem postępowano zgodnie z poniższymi wytycznymi.

1. Zaobserwować czy nastąpiła penetracja pocisku przez wkład:

- Opcja A: wkład został przebity przez pocisk,
- Opcja B: pocisk zatrzymał się na wkładzie, nastąpiła deformacja pocisku oraz deformacja podłoża balistycznego (rys. 11).



Rysunek 11. Schemat ostrzału niepenetrującego [ITB „MORATEX"]

W przypadku Opcji A badanie nie było kontynuowane, a konstrukcja wkładu ochronnego wymagała optymalizacji i ponownego badania.
W przypadku Opcji B przechodzono do kolejnego punktu.

2. Zmierzyć wymiary deformacji, powstałej w podłożu balistycznym w wyniku niepenetrującego ostrzału (rys. 11), jej głębokość, pole oraz objętość.

Konstrukcja wkładów ochronnych, która została pozytywnie zweryfikowane pod kątem bezpieczeństwa mogła zostać wykorzystana do wytworzenia partii modelowej DWKB.

Partia modelowa DWKB zweryfikowana pod kątem bezpieczeństwa kierowana była na laboratoryjne badania ergonomii i badań użytkowych z zastosowaniem symulacji warunków rzeczywistych. Wyniki przeprowadzonych badań oraz wnioski z ich analizy pozwoliły na opracowanie rekomendacji do optymalizacji procesu wytwarzania DWKB pod kątem ergonomii i bezpieczeństwa użytkowania w aspekcie wyeliminowania zagrożeń wtórnych w ramach eksperymentalnych prac rozwojowych (partia prototypowa).

2.3. Wytworzenie i przebadanie partii prototypowej

Wytworzenie partii prototypowej rozpoczęto od analizy uwag zgłoszonych przez funkcjonariuszki w ankietach podsumowujących testowanie partii modelowej DWKB oraz rekomendacje. Głównym celem było wyeliminowanie uwag dotyczących dyskomfortu, ograniczonej układalności oraz niedogodności użytkowania systemu DWKB podczas symulowanych działań w czasie pełnienia służby.

Poprawki, korekty zostały przesłane do przedsiębiorców celem weryfikacji konstrukcji systemu DWKB oraz odszycia partii prototypowej do laboratoryjnych badań ergonomicznych i użytkowych partii prototypowej systemu DWKB.

Równolegle przeprowadzono weryfikację parametrów wytrzymałościowych oraz odporności chemicznej materiałów i dodatków wytypowanych do wykonania partii prototypowej DWKB. Ponadto zoptymalizowano układy materiałowe wkładów ochronnych, które wraz ze skorygowanymi projektami oraz wytycznymi konstrukcyjnymi zostały

przekazane do przedsiębiorców celem odszycia partii prototypowej do laboratoryjnych badań ergonomicznych i użytkowych partii prototypowej DWKB [13].

Równolegle partia prototypowa wkładów ochronnych została poddana weryfikacji pod kątem bezpieczeństwa zarówno przed, jak i po procesie przyspieszonego starzenia (symulującym okres 6 lat użytkowania, zgodnie z PN-V-87001:2011). Wyniki przeprowadzonych badań potwierdziły, iż partia prototypowa wkładów ochronnych spełnia założone wymagania ochronne.

Na podstawie laboratoryjnych badań eksploatacyjnych stwierdzono, że po założeniu i prawidłowym dopasowaniu do ciała partii prototypowej systemu DWKB i wykonaniu przez funkcjonariuszki Policji serii badań statycznych i dynamicznych [13], zapewniona jest ochrona tułowia z przodu, z tyłu i na bokach. Partia prototypowa systemu DWKB nie wchodzi w niepożądaną interakcję z odzieżą wierzchnią stosowaną przez funkcjonariuszki Policji, zapewnia swobodny dostęp do broni i innych elementów wyposażenia pasa taktycznego bez względu na wykonywane czynności i pozycję ciała.

W wyniku badań użytkowych przeprowadzonych w warunkach rzeczywistych dokonano oceny prototypu systemu DWKB pod względem ergonomii, funkcjonalności i bezpieczeństwa użytkowania przez docelową grupę odbiorców – funkcjonariuszki Policji podczas wykonywania rutynowych czynności służbowych. Wnioski z przeprowadzonych badań stanowiły podstawę do opracowania rekomendacji do optymalizacji procesu wytwarzania DWKB na poziomie prac wdrożeniowych.

Opracowany samonośny i modelujący podsystem korekcyjny posiada właściwości ograniczające wstrząsy i stabilizuje biust, co pozwala na zwiększenie dopasowania kamizelki oraz zapewnia długotrwały komfort użytkowania.

W badaniach użytkowych potwierdzono także, że wkłady ochronne – wersja podstawowa oraz wersja rozszerzona zapewniają osłonę organów wewnętrznych podczas wykonywania czynności zawodowych.

W związku z powyższym potwierdzono, że prototyp systemu DWKB umożliwia funkcjonariuszkom Policji realizację działań interwencyjno-patrolowych w różnych warunkach operacyjnych.

3. Podsumowanie

W ramach pracy konsorcjum VESTA zaprojektowano, wytworzono, przebadano i potwierdzono, że system Damskiej Wielofunkcyjnej Kamizelki Balistycznej jest:

- kompatybilny;
- dopasowany do ciała użytkowniczek;
- przeznaczony do całorocznego użytkowania;
- umożliwiający kobietom realizację zadań operacyjnych w ramach obowiązków służbowych.

Podsumowując, projekt zakończono opracowaniem systemu zindywidualizowanej Damskiej Wielofunkcyjnej Kamizelki Balistycznej DWKB skrytego noszenia dla dwóch podgrup rozmiarowych: małe i średnie biusty MŚB oraz duże biusty DB, na IX poziomie gotowości technologicznej (IX TRL).

Efektem projektu jest dopasowany, ergonomiczny, kompatybilny system DWKB stanowiący pod względem funkcjonalnym jednolity zestaw wyposażenia, umożliwiający kobietom wykonywanie obowiązków służbowych w różnych warunkach operacyjnych.

Literatura

1. https://www.infosecurity24.pl/kobiety-w-sluzbach-mundurowych [data dostępu: 13.03.2023].
2. http://www.wojsko-polskie.pl/f/pages/atts/2018/2/kobiety_zolnierze2017.pdf [data dostępu: 15.07.2018].
3. https://www.defence24.pl/5-tys-zolnierzy-w-wojskach-obrony-terytorialnej [data dostępu: 13.03.2023].
4. https://www.sop.gov.pl/download/2/2/119678045126110.pdf [data dostępu: 16.06.2018]
5. http://bip.sm.szczecin.pl/chapter_86014.asp?soid=9BD4B2C1C31A4A389739745148500 5B9 [data dostępu: 13.03.2023].
6. http://www.theiacp.org/Portals/0/documents/pdfs/Femalebodyarmor.pdf [data dostępu: 13.03.2023].
7. Łuczyński W., *Sposób wytwarzania kamizelki kuloodpornej dla kobiet z profilowanym wkładem balistycznym*, Patent RP 347336, 2001.
8. PN-EN ISO 14116:2015 Odzież ochronna – Ochrona przed płomieniem – Materiały, zestawy materiałów i odzież o ograniczonym rozprzestrzenianiu płomienia.
9. PN-V-87000:2011 Osłony balistyczne lekkie. Kamizelki kulo- i odłamkoodporne. Wymagania ogólne i badania.
10. NIJ Standard 0115.00 Stab Resistance of Personal Body Armor.
11. VPAM KDIW 2004 Test Standard Stab and Impact Resistance. Requirements, classifications and test procedures.
12. PBB-12/ITB „MORATEX" ed. I z 12.2008 w zakresie odporności na uderzenie igły typ 18G (1,2mm x 40mm) i typ 21G (0,8mm x 40mm) z energią 25,0 ± 0,5 J – poziom 1 lub 2 (w oparciu o VPAM KDIW 2004 [11]).
13. Kucharska-Jastrząbek A., Suchocki P., Pieniążek D., Garus R., Dmowska-Jasek P., Kudlińska M., Olejnik M., Woźniakowska M., Gutowska A., Łężak K., Majchrzycka K., *Indywidualizacja damskiej kamizelki skrytego noszenia pod kątem ergonomii użytkowania*, [w:] Mołdoch-Mendoń I., Maciąg K. (red.), *Rozwiązania technologiczne XXI wieku – skutki i perspektywy rozwoju. Tom 4*, Wydawnictwo Naukowe TYGIEL Sp. z o.o., Lublin 2023.

VESTA – damska kamizelka skrytego noszenia – od pomysłu do projektu

Streszczenie

Obecnie w służbach mundurowych pełni służbę coraz więcej kobiet, które w realizacji zadań operacyjnych w ramach obowiązków służbowych korzystają z osłon tułowia dedykowanych i projektowanych na sylwetkę męską. Kamizelki takie mogą być relatywnie skutecznie stosowane przez kobiety o małym rozmiarowo biuście, nie zapewniając im jednak wymaganej ergonomii. Kluczowy jest fakt, że ze względu na brak pełnego dopasowania osłony tułowia do sylwetki kobiety istnieje pewność, że nie spełni funkcji ochrony narządów wewnętrznych (zwłaszcza w obszarze pomiędzy biustem a podbrzuszem lub w okolicy pach). Biorąc pod uwagę, że pod wpływem uderzenia penetratora (pocisku i/lub odłamka) w kamizelkę balistyczną następuje jej gwałtowna, chwilowa deformacja, która w konsekwencji może doprowadzić do tępego obrażenia ciała jej użytkownika, niezmiernie istotne jest dopasowanie kamizelki balistycznej do ciała użytkownika w celu zminimalizowania ryzyka obrażeń wewnętrznych. Idea zaprojektowania damskiej kamizelki skrytego noszenia uwzględniała zachowanie wymaganych poziomów bezpieczeństwa na całej powierzchni ochrony niezależnie od formy i kształtu osłony. W efekcie prace projektowe ukierunkowane były na optymalizację konstrukcji kamizelki w celu zapewnienia pełnego dopasowania kształtu do damskiej sylwetki z zachowaniem wymaganej ochrony balistycznej, a także ergonomii w czasie długotrwałego użytkowania.

Słowa kluczowe: damska kamizelka skrytego noszenia, wielofunkcyjne osłony balistyczne, technologia skanowania 3D, indywidualizacja konstrukcji, dopasowanie

VESTA – women's concealed vest – from idea to design

Abstract
Currently, more and more women serve in the uniformed services, who use ballistic vests dedicated and designed for a male silhouette in the performance of operational tasks as part of their official duties. Such vests can be relatively effectively used by women with small breasts, but they do not provide them with the required ergonomics. The key is the fact that due to the lack of a full fit of the ballistic vests to the woman's figure, there is a certainty that it will not fulfill the function of protecting internal organs (especially in the area between the breasts and the lower abdomen or in the armpit area). Considering that the impact of the penetrator (bullet and/or fragment) on the ballistic vest causes its rapid, temporary deformation, which in turn may lead to blunt injury to the user's body, it is extremely important to fit the ballistic vest to the user's body in order to minimize risk of internal injury. The idea of designing a women's concealed vest took into account the maintenance of the required levels of safety over the entire surface of the protection, regardless of the form and shape of the cover. As a result, the design work was focused on optimizing the construction of the vest in order to ensure full fit of the shape to the female figure while maintaining the required ballistic protection, as well as ergonomics during long-term use.
Keywords: women's concealed vest, multifunctional ballistic shields, 3D scanning technology, individualization of construction, fit

Paulina Dmowska-Jasek[1], Agnieszka Kucharska-Jastrząbek[2]

Indywidualizacja damskiej kamizelki skrytego noszenia pod kątem ergonomii użytkowania

1. Wprowadzenie

Niniejsze opracowanie stanowi kontynuację rozdziału pt. „VESTA – damska kamizelka skrytego noszenia – od pomysłu do projektu" zamieszczonej w niniejszej monografii [1].

W ramach realizacji projektu pt. „Indywidualizacja systemu osłony osobistej", akronim „VESTA" (POIR 04.01.04-00-0085/18), zaprojektowano, zweryfikowano i zoptymalizowano Damską Wielofunkcyjną Kamizelkę Balistyczną skrytego noszenia (DWKB), przeznaczoną do całorocznego użytkowania przez funkcjonariuszki służb mundurowych. Opracowany system DWKB stanowi jednolity zestaw wyposażenia, umożliwiający kobietom realizację zadań służbowych w różnorodnych warunkach operacyjnych. System DWKB stanowi nowatorskie podejście do projektowania osłon tułowia uwzględniając fizjologię ciała kobiety. Odpowiednie dopasowanie systemu DWKB do kształtu ciała funkcjonariuszki bezpośrednio przekłada się na: zapewnienie pełnej powierzchni ochrony balistycznej, z możliwością dodatkowej ochrony przed bronią białą, obniżenie ryzyka wystąpienia obrażeń narządów wewnętrznych tzw. efektu BABT (ang. *Behind Armour Blunt Trauma*), a także zwiększenie komfortu i ergonomii użytkowania podczas wykonywania czynności służbowych.

2. Cel pracy

Celem niniejszej pracy jest omówienie wyników badań ergonomicznych opracowanego w ramach projektu systemu Damskiej Wielofunkcyjnej Kamizelki Balistycznej skrytego noszenia (DWKB). Jednym z głównych założeń realizowanego projektu było zapewnienie pełnego dopasowania i zachowania ergonomii użytkowania podczas stosowania systemu DWKB w warunkach operacyjnych przez funkcjonariuszki służb mundurowych.

3. Przedmiot i uczestnicy badań

Celem weryfikacji zapewnienia pełnego dopasowania i zachowania ergonomii użytkowania podczas stosowania systemu DWKB w warunkach operacyjnych przez funkcjonariuszki służb mundurowych przeprowadzono dwa rodzaje badań:

1. Laboratoryjne badania ergonomiczne i użytkowe z zastosowaniem symulacji warunków rzeczywistych;
2. Badania ergonomiczne i użytkowe przeprowadzone w warunkach rzeczywistych.

[1] pjasek@moratex.eu, Instytut Technologii Bezpieczeństwa „MORATEX", Zakład Projektowy, www.moratex.eu.

[2] ajastrzabek@moratex.eu, Instytut Technologii Bezpieczeństwa „MORATEX", Zakład Projektowania Systemów Informatycznych, www.moratex.eu.

Badania laboratoryjne przeprowadzono z udziałem 12 funkcjonariuszek Policji, których sylwetki charakteryzowały się wymiarami zgodnymi z dwoma zdefiniowanymi typami:

- Typ 1: małe i średnie biusty (MŚB) – miseczka AA, A, B i C – 6 funkcjonariuszek;
- Typ 2: duże biusty (DB) – miseczka D, E, F, G i H – 6 funkcjonariuszek.

Badania laboratoryjne prowadzono dwukrotnie. Pierwsze odbyły się celem weryfikacji partii modelowej systemu DWKB. Wyniki przeprowadzonych badań oraz wnioski z ich analizy pozwoliły na opracowanie rekomendacji do optymalizacji procesu wytwarzania DWKB pod kątem ergonomii i bezpieczeństwa użytkowania w aspekcie wyeliminowania zagrożeń wtórnych w ramach eksperymentalnych prac rozwojowych (partia prototypowa). Analiza wyników badań laboratoryjnych przeprowadzonych dla partii modelowej nie została opisana w niniejszej monografii. Celem niniejszego opracowania było porównanie wyników badań prowadzonych dla tej samej partii wyrobu DWKB (partia prototypowa) w warunkach laboratoryjnych (symulujących warunki rzeczywiste) oraz w warunkach rzeczywistych.

W **badaniach** ergonomicznych i użytkowych **w warunkach rzeczywistych** uczestniczyło 50 funkcjonariuszek Policji, których sylwetki charakteryzowały się wymiarami zgodnymi ze zdefiniowanymi typami tj.:

- Typ 1: małe i średnie biusty (MŚB) – miseczka AA, A, B i C – 43 funkcjonariuszki;
- Typ 2: duże biusty (DB) – miseczka D, E, F, G i H – 7 funkcjonariuszek.

Przed przystąpieniem do badań, mających na celu ocenę systemu DWKB funkcjonariuszki zostały scharakteryzowane pod kątem:

- doświadczenia zawodowego w użytkowaniu kamizelki skrytego noszenia:

W grupie MŚB doświadczenie w korzystaniu ww. kamizelek miało łącznie 54% funkcjonariuszek (23 osoby), 47% (20 osób) nigdy nie nosiło tego typu kamizelek.

W grupie DB doświadczenie w użytkowaniu ww. kamizelek miało łącznie 72% (5 osób), 29% (2 osoby) nigdy nie nosiło tego typu kamizelek.

- rodzaju wykonywanej pracy, podczas której stosowano oceniany system DWKB:

Funkcjonariuszki z obu grup (MŚB i DB) najczęściej testowały kamizelkę podczas patrolu samochodowego.

Spośród grupy MŚB blisko połowa respondentek (21 osób) wskazała, że użytkowała ją podczas patrolu pieszego, znacznie mniej tj. 19% (8 osób) podczas pracy biurowej, a tylko 7% (3 osoby) podczas udziału w konwoju.

Z grupy DB pozostałe funkcjonariuszki testowały kamizelkę podczas pracy biurowej (29%; 2 osoby) i podczas patrolu pieszego (14%, 1 osoba).

Partia prototypowa systemu DWKB stanowiąca przedmiot niniejszych badań ergonomicznych została przebadana w dwóch wersjach:

- podstawowej – wyposażonej we wkład balistyczny kuloodporny oraz odłamkoodporny, zapewniający ochronę balistyczną w obszarze kuloodporności w klasie K2 oraz odłamkoodporności w klasie O3 zgodnie z PN-V87000:2011 [2],
- rozszerzonej – wyposażonej we wkład balistyczny kulo- i odłamkoodporny oraz wkładkę odporną na przebicie bronią białą, zapewniającą dodatkową ochronę przed uderzeniem ostrzem (S1 i P1), szpikulcem (SP) poziom 2 wg NIJ Standard 0115.00 [3] oraz igłą typ 18G (1,2 mm x 40 mm) i typ 21G (0,8 mm x 40 mm) z energią 25,0 ±0,5 J – poziom 2 (w oparciu o VPAM KDIW 2004 [4]) wg procedury PBB-12:2008 [5].

4. Laboratoryjne badania ergonomiczne i użytkowe systemów DWKB z zastosowaniem symulacji warunków rzeczywistych

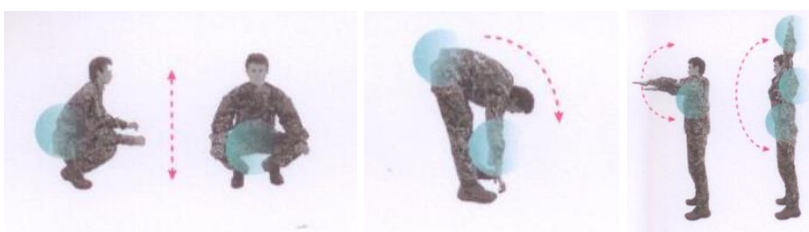
Laboratoryjne badania ergonomiczne i użytkowe prowadzone były w Laboratorium Zakładu Ochron Osobistych CIOP-PIB w Łodzi, w warunkach normalnych, tj. w temperaturze 23 ±3°C i wilgotności względnej powietrza 65 ±5%. Przed rozpoczęciem badań systemy DWKB poddano aklimatyzacji w warunkach temperatur otoczenia, przez 24 godziny.

Każdej uczestniczce badania dostarczono komplet wyposażenia przewidzianego do badań tj. odpowiednio dopasowany prototyp systemu DWKB w wersji podstawowej i rozszerzonej. Funkcjonariuszki Policji wykonywały ćwiczenia we własnej odzieży służbowej oraz obuwiu typu sportowego z antypoślizgową podeszwą.

Wykonano dwie identyczne serie badań eksploatacyjnych w odstępie tygodnia. W obu seriach badań uczestniczyły te same funkcjonariuszki Policji, ubrane w te same prototypy systemu DWKB.

Badania w obu seriach prowadzono według następującego schematu:

- zapoznanie się z instrukcją użytkowania prototypu systemu DWKB;
- założenie i dopasowanie prototypu systemu DWKB;
- dokonanie sprawdzenia poprawności założenia i dopasowania prototypu systemu DWKB przez osobę nadzorującą ćwiczenia i odnotowanie uwag i spostrzeżeń odnośnie do systemu DWKB i instrukcji użytkowania,
- wypełnienie przez uczestniczki badań ankiety odnoszącej się do zawartości instrukcji użytkowania,
- wykonanie serii ćwiczeń statyczno-ruchowych (rys. 1 i tab. 1) przez funkcjonariuszki Policji i obserwacja przez osobę nadzorującą ćwiczenia czy nie występują problemy związane z dopasowaniem i kompatybilnością elementów systemu prototypu DWKB,



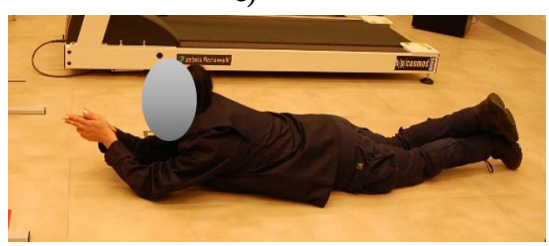
Rysunek 1. Przykładowe ćwiczenia statyczno-ruchowe: a) pozycja kuczna, b) skłon do przodu i na boki, c) wyciągnięcie rąk do przodu i na boki. Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP

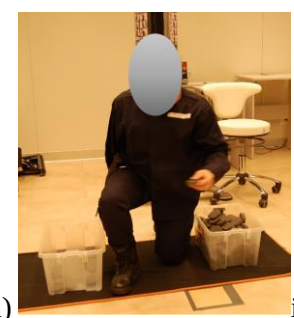
- 10 minutowy odpoczynek, a następnie założenie na system DWKB odzieży wierzchniej;
- wykonanie pod nadzorem serii zaplanowanych ćwiczeń dynamicznych (tab. 1 i rys. 2) i dokonanie przez osobę nadzorującą badanie obserwacji ewentualnych problemów w wykonaniu ćwiczeń wynikających z użytkowanych prototypu systemu DWKB;

a) b)

c)

d) e)

f) 

g) h)  i)

Rysunek 2. Prezentacja wybranych ćwiczeń wykonywanych przez funkcjonariuszki Policji podczas laboratoryjnych badań eksploatacyjnych; a) chód i bieg na bieżni, b) naskok i zeskok z przeszkody, c) przechodzenie pod przeszkodą, d) przechodzenie nad przeszkodą, e) czołganie się, f) pozycja strzelecka na leżąco z podparciem, g) pozycja strzelecka bez podparcia, h) symulacja pracy w pozycji klęczącej naprzemiennie na jedno kolano, i) siedzenie na podłożu płaskim ze skrzyżowanymi nogami. Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP

- wypełnienie przez uczestniczki badań ankiety odnoszącej się do oceny prototypu systemu DWKB i sporządzenie notatki przez osobę nadzorującą wykonywanie ćwiczeń.

Tabela 1. Harmonogram ćwiczeń fizycznych wykonywanych przez uczestniczki badań ergonomicznych

L.p.	**Czynność / ćwiczenia**	**Czas w min.**
1.	Ankieta COVID	15
2.	Pomiar uczestniczki badań	
3.	Przebranie i dopasowanie kamizelki	
	Ćwiczenia statyczne	
4.	Dopasowanie kamizelki	
	Ćwiczenia dynamiczne	
1.	Chodzenie na bieżni z prędkością 6 km/h	1
2.	Bieganie na bieżni z prędkością 8 km/h	2
3.	ZESKOK Wejścia na podwyższenie o wysokości 0,4 m i zeskok (przy ugiętych kolanach, dwiema nogami jednocześnie) – 3 powtórzenia	-
4.	Przejście w pozycji pochylonej w ograniczonej przestrzeni – 3 powtórzenia: przejścia w pozycji pochylonej pod przeszkodą ustawioną na wysokości 1 m, przejście w pozycji wyprostowanej nad przeszkodą ustawioną na wysokości 0,6 m	3
5.	ZESKOK jw. – 3 powtórzenia	-
6.	Czołganie się w tunelu o wys. 70 cm – 4 x 7 m	3
7.	ZESKOK jw. – 3 powtórzenia	-
8.	Symulacja postawy strzeleckiej z bronią krótką leżąc przodem: - przekręcanie się na boki w pozycji leżącej z rękami wyprostowanymi poziomo za głowę do pozycji strzeleckiej (broń trzymana oburącz) – 3 powtórzeń w każdą stronę o 360°	-
	Ćwiczenia statyczne Symulacja pracy w pozycji klęczącej – kolejno na lewym i prawym kolanie – przekładanie małych elementów ze skrzyni do skrzyni	4 (2x2)
10	ZESKOK jw. – 3 powtórzenia	-
	Ćwiczenia statyczne Siedzenie na podłożu płaskim ze skrzyżowanymi nogami	2
	ANKIETA – ocena zestawu	-

Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP.

4.1. Metodyka oceny statystycznej zebranych wyników ankietowych

Przeprowadzona analiza wyników badań ankietowych w szczególności dotyczyła oceny:

- stopnia zapewnienia odpowiedniej osłony powierzchni ciała z punktu widzenia ochrony balistycznej i mechanicznej z uwzględnieniem potencjalnych zmian pozycji systemu DWKB po jej prawidłowym dopasowaniu,
- poziomu akceptacji określonego dyskomfortu wynikającego z konieczności stosowania systemu DWKB, w warunkach symulujących czynności zawodowe funkcjonariuszek Policji w sytuacjach zagrożenia życia lub zdrowia,
- wyeliminowania zagrożeń wtórnych wynikających wyłącznie z charakterystyki konstrukcyjnej i materiałowej prototypu systemu DWKB.

Odpowiedzi na pytania w ankiecie miały charakter trójstopniowy:

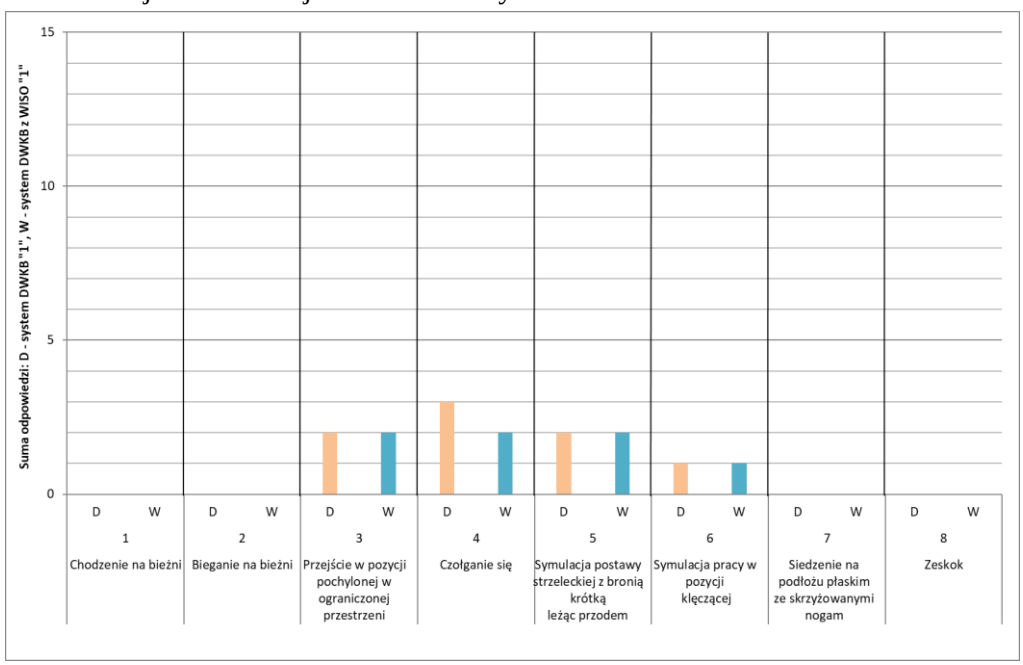
- 0 – ocena pozytywna – bez uwag;
- 1 – niewielkie trudności w zrozumieniu – opis niejasności;
- 2 – zdecydowane niezrozumiały opis/symbole – opis niejasności.

Oddzielnie oceniany był system DWKB w wersji podstawowej oraz oddzielnie w wersji rozszerzonej.

Ocena pozytywna systemu DWKB wiązała się z uzyskaniem, dla wszystkich odpowiedzi na pytania, postawione w ankiecie, sumarycznie najmniejszej liczby punktów.

4.2. Wyniki badań

Wyniki oceny sumarycznej z I i II serii badań systemu DWKB w wersji podstawowej oraz w wersji rozszerzonej zestawiono na rysunkach 3 i 4.

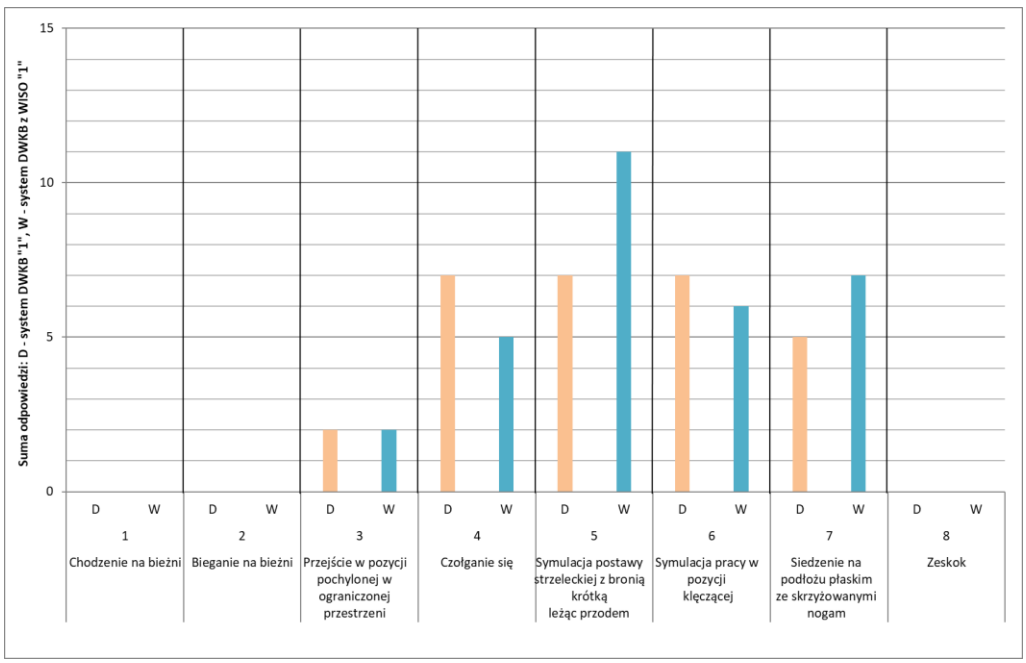


Rysunek 3. Zestawienie wyników oceny systemu DWKB w wersji podstawowej oraz w wersji rozszerzonej przez grupę funkcjonariuszek MŚB sumarycznie w I i II serii badań w zależności od wykonywanych ćwiczeń fizycznych symulujących czynności służbowe, gdzie: D – ocena „1" system DWKB w wersji podstawowej, W – ocena „1" system DWKB+WISO w wersji rozszerzonej. Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP

Wyniki oceny ankietowej systemów DWKB w zależności od wykonywanych ćwiczeń fizycznych (rys. 3 i 4) wskazują, że:

- oceniane systemy DWKB jeśli powodują jakiś dyskomfort to tylko niewielki (oceny na poziomie „1", brak ocen na poziomie „2");
- zdecydowanie więcej uwag krytycznych zgłosiły funkcjonariuszki z grupy DB;
- żadna z funkcjonariuszek z grupy MŚB, jak i DB podczas chodzenia, biegania na bieżni i zeskoków z podwyższenia nie zgłosiła uwag do obu badanych systemów DWKB;
- najwięcej uwag krytycznych do systemu DWKB od funkcjonariuszek z grupy MSB dotyczyły przejścia w pozycji pochylonej w ograniczonej przestrzeni, czołgania się i symulacji postawy strzeleckiej z bronią krótką (max. 3 uwagi krytyczne na poziomie „1");

- najwięcej uwag krytycznych do systemu DWKB od funkcjonariuszek z grupy DB dotyczyło czołgania się, symulacji postawy strzeleckiej z bronią krótką, symulacji pracy w pozycji klęczącej oraz siedzenia na podłożu płaskim ze skrzyżowanymi nogami (11 uwag krytycznych na poziomie „1");
- w przypadku funkcjonariuszek z grupy DB podczas symulacji postawy strzeleckiej z bronią krótką i siedzenia na podłożu płaskim ze skrzyżowanymi nogami system DWKB w wersji rozszerzonej powoduje większy dyskomfort niż system DWKB w wersji podstawowej.



Rysunek 4. Zestawienie wyników oceny systemu DWKB w wersji podstawowej oraz w wersji rozszerzonej przez grupę funkcjonariuszek DB sumarycznie w I i II serii badań w zależności od wykonywanych ćwiczeń fizycznych symulujących czynności służbowe, gdzie: D – ocena „1" system DWKB w wersji podstawowej, W – ocena „1" system DWKB+WISO w wersji rozszerzonej. Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP

4.3. Wnioski

Wnioski z badań ankietowych sformułowano w odniesieniu do prototypu systemu DWKB w wersji podstawowej i rozszerzonej.

Ocena systemów DWKB po założeniu i prawidłowym dopasowaniu do ciała i wykonaniu serii badań statycznych wskazał, że system DWKB w wersji podstawowej oraz rozszerzonej zapewnia osłonę balistyczną i mechaniczną tułowia z przodu, z tyłu i po bokach. Ogólna ocena systemu DWKB przez funkcjonariuszki z grupy MŚB i DB wskazywała, że system ten jest ergonomiczny, funkcjonalny i spełnia swoją funkcję ochronną. Zastosowanie dodatkowej ochrony przed bronią białą (wersja rozszerzona DWKB) nie wpłynęło na pogorszenie właściwości ergonomicznych systemu, o czym świadczy zbliżona liczba uwag dotycząca systemu DWKB w wersji podstawowej oraz rozszerzonej.

Na podstawie obserwacji i wywiadów osoby nadzorującej badania stwierdzono, że funkcjonariuszki zarówno w grupie MŚB, jak i w grupie DB podkreślały walor wyprofi-

lowania wierzchniej, przedniej części systemu DWKB. Zwrócono uwagę na ulepszoną konstrukcję prototypu kamizelki w stosunku do testowanego wcześniej modelu. Zasadnicze uwagi do ergonomii i funkcjonalności wynikały przede wszystkim z trudności w prawidłowym wyregulowaniu i dopasowaniu kamizelki do sylwetki, tak, aby nie unosiła się do góry. Stąd, kluczowe znaczenie ma dokładne zapoznanie się funkcjonariuszki z instrukcją zakładania, regulacji i dopasowywania systemu DWKB.

Na podstawie laboratoryjnych badań eksploatacyjnych stwierdzono, że po założeniu i prawidłowym dopasowaniu do ciała elementów systemu DWKB oraz wykonaniu przez funkcjonariuszki Policji serii badań statycznych i dynamicznych ubranych w system DWKB w wersji podstawowej oraz rozszerzonej, zapewniona jest ochrona balistyczna i mechaniczna tułowia z przodu, z tyłu i na bokach. System DWKB nie wchodzi w niepożądaną interakcję z odzieżą wierzchnią stosowaną przez funkcjonariuszki Policji, zapewnia swobodny dostęp do broni i innych elementów wyposażenia pasa taktycznego bez względu na wykonywane czynności i pozycję ciała.

5. Badania ergonomiczne i użytkowe w warunkach rzeczywistych

Badania przeprowadzono w warunkach rzeczywistych, tj. podczas wykonywania przez funkcjonariuszki Policji rutynowych czynności służbowych.

Każdej uczestniczce dostarczono komplet wyposażenia przewidzianego do badań: odpowiednio dobrany prototyp systemu DWKB, a także wstępną instrukcję użytkowania. System DWKB funkcjonariuszki testowały z własnym umundurowaniem służbowym.

Badania realizowano w dwóch powtarzalnych seriach z udziałem tych samych funkcjonariuszek. Podczas I serii Panie testowały system DWKB z kamizelką w wersji podstawowej. Natomiast w II turze – system DWKB w wersji rozszerzonej.

Czas trwania obu serii badań obejmował czas nie krótszy niż 20 zmian służbowych.

Po zakończeniu każdej serii badań funkcjonariuszki wypełniały ankietę online, w której oceniały poszczególne elementy systemu DWKB oraz jego instrukcję użytkowania. Ankieta zawierała część wstępną o charakterze danych dotyczących doświadczeń i wykonywanej pracy przez funkcjonariuszki Policji (omówioną w rozdziale 3) oraz część merytoryczną dotyczącą oceny właściwości ergonomicznych i użytkowych systemu DWKB. Ankieta była anonimowa.

5.1. Metodyka oceny statystycznej zebranych wyników ankietowych

Odpowiedzi na pytania w ankiecie miały charakter:

1. Pięciostopniowy:

   Ocena:
   1 Niezadowalająca
   2 Zadowalająca
   3 Dobra
   4 Bardzo dobra
   5 Doskonała

   Wybór odpowiedzi nr 1 „Niezadowalająca” uznawany był za odpowiedź negatywną. Wybór odpowiedzi od 2 do 5 uznawany jest za odpowiedź pozytywną.

   Ilość udzielonych i zinterpretowanych odpowiedzi na dane pytanie przeliczana była na wartość procentową.

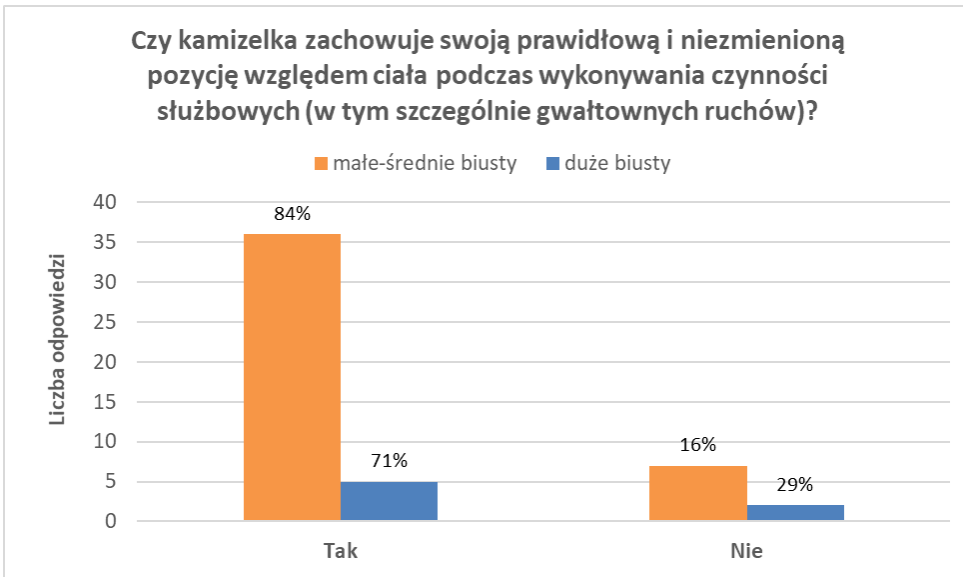
2. Pytania w skali nominalnej TAK/NIE:

Wybór odpowiedzi NIE był uznawany za odpowiedź negatywną, a wybór odpowiedzi TAK – za pozytywną.

Ilość udzielonych i zinterpretowanych odpowiedzi na dane pytanie przeliczana była na wartość procentową.

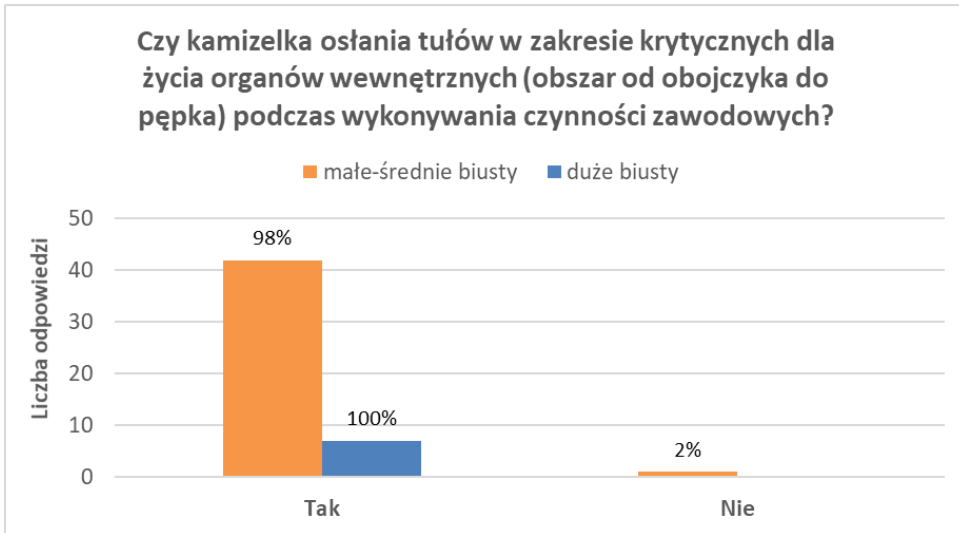
5.2. Wyniki badań

Wyniki oceny sumarycznej z I i II serii badań systemu DWKB w wersji podstawowej oraz w wersji rozszerzonej zestawiono na rysunkach 5-9.



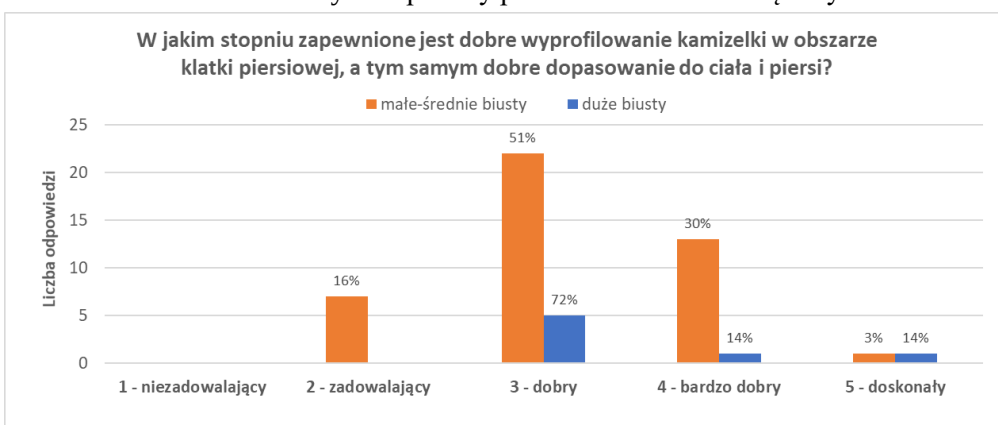
Rysunek 5. Rozkład odpowiedzi na pytanie „Czy kamizelka zachowuje swoją prawidłową i niezmienioną pozycję względem ciała użytkowniczki podczas wykonywania czynności służbowych (w tym szczególnie gwałtownych ruchów)?". Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP

Większość uczestniczek zarówno z grupy MŚB (84%) , jak i DB (71%) wskazała, że kamizelka zachowuje swoją prawidłową i niezmienną pozycję względem ciała podczas wykonywania czynności służbowych, w tym także podczas gwałtownych ruchów (rys. 5). Pozostałe 16% (tj. 7 osób) z grupy MŚB zgłosiło pojedyncze uwagi, m.in. że kamizelka przemieszczała się w trakcie biegania, pochylania się czy wsiadania do samochodu. W uwagach jedna z osób wskazała, że kamizelka znacznie ograniczała ruchy podczas kierowania pojazdem. Inna z kolei odczuwała ograniczenia ruchowe podczas zwykłego chodzenia. Wynikało to z braku kompatybilności kamizelki z bluzą służbową, gdyż po założeniu bluzy na kamizelkę funkcjonariuszka nie była w stanie jej zapiąć. Stąd wniosek, że przed użytkowaniem ocenianej kamizelki skrytego noszenia należy odpowiednio dobrać rozmiar odzieży wierzchniej, biorąc pod uwagę minimalną wielkość luzów, która zapewni swobodę ruchów. W grupie DB problem ze stabilnością kamizelki zgłosiły 2 osoby, z czego 1 wskazała, że kamizelka przemieszczała się podczas biegania.



Rysunek 6. Rozkład odpowiedzi na pytanie „Czy kamizelka osłania tułów w zakresie krytycznych dla życia organów wewnętrznych (obszar od obojczyka do pępka) podczas wykonywania czynności życiowych?". Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP

W kolejnym pytaniu weryfikowano czy kamizelka osłania tułów w zakresie krytycznych dla życia organów wewnętrznych (obszar od obojczyka do pępka) podczas wykonywania czynności zawodowych (rys. 6). Wszystkie respondentki z grupy BD oraz 98% respondentek z grupy MŚB oceniło ten aspekt pozytywnie, co w konsekwencji przekłada się na zwiększenie ochrony narządów wewnętrznych użytkowniczek. Tylko u jednej osoby kamizelka nie spełniała wymagań w tym zakresie. Respondentka wskazała, że testowana kamizelka nie okrywała połowy piersi w kierunku zewnętrznym.

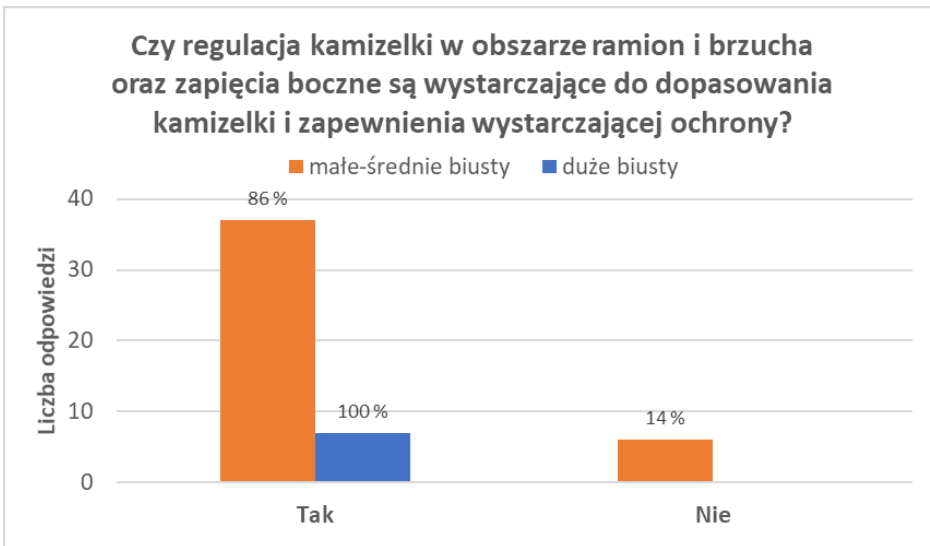


Rysunek 7. Rozkład odpowiedzi na pytanie „W jakim stopniu zapewnione jest dobre wyprofilowanie kamizelki w obszarze klatki piersiowej, a tym samym dobre dopasowanie do ciała i piersi?". Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP

Funkcjonariuszki oceniały również poprawność wyprofilowania kamizelki w obszarze klatki piersiowej, a tym samym jej dobre dopasowanie do ciała i piersi na podstawie 5-stopniowej skali, gdzie: „1" oznaczało stopień niezadawalający, „2" zadawalający, „3"

dobry, „4" bardzo dobry, a „5" doskonały (rys. 7). Najwięcej osób, bo ponad 50% (22 osoby) w grupie „małe-średnie biusty" oraz 72% (5 osób) w grupie „duże biusty" oceniło kamizelkę w aspekcie jej dopasowania do ciała użytkownika na poziomie „3". W grupie „małe-średnie biusty" 30% respondentek wystawiło ocenę „4 – bardzo dobry", 16% ocenę „2 – zadowalający" a tylko 3% (1 osoba) – ocenę „5 – doskonały". W grupie „duże biusty" po 1 osobie oceniło stopień dopasowania kamizelki do ciała na poziomie „4 – bardzo dobry" lub „5 – doskonały". W obu grupach żadna z respondentek nie wystawiła oceny najniższej („1"), wskazującej na niezadowalające wyprofilowanie kamizelki.

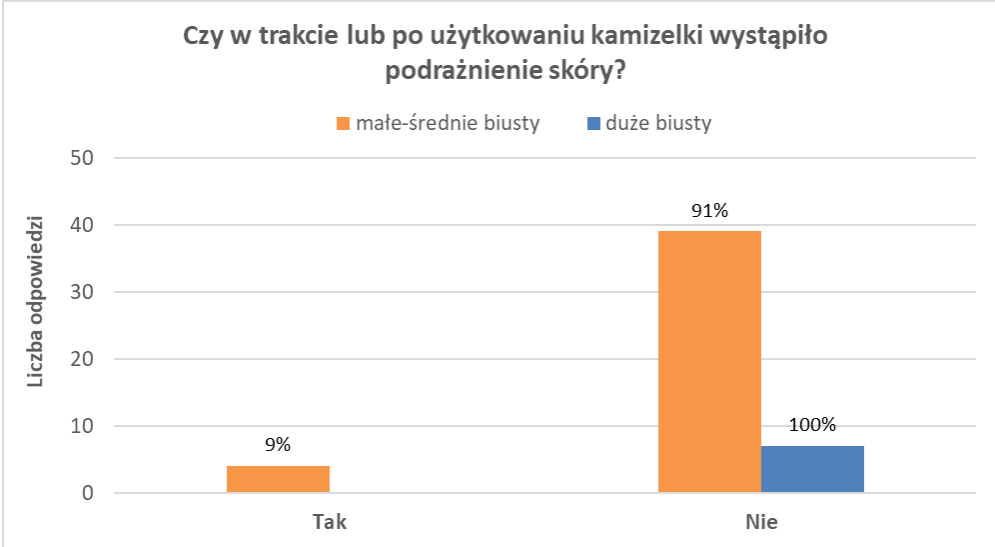
Porównując średnie oceny stopnia dopasowania badanej kamizelki do ciała można stwierdzić, że nieco lepszy wynik uzyskano w grupie „duże biusty" (ocena 3,4), aniżeli w grupie „małe-średnie biusty" (ocena 3,2). Oba wyniki, poza nielicznymi wyjątkami, świadczą o ogólnie dobrym wyprofilowaniu kamizelki w obszarze klatki piersiowej, a tym samym dobrym jej dopasowaniu do ciała i piersi.



Rysunek 8. Rozkład odpowiedzi na pytanie „Czy regulacja kamizelki w obszarze ramion i brzucha oraz zapięcia boczne są wystarczające do dopasowania kamizelki i zapewnienia wystarczającej ochrony?". Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP

Kolejne pytanie dotyczyło zakresu regulacji kamizelki. Ponad 85% respondentek z grupy „małe-średnie biusty" stwierdziło, że regulacja kamizelki w obszarze ramion i brzucha oraz zapięcie boczne są wystarczające do dopasowania kamizelki i zapewnienia wystarczającej ochrony (rys. 8). Z kolei 14% respondentek (6 osób) z tej grupy miało odmienne zdanie. Osoby te zgłosiły pojedyncze uwagi tzn. niewystarczająca regulacja w ramionach, zbyt mała regulacja w pasie, zbyt mała regulacja na brzuchu, za małe i za krótkie rzepy boczne, które zbyt słabo trzymają, kamizelka jest zbyt sztywna przez co trudno ją dopasować poprzez zapięcia boczne.

W grupie „duże biusty" wszystkie respondentki stwierdziły, że regulacja kamizelki w obszarze ramion i brzucha oraz zapięcie boczne są wystarczające do dopasowania kamizelki i zapewnienia wystarczającej ochrony.



Rysunek 9. Rozkład odpowiedzi na pytanie „Czy w trakcie lub po użytkowaniu kamizelki wystąpiło podrażnienie skóry?". Źródło: opracowanie własne na podstawie dokumentacji projektowej CIOP-BIP

W ostatnim pytaniu dotyczącym systemu DWKB funkcjonariuszki miały ocenić, czy w trakcie jego użytkowania lub po wystąpiło podrażnienie skóry. Zdecydowana większość respondentek nie stwierdziła negatywnego wpływu testowanej kamizelki na skórę (rys. 9). Spośród 50 uczestniczek badania tylko 4 osoby (z grupy MŚB) zgłosiły wystąpienie zmian skórnych w wyniku użytkowania kamizelki: otarcia na biodrach i przy ramionach w okolicy pach, podrażnienia spowodowane przez zapięcia boczne lub dolną część kamizelki, co mogło być wynikiem niewłaściwego doboru rozmiaru kamizelki do sylwetki danej uczestniczki.

5.3. Wnioski

W wyniku badań ergonomicznych i użytkowych dokonano oceny prototypu systemu DWKB pod względem ergonomii, funkcjonalności i bezpieczeństwa użytkowania przez docelową grupę odbiorców – funkcjonariuszki Policji podczas wykonywania rutynowych czynności służbowych.

Na podstawie szczegółowej analizy wyników badań stwierdzono, że opracowany system DWKB zarówno w wersji podstawowej, jak i rozszerzonej:

- posiada właściwości ograniczające wstrząsy i stabilizujące biust, co pozwala na zwiększenie dopasowania kamizelki;
- zapewnia całkowitą ochronę najbardziej newralgicznych narządów wewnętrznych; co wpływa na zwiększenie właściwości ochronnych niniejszego rozwiązania;
- zapewnia długotrwały komfort użytkowania.

6. Podsumowanie

Wyniki badań ergonomicznych i użytkowych przeprowadzonych zarówno w warunkach laboratoryjnych (symulujących warunki rzeczywiste), jak również w warunkach rzeczywistych prototypu systemu DWKB potwierdziły, iż prototyp systemu DWKB

umożliwia funkcjonariuszkom Policji realizację działań interwencyjno-patrolowych w różnych warunkach operacyjnych.

Potwierdzono dopasowanie, ergonomiczność oraz kompatybilność systemu DWKB stanowiącego pod względem funkcjonalnym jednolity zestaw wyposażenia, umożliwiający kobietom wykonywanie obowiązków służbowych w różnych warunkach operacyjnych.

Przebadany systemu zindywidualizowanej Damskiej Wielofunkcyjnej Kamizelki Balistycznej DWKB skrytego noszenia dla dwóch podgrup rozmiarowych: małe i średnie biusty MŚB oraz duże biusty DB osiągnął IX poziom gotowości technologicznej.

Literatura

1. Kucharska-Jastrząbek A., Suchocki P., Pieniążek D., Garus R., Dmowska-Jasek P., Kudlińska M., Olejnik M., Woźniakowska M., Gutowska A., *VESTA – damska kamizelka skrytego noszenia – od pomysłu do projektu*, [w:] Mołdoch-Mendoń I., Maciąg K. (red.), *Rozwiązania technologiczne XXI wieku – skutki i perspektywy rozwoju. Tom 4*, Wydawnictwo Naukowe TYGIEL Sp. z o.o., Lublin 2023.
2. PN-V-87000:2011 Osłony balistyczne lekkie. Kamizelki kulo- i odłamkoodporne. Wymagania ogólne i badania.
3. NIJ Standard 0115.00 Stab Resistance of Personal Body Armor.
4. VPAM KDIW 2004 Test Standard Stab and Impact Resistance. Requirements, classifications and test procedures.
5. PBB-12/ITB „MORATEX" ed. I z 12.2008 w zakresie odporności na uderzenie igły typ 18G (1,2mm x 40mm) i typ 21G (0,8mm x 40mm) z energią 25,0 ±0,5 J – poziom 1 lub 2 (w oparciu o VPAM KDIW 2004 [11]).

Indywidualizacja damskiej kamizelki skrytego noszenia pod kątem ergonomii użytkowania

Streszczenie

Funkcjonariuszki służb mundurowych, na chwilę obecną korzystają z osłon tułowia projektowanych na sylwetkę męską lub tzw. typu unisex. Przekłada się to na brak idealnego dopasowania kamizelki do kształtu ciała funkcjonariuszki, a w konsekwencji może powodować zwiększenie ryzyka wystąpienia urazów wewnętrznych podczas uderzenia penetratora. Ponadto, wpływa to również na odczuwanie dyskomfortu fizycznego podczas pełnienia obowiązków służbowych.

W związku z powyższym, zidentyfikowano potrzebę opracowania zindywidualizowanej kamizelki skrytego noszenia dla funkcjonariuszek służb mundurowych spełniającej zarówno wymagania bezpieczeństwa, jak również zapewniającej odpowiedni komfort podczas użytkowania w trakcie pełnienia służby. W tym celu, wytypowano grupę badawczą spośród funkcjonariuszek z całej Polski, które następnie poddano procesowi skanowania 3D, w celu jak najdokładniejszego odwzorowania kobiecej sylwetki. Umożliwiło to opracowanie specjalistycznej tabeli rozmiarowej, na podstawie której zaprojektowano i wytworzono rozwiązania uwzględniające zarówno wymiary, jak i obwody funkcjonariuszek, a tym samym umożliwiające jak najdokładniejsze dopasowanie do kobiecej sylwetki.

Kluczowym badaniem celem potwierdzenia właściwości funkcjonalnych i użytkowych zaprojektowanej damskiej kamizelki skrytego noszenia było przeprowadzenie badań ergonomicznych symulujących warunki rzeczywiste z udziałem docelowych użytkowniczek.

Słowa kluczowe: damska kamizelka balistyczna, damska kamizelka skrytego noszenia, badania ergonomiczne, dopasowanie

Individualization of the women's concealed vest in terms of ergonomics of use

Abstract

Officers of uniformed services currently use ballistic vests designed for a male figure or so-called unisex type. This translates into the lack of a perfect fit of the vest to the shape of the officer's body, and as a consequence, it may increase the risk of internal injuries during the impact of the penetrator. In addition, it also affects the feeling of physical discomfort while performing official duties.

In connection with the above, a need was identified to develop an individualized concealed vest for officers of uniformed services that meets both safety requirements and ensures appropriate comfort during use while on duty. For this purpose, a research group was selected from among female officers from all over Poland, who were then subjected to a 3D scanning process in order to reproduce the female figure as accurately as possible. This made it possible to develop a specialized size table, on the basis of which solutions were designed and manufactured, taking into account both the dimensions and circumferences of female officers, and thus enabling the most accurate fit to the female figure.

The key study to confirm the functional and operational properties of the designed women's concealed vest was to conduct ergonomic tests simulating real conditions with the participation of target users.

Keywords: women's ballistic vest, women's concealed vest, ergonomic tests, fit

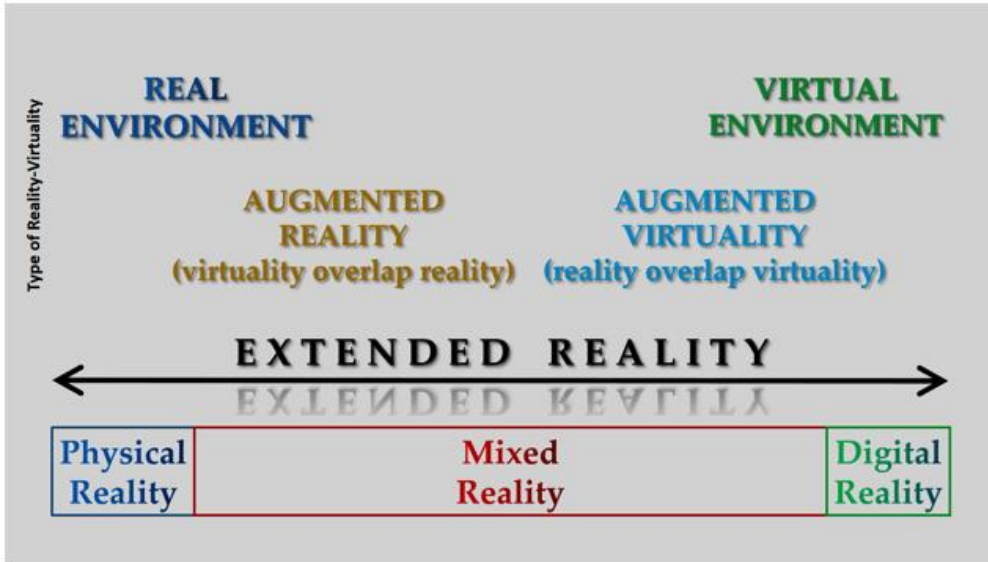
Beata Sokołowska[1]

Przykłady oceny parametrów i/lub markerów w modelowym podejściu z klasyfikacją nadzorowaną w badaniach w realnej i wirtualnej rzeczywistości

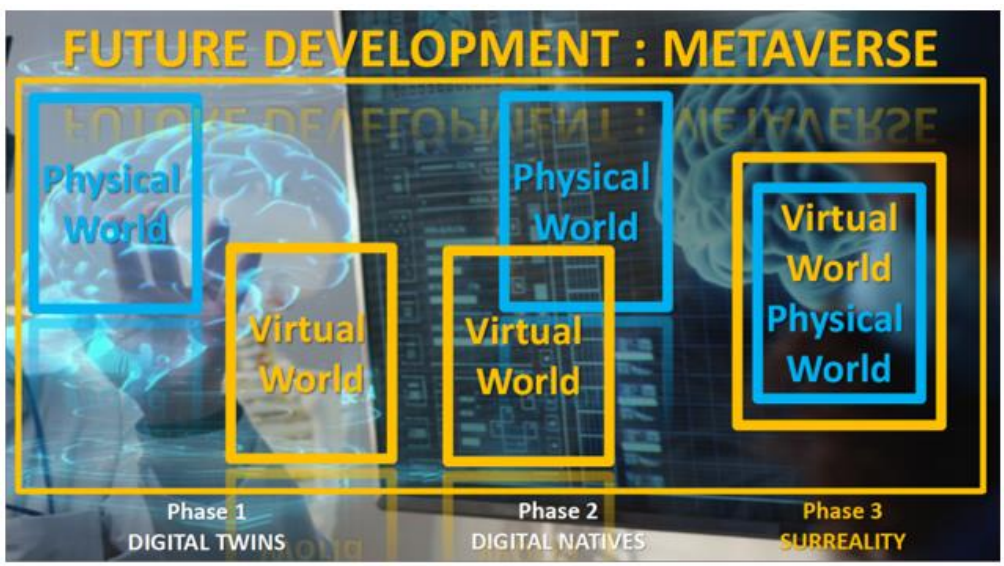
1. Wstęp

Ocena różnorodnych parametrów fizjologicznych i innych istotnych zmiennych jako (bio)markerów odgrywa ważną rolę w badaniach modelowych z udziałem osób zdrowych, jak również we współczesnej praktyce medycznej, w diagnostyce, prognostyce, w podejmowanej terapii i rehabilitacji [1-6]. Niezwykle szybki rozwój nowoczesnych technik i technologii informatycznych wraz z globalizacją Internetu i jego siecią komunikacyjno-społecznościową jest niezwykłym i obiecującym wyzwaniem dla dotychczasowych osiągnięć i odkryć w obszarze podstawowych i klinicznych neuronauk, takich jak neurobiologia, neurologia, neuropsychologia, neuropsychiatria, neuropedagogika, neurogeriatria, neuroterapia i neurorehabilitacja, czy też (neuro)(bio)informatyka. Kreowane niezwykłe wirtualne rzeczywistości, ze świata gier, reklamy, przemysłu i kultury, już na dobre wkroczyły do świata nauki i edukacji [7-11]. Stają się one istotną częścią prowadzonych badań w obszarze neuronauk podstawowych i klinicznych XXI wieku [12-16]. Obok metod opartych na dotychczasowych klasycznych „realnych" narzędziach, proponowane są nowe rozwiązania i techniki oraz środowiska odmienne od tych dotychczas znanych z realnego świata, a pochodzące właśnie z nowych kreowanych wirtualnych rzeczywistości [17-19]. Zgodnie z klasyczną koncepcją przedstawioną przez *Miligrama, Takemura, Utsumi i Kishino* [20,21] określane są one jako rozszerzona rzeczywistość (ang. *Extended Reality*, EX) z jej podstawowymi (klasycznymi) składowymi, takimi jak wirtualna i mieszana rzeczywistość (ang. *Virtual and Mixed Reality*, VR and MR), która obejmuje pogłębioną rzeczywistość i pogłębioną wirtualność (ang. *Augmented Reality and Augmented Virtuality*, AR and AV), choć ich obszary mogą się pokrywać, jak ilustruje to rysunek 1 [22]. Dalszy i niezwykle dynamiczny rozwój rzeczywistości wirtualnej i rozszerzonej oferuje najnowszą koncepcję cyfrowej przyszłości znanego nam świata, której celem jest udoskonalenie wirtualnego doświadczenia i stworzenie złożonego, interaktywnego i połączonego z realnym środowiskiem, świata określanego jako *Metaverse* [23-26]. Byłoby to cyfrowe środowisko do komunikowania się, zakupów, nauki i pracy oraz robienia wszystkiego, co normalnie byłoby planowane i realizowane na przykład w Internecie [27-31]. Rysunek 2 ilustruje tę nową koncepcję w ramach kolejnych obserwowanych etapów rozwoju IT/ICT (ang. *Information Technology and Information Communication Technology*). I choć realizowane są już projekty z nią związane, wielu badaczy wskazuje na wyłaniające się ograniczenia i możliwe zagrożenia [32-34].

[1] beta.sokolowska@imdik.pan.pl, Pracownia Bioinformatyki, Instytut Medycyny Doświadczalnej i Klinicznej im. Mirosława Mossakowskiego Polskiej Akademii Nauk (IMDiK PAN), https://imdik.pan.pl/pl/.



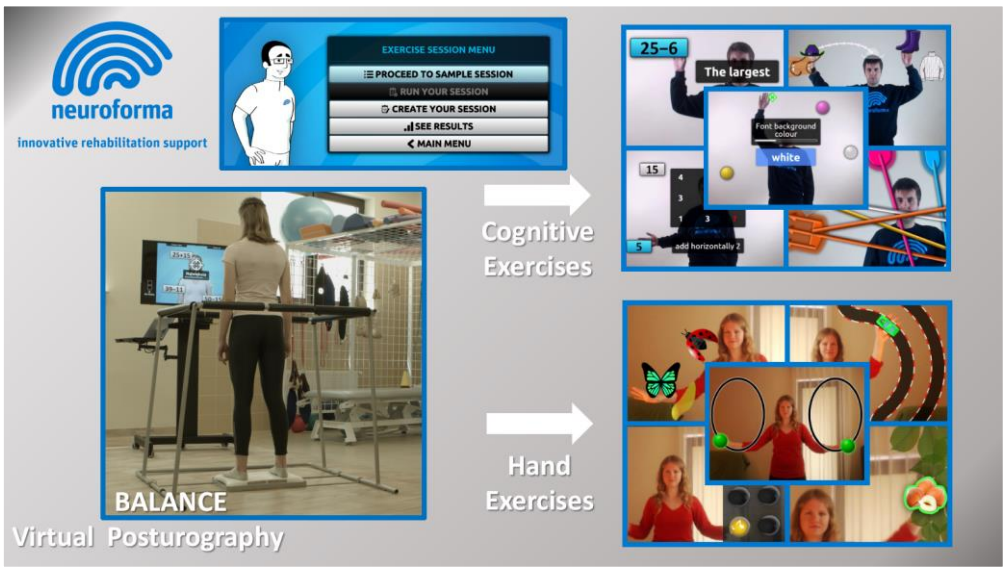
Rysunek 1. Ilustracja klasycznej koncepcji Milgrama i współpracowników [20, 21] dla rozszerzonej rzeczywistości i jej podstawowych składowych (opracowanie własne) [22]



Rysunek 2. Ilustracja koncepcji cyfrowej przyszłości, czyli środowiska „*Metaverse*", jako konsekwencji kolejnych etapów dynamicznego rozwoju IT/ICT: od wcześniej niezależnych od siebie świata realnego – fizycznego i cyfrowego – wirtualnego (faza 1), po obecnie już nakładający się świat realny i wirtualny, w którym funkcjonuje współczesny człowiek (faza 2), po kreację przyszłego surrealistycznego środowiska *Metaverse* (faza 3) (opracowanie własne)

Celem pracy jest przedstawienie wyników prowadzonych badań modelowych, eksperymentalnych i klinicznych dotyczących oceny parametrów i markerów z zastosowaniem m.in. modelowego podejścia z klasyfikacją nadzorowaną, realizowane w realnej i wirtualnej rzeczywistości. Podejście to, opierające się na znajomości przynależności

badanych obiektów do zdefiniowanych klas i/lub grup (stąd określenie klasyfikacja nadzorowana) jest bardzo skutecznym i dokładnym narzędziem, które znakomicie pomaga zarówno badaczom i klinicystom, jak i praktykom w ich pracy w wielu obszarach aktywności współczesnego człowieka, zarówno w warunkach zdrowia jak i choroby. Stanowisko wirtualnej rzeczywistości, utworzone w Pracowni Bioinformatyki IMDiK PAN, które również zastosowano w prowadzonych badaniach, prezentuje rysunek 3.

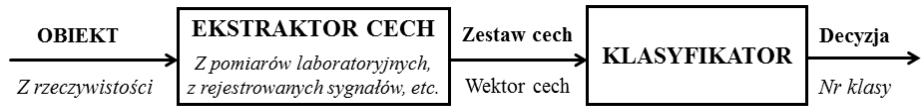


Rysunek 3. Przykład hybrydowego systemu NEUROFORMY do (tele) (neuro)rehabilitacji z nieimersyjną wirtualną rzeczywistością w technologii firmy Titanis. Podobne stanowisko badawcze zostało utworzone w Pracowni Bioinformatyki IMDiK PAN (za zgodą firmy Titanis Ltd.) [35, 36]

2. Statystyczne systemy uczące się z klasyfikacją nadzorowaną zastosowane w badaniach własnych

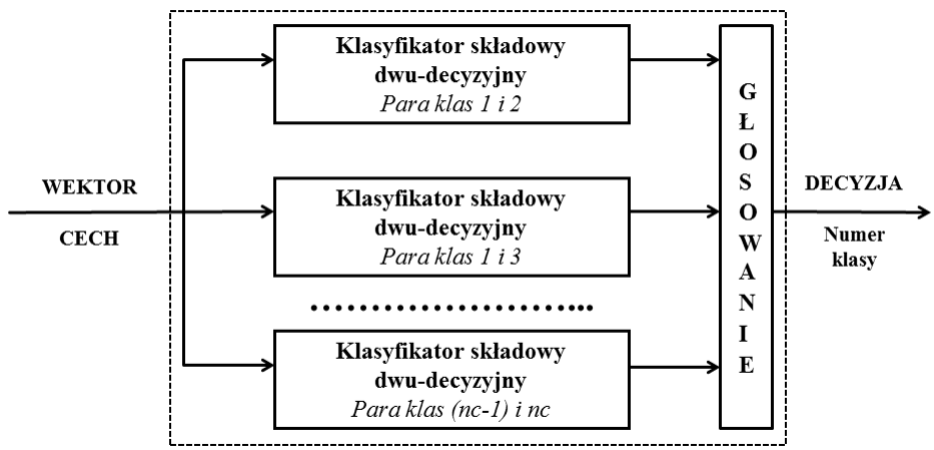
Klasyfikacja obiektów jest przedmiotem badań dziedziny określanej jako „rozpoznawanie obiektów", „rozpoznawanie obrazów" lub „rozpoznawanie wzorców" *(*ang. *statistical pattern recognition*), na przykład *Jain i współpracownicy* [37] dokonali przeglądu definicji, modeli oraz zastosowań takiego podejścia badawczo-analitycznego. Przyjmuje się, że obiekty opisane zestawem cech (wektorem cech), mogą być zmiennymi ilościowymi, jak i jakościowymi lub binarnymi. *Wektorom cech odpowiadają punkty w przestrzeni cech, którą najczęściej jest przestrzeń euklidesowa*.

W badaniach własnych, w przyjętym modelu klasyfikacji z zastosowaniem metod rozpoznawania obrazów, oparto się na *klasyfikacji nadzorowanej*, której reguły decyzyjne wyprowadzane są ze *zbioru uczącego*. Zbiór uczący to *zbiór obiektów ze znaną przynależnością do klas*. Proces obliczeniowy prowadzący do konstrukcji określonego klasyfikatora nazywany jest „*uczeniem*". W układzie rozpoznającym (w podejściu modelowym przedstawionym na rysunku 4) wyróżniamy dwa podstawowe procesy: (a) wydzielania cech (ekstraktor cech) oraz (b) konstrukcję klasyfikatora, tj. przygotowanie odpowiedniego algorytmu, procedury postępowania lub programu komputerowego, którego celem jest podjęcie decyzji co do przypisania badanego obiektu do określonej klasy.



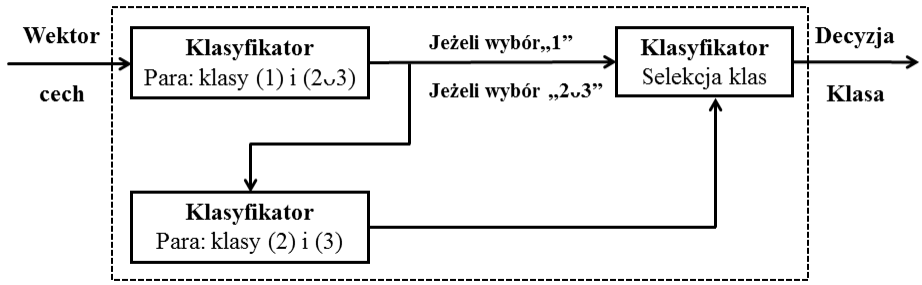
Rysunek 4. Schemat modelu klasyfikacji/ układu rozpoznającego (opracowanie własne)

Zauważmy, że klasyfikator realizuje określoną regułę decyzyjną, na podstawie której wskazywana jest jedna z dwóch klas, bądź też jedna z wielu rozważanych klas *nc* (ang. *number class*). W przypadku, dwóch klas lub dla reguły *wielo-decyzyjnej,* ale ograniczonej do pary klas, taki klasyfikator określany jest jako dwu-decyzyjny. Klasyfikatory *dwu-decyzyjne* można użyć do konstrukcji klasyfikatorów *wielo-decyzyjnych,* jak przedstawiono na rysunku 5. Wówczas klasyfikator równoległy będzie złożony z klasyfikatorów dwu-decyzyjnych, odpowiadających liczbie możliwych par klas (równej *nc*(nc-1)/2*). W takim modelu, każdy ze składowych klasyfikatorów *dwu-decyzyjnych* wskazuje poprzez głosowanie na jedną z dwóch klas, i wówczas globalny klasyfikator równoległy wskazuje na tę klasę, która otrzymała najwięcej „głosów" od klasyfikatorów składowych.



Rysunek 5. Schemat struktury klasyfikatora równoległego, utworzonego przez składowe dwu-decyzyjne klasyfikatory dla par klas, i finalnego głosowania nad wyborem właściwej klasy spośród *nc* rozważanych klas (taki układ modelowy opisano i zastosowano w badaniach własnych (np. w pracy [38])

W przypadku większej liczby klas niż dwie, możliwa jest również struktura hierarchiczna klasyfikatora, tj. *wielo-etapowa*, z utworzeniem „*makroklas*" łączących pewne klasy w zbiory klas, według ustalonego kryterium klinicznego lub modelu eksperymentalnego. Wówczas na wstępnym etapie następuje klasyfikacja do *makroklas,* na kolejnym klasyfikacja z wyborem pomiędzy klasami w obrębie każdej z *makroklas*, zaś na finalnym etapie ostateczna decyzja dotyczy już przynależności do pojedynczych klas. Hierarchiczną sekwencję klasyfikacji można otrzymać również na podstawie struktury równoległej poprzez zorganizowanie hierarchicznego głosowania składowych klasyfikatorów, jak zilustrowano na rysunku 6. Testowano ten wariant modelu klasyfikacji w badaniach własnych, np. w pracach [38, 39].



Rysunek 6. Struktura klasyfikatora hierarchicznego dla trzech klas, użytego w badaniach własnych z udziałem pacjentów ze stwardnieniem zanikowym bocznym [38]. Tutaj, makroklasę tworzą dwie grupy chorych (2ᴗ3). W pierwszym etapie klasyfikowane są obiekty-osoby badane do klasy 1 – zdrowej kontroli, a w następnym etapie klasyfikacja dotyczy przynależności już do określonej grupy pacjentów, klasy 2 – z łagodnymi objawami i klasy 3 – z poważnymi objawami choroby

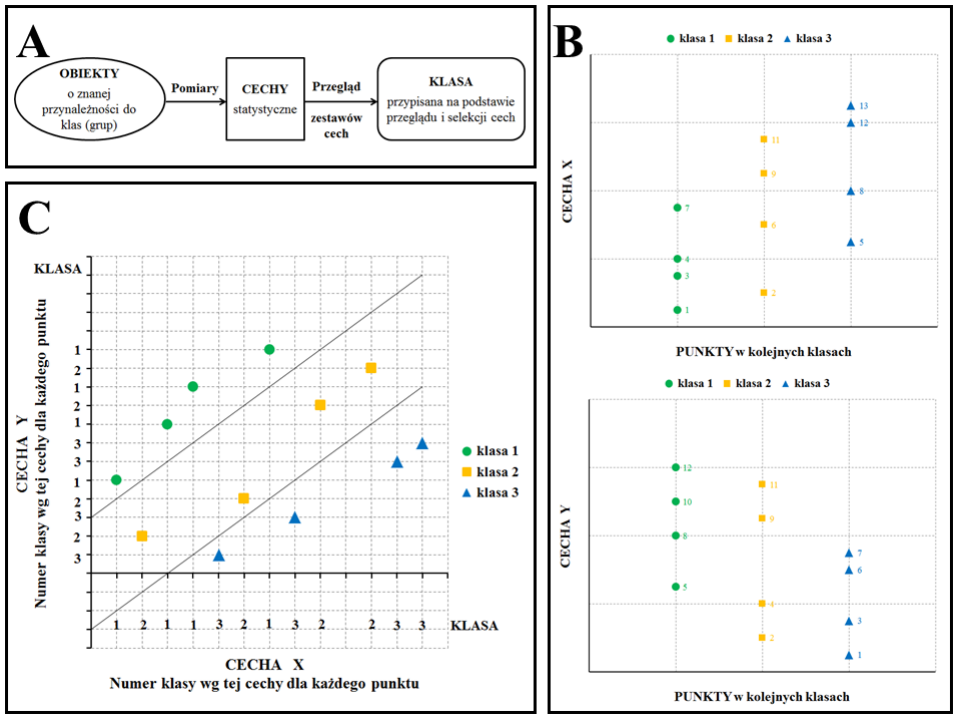
Obok założenia, że obiekt może należeć do jednej z wielu rozważanych *nc* klas (tzw. *klasyfikacja ostra*), możliwe jest zaproponowanie klasyfikacji definiowanej jako *klasyfikacja rozmyta (nieostra)* [40, 41]. Jej szczególnym przypadkiem jest właśnie klasyfikacja ostra, której składowe wektora przynależności przyjmują wartości binarne (tylko dwie) *0 lub 1*. Model klasyfikacji wg (uogólnionej) reguły rozmytej, po raz pierwszy w badaniach nad układem oddechowym, zastosowano w eksperymentalnych badaniach własnych w pracy [42], później wykorzystano go w badaniach klinicznych, dla oceny stanu pacjentów ze stwardnieniem zanikowym bocznym [43]. Model klasyfikacji rozmytej ma zastosowanie w przypadkach znaczącego pokrywania się obszarów rozważanych klas.

W procesie decyzyjnym dotyczącym przynależności do określonej klasy lub klas, kolejnym ważnym etapem jest *ocena jakości* (przeprowadzonej) klasyfikacji. Podstawowym kryterium oceny jakości klasyfikacji *jest prawdopodobieństwo mylnej decyzji*, szacowane frakcją (odsetkiem) błędnych decyzji. Naturalną metodą oszacowania prawdopodobieństwa mylnej decyzji klasyfikatora, jest zastosowanie dwóch zbiorów: *zbioru testującego T i zbioru uczącego U*. Wszystkie obiekty ze zbioru *T*, których jest m_T, klasyfikowane są zgodnie z regułą decyzyjną wyprowadzoną ze zbioru uczącego *U*. Liczba obiektów mylnie klasyfikowanych, e_T, podzielona przez liczbę wszystkich obiektów zbioru testującego *T*, m_T, czyli frakcja błędów $E_{rT}=e_T/m_T$ (ang. *error rate, E_r)* jest oszacowaniem *prawdopodobieństwa mylnej klasyfikacji*. W przypadku, gdy dysponujemy jednym zbiorem obiektów o znanej przynależności do klas, możemy podzielić ten zbiór na część uczącą *U* oraz część testującą *T*. Kolejnym rozszerzeniem tej metody jest metoda *l*-krotnej walidacji, która polega na podzieleniu zbioru *U* na *l* równych części. Każda z tych *l* części pełni kolejno rolę zbioru testującego. Szczególnym przypadkiem metody *l*-krotnej walidacji dla *l*=1 jest metoda minus jednego elementu (ang. *leave one out method*). Metodę tę stosuje się (i jest ona polecana) dla małych zbiorów uczących [44]. Ogólnie procedura dla metody minus jednego elementu polega na klasyfikacji każdego obiektu u_i ze zbioru uczącego *U*, zawierającego m_U obiektów, z użyciem klasyfikatora „wyprowadzonego” ze zbioru $U-\{u_i\}$ (tj. zbioru pomniejszonego o klasyfikowany obiekt). Podobnie postępujemy w znanej metodzie „*jackknifing*”, zaś jej zaletą jest fakt, że korzystamy wówczas z kompletnego zbioru danych, bez konieczności tworzenia podzbiorów. Oszacowaniem prawdopodobieństwa mylnej decyzji jest frakcja $E_{rU}=e_U/m_U$, gdzie e_U to

liczba popełnionych błędów. Frakcja błędów jest standardowym kryterium oceny klasyfikatora, z kolei informacja o rodzaju błędów jest przedstawiana jako *macierze przekłamań R, prawdopodobieństw P lub wiarygodności Q*. Macierze *R, P i Q,* zastosowano i opisano na przykład w badaniach własnych [38] u pacjentów ze stwardnieniem zanikowym bocznym (ang. *amyotrophic lateral sclerosis, ALS*).

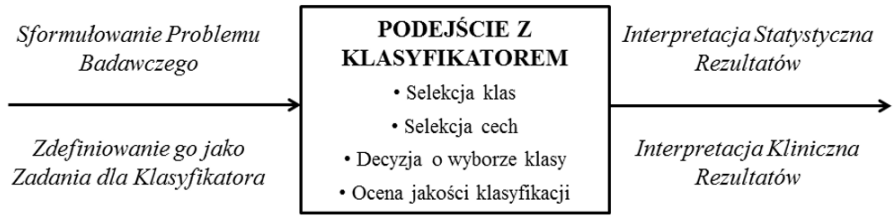
W przedstawionym modelu z klasyfikacją nadzorowaną w rozpoznawaniu obiektów, kolejnym i ważnym etapem jest *selekcja cech*. Procedura selekcji cech jest istotna w ocenie parametrów jako markerów, i została uwzględniona w proponowanym i stosowanym modelowym podejściu z klasyfikacją nadzorowaną w badaniach własnych (rys. 7A). Warto podkreślić, że spośród analizowanych cech, część z nich może nie mieć związku z rozważanymi klasami, jak też mogą to być cechy „nadmiarowe", przez co mogą wpływać na „pogorszenie" jakości klasyfikacji. Ilustruje to rysunek 7, na przykładzie *dwóch cech X i Y*, które pojedynczo nie mają żadnego znaczenia w rozpoznaniu trzech klas: klasy 1 (kółka), klasy 2 (prostokąty) i klasy 3 (trójkąty), ale razem w zestawie tworzą idealną kombinację (wektor) cech, oferując najlepszą jakość klasyfikacji, czyli perfekcyjne rozpoznanie klas, $E_r = 0$. Najlepszą i najskuteczniejszą procedurą dla selekcji cech jest *przegląd wszystkich możliwych kombinacji spośród n cech,* i dla każdej z nich oszacowanie prawdopodobieństwa mylnej decyzji (E_r) oraz wybór spośród wszystkich kombinacji tej, która oferuje najniższą wartość E_r. Warto podkreślić, że liczba możliwych kombinacji cech rośnie wraz ze wzrostem liczby *n* cech i równa jest 2^n-1. W badaniach własnych, w przyjętym modelu klasyfikacji, wykorzystano procedurę przeglądu wszystkich kombinacji cech, jako że liczba analizowanych cech, czyli parametrów/ markerów nie była aż tak duża, by wymagała oszczędnych obliczeniowo procedur przeglądu. Do takich szybkich i „oszczędnych" procedur należą algorytmy kolejnego dołączania lub odrzucania cech, bądź kombinacja obu tych procedur [44]. W badaniach własnych, na przykład w pracy [45], testowano takie podejście w przypadku proponowanych różnych modeli eksperymentalnych, co w efekcie pozwoliło na opracowanie skuteczniejszych procedur doświadczalnych. Ma to istotne znaczenie, na przykład, w warunkach braku określonych metod standaryzacji proponowanych/ stosowanych protokołów w warunkach eksperymentu biologicznego bądź klinicznego.

Kolejnym ważnym zagadnieniem, na które warto zwrócić uwagę jest pytanie o „status" nowych obiektów w zbiorze uczącym U. Klasyfikacja takich nowych obiektów, spoza aktualnego (bieżącego) zbioru uczącego, jest przeprowadzana na podstawie klasyfikatora skonstruowanego z wykorzystaniem pełnego (czyli już uzupełnionego o te obiekty) zbioru uczącego *U*. Nowe obiekty (również o znanej przynależności do klas), są włączane do zbioru uczącego, będąc podstawą do *„douczenia" klasyfikatora* i wyprowadzenia nowej reguły klasyfikacji. Taką strategię również zastosowano w badaniach prowadzonych we współpracy z klinicystami [46-48].



Rysunek 7. Ilustracja modelowej sytuacji w podejściu z klasyfikacją nadzorowaną dla obiektów pochodzących z trzech różnych klas [47]. Dwie cechy *X* i *Y* pojedynczo są nieistotne dla rozpoznania rozważanych klas, lecz wspólnie tworzą zestaw perfekcyjnie różnicujący trzy klasy: kółek – klasa 1, prostokątów – klasa 2 i trójkątów – klasa 3. Panel A to przyjęty model klasyfikacji z włączeniem przeglądu zestawów cech (w tym przypadku są to *{X}, {Y} i {X,Y}*. Panel B to wartości cechy *X* (górny panel) i cechy *Y* (dolny panel) w każdej z klas 1, 2 i 3. Obszary rozważanych klas znacząco pokrywają się, co ilustruje rzutowanie punktów (obiektów) na obie osie. Klasy są wówczas bardzo trudno rozdzielne i użycie pojedynczo którejkolwiek z tych cech nie ma sensu. Panel C prezentuje, iż analiza obu cech razem pozwala już na bezbłędnie rozpoznanie przynależności każdego obiektu do jego właściwej klasy (opracowanie własne)

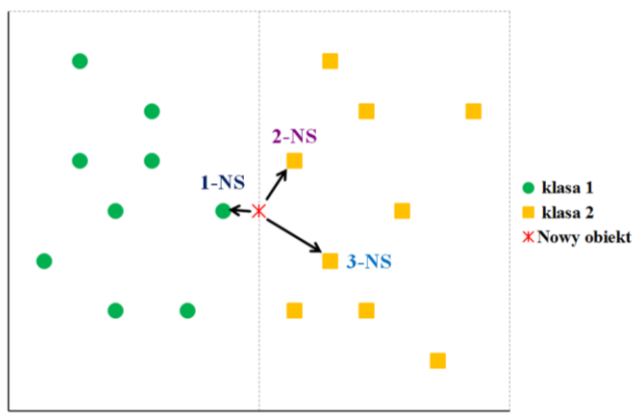
Na rysunku 8 podsumowano dotychczasowe rozważania na temat podejścia badawczo-analitycznego z zastosowaniem metod klasyfikacji nadzorowanej, które zastosowano w badaniach własnych w realnej rzeczywistości, a obecnie kontynuowanych również w modelowych badaniach z wykorzystaniem stanowiska NEUROFORMY i kreowanej nieimersyjnej wirtualnej rzeczywistości (rys. 3). Zauważmy, że w przedstawionym podejściu realizowany cel badawczy był dodatkowo formułowany jako konkretne zadanie dla klasyfikatora. Wówczas wyniki interpretowano w kontekście ich istotności statystycznej i najniższych błędów mylnej klasyfikacji dla oceny znaczenia parametrów jako potencjalnych markerów, skuteczności rozpoznawania klas (np. w celach diagnostycznych i prognostycznych) oraz znaczenia (istotności) otrzymanych rezultatów dla rozwoju prowadzonych badań i potencjalnych korzyści, na przykład dla kliniki [46-48].



Rysunek 8. Proponowane podejście badawczo-modelowe z zastosowaniem klasyfikacji nadzorowanej zastosowane w badaniach własnych (opracowanie własne)

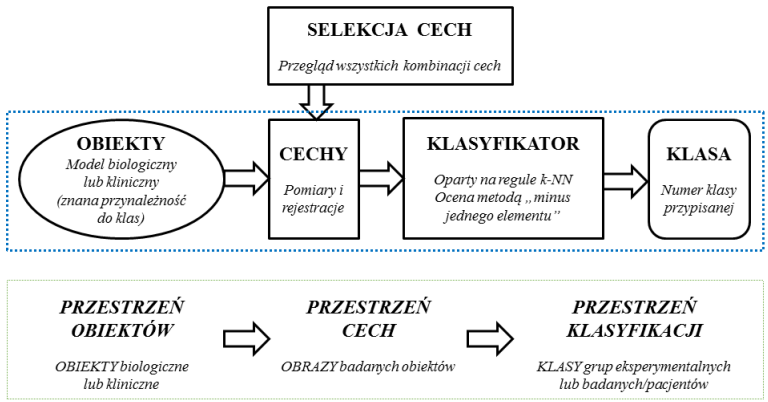
2.1. Klasyfikator k najbliższych sąsiadów

W prezentowanym podejściu badawczo-modelowym z wykorzystaniem klasyfikacji nadzorowanej zastosowano, najefektywniejszy spośród klasyfikatorów, *klasyfikator działający według reguły decyzyjnej k najbliższych sąsiadów* (ang. *k nearest neighbor*, k-NN) [49, 50], który należy do ogólniejszej grupy klasyfikatorów minimalno-odległościowych. Klasyfikator *k-NN* oferuje wysoką jakość klasyfikacji, w sensie prawdopodobieństwa poprawnej decyzji [51, 52]. Klasyfikator *k-NN* jest traktowany jako aproksymacja klasyfikatora bayesowskiego, dla którego prawdopodobieństwa nie są szacowane bezpośrednio, lecz z wykorzystaniem wzoru Bayes'a. Ponadto, klasyfikatory działające wg reguły *k-NN* są stosowane dla dowolnej liczby klas. Liczba *k najbliższych sąsiadów* wyznaczana jest eksperymentalnie z zastosowaniem jako kryterium frakcji błędów z użyciem „metody minus jednego elementu".



Rysunek 9. Ilustracja działania standardowej reguły *k-NN* i przypisania nowego obiektu (symbol gwiazdki) do określonej klasy na podstawie reguły: *1*-najbliższego sąsiada (*1-NS*), *2*-najbliższych sąsiadów (*2-NS*) oraz *3*-najbliższych sąsiadów (*3-NS*) (opisana np. w pracy [53])

Na rysunku 9 zilustrowano działanie standardowej reguły decyzyjnej *k-NN*. Klasyfikator działający wg reguły *k-NN* przyporządkowuje danemu obiektowi klasę, z której pochodzi większość spośród jego *k* najbliższych obiektów w zbiorze uczącym. Z kolei, rysunek 10 przedstawia modelowe podejście z klasyfikatorem, które w badaniach własnych oparto właśnie na regule decyzyjnej *k-NN*.



Rysunek 10. Ilustracja modelowego podejścia badawczo-analitycznego z klasyfikatorem *k-NN* zastosowanego w badaniach własnych, np. w [47, 48, 53] (opracowanie własne)

2.2. Krzywe ROC i binarny klasyfikator

W badaniach własnych, w części odnoszącej się do statystycznej analizy, włączono także znaną klasyczną metodę *krzywych ROC, ang. Receiver Operating Characteristic curve* [54-57]. Metoda wykreślania i analizy krzywych *ROC* oparta jest na klasyfikatorze binarnym i zbiorze *obiektów (np. osób badanych)* o znanym, czyli *prawdziwym statusie* przynależności do jednej z dwóch klas. Status ten jest definiowany na podstawie wyników dla „*testu złotego standardu*" bądź innego „*uznanego*" testu. Porównując wyniki dla tego testu z wynikami dla proponowanego nowego (nieznanego) testu, możemy porównać skuteczność ich obu, na podstawie analizy utworzonej tablicy kontyngencji *2 x 2*, przedstawionej w tabeli 1.

Tabela 1. Sposób przedstawiania i zliczania wyników klasyfikacji dla klasyfikatora binarnego (opracowanie własne)

DECYZJE	Prawdziwe rozpoznanie/ diagnoza badacza (uznany test)		
Nowy test (klasyfikator/algorytm)	Zmiana/ Choroba	Kontrola/ Norma	Suma
Zmiana/Choroba	*A (TP)*	*B (FP)*	*A+B*
Kontrola/Norma	*C (FN)*	*D (TN)*	*C+D*
Suma	*A+C*	*B+D*	*A+B+C+D*

Przyjęte oznaczenia:

A – przypadki zakwalifikowane przez nowy algorytm rozpoznawania i badacza jako zmiana, co oznacza liczbę przypadków prawdziwie dodatnich (ang. true positive, TP),

B – przypadki zaklasyfikowane przez algorytm jako norma, a przez badacza (lub lekarza) jako zaburzenie, są to przypadki fałszywie dodatnie (ang. false positive, FP), np. liczba osób zdrowych, błędnie zaklasyfikowanych jako osoby chore,

C – przypadki uznane przez klasyfikator za zaburzenie, a przez test diagnostyczny za normę, a zatem są to przypadki fałszywie ujemne (ang. false negative, FN), np. liczba chorych, których nowy algorytm błędnie sklasyfikował jako osoby zdrowe,

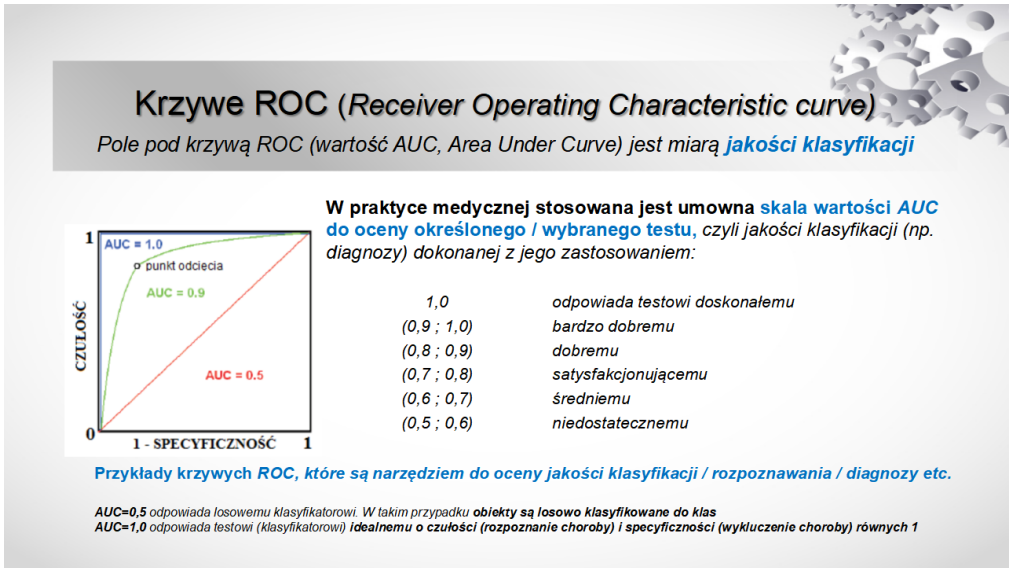
D – przypadki klasyfikowane jako norma przez klasyfikator oraz badacza, zatem są to przypadki prawdziwie ujemne (ang. true negative, TN), czyli zdrowi poprawnie sklasyfikowani jako zdrowi,

(Przy ostatecznej klasyfikacji warto zwrócić uwagę, aby możliwie najmniej przypadków pozostało w polu C, przy rozsądnej liczbie przypadków w polu B).

W prezentowanym podejściu analitycznym z zastosowaniem klasyfikatora binarnego ważne są trzy wskaźniki (parametry statystyczne) oceny klasyfikatora: (a) *jakość rozpoznania*, (b) *czułość*, czyli zdolność do wykrywania np. rzeczywiście chorych pacjentów, i (c) *swoistość (specyficzność)*, czyli np. ocena w jakim stopniu klasyfikator wskazuje prawdziwie zdrowe osoby. Zgodnie z notacją przyjętą w tabeli 1, powyższe wskaźniki można obliczyć według następujących wzorów: *jakość rozpoznania = (A+D)/(A+B+C+D), czułość = A/(A+B), swoistość = D/(C+D)*. Ponadto, zgodnie z informacją zawartą w Tabeli 1 mogą być obliczone również dodatkowe wskaźniki, takie jak: *predykcja dodatnia = A/(A+B), predykcja ujemna = D/(C+D), częstość zmiany = (A+C)/(A+B+C+D), iloraz wiarygodności = [A/(A+C)]/[B/(B+D)*. Jednakże interpretacja tych czterech ostatnich wskaźników wymaga pewnej ostrożności, szczególnie w przypadku analizy danych biomedycznych, stąd nie są one tak często stosowane, jak trzy pierwsze, czyli *jakość, czułość i specyficzność*. *Specyficzność i czułość* są podstawowymi parametrami (zmiennymi) dla wykreślania krzywych *ROC*, zaprezentowanych na rysunku 11.

Zauważmy, że analiza krzywych *ROC* pozwala weryfikować nie tylko fakt popełnienia błędów klasyfikacji, ale także oceniać i identyfikować ich rodzaj. Na przykład, w przypadku oceny i porównania różnych testów diagnostycznych, krzywe *ROC* pozwalają na porównanie skuteczności tych testów (np. w warunkach występowania tej samej choroby), poprzez obliczenie pola powierzchni pod krzywą *ROC (*parametru *AUC,* ang. *Area Under Curve*). Test z większym polem powierzchni *AUC* jest testem lepiej różnicującym, czyli rozpoznającym określoną jednostkę chorobową. Stąd analizę *ROC* wykorzystuje się w przypadku poszukiwania i testowania oraz porównywania (bio)markerów różnych chorób. Marker o większym polu *AUC* względem pozostałych markerów jest istotniej/ silniej związany z daną chorobą i/lub stanem pacjenta, co oznacza *de facto,* że będzie skuteczniejszy w jej rozpoznaniu i diagnostyce. Dodajmy też, iż przy właściwej kwalifikacji osób badanych (ogólnie obiektów) do określonych klas, grup lub kategorii (np. osoba zdrowa *versus* chora) korzysta się dodatkowo z możliwości wyznaczeniu granicznego „poziomu" badanego markera, określanego jako jego wartość odcięcia, ang. *cut-off points*. Jest to taka wartość markera, której odpowiada najmniejsza liczba błędnie zakwalifikowanych przypadków prawdziwych (*wysoka specyficzność*) oraz największa liczba prawidłowo zakwalifikowanych przypadków prawdziwych (*wysoka czułość*).

Podsumowując, zgodnie z podejściem analitycznym opartym na klasyfikatorze binarnym i krzywych *ROC*, system klasyfikacji w tym przypadku powinien charakteryzować się się jak największą *czułością* (jak najwyższą liczbą prawidłowych rozpoznań) i z możliwą do zaakceptowania *swoistością*.



Rysunek 11. Ilustracja statystycznej metody opartej na analizie krzywych *ROC*. Każdy punkt na tej krzywej ma współrzędne [1 – specyficzność; czułość]. Pole pod krzywą *ROC* (wartość *AUC*) jest miarą jakości metody klasyfikacji: pole *AUC*=0,5 oznacza losowy układ (losowe przypisanie do jednej z klas), zaś pole *AUC* = 1,0 układ idealny (bezbłędne przypisanie do właściwej klasy) (opracowanie własne)

2.3. Przykłady oceny parametrów jako markerów w badaniach własnych z zastosowaniem metod klasyfikacji nadzorowanej

Modelowe podejście badawczo-analityczne z zastosowaniem metod klasyfikacji nadzorowanej stosowane w badaniach własnych przedstawiono na rysunku 8. Zarówno w klasycznej analizie statystycznej, jak i z użyciem opisanych metod rozpoznawania obrazów, można zastosować podejście matematyczno-statystyczne, w którym: (a) obliczane są miary centralne dla pomiarów w badanych klasach/ grupach, np. w pracy [58], (b) prowadzone są niezależne analizy dla każdego indywidualnego obiektu, w pracy [59]), bądź też (c) połączenie obu podejść, np. w pracach [42, 60]. Ocenę znaczenia badanych parametrów i/lub markerów, zgodnie z modelem klasyfikacji opartym na regule decyzyjnej *k*-najbliższych sąsiadów, prowadzono dla klasyfikatora w wersji standardowej, rozmytej oraz klasyfikatora równoległego, również w systemie hierarchicznym klasyfikacji, adekwatnie do warunków podjętego zagadnienia/ problemu i celu badawczego. Warto podkreślić, iż włączenie nowoczesnych i efektywnych statystycznych metod rozpoznawania obrazów oraz znanej metody opartej na analizie krzywych *ROC* jest szczególnie przydatne w badaniach w biomedycynie. A takie właśnie badania prowadzi autorka, włączając dodatkowo zastosowanie innowacyjnych technik i metod wirtualnej rzeczywistości.

W kolejnych podrozdziałach przedstawiono przegląd wybranych badań i ich wyników dotyczących oceny i znaczenia analizowanych parametrów jako potencjalnych markerów badanych procesów i zjawisk. W tabeli 2 dokonano podsumowania tych badań i otrzymanych rezultatów wraz ze wskazaniem ich znaczenia.

2.3.1. Przykłady oceny parametrów oddechowych w badaniach nad plastycznością układu oddechowego

Interesujących i znaczących danych dotyczących plastyczności układu oddechowego dostarczają między innymi badania nad przerywaną hipoksją (ang. *intermittent hypoxia, IH*) w modelach ostrej i chronicznej ekspozycji na nią (ang. *acute or chronic intermittent hypoxia*, AIH or CIH) oraz treningi wykorzystujące *IH* (ang. *intermittent hypoxia training*, IHT) [61-66]. Badania realizowane są zarówno poprzez eksperymenty biologiczne z udziałem zwierząt, jak też modelowe i kliniczne badania w populacjach osób zdrowych i różnych grup pacjentów [67-72].

Przyjęte zaawansowane podejście modelowo-analityczne w badaniach własnych pozwoliło na opracowanie efektywnych procedur doświadczalnych dla indukcji i poznania mechanizmów plastyczności oddechowej (ang. *respiratory plasticity*); na przykład w zaobserwowanym i opisanym zjawisku związanym z długotrwałym torowaniem wentylacji płucnej *vLTF* (ang. *ventilatory long-term facilitation*) [73], jak również w zjawisku oddechowej plastyczności kompensacyjnej w eksperymentalnym modelowaniu dysfunkcji oddechowych [42, 59, 60]. Badacze zwracają uwagę na konieczność dopracowania protokołów doświadczalnych ze względu na brak standardów i stosowanie różnych protokołów i procedur eksperymentalnych [74,75].

Dla indukcji i badania zjawiska *vLTF* zaproponowano eksperymentalny model oparty na ekspozycji na ostrą przerywaną hipoksję (*AIH*), obejmujący pięć kolejnych cykli *IH* złożonych z 1-minutowej ekspozycji na umiarkowaną 14% hipoksję (faza ekspozycji bodźca, ang. *stimulus phase*) i następnie 3-minutowej fazy normoksyjnego powrotu (ang. *normoxic recovery phase*). Analizowano odpowiedź oddechową oceniając poziom wentylacji minutowej (*MV*) i jej dwóch podstawowych składowych, objętościowej (*VT*) i częstotliwościowej (*f*) [73,76]. W badaniach tych pokazano, iż *MV* wzrasta z kolejnym cyklem hipoksyjno-normoksyjnym (ang. *progressive augmentation*), i wzrastają jednocześnie obie składowe wentylacyjne, *VT* i *f*. Pogłębienie wentylacji dotyczyło zarówno fazy bodźca (hipoksji), jak i fazy normoksyjnego powrotu (normoksji) i było istotnie większe w fazie ekspozycji bodźca. Podwyższony poziom wentylacji utrzymywał się jeszcze przez 30 minut po ostatnim cyklu *AIH* (indukcja zjawiska *vLTF*). Kolejne badania pokazały podobne efekty indukowanej plastyczności oddechowej z zastosowaniem modelu *IH* u ludzi i zwierząt [72], przy czym proponowane i stosowane są różnorodne protokoły i procedury, jak na przykład cykliczna *AIH*, *CIH* z różnymi fazami powrotu normoksyjnego lub hiperoksyjnego [61,77].

Warto podkreślić, iż obserwowana plastyczność oddechowa jako pogłębienie wentylacji płucnej w odpowiedzi na ekspozycję *AIH* może być wykorzystana jako nowa propozycja w skuteczniejszej terapii oddechowej w ciężkim przebiegu *ALS*, w uszkodzeniu i/lub przerwaniu rdzenia kręgowego (na poziomie kręgów szyjnych i z zaburzeniami oddechowymi), na co wskazują już wyniki badań eksperymentalnych z udziałem zwierząt oraz pacjentów, przy czym ekspozycja na *AIH* nie powoduje żadnych niekorzystnych objawów i jest dobrze tolerowana przez chorych [78]. Ponadto, indukowane zmiany plastyczne w odpowiedzi na cykliczne epizody hipoksji pozwalają na wgląd w uruchamiane mechanizmy korzystnej poprawy oddychania w treningach *IH*, które coraz częściej stosowane są w fizjologii sportu, lotnictwie czy też treningu wspinaczkowym [70, 72].

Badania eksperymentalne i kliniczne wskazują na różną efektywność uruchamianych mechanizmów plastyczności oddechowej. Na przykład w badaniach własnych z zastoso-

waniem eksperymentalnego modelu stopniowego odnerwiania i/lub porażenia głównego mięśnia wdechowego przepony opisano uruchamiane mechanizmy wczesnej kompensacji oddechowej i ich skuteczność [42, 45, 60]. Analizowano różne zestawy parametrów oddechowych wykorzystując przedstawione podejście z klasyfikacją nadzorowaną. Opisano możliwe wzorce wczesnej odpowiedzi oddechowej, a wśród nich wskazano na te najbardziej skuteczne, zapewniające optymalny poziom wentylacji płucnej, a realizowane poprzez (a) zwiększenie ośrodkowego i obwodowego napędu oddechowego, i w efekcie następujący (b) wzrost aktywności mięśni oddechowych, których praca służy (c) zachowaniu lub zwiększeniu wentylacji płuc, a towarzyszą temu również (d) zmiany we wzorcu oddechowym, zaś po porażeniu przepony (e) obserwowano przejście na tor brzuszny oddychania.

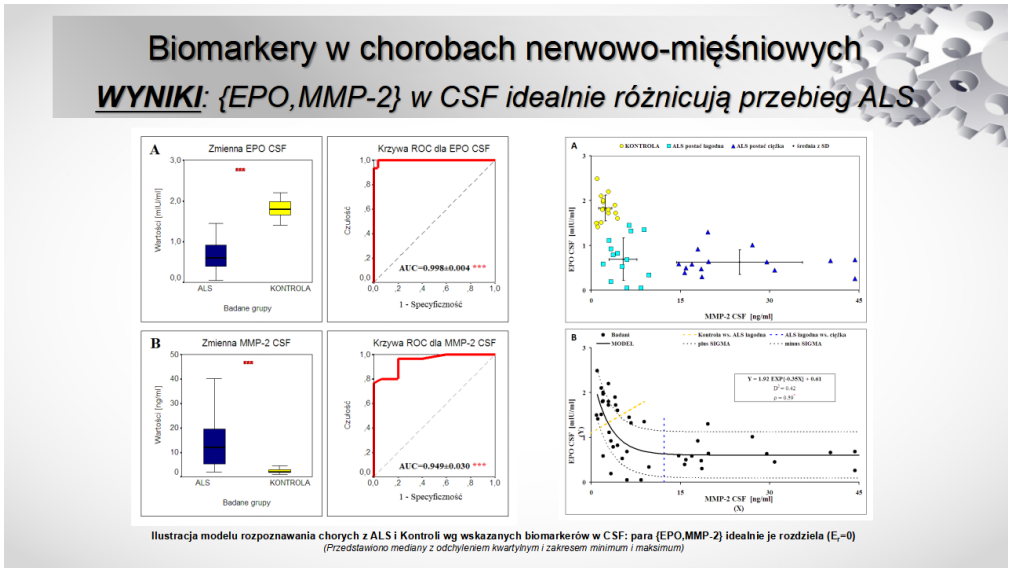
Przedstawione rezultaty mają ogromne znaczenie w klinice chorób oddechowych, dla podejmowanych różnych strategii terapeutycznych/ rehabilitacyjnych w chorobach, którym towarzyszą zaburzenia oddychania, jak już wskazano, np. w *ALS*, po uszkodzeniu rdzenia kręgowego, w dysfunkcji/ chorobach mięśni oddechowych, czy w innych poważnych schorzeniach opisanych np. w pracach [64, 65, 79, 80].

2.3.2. Przykłady oceny parametrów laboratoryjnych jako markerów w badaniach nad bardzo rzadkimi chorobami nerwowo-mięśniowymi

W badaniach klinicznych nad oceną i znaczeniem różnych markerów w bardzo rzadkich chorobach nerwowo-mięśniowych (np. jako diagnostycznych i prognostycznych biomarkerów stanu klinicznego chorych), przyjęte podejście badawczo-analityczne skutecznie wsparło klasyczne analizy statystyczne, a ponadto pozwoliło (zgodnie z procedurą selekcji cech i douczania klasyfikatora) na wskazanie bardziej efektywnych markerów, które pozwalają: (a) na perfekcyjne rozpoznanie jednostki chorobowej (potencjalne markery uszkodzenia mięśnia sercowego w przypadku różnych postaci dystrofii mięśniowej Emery'ego Dreifussa, ang. *Emery-Dreifuss muscular dystrophy, EDMD*) oraz (b) na rozpoznanie progresu procesu chorobowego (w stwardnieniu zanikowym bocznym, *ALS*).

Dystrofie mięśniowe są grupą genetycznie uwarunkowanych chorób powodujących zanik tkanki mięśniowej, jej zastąpienie tkanką tłuszczową i włóknienie, co prowadzi do postępującego osłabienia i zaniku mięśni. Dystrofia *EDMD* charakteryzuje się zajęciem mięśni szkieletowych i mięśnia serca, jest ultra rzadką chorobą, o częstości występowania 4-8 przypadków na 1000000 osób. Z kolei *ALS* jest bardzo rzadką chorobą neurozwyrodnieniową, zaliczaną do chorób neuronu ruchowego, rocznie notuje się 4-6 zachorowania na 100000 osób, u ok. 5-10% choroba ma podłoże genetyczne, u pozostałych ma postać sporadyczną *ALS* (ang. *sporadic ALS*). Oba schorzenia należą do nieuleczalnych (brak jest skutecznego leczenia przyczynowego) i przewlekłych, obniżających stopniowo sprawność fizyczną, co prowadzi do ograniczenia samodzielności, a ostatecznie i do inwalidztwa. W przypadku *ALS* postępująca niewydolność oddechowa jest najczęściej przyczyną zgonu, zaś w *EDMD* powikłania kardiologiczne mogą prowadzić do nagłego zgonu sercowego.

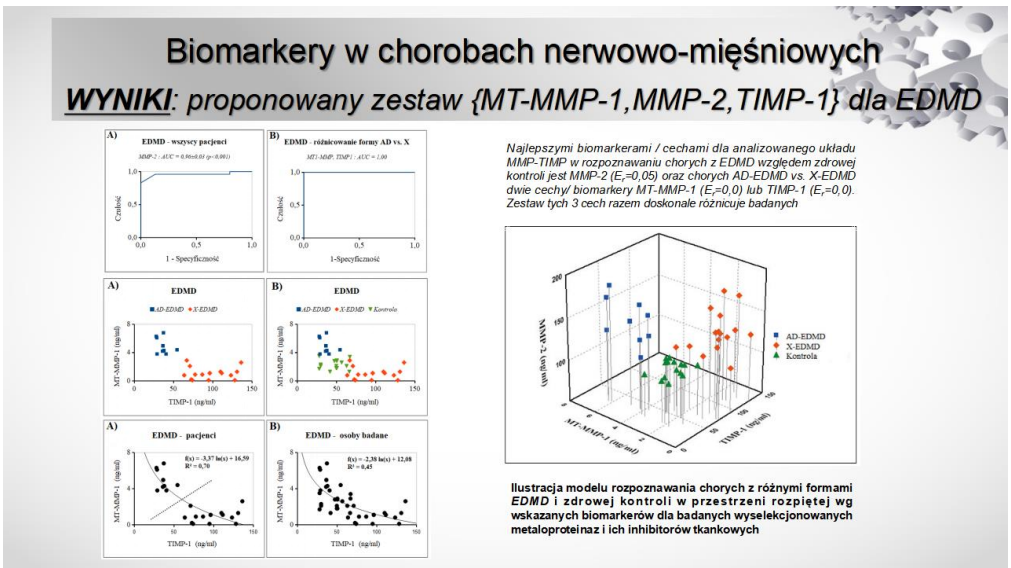
Na rysunkach 12 i 13 przedstawiono kluczowe rezultaty prowadzonych analiz z klasyfikacją nadzorowaną dla oznaczonych metodą immunoenzymatyczną ELISA różnych parametrów laboratoryjnych: (a) w surowicy krwi i płynie mózgowo-rdzeniowym (ang. *serum and cerebrospinal fluid, CSF*) w przebiegu *ALS* [58, 81] oraz (b) w surowicy dla *EDMD* i jej dwu postaci, tj. *AD-EDMD* (związanej z laminopatią) i *X-EDMD* (z *X*-emerinopatią) [82, 83]. U pacjentów z *AD-EDMD* z reguły obserwowany jest cięższy przebieg i szybciej dochodzi do niepełnosprawności.



Rysunek 12. Wyniki badań własnych nad biomarkerami w stwardnieniu zanikowym bocznym, przedstawione i szczegółowo omówione np. w pracy [47] (opracowanie własne)

Wyniki badań laboratoryjnych i/lub dane genetyczne stanowiły bazę dla wygenerowanych zbiorów i podzbiorów uczących. W przypadku pacjentów z *ALS*, u których zdiagnozowano postać sporadyczną, uwzględniono w analizie następujące zmienne: (a) wiek, postać *ALS* (opuszkową, rdzeniową), ciężkość objawów (łagodne lub poważne, ang. *mild or severe*), czas trwania choroby (są to klasyczne markery), zaś wśród oznaczonych metodą ELISA parametrów laboratoryjnych analizowano poziomy stężenia: (b) erytropoetyny (ang. *erythropoetin, EPO*), hormonu m.in. kluczowego w zapewnieniu prawidłowego utlenowania tkanek, jej poziom spada u tych chorych (kierunkiem rozwoju nowych terapii jest jego wyrównanie), ponadto oznaczano c) poziom metaloproteinaz macierzy zewnątrzkomórkowej (ang. *metalloproteinases*, MMPs) a wśród nich *MT-MMP-1* (kolagenazy transbłonowej), *MMP-2* (żelatynazy A), *MMP-9* (żelatynzy B) oraz także stężenie (d) ich inhibitorów tkankowych *TIMP-1* i *TIMP-2* (ang. *tissue inhibitors of metalloproteinases*, TIMPs). Przy czym, ich oznaczeń dokonano zarówno w surowicy, jak i płynie mózgowo-rdzeniowym, zgodnie ze standardowym postępowaniem klinicznym w *ALS*. Natomiast w przypadku pacjentów z *EDMD* poziomy parametrów laboratoryjnych były oznaczane wyłącznie w surowicy, a były to również metaloproteinazy *MT-MMP-1*, *MMP-2* i *MMP-9* oraz ich inhibitory tkankowe *TIMP-1*, *TIMP-2* i *TIMP-3*. Wybór i poszukiwanie potencjalnych markerów związanych w tych badaniach z układem *MMP-TIMP* macierzy zewnątrzkomórkowej (ang. *extracellular matrix*, ECM) wynika z obserwacji, iż w warunkach patologicznych często wzrasta aktywność *MMPs* (ich funkcją jest degradacja i przebudowa *ECM*), jednocześnie spada poziom *TIMP*, które bezpośrednio regulują aktywność *MMP* (zidentyfikowano już wiele nowych podgrup *MMPs*) [84-86]. Zmiany relacji *MMP-TIMP* mogą być wykorzystane jako prognostyczny marker dla monitorowania efektów nowych strategii terapeutycznych mających na celu przywrócenie równowagi i prawidłowej interakcji *MMP-TIMP*, a przez to i prawidłowej struktury i funkcji *ECM*, środowiska odpowiedzialnego za wiele procesów

fizjologicznych [87]. Na rysunku 12 podsumowano wyniki przeprowadzonych analiz w grupie pacjentów z *ALS*. Wskazują one, iż na przykład zestaw dwóch cech/ parametrów, takich jak poziom stężenia erytropoetyny i metaloproteinazy *MMP-2* w płynie mózgowo-rdzeniowym idealnie różnicuje pacjentów z łagodnym i ciężkim przebiegiem choroby. Z kolei na rysunku 13 podsumowano wyniki analiz w grupie pacjentów z *EDMD*. W tym przypadku zestaw trzech cech/ parametrów {*MT-MMP-1,MMP-2,TIMP-1*} perfekcyjnie różnicuje różne postacie *EDMD*, czyli *X-EDMD (EDMD1) i AD-EDMD (EDMD2)*.



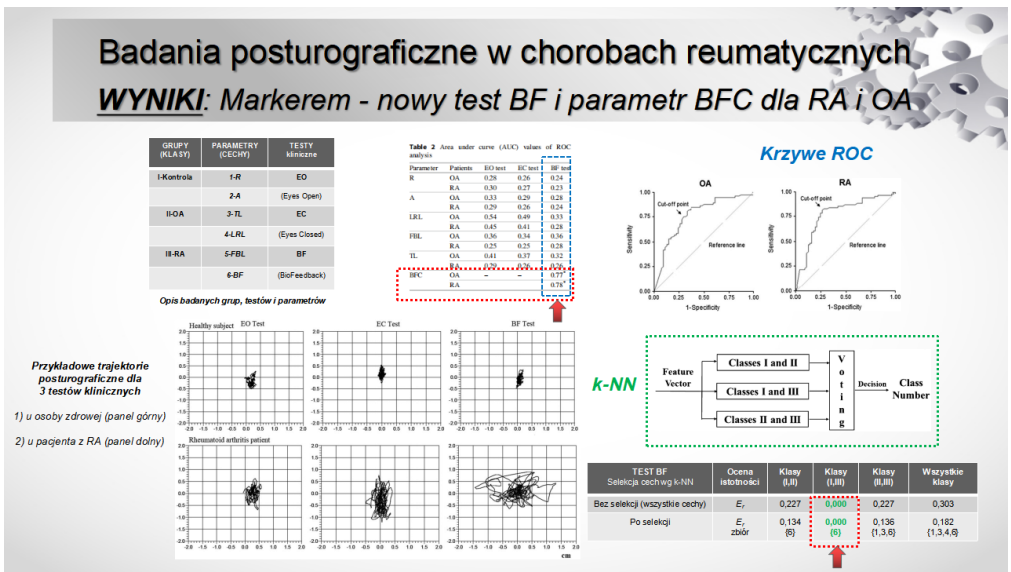
Rysunek 13. Wyniki badań własnych nad biomarkerami w dystrofii mięśniowej typu *EDMD*, przedstawione i szczegółowo omówione np. w pracy [48] (opracowanie własne)

Konkludując, badania nad biomarkerami w chorobach rzadkich mają na celu (a) wsparcie prawidłowej diagnostyki, (b) włączając i diagnostykę pozwalającą na wczesne ich wykrycie, i jak najwcześniejsze wdrożenie postępowania spowalniającego postęp choroby, (c) monitorowanie i ocenę progresu procesu chorobowego, (d) skuteczne prognozowanie i rokowanie, co do przyszłego stanu klinicznego, (e) przeciwdziałanie poważnym powikłaniom oraz groźnym incydentom kardiologicznym (*EDMD*) lub oddechowym (*ALS*), np. poprzez wszczepienie rozrusznika lub podjęcie sztucznej wentylacji.

2.3.3. Przykłady oceny parametrów posturograficznych w badaniach klinicznych nad układem ruchu

Badania dotyczyły oceny i analizy stabilności równowagi u chorych ze zmianami patologicznymi w obrębie narządu ruchu z zastosowaniem m.in. metod klasyfikacji nadzorowanej. Na przykład oceniano znaczenie parametrów opisujących trajektorie posturograficzne w testach klinicznych na statycznej platformie posturograficznej (firmy Pro-Med) u pacjentów z chorobą zwyrodnieniową stawów (ang. *osteoarthritis, OA*) oraz z reumatoidalnym zapaleniem stawów (ang. *rheumatoid arthritis, RA*). Były one kontynuacją wcześniejszych badań u pacjentów z normotensyjnym wodogłowiem oraz zanikiem mózgu (ang. *normal pressure hydrocephalus, NPH, or brain atrophy, BA*), a które wskazały na istotność zdefiniowanego parametru „indeksu wzrokowego" w rozpoznaniu

obu jednostek chorobowych, celem wsparcia diagnostyki różnicowej w *NPH* i *BA* [88-90]. Poważne zaburzenia układu ruchu towarzyszą również chorobom reumatycznym, i są coraz poważniejszym problemem współczesnych starzejących się społeczeństw. Zastosowanie klasycznych testów klinicznych na platformie posturograficznej, w szczególności zaproponowany test ze wzrokowym (wizualnym) sprzężeniem zwrotnym *BF* (ang. *biofeedback*, BF) i zdefiniowanie nowego parametru *BFC* (ang. *biofeedback coordination*) dla tego właśnie testu pokazało, iż możliwe jest precyzyjne różnicowanie pacjentów z *OA* i *RA* (chorych z *RA* charakteryzują poważniejsze objawy i cięższy przebieg choroby w porównaniu z grupą z *OA*). Wyniki wskazują, iż test *BF* wraz z proponowanym parametrem *BFC* może być skutecznym narzędziem oceny stanu klinicznego pacjentów reumatycznych. Proponowane podejście może być również narzędziem ilościowej oceny efektów terapeutycznego treningu na platformie posturograficznej w prewencji przed upadkami u osób starszych. Ostatnie badania potwierdzają takie zastosowanie [91].



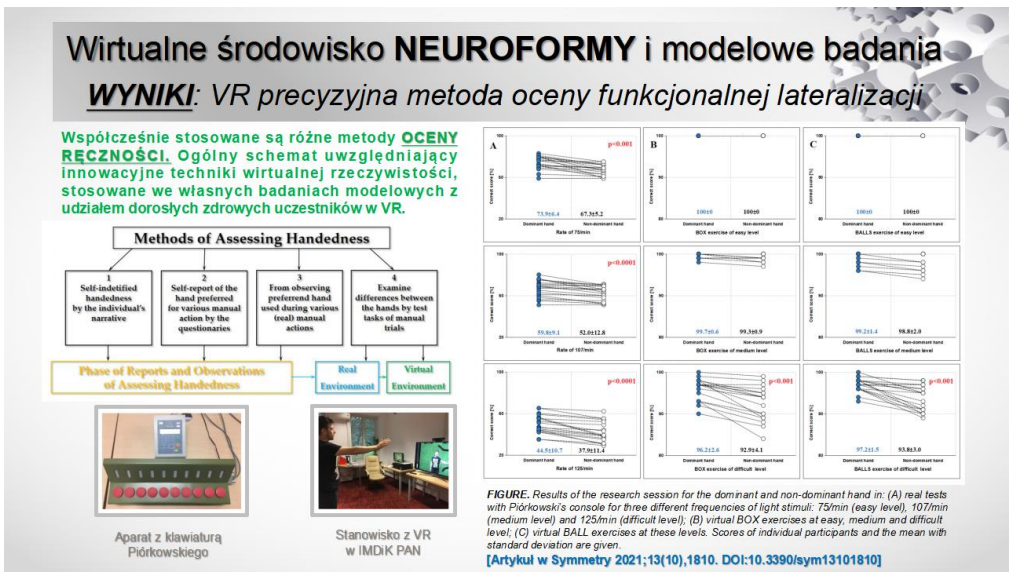
Rysunek 14. Wyniki badań własnych z zastosowaniem statycznej platformy posturograficznej w chorobach reumatycznych, szczegółowo omówionych np. w pracach [53, 77] (opracowanie własne)

2.3.4. Przykłady oceny wyników i parametrów zadań testowych realizowanych w wirtualnej rzeczywistości

Jak wspomniano we Wstępie, nowoczesne techniki i technologie oferują atrakcyjne wirtualne środowiska, które cieszą się ogromnym zainteresowaniem i odgrywają już ważną rolę w neuronaukach. Opisane metody klasyfikacji okazują się równie skuteczne w wirtualnej jak i w realnej rzeczywistości.

W modelowych badaniach własnych nad oceną funkcjonalnej lateralizacji zastosowano nieimersyjne wirtualne środowisko kreowane przez system NEUROFORMY [22, 92, 93]. Jak zilustrowano na rysunku 15, zarówno klasyczne realne testy oceny dominującej strony ciała, jak i proponowane testy realizowane w wirtualnej rzeczywistości pozwalają na równie precyzyjne określenie funkcjonalnej lateralizacji. Przy czym istotną przewagą wirtualnych testów i zadań nad klasycznymi testami neuropsychologicznymi,

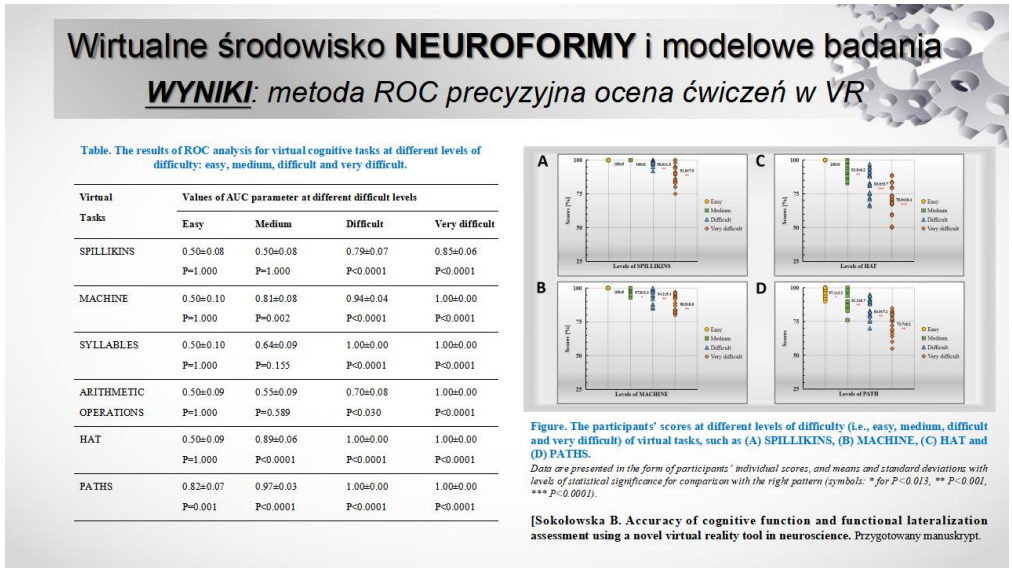
czy ogólnie „realnymi" testami klinicznymi, jest wysoka wartość ekologiczna wirtualnego środowiska oraz możliwość wykreowania praktycznie każdych warunków odpowiadających sytuacjom z rzeczywistego świata. Równie ważna jest możliwość komunikacji/ łączności on-line z osobą badaną w jej własnym środowisku, co z kolei umożliwia prowadzenie domowych treningów pod zdalnym nadzorem terapeuty lub rehabilitanta. Ponadto systemy te dostosowują poziom ćwiczeń/ sesji treningowych do aktualnych możliwości osoby ćwiczącej. Realizowane jest to najczęściej w formie interaktywnych gier komputerowych „na serio" z „biofeedbackiem", a służy temu określona algorytmiczna/ systemowa skala trudności oraz algorytm awansu lub nie na kolejny poziom zadania. Podobnie ćwiczy się w wirtualnej rzeczywistości NEUROFORMY.



Rysunek 15. Wyniki modelowych badań własnych nad funkcjonalną lateralizacją, z udziałem zdrowych dorosłych ochotników, w wirtualnej rzeczywistości kreowanej przez system NEUROFORMY, przedstawione np. w pracy [22] (opracowanie własne)

Stopniowanie poziomu trudności wirtualnego zadania testowego ma istotne znaczenie w interpretacji wirtualnych wyników, szczególne w zastosowaniach w praktyce medycznej. W odniesieniu do przedstawionej funkcjonalnej lateralizacji wzrost stopnia trudności wirtualnych zadań pozwala na ujawnienie strony dominującej (względem niedominującej) i ilościową ocenę różnicy w poziomie wykonania danego ćwiczenia w funkcji rosnącego stopnia trudności. Na przykład, na wyższych poziomach trudności uzyskiwano lepsze wyniki (i istotne statystycznie), gdy używano ręki dominującej w stosunku do ręki niedominującej.

Na rysunku 16 zaprezentowano przykład zastosowania metod klasyfikacji dla oceny i porównania stopnia trudności wirtualnych ćwiczeń realizowanych przez zdrowe dorosłe osoby.



Rysunek 16. Modelowe badania własne nad oceną poziomu wykonania wirtualnych ćwiczeń poznawczo-motorycznych w zależności od poziomu trudności. Ocenę oparto na klasyfikatorze binarnym i metodzie *ROC*, przy przyjęciu poziomu referencyjnego/ odniesienia odpowiadającego najlepszemu, czyli bezbłędnemu wykonaniu zadania, tj. 100% punktów (opracowanie własne)

Zaproponowano podejście analityczne oparte na klasyfikatorze binarnym i metodzie *ROC*, przy przyjęciu poziomu referencyjnego/ odniesienia odpowiadającego najlepszemu (bezbłędnemu) wykonaniu zadania testowego, czyli uzyskaniu 100% punktów. Porównano otrzymane wyniki na każdym kolejnym poziomie trudności, tj. od łatwego, poprzez średni, do trudnego i bardzo trudnego (ang. *easy, medium, difficult and very difficult*) dla różnych poznawczo-motorycznych ćwiczeń testowych. Były to wirtualne zadania typu: Bierki (ang. *Spillikins*), Maszyna (ang. *Machine*), Sylaby (ang. *Syllables*), Działania Arytmetyczne (ang. *Arithmetic Operations*), Kapelusz (ang. *Hat*) i Ścieżki (ang. *Paths*). Pozwoliło to na zweryfikowanie wybranego poziomu trudności wirtualnego zadania (zdefiniowanego przez systemowy algorytm) z rzeczywistymi wynikami otrzymanymi przez zdrowych uczestników. Analiza ta pokazała, iż poziomy zdefiniowane jako łatwy i średni były takie w istocie w przypadku wszystkich tych ćwiczeń. Natomiast zaobserwowano duże różnice pomiędzy ćwiczeniami na poziomach zdefiniowanych jako trudne lub bardzo trudne, gdyż zdecydowanie najtrudniejsze były przede wszystkim Ścieżki i Kapelusz, i tutaj osiągano najniższe wyniki i nie awansowano. W przeciwieństwie do Bierek, zadania w którym osiągano zawsze najlepsze wyniki na wszystkich poziomach trudności spośród wszystkich ćwiczeń i stosunkowo łatwo awansowano na kolejne poziomy tego ćwiczenia. Jest to interesujący rezultat dotyczący ogólniejszego zagadnienia, aby przy ocenie i interpretacji oraz porównywaniu rezultatów uwzględniać fakt braku standaryzacji różnych wirtualnych środowisk. Jest to ważny i szeroko dyskutowany problem wśród badaczy i klinicystów, którzy przedstawiają i proponują różne metody walidacji dla różnych stosowanych wirtualnych testów i systemów [94-97].

3. Podsumowanie wyników prowadzonych i prezentowanych badań modelowych, eksperymentalnych i klinicznych w realnej i wirtualnej rzeczywistości

Dotychczas prowadzone badania pozwoliły na podjęcie ważnych tematów z obszaru neuronauk, nauk o zdrowiu, włączając badania kliniczne, dotyczące m.in. fizjopatologii układu oddechowego, chorób neurologicznych i nerwowo-mięśniowych oraz reumatycznych. W ocenę analizowanych parametrów/zmiennych i/lub potencjalnych (bio)markerów oraz w przygotowaniu eksperymentalnych protokołów i procedur włączono również bardzo efektywne i precyzyjne metody klasyfikacji nadzorowanej, idealnie wpasowujące się w podejmowaną tematykę badawczą. Takie modelowe podejście badawczo-analityczne pozwala na testowanie i pogłębienie tematyki prowadzonych badań, w tym i poszukiwania skutecznych rozwiązań dla konkretnych problemów badawczych.

W tabeli 2 podsumowano i przedstawiono niektóre z interesujących i ważnych rezultatów badań własnych oraz badań z nimi powiązanych.

Tabela 2. Podsumowanie wyników badań nad oceną parametrów i markerów, z zastosowaniem zarówno klasycznych metod statystycznych, jak i klasyfikacji nadzorowanej i systemów uczących, realizowanych w ramach prezentowanego podejścia modelowo-badawczego (opracowanie własne)

Obszar zainteresowań badawczych/ tematyka	Opis rezultatów prowadzonych i przedstawionych badań modelowych, eksperymentalnych i klinicznych
Funkcjonalna lateralizacja oraz treningi poznawczo-motoryczne w wirtualnej rzeczywistości: [22, 92, 93, 98-104]	Zaproponowane modelowe podejście badawczo-analityczne zastosowano również w aktualnych badaniach modelowych prowadzonych z zastosowaniem metod wirtualnej rzeczywistości, na stanowisku *NEUROFORMY* w PB IMDiK PAN. Badania te pokazały, iż innowacyjne środowisko, jakim stają się wirtualne światy jest czułym narzędziem badawczym w obszarze neuronauk podstawowych i klinicznych. Dotychczasowe badania wskazują na ich wysoką wartość ekologiczną oraz wysoką dokładność, czułość i swoistość. Treningi i ćwiczenia/ zadania testowe w *VR* to znakomite narzędzie w ocenie funkcjonalnej lateralizacji, czy też poziomu wykonania wirtualnych zadań, a w efekcie również np. w poznaniu i badaniu mechanizmów procesów poznawczych i motorycznych oraz funkcji wykonawczych, np. [94, 95, 105, 106].
Choroby reumatyczne: choroba zwyrodnieniowa stawów i reumatoidalne zapalenie stawów: [53, 107-110]	W badaniach wśród tej grupy pacjentów zastosowano też zaproponowane podejście modelowe z użyciem klasyfikatorów, zarówno klasycznego klasyfikatora binarnego i opartej na nim metodzie analizy krzywych *ROC*, jak i klasyfikatora opartego na regule decyzyjnej *k-NN*. Badania te pokazały nowy rezultat związany ze stosowanymi klasycznymi testami klinicznymi. Mianowicie, wśród zaproponowanych i analizowanych w tym badaniu kluczowych testów z otwartymi i zamkniętymi oczami, włączono trzeci test z wizualnym *biofeedbackiem* (sprzężeniem zwrotnym). Okazał się on kluczowym parametrem (i zarazem markerem) w ocenie i różnicowaniu badanych grup pacjentów. Ostatnie badania również potwierdzają powyższe konkluzje z realnej rzeczywistości w tej wirtualnej [91], np. [111-114].
Choroby nerwowo-mięśniowe: stwardnienie zanikowe boczne: [38, 39, 43, 47, 58, 81, 115, 116]	W ramach podjętej współpracy z klinicystami, oceniano różne zestawy parametrów laboratoryjnych na podstawie ich oznaczeń w surowicy i płynie mózgowo-rdzeniowym. Potwierdzono wysoką skuteczność zaproponowanego modelowego podejścia badawczo-analitycznego z klasyfikacją nadzorowaną, zarówno dla klasycznego binarnego klasyfikatora (analiza krzywych *ROC*) jak i dla nieparametrycznego klasyfikatora *k-NN*. Dodatkowo testowano różne modele klasyfikacji oparte na regule decyzyjnej *k-NN*, od standardowego klasyfikatora,

	poprzez klasyfikator równoległy (o dwu-decyzyjnych składowych) i również o strukturze hierarchicznej. Zaproponowane podejście analityczno-modelowe, z kluczowym dla zagadnienia wyboru markerów, włączonym procesem selekcji cech, znakomicie sprawdziło się, zarówno w różnicowaniu zdrowej kontroli i pacjentów, jak i też wśród pacjentów z łagodnymi i ciężkimi objawami chorobowymi. Te markery badane są przede wszystkim w kontekście diagnostyki i leczenia, np. [117-125].
Choroby nerwowo-mięśniowe: dystrofia mięśniowa typu *Emery'ego-Dreifussa*: [46, 48, 82, 83, 126-131]	Podobnie jak w opisanych badaniach dotyczących *ALS*, i w przypadku pacjentów chorujących na dystrofię mięśniową typu *EDMD*, oparcie badań o analizę wg modelu z zastosowaniem metod rozpoznawania obrazów, jak i klasycznej klasyfikacji wg krzywych *ROC*, znakomicie sprawdziło się w efektywnym wyborze potencjalnych kandydatów na biomarkery, np. w bezbłędnym różnicowania grup pacjentów z postacią *X-EDMD lub AD-EDMD*. Zidentyfikowano już kolejne postacie dystrofii *EDMD* i poszerza się wiedza dotycząca puli markerów związanych z chorobami nerwowo-mięśniowymi, np. [132-138].
Choroby neurologiczne: normotensyjne wodogłowie i zanik mózgu: [88-90, 139-141]	Potwierdzono również skuteczność statystycznych metod klasyfikacyjnych w badania nad oceną równowagi i chodu, włączając tomografię komputerową i testując nowe podejście analityczne oparte na programie VisNow. Zaproponowane analizy z zastosowaniem również klasycznych metod dyskryminacyjnych oraz metod opartych na regule decyzyjnej *k-NN* są efektywnym wsparciem w diagnostycznym obrazowaniu i różnicowaniu pacjentów z wodogłowiem normotensyjnym i zanikiem mózgu. Ponadto wykazano skuteczność zaproponowanego parametru indeksu wzrokowego jako potencjalnego markera różnicującego obie jednostki chorobowe. Proponowane podejście badawczo-analityczne jest na tyle uniwersalnym, że może być ogólnie stosowane w chorobach narządu ruchu, stąd włączenie się w badania kliniczne dotyczące tych grup pacjentów. Inne przykłady, np. [142-149].
Zjawisko (neuro)plastyczności układu oddechowego (*UO*) i eksperymentalne modelowanie fizjopatologii *UO:* [42, 45, 59, 60, 73, 150-156]	W tych badaniach wypracowano model analitycznego podejścia opartego nie tylko na klasycznych metodach statystycznych, ale i uwzględniającego metody klasyfikacji nadzorowanej. Wdrożono je, na przykład, w badaniach nad poznaniem mechanizmów i strategii wczesnej kompensacji oddechowej w warunkach dysfunkcji mięśni oddechowych oraz wczesnej odpowiedzi oddechowej na ekspozycję oddechowych bodźców chemicznych w modelach zwierzęcych, u podłoża których leży neuroplastyczność *UO*. W ramach tych badań zaproponowano oddechowe eksperymentalne modelowanie i opisano zjawisko plastyczności oddechowej *vLTF*, które później zostało również potwierdzone i opisane u ludzi. W badaniach oddechowych też po raz pierwszy został zastosowany model *fuzzy k-NN*. Te badania również pozwoliły na wskazanie i opisanie różnych strategii wczesnej kompensacji oddechowej w odpowiedzi na dysfunkcję głównego mięśnia oddechowego przepony. Badania dotyczące poruszanej tematyki nie tylko potwierdzają rezultaty, ale i odsłaniają kolejne ważne własności *UO*, np. [74, 75, 157-168].

Podziękowania

Prezentowane badania realizowano m.in. w ramach dwóch autorskich tematów badawczych, dotyczących: (1) „Oceny statystycznej i klinicznej wybranych typów schorzeń oraz podjętej rehabilitacji u pacjentów" oraz (2) „Badań nad układem ruchowym z zastosowaniem rozszerzonej/ wspomaganej i wirtualnej rzeczywistości".

Autorka pragnie podziękować Pani Profesor Barbarze Więckowskiej i Panu Profesorowi Mariuszowi Panczykowi za niezwykle rzeczowe i pomocne recenzje autoreferatu pt. „Ocena istotności i znaczenia parametrów/ markerów, z uwzględnieniem statystycz-

nego modelu z klasyfikacją, w badaniach klinicznych i w modelowych eksperymentach na zwierzętach". Szczególne podziękowania autorka składa również wszystkim autorom wspólnych owocnych badań, w tym Panu Profesorowi Adamowi Jóźwikowi za algorytmiczne wsparcie w rozpoznawaniu obrazów. Dziękuje też Pani Barbarze Bugalskiej, Pani Dominice Masternak i Pani Katarzynie Nieszporskiej z Biblioteki IMDiK PAN za nieocenioną pomoc we wszelkich poszukiwaniach bibliograficznych.

Literatura

1. Abdelhafiz D., Baker T., Glascow D.A., Abdelhafiz A., *Biomarkers for the diagnosis and treatment of rheumatoid arthritis – a systematic review*, Postgrad Med, 135(3), 2023, s. 214-223. doi: 10.1080/00325481.2022.2052626.
2. Chen H., Yang Y., Odisho D., Wu S., Yi C., Oliver B.G., *Can biomarkers be used to diagnose attention deficit hyperactivity disorder?*, Front Psychiatry, 14, 1026616, 2023, doi: 10.3389/fpsyt.2023.1026616.
3. Datta N., Ghosh P.S., *Update on Muscular Dystrophies with Focus on Novel Treatments and Biomarkers*, Curr Neurol Neurosci Rep, 20(6), 2020, s. 14. doi: 10.1007/s11910-020-01034-6.
4. Pyrgelis E.S., Boufidou F., Constantinides V.C., Papaioannou M., Papageorgiou S.G., Stefanis L., Paraskevas G.P., Kapaki E., *Cerebrospinal Fluid Biomarkers in iNPH: A Narrative Review*, Diagnostics (Basel), 12(12), 2022, s. 2976. doi: 10.3390/diagnostics12122976.
5. Samad M., Malempati S., Restini C.B.A., *Natriuretic Peptides as Biomarkers: Narrative Review and Considerations in Cardiovascular and Respiratory Dysfunctions*, Yale J Biol Med, 96(1), 2023, s. 137-149. doi: 10.59249/NCST6937.
6. Sanchez-Tejerina D., Llaurado A., Sotoca J., Lopez-Diego V., Vidal Taboada J.M., Salvado M., Juntas-Morales R., *Biofluid Biomarkers in the Prognosis of Amyotrophic Lateral Sclerosis: Recent Developments and Therapeutic Applications*, Cells, 12(8), 2023, s. 1180. doi. 10.3390/cells12081180.
7. Cao Y., Ng G.W., Ye S.S., *Design and Evaluation for Immersive Virtual Reality Learning Environment: A Systematic Literature Review*, Sustainability, 15(3), 2023, s. 1964. doi: 10.3390/su15031964.
8. Chang H.Y., Binali T., Liang J.C., Chiou G.L., *Ten years of augmented reality in education: A meta-analysis of (quasi-) experimental studies to investigate the impact*, Comput Educ, 191, 2022, s. 104641. doi: 10.1016/j.compedu.2022.104641.
9. Hamad A., Jia B., *How Virtual Reality Technology Has Changed Our Lives: An Overview of the Current and Potential Applications and Limitations*, Int J Environ Res Public Health, 19(18), 2022, s. 11278. doi: 10.3390/ijerph191811278.
10. Stanney K.M., Skinner A., Hughes C., *Exercisable Learning-Theory and Evidence-Based Andragogy for Training Effectiveness using XR (ELEVATE-XR): Elevating the ROI of Immersive Technologies*, Int J Hum-Comput Interact, 2023. doi: 10.1080/10447318.2023.2188529.
11. Wang W.T., Lin Y.L., Lu H.E., *Exploring the effect of improved learning performance: A mobile augmented reality learning system*, Educ Inf Technol (Dordr), 2022, s. 1-33. doi: 10.1007/s10639-022-11487-6.
12. Ali S.G., Wang X., Li P., Jung Y., Bi L., Kim J., Chen Y., Feng D.D., Magnenat Thalmann N., Wang J., Sheng B., *A systematic review: Virtual-reality-based techniques for human exercises and health improvement*, Front Public Health, 11, 2023, 1143947. doi: 10.3389/fpubh.2023.1143947.

13. Bera K., Schalper K.A., Rimm D.L., Velcheti V., Madabhushi A., *Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology*, Nat Rev Clin Oncol, 16(11), 2019, s. 703-715. doi: 10.1038/s41571-019-0252-y.
14. Leong S.C., Tang Y.M., Toh F.M., Fong K.N.K., *Examining the effectiveness of virtual, augmented, and mixed reality (VAMR) therapy for upper limb recovery and activities of daily living in stroke patients: a systematic review and meta-analysis*, J Neuroeng Rehabil, 19(1), 2022, s. 93. doi: 10.1186/s12984-022-01071-x.
15. Milewska-Jędrzejczak M., Głąbiński A., *The Influence of Conventional and Innovative Rehabilitation Methods on Brain Plasticity Induction in Patients with Multiple Sclerosis*, J Clin Med, 12(5), 2023, 1880. doi: 10.3390/jcm12051880.
16. De Santis K.K., Mergenthal L., Christianson L., Busskamp A., Vonstein C., Zeeb H., *Digital Technologies for Health Promotion and Disease Prevention in Older People: Scoping Review*, J Med Internet Res, 25, 2023, e43542. doi: 10.2196/43542.
17. Baxi V., Edwards R., Montalto M., Saha S., *Digital pathology and artificial intelligence in translational medicine and clinical practice*, Mod Pathol, 35(1), 2022, s. 23-32. doi: 10.1038/s41379-021-00919-2.
18. Huang C.Y., Chiang W.C., Yeh Y.C., Fan S.C., Yang W.H., Kuo H.C., Li P.C., *Effects of virtual reality-based motor control training on inflammation, oxidative stress, neuroplasticity and upper limb motor function in patients with chronic stroke: a randomized controlled trial*, BMC Neurol, 22(1), 2022, s. 21. doi: 10.1186/s12883-021-02547-4.
19. Huss R., Raffler J., Märkl B., *Artificial intelligence and digital biomarker in precision pathology guiding immune therapy selection and precision oncology*, Cancer Rep (Hoboken), 2023, e1796. doi: 10.1002/cnr2.1796.
20. Milgram P., Takemura H., Utsumi A., Kishino F., *Augmented Reality: A class of displays on the reality-virtuality continuum*, [w:] *Telemanipulator and Telepresence Technologies*, SPIE: Bellingham, WA, USA, 2351, 1994, s. 282-292.
21. Skarbez R., Smith M., Whitton M.C., *Revisiting Milgram and Kishino's Reality-Virtuality Continuum*, Front Virtual Real, 2, 2021, 647997. doi: 10.3389/frvir.2021.647997.
22. Sokołowska B., *A Novel Perspective for Examining and Comparing Real and Virtual Test Tasks Performed by the Dominant and Non-Dominant Hand in Healthy Adults*, Symmetry, 13(10), 2021, s. 1810-1823. doi: 10.3390/sym13101810.
23. Dwivedi Y.K., Hughes L., Baabdullah A.M., Ribeiro-Navarrete S., Giannakis M., Al-Debei M.M., Dennehy D., Metri B., Buhalis D., Cheung C.M.K., Conboy K., Doyle R., Dubey R., Dutot Y., Felix R., Goyal D.P., Gustafsson A., Hinsch C., Jebabli I., Janssen M., Kim Y.G., Kim J., Koos S., Kreps D., Kshetri N., Kumar V., Ooi K.B., Papagiannidis S., Pappas I.O., Polyviou A., Park S.M., Pandey N., Queiroz M.M., Raman R., Rauschnabel P.A., Shirish A., Sigala M., Spanaki K., Tan G.W.H., Tiwari M.K., Viglia G., Samuel Fosso Wamba S.F., *Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy*, Int J inf Manag, 66, 2022, 102542. doi: 10.1016/j.ijinfomgt.2022.102542.
24. Lee1 L.H., Braud T., Zhou P., Wang L., Xu D., Lin Z., Kumar A., Bermejo C., Hui P., *All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda*, J Latex Class Files, 14 (8), 2021, s. 1-67. doi: 10.13140/rg.2.2.11200.05124/8.
25. Rahaman T., *Into the Metaverse - Perspectives on a New Reality*, Med Ref Serv Q, 41(3), 2022, s. 330-337. doi: 10.1080/02763869.2022.2096341.
26. Zhang X., Chen Y., Hu L., Wang Y., *The metaverse in education: Definition, framework, features, potential applications, challenges, and future research topics*, Front Psychol, 13, 2022, 1016300. doi: 10.3389/fpsyg.2022.1016300.
27. Massetti M., Chiariello G.A., *The metaverse in medicine*, Eur Heart J Suppl, 25 (Suppl B), 2023, 104-107. doi: 10.1093/eurheartjsupp/suad083.

28. Riva G,, Wiederhold B,K,, *What the Metaverse Is (Really) and Why We Need to Know About It*, Cyberpsychol Behav Soc Netw, 25(6), 2022, s. 355-359. doi: 10.1089/cyber.2022.0124.
29. Shu S., Woo B.K.P., *Pioneering the Metaverse: The Role of the Metaverse in an Aging Population*, JMIR Aging, 6, 2023, e40582. doi: 10.2196/40582.
30. Taylor S., Soneji S., *Bioinformatics and the Metaverse: Are We Ready?*, Front Bioinform, 2, 2022, 863676. doi: 10.3389/fbinf.2022.863676.
31. Yang S., *Storytelling and user experience in the cultural metaverse*, Heliyon, 9(4), 2023, e14759. doi: 10.1016/j.heliyon.2023.e14759.
32. Cho K.H., Park J.B., Kang A., *Metaverse for Exercise Rehabilitation: Possibilities and Limitations*, Int J Environ Res Public Health, 20(8), 2023, 5483. doi: 10.3390/ijerph20085483.
33. Chow Y.W., Susilo W., Li Y., Li N., Nguyen C., *Visualization and Cybersecurity in the Metaverse: A Survey*, J Imaging, 9(1), 2022, s. 11. doi: 10.3390/jimaging9010011.
34. Paquin V., Ferrari M., Sekhon H., Rej S., *Time to Think "Meta": A Critical Viewpoint on the Risks and Benefits of Virtual Worlds for Mental Health*, JMIR Serious Games, 11, 2023, e43388. doi: 10.2196/43388.
35. Pawlowski M., *Neuroforma - innovative solution for neurorehabilitation*, Journal of Alzheimers & Dementia 2019, 9th International Conference on Stroke and Cerebrovascular Diseases, July 27-28, 2020. https://www.primescholars.com/articles/neuroforma--innovative-solution-for-neurorehabilitation.pdf.
36. *Festiwal Nauki w IMDiK PAN*: https://www.youtube.com/watch?v=SCMcvr13kcs. [data dostępu: 03.04.2023].
37. Jain A.K., Duin R..PW., Mao J., *Statistical Pattern Recognition: A Review*, IEEE TRANS PATTERN ANAL MACH INTELL, 22(1), 2000, s. 1-34. doi: 10.1109/34.824819.
38. Jóźwik A., Sokołowska B., Niebroj-Dobosz I., Janik P., Kwieciński H., *Extraction of biomedical traits for patients with Amyotrophic Lateral Sclerosis using parallel and hierarchical classifiers*, Int J Biom, 2011, s. 85-94. doi: 10.1504/IJBM.2011.037716.
39. Sokołowska B., Jóźwik A., Niebrój-Dobosz I., Janik P., Kwieciński H., *Evaluation of matrix metalloproteinases in serum of patients with amyotrophic lateral sclerosis with pattern recognition methods*, J Physiol Pharmacol, 60 (Suppl 5), 2009, s. 117-120.
40. Jóźwik A., *A learning scheme for a fuzzy k-NN rule*, Pattern Recognit Lett, 1, 1983, s. 287-289.
41. Pedrycz W,, *Fuzzy sets in pattern recognition: Methodology and methods*, Pattern Recognit, 23(1-2), 1990, s. 121-146.
42. Sokołowska B., Jóźwik A., Pokorski M., *A fuzzy-classifier system to distinguish respiratory patterns evolving after diaphragm paralysis in the cat*, Jpn J Physiol, 53(4), 2003, s. 301-307. doi: 10.2170/jjphysiol.53.301.
43. Sokołowska B., Jóźwik A., Niebroj-Dobosz I., Janik P., Kwieciński H., *Analysis of matrix metalloproteinases (MMPs) in cerebrospinal fluid of patients with amyotrophic lateral sclerosis (ALS)*, Journal of Medical Informatics & Technologies, 13, 2009, s. 147-150.
44. Devijver P.A., Kittler J., *Pattern Recognition: A Statistical Approach*, Prentice-Hall, London 1982.
45. Sokołowska B., Jóźwik A., *A pattern recognition method to distinguish gradual unilateral diaphragm paralysis in the cat*, J Physiol Pharmacol, 54 (Suppl 1), 2003, s. 42-47.
46. Niebroj-Dobosz I., Sokołowska B., Madej-Pilarczyk A., Marchel M., Hausmanowa-Petrusewicz I., *Cardiovascular risk markers in dilated cardiomyopathy in Emery-Dreifuss muscular dystrophy (EDMD)*, Int J Cardiol, 173(2), 2014, s. 324-325. doi: 10.1016/j.ijcard.2014.03.058.

47. Sokołowska B., *Przykłady zastosowania statystycznego modelu z klasyfikacją w neurologii*, [w:] Maciąg M., Danielewska A. (red.), *Problematyka z zakresu Medycyny i Nauk Pokrewnych – Przegląd i Badania. Tom 1*, Wydawnictwo Naukowe TYGIEL, Lublin 2019, s. 198-227.
48. Sokołowska B., *Przykłady oceny wybranych markerów w chorobach narządu ruchu z zastosowaniem statystycznych modeli z klasyfikacją,* [w:] Maciąg M., Talarek K. (red.), *Przegląd Zagadnień z Fizjoterapii i Sportu*, Wydawnictwo Naukowe TYGIEL, Lublin 2020, s. 123-148.
49. Fix E., Hodges J.L, *Discriminatory Analysis: Small Sample Performance. Technical Report Randolph Field*, TX. USAF School of Aviation Medicine, 1952.
50. Bezdek J.C., Chuah S.K., Leep D., *Generalized K-Nearest Neighbor. Rules*, Fuzzy Sets and Systems, 18, 1986, s. 237-256.
51. Carpenter G.A., Grossberg S., *Learning, categorization, rule formation, and prediction by fuzzy neural networks*, [w:] Chen C.H. (red.), *Fuzzy logic and neural network handbook*, McGraw-Hill Series on Computer Engineering, New York 1996, s. 3-45.
52. Jóźwik A., *Nonparametric methods of supervised classification*, Journal of Medical Informatics & Technologies, 22, 2013, s. 21-32.
53. Sokołowska B., Sadura-Sieklucka T., Czerwosz L., Hallay-Suszek M., Lesyng B., Księżopolska-Orłowska K., *Estimation of Posturographic Trajectory Using k-Nearest Neighbors Classifier in Patients with Rheumatoid Arthritis and Osteoarthritis*, Adv Exp Med Biol, 1070, 2018, s. 85-95. doi: 10.1007/5584_2018_150.
54. Carter J.V., Pan J., Rai S.N., Galandiuk S., *ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves*, Surgery, 159(6), 2016, s. 1638-1645. doi: 10.1016/j.surg.2015.12.029.
55. Hajian-Tilaki K., *Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation*, Caspian J Intern Med, 4(2), 2013, s. 627-635.
56. Kumar R., Indrayan A., *Receiver operating characteristic (ROC) curve for medical researchers*, Indian Pediatr, 48(4), 2011, s. 277-287. doi: 10.1007/s13312-011-0055-4.
57. Lasko T.A., Bhagwat J.G., Zou K.H., Ohno-Machado L., *The use of receiver operating characteristic curves in biomedical informatics*, J Biomed Inform, 38(5), 2005, s. 404-415. doi: 10.1016/j.jbi.2005.02.008.
58. Niebroj-Dobosz I., Janik P., Sokołowska B., Kwiecinski H., *Matrix metalloproteinases and their tissue inhibitors in serum and cerebrospinal fluid of patients with amyotrophic lateral sclerosis*, Eur J Neurol, 17(2), 2010, s. 226-231. doi: 10.1111/j.1468-1331.2009.02775.x.
59. Sokołowska B,., Jóźwik A., *Recognition of cycles of repeated hypoxia on the basis of time periods in biological model*, [w:] Kurzyński M., Puchała E., Woźniak M. (red.), *Computer Recognition Systems 2. Advances in Intelligent and Soft Computing*, Springer- Verlag Berlin, Heiderberg, 45, 2007, s. 778-785.
60. Sokołowska B., Jóźwik A., *Ventilatory response to hypoxia in experimental pathology of the diaphragm*, J Physiol Pharmacol, 55 (Suppl III), 2004, s. 121-126.
61. Welch J.F., Nair J., Argento P.J., Mitchell G.S., Fox E.J., *Acute intermittent hypercapnic-hypoxia elicits central neural respiratory motor plasticity in humans*, J Physiol, 600(10), 2022, s. 2515-2533. doi: 10.1113/JP282822.
62. Nichols N.L., Mitchell G.S., *Mechanisms of severe acute intermittent hypoxia-induced phrenic long-term facilitation*, J Neurophysiol, 125(4), 2021, s. 1146-1156, doi: 10.1152/jn.00691.2020.
63. Perim R.R., *Differential spinal tissue oxygen profile during moderate and severe acute intermittent hypoxia*, FASEB J, 34(1), 2020, s. 1. doi: 10.1096/fasebj.2020.34.s1.03195.
64. Dougherty B.J., Terada J., Springborn S.R., Vinit S., MacFarlane P.M., Mitchell G.S., *Daily acute intermittent hypoxia improves breathing function with acute and chronic*

spinal injury via distinct mechanisms, Resp Physiol Neurobiol, 256, 2018, s. 50-57, doi: 10.1016/j.resp.2017.05.004.
65. Hassan A., Arnold B.M., Caine S., Toosi B.M., Verge V.M.K., Muir G.D., *Acute intermittent hypoxia and rehabilitative training following cervical spinal injury alters neuronal hypoxia- and plasticity-associated protein expression*, PLoS One, 13(5), 2018, 197486, doi: 10.1371/journal.pone.0197486.
66. Xi L., Serebrovskaya T.V., *Intermittent Hypoxia. From Molecular Mechanisms to Clinical Applications*, Nova Science Publishers, 2009.
67. Mendonça-Junior B.A., Fernandes M.V., Zoccal D.B., *Acute intermittent hypoxia evokes ventilatory long-term facilitation and active expiration in unanesthetized rats*, Resp Physiol Neurobiol, 294, 2021, 103768. doi: 10.1016/j.resp.2021.103768.
68. Naidu A., Peters D.M., Tan A.Q., Barth S., Crane A., Link A., Balakrishnan S., Hayes H.B., Slocum C., Zafonte R.D., Trumbower R.D., *Daily acute intermittent hypoxia to improve walking function in persons with subacute spinal cord injury: a randomized clinical trial study protocol*, BMC Neurol, 20(1), 2020, s. 273. doi: 10.1186/s12883-020-01851-9.
69. Mateika J.H., Komnenov D., *Intermittent hypoxia initiated plasticity in humans: A multipronged therapeutic approach to treat sleep apnea and overlapping co-morbidities*, Exp Neurol, 287(2), 2017, s. 113-129. doi: 10.1016/j.expneurol.2016.05.011.
70. Serebrovskaya T.V. (red.), *Intermittent Hypoxia and Human Diseases*, Springer-Verlag, London 2012. doi: 10.1007/978-1-4471-2906-6.
71. Mikulski T., *Zastosowanie treningu hipoksyjnego w warunkach nizinnych u lekkoatletów*, Sport Wyczynowy, 4(536), 2010, s. 100-105.
72. Mateika J.H., Narwani G., *Intermittent hypoxia and respiratory plasticity in humans and other animals: does exposure to intermittent hypoxia promote or mitigate sleep apnoea?*, Exp Physiol, 94(3), 2009, s. 279-296. doi: 10.1113/expphysiol.2008.045153.
73. Sokołowska B., Pokorski M., *Ventilatory augmentation by acute intermittent hypoxia in the rabbit*, J Physiol Pharmacol, 57(4), 2006, s. 341-347.
74. Navarrete-Opazo A., Mitchell G.S., *Therapeutic potential of intermittent hypoxia: a matter of dose*, Am J Physiol Regul Integr Comp Physiol, 307(10), 2014, s. 1181-1197. doi: 10.1152/ajpregu.00208.2014.
75. Mateika J.H., Sandhu K.S., *Experimental protocols and preparations to study respiratory long term facilitation*, Resp Physiol Neurobiol, 176(1-2), 2011, s. 1-11. doi: 10.1016/j.resp.2011.01.007.
76. Sokolowska B., Jozwik A., *Recognition of cycles of repeated hypoxia on the basis of time periods in biological model*, Advances in Soft Computing, 45, 2007, s. 778-785. doi: 10.1007/978-3-540-75175-5_96.
77. Vermeulen T.D., Benbaruj J., Brown C.V., Shafer B.M., Floras J.S., Foster G.E., *Peripheral chemoreflex contribution to ventilatory long-term facilitation induced by acute intermittent hypercapnic hypoxia in males and females*, J Physiol, 598(20), 2020, s. 4713-4730. doi: 10.1113/JP280458.
78. Mateika J.H., El-Chami M., Shaheen D., Ivers B., *Intermittent hypoxia: a low-risk research tool with therapeutic value in humans*, J Appl Physiol, 118(5), 2015, s. 520-532. doi: 10.1152/japplphysiol.00564.2014.
79. Fogarty M.J., Sieck G.C., *Diaphragm Muscle Adaptations in Health and Disease*, Drug Discov Today Dis Models, 2019, s. 29-30, 43-52. doi: 10.1016/j.ddmod.2019.10.002.
80. McCool F.D., Manzoor K., Minami T., *Disorders of the Diaphragm*, Clin Chest Med, 39(2), 2018, s. 345-360. doi: 10.1016/j.ccm.2018.01.012.
81. Janik P., Kwiecinski H., Sokolowska B., Niebroj-Dobosz I., *Erythropoietin concentration in serum and cerebrospinal fluid of patients with amyotrophic lateral sclerosis*, J Neural Transm (Vienna), 117(3), 2010, s. 343-347. doi: 10.1007/s00702-009-0354-2.

82. Niebroj-Dobosz I., Madej-Pilarczyk A., Marchel M., Sokołowska B., Hausmanowa-Petrusewicz I., *Matrix metalloproteinases in serum of Emery-Dreifuss muscular dystrophy patients*, Acta Biochim Pol, 56(4), 2009, s. 717-722.
83. Niebroj-Dobosz I.M., Sokołowska B., Madej-Pilarczyk A., Marchel M., Hausmanowa-Petrusewicz I., *Tissue inhibitors of matrix metalloproteinases in serum are cardiac biomarkers in Emery-Dreifuss muscular dystrophy*, Kardiol Pol, 73(5), 2015, s. 360-365. doi: 10.5603/KP.a2014.0243.
84. Molière S., Jaulin A., Tomasetto C.L., Dali-Youcef N., *Roles of Matrix Metalloproteinases and Their Natural Inhibitors in Metabolism: Insights into Health and Disease*, Int J Mol Sci, 24(13), 2023, s. 10649. doi: 10.3390/ijms241310649.
85. Cabral-Pacheco G.A., Garza-Veloz I., Castruita-De la Rosa C., Ramirez-Acuña J.M., Perez-Romero B.A., Guerrero-Rodriguez J.F., Martinez-Avila N., Martinez-Fierro M.L., *The Roles of Matrix Metalloproteinases and Their Inhibitors in Human Diseases* Int J Mol Sci, 21(24), 2020, s. 9739. doi: 10.3390/ijms21249739.
86. Rivera S., García-González L., Khrestchatisky M., Baranger K., *Metalloproteinases and their tissue inhibitors in Alzheimer's disease and other neurodegenerative disorders*, Cell Mol Life Sci, 76(16), 2019, s. 3167-3191. doi: 10.1007/s00018-019-03178-2.
87. Khalil R.A. (red.), *Matrix Metalloproteinases and Tissue Remodeling in Health and Disease: Target Tissues and Therapy*, Prog Mol Biol Transl Sci, USA, vol 148, 2017.
88. Czerwosz L., Szczepek E., Sokolowska B., Jurkiewicz J., Czernicki Z., *Posturography in differential diagnosis of normal pressure hydrocephalus and brain atrophy*, Adv Exp Med Biol, 755, 2013, s. 311-324. doi: 10.1007/978-94-007-4546-9_40.
89. Czerwosz L., Szczepek E., Sokołowska B., Jurkiewicz J., Czernicki Z., *Recognition of posture and gait disturbances in patients with normal pressure hydrocephalus using a Posturography and Computer Dynography Systems*, [w:] Sadip D. (red.), *Hydrocephalus*, InTech Publisherm, 12, 2012, s. 189-214.
90. Czerwosz L., Szczepek E., Blaszczyk J.W., Sokolowska B., Dmitruk K., Dudzinski K., Jurkiewicz J., Czernicki Z., *Analysis of postural sway in patients with normal pressure hydrocephalus: effects of shunt implantation*, Eur J Med Res, 14 (Suppl 4), 2009, s. 53-58. doi: 10.1186/2047-783x-14-s4-53.
91. Sadura-Sieklucka T., Czerwosz L.T., Kądalska E., Kożuchowski M., Księżopolska-Orłowska K., Targowski T., *Is Balance Training Using Biofeedback Effective in the Prophylaxis of Falls in Women over the Age of 65?*, Brain Sci, 13(4), 2023, s. 629. doi: 10.3390/brainsci13040629.
92. Sokołowska B., *A novel virtual reality approach for functional lateralization in healthy adults*, Brain Res, 1766, 2021, 147537. doi: 10.1016/j.brainres.2021.147537.
93. Sokołowska B., Sokołowska E., *Machine Learning algorithms to study of laterality in Virtual Reality*, The annual INCF Congress of NeuroInformatics 2019, September 1-2, Warsaw, Neuroinformatics, 2019. https://doi.org/10.12751/incf.ni2019.0053.
94. Erviana V.Y., Ghufron A., Retnawati H., *Validity and Reliability of IoJEPD Model Instruments to Improve Elementary School Teacher Competence*, Pegem J Educ Instr, 13(2), 2023, s. 200-206. doi: 10.47750/pegegog.13.02.24.
95. Msika E.F., Ehrlé N., Gaston-Bellegarde A., Orriols E., Piolino P., Narme P., *Using a Computer-Based Virtual Environment to Assess Social Cognition in Aging: An Exploratory Study of the REALSoCog Task*, Front Psychol, 13, 2022, 882165. doi: 10.3389/fpsyg.2022.882165.
96. Brassel S., Power E., Campbell A., Brunner M., Togher L., *Recommendations for the Design and Implementation of Virtual Reality for Acquired Brain Injury Rehabilitation: Systematic Review*, J Med Internet Res, 23(7), 2021, e26344. doi: 10.2196/26344.
97. Kourtesis P., Collina S., Doumas L.A.A., MacPherson S.E., *Validation of the Virtual Reality Everyday Assessment Lab (VR-EAL): An Immersive Virtual Reality*

Neuropsychological Battery with Enhanced Ecological Validity, J Int Neuropsychol Soc, 27, 2021, s. 181-196. doi: 10.1017/S1355617720000764.
98. Sokołowska B., *Impact of Virtual Reality Cognitive and Motor Exercises on Brain Health*, Int J Environ Res Public Health, 20(5), 2023, 4150. doi: 10.3390/ijerph20054150.
99. Sokołowska B., *Accuracy of cognitive function and functional lateralization assessment using a novel virtual reality tool in neuroscience*, Przygotowany manuskrypt.
100. Sokołowska B., *Przykłady zastosowania i znaczenie podejścia z wirtualną rzeczywistością w terapii, rehabilitacji i diagnostyce*, [w:] Bujalska B., Maciąg K. (red.), *Nowoczesne technologie - strategie, rozwiązania i perspektywy rozwoju. Tom 1*, Wydawnictwo Naukowe TYGIEL, Lublin 2021, s. 127-140.
101. Sokołowska B., Sokołowska E., *Przykłady funkcjonalnego modelowania i oceny testów i ćwiczeń w Wirtualnej Rzeczywistości*, [w:] Bujalska B., Maciąg K. (red.), *Nowoczesne technologie - strategie, rozwiązania i perspektywy rozwoju. Tom 1*, Wydawnictwo Naukowe TYGIEL, Lublin 2021, s. 141-155.
102. Sokołowska B., Sokołowska E., *Zastosowania i perspektywy wirtualnej posturografii*, [w:] Mołdoch-Mendoń I., Danielewska A. (red.), *Wybrane zagadnienia z zakresu psychologii i medycyny*, Wydawnictwo Naukowe TYGIEL, Lublin 2020, s. 295-311.
103. Sokołowska B., Sadura-Sieklucka T., Sokołowska E., *Przykłady badania i oceny funkcji ręki dominującej i niedominującej w warunkach zdrowia i choroby*, [w:] Maciąg M, Danielewska A (red.), *Problematyka z zakresu medycyny i nauk pokrewnych – przegląd i badania. Tom 1*, Wydawnictwo Naukowe TYGIEL, Lublin 2019, s.228-247.
104. Sokołowska B., Sokołowska E., *An Example of the Assessment of Human Laterality Using Virtual reality*, IJIRCCE, 7(7), 2019, s. 3555-3562. http://ijircce.com/admin/main/storage/app/pdf/KdxZlKhLTCPJ5Fekz5R9wfuhT7qtfDYfS9dn2Atb.pdf.
105. Ferrero L., Ortiz M., Quiles V., Iáñez E., Flores J.A., Azorín J.M., *Brain Symmetry Analysis during the Use of a BCI Based on Motor Imagery for the Control of a Lower-Limb Exoskeleton*, Symmetry, 13(9), 2021, s. 1746. doi: 10.3390/sym13091746.
106. Palomero-Gallagher N., Amunts K., *A short review on emotion processing: a lateralized network of neuronal networks*, Brain Struct Funct, 227(2), 2022, s. 673-684. doi: 10.1007/s00429-021-02331-7.
107. Szewczyk D., Sadura-Sieklucka T., Sokołowska B., Księżopolska-Orłowska K., *Improving the quality of life of patients with rheumatoid arthritis after rehabilitation irrespective of the level of disease activity*, Rheumatol Int, 41(4), 2021, s. 781-786. doi: 10.1007/s00296-020-04711-4.
108. Sadura-Sieklucka T., Sokołowska B., Prusinowska A., Trzaska A., Księżopolska-Orłowska K., *Benefits of wrist splinting in patients with rheumatoid arthritis*, Reumatologia, 56(6), 2018, s. 362-367. doi: 10.5114/reum.2018.80713.
109. Księżopolska-Orłowska K., Sadura-Sieklucka T., Kasprzak K., Gaszewska E., Rodkiewicz-Bogusławska A., Sokołowska B., *The beneficial effects of rehabilitation on hand function in patients with rheumatoid arthritis*, Reumatologia, 54(6), 2016, s. 285-290. doi: 10.5114/reum.2016.64903.
110. Sokołowska B., Czerwosz L., Hallay-Suszek M., Sadura-Sieklucka T., Księżopolska-Orłowska K., *Posturography in patients with rheumatoid arthritis and osteoarthritis*, Adv Exp Med Biol, 833, 2015, s. 63-70. doi: 10.1007/5584_2014_29.
111. Tasci Bozbas G., Sendur O.F., Aydemir A.H., *Primary knee osteoarthritis increases the risk of falling*, J Back Musculoskelet Rehabil, 30(4), 2017, s. 785-789. doi: 10.3233/BMR-150413.
112. De Lorenzis E., Crudo F., Fedele A.L., Fiorita A., Bruno D., Paludetti G., Alivernini S., Giraldi L., Picciotti P.M., Zoli A., Cadoni G., *Postural control and disability in patients*

with early rheumatoid arthritis, Clin Exp Rheumatol, 39(6), 2021, s. 1369-1377. doi: 10.55563/clinexprheumatol/hkfeur.

113. Konarzewski P., Konarzewska U., Kuryliszyn-Moskal A., Terlikowski R., Pauk J., Daunoraviciene K., Pauk K., Dakowicz A., Wojciuk M., Dzięcioł J., Dziecioł-Anikiej Z., *What Influences Proprioceptive Impairments in Patients with Rheumatic Diseases? Analysis of Different Factors*, Int J Environ Res Public Health, 20(4), 2023, s. 3698. doi: 10.3390/ijerph20043698.
114. Shen D., Li Z., Shan Z., Wang J,. Ren Y., *Relationship between the Degree of Illness in Elderly Patients with Rheumatoid Arthritis and Parameters of Musculoskeletal Ultrasound and Oswestry Dysfunction Index and Clinical Value Analysis u* Evid Based Complement Alternat Med, 2022, 4307864. doi: 10.1155/2022/4307864.
115. Sokołowska B., Jóźwik A., Niebroj-Dobosz I., Janik P., *The pair-wise linear classifier and the k-nn rule in application to ALS progression differentiation*, Journal of Medical Informatics & Technologies, 20, 2012, s. 79-83.
116. Jóźwik A., Sokołowska B., Niebroj-Dobosz I., Janik P., Kwieciński H., *Pattern recognition approach to differentiation of disease severity in patients with amyotrophic lateral sclerosis*, Journal of Medical Informatics & Technologies, 12, 2008, s. 143-147.
117. Behl T., Kaur G., Sehgal A., Bhardwaj S., Singh S., Buhas C., Judea-Pusta C., Uivarosan D., Munteanu M.A., Bungau S., *Multifaceted Role of Matrix Metalloproteinases in Neurodegenerative Diseases: Pathophysiological and Therapeutic Perspectives*, Int J Mol Sci, 22(3), 2021, s. 1413. doi: 10.3390/ijms22031413.
118. Cabral-Pacheco G.A., Garza-Veloz I., Castruita-De la Rosa C., Ramirez-Acuña J.M., Perez-Romero B.A., Guerrero-Rodriguez J.F., Martinez-Avila N., Martinez-Fierro M.L., *The Roles of Matrix Metalloproteinases and Their Inhibitors in Human Diseases*, Int J Mol Sci, 21(24), 2020, s. 9739. doi: 10.3390/ijms21249739.
119. Carilho R., de Carvalho M., Kuehl U., Pinto S., Pinto A., Kromminga A., Costa J., *Erythropoietin and amyotrophic lateral sclerosis: Plasma level determination*, Amyotroph Lateral Scler, 12(6), 2011, s. 439-443. doi: 10.3109/17482968.2011.584881.
120. Chakraborty A., Diwan A., *Biomarkers and molecular mechanisms of Amyotrophic Lateral Sclerosis*, AIMS Neurosci, 9(4), 2022, s. 423-443. doi: 10.3934/Neuroscience.2022023.
121. Chen X., Shang H.F., *New developments and future opportunities in biomarkers for amyotrophic lateral sclerosis*, Transl Neurodegener, 4, 2015, s. 17. doi: 10.1186/s40035-015-0040-2.
122. Kim H.Y., Moon C., Kim K.S., Oh K.W., Oh S.I., Kim J., Kim S.H., *Recombinant human erythropoietin in amyotrophic lateral sclerosis: a pilot study of safety and feasibility*, J Clin Neurol, 10(4), 2014, s. 342-347. doi: 10.3988/jcn.2014.10.4.342.
123. Łukaszewicz-Zając M., Mroczko B., Słowik A., *Matrix metalloproteinases (MMPs) and their tissue inhibitors (TIMPs) in amyotrophic lateral sclerosis (ALS)*, J Neural Transm (Vienna), 121(11), 2014, s. 1387-1397. doi: 10.1007/s00702-014-1205-3.
124. Sánchez-Torres J.L., Yescas-Gómez P., Torres-Romero J., Espinosa O.R., Canovas L.L., Tecalco-Cruz Á.C., Ponce-Regalado M.D., Alvarez-Sánchez M.E., *Matrix metalloproteinases deregulation in amyotrophic lateral sclerosis*, J Neurol Sci, 419, 2020, 117175. doi: 10.1016/j.jns.2020.117175.
125. Tarasiuk J., Kułakowska A., Drozdowski W., Kornhuber J., Lewczuk P., *CSF markers in amyotrophic lateral sclerosis*, J Neural Transm (Vienna), 119(7), 2012, s. 747-757. doi: 10.1007/s00702-012-0806-y.
126. Niebroj-Dobosz I., Sokołowska B., Madej-Pilarczyk A., Marchel M., Hausmanowa-Petrusewicz I., *Dysfunctional lamins as mediators of oxidative stress in Emery-Dreifuss muscular dystrophy*, Folia Neuropathol, 55(3), 2017, s. 193-198. doi: 10.5114/fn.2017.70483.

127. Sokołowska B., Jóźwik A., Niebroj-Dobosz I., Hausmanowa-Petrusewicz I., *A pattern recognition approach to Emery-Dreifuss muscular dystrophy (EDMD) study*, Journal of Medical Informatics & Technologies, 23, 2014, s. 165-171.
128. Niebroj-Dobosz I.M., Sokołowska B., Madej-Pilarczyk A., Marchel M., Hausmanowa-Petrusewicz I., *Natriuretic peptides assessment in dilated cardiomyopathy in patients withEmery-Dreifuss muscular dystrophy*, J Clin Exp Cardiol, 3(8), 2012, s. 1-7.
129. Niebroj-Dobosz I., Madej-Pilarczyk A., Marchel M., Sokołowska B., Hausmanowa-Petrusewicz I., *Osteopontin--a fibrosis-related marker-in dilated cardiomyopathy in patients with Emery-Dreifuss muscular dystrophy*, Scand J Clin Lab Invest, 71(8), 2011, s. 658-662. doi: 10.3109/00365513.2011.619272.
130. Niebroj-Dobosz I., Madej-Pilarczyk A., Marchel M., Sokołowska B., Hausmanowa-Petrusewicz I., *Circulating tenascin-C levels in patients with dilated cardiomyopathy in the course of Emery-Dreifuss muscular dystrophy*, Clin Chim Acta, 412(17-18), 2011, 1533-1538. doi: 10.1016/j.cca.2011.04.033.
131. Niebroj-Dobosz I., Marchel M., Madej A., Sokolowska B., Hausmanowa-Petrusewicz I., *Circulating autoantibodies to troponin I in Emery-Dreifuss muscular dystrophy*, Acta Myol, 27(1), 2008, s. 1-6.
132. Iskandar K., Astari F.N., Gumilang R.A., Ilma N., Shartyanie N.P., Adistyawan G., Tan G., Lai P.S., *Autosomal dominant Emery-Dreifuss muscular dystrophy caused by a mutation in the lamin A/C gene identified by exome sequencing: a case report*, BMC Pediatr, 22(1), 2022, s. 601. doi: 10.1186/s12887-022-03662-y.
133. Jędrzejowska M., Potulska-Chromik A., Gos M., Gambin T., Dębek E., Rosiak E., Stępień A., Szymańczak R., Wojtaś B., Gielniewski B., Ciara E., Sobczyńska A., Chrzanowska K., Kostera-Pruszczyk A., Madej-Pilarczyk A., *Floppy infant syndrome as a first manifestation of LMNA-related congenital muscular dystrophy*, Eur J Paediatr Neurol, 32, 2021, s. 115-121. doi: 10.1016/j.ejpn.2021.04.005.
134. Lupica A., Di Stefano V., Gagliardo A., Iacono S., Pignolo A., Ferlisi S., Torrente A., Pagano S., Gangitano M., Brighina F., *Inherited Neuromuscular Disorders: Which Role for Serum Biomarkers?*, Brain Sci, 11(3), 2021, s. 398. doi: 10.3390/brainsci11030398.
135. Łoboda A., Dulak J., *Nuclear Factor Erythroid 2-Related Factor 2 and Its Targets in Skeletal Muscle Repair and Regeneration*, Antioxid Redox Signal, 38(7-9), 2023, s. 619-642. doi: 10.1089/ars.2022.0208.
136. Madej-Pilarczyk A., *Clinical aspects of Emery-Dreifuss muscular dystrophy*, Nucleus, 9(1), 2018, s. 268-274. doi: 10.1080/19491034.2018.1462635.
137. Marchel M., Madej-Pilarczyk A., Steckiewicz R., Stolarz P., Peller M., Tymińska A., Ostrowska E., Ozierański K., Balsam P., Grabowski M., Opolski G., *Predictors of mortality and cardiovascular outcomes in Emery-Dreifuss muscular dystrophy in a long-term follow-up*, Kardiol Pol, 79(12), 2021, s. 1335-1342. doi: 10.33963/KP.a2021.0159.
138. Politano L., Carboni N., Madej-Pilarczyk A., Marchel M., Nigro G., Fidziaóska A., Opolski G., Hausmanowa-Petrusewicz I., *Advances in basic and clinical research in laminopathies*, Acta Myol, 32(1), 2013, s. 18-22.
139. Czerwosz L., Szczepek E., Nowiński K., Sokołowska B., Jurkiewicz J., Czernicki Z., Koszewski W., *Discriminant Analysis of Intracranial Volumetric Variables in Patients with Normal Pressure Hydrocephalus and Brain Atrophy*, Adv Exp Med Biol, 1039, 2018, s. 83-94. doi: 10.1007/5584_2017_75.
140. Szczepek E., Sokołowska B., Czerwosz L., Hallay-Suszek M., Nowiński K., Czernicki Z., Jurkiewicz J., *Statistical analysis of volumetric parameters using k nearest neighbors method and the VisNow software package in patients with normal pressure hydrocephalus and brain atrophy*, Post Nauk Med, XXVIII(8), 2015, s. 617-621.
141. Czerwosz L., Szczepek E., Sokolowska B., Jozwik A., Dudzinski K., Jurkiewicz J., Czernicki Z., *Recognition of gait disturbances in patients with normal pressure*

hydrocephalus using a computer dynography system, J Physiol Pharmacol, 59 (Suppl 6), 2008, s. 201-207.
142. Abram K., Bohne S., Bublak P., Karvouniari P., Klingner C.M., Witte O.W., Guntinas-Lichius O., Axer H., *The Effect of Spinal Tap Test on Different Sensory Modalities of Postural Stability in Idiopathic Normal Pressure Hydrocephalus*, Dement Geriatr Cogn Dis Extra, 6(3), 2016, s. 447-457. doi: 10.1159/000450602.
143. Lundin F., Ledin T., Wikkelsø C., Leijon G., *Postural function in idiopathic normal pressure hydrocephalus before and after shunt surgery: a controlled study using computerized dynamic posturography (EquiTest)*, Clin Neurol Neurosurg, 115(9), 2013, s. 1626-1631. doi: 10.1016/j.clineuro.2013.02.015.
144. Mendes G.A.S., de Oliveira M.F., Pinto F.C.G., *The Timed Up and Go Test as a Diagnostic Criterion in Normal Pressure Hydrocephalus*, World Neurosurg, 105, 2017, s. 456-461. doi: 10.1016/j.wneu.2017.05.137.
145. Nikaido Y., Akisue T., Urakami H., Kajimoto Y., Kuroda K., Kawami Y., Sato H., Ohta Y., Hinoshita T., Iwai Y., Nishiguchi T., Ohno H., Saura R., *Postural control before and after cerebrospinal fluid shunt surgery in idiopathic normal pressure hydrocephalus*, Clin Neurol Neurosurg, 172, 2018, s. 46-50. doi: 10.1016/j.clineuro.2018.06.032.
146. Panciani P.P., Migliorati K., Muratori A., Gelmini M., Padovani A., Fontanella M., *Computerized gait analysis with inertial sensor in the management of idiopathic normal pressure hydrocephalus*, Eur J Phys Rehabil Med, 54(5), 2018, s. 724-729. doi: 10.23736/S1973-9087.18.04949-3.
147. Rydja J., Kollén L., Hellström P., Owen K., Lundgren Nilsson Å., Wikkelsø C., Tullberg M., Lundin F., *Physical exercise and goal attainment after shunt surgery in idiopathic normal pressure hydrocephalus: a randomised clinical trial*, Fluids Barriers CNS, 18(1), 2021, s. 51. doi: 10.1186/s12987-021-00287-8.
148. Tarnaris A., Toma A.K., Kitchen N.D., Watkins L.D., *Ongoing search for diagnostic biomarkers in idiopathic normal pressure hydrocephalus*, Biomark Med, 3(6), 2009, s. 787-805. doi: 10.2217/bmm.09.37.
149. Wu D., Moghekar A., Shi W., Blitz A.M., Mori S., *Systematic volumetric analysis predicts response to CSF drainage and outcome to shunt surgery in idiopathic normal pressure hydrocephalus*, Eur Radiol, 31(7), 2021, s. 4972-4980. doi: 10.1007/s00330-020-07531-z.
150. Sokołowska B., *Układ oddechowy i przykłady (neuro)plastyczności oddechowej w warunkach zdrowia i choroby*, [w:] Kalbarczyk K., Danielewska A. (red.), *Prewencja, diagnostyka i leczenie chorób genetycznych, nowotworowych i OUN*, Wydawnictwo Naukowe TYGIEL, Lublin 2022, s. 97-124.
151. Sokołowska B., Jóźwik A., *Pattern recognition approach for analysis of metabolic response to intermittent hypoxia*, Journal of Medical Informatics & Technologies, 15, 2010, s. 177-183.
152. Sokolowska B., Rekawek A., Jozwik A., *Respiratory responses to acute intermittent hypoxia and hypercapnia in awake rats*, J Physiol Pharmacol, 59 (Suppl 6), 2008, s. 659-667.
153. Sokołowska B., Rękawek A., Jóźwik A., *Recognition of the ventilatory response to the intermittent chemical stimuli in awake animals*, [w:] Pietka E., Kawa J. (red.), *Information Technologies in Biomedicine*, Advances in Soft Computing, 47, 2008, s. 488-494.
154. Sokołowska B., Jóźwik A., *Distinguishing the strength of hypoxic stimulus in intermittent hypoxia*, J Physiol Pharmacol, 58 (Suppl 5), 2007, s. 657-663.
155. Sokołowska B., Jóźwik A., *Statistical evaluation of ventilatory patterns in response to intermittent hypoxia in the rabbit*, J Physiol Pharmacol, 56 (Suppl 4), 2005, s. 203-207.
156. Sokołowska B., Budzińska K., Pokorski M., *Elektromiogram przepony i wzorzec oddechowy po jednostronnym i obustronnym częściowym odnerwieniu przepony u kota*, Pneumonol Alergol Pol, 65(7-8), 1997, s. 500-507.

157. Dylag A.M., Raffay T.M., *Rodent models of respiratory control and respiratory system development-Clinical significance*, Respir Physiol Neurobiol, 268, 2019, s. 103249. doi: 10.1016/j.resp.2019.06.006.
158. Fiz J.A., Jané R., Lozano M., Gómez R., Ruiz J., *Detecting unilateral phrenic paralysis by acoustic respiratory analysis*, PLoS One, 9(4), 2014, e93595. doi: 10.1371/journal.pone.0093595.
159. Gerst D.G. 3rd, Yokhana S.S., Carney L.M., Lee D.S., Badr M.S., Qureshi T., Anthouard M.N., Mateika J.H., *The hypoxic ventilatory response and ventilatory long-term facilitation are altered by time of day and repeated daily exposure to intermittent hypoxia*, J Appl Physiol, 110(1), 2011, s. 15-28. doi: 10.1152/japplphysiol.00524.2010.
160. Mateika J.H., Panza G., Alex R., El-Chami M., *The impact of intermittent or sustained carbon dioxide on intermittent hypoxia initiated respiratory plasticity. What is the effect of these combined stimuli on apnea severity?*, Respir Physiol Neurobiol, 256, 2018, s. 58-66. doi: 10.1016/j.resp.2017.10.008.
161. Mateika J.H., Syed Z., *Intermittent hypoxia, respiratory plasticity and sleep apnea in humans: present knowledge and future investigations*, Respir Physiol Neurobiol, 188(3), 2013, s. 289-300. doi: 10.1016/j.resp.2013.04.010.
162. Peña-Ortega F., *Neural Network Reconfigurations: Changes of the Respiratory Network by Hypoxia as an Example*, Adv Exp Med Biol, 1015, 2017, s. 217-237. doi: 10.1007/978-3-319-62817-2_12.
163. Perim R.R., Sunshine M.D., Welch J.F., Santiago J., Holland A., Ross A., Mitchell G.S., Gonzalez-Rothi E.J., *Daily acute intermittent hypoxia enhances phrenic motor output and stimulus-evoked phrenic responses in rats*, J Neurophysiol, 126(3), 2021, s. 777-790. doi: 10.1152/jn.00112.2021.
164. Perim R.R., Gonzalez-Rothi E.J., Mitchell G.S., *Cervical spinal injury compromises caudal spinal tissue oxygenation and undermines acute intermittent hypoxia-induced phrenic long-term facilitation*, Exp Neurol, 342, 2021, s. 113726. doi: 10.1016/j.expneurol.2021.113726.
165. Perry S.F., Similowski T., Klein W., Codd J.R., *The evolutionary origin of the mammalian diaphragm*, Respir Physiol Neurobiol, 171(1), 2010, s. 1-16. doi: 10.1016/j.resp.2010.01.004.
166. Turimov Mustapoevich D., Muhamediyeva Tulkunovna D., Safarova Ulmasovna L., Primova H., Kim W., *Improved Cattle Disease Diagnosis Based on Fuzzy Logic Algorithms*, Sensors (Basel), 23(4), 2023, s. 2107. doi: 10.3390/s23042107.
167. Xing T., Fong A.Y., Bautista T.G., Pilowsky P.M., *Acute intermittent hypoxia induced neural plasticity in respiratory motor control*, Clin Exp Pharmacol Physiol, 40(9), 2013, s. 602-609. doi: 10.1111/1440-1681.12129.
168. Yokhana S.S., Gerst D.G. 3rd, Lee D.S., Badr M.S., Qureshi T., Mateika J.H., *Impact of repeated daily exposure to intermittent hypoxia and mild sustained hypercapnia on apnea severity*, J Appl Physiol, 112(3), 2012, s. 367-377. doi: 10.1152/japplphysiol.00702.2011.

Przykłady oceny parametrów i/lub markerów w modelowym podejściu z klasyfikacją nadzorowaną w badaniach w realnej i wirtualnej rzeczywistości

Streszczenie

W pracy przedstawiono podstawy modelowego podejścia badawczo-analitycznego z zastosowaniem statystycznych metod klasyfikacji nadzorowanej. W podejściu tym oparto się na bardzo efektywnych i dokładnych klasyfikatorach ze statystycznej teorii rozpoznawania obrazów oraz włączono klasyczny klasyfikator binarny i metodę krzywych *ROC*. Proponowane podejście zastosowano w prowadzonych badaniach biomedycznych, obejmujących badania modelowe, eksperymentalne i kliniczne. We własnych badaniach, m.in. opierając się właśnie na przedstawionych algorytmach klasyfikacyjnych opracowano efektywne eksperymentalne procedury, wykorzystane następnie w modelach eksperymentalnych nad badaniem plastyczności odde-

chowej, na przykład zjawiska *vLTF (ang. ventilatory long-term facilation)*, jak również plastyczności mięśni oddechowych, w eksperymentalnym modelowaniu różnych dysfunkcji oddechowych. W badaniach klinicznych nad oceną i znaczeniem różnych markerów w chorobach nerwowo-mięśniowych, przedstawione i przyjęte w badaniach podejście analityczno-badawcze skutecznie wsparło klasyczne analizy statystyczne, a ponadto pozwoliło na wskazanie wśród nich najbardziej efektywnych biomarkerów, pozwalających na perfekcyjne rozpoznanie jednostki chorobowej, stanu pacjenta (progresu choroby). Prezentowane metody klasyfikacyjne są równie skuteczne w wirtualnej rzeczywistości. Na przykład, mogą być stosowane w precyzyjnej ocenie i różnicowaniu stopni trudności wirtualnych testów i ćwiczeń. Ma to kluczowe znaczenie w interpretacji wirtualnych wyników i w badaniu procesów poznawczych i motorycznych oraz funkcji wykonawczych w neuronaukach zarówno podstawowych jak i klinicznych wykorzystujących wirtualne środowiska.
Słowa kluczowe: markery, klasyfikacja, klasyfikator *k-NN*, metoda krzywych *ROC*, wirtualna rzeczywistość

Examples of parameter and/or marker evaluation in a model-based approach with supervised classification in real and virtual reality environments

Abstract

The paper presents the basics of model testing and the analytical approach using statistically supervised classification methods. This approach is based on highly efficient and accurate classifiers from the statistical theory of pattern recognition and the classical binary classifier and the ROC curve method. The proposed approach was used in biomedical research, including modeling, experimental and clinical studies. In the research presented here, effective experimental procedures were developed based on these classification algorithms, which were then used in experimental models to study respiratory plasticity, such as the ventilatory long-term facilitation (*vLTF*) phenomenon, as well as respiratory muscle plasticity, in experimental modeling of various respiratory dysfunctions. In clinical studies on the evaluation and significance of various markers in neuromuscular diseases, the analytical and research approach presented and adopted in the study effectively supported classical statistical analyses, and also allowed to identify of the most effective biomarkers among them, allowing excellent diagnosis of the disease entity, the patient's condition (progress of the disease). The classification methods presented are equally effective in virtual reality. For example, in accurately assessing and differentiating the difficulty of virtual tests and exercises. This is crucial for interpreting virtual results and studying cognitive and motor processes and executive functions in both basic and clinical neuroscience using virtual environments.
Keywords: markers, classification, *k-NN* classifier, *ROC* curve method, virtual reality

Patryk Warzybok[1], Grzegorz Dziatkiewicz[2]

Planowanie ścieżki uczenia się i modelowanie dynamiki zmian kompetencji inżynierskich w zakresie obsługi systemów klasy CAx

1. Wprowadzenie

Zmieniające się otoczenie biznesowe oraz nieustanny rozwój projektowanych i wytwarzanych produktów stwarza potrzebę kształcenia inżynierów posiadających specjalistyczne i interdyscyplinarne kompetencje [1]. Nieustanna presja wynikająca z ograniczeń czasowych w procesach projektowania i wytwarzania, jak i konieczność minimalizacji kosztów na każdym etapie cyklu życia produktu, a także rozwój systemów wspomagających prace inżynierskie powoduje, że zakres wymaganych kompetencji inżynierskich zmienia się i nie przystaje do już uzyskanych podczas tradycyjnego akademickiego procesu kształcenia i uczenia się.

Zmiana metody wytwarzania, stosowanego oprogramowania inżynierskiego, czy samego produktu może wymagać relatywnie szybkiego nabycia nowych kompetencji na takim poziomie, który pozwoli osiągnąć zakładane przez pracodawcę cele.

Popularną i bardzo często stosowaną współcześnie metodą podnoszenia poziomu kompetencji, jak i nabywania nowych, jest szkolenie [2]. Charakteryzuje się ono intensywnym transferem wiedzy i umiejętności w stosunkowo krótkim okresie czasu, gdzie intensywność wyrażona może być poprzez liczbę lub trudność aplikowanych bodźców uczących. Dodatkowo, krótki okres czasu aplikowania bodźców, który negatywnie wpływa na podnoszenie poziomu kompetencji, może zostać wzmocniony efektem braku przerw pomiędzy poszczególnymi dawkami materiału szkoleniowego. Dlatego wydaje się istotne, aby w sposób racjonalny i obiektywny planować ścieżkę uczenia się dla danego szkolenia i związanego z nim zestawu kompetencji.

Planowanie ścieżki szkolenia to proces doboru treści wraz z ich lokalizacją czasową, a także metod ich przekazywania, które pozwolą osiągnąć zakładany zestaw kompetencji. Jest to także strukturalny sposób organizowania procesu uczenia się, który pozwala wyodrębnić i zastosować cele, zasoby, czas oraz sposoby, które mają największy wpływ na zakładane efekty kształcenia. Metody planowania takiej ścieżki mogą opierać się o ocenę cech osobowych, zachowania, czy efektów pracy inżyniera [3], gdzie w dużym stopniu poziom kompetencji nie wpływa istotnie na wybranie i zaplanowanie szkolenia indywidualnego. W takim podejściu do planowania ścieżki rozwoju, pominięcie lub zminimalizowanie znaczenia określenia poziomu kompetencji może mieć negatywny skutek na osiągnięcie zakładanego efektu kształcenia. Oparcie planowania ścieżki szkolenia o realną ocenę poziomu kompetencji jest ważnym aspektem tworzenia całego procesu rozwoju pracownika w przedsiębiorstwie, jednak istotne jest także poznanie mechanizmów i czynników wpływających na dynamikę zmian poziomu kompetencji, aby uzyskać

[1] pwarzybok@polsl.pl, Szkoła Doktorów; Politechnika Śląska; www.polsl.pl.
[2] grzegorz.dziatkiewicz@polsl.pl, Katedra Mechaniki i Inżynierii Obliczeniowej, Wydział Mechaniczny Technologiczny, Politechnika Śląska, www.polsl.pl.

maksymalny możliwy efekt wzrostu kompetencji przy ograniczeniach czasu i środków, co oznacza sformułowanie i rozwiązanie problemu optymalnego sterowania personalizowanych ścieżek kształcenia.

Określenie poziomu startowego kompetencji wymaga zastosowania odpowiedniej metody testowania. Obecnie wykorzystywane metody testowania kompetencji inżynierskich skupiają się na tradycyjnych teoriach opartych o klasyczną teorię testowania [4] (CTT, ang. *Classical Test Theory*), której głównym założeniem jest następujące stwierdzenie: prawdziwy wynik testu składa się z wyniku uzyskanego przez badanego, jak i zakładanego błędu. Dodatkowo ocenie podlega tylko poziom kompetencji oparty o całkowity wynik testu, a nie są uwzględniane i badane wyniki cząstkowe związane z rezultatami dla składowych testu kompetencji. Takie podejście nie pozwala na personalizację procesu uczenia się, który zazwyczaj zapewnia uzyskanie wyższego poziomu kompetencji. Dlatego w dalszych badaniach planuje się zastosowanie metody opartej o teorię odpowiedzi na pozycje testowe (IRT, z ang. *Item Response Theory*) [5], która pozwoli na: obiektywną identyfikację poziomu kompetencji; swobodne i optymalne projektowanie różnych typów testów oraz wdrożenie adaptacyjnego systemu testowania kompetencji wraz doborem treści dla zapewnienia optymalnego szkolenia.

Modelowanie dynamiki to trudny problem z zakresu nauk obliczeniowych. Może stanowić punkt startowy dla różnych innych zadań o dużym znaczeniu praktycznym, takich jak: predykcja zmian poziomu kompetencji dla danego poziomu zdolności i typu szkolenia, analiza wrażliwości wykonywana celem identyfikacji kluczowych parametrów modelu, ważny z punktu widzenia celu modelowania, jakim może być np. optymalizacja szkoleń; identyfikacja parametrów modelu za pomocą danych eksperymentalnych, celem projektowania szkoleń personalizowanych. W niniejszej pracy pokazano jedynie wyniki badań wstępnych dla opracowanego modelu.

Przykładem dynamicznie rozwijanej technologii wspomagania prac inżynierskich jest rozwój systemów klasy CAx (ang. *Computer-Aided Technologies*) [6]. Oprogramowanie tego typu jest obecnie scalane w ramach jednolitych platform [7] do systemowego zarządzania cyklem życia produktu, tak aby korzystając z jednego źródła danych móc efektywnie realizować kolejne etapy procesu powstawania nowych produktów – od zebrania wymagań przez koncypowanie, projektowanie, obliczenia wytrzymałościowe, przygotowanie technologii wytwarzania, opracowania planów produkcyjnych do powstania gotowego wyrobu, a nawet jego konserwacji i utylizacji. Platforma 3DExperience [8] utworzona przez firmę Dassault Systemes, wywodząca się od jednego z pierwszych programów CAD [9] (ang. *Computer-Aided Design*) [6], CATII, jest narzędziem, które inżynierowi z szerokim spektrum kompetencji pozwala znacząco zredukować czas potrzebny na wykonywanie standardowych zadań w ramach tworzenia i optymalizowania produktów.

Niniejsza praca pokazuje rezultaty wstępnych badań dotyczących klasyfikacji kompetencji inżynierskich na przykładzie wybranego szkolenia podstawowego z zakresu CATIA V5 Mechanical Design Fundamentals w celu zaprojektowania testów do badania poziomu kompetencji. Przedstawiono także propozycję modelu dynamiki zmian poziomu kompetencji w czasie szkolenia, tak aby rozpoznać wpływ wprowadzonych parametrów modelu na proces uczenia, w celu jego późniejszej optymalizacji.

2. Klasyfikacja kompetencji inżynierskich według taksonomii Blooma

Jednym z istotnych aspektów odpowiedniego planowania ścieżki uczenia jest ocena początkowego poziomu kompetencji osoby przystępującej do szkolenia. Określenie tego poziomu przed rozpoczęciem szkolenia za pomocą odpowiednio zaprojektowanych testów jest powszechnie stosowaną praktyką, zarówno w tradycyjnych, jak i komercyjnych szkoleniach.

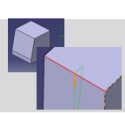
Właściwa konstrukcja bazy wiedzy dla problemu projektowania testów kompetencji wymaga wcześniejszego zdefiniowania kompetencji i ich odpowiedniego sklasyfikowania. Klasyfikacja kompetencji według taksonomii Blooma zakłada, że człowiek jest w stanie przyswajać wiedzę i umiejętności w obszarach czterech stref: kognitywnym, emocjonalnym, psychomotorycznym oraz rozwiązywania problemów [10]. Specyfika szkoleń inżynierskich zakłada, że są one realizowane w trybie nauczania indywidualnego, co oznacza możliwość pominięcia efektów kooperacji, współzawodnictwa, synergii i innych, występujących podczas tradycyjnego nauczania w grupie, w szczególności w kontakcie bezpośrednim. Ze względu na specyfikę szkoleń inżynierskich postanowiono skupić się na poziomach obszaru kognitywnego tj. wiedzy, zrozumieniu, zastosowaniu, analizie, syntezie oraz ewaluacji.

Jako bazowy obiekt analizy wybrano szkolenie podstawowe z modelowania przestrzennego 3D – CATIA V5 Mechanical Design Fundamentals. Jest to szkolenie, które dla inżynierów mechaników stanowi pierwszy etap planowanej ścieżki kształcenia, w związku z poruszanymi w trakcie jego trwania zagadnieniami z podstaw obsługi interfejsu programu, wykonywania szkiców 2D, modelowania bryłowego 3D, tworzenia złożeń, czy wykonywania dokumentacji technicznej. Na podstawie taksonomii Blooma oraz zakresu szkolenia wybrano kompetencje pokazane w tabeli 1, których poziom powinien wzrosnąć po zakończeniu procesu uczenia się podczas szkolenia.

Tabela 1. Lista badanych kompetencji na przykładzie wybranego szkolenia [opracowanie własne]

Obszar kompetencji	Nr kompetencji	Kompetencja
Infrastruktura systemu	CINF1	Znajomość infrastruktury systemu
Szkicownik	CSKT1	Potrafi tworzyć szkice
Modelowanie bryłowe	CPDG1	Potrafi modelować bryłowo
Tworzenie złożeń	CASM1	Potrafi tworzyć złożenia
Tworzenie rysunków	CGDR1	Potrafi tworzyć rysunki

Na podstawie zdefiniowanych kompetencji opracowano bazę pytań.

Kompetencja	Taksonomia	Pytanie	Pytanie - pomoc	Odp. A	Odp. B	Odp. C	Odp. D
CINF1	CAPP	Możemy wycentrować widok w miejscu kliknięcia poprzez kliknięcie na modelu klawiszem:		LPM	PPM	ŚPM	Alt+LPM
CINF1	CKNW	Schowanie/Pokazanie drzewka możemy uzyskać poprzez kliknięcie:		klawisza F3	klawisza F4	ŚPM + Hide/Show na gałęzi drzewka	LPM na dowolnej gałęzi drzewka
CPDG1	CSYN	Chamfer jak na rysunku, został uzyskany poprzez zdefiniowanie następujących wymiarów:		Length 1= 50, Angle = 80°	Length 1= 50, Angle = 10°	Length 2= 50, Angle = 10°	Length 2= 50, Angle = 80°

Rysunek 1. Przykładowe pozycje testowe dla szkolenia podstawowego z uwzględnieniem taksonomii Blooma, opracowanie własne na podstawie [10]

Na rysunku 1 pokazano przykładowe pytania utworzone w oparciu o zasady tworzenia pozycji testowych w metodzie IRT – kompetencja z obszaru infrastruktury programu (CINF1) posiada dwie pozycje testowe odpowiadające poziomowi wiedzy (CKNW, ang. *Cognitive Knowledge*) oraz zastosowaniu (CAPP, ang. *Cognitive Application*).

Wybór odpowiedniej metody testowania i przygotowanie miarodajnej bazy testów stanowi jeden z kluczowych elementów do prawidłowej oceny startowego poziomu kompetencji. Przedstawiony przykład jest wstępem do dalszych badań i posłuży do zbudowania dostrojonego systemu testowania opartego na metodzie IRT. Planowanie ścieżki szkolenia nie zależy jednak tylko od początkowego poziomu kompetencji, w związku z czym podjęto próbę stworzenia matematycznego modelu dynamiki zmian poziomu kompetencji w celu identyfikacji pozostałych parametrów wpływających na proces uczenia się.

3. Model dynamiki zmian poziomu kompetencji podczas szkolenia indywidualnego

Modelowanie dynamiki zmian, w kontekście tzw. efektu wzrostu, stanowi przedmiot zainteresowania biomatematyki [12] oraz ekonomii matematycznej [13]. W pracy zaproponowano prosty model dynamiki, oparty o model Verhulsta dynamiki wzrostu przy ograniczonych zasobach [12]. Zakłada się, że proces zmiany poziomu kompetencji odbywa się w warunkach ograniczeń związanych z cechami osobniczymi osoby szkolącej się oraz z ograniczonym programem szkolenia.

Niech $\theta(t) > 0$ oznacza poziom kompetencji w chwili t, wtedy względna zmiana poziomu kompetencji w czasie, w warunkach ograniczonych zasobów przetwarzania informacji, może zostać wyrażona tzw. uogólnionym modelem Verhulsta [12]:

$$\frac{\dot{\theta}(t)}{\theta(t)} = r_{eff}(\theta(t-\tau), u(t-\tau))\left(1 - \frac{\theta(t)}{\theta_{max}}\right); \theta(-\tau \leq t \leq 0) = \theta_0(t) \qquad (1)$$

gdzie: kropką oznaczono pochodną czasową, r_{eff} oznacza tzw. efektywną szybkość zmian poziomu kompetencji, θ_{max} to maksymalny poziom kompetencji możliwy do uzyskania przez osobę uczącą się, $\theta_0(t)$ oznacza historię zmian poziomu kompetencji przed rozpoczęciem szkolenia, dla przedziału odpowiadającego czasowi opóźnienia τ, u oznacza natężenie bodźca uczącego.

Funkcyjna postać warunku początkowego w (1) wynika z zastosowanego modelu w postaci równania różniczkowego z odchyleniem [12], odwzorowującego procesy naturalnych opóźnień w absorpcji nowych kompetencji podczas procesu szkolenia. Zazwyczaj w pierwszych badaniach zakłada się, że $\tau = 0$, wtedy $\theta(t = 0) = \theta_0$, co ma postać klasycznego warunku początkowego dla równania różniczkowego.

Niech efektywna szybkość zmian poziomu kompetencji r_{eff} jest zależna od dwóch procesów: uczenia się, zazwyczaj zapewniającego wzrost poziomu kompetencji, oraz zapominania, które prowadzi do spadku poziomu kompetencji, wtedy:

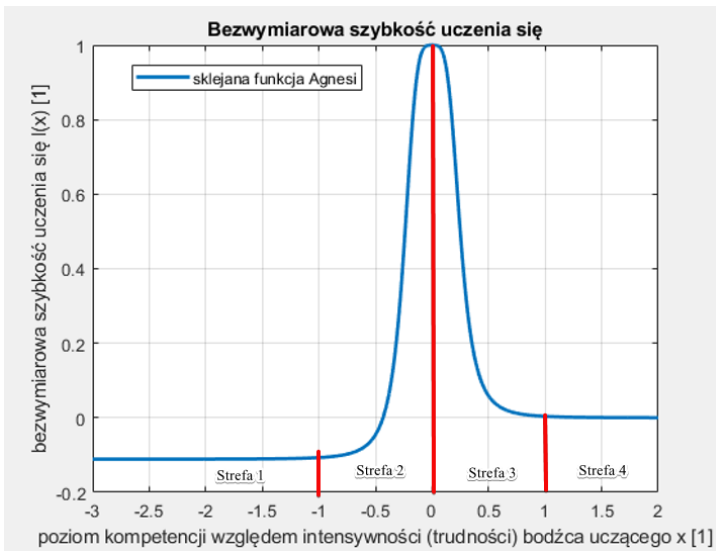
$$r_{eff} = l[\theta(t-\tau), u(t-\tau)] - f[\theta(t)] \qquad (2)$$

gdzie: przez $l[\theta(t), u(t)]$ oznaczono funkcję szybkości uczenia się [11, 14], zależną od aktualnego poziomu kompetencji $\theta(t)$ oraz od natężenia bodźca uczącego $u(t)$, który może modelować poziom intensywności lub trudności szkolenia; $f[\theta(t)]$ oznacza funkcję szybkości zapominania, zależną od aktualnego poziomu kompetencji [11].

Założono, że szybkość uczenia się, jest zależna od względnego poziomu natężenia bodźca uczącego się względem aktualnego poziomu kompetencji x = (θ – u) / θ i wykazuje cztery zakresy zmienności, odpowiadające różnej szybkości uczenia się, co przedstawia rysunek 2 z zastosowaną miarą procentową dla zmiennej niezależnej x. Wykorzystany w modelu przebieg funkcji szybkości uczenia się można przedstawić za pomocą sklejanej funkcji Agnesi, składającej się z dwóch przedziałów:

$$l(x) = \begin{cases} \frac{1}{1+c_1}\left(\frac{1}{1+(\gamma x)^{2p}} + c_1\right) & x < 0 \\ \frac{1}{1+(\gamma x)^{2p}} & x \geq 0 \end{cases} \quad (3)$$

gdzie: l(x) = l[(θ-u)/θ] i pominięto argument t, c_1 oznacza wartość przesunięcia = -0,1 w stosunku do bazowej funkcji Agnesi, γ oraz p to stałe współczynniki funkcji Agnesi równe odpowiednio: γ = 4, p = 2



Rysunek 2. Przebieg funkcji szybkości uczenia się l(x) [opracowanie własne]

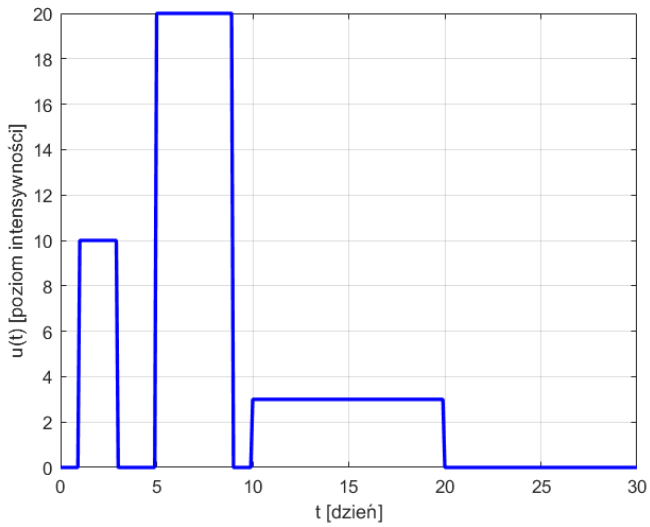
Jeśli poziom kompetencji względem intensywności bodźca jest mniejszy od -100% to proces uczenia się nie następuje, co można zaobserwować w strefie I. W strefach II i III uczenie się jest efektywne, ponieważ szybkość uczenia się l(x) > 0. Założono, że maksymalna szybkość uczenia się l_{max} odpowiada poziomowi intensywności bodźca uczącego niemal dopasowanemu do aktualnego poziomu kompetencji, tak że x_{max}~ = 0. Strefa IV odpowiada sytuacji braku bodźca uczącego (co zachodzi dla x >= 100%) lub bodźca o ujemnej intensywności, który nie powoduje inicjacji procesu uczenia się. W teorii sterowania przebieg funkcji l(x) pozwala klasyfikować model do kategorii układów z przełączeniami [13].

W niniejszej pracy założono, że przebieg bodźca uczącego u(t) można aproksymować kombinacją liniową dystrybucji bramkowych, co może stanowić model szkolenia, uwzględniający przerwy między okresami intensywnej nauki:

$$u(t) = \sum_{i=1}^{N} A_i B_i(t);\ B_i(t) = H[t - s_i] - H[t - (s_i + \delta_i)] \quad (4)$$

gdzie: A_i oznacza poziom intensywności lub trudności bodźca uczącego, a s_i i δ_i to odpowiednio chwila czasu odpowiadająca aplikacji bodźca o intensywności A_i i czas trwania bodźca.

Przykładowy przebieg funkcji u(t) przedstawiono na rysunku 3, dla danych A = [10 20 3], s = [1 5 10] oraz δ = [2 4 10].



Rysunek 3. Przykładowy przebieg bodźca uczącego u(t) [opracowanie własne]

Ponieważ u osób o wyższym aktualnym i potencjalnym poziomie kompetencji proces zapominania zachodzi wolniej, to w modelu założono, że wartości funkcji szybkości zapominania f[θ(t)] są odwrotnie proporcjonalne do aktualnego poziomu kompetencji oraz maksymalnego potencjalnego poziomu kompetencji:

$$f[\theta(t)] = c/(\theta(t) + \theta_{max}) \tag{5}$$

gdzie: c oznacza szybkość spadku kompetencji.

Ze względu na fakt, że opracowany model jest nowy i dotąd niespotykany w literaturze, to podczas badań wstępnych postanowiono sprawdzić jakościowo dynamikę poziomu zmian kompetencji, bez identyfikacji parametrów modelu. W tym celu przeprowadzono procedurę skalowania równań modelu do postaci bezwymiarowych, głównie celem redukcji liczby parametrów i celem przeprowadzenia badań symulacyjnych dla danych bezwymiarowych uzyskując równanie:

$$\frac{d\theta'(t')}{dt'} = \theta'(t')\left(l\left[\theta'(t'-\tau'), u'(t'-\tau'''')\right]\right) - \beta f[\theta'(t')]\left(1-\theta'(t')\right)$$
$$\theta'\left(-\tau' \le t' \le t_0'\right) = \theta'_0(t') \tag{6}$$

gdzie: β to współczynnik oporu dla przyrostu kompetencji (dla osób o dużych zdolnościach jego wartości są małe), θ'_0 to historia poziomu kompetencji (określa doświadczenie osoby uczącej się), a τ' to przeskalowany czas opóźnienia procesów uczenia się.

4. Testowanie modelu i omówienie wyników

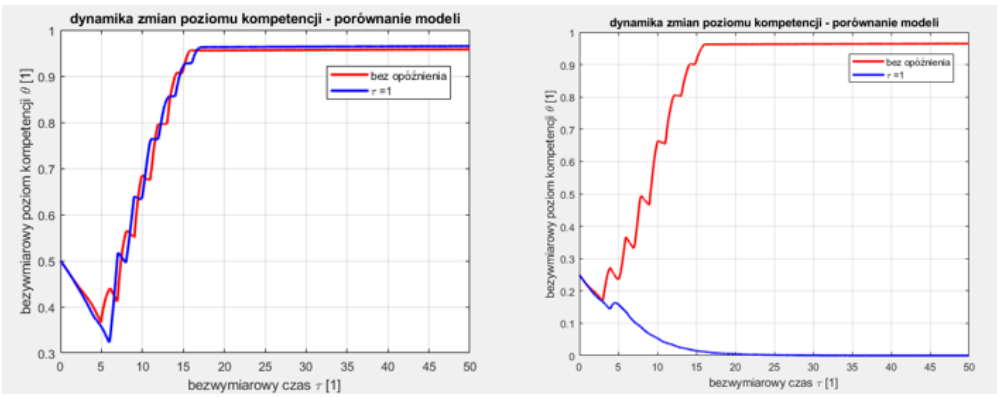
Na postawie opracowanego modelu przeprowadzono symulacje w oprogramowaniu MATLAB. Do całkowania równania dynamiki bez opóźnienia zastosowano hybrydowy jawny algorytm Rungego-Kutty IV i V rzędu wraz z metodą Dormanda-Prince'a,

z kolei dla równań z opóźnieniem hybrydowy algorytm Rungego-Kutty II i III rzędu wraz z procedurą interpolacji [15].

Jako dane dla procesu symulacji przyjęto następujące założenia:

- czas trwania szkolenia jest niezmienny;
- poszczególne elementy szkolenia, jak i przerwy pomiędzy nimi trwają stałą, niezmienną ilość czasu;
- czas opóźnienia jest stały i niezmienny, w trakcie trwania procesu uczenia się;
- poziom trudności bodźców uczących przyrasta liniowo z każdym kolejnym elementem szkolenia.

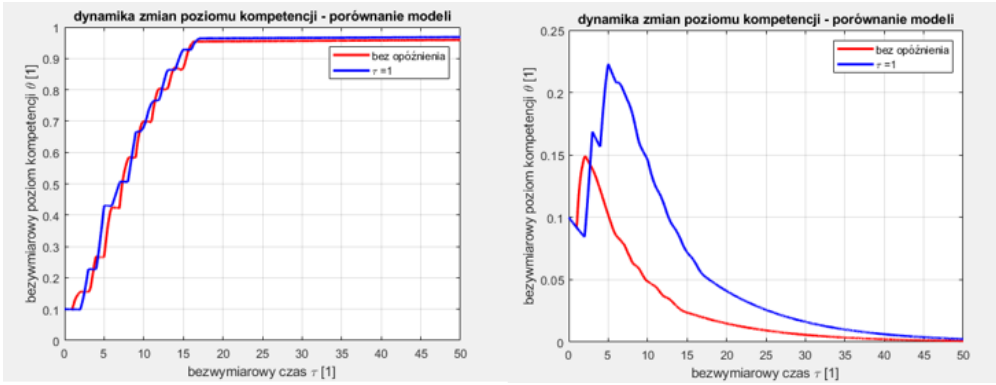
W pierwszej kolejności zbadano wpływ poziomu startowego kompetencji na dynamikę jej zmian dla przyjętego stałego współczynnika oporu (zdolności). Na rysunku 4 przedstawiono porównanie wykresów dynamiki zmian poziomu kompetencji dla dwóch różnych poziomów startowych kompetencji – 50% oraz 25% maksymalnego możliwego poziomu kompetencji dla danego szkolenia.



Rysunek 4. Porównanie wyników dynamiki zmian poziomu kompetencji dla startowego poziomu kompetencji równego 50% (po lewej stronie) oraz 25% (po prawej stronie) [opracowanie własne]

Przeprowadzone symulacje w oparciu o także inne wartości poziomu startowego kompetencji wykazały silny wpływ tego parametru na finalny poziom kompetencji po zakończeniu szkolenia. Dla poziomu startowego równego 50% można zaobserwować stabilizację poziomu kompetencji na poziomie niewiele niższym od zakładanego poziomu maksymalnego, zarówno w przypadku badań dla modelu bez opóźnienia, jak z opóźnieniem. Natomiast dla niższego poziomu startowego kompetencji (25%) stabilizacja następuje w przypadku modelu bez opóźnienia, natomiast dla modelu z opóźnieniem przyrost kompetencji finalnie nie występuje.

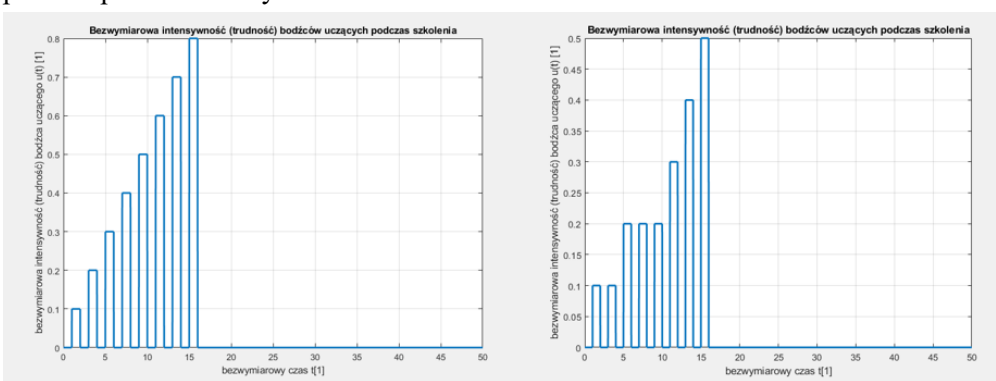
Kolejno zbadano wpływ współczynnika oporu przyrostu kompetencji dla osoby uczącej się, przy założeniu stałego, względnie niskiego poziomu startowego danej kompetencji (10%). Wyniki przedstawiono na rysunku 5.



Rysunek 5. Porównanie wyników dla współczynnika oporu równego 1% (po lewej stronie) oraz 10% (po prawej stronie) [opracowanie własne]

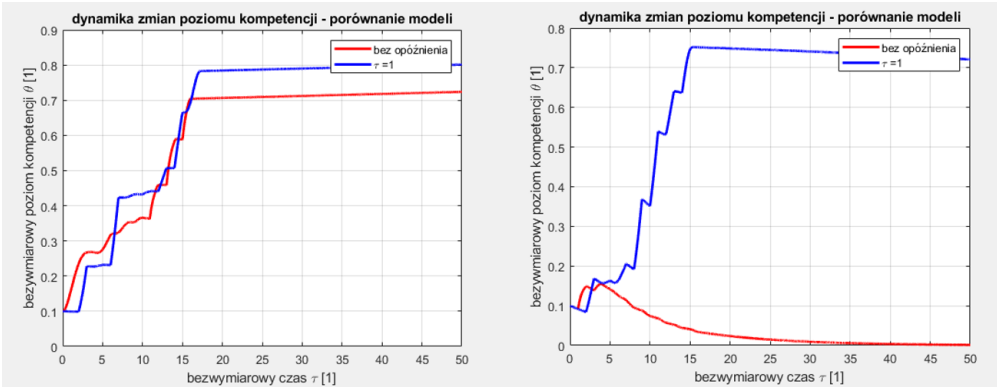
Dla badanego modelu szkolenia i jego parametrów można zauważyć, że do osiągnięcia zakładanego, zbliżonego do maksymalnego, poziomu kompetencji wymagany jest stosunkowo bardzo niski (1%) współczynnik oporu.

Obserwując, że przy zakładanym modelu bodźca uczącego proces przyrostu kompetencji nie następuje dla osób z wyższym współczynnikiem oporu postanowiono sprawdzić wpływ zmiany parametrów tegoż bodźca na osiągany poziom kompetencji. Zmniejszono trudność (intensywność) stosowanych bodźców oraz zastosowano efekt powtarzania na początkowych etapach szkolenia. Porównanie modeli bodźców uczących dla obu przypadków pokazano na rysunku 6.



Rysunek 6. Porównanie modeli bodźców uczących dla wykonywanych symulacji, po lewej stronie model początkowy, po prawej stronie model zmieniony po pierwszej iteracji obliczeń [opracowanie własne]

Przeprowadzone symulacje dla podanych danych wejściowych pozwoliły zaobserwować poprawę efektów kształcenia dla osób z większym współczynnikiem oporu, jednocześnie umożliwiając także niewielkie obniżenie efektywności procesu nauczania u osób, u których współczynnik oporu jest bardzo niewielki, co przedstawia rysunek 7.



Rysunek 7. Porównanie wyników poziomu kompetencji przy zmienionym modelu bodźca uczącego dla współczynnika oporu równego 1% oraz 10% [opracowanie własne]

W przypadku wyników przedstawionych na rysunku 7 zaobserwowano pozytywny efekt czasu opóźnienia, w porównaniu do modelu bez opóźnienia, gdzie proces uczenia się nie nastąpił.

5. Podsumowanie

Planowanie ścieżki szkolenia jest zagadnieniem wymagającym rozpoznania wielu czynników, które wpływają na osiągany efekt kształcenia. Opracowany model dynamiki zmian poziomu kompetencji i wyniki przeprowadzonych symulacji pozwalają lepiej zrozumieć, jak wiele parametrów może wpływać na proces uczenia się nowych kompetencji lub doskonalenia już nabytych. Jednym z istotnych elementów wpływających na proces uczenia jest poziom startowy kompetencji danej jednostki, co z kolei wpływa na istotność procesu testowania i weryfikowania owego poziomu.

Przeprowadzone symulacje opracowanego modelu dynamiki zmian wskazują, że opóźnienie towarzyszące procesom uczenia się ma wpływ na jakościowy przebieg krzywej poziomu kompetencji. Dla przyjętych stałych wartości czasu opóźnienia (1) oraz początkowego poziomu kompetencji (10%), ale przy różnych wartościach współczynnika oporu widać wyraźny wpływ czasu opóźnienia na osiągniecie zakładanego, stabilnego poziomu kompetencji. Również startowy poziom kompetencji w znaczącym stopniu wpływa na dynamikę procesu zmian poziomu kompetencji: przy wartości początkowej równej 50% można zaobserwować mniejszy wpływ czasu opóźnienia niż w przypadku wartości początkowej równej 25%. Opracowane wyniki pozwalają zaobserwować także istotny wpływ modelu bodźca uczącego na osiągany poziom kompetencji: dla obiektu o współczynniku oporu równemu 10% efekt nauczania nie występuje przy przedstawionych w pierwszej iteracji obliczeniach, a dla zmodyfikowanego pod względem trudności (intensywności) modelu bodźca uczącego już tak.

Zastosowanie w dalszych pracach metody opartej na teorii IRT pozwoli bardziej precyzyjnie określić poziom kompetencji z danego obszaru. Istotnym elementem dla konstrukcji testów i zawartości szkoleń jest umiejętna analiza systemowa szkolenia pod względem taksonomii Blooma, tak aby stworzyć komponenty testowe z zachowaniem wymaganego dla teorii IRT parametru niezależności pozycji testowej. Zagadnienie doboru intensywności elementów szkolenia oraz czasu ich trwania może podlegać procesowi optymalizacji w celu zwiększenia efektywności kształcenia.

Literatura

1. Grabania M.Ł., *Inżynier przemysłu 4.0*, Szybkobieżne Pojazdy Gąsienicowe, Redakcja Biuletynu SPG, Vol. 2(52), 2019, s. 47-56.
2. Serafin K., *Nowe formy budowania kompetencji zawodowych pracowników na tle rynku usług szkoleniowych*, Studia Ekonomiczne, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Vol. 322, 2017, s. 214-229.
3. Penkala D., *Ocena pracownicza oparta na kompetencjach. Wykorzystywanie wyników oceny do planowania szkoleń i rozwoju pracowników*, Zeszyty Naukowe Wyższej Szkoły Bankowej, 34, 2011.
4. DeVellis R.F, *Classical Test Theory*, Measurement in a Multi-Ethnic Society, Medical Care, Vol. 44 (11), 2006, s. 50-59.
5. Hambleton R.K., Swaminathan H., Jane Rogers H., *Fundamentals of Item Response Theory*, Sage Publications, Newbury Park 1991.
6. Lukaszewicz A, Skorulski G, Szczebiot R., *Main aspects of training in field of computer-aided techniques (CAx) in mechanical engineering*, Engineering for rural development, 2018.
7. Camba J.D., Contero M., Company P., Perez D., *On the integration of model-based feature information in Product Lifecycle Management systems*, ScienceDirect, Vol. 37 (6), 2017, s. 611-621.
8. Liu L., *Application of Dassault System 3D Experience Platform in Enterprise Digitalization*, Frontier Computing. FC 2021. Lecture Notes in Electrical Engineering, Springer, vol. 827, 2022, s. 1407-1413.
9. Michaud M., *CATIA. Narzędzia i moduły*, Helion, 2014, s. 13-24.
10. Wankat P.C., Oreovicz F.S., *Teaching engineering*, Purdue University Press, West Lafayette, 2015.
11. Bush R.R., Mosteller F., *A mathematical model for simple learning*, Springer, Vol. 58 (5), 2006, s. 231-234.
12. Foryś U., *Matematyka w biologii*, WNT, Warszawa 2005.
13. Sethi S.P., *Optimal control theory. Applications to management science and economics*, Springer, Cham 2021.
14. Wu X., Zhang Y., Wu. X., Sun W., *A nonlinear dynamic model for individual learning with plateau phenomenon*, w materiałach ICE4E „Association for Computing Machinery, DOI: https://doi.org/10.1145/3377571.3377575, 2020.
15. Szmelter J., *Metody komputerowe w mechanice*, PWN, Warszawa 1980, s. 150-157.

Planowanie ścieżki uczenia się i modelowanie dynamiki zmian kompetencji inżynierskich w zakresie obsługi systemów klasy CAx

Streszczenie

Praca przedstawia zagadnienie planowania ścieżki uczenia się podczas indywidualnych szkoleń inżynierów z zakresu kompetencji twardych, w szczególności obsługi systemów klasy CAx, wspomagających procesy inżynierskie, w szczególności proces rozwoju produktów. Zawiera próbę klasyfikacji kompetencji inżynierskich w ramach taksonomii Blooma dla przykładu wybranego szkolenia podstawowego (CATIA V5 Mechanical Design Fundamentals), jako pierwszy z etapów planowania ścieżki uczenia się, w szczególności uwzględniający psychometrycznye narzędzia oceny poziomu kompetencji w postaci testów oraz dobór odpowiednich metod i treści podczas procesu szkolenia. Opracowane, zgodnie z teorią IRT (ang. *Item Response Theory*) testy pozwolą na akwizycję danych potrzebnych do opracowania modeli oraz komponentów samouczącego się systemu optymalizacji ścieżki uczenia się wybranych kompetencji inżynierskich. W pracy przedstawiono także autorski matematyczny model dynamiki zmian poziomu kompetencji podczas szkolenia indywidualnego w postaci równania różniczkowego z opóźnieniem. Model pozwala symulować przebieg zmian poziomu kompetencji w zależności od przyjętych parametrów bodźców uczących oraz z uwzględnieniem dynamiki dwóch, zachodzących równolegle, procesów: uczenia się oraz zapominania, zależnych od poziomu kompetencji oraz poziomu natężenia bodźca uczącego względem aktualnego poziomu kompetencji. Wyniki przeprowa-

dzonych symulacji dla przykładowych parametrów modelu bezwymiarowego, uwzględniających różne poziomy początkowe kompetencji osób uczestniczących w szkoleniu oraz poziom ich indywidualnych zdolności, pozwolą lepiej zrozumieć i zobrazować istotę problemów konstrukcji optymalnych ścieżek uczenia się oraz konstrukcji testów poziomu kompetencji.
Słowa kluczowe: IRT, klasyfikacja kompetencji, model dynamiki zmian poziomu kompetencji

Planning a learning path and modeling the dynamics of changes in engineering competencies in the use of CAx class systems

Abstract
Article presents the issue of planning a learning path during individual training of engineers in the field of hard skills, in particular the use of CAx class systems, supporting engineering processes, in particular the product development process. It contains an attempt to classify engineering competences within the Bloom taxonomy for the example of a selected basic training (CATIA V5 Mechanical Design Fundamentals), as the first stage of planning a learning path, taking into account psychometric tools for assessing the level of competence in the form of tests and the selection of appropriate methods and content during the training process. The tests developed in accordance with the IRT theory (item response theory) will allow for the acquisition of data needed to develop models and components of a self-learning system for optimizing the learning path of selected engineering competencies. The article also presents an original mathematical model of the dynamics of changes in the level of competence during individual training in the form of a differential equation with a delay. The model allows to simulate the flow of changes in the level of competence depending on the adopted parameters of learning stimuli and taking into account the dynamics of two parallel processes: learning and forgetting, depending on the level of competence and the level of intensity of the teaching stimulus relative to the current level of competence. The results of the simulations carried out for sample parameters of the dimensionless model, taking into account various initial levels of competence of people participating in the training and the level of their individual abilities, will allow for a better understanding and visualization of the essence of the problems of constructing optimal learning paths and constructing competence level tests.
Keywords: IRT, competences classification, model of dynamic changes of competence level

Paweł Buchwald[1], Aleksander Dawid[2]

Metoda walidacji aplikacji internetowych przeznaczonych dla osób ze szczególnymi potrzebami

1. Wprowadzenie

Problem badawczy tego opracowania dotyczy oceny i walidacji aplikacji przeznaczonej do poprawy dostępności zintegrowanych węzłów przesiadkowych (ZWP) dla osób z niepełnosprawnościami (OzN), które posiadają szczególne potrzeby w zakresie korzystania z infrastruktury obecnej w miejscach użyteczności publicznej. W opracowaniu uwzględniono nie tylko metody oceny aplikacji pod względem jej dostępności, ale także zaproponowano kroki pozwalające na walidację rozwiązania ze względu na szczególne potrzeby osób z niepełnosprawnościami – głównie w zakresie wad powodujących dysfunkcje odbioru treści graficznych. Przeprowadzona ocena powstałego narzędzia pozwala na jego ewentualną poprawę w zakresie dostosowania treści ze względu na czytelność i dostępność dla osób ze szczególnymi potrzebami. Narzędzie aplikacyjne podlegające walidacji jest wynikiem prac Akademii Wyższej Szkoły Biznesu w Dąbrowie Górniczej, Politechniki Gdańskiej i Instytutu Technicznego Wojsk Lotniczych prowadzonych w ramach projektu „Przesiadka bez barier". W ramach prac badawczych przeprowadzona została ewidencja infrastruktury kluczowych węzłów przesiadkowych, oraz opracowano metody informatyczne z zakresu baz danych i aplikacji komputerowych, które pozwalają na przetwarzanie zebranych informacji w kontekście wyznaczania tras i ostrzegania przed niedogodnościami, jak również ułatwieniami w korzystaniu z infrastruktury. Koncepcja wprowadzania tego typu ułatwień jest zgodna z podejściem projektowania uniwersalnego [1] i może stanowić pomoc nie tylko dla osób z niepełnosprawnościami [2, 3], ale również osób starszych, osób podróżujących z bagażami lub z dziećmi w wózkach. Zebrana baza barier, ograniczeń i ułatwień, która jest podstawą działania aplikacji musi być aktualizowana, aby zachować użyteczność systemu dla pasażerów. Zakłada się, iż sami użytkownicy będą informowali o zdarzeniach i niedogodnościach przesyłając za pośrednictwem aplikacji treści opisu niedogodności. W dzisiejszych czasach kontent informacyjny wprowadzany przez użytkowników ma charakter graficzny w postaci zdjęć i wszelkiego rodzaju treści obrazkowych [4, 5]. Ze względu na szczególne potrzeby użytkowników z niepełnosprawnościami, taka informacja musi być oceniana ze względu na czytelność i ewentualnie poddana modyfikacjom w celu poprawy dostępności. Dzięki temu zagwarantować można aktualność danych systemu i jednoczesne zachowanie jego użyteczność dla osób z niepełnosprawnościami. Ważnym aspektem użyteczności systemu jest poprawne jego działanie. Jest ono zdeterminowane brakiem błędów i dysfunkcji samej aplikacji, ale również efektywnością przetwarzania danych zależną od wydajności komponentów serwerowych [6, 7]. Dostępność narzędzi aplikacyjnych jest zdetermino-

[1] pbuchwald@wsb.edu.pl, Katedra Transportu i Informatyki, Wydział Nauk Stosownych, Akademia WSB, 41-300 Dąbrowa Górnicza, ulica Cieplaka 1c.
[2] adawid@wsb.edu.pl, Katedra Transportu i Informatyki, Wydział Nauk Stosownych, Akademia WSB, 41-300 Dąbrowa Górnicza, ulica Cieplaka 1c.

wana również kompatybilnością z urządzeniami klienckimi [8, 9]. W ramach testów aplikacji należy również sprawdzić możliwość jej użytkowania z popularnymi przeglądarkami internetowymi i urządzeniami, oraz różnymi rozdzielczościami urządzeń ekranowych [10-12]. Część z elementów walidacji aplikacji wynika z założeń inżynierii oprogramowania, założeń programowania uniwersalnego i standardów WCAG [13-15]. Dodatkowymi obostrzeniami są wymogi czytelności treści w zakresie jej dostępności przez osoby niepełnosprawne, co jest dodatkowym wyzwaniem podczas projektowania systemu. Ocenę i poprawę dostępności można więc podzielić na następujące obszary walidacji:

- Walidacja w zakresie efektywności czasowej systemu.
- Walidacja w zakresie dostosowania do urządzeń i wymogów standaryzacji WCAG[3].
- Walidacja w zakresie błędów i wad wynikających z naruszenia dobrych praktyk inżynierii oprogramowania i zasad projektowania uniwersalnego.
- Walidacja dostępności treści oferowanych w ramach aplikacji pod kątem wymogów osób z niepełnosprawnościami – zwłaszcza wzrokowymi.

Przedstawioną metodę walidacji oprogramowania pod kątem deficytów widzenia barwnego sprawdzono na przykładzie serwisu webowego powstałego w ramach projektu „Przesiadka bez barier". Praca podzielona jest na cztery zasadnicze rozdziały. Pierwszy to architektura i zakres funkcjonalny systemu wspomagającego korzystanie ze zintegrowanych węzłów przesiadkowych. Drugi to metody oceny dostępności i efektywności walidowanego systemu. Trzeci opowiada o metodach oceny dostępności informacji graficznej dla osób z deficytem widzenia barwnego. W ostatnim rozdziale przeprowadzone jest analiza systemu informatycznego dla ZWP w zakresie spełnienia szczególnych potrzeb osób niepełnosprawnych.

2. Architektura i zakres funkcjonalny systemu wspomagającego korzystanie ze zintegrowanych węzłów przesiadkowych (ZWP)

Architektura systemu wspomagającego korzystanie z ZWP opiera się na centralnej bazie danych w której zdefiniowane są specjalne potrzeby podróżnych i informacje o ZWP oraz aplikacji internetowej. Baza danych działa na serwerze PostgreSQL[4] obsługującym relacyjną bazę danych w języku SQL [16]. Cyfrowy model ZWP realizowany jest zgodnie z ideą BIM (ang. Building Information Modeling) [17, 18]. Zapis taki pozwala na jednorodny dostęp do danych modelowanego budynku. Dzięki temu zapisowi wirtualny model ma wolumetryczne parametry rzeczywistego obiektu. W skład tego modelu wchodzą tablice z elementami architektonicznymi i parametrycznymi opisującymi bariery i udogodnienia dla osób ze szczególnymi potrzebami.

Główny cel badań w tym opracowaniu to analiza aplikacji internetowej pod kątem dostępności dla osób z deficytem widzenia barwnego. Interfejs graficzny aplikacji realizowanej w ramach projektu „Przesiadka bez barier" korzysta z biblioteki Three.js stworzonej do prezentacji modeli 3D na stronach www [19]. W przypadku tej aplikacji jest to model portu lotniczego Warszawa-Modlin. Funkcjonalnie aplikacja ma za zadanie

[3] WCAG (Web Content Accessibility Guidelines) to standard, który określa jak tworzyć trony internetowe i aplikacje mobilne, aby wygodnie mogły z nich korzystać osoby z niepełnosprawnościami (https://www.gov.pl/web/dostepnosc-cyfrowa/wcag-21-w-skrocie).

[4] PostgreSQL – Narzędzie serwera bazy danych oparte o model relacyjny zachowujące zgodność ze standardem języka SQL (Select Query Language).

ułatwić zmianę środka transportu dla osób ze szczególnymi potrzebami (OzSP). Podróżny jadąc z Warszawy autobusem w kierunku portu lotniczego może w tej aplikacji zapoznać się z geografią samego portu oraz sprawdzić drogę od przystanku autobusowego do dowolnego miejsca w porcie lotniczym z uwzględnieniem jego ograniczeń. Aplikacja pozwala użytkownikowi na określenie się jako osoby ze szczególnymi potrzebami i wybraniu grupy niepełnosprawności z listy przedstawionej na rysunku 1.

Rysunek 1. Grupy niepełnosprawności obsługiwane w aplikacji. Źródło: opracowanie własne

Dla osób mających problemy z poruszaniem się na schemacie portu oznaczone są potencjalne bariery jak np. schody oraz udogodnienia jak np. windy. Przy niepełnosprawnościach sprzężonych [20], gdy jedną z nich jest zaburzenie widzenia, najistotniejszą jest sprawa czytelności informacji prezentowanej przez aplikację. Aplikacja obsługuje standard WCAG (*Web Content Accessibility Guidelines*) w kontekście zmiany kontrastu barw (rys. 2).

Rysunek 2. Zmiana kontrastu związana ze ślepotą barw. Źródło: opracowanie własne

Wprowadzony do aplikacji model 3D pozwala na zorientowanie się, gdzie są odpowiednio szerokie przejścia, aby można przejechać wózkiem inwalidzkim. Wymiary przejść podane są w centymetrach. Przykładowy widok z aplikacji pokazany jest na rysunku 3. Widać tam interfejs graficzny oraz model 3D portu lotniczego. Aplikacja dopasowuje się do rozmiarów ekranu na którym jest wyświetlana, ponieważ jest wykonana zgodnie ze specyfikacją projektowania RWD (ang. *Responsive Web Design*).

Rysunek 3. Zrzut ekranu aplikacji dla urządzenia mobilnego. Źródło: opracowanie własne

Funkcjonalnie finalna wersja aplikacji powinna pozwalać na;

- interfejs głosowy;
- lokalizację i wytyczanie drogi na ZWP;
- kupowanie biletów;
- szukanie dogodnego połączenia komunikacyjnego.

Zakres urządzeń na których można będzie korzystać ze wszystkich funkcji aplikacji uzależniony jest od możliwości przetwarzania grafiki 3D na tym urządzeniu. Głównie chodzi tutaj o wsparcie w przeglądarkach internetowych dla technologii WebGL[5] [21].

3. Metody oceny dostępności i efektywności walidowanego systemu

Przedstawione oceny dostępności są adekwatne dla potrzeb walidacji treści stron statycznych, to znaczy takich, które nie ulegają modyfikacjom podczas użytkowania systemu. Dla systemów opartych na podejściu związanym z interakcją z użytkownikiem walidacja treści musi odbywać się dynamicznie. Przykładem takich systemów są portale, w których treści informacyjne są współtworzone przez użytkowników. Walidowany system zakłada również możliwość przesyłania od użytkowników do systemu informacji o utrudnieniach w węzłach przesiadkowych. Informacja ta może być reprezentowana w sposób graficzny. Dbałość o dostępność treści informacyjnych na portalu nakazuje dynamiczną ocenę treści umieszczanych przez użytkowników. Z tego powodu konieczne jest opracowanie automatycznych metod oceny treści graficznych. Metodę taką w kontekście oceny treści pod kątem ich czytelności dla osób z daltonizmem przedstawia następny rozdział opracowania.

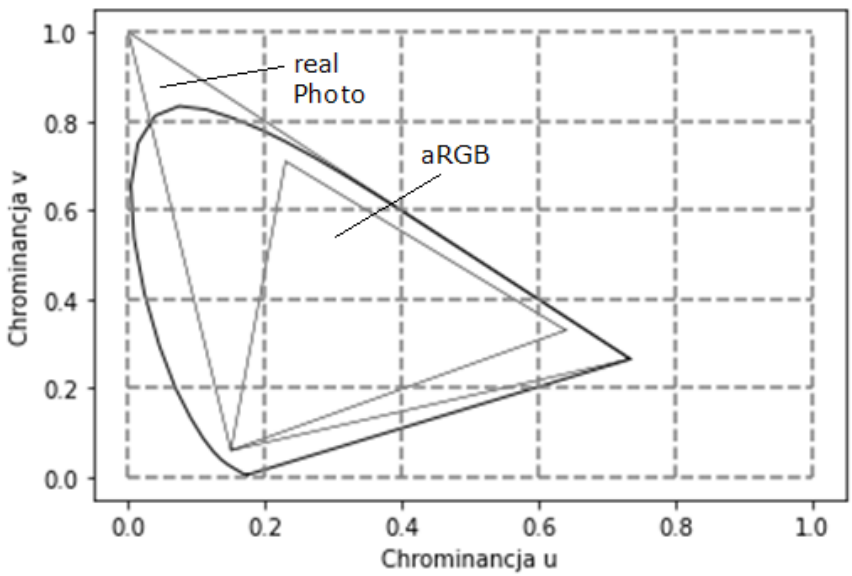
4. Metoda oceny dostępności informacji graficznej dla osób ze ślepotą barw

Dla oceny możliwości interpretacji barw przez osoby będące użytkownikami systemu teleinformatycznego wykorzystano opis barw informacji graficznej bazując na diagramie chromatyczności [22, 23]. Diagram ten przedstawia w dwuwymiarowym układzie kartezjańskim barwy, jakie mogą być interpretowane przez użytkownika, oraz wizualizowane przez urządzenia techniczne. Każda barwa jest reprezentowana przez punkt na diagramie, gdzie wartości współrzędnych tego punktu odpowiadają wartością chrominancji u, oraz chrominancji v. Punkty obrazujące poszczególne barwy przedstawia się często w odniesieniu do kształtu podkówki barw CIE[6], która reprezentuje współrzędne odpowiadające barwie światła widzialnego z zakresu od 380 nm do 700 nm. Dzięki wykresowi CIE można ocenić możliwości odwzorowania barwy przy pomocy danego urządzenia, systemu informatycznego a także możliwości percepcji użytkownika. Dla typowych używanych w grafice komputerowej przestrzeni barw budowany jest trójkąt zawierający punkty odpowiadające barwom możliwym do odzwierciedlenia. Punkty współrzędnych tego trójkąta określają graniczne barwy dla składowych kolorystycznych czerwonej (R),

[5] WebGL (*Web Graphics Library*) – jest biblioteką programistyczną rozszerzającą możliwości języka JavaScript, zapewnia dostęp do interfejsu programistycznego generowania grafiki trójwymiarowej.

[6] Skrót „CIE" oznacza „Commission Internationale de l'Éclairage", co w języku polskim można przetłumaczyć jako "Międzynarodowa Komisja Oświetlenia". Jest to międzynarodowa organizacja naukowa, która zajmuje się badaniami, standardami i rekomendacjami dotyczącymi oświetlenia i kolorów. CIE jest odpowiedzialna za opracowywanie norm i wytycznych związanych z pomiarami kolorów, oceną jakości światła, modelowaniem przestrzeni kolorów.

zielonej (G) i niebieskiej (B). Na rysunku 4 przedstawiono trójkąty dla modelu aRGB[7] [24, 25], oraz modelu real Photo [26] używanego przez rzeczywiste urządzenia cyfrowe do detekcji obrazu. W modelach istnieją osobne trójkąty widzenia barw dla osób z daltonizmem, ponieważ daltonizm to zaburzenie percepcji kolorów, które wpływa na zdolność oka do odróżniania niektórych kolorów.



Rysunek 4. Przykładowe modele używane w grafice komputerowej na wykresie chromatyczności. Źródło: opracowanie własne

Istnieją różne typy daltonizmu, ale najczęściej występującym jest **protanopia**, w której brakujące lub nieprawidłowo działające są receptory odpowiedzialne za odbieranie czerwonego koloru. Dlatego trójkąt widzenia dla osób z protanopią jest zazwyczaj zmniejszony i przesunięty w kierunku zielonego koloru, ponieważ w takim przypadku kolor czerwony jest postrzegany jako odcień zieleni. W przypadku **deuteranopii** (brakujące lub nieprawidłowo działające receptory odpowiedzialne za odbieranie zielonego koloru) trójkąt widzenia byłby przesunięty w kierunku czerwonego koloru, a dla **tritanopii** (brakujące lub nieprawidłowo działające receptory odpowiedzialne za odbieranie niebieskiego koloru) trójkąt widzenia byłby zmniejszony i przesunięty w kierunku kolorów ciepłych. Współrzędne wierzchołków przykładowych trójkątów widzenia dla osób z poszczególnymi rodzajami daltonizmu przedstawia tabela 1.

Należy zauważyć, iż ograniczenia widzenia poszczególnych barw mają charakter osobniczy i dobranie odpowiednich trójkątów widzenia wymaga przeprowadzenia wstępnych testów rozpoznawalności barw dla danego użytkownika. W prezentowanym przykładzie dobrano jednak zestaw rozróżnialnych kolorów w sposób bezpieczny, ograni-

[7] aRGB jest to model kolorów zaprojektowany przez firmę Adobe Systems. Adobe RGB ma większy zakres kolorów niż standardowe przestrzenie RGB. Umożliwia dodanie dodatkowego kanału (np. przeźroczystości) lub przekształcenia pozwalającego na odzwierciedlenie większego spektrum kolorów.

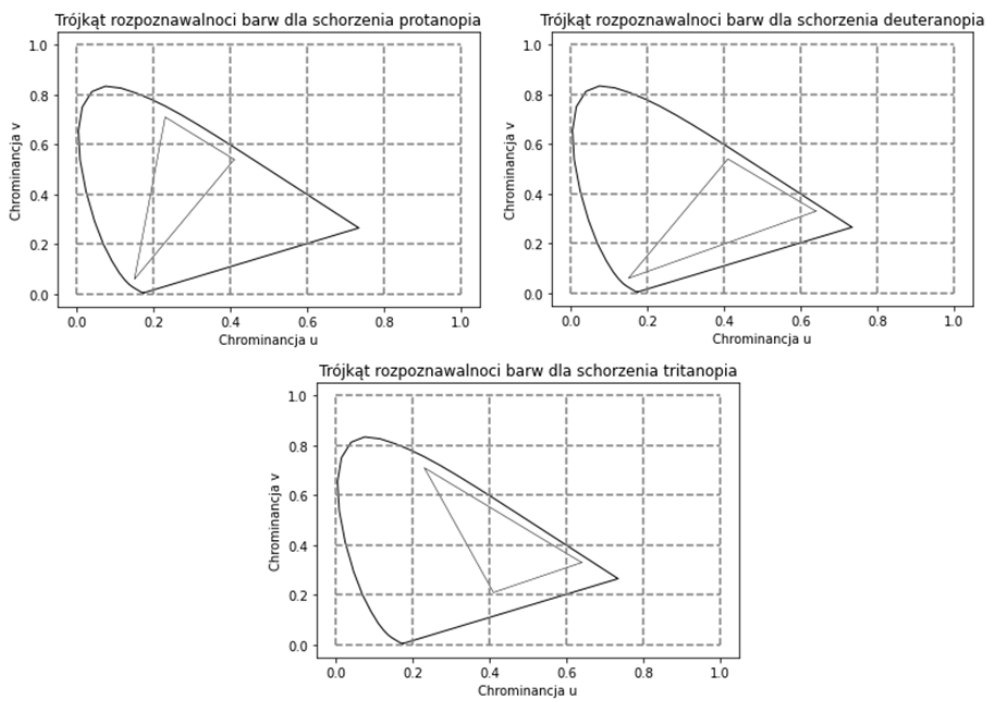
czając z nadmiarem przestrzeń trójkątów rozpoznawalnych barw. Na podstawie przynależności punktu reprezentującej barwę do danego trójkąta można stwierdzić czy jest ona możliwa do interpretacji przez osobę z rozpatrywanym rodzajem daltonizmu. Położenie na wykresie chromatyczności poszczególnych trójkątów przedstawia rysunek 5.

Tabela 1. Współrzędne wierzchołków P_1 P_2 i P_3 trójkątów widzenia dla poszczególnych dysfunkcji widzenia barwnego

Dysfunkcja wzrokowa	P_1		P_2		P_3	
	u	v	u	v	u	v
Protanopia	0.4100	0.5400	0.2300	0.7100	0.1500	0.0600
Deuteranopia	0.6400	0.3300	0.4100	0.5400	0.1500	0.0600
Tritanopia	0.6400	0.3300	0.2300	0.7100	0.4100	0.2100

Sprawdzenie czy dana barwa jest interpretowalna przez osobę cierpiącą na ślepotę barw może być dokonane na podstawie weryfikacji, czy punkt odpowiadający barwie znajduje się wewnątrz trójkąta widzenia. Aby sprawdzić, czy punkt P = (x,y) należy do trójkąta ABC o wierzchołkach A = (A_x,A_y), B = (B_x,B_y), C = (C_x,C_y), możemy skorzystać z reguły, która mówi, że punkt P należy do trójkąta ABC wtedy i tylko wtedy, gdy spełnione są jednocześnie trzy warunki:

1. Punkt P leży po tej samej stronie prostej zawierającej odcinek AB co punkt C.
2. Punkt P leży po tej samej stronie prostej zawierającej odcinek BC co punkt A.
3. Punkt P leży po tej samej stronie prostej zawierającej odcinek AC co punkt B.



Rysunek 5. Trójkąty rozpoznawalności barw na diagramie chromatyczności dla poszczególnych rodzajów daltonizmu

Warunki te wynikają wprost z własności trójkątów. Na ich podstawie autorzy utworzyli procedurę testową, która pozwoliła na programistyczne sprawdzenie możliwości interpretacji barwy

Test punktu leżącego po jednej stronie prostej polega na obliczeniu iloczynów wektorowych dwóch wektorów: $\overrightarrow{AB} \times \overrightarrow{AP}$= ($B_x$ - A_x, B_y - A_y) i (x - A_x, y - A_y) oraz dwóch wektorów: $\overrightarrow{BC} \times \overrightarrow{BP}$= ($C_x$ - B_x, C_y - B_y) i (x - B_x, y - B_y) oraz dwóch wektorów: $\overrightarrow{CA} \times \overrightarrow{CP}$= ($A_x$ - C_x, A_y - C_y) i (x - C_x, y - C_y). Jeśli iloczyny te są nieujemne (równe zero lub dodatnie), to punkt P leży po tej samej stronie prostej co dany wierzchołek. Metodę wyznaczania przynależności punktu do trójkąta można opisać poprzez kod przedstawiony poniżej.

```
if ((y - y_A) * (x_B - x_A) - (x - x_A) * (y_B - y_A) >= 0) &&
 ((y - y_B) * (x_C - x_B) - (x - x_B) * (y_C - y_B) >= 0) &&
 ((y - y_C) * (x_A - x_C) - (x - x_C) * (y_A - y_C) >= 0)
then
 punkt P należy do trójkąta ABC
else
 punkt P nie należy do trójkąta AB
```

Użycie przedstawionej metody oceny rozpoznawalności barw wymaga wyznaczenia punktu na diagramie chromatyczności CIE, który odpowiada danej barwie wykorzystanej w pliku graficznym. W wielu przypadkach dla graficznej reprezentacji obrazu rastrowego wykorzystuje się przestrzeń barw RGB. Dla modelu barw RGB używanego przez oprogramowanie Adobe przekształcenie do modelu barw CIExyz wymaga pomnożenia wektora współrzędnych punktu reprezentującego barwę w modelu wejściowym przez macierz przekształcenia.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = [M] \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

gdzie: M jest macierzą Adobe RGB (1998) o następujących współczynnikach:

0.5767309 0.1855540 0.1881852
0.2973769 0.6273491 0.0752741
0.0270343 0.0706872 0.9911085

Obliczenie współrzędnych punktu barwy P(u,v) na dwuwymiarowym wykresie chromatyczności jest możliwe z zastosowaniem poniższego wzoru.

$$u = x / (x + y + z)$$

$$v = y / (x + y + z)$$

Uzyskanie współrzędnych u,v pozwala na ocenę możliwości interpretacji barwy. Stosując iteracyjnie przedstawioną metodę sprawdzenia dla wszystkich barw wykorzystywanych w pliku graficznym można ocenić możliwość interpretacji tego pliku przez osobę cierpiącą na dany rodzaj daltonizmu. Ocena czytelności pliku graficznego w oparciu o przedstawiona metodę może być dokonana automatycznie, również dla plików graficznych dodawanych dynamicznie przez użytkowników. Za jej pomocą można również zrealizować podpowiedzi dla administratorów serwisu, którzy mogą dodać alternatywne treści graficzne (reprezentowane przez wektory transponowane $\boldsymbol{RGB^T}$ oraz $\boldsymbol{xyz^T}$ roz-

poznawalne przez osoby z danym schorzeniem ślepoty barw. Przykładowo gdy tło jest białe wtedy $\boldsymbol{RGB}^{\boldsymbol{T}}$ = [1.0, 1.0, 1.0]^{T}, obliczone $\boldsymbol{xyz}^{\boldsymbol{T}}$ = [0.9504701, 1.0000001, 1.088830]^{T}. Dla tego wektora u, v wynosi odpowiednio 0.3127266, 0.3290231. Na wykresie chromatyczności (rys. 6) jest to oznaczone literą **b**. Dla tritanopii jest to barwa nierozpoznawalna ponieważ znajduje się poza trójkątem widzenia. Dla czarnego $\boldsymbol{RGB}^{\boldsymbol{T}}$ = [0.0, 0.0, 0.0]^{T}, $\boldsymbol{xyz}^{\boldsymbol{T}}$ = [0.0,0.0,0.0]^{T} *u* i *v* nieokreślone matematycznie 0.0/0.0. Barwa ta jednak jest rozpoznawalna dla wszystkich przypadków daltonizmu. Użytkownik praktycznie widzi te informacje tekstowe, które odróżniają się od tła na którym są wyświetlane, co widzi użytkownik ze ślepotą barwy na stronie www przedstawione jest w tabeli 2.

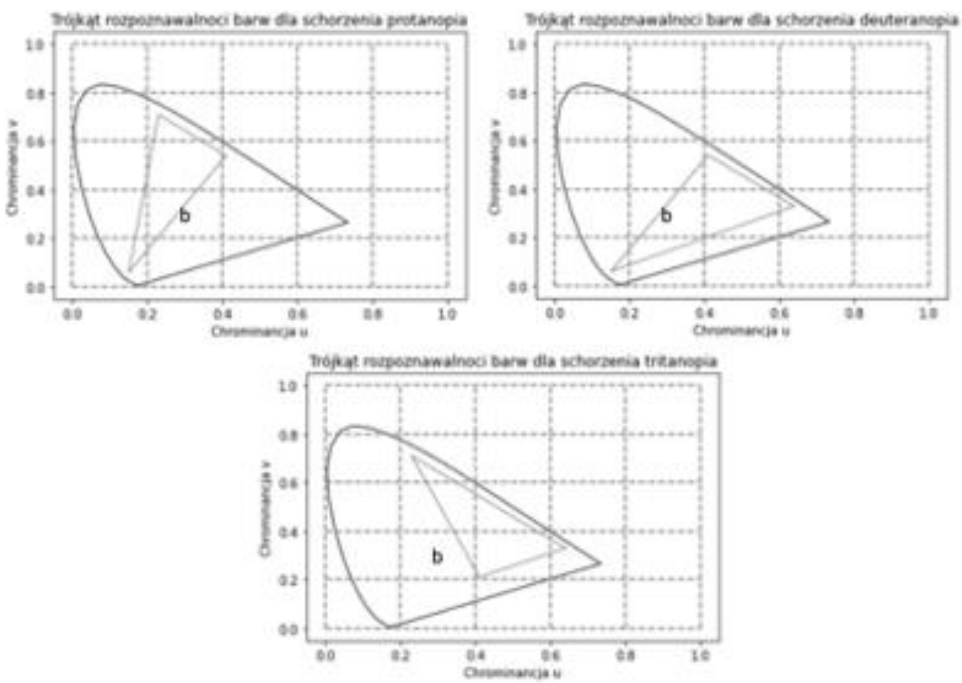
Tabela 2. Warunki widoczności *zawartości strony www (opracowanie własne)*

WIDZI

Tło	Kontent
Rozpoznaje (w trójkącie)	Rozpoznaje (w trójkącie)
Nie rozpoznaje (poza trójkątem)	Rozpoznaje (w trójkącie)
Rozpoznaje (w trójkącie)	Nie rozpoznaje (poza trójkątem)

NIE WIDZI

Tło	Kontent
Nie rozpoznaje (poza trójkątem)	Nie rozpoznaje (poza trójkątem)



Rysunek 6. Przykład dla barwy białej RGB

5. Analiza systemu informatycznego dla Zintegrowanych Węzłów Przesiadkowych (ZWP) w zakresie spełnienia szczególnych potrzeb osób niepełnosprawnych

Czytelność schematu węzła przesiadkowego można poprawić przez dobór neutralnych kolorów dla różnych przypadków ślepoty kolorów. Przykładowo na rysunku 7 przedstawiony jest rzut 3D portu lotniczego z zaznaczonymi punktami usługowymi oraz toaletami. W przypadku Protanopii te miejsca są praktycznie nierozróżnialne. W czytelności głównie zależy nam na tym, aby informacje graficzne były rozróżnialne. Niezależnie od tego jak ją widzą osoby z deficytem widzenia kolorów. Przylegające do siebie zaznaczenia w schemacie portu lotniczego powinny się wyraźnie wyróżniać. W przypadku tych trzech schorzeń postrzegania barw nie da się wybrać barw wspólnie widzialnych, ze względu na to chociażby, że trójkąty opisujące protanopie i deuteranopie nie mają wyraźnej części wspólnej, mają tylko wspólną krawędź. Jednym z rozwiązań tego problemu jest osobny zestaw barw w aplikacji webowej dla każdego z tych schorzeń.

Rysunek 7. Widok normalny i z filtrem dla protanopii [https://www.toptal.com/designers/colorfilter]

Dobór barw dla każdego schorzenia powinien zakładać, że wszystkie wybrane barwy w postaci wektora RGB mieszczą się po przeliczeniu w trójkątach na modelu CIE xyz. Najprostszym algorytmem realizującym to jest wybór wektora *u, v* i przeliczenie go na system aRGB w celu określenia koloru dla elementu strony www. Innym rozwiązaniem problemu widoczności może być miganie kilkusekundowe wybranych obiektów na schemacie.

Potrzeby w zakresie wizualizacji treści graficznych dla poszczególnych użytkowników mogą być różne i wynikają z ich spersonalizowanych preferencji oraz ograniczeń. Również dla osób ze ślepotą barw charakterystyczne osobniczo możliwości rozróżnialności i interpretacji kolorów mogą być różne. Z tego powodu informacje o możliwościach percepcji użytkownika muszą być zapisane w profilu, co ułatwi spersonalizowanie treści graficznych, tekstów i pozwoli na łatwiejszy odbiór przekazywanych informacji. Kategorie poszczególnych potrzeb w zakresie wsparcia dla osób z niepełnosprawnościami mogą jednak obejmować szersze spektrum ograniczeń i mogą dotyczyć takich użytkowników jak:

- niedowidzących i niewidomych;
- niedosłyszących i głuchych;
- osoby niepełnosprawne ruchowo;
- osoby starsze i dzieci;
- osoby cierpiące na padaczkę fotogenną (błyski animacji mogą wywołać atak epilepsji);

- osoby z niepełnosprawnością intelektualną i zaburzeniami poznawczymi;
- osoby z dysleksją;
- osoby nieobeznane z technologią lub używające starszych urządzeń o ograniczonych możliwościach wizualizacji

a także dla tych, którzy potrzebują tymczasowych udogodnień, oraz produkt dostosowany do konkretnych sytuacji.

Badania nad walidacją przedstawianego systemu wykazały, iż część użytkowników należy do wielu grup związanych ze szczególnymi potrzebami. Ograniczenia i preferencje poszczególnych grup są nierzadko rozłączne, z tego względu system spełnia najlepiej oczekiwania użytkowników w przypadku dynamicznie generowanych treści w zależności od ustawień preferencji zapisanych w profilu użytkownika. Takie podejście jest zgodne z zasadami projektowania uniwersalnego. Starając się zaspokoić potrzeby wielu użytkowników poprzez jedną uniwersalna aplikację zamiast tworzyć dedykowane wersje rozwiązań obniżamy koszty potrzebne na zbudowanie wymaganego oprogramowania i oferujemy rozwiązanie adekwatne do wymogów użytkowników.

6. Podsumowanie

Osoby ze szczególnymi potrzebami mają znacznie większe wymagania odnośnie dostarczenia im informacji o zintegrowanym węźle przesiadkowym niż osoby w pełni sprawne. W zakresie poruszania się po węźle przesiadkowym ograniczenia są szczególnie ważne dla osób z niepełnosprawnościami ruchowymi. Obecnie efektywne korzystanie z węzłów przesiadkowych wymaga jednak prawidłowej interpretacji informacji, która w wielu przypadkach jest prezentowana w sposób graficzny. Często bez możliwości korzystania z cyfrowych metod przekazywania takiej informacji użytkownicy węzła przesiadkowego nie są w stanie prawidłowo korzystać z jego infrastruktury w zakresie kupna biletu, obserwowania zmian rozkładu jazdy, lub śledzenia ważnych dla podróżujących znaków informacyjnych. W ramach projektu „Przesiadka bez barier" powstał prototyp aplikacji webowej z informacjami o wybranym Zintegrowanym Węźle Przesiadkowym (ZWP). Aplikacja dostępna jest na większość urządzeń mobilnych i stacjonarnych. Testy zgodności aplikacji ze standardami wysokiej dostępności pozwoliły na stworzenie narzędzia, które w jak najlepszy sposób zaspokaja potrzeby użytkowników ze szczególnymi ograniczeniami w zakresie interpretacji zawartości informacyjnej. Wykorzystując wiedzę zaprezentowaną w tej pracy udało się poprawić dostępność treści informacyjnych aplikacji w grupie osób z deficytem widzenia barwnego. Zaprezentowana metoda może w przyszłości pomóc projektantom i developerom stron internetowych na zwiększenie ich widoczności. Efektywne dynamiczne dostosowanie treści graficznych może być również podstawą do stworzenia uniwersalnego narzędzia, które pozwoliłoby użytkownikom ze ślepotą barw na łatwiejsze korzystanie z urządzeń cyfrowych, takich jak tablet, telefon komórkowy czy komputer. Na podstawie przedstawionej metody detekcji niedostrzegalnych dla użytkownika barw, oraz konwersji dynamicznej obrazu na alternatywną postać wykorzystującą czytelną paletę kolorystyczną można uzyskać treść dostosowaną dynamicznie do szczególnych potrzeb odbiorców. Z zastosowaniem systemów rzeczywistości rozszerzonej i urządzeń mobilnych [27] metoda ta może pomóc w interpretacji informacji przekazywanej nie tylko drogą elektroniczną, ale również w postaci wszelkiego rodzaju ogłoszeń i tablic informacyjnych umieszczanych w centrach przesiadkowych, lub innych miejscach odwiedzanych przez użytkowników z dysfunk-

cjami interpretacji barw. Metoda ta może być pomocna nie tylko dla osób z niepełnosprawnościami, ale również użytkowników urządzeń graficznych o słabszych parametrach wizualizacyjnych, które mają mniejsze możliwości w zakresie interpretacji barw. Przykładem tego mogą być wszelkiego rodzaju urządzenia oparte o starsze wyświetlacze, lub czytniki bazujące na technologii papieru elektronicznego (ang. e-Ink). W ramach kontynuacji prac badawczych nad walidacją i poprawą dostępności treści graficznych autorzy planują następujące kierunki rozwoju:

- zastosowanie przedstawionej metody do opracowania efektywnego narzędzia pozwalającego na interpretację treści kolorystycznych, które będzie działało po stronie urządzenia mobilnego użytkownika;
- rozbudowę prezentowanej metody o detekcję trójkąta prawidłowej interpretacji barw, które są charakterystyczne osobniczo i mogą być zindywidualizowane dla poszczególnych użytkowników lub grup użytkowników;
- statystyczne badania w zakresie możliwości interpretacji kolorów dla poszczególnych grup użytkowników, co może prowadzić do opracowania wytycznych w zakresie doboru palety kolorystycznej dającej lepszą dostępność informacji graficznej dla osób ze szczególnymi potrzebami.

Literatura

1. Kuryłowicz E., *Projektowanie uniwersalne*, Centrum Badawczo Rozwojowe, Warszawa 1996.
2. Rakowska A., *Język – komunikacja – niepełnosprawność, wybrane zagadnienia*, Wydawnictwo Naukowe Akademii Pedagogicznej, Kraków 2003.
3. Urbanowicz Z., *Od interdyscyplinarnego do transdyscyplinarnego spojrzenia na niepełnosprawność*, Ogrody Nauk i Sztuk, Warszawa 2012, s. 443.
4. Robbins J., *Projektowanie stron internetowych. Przewodnik dla początkujących webmasterów po HTML5, CSS3 i grafice. Wydanie V*, Helion SA, Gliwice 2020.
5. Dawid A., Buchwald P., *Performance of Modern Video Codecs in Real-Time Communication WebRTC Protocol*, [w:] Zamojski W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk J. (red.), *New Advances in Dependability of Networks and Systems*, Springer International Publishing, Berlin 2022, s. 32-41.
6. Jankowski M., Skublewska-Paszko M., *Analiza zastosowania języków Java oraz C# do budowy aplikacji mobilnej na platformę Android*, Journal of Computer Sciences Institute, 16, 2020.
7. Buchwald P., Anus A., *Industrial internet of things systems for tracking and traceability of production business processes*, Multidisciplinary Aspects of Production Engineering 3, 2020, s. 464-476.
8. Grudzień M., Korgol K. & Gutek D. *Porównanie możliwości wykorzystania oraz analiza wydajności baz danych na systemach mobilnych*, Journal of Computer Sciences Institute, 2, 2016.
9. Kułak J., *Wykorzystanie urządzeń mobilnych w zachowaniach zakupowych konsumentów w Polsce. Testowanie modelu akceptacji technologii UTAUT2*, 2018.
10. Frain B., *Responsive Web Design. Projektowanie elastycznych witryn w HTML5 i CSS3*, Helion SA, Gliwice 2014.
11. Bernacki J., *Responsive Web Design: Testing Usability of Mobile Web Applications*, [w:] Nguyen N.T., Trawiński B., Fujita H., Hong T.P. (red.), *Intelligent Information and Database Systems*, Springer, Berlin 2016, s. 257-269.
12. Hung J.C., Wang C., *Exploring the website object layout of responsive web design: results of eye tracking evaluations*, J Supercomput, 77, 2021, s. 343-365.

13. White J., *WCAG 2.1 Meets STEM: Application, Interpretation, and Opportunities for Further Standard Development*, Journal of Science Education for Students with Disabilities, 22, 2019.
14. Skierniewska M., Skroban K., *Dostępność oprogramowania użytkowego dla osób niepełnosprawnych i starszych w świetle standardów WCAG*, Roczniki Kolegium Analiz Ekonomicznych / Szkoła Główna Handlowa, 2018, s. 141-152.
15. WCAG 2.0 - Dostępność stron www. https://wcag20.widzialni.org/. [data dostępu: 10.01.2023].
16. Smith G., *PostgreSQL 9.0: High Performance*, Packt Publishing Ltd, 2010.
17. Rokooei S., *Building Information Modeling in Project Management: Necessities, Challenges and Outcomes*, Procedia - Social and Behavioral Sciences, 2015, s. 87-95.
18. Eastman C.M., Teicholz P., Sacks R., Liston K., *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors*, John Wiley & Sons, 2011.
19. Three.js – JavaScript 3D Library. https://threejs.org/. [data dostępu: 10.02.2023].
20. Szluz B., *Doświadczanie niepełnosprawności sprzężonej w rodzinie (w biografii opiekuna rodzinnego)*, Wydawnictwo Akademii Pedagogiki Specjalnej, 2018.
21. WebGL, *The Khronos Group* https://www.khronos.org//. [data dostępu: 05.01.2023]
22. Ganczarski J. CIE XYZ. (2004).
23. Molenda J., Wrona M., Siwiec E., *Zastosowanie modelu CIE Lab w badaniach barwy lotnych popiołów*, Problemy Eksploatacji, 3, 2012, s. 177-187.
24. Palus H., Bal A., *Impact of selected image acquisition parameters on image colorfulness*, Przegląd Elektrotechniczny, 84, 2008, s. 94-97.
25. Standard RGB Color Spaces, *Proc. IS&T/SID 7th Color Imaging Conference*, 1999.
26. Katoh N., *Extended colour space for capturing devices*, 2005.
27. Buchwald P.,*Urządzenia mobilne w systemach rzeczywistości wirtualnej*, Wydawnictwo Helion, Gliwice 2018.

Metoda walidacji aplikacji internetowych przeznaczonych dla osób ze szczególnymi potrzebami

Streszczenie
Przemieszczanie się z użyciem transportu publicznego jest codziennością w dzisiejszym świecie. Nie dla wszystkich jednak ludzi przebiega ono jednakowo sprawnie. Osoby ze szczególnymi potrzebami, do których zalicza się osoby niepełnosprawne, matki z dziećmi, opiekunów z osobami starszymi czy osoby podróżujące z dużym bagażem podręcznym, mogą ze względu na swoje ograniczenia nie nadążyć za informacjami związanymi z przesiadką na inny środek transportu w obrębie zintegrowanych węzłów przesiadkowych (ZWP). Przy obecnie nieograniczonym dostępie do Internetu, pomocą w rozwiązaniu problemów przesiadkowych w ZWP mogą być aplikacje internetowe. Aktualnie istnieje wiele aplikacji internetowych usprawniających transport publiczny, lecz nie są one w gruncie rzeczy przyjazne dla osób ze szczególnymi potrzebami. Jednym z rozwiązań które nastawione jest dokładnie na te potrzeby jest prototyp aplikacji internetowej realizowanej w ramach projektu badawczego NCBiR pod tytułem „Przesiadka bez barier". W ramach tej pracy opracowano jednolitą metodę walidacji aplikacji internetowych pod kątem użytkowania przez osoby ze szczególnymi potrzebami. Użyteczność metody została przebadana na aplikacji internetowej wspomagającej poruszanie się osób ze szczególnymi potrzebami po ZWP Port Lotniczy Warszawa-Modlin.
Słowa kluczowe: tritanopia, protanopia, deuteranopia, ZWP, CIExyz

A method of validating web applications designed for people with special needs

Abstract
Moving around using public transport is an everyday reality in today's world. However, it does not run smoothly for all people. People with special needs, such as people with disabilities, mothers with children, carers with the elderly or people traveling with large hand luggage may, due to their limitations, not keep up with information related to changing to another means of transport within integrated transfer hubs (ITH).

With the currently unlimited access to the Internet, Internet applications may help in solving transfer problems in ITH. Currently, there are many online applications to improve public transport, but they are not really friendly to people with special needs. One of the solutions that is focused precisely on these needs is the prototype of an internet application implemented as part of the research project of the National Center for Research and Development entitled "Transfer without barriers". As part of this work, a uniform method for validating web applications for use by people with special needs was developed. The usability of the method was tested on a web application supporting the movement of people with special needs around the ITH Warsaw-Modlin Airport.

Keywords: tritanopia, protanopia, deuteranopia, ITH, CIExyz

Anna Jackiewicz-Zagórska[1], Jakub Maksymilian Gac[2], Bartosz Nowak[3], Łukasz Werner[4]

Modelowanie procesu usuwania z powietrza cząstek aerozolowych w nowoczesnych kompozytowych materiałach polipropylen-ZnO

1. Wprowadzenie

Cząstki aerozolowe zawieszone w powietrzu stanowią jedno z najpoważniejszych zagrożeń dla zdrowia człowieka oraz środowiska. Ekspozycja na cząstki abiotyczne, tj. nieożywione (cząstki pyłów), oraz biotyczne (bakterie, wirusy) [1] stwarza niebezpieczeństwo dla człowieka zarówno na zewnątrz, jak i wewnątrz budynków. Może ona być przyczyną wielu chorób, szczególnie układu oddechowego, w tym syndromu chorego budynku [2]. Badania nad skutecznymi metodami usuwania tych cząstek z powietrza są niezwykle istotne w kontekście ochrony zdrowia publicznego oraz zapewnienia czystego i bezpiecznego środowiska życia. W ostatnich latach zainteresowanie skoncentrowało się na rozwijaniu nowoczesnych kompozytowych materiałów na bazie polimerów nie tylko o dobrych właściwościach filtracyjnych, ale też o dodatkowych właściwościach m.in. sorpcyjnych, biobójczych, fotokatalitycznych. Takie materiały jak TiO_2 (tlenek tytanu) lub ZnO (tlenek cynku) mogą być używane do oczyszczania powietrza poprzez procesy fotokatalityczne, które rozkładają zanieczyszczenia na niegroźne składniki [3]. Węgiel aktywny i jego kompozyty są często stosowane w filtrach powietrza ze względu na ich zdolność do adsorpcji wielu rodzajów zanieczyszczeń [4]. Niektóre nowoczesne techniki skupiają się na wykorzystaniu naturalnych materiałów, takich jak biokompozyty, do oczyszczania powietrza. Mogą one pochłaniać zanieczyszczenia lub uwalniać naturalne olejki eteryczne, które pomagają w poprawie jakości powietrza [5]. Dla przyszłych technologii oczyszczania powietrza bardzo obiecującymi materiałami są tzw. MOF'y (ang. *Metal-Organic Frameworks*). Są to materiały mające bardzo duże powierzchnie wewnętrzne, które mogą adsorbować duże ilości zanieczyszczeń [6].

W ramach pracy opracowano model numeryczny opisujący proces usuwania z powietrza aerozolowych cząstek abiotycznych i biotycznych, który wspomoże projektowanie wysokosprawnego kompozytowego materiału filtracyjnego polipropylen-tlenek cynku (PP-ZnO) oraz pozwoli na przewidywanie jego zachowania podczas oczyszczania powietrza z różnego rodzaju zanieczyszczeń. Bazowym materiałem w opracowywanym kompozycie jest materiał włókninowy wykonany techniką rozdmuchu stopionego polimeru *(*ang. *melt-blown*), który zostaje zmodyfikowany bezpiecznym dla ludzi i środowiska związkiem inhibitującym procesy życiowe drobnoustrojów, tj. tlenkiem cynku [7, 8]. Połączenie tych dwóch materiałów w kompozycie PP-ZnO może prowadzić do

[1] anna.jackiewicz@pw.edu.pl, Wydział Inżynierii Chemicznej i Procesowej, Politechnika Warszawska, www.ichip.pw.edu.pl.
[2] Wydział Inżynierii Chemicznej i Procesowej, Politechnika Warszawska, www.ichip.pw.edu.pl.
[3] Wydział Inżynierii Chemicznej i Procesowej, Politechnika Warszawska, www.ichip.pw.edu.pl.
[4] Wydział Inżynierii Chemicznej i Procesowej, Politechnika Warszawska, www.ichip.pw.edu.pl.

synergistycznego efektu, który zwiększa skuteczność usuwania cząstek aerozolowych różnego pochodzenia.

Projektowany materiał, poprzez swą rozbudowaną bioaktywną strukturę, ma: (i) charakteryzować się dłuższym czasem użytkowania, aniżeli jego odpowiedniki z czystego polipropylenu; (ii) zapewnić oszczędność energii z uwagi na wolniejszy przyrost spadku ciśnienia; (iii) umożliwić redukcję gabarytów urządzenia oczyszczającego powietrze, z uwagi na skuteczne działanie zarówno w przypadku cząstek nieożywionych, jak i tych ożywionych. Nowopowstały kompozyt ma wykazać się lepszymi właściwościami w porównaniu do obecnie dostępnych na rynku rozwiązań, głównie ze względu na połączenie właściwości powierzchniowych i sorpcyjnych. Zamodelowane struktury mogą znaleźć zastosowanie np. w systemach wentylacyjno-klimatyzacyjnych HVAC (ang. *Heating, Ventilation, Air Conditioning*), w filtrach samochodowych, okapach kuchennych oraz środkach ochrony osobistej. Biorąc pod uwagę specyficzne wilgotne środowisko, w którym będą zwykle pracować, sprzyjające namnażaniu się bakterii na filtrach, zastosowanie tlenku cynku na włóknach może spowolnić lub nawet zatrzymać wzrost licznych form biologicznych na ich powierzchni [9-11]. Dodatkowym atutem modelowanych materiałów będzie ich fotokatalityczna aktywność – tlenek cynku może rozkładać niebezpieczne dla zdrowia lotne związki organiczne VOC (ang. *volatile organic compounds*).

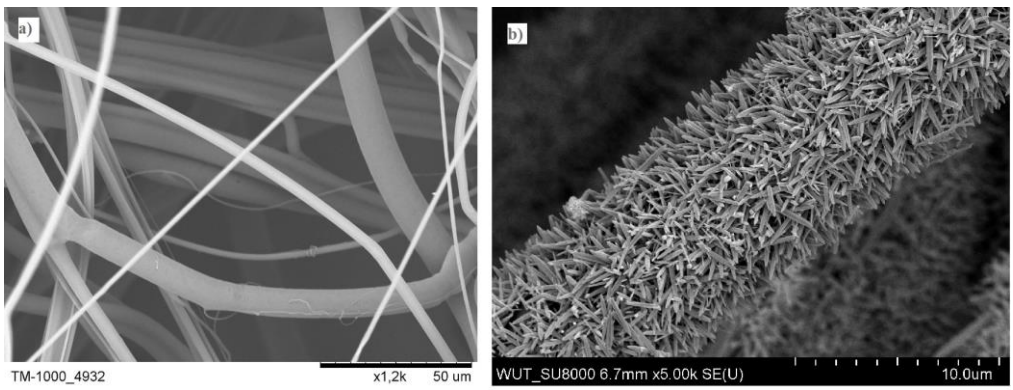
Przedstawione w pracy wyniki badań są istotne dla dalszego rozwoju nowoczesnych materiałów polimerowych o zwiększonej zdolności usuwania zarówno cząstek abiotycznych, jak i biotycznych z powietrza. Skuteczna separacja aerozoli ma znaczenie strategiczne – może przyczynić się do poprawy jakości powietrza w przemyśle, w budynkach mieszkalnych, w budynkach użyteczności publicznej i środowiskach zewnętrznych, czyli tym samym może mieć pozytywny wpływ na zdrowie publiczne.

2. Model numeryczny

Podstawą obliczeń skuteczności działania zaproponowanego filtra jest analiza oddziaływań między gazem, cząstkami i pojedynczymi włóknami polimerowymi zmodyfikowanymi tlenkiem cynku. Dzięki temu można wyznaczyć tzw. wydajność pojedynczego włókna (SFE, ang. *single fiber efficiency*), czyli wydajność usuwania cząstek na pojedynczym włóknie. W dalszej części rozdziału zostanie przedstawiony sposób wyznaczenia na tej podstawie całkowitej sprawności filtra.

Geometria pojedynczego włókna pokrytego pręcikami tlenku cynku została odwzorowana na podstawie obserwacji wyników wstępnych modyfikacji przeprowadzonych przez nasz zespół badawczy na drodze chemicznej depozycji z roztworu (CBD, ang. *Chemical Bath Deposition)* [12]. Była ona prowadzona trójetapowo: umieszczenie materiału w izopropanolu (IPA) na 24 godziny, a następnie wypłukania alkoholu z materiału wodą z użyciem sonikacji; etap nukleacji przy użyciu dwutlenku manganu i etap hydrotermalnego wzrostu pręcików w podwyższonej temperaturze. Zarodkowanie polegało na strąceniu nanocząstek tlenku manganu z roztworu nadmanganianu potasu, przy niewielkim dodatku n-butanolu i ich adsorpcji na powierzchni włókna. Następnie, w ten sposób zarodkowana powierzchnia umieszczana była w mieszaninie do wzrostu pręcików tlenku cynku. Bazowy materiał włókninowy został wykonany z polipropylenu, wykorzystując technikę rozdmuchu stopionego polimeru [13]. Jest to metoda stosunkowo tania, pozwalająca na wyprodukowanie materiału o założonych parametrach strukturalnych.

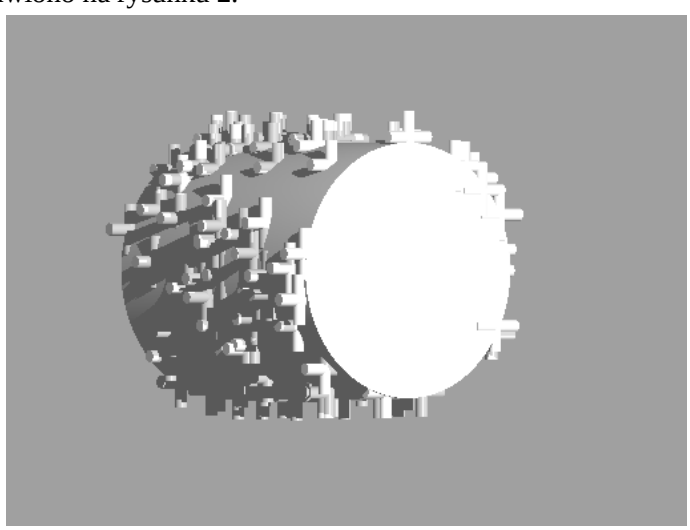
Zdjęcia bazowych włókien polipropylenowych oraz kompozytowego włókna polimer--ZnO otrzymanych wyżej opisanymi metodami wykonane za pomocą skaningowego mikroskopu elektronowego przedstawiono na rysunku 1.



Rysunek 1. Zdjęcia SEM: bazowych polimerowych włókien (a) oraz włókna zmodyfikowanego tlenkiem cynku (b) [opracowanie własne]

2.1. Domena obliczeniowa

Domena obliczeniowa to obszar cylindryczny z wewnętrznym – współosiowym – włóknem filtrującym, które jest opisane jako prosty cylinder. Średnica domeny była tak dobrana, aby zapewnić odpowiednią wartość gęstości upakowania jako stosunku objętości włókna do objętości domeny. Pręciki ZnO są reprezentowane jako małe cylindry obecne na powierzchni włókna. W celu uproszczenia opisu – nie tracąc przy tym najważniejszych cech – małe cylindry (pręciki) zorientowane są w dwóch kierunkach: poziomym i pionowym. Przykładowy widok takiego włókna zmodyfikowanego pręcikami przedstawiono na rysunku 2.



Rysunek 2. Przykładowy widok fragmentu zmodyfikowanego włókna [opracowanie własne]

2.2. Zastosowane metody numeryczne

Model numeryczny filtracji cząstek na filtrach modyfikowanych tlenkiem cynku składał się z dwóch elementów:

- model przepływu płynu wokół włókien: model gazu sieciowego Boltzmanna (ang. *lattice Boltzmann model*, LBM);
- model ruchu pojedynczej cząstki w strukturze włóknistej: równanie ruchu oparte na równaniu Langevina.

Model LBM znalazł szerokie zastosowanie jako wygodna metoda opisu pola prędkości płynu (gazu lub cieczy) oraz rozkładu ciśnienia w płynie dla obszaru o złożonej geometrii. Szczególnie przy stosunkowo niskich prędkościach liniowych przepływu (ruch laminarny) konkuruje z popularnymi i powszechnie stosowanymi metodami obliczeniowej dynamiki płynów (ang. *computational fluid dynamics*, CFD), takimi jak metoda elementów skończonych MES (ang. *finite element method*, FEM).

Model gazu sieciowego Boltzmanna (LBM) oparty jest na równaniu kinetycznym Boltzmanna, spełnianym przez funkcję gęstości prawdopodobieństwa $f\ (\vec{x},\vec{v},t)$ (ang. *probability density function*, PDF) która oznacza gęstość prawdopodobieństwa znalezienia cząstki w punkcie $\vec{x}$ i czasie *t*, o prędkości $\vec{v}$ [14, 15]. Równanie to ma postać (przy pominięciu zewnętrznych sił masowych, np. grawitacji):

$$\frac{\partial}{\partial t}\, f(\vec{x},\vec{v},t) + \vec{v}\cdot\frac{\partial}{\partial \vec{x}}\, f(\vec{x},\vec{v},t) = \int d\Omega \int d\vec{u}\sigma(\Omega)|\vec{v}-\vec{u}|\left[f\left(\vec{x},\overrightarrow{v'},t\right)\cdot f\left(\vec{x},\overrightarrow{u'},t\right) - f(\vec{x},\vec{v},t)\cdot f(\vec{x},\vec{u},t)\right] \quad (1)$$

W powyższym równaniu Ω jest kątem bryłowym rozpraszania, $\sigma(\Omega)$ – różniczkowym przekrojem czynnym na zderzenie pod kątem Ω, natomiast $\overrightarrow{v'}$ i $\overrightarrow{u'}$ to prędkosci, jakie uzyskają cząstki o prędkościach początkowych $\vec{v}$ i $\vec{u}$ po zderzeniu pod kątem Ω.

Jeśli funkcja $f\ (\vec{x},\vec{v},t)$ jest znana, można uzyskać wszystkie lokalne wartości płynu, takie jak lokalne ciśnienie, prędkość, energia wewnętrzna itp.

Główną koncepcją metody LBM [16] jest zastąpienie funkcji gęstości $f\ (\vec{x},\vec{v},t)$ jej dyskretnym odpowiednikiem $f_i\ (\vec{x},t)$. Ten dyskretny analog funkcji gęstości prawdopodobieństwa jest zależny od współrzędnej przestrzennej $\vec{x}$, czasu *t* i kierunku prędkości, oznaczonego indeksem dolnym *i*. Wszystkie trzy wartości są dyskretyzowane. Dwa kolejne etapy opisują ewolucję PDF: propagację i zderzenia, które naśladują ewolucję rzeczywistych cząstek w płynie.

Ewolucja ta przebiega zgodnie z następującym równaniem:

$$f_i\ (\vec{x}\ +\vec{c}_i\ ,t\ +1) -\ f_i\ (\vec{x},t) = \Omega_i\ (\vec{x},t) \quad (2)$$

gdzie: $\vec{c}_i$ oznacza zdyskretyzowany wektor kierunku prędkości, a $\Omega_i\ (\vec{x},t)$ jest operatorem zderzenia i odgrywa rolę całki zderzenia w równaniu kinetycznym Boltzmanna – prawa strona (1) [17]. Najpopularniejszy sposób wyrażenia tego operatora opiera się na podejściu Bhatnagara, Grossa i Krooka (BGK) [18]:

$$\Omega_i\ (\vec{x},t) = -\frac{1}{\tau}\left(f_i\ (\vec{x},t) - f_i^{eq}\ (\vec{x},t)\right) \quad (3)$$

W powyższym równaniu $f_i^{eq}\ (\vec{x},t)$ jest równowagową funkcją PDF, która zależy od kierunku *i* oraz lokalnej prędkości płynu $\vec{u}$ [16]:

$$f_i^{eq}(\vec{x},t) = \rho w_i\left(1 + 3\vec{c}_i \cdot \vec{u} - \frac{3}{2}\vec{u}^2\right) \quad (4)$$

gdzie: wartości współczynników w_i, tzw. współczynniki wagowe, zależą od geometrii siatki numerycznej i można je znaleźć m.in. w artykule Chena i Doolena [16].

Jeżeli PDF jest znana we wszystkich węzłach domeny obliczeniowej, gęstość gazu sieciowego ρ (nie mylić z gęstością płynu), ciśnienie płynu p oraz jego prędkość $\vec{u}$ są obliczane zgodnie z następującymi wyrażeniami:

$$\rho(\vec{x},t) = \sum_i f_i(\vec{x},t) \quad (5)$$

$$p(\vec{x},t) = c_s^2 \rho(\vec{x},t) \quad (6)$$

$$\rho(\vec{x},t)\vec{u}(\vec{x},t) = \sum_i f_i(\vec{x},t) \cdot \vec{c}_i \quad (7)$$

Znając pole przepływu wokół włókna, można następnie wyznaczyć dynamikę ruchu cząstek i ich osadzania na włóknie. W tym celu stosuje się metodę dynamiki Langevina. Zgodnie z tą metodą ruch każdej pojedynczej cząstki spełnia następujące równanie [19]:

$$\ddot{\vec{x}} = {-\Delta U(\vec{x})}/{m} - \gamma\dot{\vec{x}} + \sqrt{\frac{2\gamma k_B T}{m}}\Gamma(t) \quad (8)$$

gdzie: $\vec{x}$ oznacza wektor położenia cząstki, $U(\vec{x})$ – energię potencjalną cząstki (zarówno oddziaływań z innymi cząstkami, jak i między cząstką a płynem), γ – współczynnik tłumienia lepkiego i $\Gamma(t)$ – składnik stochastyczny, odpowiedzialny za dyfuzję cząstek. Ta ostatnia wielkość spełnia następujące warunki:

$$\langle\Gamma(t)\rangle = 0 \quad (9a)$$

$$\langle\Gamma(t_1)\Gamma(t_2)\rangle = \delta(t_1 - t_2) \quad (9b)$$

Biorąc pod uwagę, że jedyną zewnętrzną siłą działającą na cząstkę jest siła oporu pochodząca z przepływu gazu i zakładając, że cząstka jest kulista (a więc współczynnik tłumienia wynosi $\gamma = 3\pi\mu d$), równanie nr 8 przekształca się do postaci:

$$\ddot{\vec{x}} = {3\pi\mu d\left(\vec{u}(\vec{x}) - \dot{\vec{x}}\right)}/{m} + \sqrt{\frac{6\pi\mu d k_B T}{m}}\Gamma(t) \quad (10)$$

Cząstki o określonej średnicy (możliwy jest również rozkład średnic) były sukcesywnie wprowadzane w kierunku włókna. Ruch każdej cząstki był śledzony aż do momentu jej osadzenia na włóknie, na pręciku ZnO, na jakiejś innej wcześniej zdeponowanej cząstce lub do opuszczenia domeny obliczeniowej po minięciu włókna (na ścianach prostopadłych do osi zastosowano periodyczne warunki brzegowe). W ten sposób możliwe było śledzenie miejsca depozycji cząstek i scharakteryzowanie powstałych agregatów filtracyjnych.

Obliczenie skuteczności filtracji przeprowadzono w następujący sposób. Podczas symulacji liczono liczbę cząstek, które osiadły na włóknie lub pręciku ZnO N_{dep}. Ta liczba była podstawą do obliczenia wydajności pojedynczego włókna SFE:

$$\eta = \frac{N_{dep}}{N} \quad (11)$$

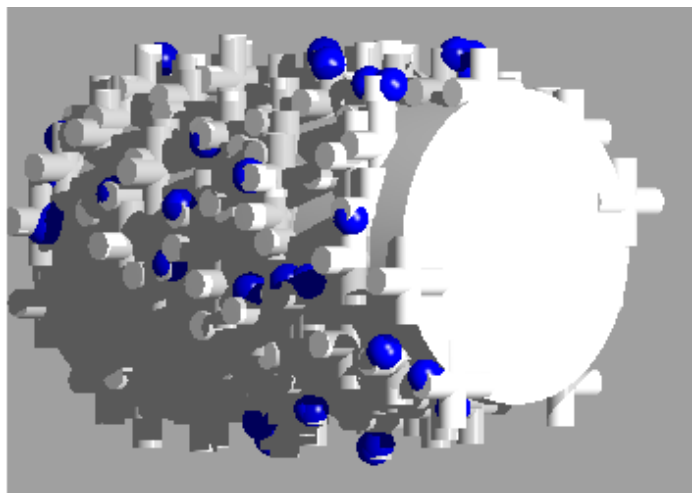
gdzie: N jest całkowitą liczbą cząstek uwolnionych do domeny obliczeniowej. Następnie obliczono całkowitą skuteczność filtracji za pomocą wzoru opracowanego na podstawie klasycznej teorii filtracji:

$$E = 1 - exp\left(-\frac{4}{\pi}\frac{\alpha}{1-\alpha}\frac{Z}{d_f}\eta\right) \tag{12}$$

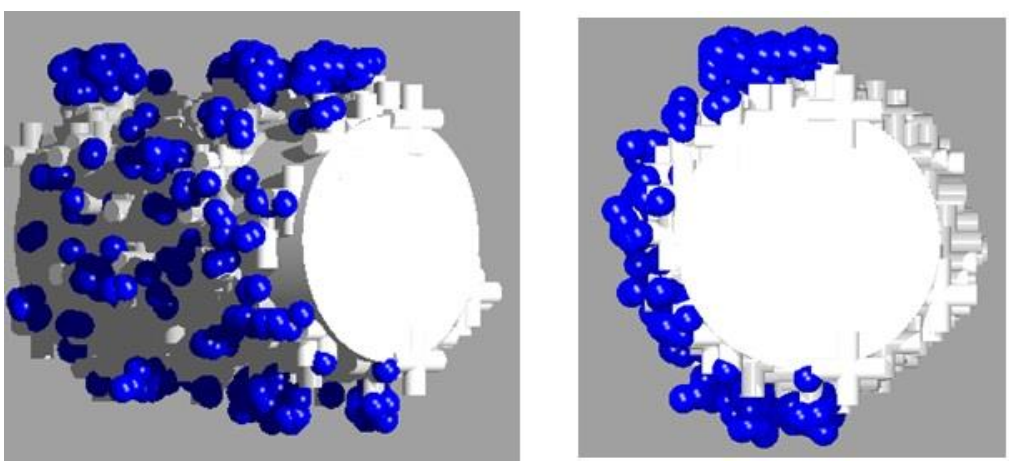
gdzie: α oznacza gęstość upakowania włókien w filtrze ($\alpha = 1 - \varepsilon$, gdzie ε oznacza porowatość filtra), Z – grubość filtra, a d_f – średnia średnica włókien filtracyjnych.

3. Wyniki symulacji numerycznych

Przykładowe wyniki obładowania włókna cząstkami (oznaczone kolorem niebieskim) przedstawiono na rysunkach 3 (dla 300 cząstek) i 4 (dla 1000 cząstek). Obliczenia wykonano dla cząstek o średnicy zbliżonej do średnicy cząstek biotycznych (bakterii, wirusów, grzybów), tj. 1 μm.



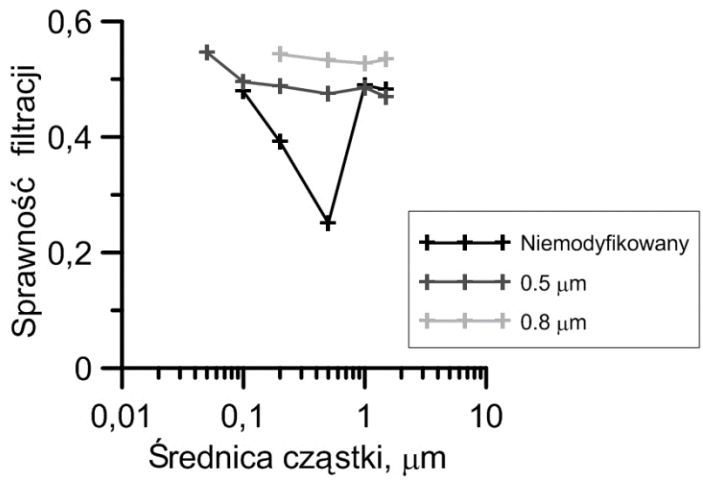
Rysunek 3. Włókno obładowane cząstkami (kolor niebieski); średnica cząstki: 1 μm, liczba cząstek: 300 [opracowanie własne]



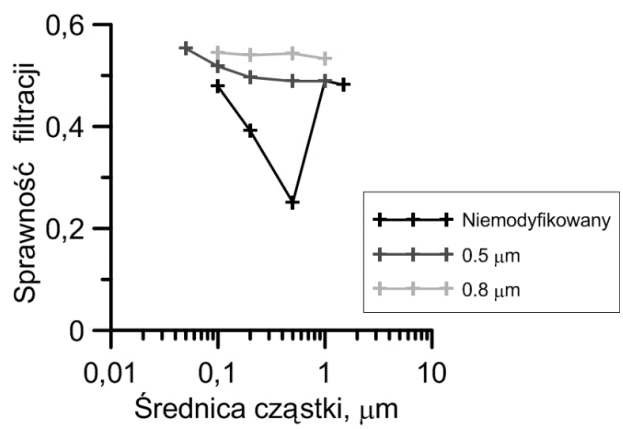
Rysunek 4. Włókno obładowane cząstkami (kolor niebieski); średnica cząstki: 1 μm, liczba cząstek: 1000 [opracowanie własne]

Na rysunku 4 widać, że zatrzymywane cząstki tworzą na powierzchni pręcików ZnO aglomeraty, tzn. świeżo napływające na filtr cząstki osadzają się na tych uprzednio zdeponowanych.

Przechodząc do obliczeń sprawności filtracji, najpierw wzięto pod uwagę gęstość liniową pręcików ZnO równą 25 pręcikom na bieżący mikrometr włókna. Średnica włókna wynosi 9 μm, a grubość całego filtra 2 mm. Wyniki przedstawiono na rysunku 5. Zaobserwowano, że całkowita wydajność filtra znacznie się poprawia w obecności pręcików ZnO. Największą poprawę obserwuje się dla cząstek najbardziej penetrujących (ang. *the most penetrating particle size*, MPPS), czyli tych, które są najsłabiej zatrzymywane przez filtr (o średnicy równej 200-400 nm). Jeśli chodzi o długość pręcików, to sprawność filtra z włóknami pokrytymi pręcikami o długości 0,8 μm jest znacznie wyższa niż tych pokrytych pręcikami o długości 0,5 μm. Tę samą tendencję obserwuje się w przypadku większej gęstości pręcików ZnO, tj. 40 pręcików na 1 μm włókna, co przedstawiono na rysunku 6.

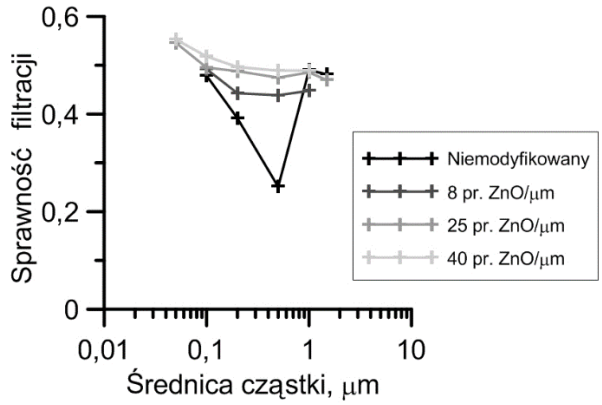


Rysunek 5. Sprawność filtracji dla filtra niemodyfikowanego oraz dla dwóch długości pręcików ZnO. Gęstość liniowa pręcików: 25 pręcików na 1 μm włókna [opracowanie własne]



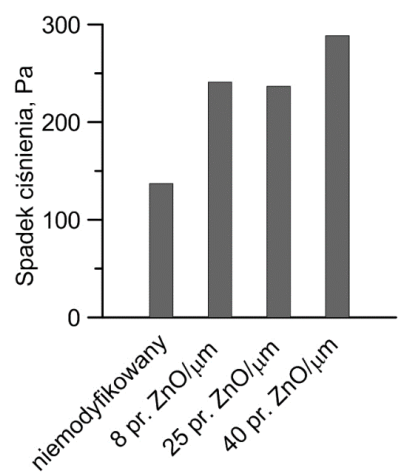
Rysunek 6. Sprawność filtracji dla filtra niemodyfikowanego oraz dla dwóch długości pręcików ZnO. Gęstość liniowa pręcików: 40 pręcików na 1 μm włókna [opracowanie własne]

Kolejnym krokiem było porównanie sprawności filtrów z włóknami pokrytymi różną liczbą pręcików ZnO. Na rysunku 7 zaprezentowano wyniki uzyskane dla trzech gęstości pokrycia pręcikami ZnO: 8, 25 i 40 pręcików na mikrometr włókna. Średnica włókien nadal wynosi 9 μm, a grubość filtra – 2 mm. Zauważono, że nawet najmniejsza gęstość liniowa pokrycia ZnO prowadzi do wyraźnej poprawy sprawności filtracji. Dalsze zwiększanie ilości pręcików ZnO nie powoduje znaczącej poprawy sprawności filtracji. W szczególności krzywe sprawności dla 25 i 40 pręcików na mikrometr włókna praktycznie się pokrywają.



Rysunek 7. Sprawność filtracji dla filtra niemodyfikowanego i dla trzech gęstości liniowych pręcików ZnO [opracowanie własne]

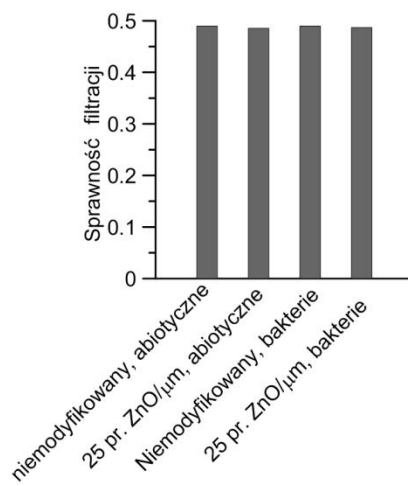
Kolejnym zagadnieniem niniejszej pracy było oszacowanie spadku ciśnienia filtrów niemodyfikowanych i modyfikowanych. Wyniki obliczeń tego parametru (w typowych warunkach prowadzenia procesu filtracji, tj. prędkości powietrza równej 0,2 m/s; długość pręcików wynosi 0,5 μm) przedstawiono na rysunku 8. Jak widać, obecność pręcików ZnO prowadzi do zwiększenia spadku ciśnienia w porównaniu z filtrem niemodyfikowanym. Jednak gęstość liniowa pokrycia pręcikami ponownie nie wpływa znacząco na ten parametr.



Rysunek 8. Spadek ciśnienia na filtrze z włóknami niemodyfikowanymi i modyfikowanymi pręcikami ZnO przy różnej gęstości pręcików [opracowanie własne]

Ostatnim podejmowanym w ramach pracy zagadnieniem było przeprowadzenie obliczeń filtracji cząstek biotycznych – komórek bakteryjnych. Rozważano tutaj bakterie typu *S. marcescens*, ponieważ dla tych bakterii dostępne są w literaturze dość szczegółowe dane dotyczące ich ruchu [20, 21]. Bakterie te są bakteriami urzęsionymi, przy czym ich rzęski mogą poruszać się w dwóch kierunkach. W przypadku, gdy wszystkie rzęski łączą się w jeden pęk i obracają w kierunku przeciwnym do ruchu wskazówek zegara, występuje tzw. stan ruchu: rzęski wytwarzają siłę o wielkości 0,48 pN. W przypadku, gdy jedna lub kilka rzęsek obraca się w kierunku zgodnym z ruchem wskazówek zegara, nie dochodzi do ich synchronizacji i powstania wypadkowej siły ciągu; wówczas komórka bakterii porusza się w sposób analogiczny, jak cząstka abiotyczna, tj. doznaje dyfuzji Brownowskiej oraz wpływu ośrodka poprzez wystąpienie siły oporu lepkiego. Badania pokazują [21], że rozkład czasu przebywania bakterii w obu tych stanach jest rozkładem Poissona, przy czym średni czas przebywania w stanie „synchronizacji rzęsek" wynosi 0,9 s, a w stanie „asynchronicznym" – 0,1 s. Ten model pozwala opisać dynamikę wielu innych typów bakterii, np. *Escherichia coli* [22], choć tu odpowiednie wartości liczbowe będą się różnić.

Zauważono, że sprawność filtracji kształtowała się na podobnym poziomie, jak w przypadku cząstek abiotycznych o podobnych rozmiarach (założono średnią średnicę komórki bakteryjnej równą 1 μm). Wynik ten przedstawia wykres na rysunku 9. Ta niewielka różnica wynika z faktu, że siła pochodząca od wici jest zbyt mała, aby skutecznie przeciwdziałać sile oporu lepkiego, która dominuje w przypadku dynamiki cząstek/komórek tej wielkości.



Rysunek 9. Porównanie sprawności filtracji cząstek abiotycznych oraz komórek *S. marcescens*, dla filtra niemodyfikowanego oraz zmodyfikowanego pręcikami ZnO [opracowanie własne]

Obecność tlenku cynku wpływała natomiast na spowolnienie rozwoju biofilmu na powierzchni włókien filtracyjnych. Stopień tego spowolnienia zależy od oddziaływania pomiędzy powierzchnią pręcików ZnO oraz błonami biologicznymi komórek. Niestety, obecne w literaturze dane nie pozwoliły na otrzymanie jednoznacznych wyników ilościowych, a jedynie jakościowe, tego spowolnienia. Temat ten wymaga dalszych badań.

4. Wnioski

Dokonano matematycznego i numerycznego opisu procesu filtracji cząstek aerozolowych na włóknach zarówno tych zmodyfikowanych, jak i niezmodyfikowanych tlenkiem cynku. Główne wnioski z pracy są następujące:

- Oparty na modelu LBM-BD opis filtracji cząstek abiotycznych na włóknach oryginalnych (niezmodyfikowanych) daje bardzo podobne wyniki jak klasyczna teoria filtracji i dlatego okazał się bardzo przydatnym narzędziem do modelowania procesu filtracji również filtrów modyfikowanych pręcikami ZnO.
- Pokrycie włókien filtracyjnych pręcikami powoduje poprawę skuteczności filtracji, zwłaszcza dla cząstek najbardziej penetrujących przez strukturę filtracyjną [23, 24] o średnicy 200-400 nm.
- Skuteczność filtracji znacznie wzrasta wraz z długością pręcików ZnO i jest bardzo słabo zależna od gęstości liniowej pręcików ZnO (liczba pręcików na długość włókna). Ten sam wniosek dotyczy spadku ciśnienia na filtrze.

Podsumowując uzyskane wyniki, można stwierdzić, że pręciki tlenku cynku pełnią podobną rolę w procesie filtracji jak osady filtracyjne. W ten sposób zwiększają początkową wydajność filtra i jednocześnie zwiększają początkowy spadek ciśnienia. Odpowiada to coraz popularniejszej metodzie obróbki osadów filtracyjnych powstałych podczas filtracji jako dodatkowych kolektorów – włókien lub pręcików, co zostało zastosowane m.in. w pracach Thomas i in. [25] oraz Kamińskiego i in. [26].

Biorąc powyższe pod uwagę, stosowanie filtrów włókninowych modyfikowanych tlenkiem cynku do usuwania cząstek abiotycznych ze strumieni gazów poprawia w znacznym stopniu skuteczność filtracji. Z kolei pręciki ZnO mogą mieć znaczenie dla filtracji gazów zawierających cząstki biotyczne (bakterie, wirusy, grzyby lub ich zarodniki), gdzie można wykorzystać biobójcze właściwości tlenku cynku. Prace nad tym tematem są kontynuowane. Zamodelowane struktury kompozytowe będą wytwarzane, badane pod kątem filtracji cząstek o różnej morfologii, a uzyskane wyniki posłużą do walidacji opracowanego modelu.

Podziękowania

Badania zostały sfinansowane przez POB Technologie Materiałowe Politechniki Warszawskiej w ramach programu „Inicjatywa Doskonałości – Uczelnia Badawcza”.

Literatura

1. Xie W., Bai Y., Li W., Hou J., Ma T., Zeng X., Zhang T., An, *The source and transport of bioaerosols in the air: A review*, Front. Environ. Sci. Eng., 15, 2021, s. 1-19, https://doi.org/10.1007/s11783-020-1336-8.
6. Nakayama Y., Nakaoka H., Suzuki N., Tsumura K., Hanazato M., Todaka E., Mori C., *Prevalence and risk factors of pre-sick building syndrome: characteristics of indoor environmental and individual factors*, Environmental Health and Preventive Medicine, 24, 2019, https://doi.org/10.1186/s12199-019-0830-8.
7. Chong M.N., Jin B., Chow C.W.K., Saint C., *Recent developments in photocatalytic water treatment technology: A review*, Water Research, 44 (10), 2010, 2997-30272010, https://doi.org/10.1016/j.watres.2010.02.039.
8. Ali A., Othman M.R., Kim J., *A review on application of activated carbons for carbon dioxide capture: present performance, preparation, and surface modification for further*

improvement, Environ Sci Pollut Res Int, 28(32), 2021, 43329-43364, doi: 10.1007/s11356-021-15121-9.
9. Jefferson J.A., Dhakal H.N., *Sustainable biobased composites for advanced applications: recent trends and future opportunities – A critical review*, Composites Part C: Open Access, 7, 2022, https://doi.org/10.1016/j.jcomc.2021.100220.
10. Li H., Wang C., Sun Y., Lollar C.T., Li J., Zhou H.C., *Recent advances in gas storage and separation using metal–organic frameworks*, Materials Today, 21(2), 2018, s. 108-121, https://doi.org/10.1016/j.mattod.2017.07.006.
11. Yamamoto O., *Influence of particle size on the antibacterial activity of zinc oxide*, International Journal of Inorganic Materials, 3, 2001, 643-646, https://doi.org/10.1016/S1466-6049(01)00197-0.
12. Zhang L., Jiang Y., Ding Y., Povey M., York D., *Investigation into the Antibacterial Behaviour of Suspensions of ZnO Nanoparticles (ZnO Nanofluids)*, Journal of Nanoparticle Research, 9, 2007, s. 479-489, https://doi.org/10.1007/s11051-006-9150-1.
13. Sharma D., Rajput J., Kaith B.S., Kaur M., Sharma S., *Synthesis of ZnO nanoparticles and study of their antibacterial and antifungal properties*, Thin Solid Films, 519(3), 2010, s. 1224-1229, https://doi.org/10.1016/j.tsf.2010.08.073.
14. Jiang Y., Zhang L., Wen D., Ding Y., *Role of physical and chemical interactions in the antibacterial behavior of ZnO nanoparticles against E. coli*, Mater. Sci. Eng. C. 69, 2016, s. 1361-1366, https://doi.org/10.1016/J.MSEC.2016.08.044.
15. Esparza-González S.C., Sánchez-Valdés S., Ramírez-Barrón S.N., Loera-Arias M.J., Bernal J., Meléndez-Ortiz H.I., Betancourt-Galindo R., *Effects of different surface modifying agents on the cytotoxic and antimicrobial properties of ZnO nanoparticles*, Toxicol. Vitr., 37, 2016, s. 134-141, https://doi.org/10.1016/j.tiv.2016.09.020.
16. Kokotov M., Hodes G., *Reliable chemical bath deposition of ZnO films with controllable morphology from ethanolamine-based solutions using $KMnO_4$ substrate activation*, Journal of Materials Chemistry, 19, 2009, s. 3847-3854, https://doi.org/10.1039/B821242B.
17. Jackiewicz A., Podgórski A., Gradoń L., Michalski J., *Nanostructured Media to Improve the Performance of Fibrous Filters*, KONA Powder and Particle Journal, 30 (30), 2013, s. 244-255, https://doi.org/10.14356/KONA.2013023.
18. He X., Luo L.S., *Theory of the lattice Boltzmann method: From the Boltzmann equation to the lattice Boltzmann equation*, Phys. Rev. E, 56(6), 1997, 6811. https://doi.org/10.1103/PhysRevE.56.6811.
19. Mohamad A.A., *Lattice Boltzmann Method. Fundamentals and Engineering Applications with Computer Codes*. Springer-Verlag London, ISBN 978-1-4471-7422-6T. Schlick, Molecular Modeling and Simulation. Springer-Verlag New York 2011, ISBN 0-387-95404-X.
20. Chen S., Doolen G.D., *Lattice Boltzmann method for fluid flows*, Ann. Rev. Fluid Mech, 30, 1998, s. 329. DOI: 10.1146/annurev.fluid.30.1.329.
21. Landau L.D., Lifshitz E.M., *Statistical Physics*, Butterworth-Heinemann, Oxford 1980.
22. Bhatnagar P.L., Gross E.P., Krook M., *A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems*, Phys. Rev., 94, 1954, s. 511, DOI: 10.1103/PhysRev.94.511.
23. Schlick T., *Molecular Modeling and Simulation*, Springer-Verlag New York 2010, ISBN 0-387-95404-X.
24. Behkam B., Sitti M., *Bacterial flagella-based propulsion and on/off motion control of microscale objects*, Appl. Phys. Lett., 90, 2007, 023902, https://doi.org/10.1063/1.2431454.

25. Arabagi V., Behkam B., Cheung E., Sitti M., *Modeling of stochastic motion of bacteria propelled spherical microbeads*, Journal of Applied Physics, 109, 2011, 114702, https://doi.org/10.1063/1.3592970.
26. Block S.M., Segall J.E., Berg H.C., *Impulse responses in bacterial chemotaxis*, Cell, 31(1), 1982, s. 215-226, ISSN 0092-8674, https://doi.org/10.1016/0092-8674(82)90421-4.
27. Lee K.W., Liu B.Y.H., *On the Minimum Efficiency and the Most Penetrating Particle Size for Fibrous Filters*, J. Air Pollut. Control Assoc., 30, 1980, s. 377-381, https://doi.org/10.1080/00022470.1980.10464592.
28. Jung C.H., Park H.S., Kim Y.P., *Theoretical study for the most penetrating particle size of dust-loaded fiber filters*, Sep. Purif. Technol., 116, 2013, s. 248-252, https://doi.org/https://doi.org/10.1016/j.seppur.2013.05.045.
29. Thomas D., Contal P., Renaudin V., Penicot P., Leclerc D., Vendel J., *Modelling pressure drop in HEPA filters during dynamic filtration*, J. Aerosol Sci., 30, 1999, s. 235-246, https://doi.org/10.1016/S0021-8502(98)00036-6.
30. Kamiński M., Gac J.M., Sobiec P., Kozikowski P., Jankowski T., *Mixture Aerosols Filtration on Filters with Wide Fibre Diameter Distribution – Comparison with Theoretical and Empirical Models*, Aerosol Air Qual. Res., 22, 2022, 220039, https://doi.org/10.4209/aaqr.220039.

Modelowanie procesu usuwania z powietrza cząstek aerozolowych w nowoczesnych kompozytowych materiałach polipropylen-ZnO

Streszczenie

Cząstki aerozolowe w powietrzu, zarówno abiotyczne (nieożywione, np. pyły) jak i biotyczne (żywe, np. bakterie, wirusy, grzyby), stanowią duże zagrożenie dla zdrowia ludzi i środowiska. W związku z tym, istotne jest opracowywanie efektywnych metod usuwania tych cząstek z gazów. W tym kontekście, coraz większe zainteresowanie budzą nowoczesne kompozytowe materiały na bazie polimerów z dodatkowymi właściwościami, takimi jak zdolność do sorpcji, działanie biobójcze czy fotokatalityczne.

W ramach pracy opracowano model numeryczny opisujący proces usuwania z powietrza cząstek aerozolowych, co pozwoli na zaprojektowanie wysokosprawnego kompozytowego materiału filtracyjnego polipropylen-tlenek cynku i przewidywanie jego działania. Materiał ten będzie miał lepsze właściwości od obecnie dostępnych na rynku rozwiązań, dzięki połączeniu właściwości powierzchniowych i sorpcyjnych. Zastosowanie filtrów z tlenkiem cynku na włóknach może spowolnić lub zatrzymać wzrost form biologicznych na powierzchni, a także rozkładać niebezpieczne dla zdrowia lotne związki organiczne.

Przeanalizowano proces filtracji cząstek aerozolowych abiotycznych i biotycznych na włóknach niezmodyfikowanych i zmodyfikowanych tlenkiem cynku. Stwierdzono, że pokrycie włókien filtracyjnych pręcikami z tlenku cynku poprawia skuteczność filtracji, zwłaszcza dla cząstek najbardziej penetrujących strukturę filtracyjną, tj. o średnicy 200-400 nm. Zauważono też, że skuteczność filtracji zwiększa się wraz z długością pręcików ZnO.

Podsumowując, zastosowanie filtrów modyfikowanych tlenkiem cynku poprawia skuteczność filtracji cząstek abiotycznych, a pręciki ZnO mogą być przydatne w filtracji gazów zawierających cząstki biotyczne, wykorzystując biobójcze właściwości tlenku cynku. Zamodelowane struktury kompozytowe będą wytwarzane, następnie badane pod kątem filtracji, a uzyskane wyniki posłużą do walidacji opracowanego modelu. Wnioski z tej pracy przyczynią się do rozwoju technologii oczyszczania powietrza i poprawy jakości powietrza w życiu codziennym, w przemyśle, budynkach mieszkalnych i środowiskach miejskich.

Słowa kluczowe: modelowanie matematyczne, filtry, materiały kompozytowe, tlenek cynku, cząstki aerozolowe

Modelling the process of aerosol particle removal from the air in modern polypropylene-ZnO composite materials

Abstract

Aerosol particles in the air, both abiotic (non-living, such as dust) and biotic (living, such as bacteria, viruses, fungi), pose a significant threat to human health and the environment. Therefore, it is crucial to develop effective methods for removing these particles. In this context, modern composite materials based on polymers with additional properties such as sorption capacity, biocidal action, or photocatalysis are gaining increasing interest.

As part of this work, a numerical model was developed to describe the process of removing aerosol particles from the air, which will allow for the design of a highly efficient composite filtration material consisting of polypropylene-zinc oxide. This material will have better properties than currently available solutions on the market, thanks to the combination of surface and sorption properties. The use of zinc oxide filters on fibers can slow down or prevent the growth of biological forms on surfaces and also degrade volatile organic compounds that are hazardous to health.

The filtration process of both abiotic and biotic aerosol particles on unmodified and zinc oxide-modified fibers was analyzed. It was found that coating the filtration fibers with zinc oxide nanorods improves the filtration efficiency, especially for particles that most easily penetrate the filtration structure, i.e., those with a diameter of 200-400 nm. It was also observed that the filtration efficiency increases with the length of ZnO nanorods.

In summary, the use of zinc oxide-modified filters improves the filtration efficiency of abiotic particles, and ZnO nanorods can be useful in filtering gases containing biotic particles, leveraging the biocidal properties of zinc oxide. The modeled composite structures will be manufactured and subsequently tested for filtration purposes, and the obtained results will be used to validate the developed model. The conclusions from this work will contribute to the development of air purification technology and the improvement of air quality in everyday life, industry, residential buildings, and urban environments.

Keywords: mathematical modelling, filters, composite materials, zinc oxide, aerosol particles

Kamila Steckiewicz[1]

Wyznaczanie temperatury błysku w układach hamulcowych

1. Wprowadzenie

Tarcie zachodzi na rzeczywistym obszarze kontaktu, który z kolei składa się z oddzielnych plam kontaktu rzeczywistego. Plamy te są ulokowane na powierzchni ciernej nominalnego obszaru kontaktu w sposób nierównomierny i skupione głównie w obszarach konturowych. Wielkość i rozmiar tych ostatnich zależy od falistości powierzchni. Rozmiar rzeczywistych plam kontaktu zależy od mikrogeometrii powierzchni ciernej, obciążenia, termicznych i mechanicznych właściwości materiałów pary ciernej.

Generacja ciepła podczas tarcia zachodzi w obszarach styku wierzchołków chropowatości, które w procesie tarcia zmieniają się i przemieszczają się po powierzchni konturowego i nominalnego obszarów kontaktu. O charakterze zmian i przemieszczeniach rzeczywistych obszarów kontaktu decydują takie parametry, jak temperatura, prędkość poślizgu, obciążenie, fizykochemiczne procesy oraz intensywność zużycia. Na dany czas wyznaczenie temperatury powstałej na skutek tarcia w układach ciernych wykonuje się według następującego schematu:

1. do wyznaczenia pola temperatury oraz temperatury nominalnego obszaru kontaktu stosuje się analityczne lub numeryczne rozwiązania odpowiednich zagadnień przewodnictwa cieplnego z uwzględnieniem generacji ciepła na skutek tarcia;
2. przy obliczeniach temperatury błysku korzysta się z rozwiązania zagadnienia przewodnictwa cieplnego dla półprzestrzeni z poruszającym się obszarem nagrzewania;
3. maksymalną temperaturę układu ciernego przedstawia się w postaci sumy powyższych składowych.

Dokonano przeglądu literaturowego dotyczącego metod analitycznych, numerycznych i doświadczalnych wyznaczenia temperatury rzeczywistego obszaru kontaktu. Następnie zaproponowano model matematyczny pozwalający na wyznaczenie zmiany w czasie hamowania temperatury błysku. Uwzględniono w nim temperaturowe zależności mechanicznych i termicznych właściwości materiałów nakładki i tarczy oraz parametry chropowatości powierzchni tarczy. Jednocześnie zaadaptowano również przybliżony model obliczeniowy do wyznaczenia temperatury błysku pozwalający na ekspresowe jej oszacowanie.

Przeprowadzono porównywalną analizę profili czasowych temperatury błysku dla wybranej pary ciernej typu nakładka-tarcza znalezionych za pomocą dwóch modeli obliczeniowych.

Na koniec sformułowane zostały najważniejsze wyniki z przeglądu literaturowego oraz analizy numerycznej.

[1] kamilasteckiewicz@gmail.com, Wydział Mechaniczny, Politechnika Białostocka, www.wm.pb.edu.pl.

2. Przegląd

W rozpatrywanym przeglądzie zostanie przeanalizowany współczesny stan wiedzy w zakresie metod wyznaczania temperatury powstałej na skutek tarcia w elementach roboczych maszyn, głównie w układach hamulcowych.

Ustalono, że całkowity wzrost temperatury powierzchni ciernej powyżej temperatury otoczenia można przedstawić w postaci sumy temperatury nominalnego obszaru kontaktu i temperatury błysku. Znane analityczne modele do obliczania temperatury nominalnego obszaru kontaktu usystematyzowano i zaprezentowano w monografiach [1-4]. Uwzględniono w nich niedoskonałość kontaktu cieplnego tarcia elementów, ich chłodzenia konwekcyjnego, wrażliwość termiczną stosowanych materiałów oraz ich strukturę kompozytową.

2.1. Temperatura maksymalna

Podczas względnego poślizgu dwóch stykających się ciał w wyniku oporu tarciu następuje przemiana energii mechanicznej w ciepło. Wiadomo, że tarcie i towarzyszące mu procesy, w tym nagrzewanie, koncentrują się w rzeczywistych obszarach kontaktu dwóch ciał [5, 6]. Prawie cała energia rozpraszana w kontaktach ciernych jest przekształcana w ciepło albo w rzeczywistym obszarze kontaktu na stykających się powierzchniach, albo w obrębie kilku mikronów pod tym obszarem [7]. Ciepło generowane na skutek tarcia może powodować znaczny wzrost temperatury pary ciernej, szczególnie w obrębie i w pobliżu rzeczywistego obszaru kontaktu. W typowym styku ślizgowym z tarciem suchym lub smarowaniem granicznym maksymalna temperatura T_{max} występuje w obszarach rzeczywistego kontaktu. Jeśli ślizgające się powierzchnie są gładkie, to obszar, w którym osiągana jest temperatura T_{max}, może być stosunkowo duży (nawet do kilkuset μm długości). W przypadku chropowatych powierzchni stykających się przy niewielkiej ilości mikronierówności, poszczególna długość obszarów kontaktu nie przekracza 10 μm. Wartości temperatury podczas tarcia mogą być bardzo wysokie (w niektórych przypadkach ponad 1000°C), ale utrzymują się przez krótki czas, szczególnie na chropowatych powierzchniach, kiedy czas styku wierzchołków mikronierówności jest bardzo krótki [8].

Podwyższona temperatura może mieć różny wpływ na zachowanie kontaktu ciernego, w tym na takie procesy jak utlenianie, zużycie, degradację smaru i niestabilność termosprężystą [7]. Prace dotyczące analizy i prognozowania temperatury powierzchni ciernych można odnieść do dwóch grup: rozwiązania analityczne otrzymane najczęściej z wykorzystaniem modelu powierzchniowego źródła ciepła lub analizy numeryczne oparte na metodach elementów skończonych (MES), różnic skończonych, transformacji całkowych itp. Metody numeryczne posiadają tę zaletę, że pozwalają uwzględniać większość parametrów wpływających na temperaturę. Niestety metody te wymagają znacznego wysiłku do opracowania poprawnego i kompletnego modelu obliczeniowego, co sprawia, że są one bardzo czasochłonne. Z drugiej strony znane rozwiązania analityczne zwykle nie uwzględniają tak ważnych cech, jak skończone wymiary elementów pary ciernej, chropowatość powierzchni kontaktu czy chłodzenie konwekcyjne.

Ogólnie rzecz biorąc, istnieją trzy czynniki przyczyniające się do wysokich temperatur w pobliżu wierzchołku ślizgającej się chropowatości. Stosunkowo niewielki wpływ na temperaturę okazuje temperatura objętościowa, której poziom najczęściej nie przekracza 100°C. W elemencie poruszającym się względem obszaru kontaktu lub względem,

którego porusza się obszar kontaktu średnia temperatura powierzchni ciernej może być wyższa niż temperatura otoczenia, nawet poza obszarem kontaktu. Spowodowane to jest akumulacją ciepła na powierzchni w wyniku wielokrotnego przechodzenia źródła ciepła tarcia. Zwiększenie temperatury powyżej wartości początkowej T_0 określa się jako nominalny (lub średni) wzrost temperatury kontaktu T_m [9, 10]. Na powierzchni, która nie porusza się względem źródła ciepła tarcia, nominalny wzrost temperatury kontaktu można zdefiniować jako uśrednioną względem zmiennych przestrzennych i czasu średnią temperatur wszystkich punktów w obrębie nominalnej powierzchni kontaktu. Z oddaleniem od powierzchni styku średnia temperatura ciała (ruchomego lub stacjonarnego) zmniejsza się do poziomu temperatury otoczenia lub objętościowej.

Większe konsekwencje mają wysokie temperatury, które występują w miejscach rzeczywistego kontaktu ślizgowego. Gwałtowny wzrost temperatury, który występuje w obszarze kontaktu wierzchołków chropowatości, jest znany jako temperatura błysku T_f [11]. Temperatura błysku może wystąpić zarówno na powierzchni w stosunku, do której źródło ciepła jest nieruchome, jak i na powierzchni poruszającej się względem tego źródła. Maksymalna całkowita temperatura kontaktu T_{max} w danym punkcie nominalnego obszaru kontaktu na powierzchni jednego ze stykających się ciał jest sumą temperatury błysku i temperatury nominalnego obszaru kontaktu [12].

Wkład temperatury błysku, spowodowanej tarciem w rzeczywistym obszarze kontaktu, jest efektem lokalnym, który występuje tylko w obrębie tej ograniczonej powierzchni styku dwóch chropowatości [7, 10]. Do czynników wpływających na poziom T_f należą warunki w strefie kontaktu (prędkość poślizgu, rozmiar i kształt rzeczywistego obszaru styku, poziom chropowatości powierzchni) oraz przewodność cieplna materiałów stykających się chropowatości.

W przeciwieństwie do gwałtownego wzrostu temperatury T_f, zwiększenie średniej temperatury T_m jest konsekwencją efektów przenoszenia ciepła na dużą skalę i zasadniczo nie mają na nią wpływu lokalne warunki w rzeczywistym obszarze kontaktu [10, 12]. Czynnikami wpływającymi na T_m są przewodzenie cieplne i chłodzenie konwekcyjne do otaczającego środowiska [10]. Jednymi z ważniejszych parametrów wpływających na T_m są wymiary (pole) nominalnego obszaru kontaktu oraz współczynnik wymiany ciepła określany oddzielnie dla ruchomych, jak i stacjonarnych ciał.

2.2. Współczynnik podziału ciepła

Ciepło, które jest wytwarzane na powierzchni tarcia, jest rozdzielane między dwa ślizgające się ciała. Charakter takiego podziału analizowano w pracach [11, 13-17]. Stwierdzono, że podział ciepła wygenerowanego podczas tarcia zależy od kształtu i wymiarów stykających się ciał, prędkości poślizgu, warunków chłodzenia konwekcyjnego, a także właściwości termicznych materiałów stykających się ciał [14].

Należy zaznaczyć, że warunek ten jest eksperymentalnie potwierdzony w przypadku gładkich powierzchni stykających się ciał [18]. Najdokładniejszą postać funkcji podziału ciepła uzyskuje się przyrównując temperatury powierzchni ciernych we wszystkich punktach obszaru kontaktu. Dokonało tego kilku badaczy, zwłaszcza w pracach [16, 19-21] opartych na zastosowaniu metod numerycznych. Większość badaczy posługuje się przybliżonym podejściem, które po raz pierwszy zaproponował Blok [11]. Postulował on, że funkcja opisująca podział ciepła jest stałą podczas tarcia i że stałą tą można oszacować, przyrównując maksymalne temperatury osiągane na powierzchniach ciernych obu ciał [11].

Obliczone na podstawie postulatu Bloka wartości współczynnika podziału ciepła i maksymalnych temperatur są zgodne z odpowiednimi rezultatami uzyskanymi za pomocą dokładniejszej analizy funkcji podziału ciepła z wykorzystaniem metod numerycznych [18-20]. Nieznaczna rozbieżność rezultatów otrzymanych z wykorzystaniem tych dwóch podejść występowała przy wysokich liczbach Pécleta.

2.3. Temperatura błysku

Większość analiz temperatury średniej powierzchni ciernych została oparta na pionierskich pracach Bloka [11] i Jaegera [15], w których obaj stosowali metodę powierzchniowego źródła ciepła, szczegółowo opisaną w monografii Carslawa i Jaegera [22]. Metoda ta jest również przydatna do analizy wzrostu temperatury błysku T_f. Rozwiązanie początkowo-brzegowych zagadnień przewodnictwa ciepła za pomocą metod, wykorzystujących powierzchniowe źródła ciepła usytuowanych na powierzchni ciała półograniczonego (półprzestrzeń) uzyskuje się poprzez całkowanie po czasie odpowiedniego rozwiązania dla źródła chwilowego [18, 22].

Wzrost temperatury błysku, szczególnie w ruchomym ciele, jest spowodowany generacją ciepła w małym obszarze kontaktu wierzchołków chropowatości i występuje głównie w bezpośrednim otoczeniu obszaru kontaktu rzeczywistego oraz w cienkiej przypowierzchniowej warstwie poniżej nagrzewanej powierzchni [23, 24]. W rezultacie na wzrost temperatury błysku nie wpływają w znaczącym stopniu wymiary stykających się ciał i stąd można wnioskować, że przyjęcie modelu o nagrzewaniu ograniczonego obszaru na powierzchni półprzestrzeni jest na ogół słuszne. Szczególny przypadek, w którym należy wziąć pod uwagę wymiary ciał, dotyczy naniesienia na powierzchnię cierną cienkiej warstwy lub powłoki ochronnej, których grubość jest tego samego rzędu co charakterystyczny wymiar (długość, promień) obszaru kontaktu. Analizę gwałtownego wzrostu temperatury w powlekanych ciałach przeprowadzono przy użyciu metod transformacji całkowej Fouriera i zaprezentowano w pracach [25-27]. Do przewidzenia maksymalnego wzrostu temperatury błysku w całym zakresie zmiany liczby Pecleta wykorzystano trójwymiarowy model lokalnego nagrzewania półprzestrzeni z naniesioną na jej powierzchnię warstwą ochronną. Eliptyczny obszar nagrzewania porusza się po powierzchni warstwy ze stałą prędkością. Rozwiązanie przedstawiono w postaci bezwymiarowej maksymalnej temperatury jako funkcja bezwymiarowych parametrów, takich jak względna grubość warstwy, rozmiary obszaru kontaktu, właściwości termiczne i liczba Pécleta zarówno materiałów warstwy, jak i podłoża (półprzestrzeni). Wyniki przeprowadzonej analizy numerycznej pokazały, że maksymalna temperatura jest osiągana na powierzchni warstwy i zbliża się do temperatury jednorodnej półprzestrzeni o właściwościach podłoża, jeśli prędkość poślizgu jest bardzo mała lub jeśli warstwa jest bardzo cienka. Oznacza to, że w takich przypadkach warstwa ma niewielki wpływ na temperaturę powierzchni ślizgowej. Jeśli warstwa jest gruba lub jeśli prędkość poślizgu jest wysoka, to temperatura nagrzewanej powierzchni warstwy jest zdominowana przez jej właściwości termiczne. Podobne rezultaty otrzymali inni autorzy, którzy badali temperatury w powlekanych ciałach za pomocą metod numerycznych [28].

Istnieją co najmniej dwa powody, dla których warto stosować założenia o stanach ustalonym lub quasi-ustalonym:

1. zakładając, że nagrzewanie cierne zachodzi w sposób ciągły, temperatura w stanie ustalonym jest najwyższą, jaka może wystąpić podczas poślizgu. Stąd przyjęcie

stanu ustalonego lub quasi-ustalonego będzie skutkowało możliwością oszacowania maksymalnej temperatury powierzchni ciernych;

2. zwykle stan ustalony jest osiągany w bardzo krótkim czasie po rozpoczęciu poślizgu, stąd większość procesu nagrzewania na skutek tarcia ciernego przebiega w stanie ustalonym.

Badania potwierdzające słuszność drugiego założenia przeprowadzili Jaeger [15], Gecim i Winer [29], Bhushan [30] oraz Yevtushenko [31]. W pracach tych oszacowano czas niezbędny do osiągnięcia stanu ustalonego oraz określono warunki, które przy tym mają zostać osiągnięte. Badając temperaturę nieustaloną spowodowaną powierzchniowym źródłem ciepła, stwierdzono, że maksymalna temperatura obszaru nagrzewania szybko zbliża się do wartości w stanie ustalonym. Prędkość takiego zbliżenia zależy od wielkości liczby Pécleta. Maksymalna temperatura osiąga blisko 90% swojej wartości w stanie ustalonym w punkcie odległym od środka obszaru nagrzewania na jedną wartość charakterystycznego rozmiaru źródła ciepła przy $Pe = 1$. Czas wymagany do osiągnięcia stanu ustalonego zmniejsza się wraz ze wzrostem prędkości poślizgu (liczby Pécleta).

W oparciu o wyżej wymienione możemy stwierdzić, że w prawie wszystkich sytuacjach nagrzewania ciernego stany ustalony lub quasi-ustalony są przeważające przy wyznaczeniu wzrostu temperatury błysku. Z innej strony należy zaznaczyć, że w przypadkach krótkotrwałych kontaktów ślizgowych może być konieczne otrzymanie rozwiązań nieustalonych zagadnień cieplnych tarcia. Przykładowo Hou i Komanduri [17] stwierdzili, że w obróbkach wykańczających z zużyciem ściernym czasy kontaktu między ziarnem ściernym a polerowaną lub szlifowaną powierzchnią są tak krótkie, że mogą być wymagane rozwiązania pozwalające oszacować temperaturę w zakresie przejściowym. Odpowiednie rozwiązania otrzymano w pracach [15, 26, 31, 32].

Jeśli pojedyncze lokalne źródło ciepła nie powtarza tej samej ścieżki na powierzchni ciernej i nagrzewane ciało jest wielkogabarytowe, to znaczący wzrost temperatury w nominalnym obszarze kontaktu nie występuje ($T_m \approx 0$). Spowodowane to jest tym, że całe ciepło wytwarzane w obszarze kontaktu rozprasza się poprzez konwekcję do otoczenia. Natomiast, jeżeli źródło ciepła przechodzi wielokrotnie nad tym samym obszarem powierzchni ciernej, to występuje w niej dodatkowy wzrost średniej temperatury nominalnego obszaru kontaktu T_m, ponieważ ciepło tarcia wytwarzane podczas jednego przejścia nie zdąży całkowicie odpłynąć z nagrzewanego obszaru przed kolejnym przejściem źródła ciepła w tym samym miejscu.

Korzystając z terminologii Ashby'ego [9], można stwierdzić, że temperatura nominalnego obszaru kontaktu jest średnią temperaturą powierzchni ciernych, gdyby ciepło tarcia było wprowadzane równomiernie na nominalną powierzchnię (o polu A_a) ciała ruchomego. Zasugerował on, że wygenerowane w nominalnym obszarze kontaktu strumienie ciepła są równomiernie rozłożone w tym obszarze i skierowane do nieruchomego i poruszającego się ciał [9].

W przypadku szybko poruszającego się źródła ciepła, które wielokrotnie przechodzi tę samą ścieżkę (klocek hamulcowy na tarczy) lub w przypadku źródła ciepła oscylującego na tej samej ścieżce kontaktu, strumień ciepła do poruszającego się ciała należy przyłożyć do całej powierzchni ciernej [12].

Przy małej prędkości wygenerowany strumień ciepła jest rozłożony tylko po nominalnej powierzchni kontaktu. Chociaż konwekcja ciepła z powierzchni wolnych do otoczenia ma niewielki wpływ na wzrost temperatury błysku T_f, to na wzrost średniej

temperatury T_m obszaru kontaktu nominalnego mechanizmy poszerzenia ciepła takie jak konwekcja lub przewodność cieplna wpływają znacząco.

Na podstawie założenia Ashby'ego [9] o przewodzeniu ciepła jako dominującego mechanizmu średnią temperaturę nominalnego obszaru kontaktu najczęściej wyznaczają z rozwiązań jednowymiarowych liniowych zagadnień brzegowych przewodzenia cieplnego.

Ilekroć konwekcja z powierzchni ciernej jest znaczna, co często ma miejsce w przypadku ruchomych elementów, rozwiązania oparte na liniowych zagadnieniach przewodnictwa cieplnego są niewystarczające, ponieważ nie uwzględniają one wymiany ciepła zależnej od prędkości poślizgu.

Odmienne podejście zaproponowali Alilat [33] oraz Laraqi [34], którzy przyjęli, że podstawowym trybem wymiany ciepła w układach ciernych jest konwekcja i określili równowagę termiczną między strumieniem ciepła w obszarze kontaktu a ciepłem odprowadzanym przez konwekcję z ruchomej powierzchni styku.

Chociaż wzory do wyznaczenia T_m z powodzeniem używano w analizie termicznej wielu praktycznych układów ciernych [9, 12, 35-37], tym niemniej bardziej precyzyjne rezultaty można otrzymać przy uwzględnieniu obu mechanizmów przekazania ciepła: przewodzenia ciepła, jak i konwekcji. Można by to zrobić rozważając bilans cieplny stykających się ciał, w których ciepło doprowadzane do każdego ciała jest ciepłem tarcia $Q_i = q_i A_r$ w rzeczywistym obszarze kontaktu, a ciepło odprowadzane jest sumą ciepła $Q_{diff,i} = q_{diff,i} A_a$ na skutek przewodzenia z nominalnej powierzchni styku i ciepła konwekcyjnego $Q_{conv,i} = q_{conv,i} A_s$ z powierzchni wolnych ciała.

Przy określaniu maksymalnej temperatury kontaktu w większości przypadków zalecane jest rozpatrzenie wzrostu ustalonej lub quasi-ustalonej temperatury błysku. Zakładając, że nagrzewanie cierne zachodzi w sposób ciągły, temperatura w stanie ustalonym jest największym wzrostem temperatury nominalnej. Natomiast są układy cierne, takie jak przykładowo nakładka-tarcza, w których ze względu na zmieniającą się z czasem hamowania prędkość poślizgu konieczna jest analiza temperatury nieustalonej.

Podczas poślizgu dwóch szorstkich powierzchni, rzeczywisty obszar kontaktu tworzą miejsca styku wierzchołków chropowatości, gdzie następuje wytwarzanie ciepła na skutek tarcia. W klasycznym przypadku pojedynczej chropowatości ślizgającej się po gładkiej, płaskiej powierzchni, do opisu procesu wytwarzania ciepła w rzeczywistym obszarze kontaktu przyjmują model powierzchniowego źródła ciepła poruszającego się po gładkiej powierzchni.

Jeżeli styki chropowatości mają różne pola obszaru kontaktu przy różnych naciskach kontaktowych, to wzrost temperatury błysku może się różnić w każdym rzeczywistym obszarze kontaktu. W rezultacie współczynniki podziału ciepła mogą nie być stałe w różnych obszarach styku chropowatości.

Na podstawie powyższych rozważań przy założeniu braku wzajemnego oddziaływania pól temperatury błysku różnych stykających się chropowatości, zaproponowano następujący schemat do wyznaczenia wzrostu temperatury rzeczywistego obszaru kontaktu chropowatych powierzchni:

1. znaleźć pole i ciśnienie nacisku w rzeczywistym obszarze kontaktu (w ogólnym przypadku z wykorzystaniem metod numerycznych);
2. wyznaczyć wzrost temperatury błysku w każdym obszarze styku chropowatości jako funkcję współczynników podziału ciepła $\alpha_k, k = 1, \dots, n$;

3. znaleźć wzrost średniej temperatury nominalnego obszaru kontaktu na podstawie rozwiązania odpowiedniego zagadnienia cieplnego tarcia dla dwóch ślizgających się ciał;
4. z warunku równości temperatur określić współczynnik podziału ciepła w obszarach styku oddzielnych chropowatości;
5. jeżeli wartości współczynnika podziału ciepła są różne w obszarach styku, to znaleźć średnią jego wartość. Pozwoli to ustalić ilość ciepła idącego na nagrzewanie każdego ciała pary ciernej;
6. oszacować wzrost średniej wzrost temperatury T_m nominalnego obszaru kontaktu, wzrost temperatury błysku T_f i maksymalną temperaturę T_{max} całkowitego rzeczywistego obszaru kontaktu.

Jeżeli obszary styku oddzielnych chropowatości są wystarczająco blisko siebie na tyle, że ich pola temperaturowe oddziaływały wzajemnie, obliczenia przeprowadzone na podstawie opisanego powyżej schematu mogłyby być niedokładne, szczególnie przy wyznaczeniu T_m. W takim przypadku do określenia ciśnień i temperatury w obszarach styku chropowatości można stosować metody numeryczne, zaproponowane podczas tarcia suchego w pracach Gao [38] oraz z uwzględnieniem smarowania Zhu i Hu [39]. Alternatywne podejście do analizy naprężeń kontaktowych i temperatury powierzchni chropowatych przy użyciu analizy fraktalnej zaproponowano w pracach Wang i Komvopoulos [33, 34].

3. Sformułowanie zagadnienia

Temperatura osiągana na jednostkowym rzeczywistym obszarze styku nosi nazwę temperatury błysku [40]. Ilość rzeczywistych obszarów styku zależy jak i od obciążenia, tak i mikrogeometrii powierzchni. W procesie hamowania kontaktujące się obszary kolejno zmieniają się. Czas istnienia rzeczywistego obszaru styku wraz z ciśnieniem, prędkością poślizgu, współczynnikiem tarcia oraz właściwościami termicznymi wyznaczają temperaturę błysku, która rozwija się na danym obszarze.

Do wyznaczenia rzeczywistego ciśnienia na jednostkowym obszarze kontaktu należy znać pole rzeczywistego obszaru styku powierzchni ciernych. W przypadku pary, która tworzy kontakt sprężysty w celu tym może być wykorzystany wzór [41]:

$$A_r = \lambda A_c \,, \tag{1}$$

$$\lambda = \left[\frac{0{,}75\ \pi\ b_0^{\nu/2} r_{av}^{1/2} (1-\mu) P}{2^{\nu/2}\ a_2 h_{max}^{1/2} E\ A_c} \right]^{2\nu/(2\nu+1)} \tag{2}$$

gdzie: P – obciążenie na nominalnym obszarze kontaktu, E – moduł Younga, μ – współczynnik Poissona, A_c – pole konturowego obszaru kontaktu wyznaczane za pomocą wzoru Hertza, r_{av}, h_{max} – odpowiednio uśredniony promień zaokrąglenia i maksymalna wysokość chropowatości, b_0 i ν – parametry krzywej powierzchni odniesienia, a_2 – współczynnik zależny od ν. Podczas obliczeń za pomocą wzorów (1), (2) przyjmuje się, że jeden element pary ciernej posiada znacznie mniejszą chropowatość i większą sztywność w stosunku do drugiego elementu.

W celu wyznaczenia pola rzeczywistego obszaru kontaktu przy plastycznym mechanizmie odkształcenia chropowatości zakłada się, że jeden element pary ciernej posiada znacznie mniejsze twardość, sztywność i chropowatość. Stosuje się następną postać parametru λ ze wzoru:

$$\lambda = \frac{P}{A_c\, C\sigma_Y} \tag{3}$$

gdzie: σ_Y – granica plastyczności materiału z mniejszą twardością po Brinellu HB i chropowatością, C – współczynnik charakteryzujący wpływ kształtu obszaru kontaktu na stan naprężeń. Pokładając we wzorze (3) $c\ \sigma_Y \approx HB$ i ze wzoru (1) otrzymano:

$$A_r = \frac{P}{HB} \tag{4}$$

Wzór (4) jest najczęściej stosowany przy wyznaczeniu pola rzeczywistego obszaru kontaktu w procesie hamowania.

Kolejną ważną charakterystyką służącą do oszacowania temperatury błysku jest uśredniona wartość średnicy d_r rzeczywistego obszaru kontaktu, którą podczas kontaktu plastycznego znajduje się ze wzoru [42]:

$$d_r = 2\sqrt{\frac{2\, r_{av} h_{max}}{\nu}\left(\frac{P}{A_c HB\ b_0}\right)^{1/\nu}} \tag{5}$$

W przypadku ogólnym, kiedy należy uwzględniać wzajemny wpływ pojedynczych źródeł ciepła rozwiązanie zagadnienia cieplnego tarcia sprowadza się do całkowego równania typu Fredholma [43]. Jednak zwykle obszary kontaktu rzeczywistego są oddalone od siebie, co powoduje brak ich wzajemnego oddziaływania.

4. Wyznaczenie temperatury błysku podczas hamowania na podstawie analitycznego rozwiązania zagadnienia cieplnego tarcia

Podczas hamowania oddziaływanie kontaktowe nakładki z tarczą nie występuje na całym nominalnym obszarze kontaktu (określonym wymiarem nakładki), ale w obszarach utworzonych przez tzw. fale chropowatości powstałe na skutek zużycia [40]. Takie fale tworzą konturowy obszar kontaktu o polu powierzchni A_c. Należy zaznaczyć, że sumaryczne pole A_r rzeczywistych obszarów kontaktu zwykle stanowi tylko niewielką część nominalnego pola A_a obszaru kontaktu: $0{,}001A_a \leq A_r \leq 0{,}01A_a$[45]. Pola A_r, A_c, A_a są powiązane pewnymi relacjami do ustalenia, których za pomocą metod teoretycznych przyjmują następne założenia upraszczające [44, 45]:

1. interakcja pomiędzy obszarami styku oddzielnych chropowatości jest pomijalnie mała;
2. chropowatości mają kształt segmentów sferycznych zlokalizowanych na powierzchni twardszego elementu (tarcza) układu hamulcowego. Powierzchnia bardziej miękkiego materiału nakładki jest gładka;
3. przyjmuje się plastyczny mechanizm oddziaływania chropowatości powierzchni tarczy z gładką powierzchnią nakładki;
4. mechanizm odkształcenia chropowatości jest plastyczny. Poślizg pojedynczej chropowatości po gładkiej powierzchni trwa do momentu przejścia jej w stan plastyczny na skutek wzrostu temperatury i zrównania się twardości materiałów tarczy i nakładki. Po czym w kontakt wstępuje inna chropowatość;
5. temperatura chropowatości przed wstąpieniem w kontakt z powierzchnią gładką nakładki jest równa średniej temperaturze T_m nominalnego obszaru kontaktu;
6. falistość powierzchni ciernych nakładki i tarczy jest niewielka ($A_c \approx A_a$).

Właściwości mechaniczne i cieplno-fizyczne materiałów pary ciernej oznaczono dolnym indeksem $i = 1{,}2$, gdzie $i = 1$ odnosi się do materiału tarczy, a $i = 2$ –

materiału nakładki. Z uwzględnieniem powyższych założeń, zmiany z czasem t pola A_r i średnicy d_r wyznaczamy ze wzorów (4) i (5) uwzględniając w nich przebieg czasowy ciśnienia kontaktowego p oraz zależność temperaturową twardości po Brinellu HB materiału nakładki:

$$A_r = \frac{p(t)A_a}{HB_2(t)}, 0 \le t \le t_s, \tag{6}$$

$$d_r(t) = 2\sqrt{\frac{2\,r_{av}h_{max}}{\nu}\left[\frac{p(t)}{HB_2(t)b_0}\right]^{1/\nu}}, 0 \le t \le t_s, \tag{7}$$

gdzie profil czasowy ciśnienia kontaktowego w przypadku ogólnym ma postać:

$$p(t) = p_0 p^*(t), p^*(t) = 1 - e^{-t/t_i}, 0 \le t \le t_s \tag{8}$$

t_s – czas hamowania, t_i – parametr, który charakteryzuje szybkość zwiększenia ciśnienia od zera do wartości nominalnej p_0, $HB_2(t) \equiv HB_2[T_m(t)]$. Parametry chropowatości r_{av}, h_{max}, b_0 i ν wyznaczają za pomocą metodyki opartej na analizie profilogramów powierzchni ciernej tarczy w kierunkach wzdłużnym i poprzecznym [42].

Zmiany z czasem hamowania temperatury wyznaczamy ze wzoru [45]:

$$T_f(t) = \frac{1{,}707\; q(t)A_a d_r(t)}{A_r(t)[4K_1(t)+\sqrt{\pi d_r(t)V(t)K_2(t)c_2(t)\rho_2(t)}}, 0 \le t \le t_s, \tag{9}$$

gdzie:

$$q(t) = fp(t)V(t), 0 \le t \le t_s, \tag{10}$$

$$K_i(t) \equiv K_i[T_m(t)], c_i(t) \equiv c_i[T_m(t)], \rho_i(t) \equiv \rho_i[T_m(t)], i = 1{,}2, \tag{11}$$

a pole A_r i średnicę d_r obszaru kontaktu rzeczywistego wyznaczamy odpowiednio ze wzorów (6) i (7).

Liniową prędkość tarczy V znajdziemy z rozwiązania następnego zagadnienie początkowego dla równania ruchu pojazdu o masie m:

$$m\frac{dV(t)}{dt} = -fp(t)A_a, 0 \le t \le t_s, V(0) = V_0. \tag{12}$$

Rozwiązanie zagadnienia początkowego (12), z uwzględnieniem profilu czasowego ciśnienia (8), ma postać:

$$V(t) = V_0 V^*(t), V^*(t) = 1 - \frac{t}{t_s^0} + p^*(t)\frac{t_i}{t_s^0}, 0 \le t \le t_s \tag{13}$$

Gdzie:

$$t_s^0 = \frac{W_0}{Q_0}, \tag{14}$$

$$Q_0 = q_0 A_a, q_0 = fp_0 V_0, \tag{15}$$

$W_0 = 0{,}5mV_0^2$ – początkowa kinetyczna energia pojazdu.

Analiza wzoru (9) wykazuje, że obliczenia za jego pomocą temperatury błysku zależy od poprawności oszacowania nie tylko właściwości termicznych materiałów i parametrów tarcia takich jak współczynnik tarcia, ciśnienie kontaktowe, prędkość poślizgu, ale i rów-

nież od charakterystyk mikrogeometrii powierzchni ciernej tarczy r_{av}, h_{max}, b_0 i ν. Należy zaznaczyć, że we wzorze (9) uwzględniono także zmiany właściwości termicznych i gęstości materiałów nakładki i tarczy od temperatury. Zależności takie dla wybranych par ciernych otrzymuje się za pomocą badań doświadczalnych.

Ustalono, że zwiększenie czasu narastania t_i ciśnienia kontaktowego we wzorze (8) powoduje obniżenie temperatury średniej T_m oraz temperatury maksymalnej T_{max} [4]. Dlatego w niniejszej pracy analiza numeryczna została przeprowadzona w przypadku hamowania ze stałym opóźnieniem ($t_i \to 0$). Przechodząc we wzorach (8), (10) i (13) do granicy $t_i \to 0$ otrzymano:

$$p(t) = p_0, \tag{16}$$

$$V(t) = V_0\, V^*(t)\,, V^*(t) = 1 - \frac{t}{t_s}, 0 \le t \le t_s, \tag{17}$$

$$q(t) = q_0\, q^*(t)\,, q^*(t) = 1 - \frac{t}{t_s}, 0 \le t \le t_s, \tag{18}$$

gdzie: czas hamowania $t_s = t_s^0$ (14).

Jak wynika ze wzoru (9), właściwości mechaniczne i termiczne materiału chropowatej powierzchni są funkcjami średniej temperatury T_m obszaru kontaktu. Ostatnia może być wyznaczona zarówno za pomocą metod doświadczalnych (termopar) lub w wyniku rozwiązania zagadnienia przewodnictwa cieplnego z odpowiednimi warunkami brzegowymi i początkowymi. W celu tym skorzystamy z dokładnego rozwiązania początkowo-brzegowego zagadnienia przewodnictwa cieplnego z uwzględnieniem generacji ciepła na skutek tarcia dla dwóch półprzestrzeni. W modelu tym przyjęto, że ciśnienie kontaktowe jest stałe i równe p_0 (16) podczas hamowania, prędkość V zmniejsza się liniowo (17), a profil czasowy gęstości mocy tarcia ma postać (18). Założono idealny kontakt cieplny tarcia pomiędzy ślizgającymi się półprzestrzeniami. Wtedy zmiana z czasem hamowania temperatury powierzchni ciernej ma postać [46]:

$$T_m(t) = T_0 + T_m^*(t), 0 \le t \le t_s, \tag{19}$$

$$T_m^*(t) = \frac{2q_0k_1^0}{K_2^0\sqrt{k_1^0}+K_1^0\sqrt{k_2^0}}\sqrt{\frac{t}{\pi}}\left(1 - \frac{2t}{3t_s}\right), 0 \le t \le t_s, \tag{20}$$

gdzie: K_i^0, k_i^0, $i = 1,2$ są to odpowiednio współczynniki przewodności cieplnej i dyfuzji cieplnej przy temperaturze początkowej T_0. Rozwiązanie (19), (20) będzie wykorzystane do wyznaczenia temperaturowej zależności właściwości materiałów we wzorach (11).

5. Wyznaczenie temperatury błysku z wykorzystaniem wzorów aproksymacyjnych

Eksperymentalne badania potwierdzają wzajemny związek pomiędzy średnią temperaturą T_m powierzchni ciernej oraz temperaturą błysku T_f [47]. W związku z powyższym aktualnym jest ustalenie funkcyjnej zależności $T_f = F(T_m)$. Analiza otrzymanych danych za pomocą badań doświadczalnych i obliczeń numerycznych wykazała, że:

1. zwiększenie średniej temperatury T_m powierzchni ciernej do wartości temperatury topnienia materiału powoduje obniżenie T_f do zera;

2. dla rozpatrywanej pary ciernej przy $T_m = T_0$ temperatura błysku T_f osiąga wartość maksymalną.

Tym dwóm wymogom odpowiada empiryczna ekspotencjalna zależność [5]:

$$T_f(t) = B(t)e^{-C(t)T_m(t)}, 0 \le t \le t_s \tag{21}$$

gdzie:

$$B(t) = \sqrt[3]{\frac{p(t)V(t)}{m_1}}, C(t) = \frac{m_2}{V(t)} + m_3. \tag{22}$$

Zmiany ciśnienia kontaktowego p i prędkości poślizgu V ustalano odpowiednio ze wzorów (16) i (17), a m_k, $k = 1,2,3$ są to współczynniki wyznaczane doświadczalnie dla wybranej pary ciernej [45-47].

6. Analiza numeryczna

Obliczenia wykonano dla hamulca tarczowego przy $A_a = 39{,}57$ cm^2, $f = 0{,}5$, $p_0 = 1{,}47$ MPa, $T_0 = 20^o$C, $V_0 = 8{,}5$ ms^{-1}, $W_0 = 98$ kJ [21]. Na podstawie tych danych ze wzorów wyznaczono: $q_0 = 6{,}25$ MWm^{-2}, $Q_0 = 24\,731$ W, $t_s = t_s^0 = 3{,}96$ s. Materiał cierny nakładek – metalo-ceramikę FMC-11 skojarzono z żeliwną tarczą hamulcową. Gęstości tych materiałów nie zmieniają się pod wpływem temperatury i wynoszą: 4700 kg m^{-3} (FMC-11) i 7100 kg m^{-3} (żeliwo).

Metalo-ceramiczny materiał cierny FMC-11 otrzymuje się metodą spiekanych proszków. Metaliczną osnowę tego materiału stanowi żelazo, które nadaje materiałowi wysoką przewodność cieplną. Natomiast niemetaliczne dodatki służą do zwiększenia współczynnika tarcia, obniżenia wytrzymałościowych właściwości i zwiększenie kruchości. Dlatego elementy cierne układów hamulcowych wykonuje się z bimetalowego materiału składającego się ze stalowej podstawy i warstwy ciernej. Skład chemiczny wybranego materiału ciernego jest następujący [5]:

FMC-11 − 64% Fe, 15% Cu, 9% C, 3% SiO_2, 3% azbest, 6% $BaSO_4$.

Zależności właściwości mechanicznych i termicznych wybranych materiałów od temperatury zapisano w postaci [5]:

$$K_i(T) = K_i^0 K_i^*(T), c_i(T) = c_i^0 c_i^*(T), \tag{23}$$

$$K_i^0 \equiv K_i(T_0), c_i^0 \equiv c_i(T_0), \tag{24}$$

$$K_i^*(T) = K_i^1 + \frac{K_i^2}{\left[(T-T_{K_1})K_i^3\right]^2+1} + \frac{K_i^4}{\left[(T-T_{K_2})K_i^5\right]^2+1}, \tag{25}$$

$$c_i^*(T) = c_i^1 + \frac{c_i^2}{\left[(T-T_{c_1})c_i^3\right]^2+1} + \frac{c_i^4}{\left[(T-T_{c_2})c_i^5\right]^2+1}, \tag{26}$$

$$HB_i(T) = HB_i^0 HB_i^*(T), \tag{27}$$

$$HB_i^*(T) = HB_i^1 + \frac{HB_i^2}{\left[(T-T_{HB_1})HB_i^3\right]^2+1} + \frac{HB_i^4}{\left[(T-T_{HB_2})HB_i^5\right]^2+1}, \tag{28}$$

gdzie: wartości współczynników aproksymacyjnych K_i^m, c_i^m, HB_i^m, $i = 1,2, m = 1,2,3,4,5$, T_{K_j}, T_{c_j}, T_{HB_j}, $j = 1,2$ są zawarte w tabelach 3.1-3.3 [5].

Tabela 3.1. Współczynniki aproksymacyjne funkcji K_i^*

K_i^*	żeliwo $i = 1$	FMC-11 $i = 2$
K^0, Wm^{-1} °C^{-1}	52,167	35,005
K^1	-2,3706	1,125
K^2	4,223	-0,6393
K^3, °C^{-1}	0,196 × 10^{-3}	2,3 × 10^{-3}
T_{K_1}, °C	-2543	900
K^4	0	0
K^5, °C^{-1}	0	0
T_{K_2}, °C	0	0

Źródło: opracowanie własne na podstawie [3].

Tabela 3.2. Współczynniki aproksymacyjne funkcji c_i^*

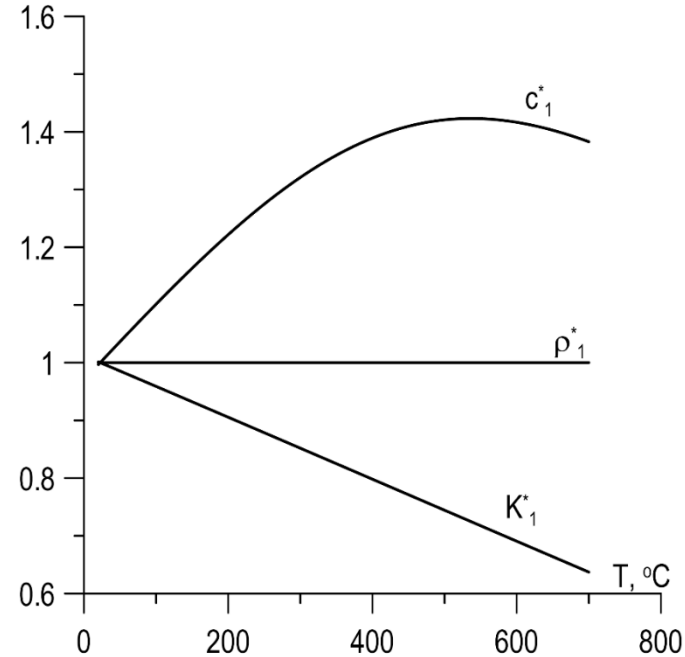
c_i^*	żeliwo $i = 1$	FMC-11 $i = 2$
c^0, J kg^{-1} °C^{-1}	444,6	478,9
c^1	-0,8461	0,7748
c^2	6,598	0,7385
c^3, °C^{-1}	0,571 × 10^{-3}	3,5 × 10^{-3}
T_{c_1}, °C	4903	1059
c^4	1,372	0,4992
c^5, °C^{-1}	1,2 × 10^{-3}	2,6 × 10^{-3}
T_{c_2}, °C	443	573

Źródło: opracowanie własne na podstawie [25].

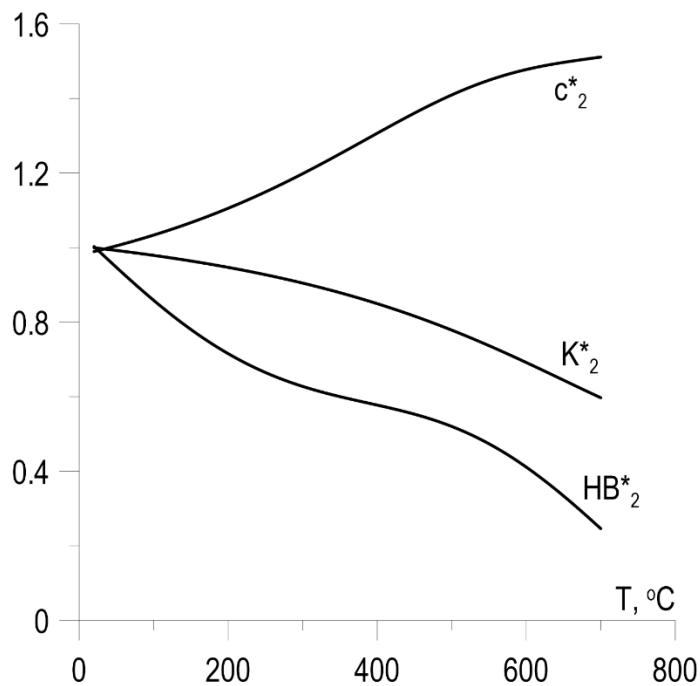
Tabela 3.3. Współczynniki aproksymacyjne funkcji HB_2^*

HB_2^*	FMC-11
HB^0, MPa	1373
HB^1	-0,926
HB^2	0,823
HB^3, °C^{-1}	2,34 × 10^{-3}
T_{HB_1}, °C	546
HB^4	2,02
HB^5, °C^{-1}	2,02 × 10^{-3}
T_{HB_2}, °C	-233

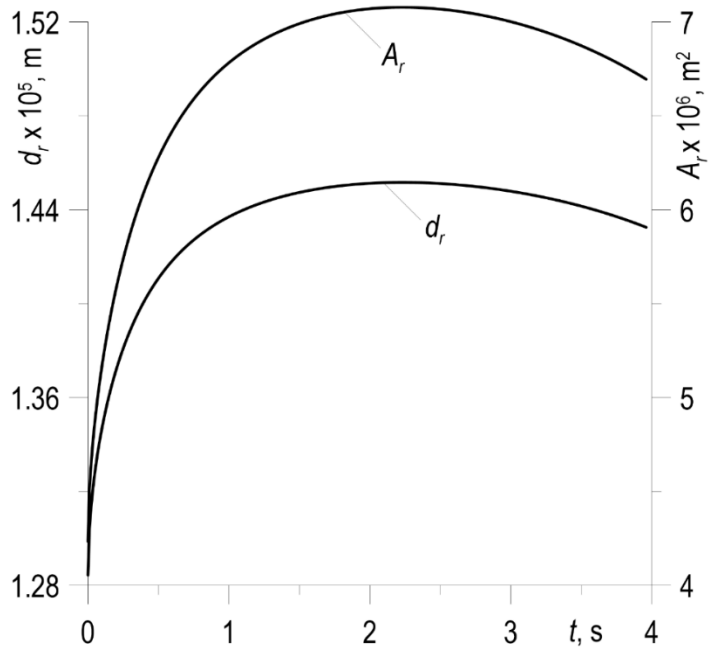
Źródło: opracowanie własne na podstawie [7].



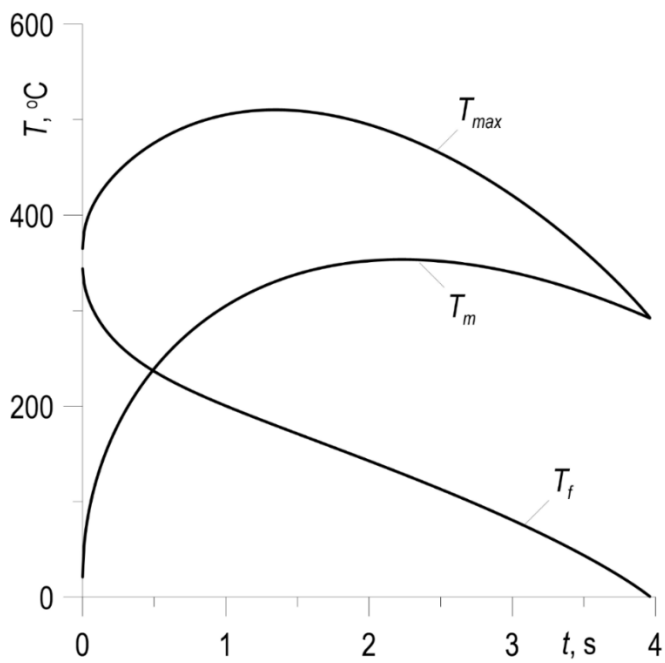
Rysunek 1. Bezwymiarowa zmiana właściwości żeliwa ChNMKh od temperatury



Rysunek 2. Bezwymiarowa zmiana właściwości FMC-11 od temperatury



Rysunek 3. Zmiana pola i średnicy rzeczywistego obszaru kontaktu podczas hamowania dla pary ChNMKh - FMC-11



Rysunek 4. Zmiany podczas hamowania temperatury średniej T_m, temperatury błysku T_f i temperatury maksymalnej T_{max} dla pary ciernej ChNMKh - FMC-11 przy $t_i = 0{,}5s$

Analizując rezultaty obliczeń numerycznych zaprezentowanych na rysunkach 1-4 należy zaznaczyć, że temperatura średnia T_m obszaru kontaktu nakładki z tarczą szybko wzrasta wraz z rozpoczęciem hamowania, osiągając swoją wartość najwyższą około połowy czasu hamowania. Po osiągnięciu tej wartości zaczyna się nieznaczne ochłodzenie powierzchni ciernych trwające do chwili zatrzymania. Przebieg czasowy temperatury błysku T_f jest inny. Temperatura ta osiąga najwyższe wartości od razu po rozpoczęciu hamowania i zmniejsza się do zera w chwili zatrzymania. Takie zachowanie T_f spowodowane jest przyjętym plastycznym mechanizmem odkształcenia chropowatości. Największa chropowatość powierzchni ciernych nakładki i tarczy ma miejsce w początkowych chwilach hamowania i zmniejsza się na skutek odkształcenia plastycznego wraz ze wzrostem temperatury średniej T_m. Profile czasowe temperatury średniej T_m i błysku T_f okazują znaczący wpływ na ewolucję temperatury maksymalnej T_{max}. W odróżnieniu od temperatury średniej T_m, temperatura maksymalna T_{max} osiąga wartość najwyższą bliżej chwili rozpoczęcia procesu hamowania. Spowodowane jest to wysokimi wartościami temperatury błysku w tym przedziale czasowym. W chwili zatrzymania temperatura maksymalna T_{max} jest praktycznie taka sama, jak i temperatura średnia T_m.

7. Podsumowanie

Podczas oceny procesów cieplnych w obszarze kontaktu trących się ciał należy rozróżniać średnią objętościową temperaturę tych ciał T_v, średnią powierzchniową temperaturę T_m, temperaturę błysku T_f i maksymalną temperaturę powierzchni ciernej T_{max}.

Za objętościową temperaturę przyjmuje się uśrednioną w czasie hamowania wartość temperatury w objętości jednego z elementów pary ciernej. Średnia powierzchniowa temperatura jest to temperatura powierzchni ciernej znaleziona z rozwiązania jednowymiarowego zagadnienia cieplnego tarcia. Temperaturowy błysk powstaje na pojedynczych plamach faktycznego kontaktu i może osiągać wartość temperatury topnienia jednego z elementów pary ciernej. Czas istnienia błysku temperaturowego zależy od prędkości poślizgu i równa się średniej od 10^{-3} do 10^{-8} sekund. Maksymalną temperaturę powierzchni ciernej z wystarczającym stopniem dokładności przedstawia się w postaci sumy: $T_{max} = T_m + T_f + T_v$. Pole obszaru kontaktu rzeczywistego stanowi około 0,01÷0,1% pola nominalnego obszaru kontaktu. Stąd wynika, że całe ciepło generowane na skutek tarcia będzie skoncentrowane na małych co do rozmiarów, krótkotrwałych plamach kontaktu rzeczywistego. Intensywność strumieni ciepła generowanych na tych plamach może być na tyle duża, że zwiększenie temperatury może osiągać setki stopni. Wymiary i czas pojawienia się błysków temperaturowych mają charakter przypadkowy, co utrudnia ich ocenę doświadczalną.

W wyniku przeprowadzonej analizy dotychczasowych osiągnięć naukowych w tematyce wyznaczenia temperatury błysku można stwierdzić, że:

1. temperatura błysku T_f w największej mierze zależy od ciśnienia, prędkości poślizgu i chropowatości powierzchni ciernych;
2. najczęściej do analitycznego obliczenia temperatury średniej stosuje się układ cierny pierścień-tarcza, a do temperatury błysku układ trzpień-tarcza;
3. otrzymano analityczne wzory do wyznaczenia temperatury błysku w warunkach nagrzewania ustalonego. Natomiast nie analizowano przebiegu czasowego temperatury błysku w przypadku nagrzewania nieustalonego.

Do przeprowadzenia analizy numerycznej uwzględniającej zmiany z czasem nagrzewania temperatury błysku zaimplementowano dokładne i przybliżone rozwiązania A.V. Chichinadze [66]. Rozwiązanie to uwzględnia profile czasowe gęstości mocy tarcia, średnice i pola rzeczywistego obszaru kontaktu, prędkości poślizgu oraz zależności temperaturowe właściwości cieplno-fizycznych i mechanicznych od temperatury średniej powierzchni kontaktu. Zależności temperaturowe ciepła właściwego, przewodności cieplnej i twardości materiałów odpowiadające danym doświadczalnym, zapisano w postaci wzorów aproksymujących.

W wyniku przeprowadzonej analizy numerycznej ustalono, że:

1. zmiany z czasem hamowania średnicy oraz pola rzeczywistego obszaru kontaktu wybranych par ciernych są różne. W przypadku nakładki FMC-11 oba te parametry zwiększają się wraz z rozpoczęciem hamowania, osiągają swoją wartość maksymalną w połowie drogi hamowania, a następnie zmniejszają się w chwili zatrzymania;
2. twardość materiału nakładki FMC-11 szybko zmniejsza się wraz z rozpoczęciem hamowania, a następnie osiąga swoją wartość ustaloną nieznacznie zwiększając się do chwili zatrzymania;
3. temperatura błysku osiąga największe wartości w początkowej chwili hamowania, a następnie obniża się aż do chwili zatrzymania. Wpływ temperatury błysku na temperaturę maksymalną jest znaczący przy korzystaniu z rozwiązania dokładnego. Natomiast przy stosowaniu rozwiązania przybliżonego wpływ ten jest znacznie mniejszy. Należy zaznaczyć, że rozwiązanie przybliżone pozwala odtworzyć tylko charakter profilu czasowego temperatury błysku. Natomiast wartości maksymalne temperatury błysku otrzymane na podstawie rozwiązania dokładnego i przybliżonego różnią się znacząco (około 120^{o}C);
4. wykorzystanie rozwiązania dokładnego wymaga przeprowadzenia badań doświadczalnych parametrów chropowatości tarczy. Natomiast w rozwiązaniu przybliżonym takie parametry nie występują.

Literatura

1. Adamowicz A., Grześ P., Jewtuszenko O., Kuciej M., Och E., *Analityczne i numeryczne modelowanie nieustalonej generacji ciepła w elementach tarciowych układów hamulcowych,* Pod red. Jewtuszenko O., Białystok: Oficyna Wydawnicza Politechniki Białostockiej, 2014.
2. Jewtuszenko O., Kuciej M., Och E., *Nieliniowe modele nieustalonego nagrzewania tarciowego w układach hamulcowych*, Białystok: Oficyna Wydawnicza Politechniki Białostockiej, 2018.
3. Jewtuszenko O., Kuciej M., Tołstoj-Sienkiewicz J., *Nieustalone i quasi-ustalone tarciowe pola temperatury*, Białystok: Oficyna Wydawnicza Politechniki Białostockiej, 2015.
4. Kuciej M., *Analityczne modele nieustalonego nagrzewania tarciowego*, Białystok: Oficyna Wydawnicza Politechniki Białostockiej, 2012.
5. Grześ P., *Sprzężone modele numeryczne generacji ciepła w hamulcach tarczowych*, Białystok: Oficyna Wydawnicza Politechniki Białostockiej, 2019.
6. Kennedy F.E., Tian X., *Modeling sliding contact temperatures including effects of surface roughness and convection*, Trans. ASME. J. Tribology, 2016, 138 (4), 042101.
7. Kennedy F.E., *Thermal and Thermomechanical Effects in Dry Sliding*, Wear, 1984, s. 453-476.
8. Quinn T.F.J., Winer W.O., *The Thermal Aspects of Oxidational Wear*, Wear, 1985, s. 102, 67-80.

9. Ashby M.F., Abulawi J., Kong H.S., *Temperature Maps for Frictional Heating in Dry Sliding*, Tribology Trans. 34, 1991, s. 577-587.
10. Barber J., *The Conduction of Heat from Sliding Solids*, Int. J. Heat Mass Transfer., 1970, s. 857-869.
11. Blok H.A., *Theoretical Study of Temperature Rise at Surfaces of Actual Contact under Oiliness Lubricating Conditions, Proc. General Discussion on Lubrication and Lubricants*, I.Mech.E., London 1937, s. 222-235.
12. Tian X., Kennedy F.E., *Contact Surface Temperature Models for Finite Bodies in Dry and Boundary Lubricated Sliding*, *ASME J. Tribol.*, 1993, s. 115, 411-418.
13. Barber J.R., *The Distribution of Heat Between Sliding Surfaces*, J. Mech. Engng Sci., 1967, s. 351354.
14. Berry G.A., Barber, J.R., *The Division of Frictional Heat - A Guide to the Nature of Sliding Contact*, ASME J. Tribol., 1984, s. 106, 405-415.
15. Jaeger J.C., *Moving Sources of Heat and the Temperature of Sliding Contacts*, Proc. R. Soc. N.S.W., 1942, s. 203-224.
16. Komanduri R., Hou Z.B., *Analysis of Heat Partition and Temperature Distribution in Sliding Systems*, Wear, 2001, s. 925-938.
17. Rashid, M., Seireg, A., *Heat Partition and Transient Temperature Distribution in Layered Concentrated Contacts. Part I and II*, ASME J. Tribology, 1987, s. 487-502.
18. Ling F.F., Lai W.M., Lucca D.A., *Fundamentals of Surface Mechanics*, Springer, New York, 2002.
19. Bansal D.M., Streator J.L., *On Estimations of Maximum and Average Interface Temperature Rise in Sliding Elliptical Contacts*, Wear, 2012, s. 18-27.
20. Bos J., Moes H., *Frictional Heating of Tribological Contacts*, ASME J. Tribol, 1995, s. 171-177.
21. Ling F.F., *A Quasi-Iterative Method for Computing Interface Temperature Distributions*, Z Agnew. Math Physik, 1959, s. 461-474.
22. Carslaw H.S., Jaeger J.C., *Conduction of Heat in Solids, 2nd Ed.*, Clarendon Press, Oxford, 1959.
23. Barber J.R., *An Asymptotic Solution for Short-Time Transient Heat Conduction Between Two Similar Contacting Bodies*, Int. J. Heat Mass Transfer, 1989, s. 943-949.
24. Gecim B., Winer W.O., *Steady Temperature in a Rotating Cylinder Subject to Surface Heating and Convective Cooling*, ASME J. Tribology, 1984, s. 120-126.
25. Ju F.D., Liu J.C., *Parameters Affecting Thermo-Mechanical Cracking in Coated Media Due to High-Speed Friction Load*, ASME J. Tribology, 1988, s. 222-227.
26. Leroy J.M., Floquet A., Villechaise B., *Thermomechanical Behavior of Multilayered Media: Theory*, ASME J. Tribology, 1989, s. 538-544.
27. Ling F.F., Yang C.F., *Surface Temperatures of Moving Layered Composites*, ASME, New York, 1969, s. 164-176.
28. Gecim B., Winer W.O., *Effect of Surface Film on the Surface Temperature of a Rotating Cylinder*, ASME J. Tribology, 1986, s. 92-97.
29. Gecim B., Winer W.O., *Transient Temperatures in the Vicinity of an Asperity Contact*, ASME J. Tribol., 1985, s. 107, 333-342.
30. Bhushan B., *Magnetic Head-Media Interface Temperatures – Part 1: Analysis*, ASME J. Tribol, 1987, s. 243-251.
31. Yevtushenko A., Ivanyk E.G., Ukhanska O.M., *Transient Temperature of Local Moving Areas of Sliding Contact*, Tribology Int'l. 30, 1997, s. 209-214.
32. Hou Z-B., Komanduri R., *Magnetic Field Assisted Finishing of Ceramics – Part 1: Thermal Model*, ASME J. Tribol, 1998, s. 645-651.

33. Wang S., Komvopoulos K., *A Fractal Theory of the Jnterfacial Temperature Distribution in the Slow Sliding Regime: Part II-Multiple Domains, Elastoplastic Contacts and Applications*, ASME J. Tribology, 1994, s. 824-832.
34. Wang, S., Komvopoulos, K., *A Fractal Theory of the Temperature Distribution at Elastic Contacts of Fast Sliding Surfaces*, ASME J. Tribology, 1995, s. 203-214.
35. Alilat N, Bairi A., Laraqi N., *Three-Dimensional Calculation of Temperature in a Rotating Disk Subjected to an Eccentric Circular Heat Source and Surface Cooling*, Numerical Heat Transfer, 2004, s. 167-180.
36. Kennedy F.E., *Determination of Contact Temperatures Resulting from Frictional Heating*, Proc. Int'l. Trib. Conf. - Nagasaki, Japan, 2000, s. 313-318
37. Laraqi N., Alilat N., Garcia de Maria J.M., Bairi, A., *Temperature and Division of Heat in a Pin-on-Disc Frictional Device – Exact Analytical Solution*, Wear, 2009, s. 765-770.
38. Gao J., Lee S.C., Ai X., Nixon H., *An FFT-Based Transient Flash Temperature Model for General Three-Dimensional Rough Surface Contacts*, ASME J. Tribology, 2000, s. 519-523.
39. Zhu D., Hu Y-Z., *Computer Program Package for the Prediction of EHL and Mixed Lubrication Characteristics, Friction, Subsurface Stresses and Flash Temperatures Based on Measured 3-D Surface Roughness*, Tribology Transactions, 2001, s. 383-390.
40. Kragelsky I.V., Dobychin, M.N., Kombalov V.S., *Friction and wear: calculation methods*, Oxford: Pergamon, 1982.
41. Kragelsky I.V., Demkin N.B., *Contact area of rough surfaces*, Wear 3.3, 1960.
42. Demkin N. B., *Contact of rough surfaces*, In Russian, Moscow: Nauka, 1940.
43. Awrejcewicz J., Pyr'yev Yu., *Nonsmooth dynamics of contacting thermoelastic bodies*, Springer-Verlag, New York, 2009.
44. Chichinadze A.V., *Polymers in friction assemblies of machines and devices: A Handbook*, New York: Allerton Press Inc., 1984.
45. Chichinadze A.V., Braun E.D. Ginsburg A.G., Ignat'eva E.V., *Calculation, testing and selection of frictional couples, In Russian*, Moscow: Nauka, 1979.
46. Mamchegov M.A., *Determination of the maximum temperature of the sliding contact*, Mashinostroeniye, 1977, s. 107-112.
47. Mamchegov M.A., Zinovyeva Z.V., Ginzburg A.G., *Questions of calculation of the maximum temperature at non-stationary friction with intensive heat generation. In: Friction and wear*, 1977, s. 46-53 (in Russian).

Wyznaczanie temperatury błysku w układach ciernych

Streszczenie
Celem pracy było opracowanie teoretycznej metody wyznaczenia temperatury rzeczywistego obszaru kontaktu (temperatury błysku) w układach ciernych. Generacja ciepła na skutek tarcia zachodzi w obszarach styku wierzchołków chropowatości powierzchni ciernych. O intensywności wytwarzania ciepła decydują takie parametry jak obciążenie zewnętrzne, prędkość poślizgu, właściwości mechaniczne i termiczne materiałów oraz intensywność zużycia. Opracowano model teoretyczny na podstawie rozwiązania A.V. Chichinadze. Uwzględnia on zmianę z czasem prędkości poślizgu i ciśnienia kontaktowego oraz zależności właściwości mechanicznych i termicznych od temperatury. Analizę numeryczną temperatury błysku przeprowadzono dla wybranej pary ciernej. Opracowano również podejście aproksymacyjne do wyznaczenie temperatury błysku. Na podstawie otrzymanych rezultatów opracowano wnioski wykazujące znaczący wpływ temperatury błysku na temperaturę maksymalną.
Słowa kluczowe: temperatura, tarcie, temperatura błysku, hamowanie, chropowatość powierzchni

Determination of flash temperature in friction systems

Abstract
The aim of the work was to develop a theoretical method for determining the temperature of the actual contact area (flash temperature) in friction systems. The generation of heat due to friction takes place in the contact areas of the peaks of the roughness of the friction surfaces. The intensity of heat generation is determined by parameters such as external load, sliding speed, mechanical and thermal properties of materials, and wear intensity. A theoretical model was developed based on the solution of A.V. Chichinadze. It takes into account the change over time of the sliding speed and contact pressure as well as the dependence of mechanical and thermal properties on temperature. The numerical analysis of the flash temperature was carried out for the selected friction pair. An approximation approach to determine the flash temperature was also developed. On the basis of the obtained results, conclusions were drawn showing a significant impact of the flash temperature on the maximum temperature.
Keywords: temperature, friction, flash temperature, braking, surface roughness

Maciej Słowik[1], Małgorzata Gulewicz[2]

Symulacyjna i eksperymentalna estymacja pozycji roweru – przykład projektu B+R SmartBrake

1. Wprowadzenie

Niniejsza praca dotyczy estymacji pozycji roweru realizowanych w ramach projektu badawczo-rozwojowego nr POIR.01.01.01-00-1056/17 pod tytułem „System SmartBrake – Inteligentny System Asekuracji Ruchu i Trasy Roweru". Projekt realizowany był w okresie od 1 listopada 2018 do 29 stycznia 2021. Przeprowadzone prace badawczo-rozwojowe zostały podzielone na cztery etapy. Pierwsze trzy etapy należały do kategorii badań przemysłowych, etap IV należał do grupy prac rozwojowych. Każdy z etapów realizowany był wspólnie przez Wnioskodawcę – firmę Moose spółka z ograniczoną odpowiedzialnością oraz Podwykonawcę – zespół z Wydziału Mechanicznego Politechniki Białostockiej.

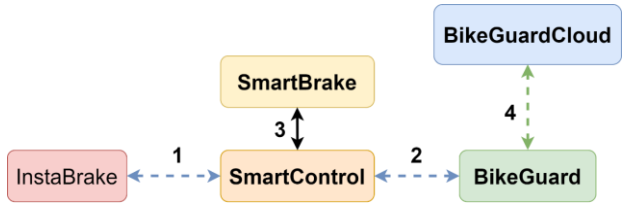
Celem projektu było opracowanie oraz implementacja innowacyjnego rozwiązania w postaci modułowego systemu SmartBrake – Inteligentnego Systemu Asekuracji Ruchu i Trasy Roweru, monitorującego oraz zwiększającego bezpieczeństwo jazdy na rowerze dla dzieci i młodzieży. System SmartBrake składa się z następujących modułów:

1. InstaBrake – moduł zdalnego sterowania, służący do wyzwalania hamowania przez rodzica, połączony bezprzewodowo z modułem SmartControl oraz telefonem z zainstalowaną aplikacją BikeGuard.
2. SmartBrake – moduł wykonawczy systemu – połączony przewodowo oraz zasilany z modułu SmartControl sterującego jego pracą.
3. SmartControl – moduł sterujący pracą hamulca SmartBrake. Połączony bezprzewodowo za pomocą modułu InstaBrake z aplikacją BikeGuard na urządzeniu mobilnym. Wyposażony w odbiornik GNSS oraz układ inercjalnej jednostki pomiarowej.
4. Aplikacja na urządzenia mobilne – BikeGuard przeznaczona do konfiguracji pracy systemu, nadzoru jazdy w trybie okręgu oraz trybie podwórka. Połączona z modułem SmartControl oraz aplikacją chmurową BikeGuardCloud.
5. Aplikacja chmurowa BikeGuardCloud – aplikacja umożliwiająca wyświetlanie historii oraz podgląd aktualnej pozycji modułu SmartControl.

Współdziałanie powyższych modułów przedstawiono na rysunku 1. Cyframi 1 i 2 oznaczono komunikację bliskiego zasięgu realizowaną za pomocą technologii Bluetooth. Przewodowe połączenie pomiędzy modułem SmartControl a SmartBrake oznaczono cyfrą 3. Komunikacja pomiędzy modułami BikeGuard i BikeGuardCloud realizowana jest natomiast za pomocą technologii GSM/LTE.

W okresie przygotowania do realizacji projektu, zespół badawczy z firmy Moose nie dotarł do informacji o rozwiązaniach z tożsamymi funkcjonalnościami. Rozwiązania o częściowo zbliżonych funkcjonalnościach to prototyp MiniBrake, umożliwiający zdalne hamowanie rowerka dziecięcego wyzwalane z pilota [1], bez dedykowanej aplikacji mobilnej i webowej, estymacji pozycji oraz możliwości zmniejszania prędkości do określonego przez rodzica poziomu.

[1] maciej.slowik@mooseinc.eu, Moose spółka z ograniczoną odpowiedzialnością.
[2] malgorzata.gulewicz@sd.pb.edu.pl, Wydział Inżynierii Zarządzania, Politechnika Białostocka.



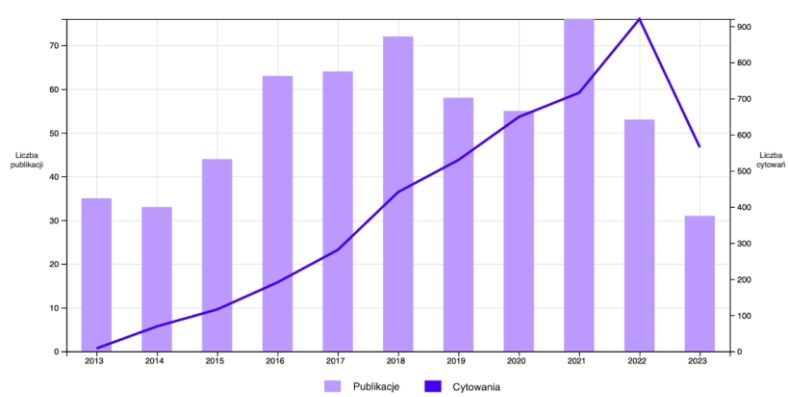
Rysunek 1. Schemat komunikacji pomiędzy poszczególnymi modułami systemu SmartBrake [opracowanie własne]

Dzięki synergicznemu połączeniu powyższych modułów zapewniono możliwość wymuszenia zatrzymania roweru dziecka przez rodzica. W przypadku wystąpienia zagrożenia opiekun dziecka mógł wywołać akcję hamowania roweru, aż do jego całkowitego zatrzymania.

Drugą funkcjonalnością jest przypadek użycia związany z ograniczeniem dystansu na który może się oddalić dziecko od rodzica. Zostały tu wykorzystane techniki tzw. geofencingu wykorzystujące lokalizację zarówno opiekuna, jak i podopiecznego. Geofencing wg Wielkiego Słownika Języka Polskiego to *wyznaczanie wirtualnych granic określonych obszarów z zastosowaniem nawigacji satelitarnej i monitorowanie ruchu obiektów takich jak pojazdy, pojawiających się w tych strefach lub opuszczających je* [2]. W celu określenia jak najdokładniejszej lokalizacji rodzica oraz dziecka wykorzystano techniki fuzji sygnałów pochodzących z modułów sensorycznych wchodzących w skład modułu SmartControl, zamontowanego na rowerze dziecka. Kolejną funkcjonalnością jest podgląd pozycji dziecka przez rodzica za pomocą aplikacji mobilnej oraz webowej.

1.1. Zagadnienia związane z lokalizacją

Problem określenia lokalizacji – estymacji pozycji jest aktualnym problemem badawczym. Ma to związek m.in. z rozwojem autonomicznych pojazdów oraz usług wykorzystujących estymację pozycji związanych z pojazdami. Przykładem tego typu zastosowań są platformy umożliwiające przewóz osób np. Bolt lub Uber. Inną grupa publikacji związana jest z użyciem nowych sposobów lokalizacji np. z wykorzystaniem technik UWB (ang. Ultra Wide Band). Na rysunku 2 przedstawiono liczbę publikacji dla frazy „position estimation" and „vehicle" na portalu Web of Science. Zapytanie obejmowało okres od 1.1.2013 do 22.8.2023. Liczba wyników wynosi 584.



Rysunek 2. Liczba publikacji z frazą „position estimation" and „vehicle" w tytule – wyniki kwerendy w Web of Science [opracowanie własne]

W tabeli 1 przedstawiono grupy tematyczne, w których zawarta jest największa liczba publikacji (433) wynoszącą ponad 74% całości.

Tabela 1. Grupy wg podziału „Citation Topics Meso" [opracowanie własne na podstawie wyniku kwerendy z Web of Science]

Nazwa grupy	Liczba publikacji
4.116 Robotics	151
4.29 Automation & Control Systems	144
4.13 Telecommunications	72
4.18 Power Systems and Electronic Vehicles	66

Problem estymacji pozycji jest często poruszany w zagadnieniach związanych z robotyką mobilną. Prace badawcze w tym obszarze dotyczą wykorzystania różnego typu czujników w które wyposażone są obiekty mobilne. Wśród stosowanych czujników możemy wyróżnić różne ich grupy. Jednymi z najczęściej spotykanych są inercjalne jednostki pomiarowe wykonane w technologii MEMS (ang. Micro-Electro Mechanical Systems). Są one oznaczane jako 6-osiowe (trój-osiowe akcelerometry oraz giroskopy) lub 9-osiowe (trój-osiowe akcelerometry, giroskopy i magnetometry). Inną grupą sensorów wykorzystywaną w robotyce mobilnej do wyznaczania pozycji są systemy wykorzystujące odometrię, oparte o różnego typu enkodery. Wykorzystywane są również systemy sensorów, których zasada działania oparta jest o efekt Dopplera – np. ultradźwiękowe czujniki odległości. Osobny zbiór stanowią systemy wizyjne wykorzystywane w nawigacji robotów mobilnych. Innym typem czujników wykorzystywanych w robotyce mobilnej są moduły nawigacyjne odbierające sygnały z globalnych satelitarnych systemów nawigacyjnych. Umożliwiają one, przy widoczności minimum 4 satelitów wyznaczenie pozycji odbiornika w układach odniesienia takich jak m.in. WGS-84 dla konstelacji GPS, PZ-90 dla konstelacji GLONASS lub GTRF dla europejskiej konstelacji Galileo [3]. Informacja o długości i szerokości geograficznej oraz wysokości położenia odbiornika najczęściej zwracana jest w formacie NMEA-0183 [4]. Format ten stosowany jest jako standard wiadomości nawigacyjnych dostępny szerokiej gamie urządzeń, od mikrokontrolerów po profesjonalne zestawy odbiorników nawigacyjnych wraz ze stacjami nawigacyjnymi [5-6].

Odczyty z każdej z powyższych grup czujników obarczone są błędami. Powstają one w wyniku budowy czujników, zasady działania ich elementów składowych – np. dryft giroskopu lub akumulacja błędu podczas całkowania danych odczytywanych z akcelerometru. Innym źródłem niedokładności odczytów mogą być opóźnienia drogi sygnału radiowego w przypadku odbiorników GNSS. Wśród możliwych źródeł błędów można wyróżnić tutaj błędy synchronizacji zegarów satelitów z zegarami odbiorników, efekty związane z odbijaniem się sygnału lub zakłóceniami związanymi z pojawianiem się chmur. Ponadto często nieznane środowiska, w których pracują czujniki wpływają na generowane błędy. Przykładem może być zmniejszona dokładność, a nawet zanik sygnału z systemów GNSS w lesie po deszczu lub zmniejszona dokładność systemów wizyjnych w środowisku w którym występuje dym lub inne zjawiska ograniczające widoczność.

Ze względu na fakt, iż nie sposób przewidzieć wszystkich potencjalnych źródeł zakłóceń oraz błędów realizowane są próby uodpornienia systemów sensorów oraz układów sterowania na zjawiska nieprzewidziane. W obszarze estymacji pozycji i położenia stosuje się techniki fuzji sygnałów sensorycznych. Polegają one na pomiarach przy użyciu

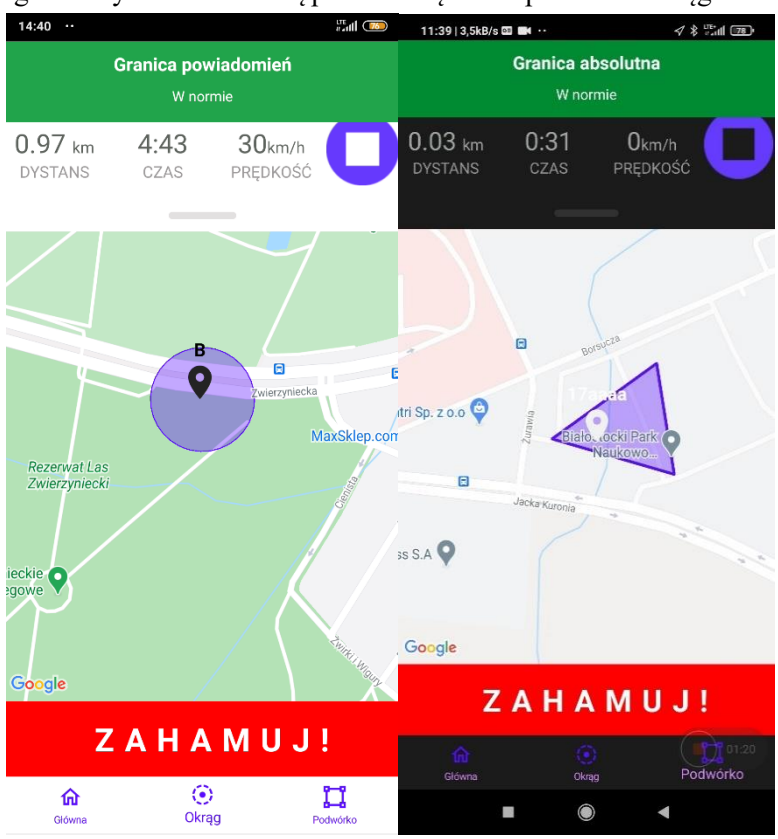
więcej niż jednego, często komplementarnego systemu sensorycznego. Umożliwia to przeciwdziałanie błędom pozycjonowania poprzez użycie uzupełniających się źródeł danych. Pozwala to na minimalizację błędów oraz uodparnia system sterowania oraz wyznaczania pozycji, prędkości i położenia na nieznane zakłócenia w zmiennych środowiskach. W pracy [7] przedstawiono propozycję zwiększenia dokładności estymacji pozycji i położenia robota mobilnego poprzez użycie dodatkowego systemu radiowego opartego o lokalizację bliskiego zasięgu z wykorzystaniem systemu UWB (ang. *Ultra Wide Band*). Istniejące sensory – inercjalna jednostka pomiarowa oraz enkodery absolutne zostały uzupełnione o dane pochodzące z systemu lokalizacji radiowej. Dzięki wykorzystaniu algorytmów fuzji danych pozwoliło to na zwiększenie dokładności pozycjonowania oraz uodporniło robota na błędy z poszczególnych sensorów. Dodanie systemu wizyjnego oraz zastosowanie rozszerzonego filtru Kalmana (ang. Extended Kalman Filter) pozwoliło na zwiększenie jakości estymacji pozycji [8]. Kolejnym obszarem stosowania technik fuzji danych są bezzałogowe statki powietrzne. Pokładowe układy sterowania wykorzystują połączenie inercjalnych jednostek pomiarowych oraz układów odbiorników sygnałów systemów globalnej nawigacji satelitarnej w celu estymacji kierunku, pozycji oraz prędkości lotu z pomocą użycia technik filtracji i fuzji danych nazywanych filtrami Kalmana [8-12].

1.2. Geofencing

Geofencing jest zbiorczą nazwą rozwiązań technicznych ograniczających dostępny obszar ruchu dla różnego typu pojazdów poprzez stworzenie wirtualnej przestrzeni. W pracy [13] przedstawiono aplikację podziału przestrzeni 3D za pomocą algorytmów do wspierania zarządzania przestrzenią powietrzną dla bezzałogowych statków powietrznych. Wykorzystanie tworzenia dynamicznych map otoczenia z wykorzystaniem wielu czujników dla inteligentnych systemów transportowych, w szczególności dla autonomicznych pojazdów oraz bezzałogowych statków powietrznych przedstawiono w pracy [14]. Użycie systemu geofencingu w przypadku wystąpienia sytuacji katastrof naturalnych jako narzędzia przydatnego służbom ratunkowym do monitorowania położenia osób przedstawiono w [15]. Autorzy proponują wykorzystanie technik geolokalizacyjnych opartych o telefony komórkowe użytkowników, połączone z dedykowaną aplikacją serwerową opartą o silnik przetwarzający i przechowujący dane o aktualnej pozycji. Dzięki porównywaniu aktualnej pozycji użytkownika z pozycjami określonymi jako współrzędne granic możliwe jest wysyłanie spersonalizowanych wiadomości o niebezpieczeństwach związanych z pobytem w strefie zagrożenia. Służby ratownicze posiadają dzięki temu informację o położeniu osób i mogą zaplanować ewakuację.

W ramach opisywanego projektu użycie technik geofencingu pozwoliło na wyznaczanie dopuszczalnej odległości pomiędzy urządzeniem mobilnym rodzica z zainstalowaną aplikacją BikeGuard, a położeniem roweru dziecka z zainstalowanym modułem SmartControl. System SmartBrake udostępnia w trybie wirtualnego podwórka dwa tryby jazdy wykorzystujące techniki geofencingu. Pierwszym z nich jest jazda po okręgu, którego środek wyznaczany jest przez położenie telefonu opiekuna. Maksymalna odległość od rodzica to promień okręgu o wartości 50 metrów. Rodzic może zmienić tą wartość w zakresie od 5 do 50 metrów. Drugim z trybów jazdy jest wirtualne podwórko, które rodzic może rysować jako wielokąt ograniczający obszar poruszania się dziecka. Oba tryby działają tak, aby jazda rowerem dopuszczalna była tylko i wyłączenie wewnątrz

wirtualnego podwórka zdefiniowanego przez rodzica. W obu trybach wyjazdu dziecka poza dozwolony obszar, wygeneruje informację o przekroczeniu granicy, którą otrzyma rodzic. W zależności od ustawień, w momencie przekroczenia granicy wyzwalana jest akcja hamowania do całkowitego zatrzymania roweru lub ograniczana jest prędkość jego jazdy. Na rysunku 3 przedstawiono widoki z aplikacji mobilnej BikeGuard umożliwiającej definiowanie parametrów wirtualnych obszarów. Na obu widokach widzimy belkę z komunikatami i statusami z widocznym napisem „Granica powiadomień". Poniżej wyświetlane są parametry aktualnej jazdy: przejechany dystans w kilometrach, czas aktualnej jazdy oraz prędkość rejestrowana przez moduł SmartControl. Poniżej użytkownik ma możliwość wyzwolenia akcji hamowania poprzez wciśnięcie przycisku „Zahamuj" widocznego na czerwonym tle. Czarny punkt na mapie zaznaczony na lewym rysunku wewnątrz okręgu oraz biały punkt wewnątrz trójkąta po prawej symbolizują położenie dziecka. W trybie okręgu położenie rodzica symbolizowane jest przez jego środek. Dopuszczalne granice symbolizowane są przez krawędzie odpowiednio okręgu lub wielokąta.



Rysunek 3. Widok z aplikacji mobilnej BikeGuard – definiowanie wirtualnego obszaru umożliwiającego, w którym dopuszczalna jest jazda rowerem. Po lewej tryb okręgu, po prawej tryb podwórka [opracowanie własne]

W niniejszej pracy przedstawione zostaną zagadnienia związane z fuzją danych nawigacyjnych w celu określenia lokalizacji roweru dziecka. Problemem badawczym jest osiągnięcie założonych parametrów kamienia milowego projektu związanego z dokładnością lokalizacji rozumianej jako określanie długości i szerokości geograficznej

polegało na opracowaniu technik pozwalających na uzyskania wyznaczania pozycji w przedziale od 3,5 do 4,5 metra wg CEP (ang. *Circular Error Probable*). Rezultat ten osiągnięto poprzez implementację rozszerzonego filtru Kalmana na module SmartControl, który estymuje aktualne położenie roweru. W pracy przedstawione zostaną:

1. Sensoryczne źródła danych wejściowych do zaimplementowanych algorytmów.
2. Opis realizacji części symulacyjnej pracy związanej z implementacją rozszerzonego filtru Kalmana w środowisku Matlab.
3. Opis implementacji algorytmu filtracji na dedykowanym mikrokontrolerze, w trakcie której przeprowadzono szereg prac eksperymentalnych oraz integracje z pozostałymi komponentami systemu SmartBrake.

Niniejsza praca składa się z pięciu rozdziałów – wstępu, z krótką charakterystyką projektu SmartBrake oraz technik związanych z lokalizacją i geofencingiem. Kolejnym rozdziałem jest opis metod badawczych użytych w trakcie osiągnięcia parametrów kamienia milowego projektu. Rozdział trzeci przedstawia wybrany algorytm fuzji danych. W rozdziale czwartym przedstawiono wyniki, natomiast rozdział piąty niniejszej pracy zawiera jej podsumowanie.

2. Metody badawcze

Przedstawione w publikacji metody badawcze tj. prace symulacyjne oraz eksperymentalne związane z integracją algorytmów fuzji danych opartych o rodzinę filtrów Kalmana wybrano ze względu na następujące ograniczenia:

1. Parametry kamienia milowego: osiągnięcie dokładności lokalizacji w przedziale od 3,5 do 4,5 metra wg CEP.
2. Wykorzystanie tanich oraz komercyjnie dostępnych modułów inercjalnej jednostki pomiarowej oraz odbiornika globalnych sygnałów nawigacji satelitarnej.
3. Ograniczenia związane z mocą obliczeniową mikrokontrolera sterującego pracą modułu SmartControl – układ z rodziny STM32L4.
4. Iteracyjny charakter realizacji prac wynikający z podziału projektu na etapy kończące się osiągnięciem kamieni milowych.

Ze względu ma powyższe ograniczenia zdecydowano się na dwutorową, równoległą realizację prac badawczych związanych z badaniami metod fuzji sygnałów pozwalających na osiągnięcie wymaganych parametrów lokalizacji modułu SmartControl.

Zespół realizujący prace związane z badaniami symulacyjnymi przeprowadził analizę literatury dotyczącej przedmiotu filtracji i fuzji danych na potrzeby nawigacji. Następnie w środowisku naukowo-inżynierskim Matlab przeprowadzono badania symulacyjne różnych wariantów filtru Kalmana: wersji liniowej, rozszerzonego filtru Kalmana oraz bezśladowego filtru Kalmana. Przeprowadzono badania związane z doborem macierzy ufności modelu oraz pomiarów. Prace symulacyjne pozwoliły na zmniejszenie kosztów związanych z przeprowadzeniem wielu iteracji badań eksperymentalnych. Wyniki badań przedstawiono na rysunkach 6-10. Ich realizacja pozwoliła następnie na przygotowanie się do przeprowadzenia prac implementacyjnych rozszerzonego filtra Kalmana na wybranej platformie mikrokontrolerowej, będącej elementem składowym kolejnych wersji prototypowych modułu SmartControl. Prace eksperymentalne polegały na doborze komponentów systemu a następnie na iteracyjnej budowie kolejnych wersji modułu SmartControl oraz realizacji testów poprawności ich pracy. Pozwoliło to na przeprowadzenie jazd rowerem w różnych warunkach terenowych oraz po różnych trasach. Miały one na celu

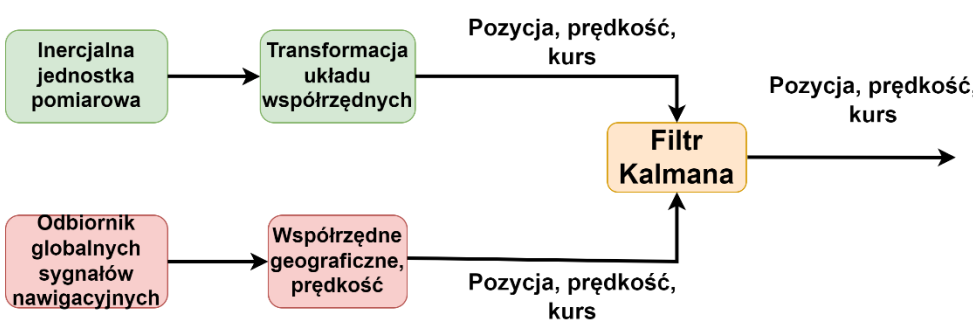
m.in.: testy poprawności estymacji pozycji (w określonej tolerancji ograniczonej przez parametry kamienia milowego), weryfikację zapisywania danych oraz zbadanie poprawności komunikacji modułu SmartControl z pozostałymi modułami systemu SmartBrake.

Dwutorowa, równoległa realizacja prac pozwoliła na optymalizację czasu trwania poszczególnych zadań w projekcie, iteracyjne podejście do ich wykonania oraz zaoszczędzenie czasu i kosztów związanych z badaniami eksperymentalnymi. Dzięki symulacjom, zespół odpowiedzialny za przeprowadzenie badań eksperymentalnych był lepiej przygotowany oraz pozyskał wiedzę w zakresie potencjalnych problemów.

3. Implementacja fuzji danych

Ze względu na rodzaj informacji zwrotnej dostarczanej przez inercjalne jednostki pomiarowe oraz układy odbiorników sygnałów nawigacyjnych, najczęściej wybieranymi metodami integracji są metody oparte o filtrację Kalmana. Jest ona wykorzystywana do powiązania wyników znajdujących się w wektorze stanu systemu INS z regularnie pojawiającymi się wyznaczeniem z systemu satelitarnego GNSS. Ze względu na typ algorytmu połączenia wyników nawigacji wyróżnić możemy integrację luźną (ang. *loosely coupled*) oraz integrację ścisłą (ang. tightly coupled). W ramach modułu SmartControl zaimplementowano algorytm fuzji danych z wbudowanego modułu inercjalnej jednostki pomiarowej oraz odbiornika globalnych sygnałów nawigacyjnych w postaci luźno powiązanej. Jest on powszechnie stosowany w przypadku stosowania komercyjnych niskokosztowych modułów sensorycznych [16]. Wpływ błędów generowanych przez elementy MEMS – akcelerometry, giroskopy oraz magnetometry jest ograniczany przez użycie drugiego źródła danych nawigacyjnych [17]. Autorzy pracy [18] wykorzystali dodatkowo oprócz pary inercjalnej jednostki pomiarowej i odbiornika GNSS algorytm lasów losowych. Pozwoliło to zmniejszyć błąd podczas zaników sygnału satelitarnego. Problemem podczas tego typu integracji jest odpowiednia transformacja układów współrzędnych, z układu bezwzględnego inercjalnej jednostki pomiarowej do układu globalnego, w którym odbierane są dane z odbiornika GNSS.

Drugim istotnym zagadnieniem jest różnica w częstotliwości odczytów danych pomiarowych. Inercjalna jednostka pomiarowa odczytuje przyspieszenia, prędkości kątowe i aktualne pole magnetyczne kilkadziesiąt razy częściej niż odbiornik nawigacyjny swoją pozycję i parametry takie jak prędkość i azymut. Ideowy schemat algorytmu fuzji danych przedstawiono na rysunku 4.



Rysunek 4. Schemat algorytmu luźnego powiązania odbiornika globalnych sygnałów nawigacyjnych z inercjalną jednostką pomiarową [opracowanie własne na podstawie [17]]

Rozwiązanie nawigacyjne pochodzące z odbiornika nawigacji satelitarnej dostarcza dane takie jak: pozycja oraz prędkość obiektu. Po odpowiedniej transformacji dane te pozwalają na ograniczenia niedokładności informacji pochodzących z dużo wrażliwszej inercjalnej jednostki pomiarowej. Dane z obu źródeł trafiają do filtra Kalmana, gdzie następuje porównanie prognozowanych błędów pozycji z różnicami pomiędzy wynikami z odbiornika sygnałów nawigacji satelitarnej oraz inercjalnej jednostki pomiarowej. Obliczone estymaty wykorzystywane są jako poprawki wyniku obliczeń dla układu inercyjnego. Aktualizacja dla błędów pozycjonowania inercjalnej jednostki pomiarowej jest dostępna jedynie w czasie, w którym istnieje rozwiązanie nawigacyjne układu odbiornika sygnałów nawigacji satelitarnej (dostępny jest sygnał z przynajmniej czterech satelitów w tym samym czasie).

Charakterystyka algorytmu luźnej integracji:

- niezależność pracy, możliwe do uzyskania są trzy rozwiązania nawigacyjne (z inercjalnej jednostki pomiarowej, modułu odbiornika nawigacji satelitarnej, rozwiązanie mieszane);
- efekt fuzji występuje tylko wtedy, gdy otrzymujemy sygnał z modułu odbiornika globalnych sygnałów nawigacji satelitarnej.

Poniższe równania filtra Kalmana w integracji luźnej przedstawiono w pracach [16-26]:
Model predykcji w integracji luźnej

$$\delta \hat{x}_t = A\delta x_{t-1} + Bu \tag{1}$$

Wektor stanu składa się z błędów pozycji, prędkości, orientacji zewnętrznej, giroskopu oraz akcelerometru:

$$\delta \hat{x}_{t\,15x1} = [\delta r_{3x1}, \delta \upsilon_{3x1}, \delta \varepsilon_{3x1}, \delta \omega_{3x1}, \delta a_{3x1}]^T$$
$$\delta \hat{x}_{t\,15x1} = [\delta_{r_{3x1}}, \delta_{\upsilon_{3x1}}, \delta_{\varepsilon_{3x1}}, \delta_{\omega_{3x1}}, \delta_{a_{3x1}}]^T \tag{2}$$

gdzie:
δr = [δφ, δλ, δh]T – wektor błędów pozycji (szerokości, długości i wysokości),
δυ = [δυe, δυn, δυh]T – wektor błędów prędkości w układzie lokalnym,
δε = [δθ, δφ, δψ]T – wektor błędów kątów Eulera (θ – kąt nutacji, φ – kąt precesji, ψ – kąt obrotu własnego),
δω = [δωx, δωy, δωz]T – wektor błędów giroskopu (składający się z dryftów),
δa = [δax, δay, δaz]T – wektor błędów akcelerometru (składający się z błędów stałych).

Model pomiaru integracji luźnej można zapisać jako:

$$\delta z_t = H_t \delta x_t + \eta_t \tag{3}$$

W wektorze stanu modelowane są błędy inercyjnej jednostki pomiarowej, zatem w odpowiadającym mu wektorze pomiarowym znajdują się różnice pomiędzy rozwiązaniami algorytmów nawigacji satelitarnej i nawigacji inercyjnej INS:

$$\delta z_t = \begin{bmatrix} r_{INS} - r_{GPS} \\ \upsilon_{INS} - \upsilon_{GPS} \end{bmatrix} = \begin{bmatrix} \varphi_{INS} - \varphi_{GPS} \\ \lambda_{INS} - \lambda_{GPS} \\ h_{INS} - h_{GPS} \\ \upsilon_{e\,INS} - \upsilon_{e\,GPS} \\ \upsilon_{n\,INS} - \upsilon_{n\,GPS} \\ \upsilon_{h\,INS} - \upsilon_{h\,GPS} \end{bmatrix} \tag{4}$$

Wektor $\boldsymbol{\eta_t}$ jest wektorem niepewności pomiarowych odpowiedzi z systemu satelitarnego. W luźnej integracji macierz H będzie wyglądać następująco:

$$H_{6x15} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{5}$$

Macierz kowariancji pomiaru R, jest obliczana wewnątrz algorytmu pozycjonowania nawigacji satelitarnej i ma formę diagonalną:

$$R_{6x6} = diag(\sigma_{\varphi}^2, \sigma_{\lambda}^2, \sigma_{h}^2, \sigma_{\upsilon_e}^2, \sigma_{\upsilon_n}^2, \sigma_{\upsilon_h}^2) \tag{6}$$

Błędy otrzymane na wyjściu filtra są podstawą do poprawienia wskazań nawigacji inercyjnego systemu pomiarowego.

3.1. Rozszerzony filtr Kalmana

Ze względu na nieliniowy charakter danych otrzymywanych z modułów sensorycznych użyto rozszerzonego filtru Kalmana. Równanie dla procesów nieliniowych opisane przez zależności różnicowe [16, 27]:

$$\begin{aligned} x_{k+1} &= f(x_k) + w_k \\ y_{k+1} &= h(x_{k+1}) + v_{k+1} \end{aligned} \tag{7}$$

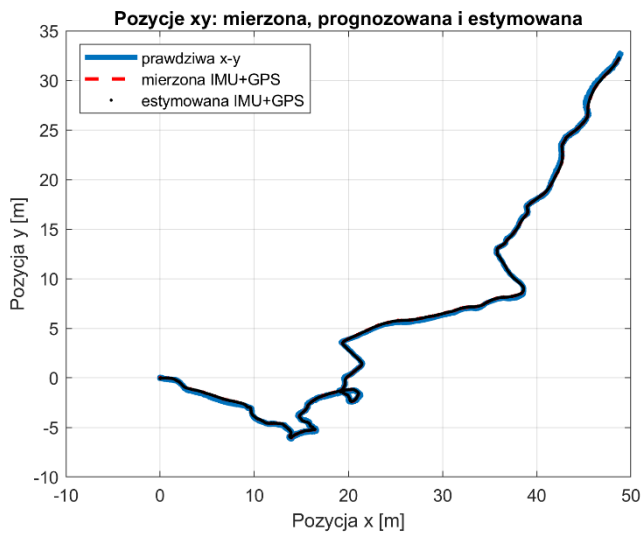
gdzie:
x_{k+1} – wektor stanu;
y_{k+1} – wektor pomiarowy;
w_k – wektor zmiennej losowej szumu procesu;
v_{k+1} – wektor zmiennej losowej szumu pomiaru;
f() – nieliniowa funkcja stanu wiążąca wektor stanu w kroku *k* z wektorem stanu w kroku *k+1*;
h() – nieliniowa funkcja pomiarowa wiążącą stan z pomiarami w tym samym kroku *k*.

Rozszerzony filtr Kalmana posiada następujące ograniczenia:

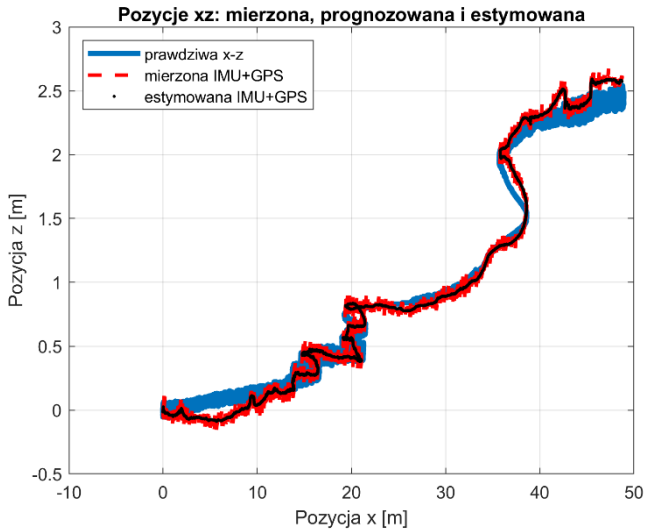
- działa dobrze tylko z małymi nieliniowościami, ponieważ opiera się na lokalnym przybliżeniu liniowym;
- dla układów wysoce nieliniowych model EKF nie jest estymatorem optymalnym;
- ze względu na Jacobiany, modele pomiaru i dynamiki muszą być zróżniczkowane;
- dozwolone są tylko szumy procesowe addytywne i gaussowskie.

4. Wyniki

Wyniki symulacyjne zaimplementowanego filtru Kalmana przedstawiono na rysunkach 5-9. W celu przeprowadzenia badań w środowisku Matlab zasymulowano odczyty odbiornika sygnałów nawigacji satelitarnej oraz inercjalnej jednostki pomiarowej dla ruchu punktu materialnego w układzie kartezjańskim XYZ. Widoczne oscylacje wartości na rysunku 6 wynikają z poziomu szumu pomiarowego inercjalnej jednostki pomiarowej. Zaimplementowano m.in. równania 1-7 oraz ustalono krok i czas symulacji. Na rysunkach 5-9 przedstawiono wykresy dla częstotliwości 50 Hz oraz czasu symulacji trwającego 150 sekund (czas kroku = 0,02 s, liczba kroków = 7500).

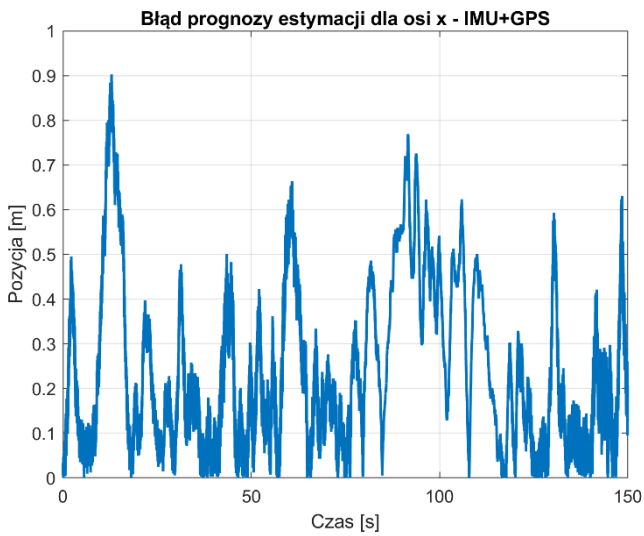


Rysunek 5. Wykres przedstawiający zmianę pozycji obiektu w płaszczyźnie XY [opracowanie własne]

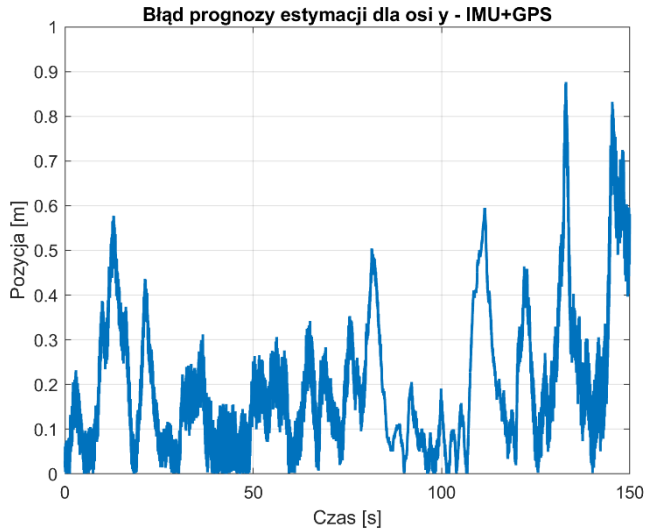


Rysunek 6. Wykres przedstawiający zmianę pozycji obiektu w płaszczyźnie XZ [opracowanie własne]

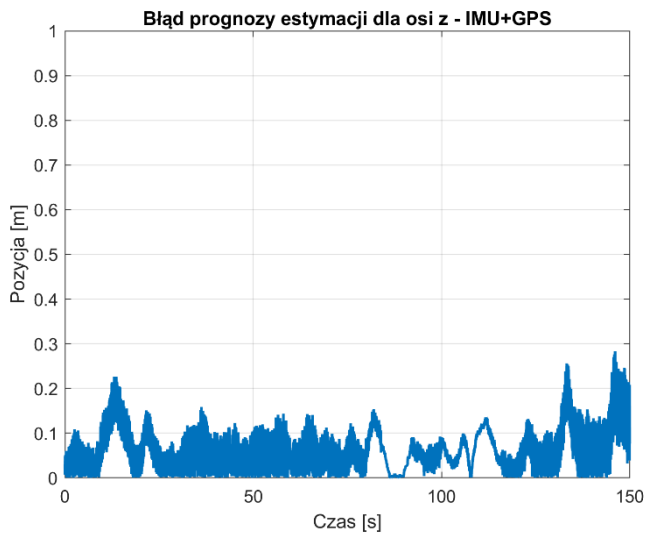
Wykresy błędów prognoz przedstawiono na rysunkach 7-9. Przedstawiono na nich różnicę pomiędzy rzeczywistą pozycją a jej estymacją dla każdej z trzech osi ruchu. Wartości błędów dla każdej z osi nie przekraczają wartości 1 metra co oznacza spełnienie wymogu wynikającego z parametrów kamienia milowego.



Rysunek 7. Błąd prognozy estymacji osi X [opracowanie własne]



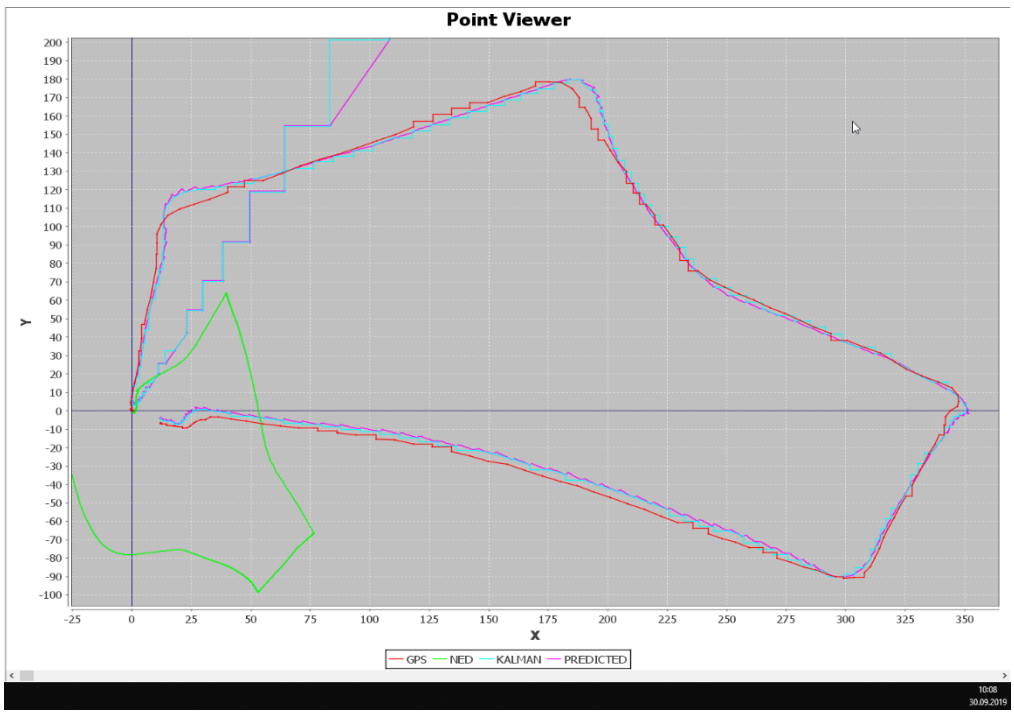
Rysunek 8. Błąd prognozy estymacji osi Y [opracowanie własne]



Rysunek 9. Błąd prognozy estymacji osi Z [opracowanie własne]

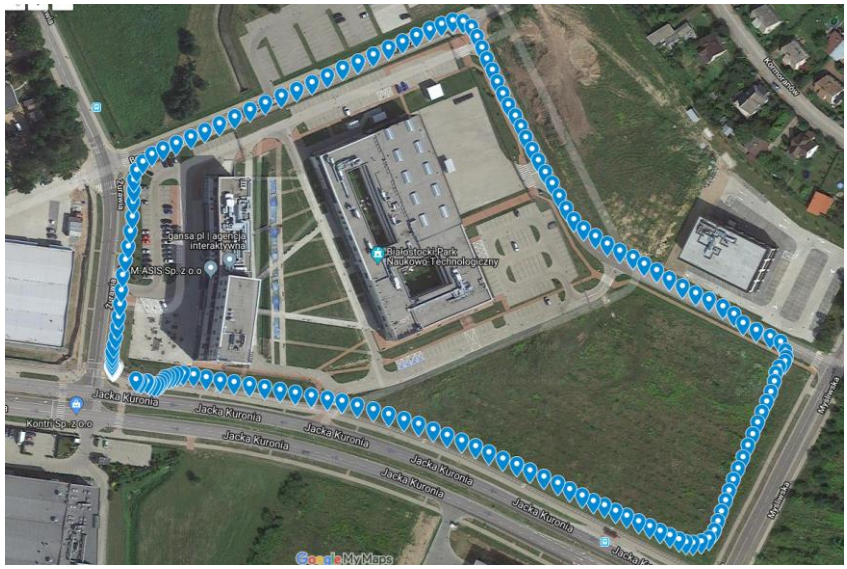
Równolegle do przeprowadzonych badań symulacyjnych realizowane były prace związane z implementacją filtra Kalmana oraz fuzji danych na platformie mikroprocesowej – mikrokontrolerowym układzie produkcji firmy STM z rodziny L4. Dane z inercjalnej jednostki pomiarowej pochodziły z modułu LSM9DS1. Do pozyskiwania danych z globalnej nawigacji satelitarnej wykorzystany został układ CD-Top Technology PA1616D. Dzięki realizacji fazy badań symulacyjnych, podczas prac eksperymentalnych wykorzystano wiedzę oraz zdobyte doświadczenia, co pozwoliło na zmniejszenie liczby eksperymentów. Eksperymenty przeprowadzone podczas jazdy rowerem wyposażonym w moduł SmartControl potwierdziły poprawność implementacji filtra Kalmana na platformie mikroprocesorowej. Przykładem takiego eksperymentu jest wykres przedstawiony

na rysunku 11, gdzie widoczny jest zapis wskazań pozycji roweru z wykorzystaniem układu inercjalnej jednostki pomiarowej, odbiornika sygnałów nawigacyjnych oraz dwóch przypadków ich fuzji. Linia niebieska to przykład działania fuzji sygnałów bez oczekiwania na pojawienie się sygnału nawigacyjnego z odbiornika GNSS, linia fioletowa to fuzja, gdzie mamy dostępne dane z obu źródeł informacji pomiarowej. Kształt linii zielonej wynika z problemów numerycznych związanych z akumulacją błędów całkowania przyspieszeń.



Rysunek 11. Wykres zmian pozycji uzyskanej podczas eksperymentu. Linia fioletowa – predykcja filtru Kalmana, linia czerwona – odbiornik sygnałów nawigacyjnych, linia zielona – inercjalna jednostka pomiarowa, linia niebieska filtr Kalmana (zawsze), linia fioletowa predykcja gdy dostępny odczyt z modułu odbiornika sygnałów nawigacji satelitarnej [opracowanie własne]

Na rysunku 12 w celu porównania przedstawiono przebieg trasy zapisany na mapie satelitarnej. Przebieg zapisanej pozycji jest zbieżny z wykresami przedstawionymi na rysunku 11. Jedynym punktem gdzie występuje zanik sygnału dla wszystkich użytych metod jest dolny lewy punkt (współrzędne [0,0]). Jest to miejsce gdzie ze względu na ukształtowanie terenu oraz budynki obok występuje miejscowy brak sygnału z odbiornika nawigacyjnych sygnałów satelitarnych.



Rysunek 12. Wizualizacja przebiegu pozycji modułu SmartControl naniesiona na mapę satelitarną [opracowanie własne]

Zarówno wyniki prac symulacyjnych jak i badań eksperymentalnych potwierdziły osiągnięcie parametru kamienia milowego związanego z dokładnością lokalizacji w przedziale od 3,5 do 4,5 metra wg CEP (rysunek 7-9). Pozwoliło to na dalszą realizację prac w projekcie oraz integrację z aplikacją BikeGuard w celu realizacji trybów jazdy takich jak jazda w okręgu oraz jazda w trybie podwórka (przedstawione na rysunku 3). Wykorzystania w projekcie metodyka prowadzenia badań – tj. równoległe, dwutorowe realizowanie badań symulacyjnych oraz eksperymentalnych w sposób iteracyjny umożliwiło osiągnięcie wymaganych parametrów. Wyróżnić tu można dobór zapisu równań lub dobór wag macierzy ufności dla modelu i pomiarów. Przeprowadzone prace symulacyjne zaoszczędziły czas potrzebny na określenie tych parametrów podczas jazd eksperymentalnych.

5. Podsumowanie

W publikacji przedstawiono wybrane wyniki prac nad projektem „System SmartBrake – Inteligentny System Asekuracji Ruchu i Trasy Roweru”. Dotyczyły one zapewnienia poprawnej lokalizacji – estymacji pozycji poprzez użycie fuzji sygnałów. Zaimplementowany algorytm tzw. luźnej integracji inercjalnej inercyjnej jednostki pomiarowej oraz układu odbiornika globalnych sygnałów nawigacji satelitarnej zwiększa dokładność estymacji pozycji względem metod opartych o pojedyncze źródło sygnału. Jak przedstawiono w rozdziale czwartym – użycie odbiornika sygnałów nawigacji satelitarnej zmniejsza błędy wynikające z dryftu giroskopów oraz błędów całkowania akcelerometrów. Na rysunku 12 przedstawiono dodatkową modyfikację w implementacji na rzeczywistym układzie sterowania – obliczenie pozycji realizowane jest tylko w przypadku dostępności obliczonej pozycji z modułu odbiornika sygnałów nawigacyjnych. Pozwala to ograniczyć efekty nagłej zmiany pozycji (rysunek 11 – kolor zielony) gdy estymacja położenia oparta jest tylko o moduł inercyjnej jednostki pomiarowej. Realizacja prac zapewniła poprawne wykorzystanie technik geofencingu co umożliwiło osiągnięcie zakładanych rezultatów

projektu – odpowiednią dokładność wskazań lokalizacji oraz płynność działania trybów okręgu oraz podwórka.

Miarą liczbową określającą poprawność estymacji pozycji przez moduł SmartControl było spełnienie warunków dotyczących parametrów lokalizacji wyszczególnionych we wniosku o dofinansowanie tj. osiągnięcie dokładności mniejszej niż 3,5-4,5 m wg CEP.

Dalsze kierunki prac obejmują kontynuację rozwoju modułów systemu SmartBrake oraz ich komercjalizację. Planowane jest wykorzystanie fuzji danych z filtru Kalmana wbudowanego w moduł SmartControl oraz przy pomocy danych zapisywanych z urządzenia mobilnego na którym zainstalowana jest aplikacja BikeGuard.

Literatura

1. https://www.indiegogo.com/projects/minibrake-to-make-cycling-safer-for-kids-all-over-the-world#/ [data dostępu: 19.08.2023]
2. https://wsjp.pl/haslo/podglad/107367/geofencing [data dostępu: 19.08.2023]
3. https://gssc.esa.int/navipedia/index.php/Reference_Frames_in_GNSS [data dostępu: 15.06.2023].
4. Dang S., Huang H., Li X., *About the Parsing of NMEA–0183 Format Data Streams in GPS, Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*, ICNC-FSKD 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 153. Springer, Cham., s. 1282-1289.
5. https://receiverhelp.trimble.com/alloy-gnss/en-us/NMEA-0183messages_MessageOverview.html [data dostępu: 19.08.2023]
6. https://docs.novatel.com/OEM7/Content/Logs/NMEA_Standard_Logs_.htm [data dostępu: 19.08.2023]
7. Li J., Xue J., Fu D., Gui C., Wang X., *Position Estimation and Error Correction of Mobile Robots Based on UWB and Multisensors*, Journal of Sensors, 2022, s. 1-18.
8. Alatise M.B., Hancke G.P., *Pose Estimation of a Mobile Robot Based on Fusion of IMU Data and Vision Data Using an Extended Kalman Filter*, Sensors, 17 (10), 2017, s. 1-22.
9. Lin F., Miao X.L., Qu X.G., *Kalman Filter and Complementary Filter in Attitude Estimation of APM*, Applied Mechanics and Materials, 577, 2014, s. 794-797.
10. Salwa M., Krzysztofik I., *Application of Filters to Improve Flight Stability of Rotary Unmanned Aerial Objects*, Sensors, 22 (4), 2022, s. 1-18.
11. Mao G., Drake S., Anderson B.D.O., *Design of an Extended Kalman Filter for UAV Localization, In Proceedings of the Information*, Decision and Control, Adelaide, Australia, 12–14 Lutego 2007, s. 224-229.
12. Valade A., Acco P., Grabolosa P., Fourniol J., *A Study about Kalman Filters Applied to Embedded Sensors*, Sensors, 17, 2810, 2017, s. 1-18.
13. Kim J., Atkins E., *Airspace Geofencing and Flight Planning for Low-Altitude*, Urban, Small Unmanned Aircraft Systems, Applied Sciences, 12 (2), 2022, s. 1-24.
14. Maiouak M., Taleb T., *Dynamic Maps for Automated Driving and UAV Geofencing*, IEEE Wireless Communications, 26 (4), 2019, s. 54-59.
15. Suyama A., Inoue U., *Using Geofencing for a Disaster Information System*, IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), 2016, s. 1-5.
16. Falco G., Pini M., Marucco G., *Loose and Tight GNSS/INS Integrations: Comparison of Performance Assessed in Real Urban Scenarios*, Sensors, 17 (2), 2017, s. 1-25.
17. Quinchia A.G., Falco G., Falletti E., Dovis F., Ferrer C.A, *Comparison between Different Error Modeling of MEMS Applied to GPS/INS Integrated Systems*, Sensors, 13 (8), 2013, s. 9549-9588.

18. Adusumilli S., Bhatt D., Wang H., Bhattacharya P., Devabhaktuni V.A, *Low-Cost INS/GPS Integration Methodology Based on Random Forest Regression*, Expert Systems with Applications, 40 (11), 2013, s. 4653-4659.
19. Al Bitar N., Gavrilov A.I., *Comparative Analysis of Fusion Algorithms in a Loosely-Coupled Integrated Navigation System on the Basis of Real Data Processing*, Gyroscopy Navigation, 10 (4), 2019, s. 231-244.
20. Noureldin A., Karamat T. B., Georgy J., *Fundamentals of inertial navigation, satellite-based positioning and their integration*, Springer Science & Business Media, Nowy Jork, USA, 2012.
21. Solimeno A., *Low-cost INS/GPS data fusion with extended Kalman filter for airborne applications*, Praca Magisterska, Universidade Technica de Lisboa, 2007.
22. Syed Z., *Design and implementation issues of a portable navigation system, Department of Geomatics Engineering*, Praca Doktorska, The University of Calgary, Calgary, Alberta, Kanada, 2009.
23. Syed Z., Aggarwal P., Goodall C., Niu X., El-Sheimy N., *A new multi-position calibration method for MEMS inertial navigation systems*, Measurement Science and Technology, 18(7), 2007, s. 1897-1907.
24. Tomaszewski D., *Doskonalenie metod pozycjonowania z wykorzystaniem zintegrowanych pomiarów INS/GPS*, Praca doktorska, Wydział Geodezji, Inżynierii przestrzennej i Budownictwa Uniwersytetu Warmińsko-Mazurskiego w Olsztynie, 21 listopada 2017.
25. Wang W., Liu Z.Y., Xie R.R., *Quadratic extended Kalman filter approach for GPS/INS integration*, Aerospace science and technology, 10(8), Elsevier, 2006 s. 709-713.
26. Xie F., Liu J., Li R., Jiang B., Qiao L., *Performance analysis of a federated ultra-tight global positioning system/inertial navigation system integration algorithm in high dynamic environments*, Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 229(1), 2015, s. 56-71.
27. https://stanford.edu/class/ee363/lectures/ekf.pdf [data dostępu: 03.07.2023].

Symulacyjna i eksperymentalna estymacja pozycji roweru – przykład projektu B+R SmartBrake

Streszczenie

W niniejszej pracy przedstawiono wyniki prac symulacyjnych oraz eksperymentalnych dotyczących opracowania metody wyznaczania pozycji i prędkości roweru dziecka. W pierwszym rozdziale przedstawiono projekt badawczo-rozwojowy w ramach którego przeprowadzono badania. Następnie autorzy opisali zagadnienia związane z estymacją pozycji w oparciu o dane odczytywane z czujników. W kolejnym rozdziale przedstawiono zagadnienia związane z wybraną metodą fuzji sygnałów z inercjalnej jednostki pomiarowej oraz odbiornika sygnałów z globalnych satelitarnych sieci nawigacyjnych. Następnie przedstawiono dobór układów sprzętowych oraz wyniki badań symulacyjnych oraz eksperymentalnych. W ostatnim rozdziale przedstawiono podsumowanie realizacji tej części projektu oraz wnioski.

Słowa kluczowe: estymacja pozycji, filtry Kalmana, fuzja sygnałów, integracja INS/GNSS

Simulation and experimental estimation of bicycle position – an example of SmartBrake R&D project

Abstract

This paper presents the results of simulation and experimental work on developing a method for determining the position and attitude of a child's bicycle. The first chapter presents the research and development project under which the research was conducted. Then the authors described issues related to position estimation based on data read from sensors. The next chapter presents issues related to the chosen signal fusion method that uses signals from an inertial measurement unit and a global navigation satellite signals receiver. Then the selection of hardware systems and the results of simulation and experimental studies are presented. The final chapter contains a summary of the implementation of this part of the project and conclusions.

Keywords: position estimation, Kalman filters, signal fusion, INS/GNSS integration

Grzegorz Drozdowski[1], Paweł Dziekański[2]

Wpływ zachowań przedsiębiorczych pracowników na funkcjonowanie przedsiębiorstw w dostosowaniu się do otoczenia nowych technologii

1. Wprowadzenie

Dynamiczne otoczenie zewnętrzne i wzrost wymagań interesariuszy powoduje konieczność wprowadzania zmian w funkcjonowaniu przedsiębiorstw. Narastanie konkurencji globalnej oraz coraz większe koszty rozwoju gospodarki wiedzy niezbędnej do dorównania standardom najlepszych praktyk zarządzania w Europie i na świecie spowodowały konieczność ciągłego doskonalenia strategii przedsiębiorstw. Warunki działania firm wymagają także nowych umiejętności potrzebnych do osiągnięcia sukcesu rynkowego w okresie dostosowania się do środowiska nowych technologii [1].

Przedsiębiorstwa znajdują się pod szczególną presją czasu, ze względu na konieczność ciągłego dostosowania zasad swojego działania do warunków pojawiających się w nowych trendach gospodarki globalnej. Doświadczenie krajów wysoko rozwiniętych pokazuje, że przedsiębiorstwa tych krajów charakteryzują się otwartością na zmiany. Wynika to w dużej mierze z panującego przeświadczenia, że w świecie konkurencji globalnej, każda zdobyta przez przedsiębiorstwo pozycja ma charakter przejściowy i może w każdym momencie przekształcić się w sytuację niekorzystną. Prawdopodobieństwo wystąpienia takiej sytuacji pojawia się wówczas, kiedy imperatyw uczenia się, czyli ciągłego rozwoju kompetencji, związany chociażby z racjonalnym podejmowania decyzji, uruchomieniem przedsięwzięć antycypacyjnych, wprowadzaniem innowacji, organizowaniem procesów gospodarczych, zostanie przez kapitał ludzki firm zaniedbany lub zlekceważony [2]. Ten obraz praktyki gospodarczej, kreuje nowe reguły gry rynkowej, które przyjmują rolę paradygmatów w zmieniającym się środowisku konkurencyjnym w ramach adaptacji rozwiązań nowych, innowacyjnych technologii.

Otoczenie gospodarcze, w jakim funkcjonują dzisiejsze przedsiębiorstwa, ulega dramatycznej zmianie. Pojawiła się ostra rywalizacja między przedsiębiorstwami i ulega skróceniu długość życia firm. Mamy do czynienia ze zmianami w technologii i koniecznością reagowania na wyzwania strategiczne. Zaczyna się dążenie do zniżki kosztów stałych firm. Klienci oczekują coraz to nowszych towarów i usług o najwyższej jakości. Na rynku pracy z jednej strony trudno jest znaleźć pracodawcę gwarantującego długotrwałe zatrudnienie, z drugiej pracownicy oczekują, że będzie im przekazywanych więcej uprawnień i odpowiedzialności. Następuje zmiana przedsiębiorstw hierarchicznych w kierunku struktury projektowej [3]. Zatem, wraz ze wzrostem oczekiwań klientów, globalną konkurencją oraz malejącą rolą państwa, konkurencyjne i stale zmieniające się otoczenie może stanowić wyjątkowe wyzwanie dla kapitału ludzkiego przedsiębiorstw.

[1] grzegorz.drozdowski@ujk.edu.pl, Uniwersytet Jana Kochanowskiego w Kielcach, Wydział Prawa i Nauk Społecznych.
[2] pawel.dziekanski@ujk.edu.pl, Uniwersytet Jana Kochanowskiego w Kielcach, Wydział Prawa i Nauk Społecznych.

Dostosowanie systemu gospodarczego do wyzwań otoczenia globalnego implikowało konieczność wielowymiarowych zmian w polskich przedsiębiorstwach. Towarzyszące tym zmianom wyzwania objęły swoim zasięgiem niemal każdy aspekt przedsiębiorstwa. W szczególny sposób dotykają one jednak społecznych poziomów hierarchii przedsiębiorstw: od pracowników wykonawczych do kierowników najwyższego szczebla. Obserwując te przemiany z perspektywy przedsiębiorstw można zauważyć, że kapitał ludzki stanowi podstawowy podmiot określający zarówno możliwości przeprowadzenia zmian, jak i – co szczególnie istotne – ich utrwalenia i kontynuacji w przyszłości. Zagadnienie podstawowe zawierające się w problematyce powyższego zjawiska uwidocznione jest zwłaszcza na płaszczyźnie kompetencji kapitału ludzkiego przedsiębiorstw.

Powyższe założenie nabiera nowego znaczenia, jeżeli odniesiemy ją do elastyczności jako podstawowego atrybutu kompetencji pracowników firm. W świetle zadań stawianych pracownikom przez przedsiębiorstwo i jego otoczenie, coraz częściej i z większą siłą pojawia się wymaganie elastyczności kompetencyjnej. Wysoka elastyczność kompetencji staje się synonimem efektywnego zarządzania i podstawowym warunkiem rozwoju i konkurencyjności przedsiębiorstwa. Gotowość pracowników do płynnej oraz rozważnej reakcji na otaczający świat, wynikającej chociażby z nowych wyzwań o charakterze zachowań przedsiębiorczych jest cechą łączącą przedsiębiorstwa z wymaganiami determinującymi przyszłość ich działania. Zmienne warunki, w których funkcjonują firmy wymuszają potrzebę szybkiego i elastycznego sprostania pojawiającym się wyzwaniom. Są jednocześnie podstawowym wyznacznikiem efektywności kapitału ludzkiego w zakresie umiejętności radzenia sobie z problemami komplikującymi byt rynkowy organizacji gospodarczych. Ponadto, aby móc antycypować zmiany, tworzyć oryginalne rozwiązania oraz systematycznie rozwijać właściwe kompetencje, żaden z pracowników nie może się opierać jedynie na katalogu posiadanych kompetencji. Dynamiczny charakter kompetencji i fakt, że odpowiadają one na konkretne sytuacje może decydować o efektywnym zarządzaniu przedsiębiorstwem i czynieniu go konkurencyjnym na europejskim czy światowym rynku. Z tego punktu widzenia indywidualny bilans atrybutów kompetencyjnych pracowników musi być regularnie aktualizowany w zależności od potrzeb narzucanych przez szybko zmieniające się otoczenia nowych technologii. Bez wątpienia wyznacznikiem takich parametrów kompetencyjnych pracownika w dzisiejszych czasach są zachowania przedsiębiorcze. Celem niniejszej pracy jest zatem ukazanie roli zachowań przedsiębiorczych pracowników w funkcjonowaniu przedsiębiorstw działających w otoczeniu nowych technologii. By zrealizować tak zdefiniowany cel autorzy wykorzystali w pierwszej kolejności analizę literatury przedmiotu. Analiza ta stanowi podstawę kolejnych dociekań naukowych opartych na przeprowadzanych badaniach własnych, szczególnie w odniesieniu do działań przedsiębiorczych pracowników firm w czasie popandemicznym. Autorzy, uzupełniając cel główny, zdecydowali się na sformułowanie pytań badawczych: jak prezentują się zachowania przedsiębiorcze pracowników w zmieniającym się otoczeniu nowych technologii? Jaka jest wzajemna zależność między dynamicznie zmieniającym się otoczeniem i przedsiębiorczością pracowników? Przedstawiony opis zachowań przedsiębiorczych pracowników ma charakter przeglądowy i daje podstawę do przeprowadzona dalszych badań.

2. Uwarunkowania funkcjonowania przedsiębiorczych pracowników w środowisku nowych technologii

Wymagania kompetencyjne stawiane pracownikom przez turbulentne otoczenie adresowane są do cech osobowościowych, wiedzy oraz umiejętności [4]. Zatem otoczenie w swoisty sposób wpływa na wizerunek i sposób działania pracowników, co widoczne jest gdy zmieniające się warunki ekonomiczne czy technologiczne wymuszają uzupełnienie referencyjnego wykazu kompetencji. Permanentne zmiany niosą dodatkowo za sobą możliwość pojawienia się mechanizmów zakłócających pełnienie powierzonych zadań. Zakłócenia przejawiające się chociażby w postaci imperatywów gospodarki globalnej, konkurencyjności i efektywności przedsiębiorstwa, elastyczności struktur organizacyjnych oraz identyfikacji właścicielskiej [5], mogą odgrywać rolę pobudzającą zmiany, podnosząc efektywności działania instytucji gospodarczych. Mogą przy tym jednakże przyczynić się do ograniczonego wykorzystania kompetencji kapitału ludzkiego lub wywołać nawet stany stresowe. Z badań laboratoryjnych przeprowadzonych przez S. Tokarskiego wynika, że stres u znacznej większości osób (3/4 osób badanych) wpływa mobilizująco. Stają się oni aktywniejsi, wykazują pozytywny stosunek do swojej pracy a pokonanie stresu przynosi im satysfakcję [6].

Nowa rzeczywistość gospodarcza, zmieniająca dotychczasowe struktury, sposoby i formy działania, stawia pracownikom przedsiębiorstw bardzo konkretne wymagania kompetencyjne. Próba wyjaśnienia koncepcji kompetencji kapitału ludzkiego w rozumieniu różnych autorów [7] dała asumpt do zakreślenia profilu współczesnego pracownika. Jest to osoba, która pełni rolę mediatora kulturowego i agenta zmian, ma zdolności do wieloaspektowej oceny sytuacji przedsiębiorstwa i jego otoczenia, radzi sobie w stresujących sytuacjach, posiada umiejętności kierowania własną efektywnością działania, ma zdolność do podejmowania ryzyka, jest kreatywna oraz przedsiębiorcza. Kompetentny personel potrzebuje dzisiaj nie tylko wizji, energii, autorytetu i zdolności kierowania emocjami. Aby móc osiągnąć wysoką efektywność pracy, kapitał ludzki firm powinien odznaczać się wieloma metacechami, rozpoznawalnymi w płaszczyźnie zachowań (podejmowanie inicjatyw w ryzykownych i niepewnych warunkach), percepcyjnej (zdolność do systematycznego zbierania i przetwarzania informacji), emocjonalnej (umiejętność stawiania się na miejscu innych ludzi i rozwiązywania konfliktów międzyludzkich) i symbolicznej (zdolność do percepcji przedsiębiorstwa jako systemu).

Oczywistość tych podstawowych wymagań staje się szczególnie wyraźna, gdy rozpatruje się je na tle tak radykalnych, wieloaspektowych i głębokich zmian jaką jest przemiana globalna. Ustalenie klasyfikacji bądź typologii, a także hierarchizacji owych przeobrażeń z punktu widzenia jakiegokolwiek kryterium jest bardzo złożone i trudne, gdyż w każdym przedsiębiorstwie procesy te ma swoją specyfikę. Niemniej jednak można przyjąć, że nowe reguły gospodarowania eksponują pewne tendencje, powodujące odmianę struktury zatrudnienia i przypisujące większe znaczenie kompetencjom. Pod rozwagę należy tu wziąć:

- ograniczenie liczby szczebli hierarchicznych;
- fuzje i wykup przedsiębiorstw, co wywołuje głębokie procesy restrukturyzacyjne;
- wzrost liczby małych przedsiębiorstw, bardziej wytrzymałych na zmiany ekonomiczne i tworzących więcej miejsc pracy;

- częste zastępowanie w dużych firmach służb funkcjonalnych o sztywnej strukturze zespołami projektowymi i zadaniowymi [8].

Warto w tym miejscu szerzej opisać model cech osobowościowych zaproponowany przez F. Havaleschkiego który uważał, że należy koncentrować się na tych cechach, których posiadanie ma związek z sukcesami ludzi w konkretnych zawodach. W swoim modelu, nazwanym „Głowa, serce, nogi", zaproponował 16 cech, które podzielił na 3 grupy, w zależności od wymagań zawodowych różne jest bowiem znaczenie posiadanych przez jednostkę cech psychologicznych. W sytuacji gdy praca wiąże się z częstym podejmowaniem decyzji szczególne znaczenie mają cechy „głowy" (elastyczne myślenie, kompleksowe spojrzenie na sytuację, abstrakcyjne myślenie, podejmowanie ryzyka, samokontrola), gdy ważne są umiejętności społeczne, liczą się cechy „serca" (empatia, kontakt społeczny, elastyczność społeczna, wsparcie, pewność, zaufanie), natomiast gdy mocno trzeba stać na ziemi, w trudnych do przewidzenia okolicznościach, potrzebne są cechy z grupy „nogi" (rywalizacja, wywieranie wpływu na ludzi, pewność siebie, niezależność, odporność na stres, energiczność). Aby można było odnosić sukcesy, potrzebne są cechy należące do wszystkich 3 grup, chociaż w różnym stopniu [9].

Przedstawione na gruncie literatury propozycje kompozycji cech składających się na „wzorcową" sylwetkę personelu przedsiębiorstwa poszerza się o kompetencje kształtowane przez procesy globalizacji gospodarki. Istotnym warunkiem jest zatem profesjonalne przygotowanie ludzi, czyli umiejętność prowadzenia działalności gospodarczej oraz duża elastyczność i szybkość działania. Ponadto muszą oni charakteryzować się otwartym sposobem myślenia, umiejętnością pracy w zespołach oraz szanowaniem różnic poglądów i osobowości. W przypadku funkcjonowania przedsiębiorstw na rynku wysokich technologii potrzebni są pracownicy o szerokim spojrzeniu. Większość badaczy problemów zarządzania kapitałem ludzkim widziałoby dzisiaj pracownika umiejącego sprawdzać się w kluczowych wymiarach, które obejmujących: globalne myślenie; przewidywanie przyszłych możliwości; kreatywne myślenie; akceptowanie różnic kulturowych; budowanie partnerstwa; sprzyjanie zmianom; entuzjazm dla nowych technologii; wspieranie konstruktywnych wyzwań; zapewnianie zadowolenia klientów; osiąganie przewagi konkurencyjnej; okazywanie osobistej siły, działanie zgodne z głoszonymi wartościami [10].

Z przedstawionego przeglądu koncepcji kompetencji przedsiębiorczych kapitału ludzkiego wynika, że era globalizacji i informacji ukształtuje i utrwali nowy stereotyp pracownika. Ch. Handy ujął to następująco:

> *Sądziliśmy, że wiemy jak kierować organizacją. Dziś wiemy jeszcze lepiej. Wiemy lepiej niż kiedykolwiek, że muszą one być dzisiaj jednocześnie globalne i lokalne, muszą być niekiedy małe a czasem wielkie, czasem scentralizowane, a przeważnie zdecentralizowane* [...] *Przedsiębiorstwa, które odnosiły sukces, potrafią sobie poradzić z paradoksami, czy jak to one nazywają – dylematami. Przedsiębiorstwa te muszą być ściśle zaprojektowane, ale elastyczne; muszą być zróżnicowane i jednocześnie zintegrowane, muszą działać na rynku masowym, ale troszczyć się o wiele nisz rynkowych; muszą wprowadzać nową technologię, ale też pozwalać pracownikom, by sami decydowali o swojej roli; muszą znaleźć sposoby, by produkować rzeczy różnorodne, dobrej jakości i zgodne z modą, wszystko po niskich kosztach; w skrócie – muszą znaleźć sposoby na pogodzenie sprzeczności, a nie wybierać między nimi* [11].

Dlatego, wobec determinującej roli otoczenia globalnego w kreowaniu sukcesu przedsiębiorstwa, pracownicy powinni uświadomić sobie konieczność rozwijania zdolności adaptacyjnych i wprowadzania nowych rozwiązań dostosowanych do pojawiających się szans i zagrożeń.

Wymagania stawiane pracownikom mogą stanowić impuls do zdiagnozowania ich na gruncie empirycznym i naukowym. Nie jest to jednak zadanie proste, na co mogą wskazywać z jednej strony, liczne przykłady opisujące sylwetkę pracownika, a z drugiej pewna dowolność w interpretacji obserwowanych zachowań jako przejawu określonych kompetencji osób, które odgrywają swoje role zawodowe. Dzieje się tak gdyż, identyfikacja kompetencji kapitału ludzkiego napotyka na swoiste trudności:

- żadne stanowisko pracy nie ma niezmiennej w czasie treści;
- w aktualnym kontekście ekonomicznym i technologicznym troska o elastyczność ma priorytetowy charakter;
- na takim stanowisku pracy zadania mogą być realizowane w różny sposób przez różne podmioty,
- nie należy przywiązywać wagi do nazw stanowisk pracy: z *tą samą nazwą, w tym samym przedsiębiorstwie wiązać się bowiem mogą różne kompetencje, ponieważ sytuacja i środowisko są odmienne* [12].

Z tego punktu widzenia nazwa nadawana określonej kompetencji w jednym przedsiębiorstwie niekoniecznie oznacza tę samą treść w innym. W analogiczny sposób specyficzna kompetencja może odwoływać się do różnych zachowań w zależności od sytuacji. Na przykład, tolerancja na odczuwany stres wobec konfliktów pracowniczych różni się od tolerancji na stres wobec niezadowolonych klientów lub od takiej samej kompetencji, kiedy chodzi o zaadaptowanie się do dużego przeciążenia pracą. Toteż w konsekwencji trzeba stwierdzić, że obok list zawierających ogólne cechy osobowościowe i umiejętności pracowników odczuwa się potrzebę opracowania zestawów kompetencji, które odpowiadałyby na zapotrzebowanie określonych uwarunkowań, strategii danego przedsiębiorstwa czy wymagań otoczenia globalnego. Wykazy powinny zawierać z pewnością przydatne kompetencje przy założeniu, że zostaną dokładnie i konkretnie zdefiniowane oraz zapewni się nim użyteczność praktyczną.

3. Przedsiębiorczość jako atrybut kompetentnego pracownika

Z wielu list charakteryzujących cechy kompetentnego pracownika na specjalną uwagę zasługuje jego przedsiębiorczość. Jest ona istotnym elementem kompetencji, która pozwala na wykorzystanie okazji do praktycznego wdrożenia nowych rozwiązań zmieniających istniejący stan rzeczy. Istota przedsiębiorczości tkwi w zdolności tworzenia nowych sposobów zaspokajania potrzeb oraz umiejętności kreowania nowego popytu. Znaczenie przedsiębiorczości w warunkach zmieniającej się gospodarki, gdzie personel napotyka na wiele trudności o charakterze zarządczym, wydaje się bardzo ważne. Przedsiębiorczość bowiem uznawana jest powszechnie za główną siłę rozwojową każdej organizacji, wpisaną w jej system zarządzania. Przedsiębiorczość jest jedną z najważniejszych cech i umiejętności skutecznie działających pracowników, będących odzwierciedleniem nowego systemu gospodarowania, w którym kierujący uświadamiają sobie, że nowy system wymaga nowych kompetencji. Przedsiębiorczość powinna stać się zatem dzisiaj główną siłą kreatywną każdej firmy wpisaną na trwałe w jej system zarządzania i kulturę organizacyjną. J. Penc podkreślał, że *bez przedsiębiorczości, bez*

elastycznego reagowania na wyzwania rynku, bez wolności wyboru i samodzielnych działań na własny rachunek i na własną odpowiedzialność polska gospodarka pozostanie rynkową tylko z nazwy i nie wyjdzie na drogę efektywności, racjonalności i oszczędności [13]. Jednak stosowanie się do podanych przez niego reguł formułowanych w związku z potrzebą działań przedsiębiorczych musi uwzględniać następujące uwarunkowania: kształtowanie przedsiębiorczej orientacji firmy, popieranie nowych pomysłów, popieranie prób i błędów, gotowość do podejmowania ryzyka, ułatwienie dostępu do zasobów finansowych i rzeczowych przedsiębiorstwa, budowę systemu sponsorowania przedsięwzięć, popieranie pracy zespołów wielodyscyplinarnych, ocena efektów poszczególnych przedsięwzięć z dłuższej perspektywy czasowej, przyjęcie zasady, że każdy pracownik może być kreatorem nowego przedsięwzięcia, przygotowanie odpowiedniego systemu motywacyjnego, wspieranie działań przedsiębiorczych przez naczelne kierownictwo, tworzenie elastycznych struktur [14].

Pojęcie przedsiębiorczości nie należy do jednoznacznie interpretowanych i nie ma ostatecznie ustalonego zakresu stosowania. Na problemy związane z konceptualizacją terminu *przedsiębiorczość* zwrócił uwagę K. Wach, stwierdzając, że zarówno w języku potocznym, jak i naukowym występuje mnóstwo różnych interpretacji tego pojęcia [15]. Mimo wielu opracowań naukowych nie doprowadzono do wypracowania spójnej i akceptowanej przez wszystkich definicji przedsiębiorczości, a w literaturze można spotkać rozmaite próby interpretacyjne. Przedstawione w literaturze przedmiotu definicje przedsiębiorczości i przedsiębiorcy można klasyfikować według trzech podstawowych kategorii:

- definicje przedstawiane na podstawie funkcjonalnego opisu działań przedsiębiorczych, gdzie odnoszą się do ekonomicznych wymiarów wykorzystania zasobów;
- definicje skonstruowane na cechach osobowościowych przedsiębiorców;
- definicje odwołujące się do specyficznego sposobu zachowania poszczególnych pracowników, gdzie wykorzystywane są posiadane umiejętności [16].

Pierwsza z wymienionych kategorii pojęciowych przedsiębiorczości przyjmuje za punkt wyjścia działania związane z podejmowaniem aktywności gospodarczej, co w konsekwencji odnosi się do procesu, w którym następuje inicjowanie i wprowadzanie zmian. Dał temu wyraz P. Drucker, który stwierdził, że *dzisiejsze przedsiębiorstwa, zwłaszcza duże, nie mają szans na przetrwanie w dobie gwałtownych innowacji, jeśli nie zdobędą kompetencji przedsiębiorczych* [17]. Autor dalej zaznaczył, że *przedsiębiorca zawsze poszukuje zmiany, reaguje na nią i wykorzystuje ją jako okazję i chociaż zazwyczaj sam jej nie powoduje, to stanowi ona dla niego bodziec do aktywności* [17].

Dla ilustracji pierwszego zbioru definicji można przytoczyć określenie P. Zbierowskiego, który podkreślał, że klasyczny przedsiębiorca to osoba, która zmierza do stworzenia przedsięwzięcia gospodarczego i świadomie planuje jego przyszłość [18]. Funkcjonalne rozumienie przedsiębiorczości dostrzegł także H. Bieniok i wskazał, że przedsiębiorca to człowiek, który tworzy przedsiębiorstwo lub obejmuje nad nim kontrolę, dostrzegając sprzyjające interesom okazje oraz wykorzystuje je do dokonania zmian i postępu w gospodarce [19]. Podobną interpretację przedstawił K. Safin, który wyodrębnił ujęcie atrybutowe i czynnościowe przedsiębiorczości. Pierwsze ujecie przyjmuje przedsiębiorczość jako zestaw cech, a drugie – wyjaśnia istotę przedsiębiorczość z punktu widzenia działań związanych z podejmowaniem działalności gospodarczej [20]

Podobną konwencję interpretacyjną wykorzystał R.W. Griffin, który potraktował przedsiębiorczość jako proces tworzenia oraz prowadzenia nowego przedsiębiorstwa, zauważając jednak, że organizowanie i prowadzenie działalności gospodarczej związane jest nierozerwalnie z ryzykiem oraz nastawieniem na sukces [21]. Podobnego zdania jest S. Tokarski, który zaznaczył, że *przedsiębiorczość manifestuje się w sposób najbardziej spektakularny w dostrzeganiu i wykorzystywaniu okazji wyłaniających się za sprawą okoliczności zewnętrznych* [6].

Tak rozumiana przedsiębiorczość wiąże się ze sztuką kalkulacji, ale kalkulacji dokonywanych kompleksowo z uwzględnieniem otwartości systemu, niepewności i wielkości parametrów. W tym rozumieniu praktycznym wyznacznikiem działań przedsiębiorczych jest uwzględnienie w kalkulacji takich czynników, jak: umiejętność i zdolność ludzi do podejmowania wyzwań związanych z wprowadzaniem zmian, elastyczność w działaniu i przyjmowaniu nowych wzorców oraz możliwość uczenia się nowych zachowań, które ułatwiają rozwiązywanie problemów przy użyciu posiadanych środków. Dodatkowo, H. Landström wskazał, że przedsiębiorczość to proces tworzenia nowych przedsięwzięć i odkrywania oraz wykorzystywania nadarzających się okazji [22].

Definicje przedsiębiorczości składające się na drugą kategorię tworzą szeroką i niejednorodną strukturę. Wyznaczenie typowych cech osób przedsiębiorczych należy do zagadnień bardzo rozległych, gdyż przedsiębiorca posiada różne cechy w zależności od tego, czy charakteryzują one liderów gospodarki, właścicieli firm, innowatorów, czy też administratorów różnych dziedzin życia społecznego. Interpretacje terminu w tej konwencji terminologicznej można odnaleźć u S. Tokarskiego, który zidentyfikował osobę przedsiębiorczą jako osobę rzutką, energiczną, posiadającą ducha inicjatywy oraz *wkraczającą do akcji tam, gdzie stare przyzwyczajenia i niechęć do ryzyka dyktują powtarzanie w kółko tego, co stało się wygodnym przyzwyczajeniem* [6]. Według W. Szymańskiego w literaturze skupia się uwagę przede wszystkim na cechach tych przedsiębiorców, którzy są zdolni wprowadzać zmiany pozbawione jakiejkolwiek regularności. Jego zdaniem przedsiębiorca to marzyciel, który zawsze poszukuje zmian, reaguje na nie i wykorzystuje je jako okazję. Odznacza się potrzebą samorealizacji oraz wiarą w możliwości zrealizowania czegoś nowego. Działa w warunkach, w których niezbędne jest wykorzystanie intuicji [23].

Przykład reprezentujący drugą grupę terminologiczną przedstawił również H. Bieniok. Autor ten przedsiębiorczą sylwetkę scharakteryzował przez: silną potrzebę osiągnięć, wytrwałość i determinację, potrzebę autonomii, umiarkowaną skłonność do ryzyka, zaufanie do własnych zdolności kontrolowania zdarzeń, zorientowanie na okazję, umiejętności przyjmowania porażki i wyciągania z niej wniosków, optymizm życiowy i entuzjazm w pracy, kreatywność i innowacyjność oraz przez zdolności przywódcze [19].

Znaczną część trafnie dobranych cech osoby przedsiębiorczej zawiera propozycja W.G. Nickelsa, który wymienił:

- dojrzałość w kierowaniu swoją osobą, czyli zdolność do wypracowania samodyscypliny przy jednoczesnym poczuciu komfortu psychicznego;
- wiarę we własne siły, sprowadzającą się do umiejętności przeforsowania własnych pomysłów oraz podtrzymywania uczucia entuzjazmu zawodowego;
- orientację na działanie widzianą jako zdolność do realizacji i dostosowania do potrzeb wypracowanych projektów;
- energię psychiczną i fizyczną do wykonywania ciężkiej i długotrwałej pracy;

- liczenie się z ryzykiem.

W przedstawionej propozycji podziału cech osoby przedsiębiorczej autor dostrzegł zarówno zdolności do przejawiania wiary we własne pomysły, jak też umiejętności wykorzystywania pojawiających się szans, co wiąże się bezpośrednio z ich praktyczną realizacją, do której przedsiębiorca potrafi się przygotować psychicznie i fizycznie [24].

Dokonuje się też podziału cech przedsiębiorcy na cechy wrodzone i nabyte[3]. Wśród tych pierwszych na uwagę zasługuje bystrość umysłu, inteligencja, energia i wrodzone przywództwo, odwaga w podejmowaniu trudnych i ryzykownych decyzji oraz intuicja. Do cech nabytych z kolei zalicza się zdolność szybkiej oceny sytuacji i przewidywania przyszłych zdarzeń. Według U. Nowackiej najważniejszymi cechami osoby przedsiębiorczej są: nastawienie na osiągnięcia, inicjatywa, pragmatyzm, inteligencja emocjonalna, duża doza indywidualizmu oraz rozbudowana potrzeba autonomii [25]. Inny autor, M. Strużycki, stwierdził, że podstawą zachowań przedsiębiorczych są następujące cechy: wiedza, intelekt, charakter, odwaga, zapał, optymizm, zdolność do ryzyka, wytrwałość, umiejętność doboru pracowników i tworzenie organizacyjne całości [26].

Przedsiębiorczego pracownika nie identyfikuje się tylko z jego cechami osobowościowymi, które można mieć, ale nie wykorzystywać w praktycznym działaniu. Stąd przedsiębiorczego człowieka można przedstawić za pomocą określonych cech psychofizycznych, niemniej jednak jego przedsiębiorczość wyraża się przede wszystkim w praktycznym działaniu, gdzie cechy te mogą przyczynić się do osiągnięcia racjonalnego efektu tegoż działania.

W tym miejscu należy przejść do trzeciej wyeksponowanej kategorii definicji przedsiębiorczości, rozumianej jako specyficzny sposób działania. Przedsiębiorczość to taki rodzaj zachowania, który charakteryzuje się gotowością do rozpoczynania i kontynuacji niewypróbowanych jeszcze projektów, do rozpoznawania oraz wykorzystywania pojawiającej się szansy osiągnięcia czegoś wartościowego w sytuacji, gdy inni tkwią jeszcze w rutynie [6]. Pojęciem przedsiębiorczości posługujemy się również w odniesieniu do działań, które mają spowodować jakiś stan rzeczy oraz do decyzji lub wyboru, z których ma coś wynikać w przyszłości. W rozwoju tych działań autor rozróżnił cztery fazy:

- faza pierwsza – wybór kierunku działania. Bierze się w niej pod uwagę niesztablonowe i ryzykowne możliwości;
- faza druga – nabieranie wprawy. Szuka się nowych metod szybszego osiągnięcia postępów w działaniu;
- faza trzecia – nabieranie rutyny. Przynosi ona komfort działania w sposób biegły i wprawny;
- faza czwarta – regres. Może się ona niekiedy pojawiać na skutek przestarzałych technologii, zanikającego popytu czy poczucia nudy, znużenia i niezadowolenia osób [6].

Z kolei przez M. Kondracką-Szalę i J. Malinowską przedsiębiorczość rozumiana jest jako poszukiwanie i znajdywanie nowych możliwości prowadzących do wyznaczonych celów, znajdowanie nowych możliwości oraz rozpoczęcia udanego przedsięwzięcia [27]. Definicję przedsiębiorczości omówiła podobnie U. Nowacka, określając ją jako aktywne zachowanie się osób wobec oczekujących zadań. Aktywność ta sprowadza się do umie-

[3] Według K. Safina w literaturze przedmiotu nie ma zgodności co do natury cech przedsiębiorcy, jednak powszechnie uznaje się, że określone warunki pozwalają się tym cechom uzewnętrznić [20].

jętności wykorzystania istniejących warunków oraz podejmowania działań dla zmiany tych warunków w pożądanych kierunkach [28].

Przedsiębiorczość jest też rozumiana jako działalność wyróżniająca się aktywnością i dynamizmem, innowacyjnością, poszukiwaniem zmian i reagowaniem na nie, postrzeganiem szans oraz ich wykorzystaniem, niezależnie od posiadanych zasobów, oraz gotowością do podejmowania ryzyka, której głównym motywem działania jest pomnażanie kapitału [19]. W poglądach nad przedsiębiorczością dostrzega się wyeksponowany tzw. geniusz przedsiębiorczości, który jest specyficznym typem zdolności, na którego powstanie złożyć się musi wiele przyczyn zewnętrznych i wewnętrznych. Przedstawia się tu trzy właściwości kierownika-przedsiębiorcy: zdobywca (posiada zdolność narzucania planów, wytrwałość, energię), organizator (wykorzystuje umiejętność zarządzania ludźmi i rzeczami) i kupiec (musi umieć budzić zainteresowanie i zaufanie) [19].

W pewnym stopniu do powyższej propozycji przybliża się koncepcja B. Wawaka, który określił przedsiębiorczość jako proces aktywnego reagowania na wszelkie zmiany, jakie niosą procesy globalizacji stanowiące wyzwanie dla przedsiębiorców. Pomocne w tym zakresie będzie wykształcenie wśród pracowników nowych kompetencji charakterystycznych dla przedsiębiorczego profesjonalisty, takich jak: zdolność do wprowadzania zmian; umiejętność uczenia się od innych i adaptowanie ich rozwiązań do sytuacji, w których się dana firma znajduje; uczenia się przez całe życie; umiejętność komunikowania się oraz umiejętność kierowania ludźmi [29]. Obraz tego zjawiska jest szczególnie widoczny w warunkach kryzysowych [30], w których pracownicy ukierunkowani na otwartość na zmiany, wzbogacanie wiedzy i umiejętności mogą wnieść istotny wkład w powstanie dynamicznych i trwałych gospodarczo przedsiębiorstw. Wiele zjawisk kryzysowych można zatem lepiej opanować, sprzyjając rozwojowi zachowań przedsiębiorczych w dostosowaniu się do współczesnego środowiska nowych technologii.

4. Podsumowanie

Jak można zauważyć, istnieje wiele badań dotyczących przedsiębiorczości pracowników. Jednak nasze badania pozwoliły ukazać wybrane paradygmaty zachowań przedsiębiorczych pracowników firm funkcjonujących w otoczeniu nowych technologii. Ponadto, badania pozwoliły ustalić zachowania przedsiębiorcze pracowników w zmieniającym się otoczeniu nowych technologii. Zidentyfikowaliśmy również, jak przedsiębiorczość pracowników jest determinowana przez wyzwania nowego otoczenia przedsiębiorstw. Można zatem stwierdzić, że firmy powinny być bardziej aktywne w identyfikowaniu nowatorskich zachowań przedsiębiorczych pracowników poprzez analizę sygnał płynących z otoczenia nowych technologii.

Przedsiębiorczość stanowi w pracy każdego z nas fundamentalny element działania nastawionego na zmiany [31]. Umożliwia praktyczną realizację nowatorskich koncepcji prowadzących do tego, że każdy szuka takiego wykorzystania swojej pracy, która przyniesie mu najwięcej korzyści. Promowanie zachowań przedsiębiorczych wśród kadry powinno być ukierunkowane na pracowników w ten sposób, aby aktywizować ich działalność, która wymaga swobody pozwalającej na eksperymentowanie oraz urzeczywistnienie czegoś nowego [32]. Stąd przedsiębiorczość siłą rzeczy, narusza *status quo* funkcjonowania przedsiębiorstwa tworząc nierzadko twórczy chaos lub „kreatywną konstrukcję" w ten sposób, aby kompetencją wyróżniającą zdolnych pracowników była umiejętność przewidywania nowych realiów oraz tworzenia wizji na drodze praktycznych ekspery-

mentów. A zatem konsekwencją przedsiębiorczego zarządzania jest podejmowanie takich działań, które w swojej treści zakładają powodzenie podjętych zmian będących odpowiedzią na pojawiające się szanse w środowisku nowych technologii.

Przeprowadzona analiza dotycząca wpływu zachowań przedsiębiorczych na funkcjonowanie współczesnych przedsiębiorstw środowiska nowych technologii wymaga zmiany sposobu myślenia u pracowników. Biorąc pod uwagę poruszane w artykule zagadnienia, należy stwierdzić, że personel kierowniczy kreujący jakość polityki kadrowej powinien reagować elastycznie na pojawiające się wyzwania współczesnego otoczenia, zwłaszcza technologicznego [33]. Przyjmując bowiem, że kadra kierownicza podejmuje szereg decyzji w oparciu o dane zewnętrzne, a więc takie, na które nie ma większej możliwości oddziaływania, ich uwaga powinna zostać skupiona na zapewnieniu wysokiej konkurencyjności przedsiębiorstw poprzez racjonalne decyzje podejmowane w zakresie adaptacji wymagań środowiska nowych technologii. Konsekwencją przyjętego założenia jest zwrócenie uwagi na umiejętności pracowników w adaptacji nowych umiejętności w zakresie wykorzystywania spółczesnych technologii. Podejście to jest szczególnie istotne w tworzeniu standardów prowadzonej polityki personalnej, gdyż każda podejmowana decyzja kadrowa jest wypadkową czynników, jakie wynikają z logiki danego punktu odniesienia.

Literatura

1. European Innovation Scoreboard 2022, https://op.europa.eu/en/publication-detail/-/publication/f0e0330d-534f-11ed-92ed-01aa75ed71a1/language-en/format-PDF/source-272941691 [data dostępu: 4.02.2023].
2. Lewicka D., *Zarządzanie kapitałem ludzkim w polskich przedsiębiorstwach*, PWN, Warszawa 2010.
3. Morawski M., *Zarządzanie wiedzą. Organizacja – system – pracownik*, Wyd. Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 2006.
4. Kęsy M., *Kształtowanie kompetencji menedżerskich personelu medycznego w szpitalach*, Wyd. Uniwersytetu Jagiellońskiego, Kraków 2013.
5. Suszyński C., *Restrukturyzacja przedsiębiorstw*, PWE, Warszawa 1999.
6. Tokarski S., *Kierownik w organizacji*, Difin, Warszawa 2006.
7. Sojda P., Lepiarska E., *Zarządzanie kompetencjami pracowników przedsiębiorstwie usługowym*, Zeszyty Naukowe Politechniki Częstochowskiej, Nr 34, 2019.
8. Koźmiński A.K, Piotrowski W., *Zarządzanie. Teoria i praktyka*, PWN, Warszawa 2023.
9. Rakowska A., *Kompetencje menedżerskie kadry kierowniczej we współczesnych organizacjach*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 2007.
10. Armstrong M., *A Handbook of human resource management*. Practice, Kogan Page Limited, London 2006.
11. Handy Ch., *Wiek paradoksów. W poszukiwaniu sensu przyszłości*, Dom wydawniczy ABC, Warszawa 1996.
12. Levy-Leboyer C., *Kierowanie kompetencjami*, Poltext, Warszawa 1997.
13. Penc J., *Zarządzanie w biznesie i administracji. Zasady działania operacyjnego i strategicznego*, Wydawnictwo AHE w Łodzi, Łódź 2017.
14. Pietruszyński P., *Pokolenie informacji zmieni środowisko biznesu*, Computerworld, 20, 2015.
15. Wach K., *Przedsiębiorczość jako czynnik rozwoju społeczno-gospodarczego przegląd literatury*, Przedsiębiorczość – Edukacja, Vol. 11, 2015.
16. Glinka B., Gudkova S., *Przedsiębiorczość*, Wolters Kluwer, Warszawa 2011.
17. Drucker P., *Menedżer skuteczny*, Nowoczesność, Warszawa 1994.
18. Zbierowski P., *Determinanty intencji przedsiębiorczej studentów – wyniki badań*, Horyzonty Wychowania, 13(26), 2014.

19. Bieniok H., *Podstawy zarządzania przedsiębiorstwem*, Wyd. AE, Katowice 2003.
20. Safin K., *Zarządzanie małą firmą*, Wyd. AE we Wrocławiu, Wrocław 2003.
21. Griffin R.W, *Podstawy zarządzania organizacjami*, PWN, Warszawa 2022.
22. Landström H., *Pioneers in Entreprenership and Small Business Research*, Springer, Heidelberg 2010.
23. Szymański W., *Przedsiębiorca i przedsiębiorczość*, [w:] Tkaczyk T.P. (red.), *Przedsiębiorczość a strategie konkurencji*, SGH, Warszawa 2000.
24. Nickels W.G., *Zrozumieć biznes*, Dom wydawniczy „Rytm", Warszawa 2000.
25. Nowacka U., *Entrepreneurship as a Key Competence-Implications for the Education Process in Poland*, Society, Integration, Education. Proceedings Of The International Scientific Conference, Vol. 1, 2015.
26. Strużycki M., *Przedsiębiorstwo a rynek*, PWE, Warszawa 1992.
27. Kondracka-Szala M., Malinowska J., *Studiowanie – czas na rozwijanie postawy przedsiębiorczej nauczycieli dziecka młodszego? Analiza porównawcza programów kształcenia w zakresie przedsiębiorczości na polskim i fińskim uniwersytecie*, E-mentor, 3(65), 2016.
28. Nowacka U., *Kreowanie przedsiębiorczości intelektualnej*, Prace Naukowe Akademii im. Jana Długosza w Częstochowie, z. 5, 2011.
29. Wawak T., *Zarządzanie w szkołach wyższych i innowacje w gospodarce*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2016.
30. Audretsch D.B., Heger D., Veith T., *Infrastructure and entrepreneurship*, Small Bus Econ, nr 44, 2015.
31. Di Vaio A., Hassan R., Chhabra M., Arrigo E., Palladino R., *Sustainable entrepreneurship impact and entrepreneurial venture life cycle: A systematic literature review*, Journal of Cleaner Production, Volume 378, 2022.
32. Crawford G.C., Skorodziyevskiy V., Frid C.J., Nelson T.E., Booyavi Z., Hechavarria D.M., Li X., Reynolds P.D., Teymourian E., *Advancing Entrepreneurship Theory Through Replication: A Case Study on Contemporary Methodological Challenges, Future Best Practices, and an Entreaty for Communality*, Entrepreneurship Theory and Practice, 46(3), 2022.
33. Dimov D., Maula M., Romme A.G.L., *Crafting and Assessing Design Science Research for Entrepreneurship*, Entrepreneurship Theory and Practice, 0(0), 2022.

Wpływ zachowań przedsiębiorczych pracowników na funkcjonowanie przedsiębiorstw w dostosowaniu się do otoczenia nowych technologii

Streszczenie

W dobie turbulentnego i mniej przewidywalnego otoczenia biznesowego, umiejętność efektywnego wykorzystania potencjału przedsiębiorczego pracowników nabiera szczególnego znaczenia. Istnieje potrzeba zmniejszania niszy kompetencyjnej pracowników, a problem niedopasowania kompetencyjnego do wymagań współczesnego otoczenia nowych technologii stał się zauważalny. Warunki działania firm determinują posiadanie nowych umiejętności potrzebnych do osiągnięcia sukcesu rynkowego. Gotowość pracowników do płynnej oraz rozważnej reakcji na otaczający świat, wynikającej chociażby z nowych wyzwań o charakterze zachowań przedsiębiorczych jest cechą łączącą przedsiębiorstwa z wymaganiami determinującymi przyszłość ich działania. Zmienne warunki, w których funkcjonują firmy wymuszają potrzebę szybkiego i elastycznego sprostania pojawiającym się wyzwaniom. Są jednocześnie podstawowym wyznacznikiem efektywności kapitału ludzkiego w zakresie umiejętności radzenia sobie z problemami komplikującymi byt rynkowy wielu firm funkcjonujących w otoczeniu nowych technologii.

Słowa kluczowe: przedsiębiorczość, pracownik, otoczenie, kompetencje

Influence of entrepreneurial behaviour of employees on the functioning of enterprises in adapting to the environment of new technologies

Abstract

In a turbulent and less predictable business environment, the ability to effectively use the entrepreneurial potential of employees becomes particularly important. There is a need to reduce the competence niche of employees, and the problem of competence mismatch to the requirements of the modern environment of new technologies has become noticeable. The conditions of operation of companies determine the possession of new skills needed to achieve market success. The readiness of employees to react smoothly and prudently to the surrounding world, arising for example from new challenges of the nature of entrepreneurial behaviour, is a feature that links companies with the requirements that determine the future of their operations. The changing conditions in which companies operate to force the need to respond quickly and flexibly to emerging challenges. At the same time, they are a fundamental determinant of the effectiveness of human capital in terms of the ability to deal with problems that complicate the market life of many companies operating in the environment of new technologies.

Keywords: entrepreneurship, employee, environment, competences

Bożena Gajdzik[1]

Adaptacyjność przedsiębiorstw w Polsce w ostatnich trzech dekadach. Od restrukturyzacji naprawczej po inteligentne technologie

1. Wprowadzenie

Determinantą wyboru tematu były zmiany poczynione w przedsiębiorstwach produkcyjnych w Polsce w ciągu ostatnich trzech dekad. Za początek przemian wybrano rok 1989, to wtedy w Polsce, pakietem reform systemowych, zainicjowano przejście od gospodarki centralnie sterowanej do gospodarki wolnorynkowej. Można przyjąć, że od 1990 roku budzimy w Polsce przedsiębiorczość, a zrealizowana w latach 90. i późniejszych prywatyzacja przedsiębiorstw państwowych (dawnych kombinatów) radyklanie zmieniła rynek przemysłowy w Polsce. Cały okres transformacji systemu gospodarczego do gospodarki wolnorynkowej w Polsce, to wiele programów głębokiej i radyklanej restrukturyzacji poszczególnych branż przemysłu, które z uwagi na radykalność zmian nazywano programami naprawczymi. Restrukturyzacja była początkiem przemian na drodze do współczesnego sukcesu przedsiębiorstw produkcyjnych. Obecnie przedsiębiorstwa wprowadzają zmiany, które budują smart manufacturing na bazie technologii Czwartej Rewolucji Przemysłowej. Po trzech dekadach od restrukturyzacji, przedsiębiorstwa w Polsce mają szansę uczestniczyć we współtworzeniu Industry 4.0 (Przemysłu 4.0), w skrócie I 4.0.

Praca jest rezultatem rozważań konceptualnych, które rozbudowano o przykłady (dane statystyczne) dotyczące transformacji przemysłu. Celem pracy było nakreślenie drogi, jaką musiały pokonać polskie przedsiębiorstwa, od restrukturyzacji naprawczej w latach 90., aby móc prowadzić działalność gospodarczą w warunkach wolnorynkowych, do chwili obecnej. Zgodnie z przyjętym tematem pracy, proces restrukturyzacji przedsiębiorstw w Polsce przedstawiono w kontekście ich adaptacyjności do dynamicznych warunków prowadzenia biznesu. Pojęcie adaptacyjności zostało użyte w rozumieniu przystosowania się obiektu do warunków środowiska (S łownik PWN) [1]. Metodologia praca opiera się na: (1) wiedzy eksperckiej o restrukturyzacji przemysłu tradycyjnego, (2) analizie statystyczne przemian, jak (3) opracowaniu konceptualnej, holistycznej ścieżki restrukturyzacji polskich przedsiębiorstw w okresie ostatnich trzech dekad. Ścieżkę wyznaczają następujące etapy (1) restrukturyzacja naprawcza, w podziale na okres przedakcesyjny Polski do UE i okres akcesyjny, (2) restrukturyzacja rozwojowa, (3) restrukturyzacja kryzysów, (4) restrukturyzacja cyfrowa, (5) restrukturyzacja zielona (głęboka). Poszczególne etapy na ścieżce restrukturyzacji są sekcjami w niniejszej pracy. Ponadto do podsumowania dołączono tabele z własnymi definicjami poszczególnych typów restrukturyzacji.

2. Do początku, czyli restrukturyzacji naprawczej przemysłu

Po 1989 roku polski rząd przyjął pakiet reform systemowych. Lata 90. ubiegłego wieku nacechowane były radykalnością przemian gospodarczych. Polska z gospodarki centralnie sterowanej transformowała do gospodarki wolnorynkowej. Czym różniła się krajowa re-

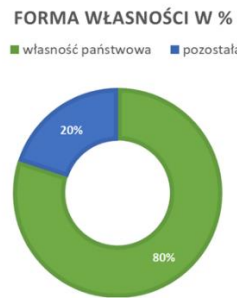
[1] Bozena.Gajdzik@polsl.pl, Politechnika Śląska, Gliwice, ORCID: https://orcid.org/0000-0002-0408-1691.

strukturyzacja przemysłu od restrukturyzacji przemysłów w innych krajach? Przede wszystkim tym, że kraje spoza Europy Środkowo-Wschodniej funkcjonowały w warunkach gospodarki rynkowej, gdy tymczasem Polska, jak i kraje sąsiednie z byłego bloku socjalistycznego, musiały dopiero stworzyć warunki dla przedsiębiorczości. Jak należy rozumieć restrukturyzację? Przyjmuje się, że pojęcie restrukturyzacji odnoszone jest do zmian, których realizacja przyczynia się do radyklanej przebudowy przedsiębiorstwa. Przymiotnik „radykalna" pochodzi z języka łacińskiego „radix" i oznacza „korzeń". Radykalna restrukturyzacja wymaga ponownego przejścia do „korzenia rzeczy", czyli odszukania istoty prowadzonej działalności gospodarczej [2]. Restrukturyzacja nie skupia się na fragmentach biznesu lecz rozpatruje go całościowo. W Polsce, w latach 90., fragmentaryczne zmiany w funkcjonowaniu przedsiębiorstw byłyby niewystarczające, trzeba było porzucić wszystkie dotychczasowe zasady i wprowadzić zupełnie nowe i oryginalne rozwiązania dla przedsiębiorczości, aby przedsiębiorstwa mogły samodzielnie podejmować decyzje i prowadzić działalność gospodarcza na własne ryzyko. Restrukturyzacja, w Polsce, była, w dużym stopniu, odkrywaniem na nowo biznesu i uczeniem się przedsiębiorczości. Pomimo, że w tamtym okresie działały firmy prywatne, to ich znaczenie było marginalne dla gospodarki. Według danych statystycznych (cyt. za: [3, s. 330]) własność inna niż państwowa stanowiła wówczas tylko 20%, przy czym w tej mniejszości była również własność spółdzielcza.

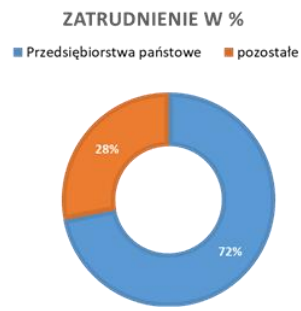
Celem restrukturyzacji, w ogólnym rozumieniu, jest stworzenie warunków do wzrostu wartości przedsiębiorstwa [4], co w warunkach transformacji gospodarczej krajów Europy Środkowo-Wschodniej, oznaczało właściwie sprywatyzowanie przedsiębiorstw. W krajach wysokorozwiniętych wartość biznesu była umacniana i poszerzana zgodnie z zasadą (ang.) *Social Responsibility* (SR), że (ang.) *business in not only business*, a tymczasem w Polsce skoncentrowano się na poszukiwaniu inwestorów strategicznych dla prywatyzowanych przedsiębiorstw. Przejdźmy do danych statystycznych przedstawiających zastaną sytuację w Polsce w latach 90. Rok 1989 (rys. 1), ponad 80% ogółu przedsiębiorstw stanowiły przedsiębiorstwa państwowe, sektor prywatny był mniejszościowy (tylko 20% firmy, najczęściej drobnych). Sektor podmiotów państwowych zatrudniał ponad 72% ogólnej liczby pracowników, a sektor prywatny tylko 28% (rys. 2). Największe przedsiębiorstwa państwowe były wielkimi kombinatami, które podlegały bezpośrednio resortom i właściwym ministerstwom. Na rysunku 3 przedstawiono wytworzoną wartość dodaną przez przedsiębiorstwa należące do Państwa. Przedsiębiorstwa państwowe wytwarzały wówczas ponad 82% PKB. Zatem tylko wielkie podmioty gospodarcze, należące do Państwa, liczyły się w gospodarce centralnie sterowanej. Szczególne znaczenie przywiązywano do podmiotów przemysłu tradycyjnego, czyli górnictwa i hutnictwa. Wokół tych sektorów rozwijały się miejsca pracy i sfera socjalno-bytowa (osiedla mieszkaniowe, domy wczasowe, domu kultury, tymczasowe formy zakwaterowania typu dom górnika, dom hutnika, stołówki zakładowe i wiele innych organizacji).

Restrukturyzacja przemysłu w Polsce miała charakter nakazowy (odgórny, ustawowy). W procesie restrukturyzacyjnym przemysłu w Polsce można było wyróżnić dwa etapy:

1. Etap przedakcesyjny Polski do Unii Europejskiej.
2. Etap akcesyjny. Zakres akcesyjny był silnie skoncentrowany na budowaniu opłacalności ekonomicznej prowadzonej działalności gospodarczej (ang. *viability*).



Rysunek 1. Struktura własnościowa przedsiębiorstw w Polsce przed restrukturyzacją naprawczą
Źródło: opracowanie własne na podstawie [2]

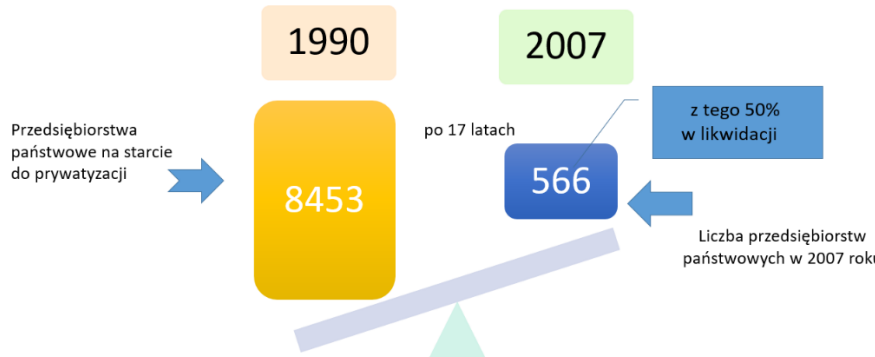


Rysunek 2. Miejsce zatrudnienia w Polsce przed restrukturyzacją naprawczą. Źródło: opracowanie własne na podstawie [2]

Rysunek 3. Miejsce wytwarzania wartości dodanej w Polsce przed restrukturyzacją naprawczą.
Źródło: opracowanie własne na podstawie [3]

Zakres restrukturyzacji przemysłu był bardzo rozległy, nie było sektora przemysłu, który nie były zrestrukturyzowany. Restrukturyzacja była realizowana na wielu poziomach organizacji kraju: tj. przedsiębiorstw (mikro ujęcie), branż, np. restrukturyzacja przemysłu wydobywczego, hutniczego, stoczniowego, odzieżowego, włókienniczego itd.) i gospodarki (makro). Skala przemian była ogromna od naprawy działalności statusowej tzw. (ang.) *core business* po likwidację działalności (ang.) *non-core*, np. zakładowych orkiestr dętych, chórów i innych organizacji, które działały przy przedsiębiorstwach państwowych. Pojawił się (ang.) *outosuring* jako nowa forma reorganizacji w odniesieniu do wielu obszarów działalności pozabiznesowej, czyli wydzielania aktywności, która nie mieściła się w działalności statutowej lub dalsze jej prowadzeni było nieopłacalne.

Restrukturyzacja nie byłaby możliwa bez prywatyzacji przedsiębiorstw państwowych. Prywatyzacja największych przedsiębiorstw w Polsce trwała ponad 7 lat. Na wejściu do programu było prawie 8500 podmiotów, w rezultacie w 2007 roku sprywatyzowano ponad 500 podmiotów, z tego połowa była przeznaczona do likwidacji (rys. 4). Zlikwidowano podmioty, które nie były w stanie pomyślnie przejść restrukturyzacji lub którymi nie byli zainteresowani inwestorzy.

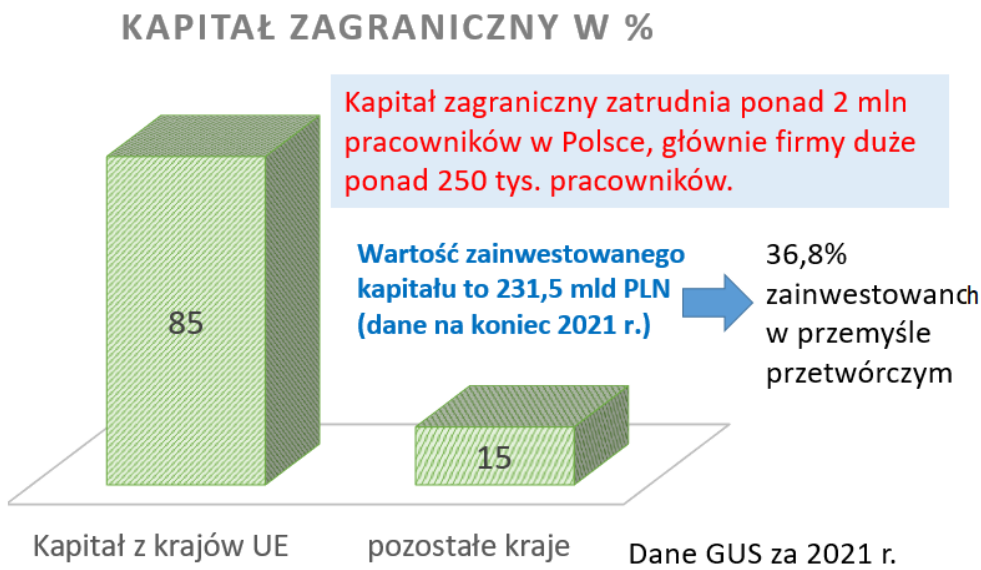


Rysunek 4. Prywatyzacja przedsiębiorstw państwowych (pp) w Polsce trwała 1990-2007 (liczba pp).
Źródło: opracowanie własne na podstawie [5]

W trakcie prywatyzacji, po raz pierwszy na tworzonym rynku pojawił się inwestor zagraniczny spoza „bloku wschodniego", który nabył (zakupił) wiele dużych podmiotów, tworząc „prawie monopol" na rynkach branżowym. Jako przykład można podać sektor stalowy w Polsce, gdzie kapitał zagraniczny decyduje o ponad 80% zdolnościach produkcyjnych, z tego ponad 70% należy do jednego inwestora zagranicznego – międzynarodowej grupy kapitałowej. Prywatyzacja była realizowana równolegle z restrukturyzacją.

Umownie przyjmuje się, że w 2007 roku polskie przedsiębiorstwa przemysłowe zakończyły etap dostosowawczy do wymogów gospodarki rynkowej. Krajowe firmy mogły konkurować z zagranicznymi na równoprawnych zasadach. Kapitał zagraniczny, głównie z krajów UE, który zatrudnia ponad 2 mln pracowników, a który przejął głównie firmy duże ponad 250 tys. pracowników, zainwestował w Polsce ponad (230 miliarda złotych) 231,5 mld PLN (dane na koniec 2021 roku) – rysunek 5.

Polski biznes, przy silnym wsparciu inwestorów z zagranicy, latami uczył się przedsiębiorczości budując wartość swoich przedsiębiorstw. Na wartość przedsiębiorstwa składają się aktywa, marka, reputacja, sytuacja rynkowa firmy i finansowa, udział w rynku, portfel produktów i usług, wiele relacji z klientami i kontrahentami, jak i wieloma interesariuszami. Interesariuszami są pracownicy, akcjonariusze, związki zawodowe organizacje branżowe, społeczność lokalna, instytucje publiczne, np. ZUS, Urząd Skarbowy, banki, firmy ubezpieczeniowe itd. W dotychczasowych rozważaniach restrukturyzację dzielono na naprawczą i rozwojową. Restrukturyzacja naprawcza była determinantą do dalszych działań biznesowych w ramach restrukturyzacji rozwojowej, której efektem jest ciągłe narastanie wartości. O ile restrukturyzacja naprawcza zorientowana była przede wszystkim na teraźniejszość i miała być gwarancją przetrwania przedsiębiorstw w okresie transformacji systemu gospodarczego, to restrukturyzacja rozwojowa ukierunkowana jest na przyszłe możliwości rozwoju przedsiębiorstw i konkurowanie na rynku globalnym.



Rysunek 5. Kapitał zagraniczny w Polsce. Źródło: opracowanie własne na podstawie [6]

3. W kierunku restrukturyzacji rozwojowej

Rozważając pojęcie restrukturyzacji przemysłu w kontekście celów biznesu, wskazać można na jej ujęcie jako procesu rekonstrukcji lub przebudowy struktury. Restrukturyzacja jest tożsama ze zmianami w strukturze bazy materialnej przedsiębiorstw i ich reorganizacją. W restrukturyzacji stawia się na zmiany, które prowadzą do innowacyjności. Zmiany dotyczą zatem zwiększenia roli wszystkich tych elementów struktury biznesu, które reprezentują nowoczesność i z reguły wyższą efektywność od rozwiązań dotychczas stosowanych w przedsiębiorstwie. Rozwój jest strategicznym celem funkcjonowania przedsiębiorstw. Rozwój jest „procesem dialektycznym, a zmiany, które leżą u jego podstawy, nie przebiegają skokowo, lecz układają się w pewne sekwencje i ciągi – stanowiąc proces, w toku którego poszczególne całości łączą się w układy bardziej skomplikowane, które mają nowe właściwości i prawidłowości dotychczas niespotykane” [7]. W ujęciu systemowym rozwój przedsiębiorstw wiąże się ze wzrostem złożoności biznesu. Współczesny biznes jest „czymś” więcej niż tylko działalnością zarobkową (profit), jest także organizacją ludzi (ang. *people*) i wiedzy (ang. *know-how*). Na rysunku 6 przedstawiono szeroki zakres innowacji biznesowych.

Rysunek 6. Szeroki zakres innowacji biznesowych. Źródło: opracowanie własne [8]

Współczesny rozwój przedsiębiorstw musi być odpowiedzialny (ang. *Corporate Social Responsibility*, skrót: CSR) i zrównoważony (ang. *sustainability*). Istota rozwoju przedsiębiorstw jest wielowymiarowa i złożona. Przedsiębiorczość jest kształtowana przez warunki dla rozwoju biznesu, w tym przez wiele czynników i procesów budujących działania według praw ekonomii, a także będących ich efektami korzyściami ekonomicznymi dla prowadzących biznes i całej gospodarki. Czynniki ekonomiczne i pozaekonomiczne oddziaływują na rozwój i efektywność przedsiębiorstw. Kwantyfikowalnym wymiarem rozwoju przedsiębiorstw są produkty, a właściwie zawarta w nich wartość dodana [9]. Droga do sukcesu restrukturyzacji wiedzie przez innowacje, które tworzą wartość dla klienta. Innowacje należy rozumieć bardzo szeroko, jako zmiany technologiczne, organizacyjne, procesowe, produktowe, usługowe itd. Wiele innowacji, to po prostu udoskonalenia lub modyfikacje, np. dotychczasowy wyrób po modyfikacji uzyskał nowe parametry, które zwiększyły jego użyteczność i funkcyjność. Innowacje powstają w biurach i departamentach Badania i Rozwoju (skrót: B+R lub R+D). Sporo innowacji przedsiębiorstwa nabywają z transferu wiedzy i przez zakup licencji, np. informatyczno-komputerowe systemy wsparcia procesowego. Inwestycje, które zrealizowały przedsiębiorstwa w okresie radykalnej restrukturyzacji dotyczył zarówno technologii, jak i usprawnień procesowych. Procesy zaczęto realizować z użyciem nowych analiz i metod zarządzania, np. benchmarking, TQM, Kaizen (czyli małymi krokami), Lean, Just in Time (tab. 1). W restrukturyzacji naprawczej został dowartościowany człowiek. Czynnik ludzki urósł do rangi kapitału intelektualnego przedsiębiorstw [10].

Tabela 1. Zakresy zmian restrukturyzacji przemysłu w Polsce

Pola zmian	Opis zmian
Technologia	W okresie restrukturyzacji wycofano z eksploatacji technologie (maszyny), które były nieopłacalne ekonomicznie lub nie spełniały radykalnych norm ochrony środowiska), np. w sektorze stalowym całkowicie wycofano technologię wytopu stali w piecach martenowskich (2002 roku). Wraz z radykalną redukcją zatrudnienia poprawiono wydajności pracy w przedsiębiorstwach produkcyjnych. Wydajność w wielu przypadkach poprawiła się samoistnie bo redukując zatrudnienie rosła wydajność, liczona jako wolumen produkcji przypadający na jednego pracownika. Przykład w sektorze stalowym, w 1990 roku wyprodukowano 13,5 mln. ton stali przy zatrudnieniu 147 tys. osób, obecnie 9 mln. ton przy zatrudnieniu 20 tys. osób. Wprowadzono automatyzację produkcji, która rozwija się w kierunku (ang.) *smart manufacturing*. Systemy informatyczno-komputerowe (ERP, MES, itd.) były (i nadal są) wsparciem procesowym w przedsiębiorstwach.
Produkt	Organizacja produkcji została zorientowana na system ssania (ang. *pull*), co wymagało usprawnień w zakresie harmonogramowania produkcji i elastyczności systemów wytwarzania (ang. FMS). Wprowadzono orientację na klienta, które stał się aktywnym uczestnikiem produkcji. Zadbano o jakość wyrobów i rozwój usług przed i po sprzedaży. Przedsiębiorstwa wprowadziły systemy zarządzania jakością (ISO 9001). W Czwartej Rewolucji Przemysłowej w inteligentnych fabrykach produkcja jest coraz bardziej spersonalizowana. Smart fabryki będą w stanie sprostać każdemu oczekiwaniu klienta, a indywidualna personalizacja stanie się personalizacją masową.
Zaopatrzenie	W sferze zaopatrzenia zadbano przede wszystkim o wybór dostawców (plasowanie dostawców) i wprowadzono koncepcję (jap.) *Just in Time* (dokładnie na czas), co oznaczało ograniczenie do minimum czynności magazynowych i elastyczność produkcji.
Produkcja	Sposób organizacji produkcji i jej przebieg zgodny z filarami produkcji klasy światowej, które nazywamy WCM od (ang.) *World Class Manufacturing*.10 filarów WCM zaczerpnięto z Systemu Produkcji Toyoty: bezpieczeństwo pracy, analiza kosztów, skoncentrowana poprawa, utrzymanie ruchu: autonomiczne i profesjonalne (ang. TPM), rozwój pracowników, wczesne zarządzanie wyposażeniem, jakość produktów, obsługa klienta, środowisko i odpowiedzialność społeczna.
Montaż	Wprowadzono zasadę racjonalnego gospodarowania zasobami i usprawniono logistykę wewnętrzną. Linie produkcyjne zostały silnie zautomatyzowane, a rola czynności fizycznych wykonywanych przez człowieka została ograniczona do minimum. Człowiek kontroluje przebieg procesu.

Źródło: opracowanie własne.

Po restrukturyzacji biznes osiągnął, w wielu sektorach, wymogi produkcji klasy światowej (ang. *World Class Manufacturing*). Menadżerowie nauczyli się zarządzać wieloma obszarami aktywności gospodarczej, począwszy od zarządzania produkcją, po zarządzanie zasobami ludzkimi, przez zrządzanie finansami, zarządzanie marketingowe, zarządzanie zmianą, zarządzanie marką, zarządzanie projektami, zarządzanie talentami itd. Wykazane w tabeli 1 zmiany są należy rozpatrywać jako przykłady zmian restrukturyzacyjnych, które dokonały się, w okresie ostatnich trzech dekady, w restrukturyzowanych przedsiębiorstwach. Przedsiębiorczy zrobili wiele w zakresie organizacji i zarządzania, a teraz stawiają na technologie komputerowe i inteligentne (AI). Na rynku jest już pokolenie „Z", które jest kreatywne i odważne, a przede wszystkim chętne do współpracy z nowymi technologiami komputerowymi i zainteresowane komunikacją online.

3.1. Od stabilności do labilności otoczenia: restrukturyzacja kryzysów

Obecnie przedsiębiorstwa są zmuszone do ciągłego dostosowywania się do otoczenia. Otoczenie. ze względnie (quasi) stabilnego, przekształciło się w dynamiczne. Współczesne otoczenie coraz częściej zaskakuje przedsiębiorców, a nawet zdarza się, że jest labilne. Labilność otoczenia, przez N.N. Taleba, została porównana do „czarnego łabędzia" [11]. Termin ten jest używany do opisywania nieoczekiwanych zdarzeń, których (prawie) nikt nie jest w stanie przewidzieć. Termin były już używany w XVI wieku, ale w obecnym znaczeniu upowszechnił się po tym, jak libańsko-amerykański naukowiec Nassim Nicholas Taleb użył go w 2007 roku w swojej książce zatytułowanej „The Black Swan: The Impact of the Highly Improbable". W swojej publikacji N.N. Taleb przewiduje, że „czarne łabędzie: w coraz większym stopniu będą zmieniać bieg wydarzeń na świecie. Takie przypuszczenie autor wysuwa z rosnącej złożoności społeczeństw i gospodarek. Wzajemne oddziaływanie czynników często jest lekceważone, a prognozy opracowywane są w oparciu o istniejące wzory i modele. Według N.N. Taleba pandemia COVID-19 nie była typowym czarnym łabędziem. Sądzi on, że aktualny kryzys jest „białym łabędziem" albo „szarym nosorożcem". Jest to zdarzenie, które ma ogromny wpływ na rzeczywistość, ale można je było przewidzieć. Charakterystyczną cechą „szarego nosorożca" jest ignorowanie zagrożeń. Poza pandemią COVID-19, która mocno wstrząsnęła biznesem, np. problemy z zapewnieniem ciągłości produkcji w okresie *lockdown* spowodowane przerwaniem łańcuchów dostaw, w kolejnym roku (2022 roku) przemysł musiał zareagować na rosnące ceny energii. Kryzys energetyczny przeniósł się na rynki surowców, a w rezultacie spowodował szybki wzrost cen wyrobów finalnych. Kryzys ten miał inne oblicze, niż kryzys amerykański, który pojawił się na rynkach nieruchomości i bankowości w USA w 2008 roku, a który do Polski kryzys zawitał w 2009 roku. Skutkiem tego kryzysu był wzrost cen kredytów udzielanych w walutach zagranicznych, a skutkiem kryzysu finansowego był spadek produkcji w wielu branżach przemysłu [12].

3.2. W kierunku Przemysłu 4.0: restrukturyzacja cyfrowa

Od ponad dekady popularyzowana jest nowa koncepcja rozwoju nazywana Przemysłem 4.0 (ang. *Industry 4.0*) w skrócie: I 4.0. Przemysł 4.0 jest rezultatem dokonującego się postępu technologicznego w dziejącej się Czwartej Rewolucji Przemysłowej. Pojęcie: *Industrie 4.0* (niem.) powstało na targach w Hanowerze w 2011 roku i do dzisiaj jest bardzo popularne (ponad 20 tys. publikacji naukowych zidentyfikowanych, tylko w jednej, bazie naukowej, Web of Science). Czym jest Przemysł 4.0? Przemysł 4.0 korzysta z osiągnięć cyfrowych technologii Trzeciej Rewolucji Przemysłowej ale wzbogaca je nowymi rozwiązaniami, które wpisują się w sztuczną inteligencję (ang. *Artificial Intelligence*, skrót: AI) [13]. Te nowe technologie korzystają z algorytmów rozumowania, myślenia i działania ludzi. Aby zapewnić AI samodzielność, musi ona mieć dostęp do Big Data i cloud oraz Internetu Rzeczy (ang. *Internet of Things*, skrót: IoT) a właściwie już Internetu Wszechrzeczy (ang. *Internet of Everythings*, skrót: IoE). Zmiany są wszechobecne, dzieją się w fabrykach, transporcie, miastach, energetyce, komunikacji itd. W fabrykach *Industry 4.0* jest pogłębioną cyfryzacją biznesu oraz postępującą automatyzacją procesów przemysłowych. Technologie zyskują nowe funkcje dzięki wykorzystaniu w przedsiębiorstwach zaawansowanych systemów informatycznych, sztucznej inteligencji (AI), Internetu Rzeczy (IoT) i dostępu do danych w czasie rzeczywistym. Maszyny są w stanie komunikować się ze sobą (M2M) i z ludźmi (M2P). Algorytm my-

ślenia człowieka jest przekazywany technologiom, a człowiek jest uczestnikiem cyber-fizycznych systemów (ang. *Cyber-Physical Systems*, skrót: CPSs) [14]. Biznes liczy na to, że inteligentna technologia wesprze go, w tym co robi i jak funkcjonuje. Dzięki cyfryzacji udało się przedsiębiorstwom przetrwać okres lockdown (u). Wiele firm uwierzyło w komputer, Internet i pracę z domu (ang. *home office*). W centrum, dokonujących się przemian, jest człowiek (ang. *Human-Cyber-Physical Systems*, w skrócie: H-CPS) [15], pod warunkiem, że będzie w stanie współpracować z wysokimi technologiami. Na rynku pracy jest pokolenie „X" i „Y", które musi się zmienić. Pracownicy z pokoleń X i Y muszą nabyć nowe umiejętności cyfrowe i polubić sztuczną inteligencję. Nie ma z tym problemu pokolenie „Z", które urodziło się podczas boomu na informatyczno-komputerowe technologie (ang. ICTs). W dokumentach rządowych pojawia się wiele modeli dotyczących rozwoju czynnika ludzkiego (ang. *Human Factors*, skrót: HF), jednym z nich jest model w kształcie litery „T", w którym nacisk kładziony jest na łączenie wiedzy i umiejętności twardych (ang. *hard skills*) z umiejętnościami miękkimi (ang. *soft skills*) [16]. Pracownicy muszą być elastyczni, kreatywni, komunikatywni, itd. [10, 17-18].

3.3. Strategiczne kierunki rozwoju biznesu w Przemyśle 5.0

W styczniu 2021 roku, Komisja Europejska, opublikowała dokument pt.: *Industry 5.0. Towards a sustainable, human centric and resilient European industry* [19]. Wraz z pojawieniem się „Przemysłu 5.0" przyszła kolej na kolejny krok w rozwoju przemysłu. Jeszcze przed okrągłą rocznicą powstania Przemysłu 4.0 nowy termin był już w obiegu, a pierwsze artykuły na ten temat pojawiły się w 2016 roku, kiedy proces transformacji do wersji 4.0 Inteligentnego Przemysłu był jeszcze w pełnym rozkwicie [20]. Według Komisji Europejskiej Przemysł 5.0 uzupełnia istniejący paradygmat Przemysłu 4.0 poprzez podkreślenie znaczenia człowieka, zrównoważoności i odporności (stabilności biznesu). Technologie Przemysłu 4.0 muszą być ściślej powiązane ze społeczeństwem i środowiskiem. Według Komisji Europejskiej obecny rozwój Przemysłu 4.0 w coraz większym stopniu odchodzi od pierwotnych założeń równości społecznej i zrównoważonego rozwoju, a bardziej dąży do cyfryzacji i sztucznej inteligencji (AI) w celu zwiększenia wydajności i elastyczności produkcji. Pojawia się również kwestia wpływu transformacji na ludzi – jakie miejsce zajmują ludzie w przemyśle, w którym spora część zadań jest zautomatyzowana i wykonywana przez maszyny? Ponadto, po okresie pandemicznym, pojawiło się również pytanie: Jak zabezpieczyć biznes przed nieprzewidywalnymi zdarzeniami? Systemy zarządzania zdarzeniami mogą stanowić integralną część maszyny, być zbiorem maszyn czy osobnym systemem.

Przemysł 5.0 jest rezultatem popularyzacji japońskiej koncepcji Społeczeństwa 5.0, czyli społeczeństwa, w którym zaawansowane technologie informatyczne, Internet Rzeczy, roboty, sztuczna inteligencja i rzeczywistość rozszerzona są aktywnie wykorzystywane w życiu codziennym, przemyśle, opiece zdrowotnej i innych sferach działalności, przy czym nie chodzi tu głównie o korzyści ekonomiczne, ale o dobro i wygodę każdego obywatela. Przemysł 5.0 łączy przemysł, środowisko i społeczeństwo za pomocą technologii i na tej podstawie opracowuje zasady globalnego rozwoju, które są skoncentrowane na humanizacji i zrównoważoności biznesu [21].

Industry 4.0 jest teraźniejszością – przedsiębiorstwa dopiero transformują do poziomu 4.0. Przedsiębiorstwa wprowadzają technologie, które pozwolą im uczynić z produkcji, produkcję „smart". Elementami krytycznymi są czas i tempo wprowadzania zmian

[22, 23]. Będą one kluczowe zarówno w obecnej, jak i w przyszłej rewolucji przemysłowej. *Industry 5.0* jest przyszłością, w której uczenie maszynowe (ang. *machine learning*) i sztuczna inteligencja wsparta danymi, będą się rozwijać. Będzie to też czas nowych systemów wytwarzania, które będą łączyć ze sobą obiekty z wykorzystaniem idei blockchain [24]. Obecna rewolucja nie jest jeszcze właściwie wdrożona i wykorzystana, pozostaje mieć nadzieję, że przedsiębiorstwa sprostają następnej.

Pomimo, że Przemysł 5.0 jest przyszłością, to już pojawiają się pierwsze artykuł o Przemyśle 6.0, zatem postęp technologiczny przyspiesza. Na rynku już jest pokolenie „Z", które jest zainteresowane technologiami zwinnymi, a za parę lat będzie pokolenie „alfa", którego oczekiwań jeszcze dobrze nie znamy, a które urodziło się już w okresie popularyzacji komunikacji online i inteligentnych technologii.

4. Cykl się zamyka: głęboka zielona restrukturyzacja

Pokazując historię przemian, warto jeszcze zatrzymać się na zrównoważonym rozwoju. Od naprawy biznesu po zrównoważony biznes, tak można by zatytułować ten rozdział. W kontekście *sustainability* restrukturyzacja naprawcza stanowi podstawę do dalszych działań, tj. restrukturyzacji rozwojowej, której efektem jest zrównoważony rozwój. Trzy kluczowe koncepcje działań przedsiębiorstw można przedstawić w formie piramidy, w której podstawę stanowi restrukturyzacja naprawcza, a wierzchołek to zrównoważony biznes. Łącznikiem pomiędzy restrukturyzacją naprawczą a zrównoważonym biznesem jest restrukturyzacja rozwojowa (rys. 7). O ile restrukturyzacja naprawcza zorientowana jest przede wszystkim na teraźniejszość i ma być gwarancją przetrwania przedsiębiorstwa, to rozwojowa ukierunkowana jest na przyszłe możliwości rozwoju przedsiębiorstwa (działania ukierunkowane na wyprzedzanie zmian w otoczeniu), w tym na zrównoważony biznes. Zrównoważony rozwój jest silnym balansowaniem biznesu między ekonomią, ekologią a etyką. Społecznie odpowiedzialny biznes stosuje wskaźników GRI w raportowaniu swojej aktywności na rynku.

Rysunek 7. Etapowość restrukturyzacji przemysłu [opracowanie własne].

Znaczenie zrównoważonego rozwoju i społecznej odpowiedzialności w przemyśle ciągle rośnie ponieważ coraz bardziej restrykcyjne są wymogi Unii Europejskiej i biznesu światowego. Do 2050 gospodarki chcą uzyskać zerową emisje CO_2, co oznacza, że kwestie środowiska będą wyznaczać, w kolejnych latach, kierunki inwestycji, a sztuczna inteligencja będzie pomagać w tworzeniu (rozwoju) *smart city*, *smart factory*, *smart product* (ang.)

Czy będzie kolejna restrukturyzacja przedsiębiorstw w Polsce? W warunkach radykalności polityki na rzecz ochrony środowiska może dojść do kolejnej „głębokiej" restruk-

turyzacji, czyli zostanie zamknięta pętla zmian. Politycy dążą, co jest słuszne, do radykalności działań na rzecz ochrony środowiska. W Polsce 70% energii otrzymuje się z węgla, a dąży się do emisji zerowej CO_2, a zatem branże, takie jak: energetyka, górnictwo, hutnictwo i inne muszą radykalnie zmienić technologie i źródła energii. Do przemysłu wkroczą nowe zeroemisyjne technologie, które są już na etapie testowane. Zmieni się otoczenie biznesu, samochody będą zasilane energią, miasta, będą budować inteligentność, urzędy będą świadczyć usługi w formie *homeoffice* (ang.). W zakresie zasilania energetycznego, przewiduje się, że przyszłość będzie należeć do wodoru. Pojawi się nowa wartość, w której „green" znaczyć będzie więcej niż tylko ekologia, to będzie nowa strategia biznesu.

5. Podsumowanie

Dynamika zachodzących zmian wynika przede wszystkim z faktu pojawienia się zupełnie nowych warunków prowadzenia biznesu, które wymuszają rozwój nowych technologii, produktów, kompetencji itd. Po restrukturyzacji naprawczej, w latach 90., przemysł etapami tworzył nowe obszary aktywności biznesowej. Obecnie, zgodnie z ideą Przemysłu 4.0, uczestniczy w budowaniu środowiska smart [23]. Przemysłowy Internet Rzeczy (IoT), inteligentne czujniki, roboty przemysłowe, drukarki 3D to elementy obecnej transformacji cyfrowej. Te technologie będą kluczowe dla kolejnego kroku transformacji cyfrowej, która zbuduje Industry 5.0. Ta nowa koncepcja rozwoju przemysłu ma być zorientowana na człowieka, zrównoważony rozwój i odporność przedsiębiorstw na silne wstrząsy w otoczeniu. Nie uda jej się jednak osiągnąć bez ponownej radyklanej restrukturyzacji, tym razem „green" restrukturyzacji, która pozwoli światu osiągnąć neutralność klimatyczną do 2050 roku. W wielu publikacjach podkreśla się, że wpływ polityki ochrony środowiska na innowacje, w przedsiębiorstwach, określane jako „clean" jest już widoczny [25-28]. Konieczność restrukturyzacji ekologicznej wynika bezpośrednio z przepisów prawa ochrony środowiska i polityki środowiskowej UE. Zgodnie z długoterminową polityką UE (strategia do 2050 roku) [29], kraje członkowskie przyjęły kierunek rozwoju określany jako „net zero", czyli zerowa emisja CO_2 do 2050 roku. Ten kierunek wymusza już na przedsiębiorstwach zmiany w technologiach wytwarzania [30, 31], jak i w źródłach zasilania energetycznego. Przyjmując, za wieloma naukowcami [32-36], że restrukturyzacja jest procesem ciągłym, któremu towarzyszą zmiany radyklane w funkcjonowaniu przedsiębiorstw, można z dużą pewnością zakładać, że „green" restrukturyzacja będzie realizowana w przedsiębiorstwa w następnych trzech dekadach.

Dotychczasową ścieżkę restrukturyzacji można opisać przez kluczowe zakresy przemian adaptacyjnymi przedsiębiorstw do warunków środowiska biznesowego. Zakresy zmian są wyznacznikami typów restrukturyzacji, które ujęto w tabeli 2, będących formą podsumowania opisanej w niniejszej pracy ścieżki przemian restrukturyzacyjnych przedsiębiorstw w Polsce.

Tabela 2. Typy restrukturyzacji przedsiębiorstw w Polsce na ścieżce adaptacji do warunków otoczenia

Restrukturyzacja naprawcza	Restrukturyzacja przeprowadzona w Polsce w latach 90. i późniejszych. Jej celem było stworzenie warunków do prowadzenia biznesu w gospodarce rynkowej poprzez prywatyzację przedsiębiorstw i radykalną poprawę niekorzystnych warunków ekonomicznych poprzez eliminację niekorzystnych czynników, prowadzących przedsiębiorstwo do likwidacji [2, 32, 33, 37].
Restrukturyzacja rozwojowa	Zapoczątkowana po zakończeniu restrukturyzacji naprawczej, jej celem było budowanie wartości przedsiębiorstwa. Za podstawę zmian przyjmuje się decyzje strategiczne podejmowane w przedsiębiorstwie o charakterze innowacyjnym, rozwojowym [2, 4, 32, 33, 37, 38].
Restrukturyzacja kryzysów	Restrukturyzacja jest rezultatem reakcji przedsiębiorstw na zdarzenia kryzysowe – światowe kryzysy ekonomiczne. Działania naprawcze mają charakter reaktywny [2].
Restrukturyzacja cyfrowa	Restrukturyzacja przemysłu realizowana w celu dostosowania przedsiębiorstw do wymogów przemysłowych koncepcji rozwoju, tj. Industry 4.0 i Industry 5.0. Restrukturyzacja obejmuje zmiany w zakresie aktywów przedsiębiorstwa. Zawiera w sobie głównie obszary: zaopatrzenie, produkcja, marketing i sprzedaż. Celem restrukturyzacji marketingowej jest uzyskanie wymogów cyfrowego biznesu.
Restrukturyzacja ekologiczna (zielona)	Podstawą restrukturyzacji jest strategia zrównoważonego rozwoju. Przedsiębiorstwa, od lat, dostosowują się do wymogów polityki ochrony środowiska lecz w ostatnich latach odnotowano nasilenie działań w związku z koncepcją „Zielonego Ładu". Restrukturyzacja zielona, czy też głęboka dekarbonizacji oznacza radykalne ograniczenie emisji CO2 w dążeniu do celu, jakim jest zerowa emisja tego gazu do 2050 roku.

Źródło: przytoczone definicje są opracowaniem własnym, przy czym jako przypisy podano inne publikacje, w których również są definicje restrukturyzacji.

Uwagi ogólne

Źródło finansowania: Badania Statutowe Politechniki Śląskiej: 11/040/BK_223/0029.

Literatura

1. Adaptacyjność. Wyjaśnienie hasła na podstawie słownika PWN, https://sjp.pwn.pl/slowniki/adaptacyjno%C5%9B%C4%87.html
2. Gajdzik B., *Restrukturyzacja przedsiębiorstw w warunkach destabilizacji otoczenia na przykładzie branży hutniczej*, Wydawnictwo Naukowe Difin, Warszawa 2012.
3. Savas E.S., *Prywatyzacja. Klucz do lepszego rządzenia*, PWE, Warszawa 1992.
4. Gajdzik B., *Przedsiębiorstwo hutnicze po restrukturyzacji*, Wydawnictwo Politechniki Śląskiej, Gliwice 2010, ISBN 978-7335-543-2.
5. *Prywatyzacja przedsiębiorstw państwowych*, GUS, Warszawa 2009.
6. GUS, Raport. *Rocznik statystyczny przemysłu*, 2021
7. Fabiańska K., Rokita J., *Planowanie rozwoju przedsiębiorstwa*, PWE, Warszawa 1986, s. 14.
8. Gajdzik B., *Kompleksowe spojrzenie na zmiany restrukturyzacyjne w sektorze hutnictwa żelaza i stali w Polsce*, [w:] Pytka J. (red.), *Nowoczesność przemysłu i usług. Dynamika zmian w polskim przemyśle i usługach*, TNOiK, ISBN 978-83-85587-28-6, Katowice 2011, s. 30-44.
9. Marcinkowska M., *Kształtowanie wartości firmy*, Wydawnictwo Naukowe PWN, Warszawa 2000, s. 18.
10. Czerwińska-Lubszczyk A. Gajdzik B., Grebski M., Grebski W., Jagoda-Sobolak D., Krawczyk D., Kuzior A., Wolniak R., *Zarządzanie kreatywnością i innowacyjnością w przedsiębiorstwie*, Towarzystwo Naukowe Organizacji i Kierownictwa. Dom Organizatora, 2022. ISBN 978-83-67153-19-5.
11. Taleb N.N., *The Black Swan*: *The Impact of the Highly Improbable*, 2007.

12. Gajdzik B., *Reakcja przedsiębiorstw hutniczych a światowy kryzys gospodarczy*, VIII Konferencja Naukowa pt. *Procesy internacjonalizacji w gospodarce światowej*, organizator: Katedra Międzynarodowych Stosunków Ekonomicznych Uniwersytetu Ekonomicznego w Katowicach, 2009.
13. Kagermann H., Wahlster W., Helbig J. (red.), *Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0*. Working Group: Industrie 4.0: Mit dem Internet der Dinge auf dem Weg zur 4. industriellen Revolution, VDI-Nachrichten, Acatech-National Academy of Science and Engineering: München, Germany, [online:] http://forschungsunion.de/pdf/industrie_4_0_final_report.pdf.Lee J., BagheriB., Kao H., *Research Letters: A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems*, Manuf. Lett., 2015, 3, 18-23.
14. Romero D., Noran O., Stahre J., Bernus P., Fast-Berglund Å., *Towards a human-centred reference architecture for next generation balanced automation systems: human-automation symbiosis*, In IFIP International Conference on Advances in Production Management Systems−APMS. Advances in Production Management Systems: Initiatives for a Sustainable World, 2015, s. 556-566. Retrieved from https://link.springer.com/chapter/10.1007/978-3-319-22759-7_64 .
15. Blueprint for sectoral cooperation on skills: towards an EU strategy addressing the skills needs of the steel sector. European vision on steel-related skills and supporting actions to solve the skills gap today and tomorrow in Europe, May 2020 European Commission Executive Agency for Small and Medium-sized Enterprises (EASME) Unit A1 – COSME, Brussels 2020, s. 73 -74.
16. Gajdzik B., Wolniak R., *Smart Production Workers in Terms of Creativity and Innovation: The Implication for Open Innovation*, J. Open Innov. Technol. Mark. Complex., 8, 2022.
17. Flores E., Xu X., Lu Y., *Human Cyber-Physical Systems: A skill-based correlation between humans and machines*, 16th IEEE International Conference on Automation Science and Engineering (CASE) August 20-21, Online Zoom Meeting, 2020, s. 1313-1318.
18. *Industry 5.0 Towards a sustainable, human centric and resilient European industry*. European Commission, Brussels, Manuscript completed in January 2021.
19. Gajdzik B., *Industry 5.0 as a new concept of development within high volatility environment: about the Industry 5.0 based on political and scientific studies*, Zeszyty Naukowe Politechniki Śląskiej. Organizacja i Zarządzanie, vol. 169, 2023, s. 255-279. DOI:10.29119/1641-3466.2023.169.15.
20. Grabowska S., Saniuk S., Gajdzik B., *Industry 5.0: improving humanization and sustainability of Industry 4.0*, Scientometrics, 127(6), 2022, s. 3117-3144.
21. Gajdzik B., *Frameworks of the Maturity Model for Industry 4.0 with Assessment of Maturity Levels on the Example of the Segment of Steel Enterprises in Poland*, J. Open Innov. Technol. Mark. Complex, 8, 2022, s. 77. https://doi.org/10.3390/joitmc8020077.
22. Schumachera A., Erolb S., Sihna W., *A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises*, Procedia CIRP 52, 2016, s. 161-166. Elsevier [online] http://creativecommons.org/licenses/by-nc-nd/4.0/.
23. Ayan B., Güner E., Son-Turan S., *Blockchain Technology and Sustainability in Supply Chains and a Closer Look at Different Industries: A Mixed Method Approach*, Logistics, 6, 2022, 85. https://doi.org/10.3390/logistics 6040085.
24. Gajdzik B., *Diagnoza kierunków transformacji przemysłu stalowego w Przemyśle 4.0*, ISBN 978-83-7880-850-3 Wydawnictwo Politechniki Śląskiej Gliwice 2022, Monografia 945.
25. Eugster J., *The Impact of Environmental Policy on Innovation in Clean Technologies*, IMF Working Papers, 213, 2021.
26. IMF.*Reaching Net Zero Emissions*, G20 Note Washington, DC, International Monetary Fund, 202.

27. Black S., Chateau F., Jaumotte I., Parry G., Schwerhoff S., Thube D., Zhunussova K., *Getting on Track to Net Zero: Accelerating a Global Just Transition in This Decade*, Staff Climate Notes, 10, 2022.
28. Lahouirich M.W., Oulfarsi S., Eddine A.S., Sakalli H.E.B., Boutti R., *From financial performance to sustainable development: A great evolution and an endless debate*, Financ. Mark. Inst. Risks, 6, 2022, s. 68-79.
29. European Green Deal. https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2050-long-term-strategy_pl.
30. Miskiewicz R., *Efficiency of electricity production technology from postprocess gas heat: Ecological, economic and social benefits*, *Energies*, 13, 2020, 6106.
31. Gajdzik B., Wolniak R., Grebski W., *Process of Transformation to Net Zero Steelmaking: Decarbonisation Scenarios Based on the Analysis of the Polish Steel Industry*, Energies, 16(8), 2023, s. 3384. https://doi.org/10.3390/en16083384.
32. Bitkowska A., Wójcik G., *Restrukturyzacja przedsiębiorstw jako ciągły proces dostosowawczy*, [w:] Bitkowska A. (red.), *Procesy restrukturyzacji warunkiem poprawy konkurencyjności przedsiębiorstwa*, Difin, Warszawa 2010.
33. Borowiecki R., Nalepka A., *Restrukturyzacja w procesie funkcjonowania i rozwoju przedsiębiorstw*, [w:] Borowiecki R. (red.), *Zarządzanie restrukturyzacją procesów gospodarczych: aspekt teoretyczno-praktyczny*, Difin, Warszawa, 2003.
34. Wiśniewska E., *Restrukturyzacja organizacyjna jako zmian radykalna*, Zeszyty Naukowe Politechniki Śląskiej. Organizacja i Zarządzanie, z. 67, 2013.
35. Borowiecki R., Jaki A. (red.), *Procesy restrukturyzacji a konkurencyjność w warunkach globalizacji*, AE - Fundacja AE, Kraków 2006.
36. Nalepka A., *The Range and Conditions of Organizational Restructuring of an Enterprise* Zeszyty Naukowe / Akademia Ekonomiczna w Poznaniu, nr 88, 2007, s. 71-81.
37. Borowiecki R. (red.), *Zarządzanie restrukturyzacją przedsiębiorstw i gospodarki. Uwarunkowania – Procesy - Efekty*, Fundacja Ekonomicznego Uniwersytetu w Krakowie, Kraków 2014.
38. Suszyński C., *Restrukturyzacja przedsiębiorstw. Proces zarządzania zmianami*, PWE, Warszawa 1999, s. 67.

Adaptacyjność przedsiębiorstw w Polsce w ostatnich trzech dekadach. Od restrukturyzacji naparwczej po inteligentne technologie

Streszczenie
W pracy przedstawiono transformację przemysłu w Polsce w okresie ostatnich trzech dekad. Praca składa się z kilku części, które zostały poświęcone zmianom przemysłowym, począwszy od transformacji gospodarczej w Polsce do chwili obecnej. W pracy uwzględniono takie zagadnienia, jak: restrukturyzacja naprawcza, rozwój, innowacyjność, dynamika otoczenia, zrównoważoność. Praca jest rezultatem badań własnych i stanowi ich podsumowanie konceptualne.
Słowa kluczowe: restrukturyzacja, zmiana, Przemysł 4.0, zrównoważoność, rozwój

Adaptability of enterprises in Poland in the last three decades. From repair restructuring to smart technologies

Abstract
The paper presents the transformation of industry in Poland over the last three decades. The work is divided into several parts, which are devoted to industrial change from the economic transformation in Poland to the present. The paper considers issues such as remedial restructuring, development, innovation, environmental dynamics, sustainability. The work is the result of own research and is a conceptual summary of it.
Keywords: restructuring process, change, Industry 4.0, sustainability, development

Barbara Wieczorek[1]

Zabezpieczenie połączenia płyta-słup konstrukcji żelbetowych przed katastrofą postępującą w świetle przepisów normowych

1. Wprowadzenie

Ustroje płytowo-słupowe charakteryzuje bardzo duża popularność stosowania w budownictwie mieszkaniowym, przemysłowym czy też jako obiektów biurowych, użyteczności publicznej, magazynowych lub parkingowych. Główną ich cechą jest bowiem prostota kształtowania, walory architektoniczne, szybkość wykonania oraz bezproblemowe użytkowanie, czy bieżące remonty. Niestety konstrukcje tego typu wykazują bardzo dużą podatność na uszkodzenia wynikające z pękania betonu na skutek przeciążenia elementów konstrukcji lub innych czynników zewnętrznych. Istotnym aspektem zatem pozostaje zagadnienie zapewnienia bezpieczeństwa w trakcie ich użytkowania.

Omówione wybrane awarie obiektów budowlanych wykonanych jako ustroje płytowo-słupowe wskazują najczęstsze przyczyny, które spowodowały ich wystąpienie. W wielu przypadkach, uszkodzenie jednego z elementów konstrukcji doprowadziło do rozprzestrzenienia się zniszczeń na kolejne fragmenty konstrukcji i w konsekwencji przyczyniło się do zawalenia się części tego obiektu, a w niektórych przypadkach objęło swoim zasięgiem nawet cały obiekt. Skutki takich katastrof obejmują nie tylko straty materialne, ale niejednokrotnie niosą za sobą ofiary śmiertelne w ludziach.

W przypadku monolitycznych konstrukcji żelbetowych ich awarie najczęściej związane są z przebiciem płyty wokół słupa na skutek wyczerpania nośności betonu stropu. Zatem kluczową jest kwestia zastosowania odpowiednio skonstruowanego zbrojenia, które zapewni nośność konstrukcji nawet po wystąpieniu lokalnego jej uszkodzenia. Przedstawiono wytyczne odnośnie projektowania strefy podporowej według Eurokodu PN-EN 1991-1-7:2008 [1] i Eurokodu PN-EN 1992-1-1:2008 [2]. Dodatkowo omówiono przepisy zawarte w zagranicznych normach dotyczące stosowania dodatkowego dolnego zbrojenia krzyżującego się bezpośrednio nad słupem, aby zapewnić nośność połączenia płyty ze słupem po jej uszkodzeniu wskutek przebicia. W szczególności przedstawiono zapisy normy kanadyjskiej CSA A23.3 [3], rekomendacji do normy amerykańskiej ACI 352.1R [4], wprowadzeniu do normy szwajcarskiej SIA 262 [5] oraz Model Code 2010 [6].

2. Wady i zalety konstrukcji płytowo-słupowych

W budynkach wykonanych w postaci ustroju płytowo-słupowego, ich konstrukcja składa się płyty stropowej podpartej bezpośrednio na słupach bez zastosowania belek pośrednich lub głowic (rys. 1).

[1] barbara.wieczorek@polsl.pl, Politechnika Śląska, Wydział Budownictwa, www.polsl.pl.

Rysunek 1. Przykładowy widok żelbetowej konstrukcji płytowo-słupowej [7]

Niewątpliwą zaletą tego typu ustroju konstrukcyjnego są między innymi prostota oraz szybkość wykonania konstrukcji ze względu na nieskomplikowany kształt i szybkie do wykonania deskowanie, a także możliwość bezproblemowej adaptacji danej kondygnacji po zakończeniu procesu betonowania kondygnacji wyższej. Dodatkowo konstrukcja stropów nie stawia żadnych ograniczeń w zakresie kształtowania i rozmieszczenia powierzchni do planowanego zakresu użytkowania. Brak belek wewnętrznych pozwala na obniżenie wysokości kondygnacji, a tym samym możliwe jest zwiększenie liczby kondygnacji, przy jednoczesnym niezwiększaniu całkowitej wysokości budynku. Płaska powierzchnia stropów zapewnia łatwość przeprowadzania instalacji pod sufitem, co umożliwia oszczędność materiałów użytych do budowy, a w czasie eksploatacji obiektu łatwość w prowadzeniu konserwacji i remontów. Ponadto dostępny jest szeroki zakres metod projektowania, przepisów normowych i wytycznych technicznych oraz prostych i skutecznych narzędzi numerycznych do projektowania ustrojów płytowo-słupowych.

Niestety tego typu ustrój konstrukcyjny posiada też wady. Zagrożeniem dla konstrukcji płytowo-słupowych jest głównie możliwość przebicia stropu w rejonie słupa, które powstaje w wyniku wyczerpania wytrzymałości betonu na rozciąganie.

Awarie wywołane przebiciem występują prawie bez wstępnych objawów, ponieważ powstałe ugięcia są małe, a zarysowania na górnej stronie płyty są zwykle niewidoczne z uwagi na zastosowane warstwy posadzkowe. Skutkiem wystąpienia przebicia płyty w otoczeniu jednego słupa jest zwiększenie wartości siły poprzecznej na sąsiednich słupach, co w efekcie może doprowadzić do powstawania zniszczeń elementów w kolejnych

częściach konstrukcji. Dlatego konieczne jest stosowanie dodatkowego usztywnia konstrukcji za pomocą żelbetowych ścian wewnętrznych i zewnętrznych.

3. Przykłady katastrof wywołanych przebiciem płyty

W literaturze technicznej można znaleźć informacje o kilkunastu katastrofach ustrojów płytowo-słupowych, w przypadku których zaprojektowane zbrojenie podporowe nie zapobiegło zniszczeniu obiektu.

W 1997 roku całkowitemu uszkodzeniu uległa część stropu na najwyższym piętrze wielopoziomowego parkingu Piper's Row Car w Wielkiej Brytanii (rys. 2). Wiele czynników (projektowych, wykonawczych, użytkowych) przyczyniło się do początkowego przebicia jednego z połączeń płyta-słup, które zainicjowało postępującą katastrofę. Jako bezpośrednią przyczynę wskazano w raporcie złą konserwację, co doprowadziło do uszkodzeń spowodowanych mrozem i utraty nośności [8]. Chociaż obiekt był w tym czasie użytkowany, nikt nie został ranny, gdyż zdarzenie miało miejsce w nocy

Rysunek 2. Katastrofa obiektu – Piper's Row Car, Wielka Brytania, Wolverhampton 1997 [9]

Największa w historii awaria ustroju płytowo-słupowego, a mianowicie zawalenie się galerii handlowej (rys. 3), miała miejsce w Seulu w 1995 roku. Przyczyną katastrofy było przeciążenie strefy przysłupowej [10]. W efekcie doszło do przebicia, które następnie lawinowo rozprzestrzeniło się na sąsiednie słupy i doprowadziło do zawalenia prawie całego obiektu. Budynek praktycznie w całości uległ zniszczeniu, zginęło 501 osób, a 937 zostało rannych. Główną przyczyną zdarzenia był błąd projektowy, polegający na usunięciu kilku słupów nośnych na niższych piętrach w celu zapewnienia miejsca dla schodów ruchomych. Zwiększone ponad przewidywane obciążenie dachu spowodowało wystąpienie postępującego przebicia kolejnych połączeń płyta-słup.

Rysunek 3. Katastrofa obiektu – Sampoong Department Store, Korea Południowa, Seul 1995 [11]

W 2003 roku, zawaleniu uległ będący w budowie garaż znajdujący się w mieście Atlantic w stanie New Jersey (rys. 4). Garaż był częścią projektu rozbudowy Tropicana Casino and Resort. Budynek zawalił się w trakcie budowy, obejmując swoim zasięgiem pięć kondygnacji parkingowych. W wyniku katastrofy zginęło czterech pracowników budowlanych, a rany odniosło ponad 30 innych osób. Powodem katastrofy były błędy popełniane przy wykonywaniu obiektu [12].

Rysunek 4. Katastrofa obiektu – Tropicana Casino parking garage, USA, Atlantic City 2003 [13]

Segment budynku garażowo-magazynowego w Warszawie uległ zawaleniu w 2003 roku. Głównym czynnikiem, który przyczynił się do wystąpienia zniszczenia tego obiektu był brak odpowiedniej pielęgnacji betonu na budowie po jego ułożeniu w deskowaniu [14]. Prace budowlane były prowadzone w temperaturze poniżej zera, co przyczyniło się do poważnych uszkodzeń struktury materiału. Wytrzymałość betonu okazała się niedostateczna w częściach konstrukcji obiektu, narażonych na przemrożenie w pierwszym okresie jego dojrzewania. Na zaistniały stan miał wpływ brak zastosowania odpowiednich zabezpieczeń w postaci okrycia świeżego betonu odpowiednimi matami czy zastosowania cieplaków lub ogrzewania pomieszczeń ze świeżym betonem. Ponadto bez prowadzenia kontroli wytrzymałości betonu w stropach i słupach, były prowadzone prace związane ze zdejmowaniem stojaków rusztowań na kolejnych piętrach, przy równoczesnym kontynuowaniu robót betonowych na innych kondygnacjach.

Rozwinięcie tego tematu, z omówieniem większej liczby katastrof obiektów o konstrukcji płytowo-słupowej, zostało zamieszczone w publikacji [15].

4. Przyczyny katastrof i zabezpieczenia

Badania, ekspertyzy prowadzone po katastrofach wskazują, że zarówno głównymi przyczynami katastrof są błędy popełnione na etapie projektowania oraz w czasie wykonywania obiektów. Większości przypadków tych awarii można było uniknąć, gdyby prowadzone były dokładniejsze kontrole projektów, a na budowie materiałów i detali konstrukcyjnych.

Wśród głównych czynników powodujących katastrofy postępujące należy wskazać czynniki technologiczne:

- drastycznie zaniżona wytrzymałość betonu – błąd dostawcy;
- niewłaściwa pielęgnacja betonu – przemrożenie betonu, skurcz betonu;
- zbyt wczesne rozdeskowanie konstrukcji;
- zbyt wczesne obciążenie elementów konstrukcyjnych;
- inne nieumyślne działania.

Najczęściej przyczyną katastrofy jaka może wystąpić w przypadku obiektów o konstrukcji płytowo-słupowej jest zniszczenie płyty stropowej w strefie podporowej wokół słupa. Stąd konieczne jest odpowiednie zabezpieczenie tego obszaru konstrukcji przed ewentualnym uszkodzeniem.

Wśród różnych metod zabezpieczających można wyróżnić:

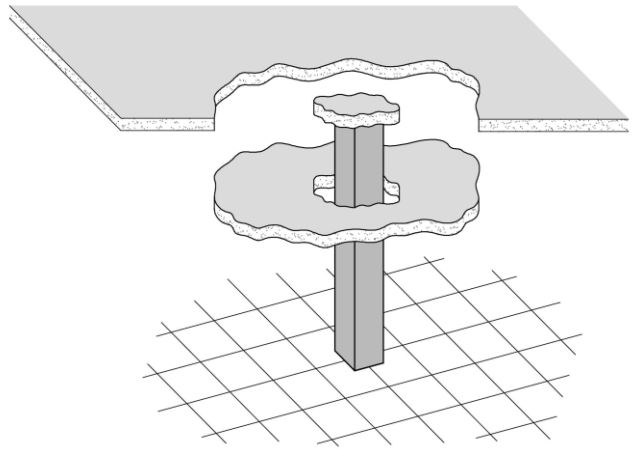
- *korzystanie z głowic w celu zwiększenia grubości stropu tak, aby rozszerzyć zakres strefy przebicia i ograniczyć niekorzystne naprężenia rozciągające*
  Rozwiązanie to wyraźnie zwiększa wytrzymałość połączenia na przebicie płyty. Jednak rozwiązanie to również zwiększa koszty budowy z kilku powodów (większa ilość betonu, bardziej skomplikowane deskowanie), a ponadto przeważnie jest sprzeczne z kryteriami architektonicznymi, w których dodatkowe poszerzenia stropu są niepożądane, w sytuacji gdy sama płyta tworzy sufit.
- *projektowanie układu konstrukcyjnego i rozkładu dylatacji tak, aby zapobiec rozprzestrzenianiu się miejscowego uszkodzenia*
  Rozwiązanie to nie jest zasadniczo skuteczne z powodu problemów ekonomicznych i ograniczeń funkcjonalnych. Stosowanie dylatacji wprowadza do konstrukcji znaczną liczbę stopni swobody skutkując koniecznością stosowania dodatkowych elementów usztywniających. Ponadto same dylatacje znacznie podnoszą koszt inwestycji i czasem mogą ograniczać funkcjonalność.
- *ograniczenie rozpiętości przęseł stropu*
  Podobnie jak w powyższym rozwiązaniu, nie zawsze można ograniczyć rozpiętość stropu, a przy pewnych jego proporcjach jest to nawet nieuzasadnione ergonomią obiektu i powszechnie stosowaną praktyką projektową.
- *zastosowanie prętów odgiętych w celu zwiększenia nośności połączenia płyta-słup na przebicie*
  Rozwiązanie to wydaje się być rozwiązaniem najbardziej ekonomicznym. Znaczna skuteczność tego typu zbrojenia potwierdzona w badaniach jest niwelowana dużymi trudnościami praktycznymi z wprowadzeniem takiego zbrojenia nad słup, gdyż wymaga skomplikowanej sekwencji jego umieszczania.
- *zwiększenia nośności strefy podporowej na przebicie*
  Metody, stosowane głownie przy zwiększaniu projektowanej nośności płyty na przebicie uwzględniają zastosowanie dodatkowych elementów przy projektowaniu połączenia tj. sztywne wkłady wykonane z profili stalowych, ukryte belki zbrojone intensywnie poprzecznie strzemionami, górne zbrojenie odgięte w dół (kotwione w przęśle płyty), trzpieniowe systemy zbrojenia stref podporowych na przebicie. Rozwiązanie to również wiąże się z dodatkowymi kosztami i wydłużeniem czasu realizacji prac budowlanych.

Wskazane powyżej metody pozwalają jedynie na zabezpieczenie połączenia płyta-słup przed wystąpieniem przebicia płyty. Jednak zastosowanie dodatkowych wzmocnień strefy podporowej nie wpływa na pracę tego połączenia w sytuacji, gdy mimo wszystko dojdzie do jego awarii.

5. Ustroje płytowo-słupowe – zbrojenie integrujące

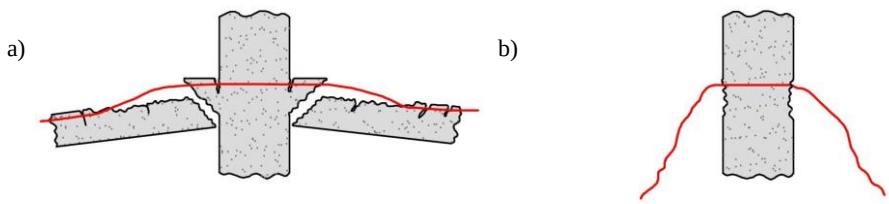
Przeprowadzone ekspertyzy po zniszczeniu obiektów wykazały, że awarie występowały wskutek wyczerpania nośności płyty w strefie podporowej. W każdym z tych przypadków do awarii dochodziło zawsze w chwili, gdy następowała utrata spójności pomiędzy

płytą i słupem, a słup przemieszczał się ponad powierzchnię górną płyty (rys. 5). Analizy różnych sytuacji, w których połączenia ulegały uszkodzeniu, pozwoliły określić mechanizm, który był przyczyną przedstawionych katastrof.



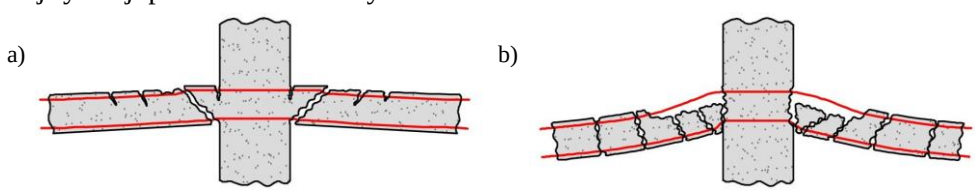
Rysunek 5. Schematyczne zachowanie się płyty po zniszczeniu połączenia ze słupem [opracowanie własne]

Zniszczenie strefy podporowej ustroju płytowo-słupowego przez przebicie płyty niezależnie od przyczyn powoduje, że w trakcie opadania stropu odrywane jest zbrojenie górne, co prowadzi do opadnięcia stropu. Zniszczenie tego typu zaprezentowano na rysunku 6.



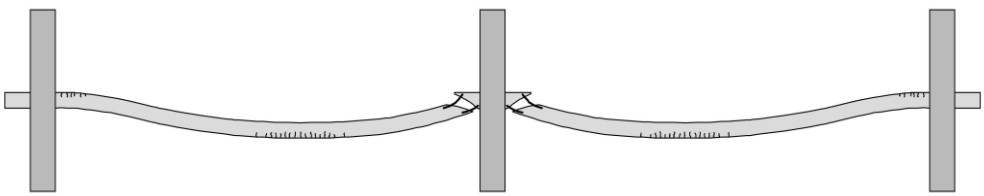
Rysunek 6. Fazy zniszczenia ustroju płytowo–słupowego po przebiciu – połączenie płyta-słup tylko ze zbrojeniem górnym: a) odrywanie prętów zbrojenia górnego bezpośrednio po przebiciu płyty, b) utrata podparcia przez pręty zbrojenia górnego i opadnięcie stropu [15]

Jedynym elementem zdolnym powstrzymać rozwój katastrofy jest zbrojenie dolne przechodzące nad słupem. Wówczas całkowite zniszczenie i opadnięcie stropu może wystąpić dopiero po zerwaniu dolnych prętów zbrojenia. Mechanizm pracy połączenia w tej sytuacji przedstawiono na rysunku 7.



Rysunek 7. Fazy zniszczenia ustroju płytowo-słupowego po przebiciu – połączenie płyta-słup zawierające zbrojenie górne i ciągłe zbrojenie dolne: a) bezpośrednio po przebiciu płyty, b) zapewnienie nośności przez ciągłe zbrojenie dolne i zawiśnięcie płyty [15]

W sytuacji awaryjnej wywołanej przebiciem płyty, dolne zbrojenie płyty przebiegające w sposób ciągły nad obszarem słupa, pozwala na przekształcenie się ustroju płytowo-słupowego w ustrój cięgnowy (membranowy), pod warunkiem, że wykonane jest ze stali o wysokiej ciągliwości (rysunek 8). Rozwiązanie takie stwarza w największym stopniu szanse na praktyczne zwiększenie nośności połączenia płyta-słup po uszkodzeniu płyty w wyniku utraty nośności na przebicie. Pręty zbrojenia integrującego ulegną odkształceniu, a następnie dzięki pracy cięgnowej zapewnią podwieszenie płyty na słupach. W przypadku braku tego zbrojenia lub jego złego wykonania nastąpi opadnięcie płyt stropowych na niższą kondygnację. Jednocześnie przedstawiony sposób zabezpieczenia konstrukcji przed rozwojem katastrofy postępującej, w przeciwieństwie do innych rozwiązań, jest szybki, tani i prosty w realizacji.



Rysunek 8. Zachowanie konstrukcji żelbetowej po zniszczeniu przez przebicie – praca cięgnowa [16]

Metody projektowania połączeń płyta-słup ze względu na przebicie doczekały się szeregu opracowań naukowych, na podstawie których opracowano procedury obliczeniowe zawarte obecnie w wytycznych normowych. Natomiast informacje w zakresie pracy połączenia płyty ze słupem po jej przebiciu są dostępne jedynie w bardzo nielicznych zagranicznych opracowaniach.

6. Wytyczne konstruowania zbrojenia integrującego

W sposób bezpośredni zagadnienie, dotyczące projektowania połączenia płyta-słup przy uwzględnieniu jego pracy po przebiciu płyty, nie zostało sprecyzowane w PN-EN 1991-1-7:2008 [1], PN-EN 1992-1-1:2008 [2], ani w nieobowiązującej już normie PN-B-03264:2002 [17]. Oprócz wymogu poprowadzenia dwóch prętów ciągłych nad słupem w każdym z dwóch prostopadłych kierunków nie podano żadnych innych wytycznych.

Procedury dotyczące obliczania pola powierzchni przekroju poprzecznego zbrojenia integrującego w połączeniach płyta-słup można należć jedynie w zagranicznych wytycznych:

- norma kanadyjska A23.3 [3];
- rekomendacja do normy amerykańskiej ACI 352.1R [4];
- wprowadzenie do normy szwajcarskiej SIA 262 [5];
- pre-norma europejska Model Code 2010 [6].

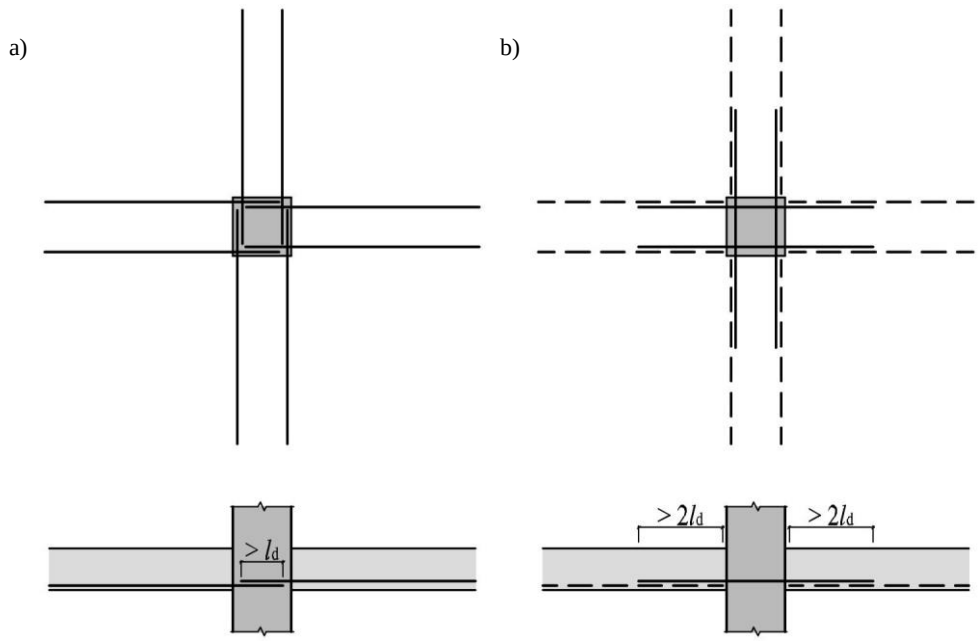
6.1. Norma kanadyjska CSA A23.3

Zbrojenie integrujące wg normy kanadyjskiej CSA A23.3 (*Canadian Standard Association*) [3] powinno składać się z co najmniej czterech krzyżujących się prętów lub ośmiu prętów niepołączonych ze sobą, ale odpowiednio zakotwionych w słupie z co najmniej dwóch prętów, które należy umieścić nad obrysem słupa i poprowadzić w każdym z dwóch kierunków. Jeżeli zbrojenie nie jest uciąglone mechanicznie, norma zaleca stosowanie odpowiedniej długości jego zakotwienia l_d zgodnie z rysunkiem 9.

Całkowite pole powierzchni przekroju poprzecznego prętów A_{sb} określa się na podstawie wzoru:

$$A_{sb} = \frac{2\,V_{se}}{f_{yk}} \tag{1}$$

gdzie: jako V_{se} należy przyjąć reakcję przekazywaną ze stropu na słup, wywołaną przez działające obciążenia o wartości nie mniejszej niż podwójny ciężar własny płyty, oraz uwzględnić charakterystyczną granicę plastyczności stali zbrojeniowej f_{yk}.



Rysunek 9. Szczegóły zakotwienia prętów zbrojenia integrującego wg [3]: a) zakład nad słupem, b) zakład poza słupem

6.2. Rekomendacja do normy amerykańskiej ACI 352.1R

Norma amerykańska ACI 318 [18] (American Concrete Institute) zawiera wyłącznie ogólne wytyczne do konstruowania zbrojenia dolnego, podobnie jak normy europejskie. Szczegółowe zalecenia zamieszczone są natomiast w rekomendacji do tej normy ACI 352.1R [4].

Zbrojenie integrujące wyznacza się na podstawie obliczeniowej wartości obciążenia równomiernie rozłożonego q_d (nie mniejszego niż dwukrotny ciężar własny płyty) z obszaru oddziaływania równego $l_1 \times l_2$, gdzie jako l_1 i l_2 stanowią rozpiętości przęseł w odpowiednich kierunkach. Pole powierzchni przekroju poprzecznego zbrojenia A_{sm}, wymaganego na jeden kierunek, wyznacza się zgodnie ze wzorem:

$$A_{sm} = \frac{0{,}5\,q_d\,l_1\,l_2}{\Phi\,f_{yk}} \tag{2}$$

gdzie: f_{yk} jest charakterystyczną granicą plastyczności stali zbrojeniowej.

Dodatkowo w obliczeniach zalecane jest przyjęcie współczynnika redukcji o wartości $\Phi = 0{,}9$. Analogicznie do wymagań normy [3] dolne pręty o polu powierzchni A_{sm} można uznać za ciągłe, jeżeli spełniają założenia przedstawione na rysunku R101.

6.3. Norma szwajcarska SIA 262

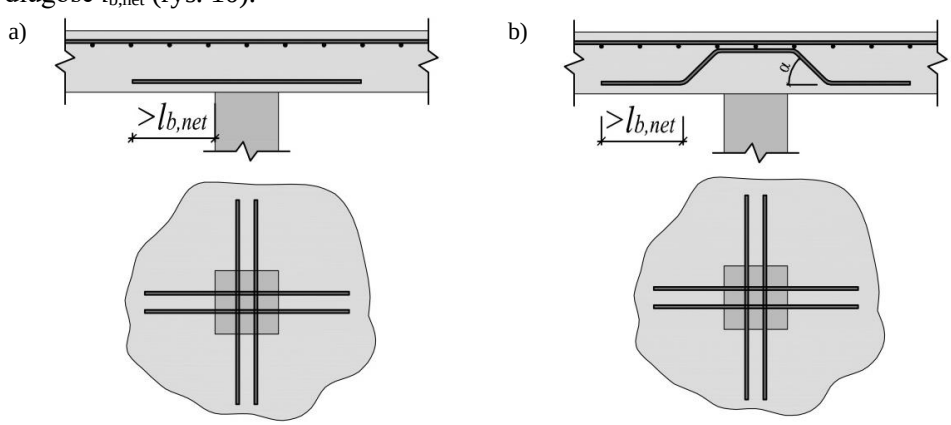
Norma szwajcarska SIA 262 [19] (*Schweizerischer Ingenieur Und Architekten-verein*) zawiera podobne podejście dotyczące stosowania zbrojenia dolnego w strefie ściskanej płyty jak wcześniej przytoczone wymagania, przy czym uszczegółowienie zapisów znajduje się we wprowadzeniu do tej normy [5]. Zbrojenie integrujące składające się z co najmniej dwóch prętów na każdym kierunku powinno być rozmieszczone nad obszarem słupa i zwymiarowane według wzoru:

$$A_{sb} = 1{,}5\frac{V_d}{f_{yd}} \tag{3}$$

gdzie: V_d jest obliczeniową wartością reakcji przekazywanej na słup, natomiast f_{yd} jest obliczeniową granicą plastyczności. Dodatkowo zbrojenie integrujące należy łączyć na zakład z dolnym zbrojeniem przęsłowym płyty.

6.4. Model Code 2010

W ostatniej wersji Model Code 2010 [6] rozszerzono, w stosunku do wcześniejszego wydania, zagadnienia związane ze zbrojeniem zabezpieczającym połączenie płyta-słup po przebiciu płyty. Zbrojenie integrujące, podobnie jak w innych wytycznych, powinno składać się z prętów poprowadzonych przez słup w każdym kierunku (minimum po dwa pręty), przy czym może być wykonane z prętów prostych lub odgiętych, zakotwionych na długość $l_{b,net}$ (rys. 10).



Rysunek 10. Szczegóły zakotwienia prętów zbrojenia integrującego wg [6]: a) z prętów prostych, b) z prętów odgiętych

Nośność połączenia płyta-słup po przebiciu wyraża się wzorem:

$$V_{Rd,int} = A_{sb} f_{yd} \left(\frac{f_t}{f_y} \right)_k \sin \alpha_{ult} \tag{4}$$

gdzie: A_{sb} jest polem powierzchni przekroju poprzecznego zastosowanych prętów, a f_{yd} obliczeniową granicą plastyczności stali zbrojeniowej.

Dodatkowo stosunek wartości wytrzymałości charakterystycznych wytrzymałości zbrojenia na rozciąganie i granicy plastyczności $(f_t/f_y)_k$ oraz charakterystycznej wartości odkształceń przy maksymalnym naprężeniu ε_{uk} określają trzy klasy ciągliwości stali zbrojeniowej:

klasa A: $(f_t/f_y)_k \geq 1{,}05$ i $\varepsilon_{uk} \geq 2{,}5\%$,

klasa B: $(f_t/f_y)_k \geq 1{,}08$ i $\varepsilon_{uk} \geq 5\%$,

klasa C: $1{,}35 \geq (f_t/f_y)_k \geq 1{,}15$ i $\varepsilon_{uk} \geq 7{,}5\%$,

na podstawie których należy ustalić teoretyczny kąt nachylenia prętów α_{ult} przechodzących nad obrysem słupa przyjmując:

$\alpha_{ult} = 0°$ – pręty proste, klasa A,

$\alpha_{ult} = 20°$ – pręty proste, klasa B,

$\alpha_{ult} = 25°$ – pręty proste, klasa C,

$\alpha_{ult} = \alpha \leq 40°$ – pręty odgięte, klasa C (α – pierwotny kąt odgięcia prętów).

7. Podsumowanie

Zawarte w powyższych normach wymagania wskazują na różne sposoby opisu zagadnienia, a w szczególności różne metody obliczeń. W efekcie tego otrzymywane są wzajemnie niespójne wyniki. Widoczny brak korelacji pomiędzy przepisami wynika nie tyle z nowości problemu, ale głównie z braku odpowiedniego rozeznania badawczego.

W tabeli 1 zestawiono zależności pozwalające na oszacowanie nośności połączenia płyta-słup po przebiciu płyty V_{pp}, wyprowadzone na podstawie szczegółowych wytycznych zawartych w normach. W wszystkich wzorach nośność połączenia uzależniona jest od pola powierzchni zbrojenia integrującego A_{sb} i własności materiałowych stali zbrojeniowej. Dodatkowo należy podkreślić, że tylko dolne zbrojenie, ciągłe i rozmieszczone bezpośrednio nad obrysem słupa, jest zdolne do zachowania znacznej nośności połączenia po przebiciu płyty.

Tabela 1. Zestawienie wzorów do obliczania nośności zbrojenia integrującego

Źródło	Określenie nośności połączenia po przebiciu płyty
CSA A23.3 [3]	$V_{pp} = \dfrac{A_{sb}\, f_{yk}}{2}$
ACI 352.1R [4]	$V_{pp} = \dfrac{A_{sb}\, \Phi\, f_{yk}}{2}$, przy czym $\Phi = 0{,}9$
SIA 262 [5]	$V_{pp} = \dfrac{A_{sb}\, f_{yd}}{1{,}5}$
MC 2010 [6]	$V_{pp} = A_{sb}\, f_{yd} \left(\dfrac{f_t}{f_y} \right)_k \sin \alpha_{ult}$

Źródło: opracowanie własne.

Wstępna analiza podanych zależności wskazuje, że obciążenie, które powinny przenieść pręty zbrojenia integrującego po przebiciu płyty, oscyluje około dwukrotnej wartości obliczeniowej siły występującej w połączeniu. Problematyczne pozostaje jednak ustalenie, jaką wartość tej siły należy przyjąć do obliczeń, co wbrew pozorom nie jest oczywiste.

Ze względu na fakt, iż każda z zależności zamieszczonych w tablicy 1, nośność połączenia płyta-słup po przebiciu płyty w inny sposób uwzględnia własności materiałowe stali zbrojeniowej trudno jednoznacznie wskazać, która z nich może stanowić najdokładniejsze oszacowanie. Z tego względu należy indywidualnie rozpatrywać każdą sytuację projektową, ustalając dla niej poszczególne wielkości: wytrzymałość na zerwanie stali f_{tk}, charakterystyczną granicę plastyczności stali f_{yk}, obliczeniową granicę plastyczności stali f_{yd}.

Ze wszystkich przytoczonych norm, jedynie w przypadku Model Code 2010 dokładnie został określony wpływ różnych parametrów wytrzymałościowych stali zbrojeniowej na nośność, którą zapewniają pręty zbrojenia integrującego po przebiciu płyty. Na tej podstawie można wnioskować, iż stal zbrojeniowa powinna charakteryzować się jak najwyższą granicą plastyczności i dużym współczynnikem wzmocnienia f_{tk}/f_{yk} oraz wydłużalnością powyżej 7,5%. Dodatkowo należy podkreślić, że pręty zbrojenia integrującego powinny być ciągłe lub zostać uciąglone mechanicznie, zwłaszcza w przypadku zastosowania dużych średnic.

Literatura

1. PN-EN 1991-1-7:2008/NA:2015-02 – Eurokod 1, *Oddziaływania na konstrukcje, Część 1-7: Oddziaływania ogólne – Oddziaływania wyjątkowe*.
2. PN-EN 1992-1-1:2008/NA:2018-11 – Eurokod 2, *Projektowanie konstrukcji z betonu, Część 1-1: Reguły ogólne i reguły dla budynków*.
3. CSA Standard A23.3-14, *Design of concrete structures*, Canadian Standard Association, Canada.
4. ACI 352.1R-89 (Reapproved 1997), *Recommendations for design of slab-column connections in monolithic reinforced concrete structures*, American Concrete Institute, ACI-ASCE Committee 352, USA.
5. *Construction en béton, Introduction à la norme SIA 262*, Extrait de la documentation D 0182, Société suisse des ingénieurs et des architectes, Zürich, Switzerland, 2003.
6. CEB-FIB – *Model Code 2010, Final draft*, Volume 1 and 2, Bulletin 65 and Bulletin 66 of the fib Model Code for Concrete Structures, 2010.
7. https://builderpolska.pl/2019/01/07/stal-w-architekturze/ [data dostępu: 19.01.2023]
8. Wood J.G.M., *Pipers Row Car Park, Wolverhampton – Quantitative Study of the Causes of the Partial Collapse on 20th March 1997*.
9. http://news.bbc.co.uk/2/hi/uk_news/england/west_midlands/3392397.stm [data dostępu: 19.01.2023].
10. Gardner N.J., Jungsuck H., Chung L., *Lessons from the Sampoong department store collapse, Cement and Concrete Composites*, Vol. 24, No. 6, 2002, s. 523-529.
11. https://en.wikipedia.org/wiki/Sampoong_Department_Store_collapse [data dostępu: 19.01.2023].
12. *Investigation of October 30, 2003 Fatal Parking Garage Collapse at the Tropicana Casino Resort*, Atlantic City, NJ, U.S. Department of Labor Occupational Safety and Health Administration, 2004.
13. https://ars.els-cdn.com/content/image/1-s2.0-S1687404812000077-gr1.jpg [data dostępu: 19.01.2023].
14. Brandt A.M., Glinicki M.A., *Ocena jakości betonu w budynku, który uległ katastrofie w wyniku przemrożenia beto*nu, Przegląd Budowlany, 11, 2007, s. 50-54.
15. Wieczorek B., *Wewnętrzna strefa podporowa żelbetowego ustroju płytowo-słupowego w stanie awaryjnym wywołanym przebiciem płyty*, Wydawnictwo Politechniki Śląskiej, Gliwice 2019.

16. Mitchell D., Cook W.D., *Preventing progressive collapse of slab structures*, Journal of Structural Engineering, Vol. 110, No. 7, 1984, s. 1513-1532.
17. PN-B-03264:2002, *Konstrukcje betonowe, żelbetowe i sprężone. Obliczenia statyczne i projektowanie*.
18. ACI 318-19, *Building Code Requirements for Structural Concrete*, ACI 318M-14, American Concrete Institute, ACI Committee 318, USA.
19. SIA 262:2013, *Construction en béton*, Swiss Standards Association, Switzerland.

Zabezpieczenie połączenia płyta-słup konstrukcji żelbetowych przed katastrofą postępującą w świetle przepisów normowych

Streszczenie

W artykule przedstawiono zagadnienia dotyczące połączenia płyty ze słupem w monolitycznych żelbetowych konstrukcjach w sytuacji awaryjnej. Konstrukcje tego typu wykazują bardzo dużą podatność na uszkodzenia wynikające z pękania betonu na skutek przeciążenia elementów konstrukcji lub innych czynników zewnętrznych. Ich awarie najczęściej związane są z przebiciem płyty wokół słupa na skutek wyczerpania nośności betonu stropu. Omówiono wybrane awarie obiektów budowlanych, w których uszkodzenie jednego z elementów konstrukcji doprowadziło do rozprzestrzenienia się zniszczeń na kolejne fragmenty obiektu i w konsekwencji przyczyniło się do zawalenia się części tego obiektu, a w niektórych przypadkach objęło swoim zasięgiem nawet cały obiekt. Wskazano najczęstsze przyczyny, które spowodowały wystąpienie tych katastrof i ich skutki. Określono możliwe sposoby zabezpieczenia połączenia płyta-słup przed wstąpieniem przebicia w strefie podporowej. A także podano rozwiązanie dotyczące stosowania dodatkowego zbrojenia integrującego, które zapewnia nośność połączenia płyta-słup nawet po wystąpieniu jego uszkodzenia. Przedstawiono wytyczne odnośnie projektowania strefy podporowej według Eurokodu PN-EN 1991-1-7:2008 i Eurokodu PN-EN 1992-1-1:2008. Dodatkowo omówiono przepisy zawarte w zagranicznych normach dotyczące stosowania dodatkowego dolnego zbrojenia krzyżującego się bezpośrednio nad słupem, aby zapewnić nośność połączenia płyty ze słupem po jej uszkodzeniu wskutek przebicia. W szczególności przedstawiono zapisy normy kanadyjskiej CSA A23.3, rekomendacji do normy amerykańskiej ACI 352.1R, wprowadzenia do normy szwajcarskiej SIA 262 oraz Model Code 2010.

Słowa kluczowe: konstrukcja płytowo-słupowa, połączenie płyta-słup, zbrojenie integrujące

Protection of the slab-column connection of reinforced concrete structures against a progressive catastrophe in relation to standard regulations

Abstract

The article presents issues related to the connection of a slab with a column in monolithic reinforced concrete structures in an emergency situation. Structures of this type are very susceptible to damage resulting from cracking of concrete as a result of overloading the structure elements or other external factors. Their failures are most often associated with puncture of the slab around the column as a result of exhausting the bearing capacity of the concrete floor. Selected failures of buildings were discussed, in which damage to one of the structural elements led to the spread of damage to other parts of the building and, as a consequence, contributed to the collapse of a part of this building, and in some cases even covered the entire building. The most common causes that caused the occurrence of these disasters and their consequences were indicated. Possible methods of securing the slab-column connection against puncturing in the support zone were specified. A solution was also given for the use of additional integrating reinforcement, which ensures the load-bearing capacity of the slab-column connection even after its failure. Guidelines for the design of the support zone according to Eurocode PN-EN 1991-1-7:2008 and Eurocode PN-EN 1992-1-1:2008 are presented. In addition, the provisions contained in foreign standards regarding the use of additional bottom reinforcement crossing directly above the column to ensure the load-bearing capacity of the connection between the slab and the column after its damage due to puncture are discussed. In particular, the provisions of the Canadian CSA A23.3 standard, recommendations to the American standard ACI 352.1R, introduction to the Swiss standard SIA 262 and Model Code 2010 were presented.

Keywords: slab-column structure, slab-column connection, integrity reinforcement

Barbara Wieczorek[1]

Przegląd i porównanie wyników badań eksperymentalnych połączenia płyta-słup monolitycznych konstrukcji żelbetowych w odniesieniu do wytycznych normowych

1. Wprowadzenie

Ze względu na estetykę zewnętrzną i wewnętrzną obiektu, a także nieskomplikowany kształt pozwalający na szybką realizację inwestycji, konstrukcje żelbetowe płytowo-słupowe znajdują powszechne zastosowanie w budownictwie. Niestety monolityczne ustroje tego typu są bardzo podatne na uszkodzenia związane z pękaniem betonu w wyniku przeciążenia elementów konstrukcji, nieodpowiedniej wytrzymałości betonu lub innych czynników zewnętrznych.

Przegląd awarii obiektów budowlanych [1-6] wykonanych jako ustroje płytowo-słupowe uwidocznił, iż w wielu przypadkach ich przyczyną było przebicie płyty wokół słupa na skutek wyczerpania nośności betonu stropu. W wyniku uszkodzenia jednego z połączeń płyta-słup nastąpiło rozprzestrzenienie się zniszczeń na kolejne fragmenty konstrukcji, co w konsekwencji wywołało zawalenie znacznej części obiektu tj. katastrofę postępującą. Zatem istotne jest zapewnienie możliwości wytworzenia wtórnego ustroju konstrukcyjnego po lokalnym jej uszkodzeniu, a w rozpatrywanym przypadku zapewnienie nośności połączenia płyty ze słupem nawet po wystąpieniu jego awarii przez przebicie płyty.

Jednym ze sposobów zapewnienia nośności połączenia płyta-słup również po jego awarii spowodowanej przebiciem płyty jest zastosowanie odpowiednio skonstruowanego zbrojenia dolnego w postaci prętów krzyżujących się bezpośrednio nad słupem i odpowiednio zakotwionych w płycie.

Głównym celem pracy jest omówienie, na podstawie dostępnej literatury, przeprowadzonych dotychczas badań eksperymentalnych żelbetowych modeli połączeń płyta-słup, w których zastosowano pręty zbrojenia integrującego. Zamieszczono opis modeli badawczych i uzyskane rezultaty badań w stanie awaryjnym po ich zniszczeniu przez przebicie płyty.

Zestawienie ograniczono wyłączne do tych badań eksperymentalnych, które dotyczyły wewnętrznego połączenia płyta-słup, tak aby zachować odpowiedniość z istniejącymi wytycznymi. W przypadku niektórych badań wystąpiły okoliczności, które mogły w istotny sposób wpłynąć na uzyskanie wyników obarczonych błędem, na co wskazywali ich autorzy.

Przedstawione wyniki badań eksperymentalnych odniesiono do wytycznych zawartych w zagranicznych normach, dotyczących stosowania zbrojenia integrującego, w szczególności zapisów normy kanadyjskiej CSA A23.3 [7], rekomendacji do normy amerykańskiej ACI 352.1R [8], wprowadzenia do normy szwajcarskiej SIA 262 [9] oraz Model Code 2010 [10].

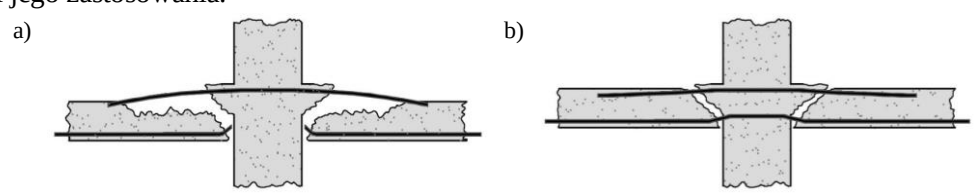
[1] barbara.wieczorek@polsl.pl, Politechnika Śląska, Wydział Budownictwa, www.polsl.pl.

2. Zabezpieczenie konstrukcji płytowo-słupowych – zbrojenie integrujące

W literaturze przedmiotu znajduje się wiele opracowań dotyczących bezpośrednio zjawiska przebicia, do którego dochodzi w płytach żelbetowych wskutek przeciążenia konstrukcji, czy też błędów projektowych lub technologicznych. Skupiają się one głównie na analizie zjawiska przebicia płyty oraz metodach wzmocnienia strefy podporowej wokół słupa. Ich zastosowanie nie zabezpiecza jednak przed katastroficznymi skutkami, jeśli mimo wszystko wystąpi utrata nośności płyty ze względu na uszkodzenia betonu.

Niezależnie od przyczyny zniszczenie strefy podporowej ustroju płytowo-słupowego przez przebicie płyty powoduje, że w trakcie opadania stropu odrywane jest zbrojenie górne od powierzchni płyty, co w konsekwencji prowadzi do utraty spójności pomiędzy płytą i słupem. Jedynym możliwym rozwiązaniem zabezpieczającym przed wystąpieniem katastrofy jest wprowadzenie dodatkowego zbrojenia integrującego w postaci prętów zbrojenia dolnego przechodzących na śladem słupa. Wówczas całkowite zniszczenie i opadnięcie stropu nastąpi dopiero po zerwaniu tych dolnych prętów zbrojenia.

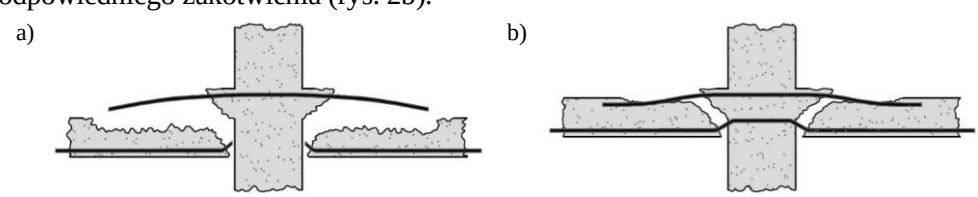
Na rysunku 1 przedstawiono zachowanie się połączenia płyta-słup w sytuacji awaryjnej wywołanej przebiciem płyty, w przypadku braku użycia zbrojenia integrującego, jak i jego zastosowania.



Rysunek 1. Stan połączenia płyta-słup bezpośrednio po przebiciu płyty a) tylko ze zbrojeniem górnym, b) zawierającego zbrojenie górne i ciągłe zbrojenie dolne [11]

W połączeniach, które zawierały wyłącznie zbrojenie górne, bez prętów ciągłych zbrojenia dolnego bezpośrednio nad słupem, po zniszczeniu przez przebicie pręty zbrojenia górnego są odrywane od górnej powierzchni płyty, co prowadzi do całkowitej utraty nośności płyty w tej strefie, co prowadzi do opadnięcia stropu (rys. 2a).

W przypadku zniszczenia przez przebicie połączenia płyta-słup z dolnym zbrojeniem integrującym, w stanie awaryjnym całkowite obciążenie będzie przenoszone przez dolne zbrojenie podłużne. Pręty zbrojenia integrującego ulegną odkształceniu, a następnie zapewnią podwieszenie płyty na słupach. Natomiast górne zbrojenie podporowe może potencjalnie przyczynić się do zwiększenia nośności tego połączenia pod warunkiem odpowiedniego zakotwienia (rys. 2b).



Rysunek 2. Zachowanie się połączenia płyta-słup po przebiciu: a) opadnięcie stropu po wyrwaniu prętów zbrojenia górnego, b) zawiśnięcie płyty zapewnione przez ciągłe zbrojenie dolne [11]

Zatem w sytuacji awaryjnej będącej skutkiem przebicia płyty, zbrojenie integrujące przebiegające w sposób ciągły nad obszarem słupa, pozwala na przekształcenie się ustroju płytowo-słupowego w ustrój cięgnowy, a tym samym zabezpiecza przed wystąpieniem katastrofy postępującej.

3. Wytyczne konstruowania zbrojenia integrującego

W obowiązujących normach PN-EN 1991-1-7:2008 [12] i PN-EN 1992-1-1:2008 [13] zamieszczono jedynie zalecenie poprowadzenia dwóch prętów ciągłych nad słupem w każdym z dwóch prostopadłych kierunków, natomiast nie podano żadnych innych wytycznych. Bardziej szczegółowe zapisy można znaleźć jedynie w zagranicznych opracowaniach:

- norma kanadyjska Canadian Standard Association A23.3 [7];
- rekomendacja do normy amerykańskiej Recommendations for American Standard ACI 352.1R [8];
- wprowadzenie do normy szwajcarskiej SIA 262 [9];
- pre-norma europejska Model Code 2010 [10].

Zawierają one szczegółowe zasady dotyczące obliczania pola powierzchni przekroju poprzecznego zbrojenia integrującego, jakie należy zastosować w połączeniu płyta-słup na wypadek wystąpienia sytuacji awaryjnej wynikającej z przebicia płyty.

Podobnie, jak podaje Eurokod [13], zbrojenie integrujące powinno składać się z co najmniej dwóch prętów, które należy umieścić nad obrysem słupa i poprowadzić w każdym kierunku. W przypadku, gdy zbrojenie nie jest uciąglone mechanicznie zalecane jest stosowanie odpowiedniej długości jego zakotwienia w płycie.

Przy wymiarowaniu prętów integrujących należy uwzględnić wartość reakcji przekazywanej ze stropu na słup, wywołanej przez działające obciążenia na konstrukcję o wartości nie mniejszej niż podwójny ciężar własny płyty. Dodatkowo należy uwzględnić granicę plastyczności stali zbrojeniowej.

Zależność pozwalającą na oszacowanie nośności połączenia płyta-słup po przebiciu płyty V_{pp}, wyprowadzono na podstawie szczegółowych wytycznych zawartych w [7-10]. We wszystkich wzorach nośność połączenia po przebiciu płyty uzależniona jest od pola powierzchni zbrojenia integrującego A_{sb} i własności materiałowych stali zbrojeniowej. Z podanych zależności wynika, że obciążenie, które powinny przenieść pręty zbrojenia integrującego po przebiciu płyty, odpowiada w przybliżeniu około dwukrotnej wartości obliczeniowej siły oddziałującej na połączenie płyty ze słupem.

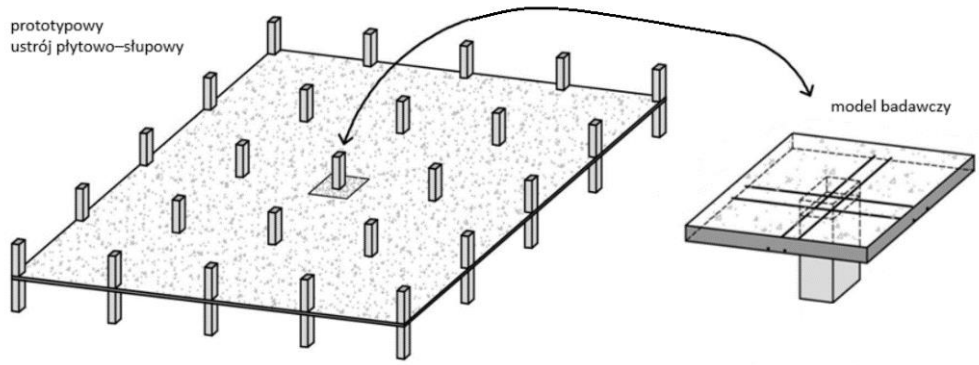
4. Badania eksperymentalne połączenia płyta-słup

Zagadnienie dotyczące określenia siły powodującej wyczerpanie nośności płyty na przebicie w monolitycznych konstrukcjach żelbetowych było przedmiotem wielu badań eksperymentalnych m.in. [14-19] obliczeń numerycznych oraz analiz teoretycznych zarówno za granicą, jak i w Polsce. Uzyskanie rezultaty stanowiły podstawę do opracowania wytycznych do projektowania zawartych w normach.

Natomiast w pracach naukowych praktycznie pomijany był aspekt zachowania się połączenia płyta-słup i jego dalszej pracy w stanie awaryjnym wywołanym przebiciem, gdzie obciążenie będzie przenoszone wyłącznie przez zbrojenie integrujące. W literaturze rozważania w tym zakresie są mało liczne i w wielu przypadkach zawierają niekompletne informacje.

4.1. Model badawczy i stanowisko

W dostępnej literaturze i omówionych poniżej badaniach zagadnienie nośności monolitycznych połączeń płyty ze słupem przeprowadzane było na elementach stanowiących fragment rzeczywistego ustroju płytowo-słupowego (rys. 3). Model badawczy przyjmowano jako część płyty wyciętą wokół słupa wewnętrznego w odległości od jego osi wynoszącej 0,2÷0,22 [18] rzeczywistej rozpiętości siatki słupów, z usytuowanym od dołu słupem.



Rysunek 3. Prototypowy ustrój płytowo-słupowy i model badawczy [11]

Zbrojenie płyty w modelu badawczym ustala się, jak w przypadku rzeczywistej konstrukcji płytowo-słupowej przy uwzględnieniu całkowitego obciążenia projektowego. Dodatkowo nad słupem wprowadza się pręty zbrojenia integrującego po dwa w każdym kierunku. Również parametry betonu ustala się, jak typowych konstrukcji żelbetowych.

Badania przeprowadzane są w warunkach laboratoryjnych na odpowiednio skonstruowanym stanowisku. Model zostaje mocowany na obwodzie, aby zapewnić blokadę przemieszczenia płyty, a obciążenie przykładane jest u podstawy słupa i wywierane za pomocą siłownika. Proces obciążania obejmuje dwie fazy: pierwszą do uzyskania przebicia płyty oraz drugą, w której obciążenie zwiększa się aż do zerwania prętów zbrojenia dolnego krzyżującego się nad słupem.

Wartość obciążenia, przy której dochodzi do zniszczenia zbrojenia integrującego przyjmuje się jako nośność połączenia płyta-słup po przebiciu płyty.

Oczywiście powyższy opis prowadzenia badań jest bardzo ogólnikowy i praktycznie w każdych z poniżej przedstawionych badań był analogiczny. W zależności od przyjętej metodologii badawczej może jedynie różnić się szczegółami technicznymi w zakresie wymiarów geometrycznych modelu (płyty i słupa), parametrami wytrzymałościowymi stali zbrojeniowej i betonu, a także sposobu mocowania modelu na stanowisku i zastosowanego systemu obciążania. Z tego względu dokonanie porównania rezultatów badań prowadzonych w różnych ośrodkach badawczych wymaga bardzo szczegółowych danych.

4.2. Przykłady badań eksperymentalnych

W literaturze dostępne są opisy tylko nielicznych serii badań, które zostały przeprowadzone w celu poznania zachowania połączenia płyty ze słupem z zastosowanym zbrojeniem integrującym. Wszystkie zostały wykonane na modelach stanowiących fragment płyty żelbetowej, w których w pierwszej kolejności wywołano stan awaryjny w postaci przebicia płyty i określono wartość siły niszczącej połączenie V_p, a następnie wyznaczono wartość siły V_{pp} przy której nastąpiło zerwanie prętów zbrojenia poprowadzonego na słupem.

Wśród opublikowanych dotąd prac w omawianym zakresie na uwagę zasługują badania przeprowadzone przez:

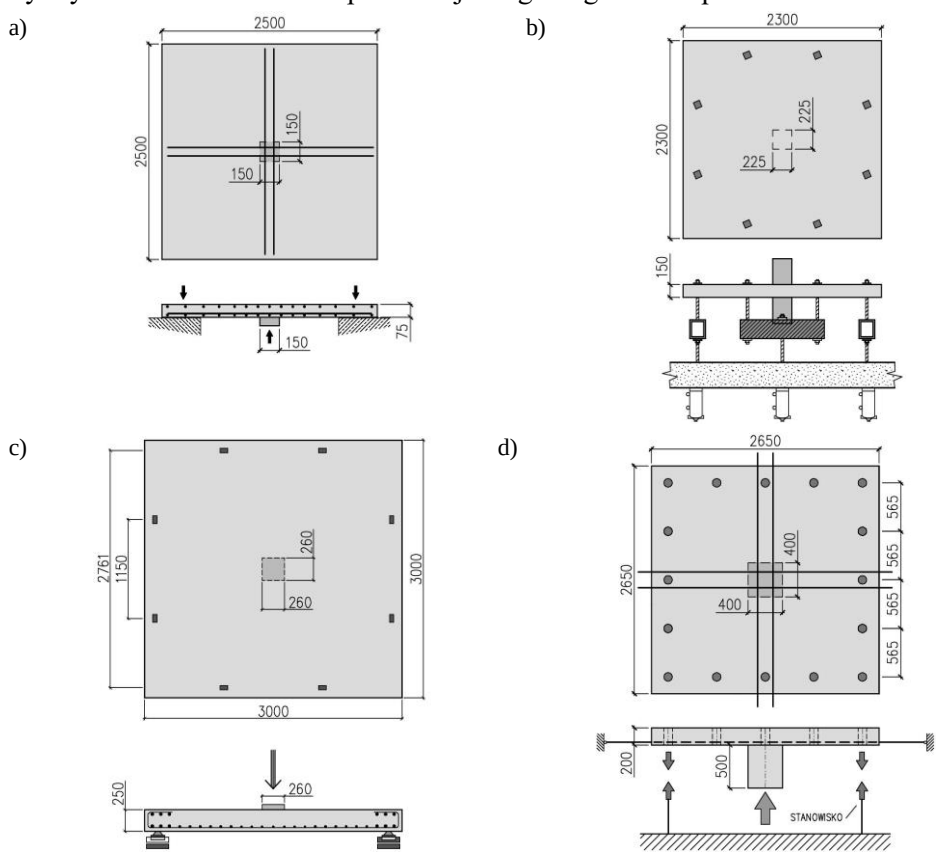
- Melo i Regan (1998) [20]

Badania obejmowały pięć modeli wykonanych w skali 1:4. Modele składały się z kwadratowej płyty o boku 2,5 m i grubości 75 mm, natomiast w środku modeli znaj-

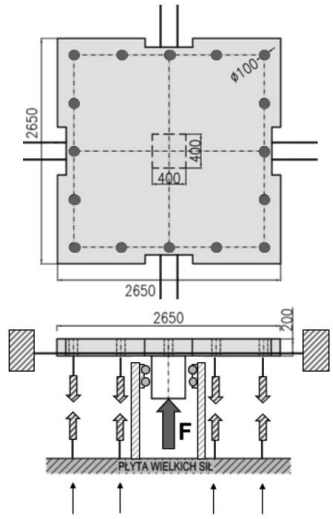
dował się słup o kwadratowym przekroju 150 × 150 mm (rys. 4a). Podparcie zastosowano na słupie oraz na krawędziach płyty. Obciążenie przyłożono w 16 punktach na płycie. W każdym z modeli zastosowano identyczne zbrojenie główne, które zaprojektowano zgodnie z wymaganiami normy brytyjskiej. Parametry zbrojenia integrującego w każdym z modeli były inne. Składało się z prętów czterech prętów ø6 lub ø8 o różnych parametrach stali zbrojeniowej. Dodatkowo przeprowadzono badania dwóch modeli bez zbrojenia integrującego, w celach porównawczych. Na podstawie badań stwierdzono, że we wstępnej fazie przebicia płyty nie zaobserwowano znaczącej różnicy między modelami ze zbrojeniem integrującym lub bez zbrojenia integrującego. W fazie po przebiciu płyty modele zawierające zbrojenie integrujące miały znacznie większą nośność.

- Ghannoum (1998) [21]

W ramach badań wykonano sześć modeli w skali 3:4, które składały się z kwadratowych płyt o wymiarach 2,3 × 2,3 m i grubości 150 mm ze słupem o przekroju 225 × 225 mm (rys. 4b). W każdym z nich jako zbrojenie górne zastosowano taką samą liczbę prętów zbrojeniowych, przy czym zastosowano dwa różne sposoby ich rozmieszczenia. We wszystkich modelach jako zbrojenie integrujące zastosowano po trzy pręty w dwóch kierunkach o przekroju poprzecznym 100 mm^2, z tej samej stali zbrojeniowej. We wszystkich przypadkach nośność połączenia po przebiciu płyty uzyskano praktycznie taką samą. Natomiast otrzymano różne wartości siły przebijającej, na co miała wpływ wytrzymałość betonu oraz stopień zbrojenia górnego nad słupem.



e)



Rysunek 4. Schemat modeli i stanowiska w badaniach: a) Melo i Regan [20], b) Ghannoum [21], c) Mirzaei [22], d) Starosolski i in. [23], e) Wieczorek [11]

- Mirzaei (2008) [22]

Przeprowadzono badania modeli połączenia płyta-słup wykonanych w skali 1:2. Wszystkie modele wykonano jako kwadratowe płyty o boku 1,5 m i grubości 125 mm, bez słupa (rys. 4c). Płyty były podparte na ośmiu łożyskach tocznych i obciążane pionowo przez siłownik na górnej ich powierzchni. We wszystkich modelach zastosowano pręty ø8 mm jako zbrojenie górne. Zbrojenie integrujące składało się z prętów prostych (PM-9÷PM-12, PM-25÷PM28) odpowiednio o średnicach ø8, ø10, ø12 i ø14 oraz (PM-21÷PM22) o średnicach ø8 i ø10, które różniły się parametrami zbrojeniowej. Zaobserwowano, że na nośność połączenia po przebiciu płyty miała bardzo duży wpływ nośność prętów na zerwanie. W przypadku modeli serii PM-25÷PM28 podczas prowadzenia badań wystąpiło uszkodzenie kotwienia prętów lub stwierdzono wadę fabryczną prętów.

- Starosolski i in. (2009) [23]

Wykonano dwa identyczne modele, składające się z płyty o wymiarach 2,65 × 2,65 m i grubości 20 cm z kwadratowym słupem 40 × 40 cm długości 50 cm (rys. 4d). W jednym modelu PI/16-1 zastosowano w każdym kierunku po dwa pręty ø16 ze stali walcowanej na gorąco, w drugim PII/12-1 – czterech pręty ø12 ze stali zimnowalcowanej. Podczas prowadzenia badań modelu PII/12-1 dokonano odkręcenia śrub mocujących model do stanowiska. Ze względu na problemy występujące w czasie prowadzenia badań modelu PI/16-1 konieczne było odcięcie prętów zbrojenia górnego, a następnie przerwanie badań z uwagi na przechylenie słupa. Zerwanie prętów zbrojenia integrującego uzyskano przy sile o ok. 10% mniejszej niż ta podana w tabeli 1.

- Wieczorek (2012) [11]

Badania prowadzono na trzech modelach wykonanych w skali 1:1. Model stanowiła płyta żelbetowa o wymiarach 2,65 × 2,65 × 0,2 m z umieszczonym od dołu słupem o wysokości 50 cm i przekroju kwadratowym 40 × 40 cm (rysunek 4d). Zbrojenie górne i dolne było identyczne we wszystkich modelach. Zbrojenie integrujące stanowiły dwa pręty ø16 w każdym kierunku z tej samej stali zbrojeniowej. Modele były badane na

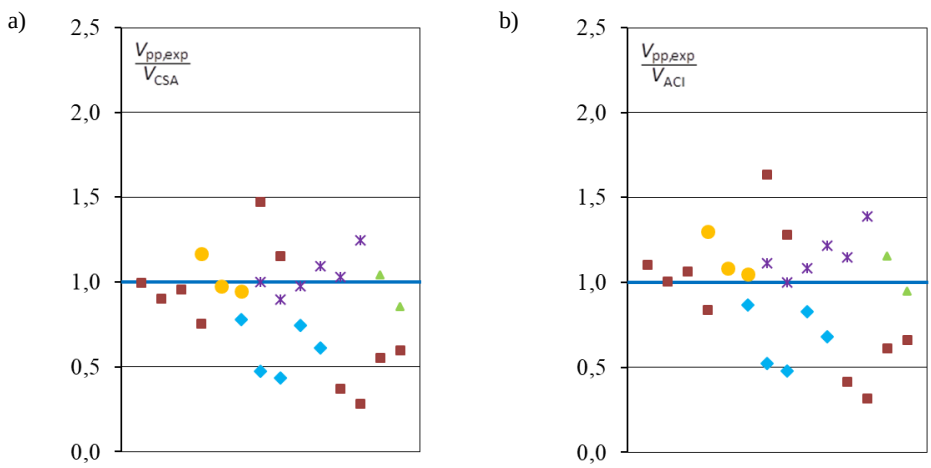
stanowisku, którego konstrukcja zapewniała pionowe przemieszczanie słupa ponad płytę, w fazie po przebiciu płyty. Obciążenie było wywierane u podstawy słupa za pomocą siłownika hydraulicznego. W każdym z modeli zastosowano inną metodologię obciążania po przebiciu płyty, co wpłynęło na wartości sił, przy których nastąpiło zerwanie prętów zbrojenia integrującego. Stwierdzono, że na nośność połączenia po przebiciu płyty, zapewnioną przez pręty zbrojenia integrującego, miały wpływ parametry stali zbrojeniowej, która była klasy C o dużej ciągliwości.

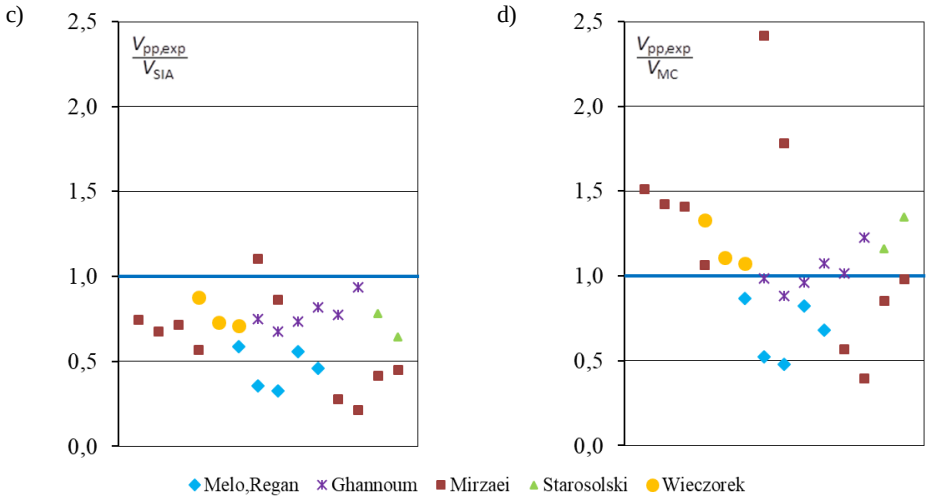
5. Porównanie wyników badań eksperymentalnych i oszacowań normowych

Na podstawie informacji o przedstawionych badaniach eksperymentalnych dokonano porównania ich rezultatów z wartościami sił, jakie po przebiciu płyty powinny przenieść pręty zbrojenia integrującego według procedur określonych w normach [7-10].

W celu przeprowadzenia analizy określono wartości sił $V_{pp,norm}$, które wyznaczono jako nośność połączenia według zależności (1) – (4) odpowiednio wg wytycznych normy: CSA A23.3 [7] (V_{CSA}), ACI 352.1R [8] (V_{ACI}), SIA 262 [9] (V_{SIA}) i MC 2010 [10] (V_{MC}). Do obliczeń przyjęto całkowite pole powierzchni prętów integrujących oraz granicę plastyczności stali na podstawie wyników badań materiałowych podanych w publikacjach.

Powyższe siły $V_{pp,norm}$ odniesiono do uzyskanych w ramach badań eksperymentalnych nośności połączeń, jakie otrzymano po przebiciu płyty $V_{pp,exp}$, które odpowiadały zerwaniu prętów zbrojenia integrującego (tab. 1) i przestawiono na rysunku 4.





Rysunek 4. Porównanie nośności połączenia, określonej eksperymentalnie i na podstawie normy, a) CSA A23.3 [7], b) ACI 352.1R [8], c) SIA 262 [9], d) MC 2010 [10] [opracowanie własne]

Tabela 1. Zestawienie nośności połączeń wyznaczonych na podstawie wzorów normowych dla modeli z badań eksperymentalnych

Oznaczenie modelu	Skala modelu	$V_{pp,exp}$ [kN]	V_{CSA} [kN]	V_{ACI} [kN]	V_{SIA} [kN]	V_{MC} [kN]
Melo i Regan [20]						
MiR2	1:4	64	85,8	77,2	114,4	77,6
MiR3	1:4	81	171,6	154,4	228,8	155,1
MiR4	1:4	66	152,5	137,3	203,4	137,9
MiR5	1:4	65	106,3	95,7	141,7	95,5
MiR7	1:4	83	106,3	95,7	141,7	95,5
Ghannoum [21]						
S1-U	3:4	273	272,4	245,2	363,2	277,3
S1-B	3:4	245	272,4	245,2	363,2	277,3
S2-U	3:4	266	272,4	245,2	363,2	277,3
S2-B	3:4	298	272,4	245,2	363,2	277,3
S3-U	3:4	281	272,4	245,2	363,2	277,3
S3-B	3:4	340	272,4	245,2	363,2	277,3
Mirzaei [22]						
PM-9	1:2	123	123,8	111,4	165,1	81,2
PM-10	1:2	159	175,8	158,3	234,5	111,8
PM-11	1:2	237	247,8	223,0	330,4	168,0
PM-12	1:2	245	324,3	291,9	432,4	230,1
PM-21	1:2	185	125,6	113,0	167,5	76,6
PM-22	1:2	219	190,0	171,0	253,3	122,8
PM-25*	1:2	75	125,6	113,0	167,5	76,6
PM-26*	1:2	105	190,0	171,0	253,3	122,8
PM-27*	1:2	94	252,8	227,5	337,0	166,1
PM-28*	1:2	101	355,7	320,2	474,3	254,3
Starosolski [23]						
PI/16-1**	1:1	445	439,7	395,7	586,3	386,2
PII/12-1**	1:1	386	439,7	395,7	586,3	386,2
Wieczorek [11]						

P-I/M	1:1	514	439,7	395,7	586,3	386,2
P-I/N	1:1	428	439,7	395,7	586,3	386,2
P-I/MN	1:1	415	439,7	395,7	586,3	386,2

* uszkodzenie kotwienia prętów lub ich wada fabryczna, ** odkręcenie modelu od stanowiska lub odcięcie prętów zbrojenia górnego. Źródło: opracowanie własne.

Na podstawie wyników zamieszczonych w tabeli 1 stwierdzono, że wśród obliczeń, zgodnie z wzorami normowymi (tab. 1), najlepsze zgodności z wynikami przedstawionych serii badań eksperymentalnych uzyskano według normy ACI 352.1R [8], gdzie średni stosunek $V_{pp,exp}/V_{pp,norm}$ wyniósł 1,0140 (rysunek 4b). Jednak wynik ten uzyskano przy bardzo dużym współczynniku zmienności, równym prawie 34%. Wartości obciążeń niszczących wyznaczone wg norm CSA A23.3 [7] i SIA 262 [9] były mniejsze od wyników badań eksperymentalnych, odpowiednio na poziomie 0,9126 (rys. 4a) i 0,6844 (rys. 4c). Średni stosunek nośności połączenia uzyskanego w badaniach eksperymentalnych do obliczonej według MC 2010 [10] był większy od jedności i wyniósł 1,0080 (rys. 4d), przy czym współczynnik zmienności zwiększył się do 40%. Może być to związane z faktem, iż w MC 2010 [10] w większym zakresie uwzględniane są parametry wytrzymałościowe stali zbrojeniowej, niż w ACI352.1R [8].

Na otrzymany rezultat mógł wpłynął również fakt, iż na populację wyników składały badania modeli, które zostały wykonane przez różnych badaczy. Ponadto badania te były prowadzone w różnorodnych metodologiach badawczych, a także przy zastosowaniu odmiennych kształtów i wymiarów modeli oraz użyciu różnych materiałów i przyrządów pomiarowych. Powyższe uzasadnia spostrzeżenie, że poszczególne serie badawcze były oczywiście obarczone mniejszymi rozrzutami.

6. Podsumowanie

Normy CSA A23.3 [7], ACI 352.1R [8] i SIA 262 [9] określają nośność połączenia płyta-słup po przebiciu płyty $V_{pp,norm}$ na podstawie wartości siły jaką mogą przenieść na słup pręty zbrojenia integrującego. W zależności od podejścia, wartość tej siły jest korygowania współczynnikiem, który wynosi odpowiednio „2,0" w CSA A23.3 [7], „2,22" w ACI 352.1R [8] oraz „1,5" w SIA 262 [9]. Stąd widoczne jest, iż zapisy normy ACI 352.1R [8] są najbardziej restrykcyjne. Z tego też względu uzyskano najlepsze zgodności wyników badań eksperymentalnych z oszacowaniami normowymi. Odmienne podejście zastosowano w wytycznych MC 2010 [10], w których – w przeciwieństwie do wcześniej wspomnianych norm – nośność połączenia po przebiciu płyty uzależniona została od wartości parametru stali, jakim jest ciągliwość. Ponadto jedynie w MC 2010 [10] uwzględnia się obliczeniową granicę plastyczności stali f_{yd}, która może stanowić o dodatkowej nośności połączenia po przebiciu płyty jaką zapewniają pręty zbrojenia integrującego.

Ze wszystkich przytoczonych wytycznych projektowych [7]÷[10], jedynie w oparciu o przepisy Model Code 2010 [10] uzyskuje się najbezpieczniejsze oszacowanie nośności połączenia płyta-słup po przebiciu– lecz pod warunkiem zastosowania zbrojenia ciągłego.

Na podstawie przedstawionej analizy można zauważyć, iż według norm ilość zbrojenia integrującego, jaką należałoby zastosować w konstrukcji w celu zapewnienia odpowiedniej nośności połączenia po przebiciu płyty, była w wielu przypadkach zaniżona. Może to prowadzić do niebezpiecznego zawyżenia nośności obliczeniowej dla tej fazy pracy konstrukcji.

Niemniej jednak uzyskane rezultaty badań wskazały na konieczność stosowania dodatkowego zbrojenia poprowadzonego bezpośrednio nad słupem na wypadek sytuacji awaryjnej w połączeniu płyta-słup po wystąpieniu przebicia płyty.

Literatura

1. Wood J.G.M., *Pipers Row Car Park*, Wolverhampton – Quantitative Study of the Causes of the Partial Collapse on 20th March 1997.
2. Gardner N.J., Jungsuck H., Chung L., *Lessons from the Sampoong department store collapse*, Cement and Concrete Composites, 24(6), 2002, s. 523-529.
3. Investigation of October 30, 2003 Fatal Parking Garage Collapse at the Tropicana Casino Resort, Atlantic City, NJ, U.S. Department of Labor Occupational Safety and Health Administration, 2004.
4. Brandt A.M., Glinicki M.A., *Ocena jakości betonu w budynku, który uległ katastrofie w wyniku przemrożenia betonu*, Przegląd Budowlany, 11, 2007, s. 50-54.
5. Subramanian N., *The failures that changed the perception of our designs*, The Bridge and Structural Engineer, 45(4), 2014, s. 23-44.
6. Subramanian N., *Alternative punching shear reinforcement for RC flat slabs*, The Indian Concrete Journal, 88(1), 2014, s. 33-44.
7. (1) CSA Standard A23.3-14, *Design of concrete structures*, Canadian Standard Association, Canada.
8. ACI 352.1R-89 (Reapproved 1997), *Recommendations for design of slab-column connections in monolithic reinforced concrete structures*, American Concrete Institute, ACI-ASCE Committee 352, USA.
9. *Construction en béton, Introduction à la norme SIA 262*, Extrait de la documentation D 0182, Société suisse des ingénieurs et des architectes, Zürich, Switzerland, 2003.
10. CEB-FIB – *Model Code 2010: Final draft*, Volume 1 and 2, Bulletin 65 and Bulletin 66 of the fib Model Code for Concrete Structures, 2012.
11. Wieczorek B., *Wewnętrzna strefa podporowa żelbetowego ustroju płytowo-słupowego w stanie awaryjnym wywołanym przebiciem płyty*, Wydawnictwo Politechniki Śląskiej, Gliwice, 2019.
12. PN-EN 1991-1-7:2008/NA:2015-02 – Eurokod 1, *Oddziaływania na konstrukcje, Część 1-7: Oddziaływania ogólne – Oddziaływania wyjątkowe.*
13. PN-EN 1992-1-1:2008/NA:2018-11 – Eurokod 2, *Projektowanie konstrukcji z betonu, Część 1-1: Reguły ogólne i reguły dla budynków.*
14. Kinnunen S., Nylander H., *Punching of concrete slabs without shear reinforcement*, Transactions of the Royal Institute of Technology, No. 158, Stockholm 1960.
15. Moe J., *Shearing strength of reinforced concrete slabs and footings under concentrated loads*, Development Department Bulletin, No. D47, Portland Cement Association, Skokie, 1961.
16. Muttoni A., *Punching shear strength of reinforced concrete slabs without transverse reinforcement*, ACI Structural Journals, 105(4), 2008, s. 440-450.
17. Godycki-Ćwirko T., *Mechanika betonu*, Arkady, Warszawa 1982.
18. Urban T.S., *Przebicie w żelbecie*, Wybrane zagadnienia, Wydawnictwo Politechniki Łódzkiej, Łódź 2005.
19. Hulimka J., *Strefa podporowa żelbetowego stropu płaskiego o podwyższonej nośności na przebicie*, Wydawnictwo Politechniki Śląskiej, Gliwice 2009.
20. Melo G.S., Regan P.E., *Post-punching* resistance *of connections between flat slabs and interior columns*, Magazine of Concrete Research, 50(4), 1998, s. 319-327.
21. Ghannoum C.M., *Effect of high-strength concrete on the performance of slab-column specimens*, Department of Civil Engineering, McGill University, Montreal, QC, 1998.

22. Mirzaei Y., Muttoni A., *Tests of the post-punching behavior of the reinforced concrete flat slabs*, IS-BETON, Switzerland 2008.
23. Jasiński R., Kupczyk R., Starosolski W, Wieczorek M., *Badania zachowania się strefy podporowej monolitycznych ustrojów płytowo – słupowych w stadium awaryjnym*, Biuletyn Informacyjny Centrum Promocji Jakości Stali, Nr 2, Warszawa 2009.

Badania eksperymentalne połączenia płyta-słup monolitycznych konstrukcji żelbetowych w odniesieniu do wytycznych normowych

Streszczenie
W artykule omówiono opublikowane w literaturze prace badawcze, które zostały przeprowadzone w celu poznania zachowania połączenia płyty ze słupem, w którym zastosowano zbrojenie integrujące. Badania te przeprowadzane były na modelach stanowiących fragment rzeczywistego ustroju płytowo-słupowego, który obejmował wewnętrzną strefę przysłupową. Przytoczono wytyczne normowe dotyczące stosowania zbrojenia integrującego, w celu zapewnienia nośności połączenia płyta-słup po jego uszkodzeniu wskutek przebicia płyty. W szczególności przedstawiono zapisy normy kanadyjskiej CSA A23.3, rekomendacji do normy amerykańskiej ACI 352.1R, wprowadzenia do normy szwajcarskiej SIA 262 oraz Model Code 2010. Na podstawie dostępnych informacji o badaniach eksperymentalnych dokonano porównania ich rezultatów z wartościami sił, jakie po przebiciu płyty powinny przenieść pręty zbrojenia integrującego według procedur określonych w normach. Stwierdzono, że ilość zbrojenia integrującego, jaką należałoby zastosować w celu zapewnienia odpowiedniej nośności połączenia po przebiciu płyty przez pręty zbrojenia integrującego, była w wielu przypadkach zaniżona. Uzyskane rezultaty badań wskazały na konieczność stosowania tego dodatkowego zbrojenia poprowadzonego bezpośrednio nad słupem na wypadek ewentualnej sytuacji awaryjnej.
Słowa kluczowe: konstrukcja płytowo-słupowa, połączenie płyta-słup, zbrojenie integrujące

Experimental tests of slab-column connection of monolithic reinforced concrete structures in relation to standard guidelines

Abstract
The article discusses the research works published in the literature, which were carried out in order to understand the behavior of the connection between the slab and the column, in which integrity reinforcement was used. support zone. Standard guidelines for the use of integrity reinforcement have been quoted in order to ensure the load-bearing capacity of the slab-column connection after its damage due to punching slab. In particular, the provisions of the Canadian CSA A23.3 standard, recommendations to the American standard ACI 352.1R, introduction to the Swiss standard SIA 262 and Model Code 2010 were presented. On the basis of the available information on the experimental tests, the results were compared with the values of forces which, after punching the slab, should be transferred by the integrity reinforcement bars according to the procedures specified in the standards. It was found that the amount of integrity reinforcement that would have to be used was in many cases underestimated in order to ensure the proper load-bearing capacity of the connection after punching by the integrity reinforcement bars. The obtained test results indicated the need to use this additional reinforcement placed directly above the column in the event of an emergency situation.
Keywords: slab-column structure, slab-column connection, integrity reinforcement

Indeks Autorów

Buchwald P. 241
Bugaj M. 75
Dawid A. 241
Dmowska-Jasek P. 173, 184
Drozdowski G. 302
Dziatkiewicz G. 230
Dziekański P. 302
Gac J.M. 254
Gajdzik B. 314
Gulewicz M. 286
Jackiewicz-Zagórska A. 254
Jermacz W.J. 152
Kałdoński T. 32
Konieczna M. 95
Krukowski M. 55, 105
Kucharska-Jastrząbek A. 173, 184
Lewandowski A. 152
Marczak H. 130
Meisner N. 55
Nowak B. 254
Nowak J. 85
Różańska A. 17, 141
Senyk S. 32
Słowik M. 286
Sokołowska B. 198
Steckiewicz K. 267
Szumielewicz K. 75
Turoboś F. 55, 105
Walkowicz M. 17, 141
Wartecki A. 152
Warzybok P. 230
Waśniowska J. 7
Werner Ł. 254
Wieczorek B. 328, 340
Wilczek A. 85
Wróbel K. 75