30 września 2024 r.

dr hab. inż. Artur Klepaczko, prof. ucz.
Instytut Elektroniki
Politechniki Łódzkiej

RECENZJA ROZPRAWY DOKTORSKIEJ
DLA RADY DYSCYPLINY INŻYNIERIA BIOMEDYCZNA
W POLITECHNCE ŚLĄSKIEJ

Tytuł rozprawy: Opracowanie metodyki przetwarzania sygnałów akustycznych i danych obrazowych dla celów komputerowego wspomagania diagnostyki logopedycznej z wykorzystaniem technik sztucznej inteligencji

Autor rozprawy: mgr inż. Agata Sage
Promotor: dr hab. inż. Paweł Badura, prof. PŚ

1. Ocena układu rozprawy doktorskiej, w tym informacje o jej poszczególnych częściach składowych

Przedłożona do recenzji rozprawa obejmuje 9 rozdziałów, bibliografię oraz 2 dodatki zawierające szczegółowe wyniki doświadczeń. Układ rozdziałów jest przejrzysty. Autorka spójnie wyjaśnia cele, motywację, zastosowany aparat badawczy oraz wyniki kolejnych etapów badań. We wprowadzeniu zamieszczonym w rozdziale 1. jasno sformułowano podjęte w rozprawie zagadnienie badawcze, jakim jest komputerowe wspomaganie diagnostyki logopedycznej u dzieci z wadami wymowy tzw. głosek dentalizowanych. Problem ten jest ważny i aktualny, ponieważ dotyczy znaczącego odsetka populacji, natomiast niewyeliminowane w wieku rozwojowym przeszkody językowe wpływają negatywnie na kondycję intelektualną i psychiczną danej osoby w całym okresie dojrzewania, a także w wieku dorosłym. Diagnostyka logopedyczna wymaga zbadania wielu złożonych czynników, które powodują powstanie zaburzeń. Stąd omawiane przez Autorkę rozwiązania poszerzają katalog metod, jakimi mogą posługiwać się lekarze logopedzi podczas leczenia.

W rozdziale 2. sformułowano tezy badawcze (tzn. tezę główną oraz tezy pomocnicze), cel pracy doktorskiej, a także w sposób syntetyczny podsumowano oryginalne osiągnięcia rozprawy. Osiągnięcia te potwierdzają zamieszczone w kolejnych rozdziałach wyniki badań. Ponadto opisano zastosowaną w pracy metodę badawczą oraz zaprojektowany przepływ danych przez kolejne kroki ekstrakcji cechy i ich analizy prowadzące do osiągnięcia zamierzonego celu. W rozdziale 3. opisano przygotowany na potrzeby pracy wielomodalny zbiór danych, jak również skonstruowany w tym celu audio-wizyjny system do ich pozyskiwania. W tym miejscu wskazano również zakres udziału Doktorantki w projektowaniu urządzenia. Opisano także protokoły rejestracji danych wielomodalnych oraz towarzyszących im badań logopedycznych przeprowadzanych przez lekarzy specjalistów. Badania te przeprowadzono w ramach projektu badawczego finansowanego przez Narodowe Centrum Nauki.

Rozdziały 4–6 zawierają opisy metod zastosowanych do wykonania postawionych zadań. W szczególności, rozdział 4. dedykowany danym obrazowym opisuje m.in. metody detekcji i segmentacji artykulatorów mowy z wykorzystaniem sztucznych sieci neuronowych, a następnie przedstawia różnorodne modele cech (tekstury, kształtu, itp.), które zastosowano do analizy wyróżnionych w obrazach obszarów zainteresowania. Podobnie, rozdział 5. poświęcono pozyskanym sygnałom akustycznym. Rozdział 6. z kolei przedstawia teoretyczne podstawy analizy statystycznej wykonanej dla pozyskanych danych pomiarowych.

W rozdziale 7. zamieszczono wszechstronną i wnikliwą prezentację uzyskanych wyników, których krytyczną dyskusję przeprowadzono w rozdziale 8. Następnie w podsumowującym rozdziale 9. oprócz odniesienia uzyskanych wyników do celów i tez pracy, wskazano możliwe kierunki dalszych prac oraz potencjalne szersze zastosowania proponowanego podejścia do analizy obrazów i sygnałów audio.

2. Ocena zastosowanego piśmiennictwa w ramach rozprawy doktorskiej

Zamieszczony w pracy przegląd literatury przedmiotu jest stosunkowo obszerny. Znajdują się tu odniesienia zarówno do prac dedykowanych diagnostyce logopedycznej, jak i do badań dotyczących analizy sygnału mowy w szerszym ujęciu, wykorzystujących różnorodne algorytmy rozpoznawania mowy czy wspomagania uczenia prawidłowej wymowy w danym systemie językowym. Z przeprowadzonego przeglądu wynika także, że podejście przyjęte przez Autorkę, polegające na ekstrakcji i analizie cech jednocześnie obrazowych i akustycznych stanowi oryginalny kierunek badań. Istotnym walorem proponowanego podejścia na tle cytowanego piśmiennictwa jest jego nieinwazyjność. Nie wymaga ono umieszczania w jamie ustnej badanego żadnych czujników, jak ma to miejsce w przypadku innych rozwiązań.

Ponieważ postulat wykorzystania cech obrazowych wymaga m.in. użycia metod segmentacji artykulatorów, osobną część przeglądu poświęcono temu właśnie zagadnieniu. Tu na pierwszy plan wysuwają się algorytmy głębokiego uczenia, w tym w szczególności sztuczne sieci neuronowe typu U-Net lub wykorzystujące jednostki GRU. Warto zaznaczyć, że omówienie tej technologii w dysertacji pojawia się tylko właśnie w kontekście segmentacji obrazów. Tymczasem, istotną częścią pracy jest ekstrakcja cech, a ta stanowi o sile algorytmów głębokiego uczenia, niezależnie od ich końcowego zastosowania. Cechy utajone obliczane przez przykładowo wariacyjne sieci autoenkodujące, choć nie posiadają bezpośredniej formalnej interpretacji, wykazują zdolności do rozróżniania semantycznych wzorców w analizowanych obrazach. **Czy znane są Autorce podobne podejścia w specyficznym zastosowaniu do przetwarzania obrazów twarzy, w tym części aparatu mowy?**

3. Wskazanie oraz ocena celu pracy kandydata

W odpowiedzi na wyzwania diagnostyki logopedycznej w rozprawie zarysowano potrzebę zaprojektowania narzędzi, które w sposób nieinwazyjny umożliwiłyby zautomatyzowanie i zobiektywizowanie oceny cech wymowy. Autorka postuluje zastosowanie w tym celu szeregu cech obrazowych oraz sygnałów akustycznych. Przeprowadzone badania prowadzą de facto do zidentyfikowania deskryptorów numerycznych pozwalających na rozróżnianie grup o określonym typie zaburzenia wymowy (a także grupy wymowy normatywnej) w ramach tzw. szeregów głosek dentalizowanych – syczącego, szumiącego i ciszącego. Zaprojektowany na

potrzeby realizacji tego zadania tor przetwarzania danych w zamierzeniu Autorki ma stanowić realizację celu nadrzędnego pracy, który sformułowano w rozdziale 2. jako:

> *Opracowanie metodyki przetwarzania sygnałów akustycznych i danych obrazowych z wykorzystaniem metod sztucznej inteligencji.*

Osiągnięcie tego celu ma z kolei służyć potwierdzeniu tezy głównej rozprawy o następującym brzmieniu:

> *Istnieją istotne statystycznie różnice w cechach sygnałów akustycznych i danych obrazowych prezentujących mowę dzieci z różnymi (normatywnymi i nienormatywnymi) cechami mowy,*

wspartej dwiema tezami pomocniczymi:

> *1. Możliwa jest wiarygodna segmentacja wybranych artykulatorów w obrazach twarzy z wykorzystaniem metod sztucznej inteligencji.*
>
> *2. Ekstrakcja i analiza cech obrazowych 2D i 3D oraz parametrów akustycznych pozwala na określenie różnic między grupami w wybranych cechach artykulacyjnych.*

O ile treść tezy głównej oraz tez pomocniczych nie budzi większych zastrzeżeń (poza wątpliwością związaną z użyciem jakościowej i względnej kategorii wiarygodności w odniesieniu do wyników segmentacji, która to kategoria pozostawia pole do swobodnej interpretacji), o tyle sformułowanie celu pracy wymagałoby moim zdaniem korekty. Wydaje się, że intencją Autorki jest zaproponowanie zestawu metod do przetwarzania obrazów i sygnałów akustycznych, umożliwiających efektywną ekstrakcję cech danych wielomodalnych o dużej zdolności dyskryminacyjnej. Ponadto, w takim ujęciu, zaproponowany sposób postępowania mógłby być uogólniony i zastosowany w innych kontekstach niż rozważana w pracy diagnostyka logopedyczna. Jednakże, Autorka koncentruje się na konkretnych metodach przetwarzania na każdym kolejnym etapie ścieżki przepływu danych. **W tym sensie, przedstawione opracowanie stanowi metodę, a nie metodykę, zgodnie z poniższymi definicjami**[1]:

> ***Metoda**. 1) Sposób postępowania, dobór rodzaju działania (dobór i kolejność czynności składowych działania złożonego) świadomie stosowany z możliwością powtórzenia go we wszystkich przypadkach danego typu; 2) zespół czynności i środków zastosowanych w określony sposób dla osiągnięcia określonego celu; 3) sposób wykonania danego zadania praktycznego; 4) zespół założeń ogólnych przyjmowanych jako ramy lub wytyczne badania. (...) Metodę w czwartym, najbardziej wąskim znaczeniu, określa się jako metodę naukową – sposób badania, wyznaczający etapy postępowania uczonych w danej dziedzinie naukowej.*
>
> ***Metodyka**. W znaczeniu ogólnym – zespół wytycznych dotyczących sposobów postępowania, efektywnych ze względu na określony cel. W metodologii nauk – używana niekiedy nazwa norm (przepisów) określających właściwe danej nauce sposoby postępowania.*

[1] W. Baturo i in. (red.), *Nowa encyklopedia powszechna PWN*, t. 5, Wydawnictwo Naukowe PWN, Warszawa 2004.

W pracy nie pojawiają się żadne rekomendacje, czy wytyczne do projektowania metod przetwarzania danych akustyczno-obrazowych. Cel taki byłby ryzykowny i nader ambitny. Zrealizowano natomiast konkretny algorytm, na który składają się etapy pozyskiwania danych, badania logopedyczne, metody segmentacji, ekstrakcji cech itp. Biorąc jednak pod uwagę kontekst zarysowany we wprowadzeniu do rozprawy i wynikający z treści tez, intencja stojąca za tak zdefiniowanym celem pracy wydaje się jasna i sformułowanie to może być przyjęte, pod warunkiem rozszerzenia definicji metodyki o sekwencję metod.

4. Wskazanie oraz ocena zastosowanych metod badawczych

We wprowadzeniu do rozprawy Autorka w sposób kompetentny i na odpowiednim poziomie szczegółowości, przedstawiła cechy wymowy normatywnej, a także możliwe odstępstwa od niej. Na uwagę zasługuje bardzo jasne przedstawienie techniki artykulacji głosek oraz znaczenie w tym procesie poszczególnych artykulatorów – ust, warg, języka oraz zębów. Zamieszczone opisy, choć syntetyczne, świadczą o dogłębnym zrozumieniu przez Autorkę dziedziny, w której przeprowadziła badania. Ponadto, informacje tu zawarte uzasadniają na poziomie ogólnym dobór metod przetwarzania sygnałów i obrazów dokonany na potrzeby rozwinięcia własnego rozwiązania wspomagającego diagnostykę logopedyczną.

Zastosowane metody badawcze zostały szczegółowo opisane w rozdziałach 4–6. Pierwszym etapem przetwarzania (po pozyskaniu danych) jest segmentacja artykulatorów. Do tego celu, Autorka proponuje w pierwszej kolejności użyć tzw. sieci YOLOv6. Zadaniem tego algorytmu jest detekcja i określenie prostokątnych obszarów obejmujących wargi, język i zęby. Dokładna segmentacja poszczególnych artykulatorów odbywa się za pomocą sieci DeepLabv3+, której trening przebiega dwuetapowo. Wewnątrz obszarów znalezionych przez sieć YOLO przeprowadzana jest zgrubna segmentacja metodą zbiorów poziomicowych. Wyniki tej segmentacji posłużyły jako etykiety treningowe w uczeniu sieci segmentacyjnej, która w drugim etapie dostrajana była za pomocą precyzyjnych adnotacji wykonanych przez ekspertów. Procedura ta w mojej ocenie jest poprawna, choć nie jedyna możliwa.

Na uznanie zasługuje z powodzeniem zrealizowana koncepcja uczenia sieci DeepLabv3+ w trybie słabego nadzoru, a następnie zastosowanie uczenia transferowego. Nie budzi zastrzeżeń także sposób przygotowania danych, w tym podziału pacjentów na 3 grupy, osobno do uczenia sieci detekcyjnej, segmentacyjnej oraz testowej. **Niejasne jedynie pozostaje, czy w ramach np. uczenia sieci DeepLab v3+, do uczenia której użyto danych z podzbioru oznaczonego w pracy literą B (1092 ramki pochodzące od 25 dzieci), zachowano rozdział między pacjentami należącymi do zbiorów treningowych i walidacyjnych**. Ponadto, zasadniczym zadaniem realizowanym w omawianym kroku analizy jest semantyczna segmentacja lub segmentacja instancji w obrazach twarzy. **Powstaje pytanie, czy wstępny krok detekcji obiektów realizowany przez sieć YOLO jest w ogóle konieczny.** Współczesne architektury transformerowe, jak np. SegFormer, OneFormer, wykorzystujące mechanizmy uwagi krzyżowej lub własnej (ang. *cross-, self-attention*) pozwalają uzyskać wyniki o bardzo dużej zgodności z rzeczywistymi maskami instancji, nawet jeśli liczba pikseli im odpowiadająca jest dużo mniejsza od obszarów tła. Wydaje się, że nawet klasyczne już sieci typu U-Net mogłyby być wytrenowane tak, aby znalezienie dokładnych granic obszarów poszukiwanych obiektów odbywało się w jednym kroku.

Po znalezieniu masek artykulatorów następuje ekstrakcja cech. Jak zwróciłem uwagę w pkt. 1 recenzji, Autorka skupia się w tym wymiarze pracy na cechach kształtu i tekstury. W tym drugim przypadku, wykorzystano szereg różnorodnych modeli, obejmujących cechy pierwszego i wyższych rzędów, czyli parametry wyliczane na podstawie poziomów szarości, histogramu jasności, macierzy zdarzeń, macierzy długości serii, macierzy jednorodnych stref pikseli oraz macierzy różnic jasności w sąsiedztwie. Wszystkie wybrane modele są uznanymi metodami opisu tekstury**. Jednakże koncepcja, iż określone wady wymowy mogą kształtować teksturę obrazu nie wydaje się być podparta jednoznacznymi przesłankami wynikającymi z procesów biofizycznych zachodzących podczas artykulacji głosek.** Hipoteza ta nie znalazła także mocnego potwierdzenia w wynikach doświadczeń, gdzie parametry tekstury wykazały zdolność dyskryminacyjną tylko w przypadku jednej głoski z szeregu syczącego i dwóch głosek w szeregu szumiącym, przy czym wielkość efektu w teście Kruskalla-Wallisa w każdym przypadku była znacznie mniejsza niż dla cech innych niż teksturowe.

Kolejną wątpliwość stanowi zespół cech kształtu 3D. Dane trójwymiarowe analizowane w pracy powstały przez złożenie sąsiednich ramek zapisu wideo. Jako takie więc nie tworzą rzeczywistego obrazu trójwymiarowego. Z treści rozprawy nie wynika także, aby modele trójwymiarowe badanych organów były rekonstruowane na podstawie serii obrazów płaskich, co poniekąd byłoby możliwe (i interesujące!) z wykorzystaniem nowoczesnych modeli sztucznej inteligencji. **Dlatego nie do końca jasna dla mnie pozostaje interpretacja cech kształtu 3D, takich jak objętość, sferyczność, maksymalna średnica w płaszczyźnie osiowej, czy strzałkowej.**

Druga grupa parametrów opisana w rozdziale 5. była obliczana na podstawie sygnału akustycznego. Tutaj, wyznaczono szereg cech w dziedzinie czasu, parametry częstotliwościowe (zarówno spektralne, jak i melcepstralne) oraz parametry w pasmie szumu. Zarówno dobór parametrów, jak i przetwarzanie wstępne danych opisano w sposób klarowny i nie budzący wątpliwości.

W rozdziale 6. z kolei opisano metody statystyczne, które posłużyły do oceny zdolności poszczególnych atrybutów akustycznych i obrazowych do rozróżniania grup mówców w obrębie poszczególnych cech artykulacyjnych. W pierwszej kolejności wykonano badanie rozkładów wszystkich cech za pomocą testu normalności Shapiro-Wilka, który wykazał asymetryczność większości rozkładów. W drugiej części badań wykonano szereg testów do oceny jednorodności rozkładów. Tutaj zastosowano test U Manna-Whitneya (dla podziału danych na 2 grupy) lub Kruskala-Wallisa (w przypadku podziału na 3 lub więcej grup). Zastosowana metoda badawcza w sposób pełny i rzetelny pozwala ocenić statystyczną zdolność pojedynczych parametrów akustyczno-obrazowych do dyskryminacji grup wewnątrz określonej cechy artykulacyjnej. Niewątpliwą zaletą pracy jest wykonanie obliczeń w obrębie 4 oddzielnych szeregów głosek dentalizowanych, co ma istotny wpływ na zastosowanie metody w diagnostyce logopedycznej. Słabą stroną jednak tego podejścia jest ujęcie poszczególnych cech w izolacji od pozostałych. Bardzo często w uczeniu maszynowym spotykamy się z sytuacją, w której pojedyncze cechy ujawniają swoją zdolność dyskryminacyjną dopiero, gdy towarzyszą im inne cechy niezależne. **Czy były podejmowane próby (nieparametrycznej) analizy wieloczynnikowej?**

5. Ocena części rozprawy doktorskiej dotyczącej omówienia wyników badań

Autorka opisała swoje osiągnięcia badawcze na odpowiednim poziomie szczegółowości stawianym rozprawom doktorskim. Wyniki rzetelnie przedstawione w rozdziale 7 oraz w dodatkach A i B w sposób przekonujący dowodzą słuszności postawionych tez. Zamieszczone w tych częściach pracy tabele pozwalają zidentyfikować cechy akustyczne i obrazowe, które – parafrazując tezę główną i tezę pomocniczą nr 2 – wykazują *istotne statystycznie różnice między grupami w wybranych cechach artykulacyjnych*. Na bardzo dobrą ocenę zasługuje wszechstronne, wyczerpujące i systematyczne zebranie najważniejszych wyników ilościowych przeprowadzonych testów statystycznych. Dzięki temu praca nie pozostawia żadnych wątpliwości dotyczących wiarygodności wniosków końcowych rozprawy.

Pozytywne wyniki uzyskane przez Doktorantkę również w zakresie segmentacji obrazów bronią zaproponowanego dwuetapowego podejścia opartego na metodach sztucznej inteligencji, co dowodzi słuszności tezy pomocniczej nr 1. **Drobne zastrzeżenia budzi jedynie warstwa prezentacyjna wykresów zamieszczonych na rys. 7.5b i 7.7b**. Osiągnięte rezultaty dokładności dla wszystkich artykulatorów, zastosowanych metod oraz sieci szkieletowych są na tyle duże, że umieszczenie ich na wykresach w całej rozciągłości możliwej skali (0-1) pozbawia te wykresy czytelności i utrudnia porównanie ze sobą poszczególnych wariantów. Prawdopodobnie korzystniejsze byłoby umieszczenie tych wyników w tabelach.

Doktorantka przeprowadziła również szeroką dyskusję uzyskanych wyników w rozdziale 8. Rozdział ten świadczy o umiejętności krytycznej oceny własnej pracy przez Doktorantkę, co jest niezwykle istotne w pracy naukowej.

6. Informacje dotyczące praktycznego zastosowania uzyskanych wyników badań

Wspomaganie diagnostyki logopedycznej, jak wskazano w niniejszej recenzji oraz w samej rozprawie, stanowi bardzo ważne zagadnienie w wymiarze społecznym. Dlatego praktyczna implementacja algorytmów rozróżniania cech wymowy normatywnej i zaburzonej, opartych na zidentyfikowanych przez Autorkę parametrach obrazowo-akustycznych ma potencjalnie duże znaczenie dla zwiększenia skuteczności badań diagnostycznych. Z tego względu należy pozytywnie ocenić współpracę zespołu z Katedry Informatyki Medycznej i Sztucznej Inteligencji, w skład którego wchodzi Doktorantka z ekspertami w dziedzinie logopedii, potencjalnymi odbiorcami opracowanej technologii.

7. Informacja o ewentualnych nieprawidłowościach, które pojawiły się w ocenianej rozprawie doktorskiej

Praca jest napisana w sposób bardzo poprawny. Do jej słabszych strony można zaliczyć wątpliwości wyrażone powyżej:

1) Brak odniesień w przeglądzie literatury do podejść wykorzystujących tzw. cechy utajone w analizie obrazów twarzy.
2) Brak informacji o rozkładzie pacjentów w podziałach na zbiór treningowy i walidacyjny przy uczeniu sieci detekcyjnej i segmentacyjnej.
3) Brak uzasadnienia dla wykonywania wstępnej detekcji artykulatorów mowy przed ich właściwą segmentacją.

4) Brak uzasadnienia dla obliczania cech teksturowych.
5) Niejasna interpretacja obliczanych cech kształtu 3D.
6) Brak statystycznej analizy wieloczynnikowej na etapie oceny zdolności dyskryminacyjnej cech akustyczno-obrazowych.

Język tekstu rozprawy nie budzi zastrzeżeń. Można doszukać się jedynie nielicznych błędów językowych lub logicznych o małej wadze:

a) s. 14: *Drugi etap dotyczył segmentacji obiektów stosując sieć DeepLabv3+.*
Błąd logiczny (anakolut). Zdanie to w obecnym kształcie sugeruje, że etap dotyczy segmentacji **poprzez** stosowanie sieci DeepLabv3+. Tymczasem etap dotyczy segmentacji i zastosowano w nim taki, a nie inny rodzaj sieci neuronowej.
b) s. 16: *Każde z istotnych zagadnień rozpoczyna się jest od wprowadzenia teoretycznego.*
Powinno być: … *rozpoczyna się od …*
c) s. 20: *Charakter urządzenia pozwala na wykonywanie ruchów głowy, zachowując jednocześnie stabilność położenia czujników…*
Błąd logiczny (anakolut). Podmiotem zdania jest charakter, ale to nie on zachowuje stabilność położenia.
d) s. 22: …*wykorzystywano urządzenie wizyjno-akustycznego…*
Powinno być: *wizyjno-akustyczne…*
e) s. 46: *Obiektywne cechy opisujące dane…*
Czy mogą występować cechy **nieobiektywne**?
f) s. 48: *W przeciwieństwie do cech związanych z kształtem, parametry teksturowe zamiast pojedynczych pikseli wykorzystują ich zbiorowiska.*
Cechy kształtu również oblicza się na podstawie zbiorów pikseli.
g) s. 87: *Niższa wartość dla detekcji zębów nie jest istotna ze względu na najmniejszą istotność i wkład tego artykulatora w kolejne etapy przetwarzania.*
Powtórzenie – *istotna/istotność.* Przymiotniki wysoki, niski odnoszą się do wysokości (np. budynku) lub czyjegoś wzrostu. W przypadku wartości wielkości fizycznych czy wartości numerycznych używa się kategorii *mały*, *duży*. Powinno być: *Mniejsza wartość…*

8. Ocena, czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego

Oryginalność rozprawy ogniskuje się moim zdaniem wokół trzech osi badawczych:

1) Zastosowanie wielomodalnych danych w ocenie sygmatyzmu u dzieci. Należy podkreślić wykorzystanie danych obrazowych obok sygnałów akustycznych, najczęściej stosowanych we wcześniejszych badaniach.
2) Opracowanie autorskiej metody segmentacji artykulatorów mowy z wykorzystaniem sieci neuronowych głębokiego uczenia oraz technik uczenia słabo nadzorowanego i transferowego.
3) Wszechstronna analiza statystyczna parametrów akustyczno-obrazowych, dzięki której możliwe było zidentyfikowanie atrybutów najbardziej znaczących dla dyskryminacji grup w cechach artykulacyjnych dla poszczególnych głosek dentalizowanych.

W każdej z wymienionych osi, osiągnięcia Doktorantki stanowią jej autorski wkład do stanu wiedzy światowej i jako całość stanowią oryginalne rozwiązanie zdefiniowanego w rozprawie problemu badawczego.

9. Ocena, czy rozprawa doktorska prezentuje ogólną wiedzę teoretyczną kandydata w dyscyplinie albo dyscyplinach oraz umiejętność samodzielnego prowadzenia pracy naukowej lub artystycznej

Zarówno zawartość merytoryczna rozprawy, jak i jej logiczny układ świadczą o osiągnięciu przez Autorkę umiejętności do samodzielnego prowadzenia prac naukowych oraz pozyskania odpowiedniej wiedzy w dyscyplinie inżynieria biomedyczna. Treść kolejnych rozdziałów wskazuje ponadto, że Doktorantka posiada bardzo dużą wiedzę z zakresu przetwarzania sygnałów i obrazów, uczenia maszynowego, eksploracji danych oraz statystki matematycznej, a więc posiada warsztat badawczy gwarantujący odpowiednią jakość ewentualnych przyszłych prac naukowych bądź innych aktywności zawodowych wymagających podobnego doświadczenia.

10. Wnioski końcowe

Stwierdzam, że przedłożona do recenzji rozprawa doktorska spełnia wymagania Ustawy Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 roku, Dziennik Ustaw Nr 2018, poz. 1668 z późn. zm. Wnioskuję o dopuszczenie rozprawy doktorskiej mgr inż. Agaty Sage do publicznej obrony. Ponadto, biorąc pod uwagę 1) istotny dorobek publikacyjny Autorki – wg bazy Scopus obejmuje on 9 publikacji cytowanych łącznie 81 razy (bez autocytowań), 2) nowatorski charakter i skuteczność zaproponowanej metody na tle osiągnięć opisanych w aktualnej literaturze przedmiotu (zastosowanie danych obrazowych obok danych akustycznych), 3) rzetelną i wszechstronną analizę uzyskanych wyników oraz 4) bardzo staranny naukowy styl tekstu rozprawy składam wniosek o jej wyróżnienie.

Łódź, 30 września 2024 r.