

Szczecin, 2.09.2024 r.

prof. dr hab. inż. Krzysztof Okarma
Wydział Elektryczny
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

**RECENZJA ROZPRAWY DOKTORSKIEJ
dla Senatu Politechniki Bydgoskiej im. Jana i Jędrzeja Śniadeckich**

*opracowana na podstawie pisma Przewodniczącego Rady Dyscypliny
Informatyka Techniczna i Telekomunikacja Politechniki Bydgoskiej im. Jana i Jędrzeja Śniadeckich
dr. hab. inż. Tomasza Talaśki, prof. PBŚ*

Tytuł rozprawy: **Badania algorytmów sztucznej inteligencji i ich odpowiednich modyfikacji w procesie modelowania komórek biologicznych oraz wykrywania wybranych chorób onkologicznych i kardiologicznych**

Autor rozprawy: **mgr inż. Krzysztof Pałczyński**

Dyscyplina naukowa: **informatyka techniczna i telekomunikacja**

Promotor: **dr hab. inż. Tomasz Talaśka, prof. PBŚ**

I. TEMATYKA, TEZA NAUKOWA I CEL ROZPRAWY

Rozprawę doktorską mgr. inż. Krzysztofa Pałczyńskiego stanowi zbiór 10 współautorskich publikacji wraz z autoreferatem, którego uzupełnienie stanowią załączone kopie oświadczeń współautorów prac oraz informacje o kodach źródłowych udostępnionych w repozytorium GitHub. Dotyczy ona zagadnień o charakterze silnie interdyscyplinarnym, związanych nie tylko z informatyką techniczną i telekomunikacją, ale także w dużej mierze z inżynierią biomedyczną, czy medycyną. Ze względu na dynamiczny rozwój sztucznej inteligencji, a w szczególności metod związanych ze sztucznymi sieciami neuronowymi, a także ich liczne zastosowania w różnych gałęziach nauki oraz techniki, wybór tematyki badań należy uznać za trafny, a tematykę przedłożonej do oceny rozprawy za aktualną.

Tematyka rozprawy skupia się wokół badań algorytmów sztucznej inteligencji używanych zarówno w procesie modelowania komórek biologicznych (pierwsza część zbioru obejmująca 5 publikacji stanowiących efekt realizacji interdyscyplinarnego grantu NCN, w którym Doktorant odpowiadał za jego część informatyczną), jak również wykrywania wybranych chorób onkologicznych oraz kardiologicznych (druga część zbioru, również obejmująca 5 współautorskich pozycji).

Za główny cel rozprawy Doktorat postawił sobie opracowanie modyfikacji algorytmów sztucznej inteligencji, które umożliwiłyby modelowanie komórek biologicznych, a także były użytecznym narzędziem w diagnozie niektórych chorób kardiologicznych oraz onkologicznych. Teza naukowa rozprawy, wiążąca się z tak postawionym celem, została podzielona na pięć tez składowych dotyczących:

- potwierdzenia celowości stosowania teorii kolejek do modelowania szlaków metabolicznych jako alternatywy dla równań różniczkowych,
- zweryfikowania możliwości zastosowania teorii kolejek oraz metod sztucznej inteligencji wraz ze wstępnym przetwarzaniem wektorów parametryzujących do symulacji komputerowej komórek biologicznych,
- potwierdzenia możliwości poprawy dokładności procesu rozpoznawania wybranych chorób onkologicznych i kardiologicznych dzięki zastosowaniu wstępnego przetwarzania wektorów uczących w procesie klasyfikacji,
- weryfikacji skuteczności użycia modeli wykorzystujących mniejszą liczbę parametrów bez utraty dokładności wykonywanego zadania w rozważanych zastosowaniach.
- poprawy procesu klasyfikacji dzięki wykorzystaniu hybrydowych metod uczenia maszynowego w analizowanych zastosowaniach.

Zarówno cel rozprawy, jak również postawione tezy, uznać można za właściwe i w pełni zgodne z zakresem pracy. Pewne zastrzeżenie może dotyczyć dość ogólnego sformułowania dwóch ostatnich tez, gdzie Doktorant nie sprecyzował, iż chodzi o rozważane w rozprawie zastosowania. W przypadku ogólnym udowodnienie czwartej tezy mogłoby być kłopotliwe, gdyż w wybranych zastosowaniach ograniczenie liczby parametrów zapewne prowadziłoby do pogorszenia dokładności klasyfikacji. Z kolei ostatnia teza wydaje się w przypadku ogólnym dość oczywista, choć w konkretnym zastosowaniu jej udowodnienie można uznać za celowe.

Warto zauważyć, iż wszystkie prace stanowiące zbiór opublikowanych i powiązanych tematycznie artykułów naukowych przedstawiony do oceny, zostały wydane w latach 2020-2023, co dobrze świadczy o aktualności przedstawionych w nich wyników badań.

II. ZAWARTOŚĆ MERYTORYCZNA ROZPRAWY

Rozprawa doktorska mgr. inż. Krzysztofa Pałczyńskiego pt. *„Badania algorytmów sztucznej inteligencji i ich odpowiednich modyfikacji w procesie modelowania komórek biologicznych oraz wykrywania wybranych chorób onkologicznych i kardiologicznych”*, stanowiąca zbiór 10 opublikowanych i powiązanych tematycznie artykułów naukowych, charakteryzuje się silną interdyscyplinarnością, co w znacznej mierze wynika z trendów rozwojowych metod sztucznej inteligencji oraz ich zastosowań.

Zbiór ten został podzielony na dwie części, które są jednakże ze sobą powiązane tematycznie (zgodnie z wymogami Ustawy *Prawo o szkolnictwie wyższym i nauce*). Ze względu na fakt, iż są to

wyłącznie publikacje współautorskie, pewną trudność stanowi w tym przypadku wyodrębnienie wkładu merytorycznego Doktoranta, związanego ściśle z dyscypliną naukową *informatyka techniczna i telekomunikacja*. Docenić należy fakt publikowania wyników badań w wysoko punktowanych czasopismach naukowych, a także umiejętność współpracy w zespołach badawczych, co jest coraz częściej nieodzownym elementem działalności naukowej. Jednakże, liczba autorów każdej z publikacji w pierwszej części zbioru (P1-P5) to średnio 8 osób (od 7 do 9), gdzie Doktorant występuje zawsze na drugiej pozycji, nie będąc nigdy autorem korespondencyjnym. Szkoda, że przedstawione oświadczenie dotyczące udziału Doktoranta w poszczególnych publikacjach zawiera je w kolejności odwrotnej aniżeli w przedstawionym do oceny zbiorze (a nawet z pewnymi odstępstwami od tej reguły). Niestety zawiera ono pewne niespójności względem zapisów w poszczególnych publikacjach np. w pozycji [P3] w sekcji *CRedit authorship contribution statement* znaleźć można zapis: „Krzysztof Pałczyński: Methodology, Software, Validation, Visualization”, podczas gdy w oświadczeniu znajduje się także wzmianka o uczestnictwie w pisaniu publikacji (w publikacji taka wzmianka tj. „Writing – review & editing” znajduje się z kolei przy wszystkich pozostałych autorach). Nie jest to jednak element kluczowy, ponieważ wkład merytoryczny Doktoranta został dobrze opisany, a problem ten nie występuje przy innych publikacjach w tym samym zespole ([P2],[P5]) – można to zatem uznać nawet za zwykłą pomyłkę. Wyjaśnienia wymaga jednakże brak podpisów w oświadczeniach autorów zagranicznych (Marissa Miller – być może jest tam znak graficzny podpisu elektronicznego, Paul Davis – brak widocznego podpisu). Jak można się domyślić z oświadczenia Marissa Nitz oraz Marissa Miller to ta sama osoba, a zatem wszystkie pozostałe oświadczenia uznaję za kompletne, opisujące w wystarczająco szczegółowy sposób wkład poszczególnych Autorów w powstanie poszczególnych publikacji, z niewielkim wyjątkiem wskazanym w dalszej części recenzji.

W drugiej części zbioru współautorami publikacji są wyłącznie inne osoby aniżeli promotor oraz promotor pomocniczy, co jest nieco zaskakujące, jednak w trzech przypadkach Doktorant jest pierwszym autorem a w dwóch pozostałych występuje na drugiej pozycji. W tej części zbioru publikacji występuje zauważalnie mniejsza liczba autorów każdej z nich (od 3 do 6), co wpłynęło na większy udział merytoryczny Doktoranta w ich powstawaniu, jak wynika także z oświadczeń współautorów oraz zapisów w sekcji *Author Contributions* każdego z artykułów. Wszystkie te publikacje ukazały się w dwóch czasopismach wydawnictwa MDPI: *Sensors* oraz *Entropy*, co z jednej strony jest zrozumiałe, biorąc pod uwagę długość cyklu wydawniczego, ale z drugiej strony budzi pewien niedosyt z punktu widzenia „dywersyfikacji” miejsc publikowania dorobku, skompensowany jednak w pewnym stopniu pierwszą częścią zbioru publikacji.

W przedstawionej dokumentacji (zarówno w wersji elektronicznej, jak też papierowej) brak jest niestety pełnego tekstu pozycji [P9], zapewne przez pomyłkę, gdyż dwukrotnie zamieszczony został tekst publikacji [P8]. Nie najlepiej niestety świadczy to – także w zestawieniu z dość licznymi uwagami językowymi i technicznymi dotyczącymi autoreferatu, czy wspomnianymi już uwagami dotyczącymi oświadczeń współautorów – o sposobie przygotowania dokumentacji przez Kandydata. Jedną z uwag tego rodzaju dotyczyć może także tytułu rozprawy, gdyż wątpliwości budzi nie tylko jego długość, ale przede wszystkim sformułowanie „odpowiednich modyfikacji” (nie jest jasne, co Autor rozumie przez tę „odpowiedniość”).

Wspomniana publikacja [P9] jest na szczęście dostępna w formule Open Access, a zatem jej brak w dostarczonej dokumentacji nie wpływa w żaden sposób na ocenę rozprawy. Wątpliwości budzi

jednakże przedstawiony w publikacjach [P8] oraz [P9] wkład poszczególnych współautorów, który jest identyczny, co nie jest w pełni spójne z załączonymi oświadczeniami. Ich analizę dodatkowo utrudnia fakt, iż w niektórych przypadkach udział poszczególnych współautorów jest (słusznie zresztą) doprecyzowany w oświadczeniach, jednak zdarzają się przypadki (np. przy wspomnianej już publikacji [P9]), dla której sformułowanie „walidacja badań i analiza wyników” pojawia się w oświadczeniach każdego ze współautorów.

Szkoda, iż w przedstawionym zbiorze publikacji nie pojawiła się żadna praca samodzielna, ani też żaden artykuł opublikowany wyłącznie z promotorem i ewentualnie z promotorem pomocniczym. Obniża to niewątpliwie ocenę osiągnięcia, utrudniając dodatkowo precyzyjne określenie wkładu Kandydata w powstanie poszczególnych publikacji związanego z rozwiązaniem problemu naukowego w dyscyplinie *informatyka techniczna i telekomunikacja*.

Odnosząc się do poszczególnych pozycji z pierwszej części zbioru publikacji opisanych w autoreferacie, powstałych w wyniku realizacji projektu NCN OPUS pt. „*Telekomunikacyjne podejście do modelowania metabolizmu komórek*” nr UMO-2019/33/B/ST6/00875 kierowanego przez prof. dr. hab. inż. Tadeusza Wysockiego, warto podkreślić, iż Doktorant odpowiedzialny był w nim – jak sam zauważa w autoreferacie – za część informatyczną projektu i odpowiednie przygotowanie modeli z wykorzystaniem zmodyfikowanych algorytmów sztucznej inteligencji, po wcześniejszym odpowiednim przetworzeniu wzorców uczących. Pozwoliło to na opracowanie modeli charakteryzujących się potencjalną przydatnością we wczesnej diagnostyce różnego rodzaju schorzeń, wśród których Doktorant wskazał cukrzycę, otyłość, czy niektóre nowotwory. Stanowi to niewątpliwie oryginalne osiągnięcie Kandydata związane z dyscypliną *informatyka techniczna i telekomunikacja*, biorąc w szczególności pod uwagę fakt zaproponowania zastosowania teorii kolejek przy tworzeniu modeli. Ich zastosowanie wyjaśnia fakt, iż przedstawione podejście do modelowania metabolizmu komórek wywodzi się z telekomunikacji a nie ma charakteru stricte informatycznego, co w świetle połączenia informatyki technicznej i telekomunikacji w jedną dyscyplinę nie ma formalnego znaczenia, jednak bardzo dobrze wpisuje się w związane z tym trendy rozwojowe tejże dyscypliny.

We wstępnej części autoreferatu Doktorant przedstawił motywację badań z zakresu sztucznej inteligencji w zastosowaniach medycznych, wskazując potrzebę wspomaganie pracy personelu medycznego przez algorytmy sztucznej inteligencji. Zauważył także, iż można je wykorzystać do modelowania i symulacji procesów biochemicznych, które mają miejsce w organizmach. Za istotny element przemawiający za użyciem sieci neuronowych uznał możliwość pracy na danych nieustrukturyzowanych, zauważając także potencjał tkwiący we wstępnym przetwarzaniu danych, co prowadzić powinno do ograniczenia liczby parametrów przy zachowaniu dokładności klasyfikacji.

Następnie Doktorant przedstawił cel pracy i hipotezy badawcze, a także syntetyczne informacje dotyczące najważniejszych osiągnięć w poszczególnych artykułach stanowiących zbiór publikacji powiązanych tematycznie przedstawionych do oceny. Wskazał także, iż we wszystkich wykazanych pracach był odpowiedzialny za opracowanie i przygotowanie algorytmów sztucznej inteligencji oraz ich modyfikacji, ja również wstępne przetwarzanie danych uczących, implementację modeli i algorytmów, a także testowanie i badania uzyskanych rozwiązań. Pomimo wskazanych wcześniej

uwag natury formalnej, uznaję, iż wkład ten został w wystarczający sposób udokumentowany oraz potwierdzony w oświadczeniach współautorów i notkach zawartych w poszczególnych publikacjach.

W publikacji [P1] dokonano zamodelowania szlaku metabolicznego cyklu Krebsa z użyciem sieci neuronowej. Jest to proces, dzięki monitorowaniu którego możliwa jest ocena skuteczności działania wybranych leków przeciwnowotworowych. W Autoreferacie Doktorant przedstawił bardzo szczegółowy opis tego cyklu od strony biochemicznej, co jednakże nie jest ściśle związane z dyscypliną naukową *informatyka techniczna i telekomunikacja*. Tym niemniej, szczegółowe opisy zamieszczone w autoreferacie mogą być pomocne w zrozumieniu stopnia złożoności zagadnienia i specyfiki procesów modelowanych za pomocą algorytmów sztucznej inteligencji. Podobnie szczegółowy opis dotyczy pozostałych osiągnięć wchodzących w skład pierwszej części zbioru publikacji dotyczących zamodelowania innych procesów w oparciu o teorię kolejek oraz równania Michaelisa-Menten. W publikacji [P2] przedstawiono model szlaku pentozofosforanowego (PPP), istotnego w diagnostyce wybranych nowotworów, zamodelowanego właśnie w ten sposób. Uzyskane efekty symulacji zostały porównane z wynikami badań empirycznych, co pozwoliło na stwierdzenie poprawności uzyskanego modelu.

Z kolei publikacja [P3] dotyczy modelowania procesu beta-oksydacji kwasów tłuszczowych w celu odzwierciedlenia wartości stężeń metabolitów zawierających łańcuchy węglowe o różnej liczbie atomów węgla. W tym przypadku konieczna okazała się zmiana funkcji strat w zastosowanym algorytmie genetycznym zastosowanym w celu odwzorowania stężeń metabolitów w zakładanych punktach pomiarowych określonych jako początek i koniec badania oraz po upływie 1/3 czasu symulacji. Połączenie tego modelu z modelami cyklu Krebsa oraz szlaku pentozofosforanowego zostało przedstawione w publikacji [P5], prowadząc do opracowania pełniejszego modelu komórki biologicznej. Praca [P4] skupia się z kolei na modelu odpowiedzi komórkowej na insulinę, co jest istotne w diagnostyce i leczeniu cukrzycy. Wytrenowany model odwzorowuje szlak sygnalizacyjny tej odpowiedzi, co zostało również szczegółowo przedstawione w autoreferacie.

Publikacje [P1]-[P4] oparte są na idei Doktoranta związanej z użyciem teorii kolejek, z kolei publikacja [P5] ma charakter podsumowujący i łączący poprzednie osiągnięcia w jeden spójny model.

W kolejnej części autoreferatu Kandydat skupił się na opisie osiągnięć zawartych w drugiej części zbioru publikacji ([P6]-[P10]), dotyczących zastosowania algorytmów sztucznej inteligencji do wspomagania wykrywania wybranych chorób onkologicznych i kardiologicznych.

W publikacji [P6] przedstawiono możliwość użycia wybranych sieci neuronowych do klasyfikacji zdjęć mikroskopowych limfocytów w celu diagnostyki ostrej białaczki limfoblastycznej (ALL) na podstawie obrazów z bazy danych ALL-IDB2 zawierającej 260 zdjęć. Wstępnej klasyfikacji dokonano przy użyciu prostej homogenicznej sieci spłotowej (CNN) oraz sieci MobileNet V2, w tym wstępnie wytrenowanej na zbiorze ImageNet. Przeanalizowano także możliwość zastosowania wstępnego przetwarzania danych wejściowych w celu redukcji wymiarowości sieci, a także augmentacji danych. Dodatkowo w publikacji [P7] zastosowano algorytm XGBoost bazujący na drzewach decyzyjnych, a także przeanalizowano możliwość użycia modelu HSV zamiast RGB do odwzorowania kolorów. Dzięki wstępnemu przetwarzaniu danych z zastosowaniem histogramów koloru uzyskano poprawę dokładności klasyfikacji w porównaniu z przetwarzaniem danych „surowych”. Wskazano także, iż istotne informacje umożliwiające wykrywanie ALL znajdują się w otoczeniu limfocytów.

Kolejne publikacje opisane w autoreferacie dotyczą wykrywania wybranych chorób serca na podstawie analizy elektrokardiogramów uzyskanych z publicznie dostępnej bazy PTB-XL, zawierającej dane zebrane od prawie 19 tysięcy pacjentów z etykietami opisującymi 4 lub 19 chorób, co umożliwia nie tylko klasyfikację binarną, ale także 5-klasową oraz 20-klasową, której wyniki przedstawiono w pracach [P8, P9, P10]. W pierwszej z nich zastosowano sieć splotową (CNN), następnie zaproponowano użycie dodatkowych cech bazujących na entropii, a także modyfikacje sieci oraz funkcji strat, co pozwoliło zauważalnie poprawić uzyskiwane wyniki klasyfikacji.

Autoreferat kończy się podsumowaniem i wnioskami, dołączono do niego także bibliografię, zestaw oświadczeń współautorów oraz kopie 9 z 10 publikacji przedstawionych do oceny.

III. OGÓLNA OCENA ROZPRAWY i UWAGI DYSKUSYJNE

Porównując obie części zbioru publikacji przedstawionych do oceny, głównych osiągnięć Doktoranta z punktu widzenia dyscypliny naukowej *informatyka techniczna i telekomunikacja* upatrywałbym w jego drugiej części. Z jednej strony ma to związek z mniejszą liczbą współautorów tych prac, z drugiej strony publikacje [P1]-[P5] mają silniejsze zabarwienie biomedyczne, nieco mniej skupiając się na istocie modyfikacji algorytmów sztucznej inteligencji, które są zasadniczym osiągnięciem Kandydata. Wszystkie te osiągnięcia mają charakter eksperymentalny, gdyż podbudowa teoretyczna przedstawiona w pracach dotyczy przede wszystkim aspektów biomedycznych, czy też nawet biochemicznych, co stanowi wprawdzie podbudowę do opracowywania modeli procesów czy też sieci neuronowych, jednak nie jest ściśle związane z dyscypliną naukową, w której Kandydat ubiega się o stopień doktora nauk inżynierjno-technicznych.

Biorąc pod uwagę fakt, iż wszystkie publikacje ukazały się w recenzowanych czasopismach zagranicznych posiadających współczynnik Impact Factor oraz wysoką punktację MNiSW, uwagi dyskusyjne przedstawione poniżej dotyczą przede wszystkim autoreferatu a nie samych tekstów publikacji, które były już zresztą poddane niezależnym recenzjom.

Ogólna ocena rozprawy pozostawia pewien niedosyt – z jednej strony związany z brakiem samodzielnych publikacji, jak również faktem publikowania znacznej części dorobku w zespołach innych aniżeli promotor i promotor pomocniczy, co jest raczej rzadkością; z drugiej strony dorobek jest bardzo interdyscyplinarny, co widać chociażby w nieco zbyt szczegółowych opisach aspektów biochemicznych i biomedycznych w porównaniu z mniejszą szczegółowością elementów informatycznych. Wydaje się, iż proporcje te zostały nieco zaburzone.

Z punktu widzenia wkładu w dyscyplinę naukową, niektóre z publikacji mają charakter dość „przyczynkarski”, skupiając się jedynie na wybranym aspekcie modyfikacji algorytmów sztucznej inteligencji w konkretnym zastosowaniu, niekiedy dla relatywnie małej bazy danych uczących oraz testowych. Jednakże, biorąc pod uwagę wymogi publikacyjne związane z koniecznością rozliczenia grantu NCN, a także współpracę z innymi naukowcami, osiągnięcia Kandydata powinny być w tym przypadku analizowane całościowo.

Po analizie prac oraz oświadczeń współautorów można wysnuć wniosek, iż udział Kandydata w ich powstaniu był istotny ze względu na odpowiedzialność za elementy związane z zastosowaniem

oraz modyfikacjami algorytmów sztucznej inteligencji. Tym niemniej wśród współautorów także znajdują się osoby, które tymi aspektami w niemałym stopniu się zajmowały.

Wśród oryginalnych osiągnięć Kandydata pozwalających wypełnić wymagania do uzyskania stopnia doktora nauk inżynieryjno-technicznych w dyscyplinie *informatyka techniczna i telekomunikacja* upatruję głównie:

- zastosowanie teorii kolejek do modelowania procesów i komórek biologicznych,
- zaproponowane i zweryfikowane modyfikacje algorytmów sztucznej inteligencji, w szczególności poprzez zastosowanie wstępnego przetwarzania danych oraz wektorów uczących, w celu poprawy klasyfikacji w zadaniach wykrywania wybranych chorób serca oraz onkologicznych,
- dobór funkcji strat w procesie trenowania sieci neuronowych pozwalających na poprawę dokładności klasyfikacji w zadaniach rozważanych w rozprawie,
- opracowanie i implementację hybrydowych metod uczenia maszynowego wspomagających proces diagnostyki wybranych schorzeń kardiologicznych oraz onkologicznych.

Pomimo, iż wniosek dotyczący wyższej precyzji uzyskiwanych wyników (choć być może właściwsze byłoby słowo „dokładności”) przy zastosowaniu wstępnego przetwarzania wektorów uczących przy jednoczesnej redukcji liczby parametrów w używanych modelach wydaje się oczywisty, istotnym osiągnięciem Doktoranta jest weryfikacja tej hipotezy w konkretnych zastosowaniach rozważanych w badaniach. Kandydat słusznie podkreślił, iż wstępne przetwarzanie danych uczących pomaga wyekstrahować pożądaną informację jednocześnie eliminując dane nieprzydatne, pozwalając ponadto na realizację hybrydowych modeli umożliwiających przetwarzanie różnego rodzaju danych, w tym łączących różne algorytmy sztucznej inteligencji.

Przekonującym elementem dorobku jest również poprawa dokładności klasyfikacji uzyskana dzięki wprowadzonym modyfikacjom algorytmów SI, np. wskazana w publikacji [P9] rzędu 5 punktów procentowych, czy też znaczna redukcja liczby parametrów modelu (kilkusetkrotna, jak wskazano w publikacji [P7]). Na pozytywną ocenę zasługuje także fakt weryfikacji opracowanych modeli pod kątem ich zgodności z danymi literaturowymi, np. dotyczących cyklu Krebsa [P1]. Dobrym pomysłem jest także zastosowanie dodatkowych cech bazujących na entropii przy klasyfikacji sygnałów EKG.

Lektura autoreferatu nasuwa jednakże pewne wątpliwości i skłania do sformułowania kilku pytań dotyczących poszczególnych osiągnięć, jak również uwag szczegółowych wymienionych w dalszej części recenzji. Z kolei uwagi natury formalnej, w tym dotyczące oświadczeń współautorów zostały sformułowane wcześniej.

W wyjaśnieniu spójności „monotematycznego cyklu publikacji”, czyli właściwie zbioru publikacji powiązanych tematycznie (zgodnie z aktualnymi zapisami w Ustawie *Prawo o szkolnictwie wyższym i nauce*, gdyż pojęcie „cyklu publikacji” dotyczy postępowań habilitacyjnych), zastanawiające jest stwierdzenie „odpowiednie modyfikacje”, pojawiające się zresztą już w tytule pracy. Nie jest jasne, co Autor rozumie przez tę „odpowiedniość” i w jaki sposób jest ona określana.

Nie jest również jasne, co Doktorant rozumie przez „uzyskanie wyższej o średnio 2.7 punktu procentowego jakości przetwarzania” (pomijając już użycie kropki dziesiętnej zamiast przecinka) na stronie 9. Czy ta „jakość przetwarzania” stanowi dokładność (*accuracy*) klasyfikacji czy jest określana w jakiś inny sposób?

Nie jest również jasne, co Autor rozumie przez pojęcie „ilości stężenia substancji” znajdujące się na str. 19.

Na str. 20-22 znajduje się dość krótki opis znanych algorytmów genetycznych, jednak celowe byłoby wyjaśnienie, dlaczego w tym przypadku Doktorant zastosował właśnie tego rodzaju metody zamiast klasycznych technik optymalizacyjnych, a także czy zostało przeprowadzone porównanie uzyskiwanych wyników różnymi metodami wraz z czasem obliczeń.

Zdanie zawarte na str. 22 „Założenie to bierze się z braku mechanizmu w algorytmie genetycznym przeprowadzającym analizę wpływu zmian wartości genów na wynik” wymaga wyjaśnienia.

Koncepcje zastosowania algorytmów sztucznej inteligencji przedstawione w pracach [P1]-[P4] są dość podobne – warto byłoby w syntetyczny sposób przedstawić podobieństwa i różnice pomiędzy zastosowanymi w tych pracach podejściami.

Jak stwierdzono na str. 27, zsumowanie wartości stężeń dwóch metabolitów (w pracy [P1]) zwiększyło stabilność modelu. Celowe byłoby wyjaśnienie, w jaki sposób ta stabilność została określona i w jakim stopniu została zwiększona.

Proszę o wyjaśnienie, co Autor miał na myśli, stwierdzając na str. 44, iż „przy tak dużej ilości danych zwiększona jest podatność algorytmów AI na przeuczenie”, gdyż nadmierne dopasowanie modelu do danych typowo ma miejsce przy dużej liczbie parametrów i mniejszej ilości danych.

Doktorant nie umotywił wystarczająco doboru metod augmentacji danych wskazanych na str. 45. Czy były prowadzone eksperymenty dotyczące zastosowania innych metod rozszerzających zbiór uczący? Co było powodem wyboru właśnie tego rodzaju operacji na obrazach?

Wyjaśnienia wymaga także motywacja wyboru histogramu wartości koloru jako sposobu kodowania informacji obrazowych (str. 47), który powoduje przecież całkowitą utratę informacji dotyczących kształtu obiektów widocznych na obrazie. Nie jest również jasne, z jakiego powodu wybrany został model HSV i czy były prowadzone badania eksperymentalne z użyciem innych modeli np. CIELAB, a także nawet konwersji do skali szarości (można byłoby się spodziewać w takim przypadku dość podobnych wyników do zastosowanie kanału zielonego).

Analizując obrazy przedstawione na rys. 15 oraz 16, wydaje się, iż użycie metod sztucznej inteligencji do klasyfikacji tego rodzaju obrazów jest przesadą. „Na pierwszy rzut oka” zupełnie wystarczające powinno być wyznaczenie wariancji lub entropii obrazu.

Nie jest także jasne, czy przy zaledwie 260 zdjęciach oraz 51 parametrach (skąd ta liczba?) nie występuje niebezpieczeństwo przetrenowania modelu, choć istotnie zmniejszenie rozmiaru modelu zmniejsza to ryzyko.

IV. UWAGI SZCZEGÓLWE

Przewodnik po publikacjach (autoreferat) jest zredagowany na ogół poprawnym i zrozumiałym językiem. Autor nie ustrzegł się jednak pomyłek, nieścisłości i błędów, spośród których niektóre zauważone w trakcie lektury wymieniono poniżej (także te o charakterze dyskusyjnym):

- używanie wielkich liter w sformułowaniu „Teoria Kolejek” nie jest właściwe – nie jest to nazwa własna,
- skrót ASI w odniesieniu do algorytmów sztucznej inteligencji nie jest powszechnie używany, znacznie częściej spotyka się skrót SI (lub anglojęzyczny AI) w odniesieniu do sztucznej inteligencji, a zatem bardziej naturalne byłoby określenie „algorytmy SI”, ale na pewno nie „algorytmy ASI”, jak ma to miejsce np. na stronie 7,
- wątpliwości budzi także stwierdzenie na str. 8: *„To z kolei może się przełożyć na opracowywanie lepszych leków i terapii, co może znacznie poprawić skuteczność leczenia i jakość życia pacjentów”* – tryb przypuszczający nie powinien być raczej stosowany w pracach naukowych,
- o ostatnim akapicie na str. 7 Doktorant zacytował wszystkie 10 prac z przedstawionego zbioru, stwierdzając, iż *„powstaje wiele artykułów naukowych o nowych sposobach...”* – w tym miejscu powinno być zacytowanych przynajmniej kilka publikacji innych autorów,
- formułowanie *„zaadresować fundamentalne różnice”* na stronie 9 stanowi kalkę językową, która nie powinna się pojawiać w tekstach naukowych, w kolejnym wierszu znajduje się z kolei sformułowanie *„w miejscu tym”* zamiast *„w tym miejscu”*,
- sformułowanie *„efekty tych prac zostały wyeksportowane do pięciu publicznie dostępnych repozytoriów”* nie jest właściwe w odniesieniu do cytowanych publikacji, gdyż na platformie GitHub znajdują się raczej kody źródłowe i dane; ponadto sformułowanie *„cykl badań składający się z dziesięciu artykułów”* również nie znajduje uzasadnienia, gdyż w artykułach przedstawione są wyniki przeprowadzonych badań,
- sposób przedstawienia niektórych równań biochemicznych z użyciem nazw substancji (np. jabłczan czy szczawiooctan) zamiast wzorów nie jest typowy, w pracy z dziedziny nauk inżyneryjno-technicznych spodziewać się należy raczej wzorów, choć w tym przypadku wymagałyby one zapewne dodatkowego wyjaśnienia,
- sformułowanie *„opartych o równanie Michaelisa-Menten”* na str. 18 jest błędem językowym, podobnie jak *„zmiany te opierają się o [...] przetwarzanie wektorów”* na str. 22,
- schemat PPP przedstawiony na rysunku 6 mógłby być przetłumaczony na język polski, jego jakość (niestety także w pliku PDF) jest zbyt niska, podobnie jak rysunku 8, co nieco dziwi, biorąc pod uwagę dyscyplinę, w której Doktorant ubiega się o uzyskanie stopnia naukowego, niektóre pozostałe rysunki również są zbyt silnie skompresowane,
- sformułowanie *„ilość początkowego stężenia”* na str. 37 nie jest precyzyjne językowo,
- sformułowanie *„liczba danych”* na str. 44 powinno być raczej zastąpione przez *„ilość danych”*, jak ma to miejsce w kolejnym wierszu,
- sformułowanie *„specjalnie utworzone techniki w swej strukturze”* (str. 44) w sieci MobileNet V2 nie jest najtrafniejsze,
- nagłówek tabeli 11 jest nieprawidłowy – użycie słowa *„przetrenowanie”* w odniesieniu do faktu douczenia sieci wstępnie wytrenowanej nie jest zasadne, gdyż przetrenowanie dotyczy

nadmiernego dopasowania do danych uczących (*overfitting, overtraining*), zatem nagłówek taki jest mylący,

- nie jest zbyt szczęśliwe ciągłe używanie sformułowania *data augmentation* w polskim tekście – w literaturze dotyczącej sztucznej inteligencji przyjęł się już polski odpowiednik: augmentacja danych, choć jest on w istocie kalką językową,
- szyk sformułowania „co zwiększa ich odporność na przetrenowanie i interpretowalność” nie jest również najtrafniejszy, gdyż nie jest jasne, czy chodzi o zwiększenie interpretowalności (jak ma to faktycznie miejsce) czy odporności na interpretowalność,
- podpisy rys. 14 są nieprecyzyjne – zamiast słów „odwrócenie” a tym bardziej „obrócenie”, lepiej byłoby użyć sformułowania „odbicie symetryczne” lub nawet po prostu „odbicie”.

Pomimo dość dużej liczby usterek w przedstawionym przewodniku po publikacjach, który z założenia prezentować powinien własne dokonania Kandydata, można uznać, iż w większości nie są to błędy rażące. Zauważone usterki i błędy nie utrudniają w znaczący sposób lektury i nie umniejszają ogólnie pozytywnej oceny rozprawy. Pomimo wskazanych wątpliwości i uwag, a także braku publikacji samodzielnych, można uznać, iż założone cele rozprawy zostały osiągnięte a tezy udowodnione.

V. WNIOSKI KOŃCOWE

Opiniowana rozprawa, a właściwie wkład Kandydata w przedstawione współautorskie publikacje, stanowi samodzielne rozwiązanie problemu naukowego mieszczącego się w zakresie dyscypliny naukowej *informatyka techniczna i telekomunikacja*, zawierając także pewne elementy interdyscyplinarne, na pograniczu inżynierii biomedycznej oraz medycyny. W pracy Autor wykazał się znajomością zagadnień dotyczących algorytmów sztucznej inteligencji, jak również wybranych zagadnień z zakresu inżynierii biomedycznej oraz medycyny. Uwidocznił także swoje umiejętności zastosowania i dostosowania algorytmów sztucznej inteligencji do rozwiązywania problemów z tego zakresu, potwierdzając umiejętność samodzielnego rozwiązywania problemów naukowych na pograniczu ww. dyscyplin naukowych.

Stwierdzam, iż przedstawiona do recenzji rozprawa doktorska **mgr. inż. Krzysztofa Pałczyńskiego** pt. ***Badania algorytmów sztucznej inteligencji i ich odpowiednich modyfikacji w procesie modelowania komórek biologicznych oraz wykrywania wybranych chorób onkologicznych i kardiologicznych***, której promotorem jest dr hab. inż. Tomasz Talaśka, prof. PBŚ, spełnia, choć zaledwie z nieznacznym nadmiarem, wymagania stawiane rozprawom doktorskim przez *Ustawę Prawo o szkolnictwie wyższym i nauce* z dnia 20 lipca 2018 roku (tekst jednolity Dz. U. 2023 poz. 742, z późn. zm.). **Wnioskuje o jej przyjęcie i dopuszczenie do publicznej obrony.**

