

Politechnika Poznańska
Wydział Informatyki i Telekomunikacji,
Instytut Sieci Teleinformatycznych

RECENZJA ROZPRAWY DOKTORSKIEJ

przygotowana dla Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja
Politechniki Bydgoskiej im. Jana i Jędrzeja Śniadeckich

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**

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

Promotor pomocniczy: dr inż. Tomasz Marciniak, prof. PBŚ

Dziedzina: nauki inżyniersko-techniczne

Dyscyplina: informatyka techniczna i telekomunikacja

- **Jakie zagadnienie naukowe jest rozpatrzone w pracy (teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez autora?**

Podstawowym celem, który postawił przed sobą Kandydat jest „*opracowanie modyfikacji algorytmów sztucznej inteligencji, które będą umożliwiły modelowanie komórek biologicznych oraz będą wspomagały proces wykrywania wybranych chorób onkologicznych i kardiologicznych*”. Doktorant uszczegółowił cel poprzez zdefiniowanie zbioru tez, która praca ma wykazać.

Tezy przedstawione przez Kandydata zakładają, że:

- *Teoria kolejek może stanowić skuteczną alternatywę dla równań różniczkowych stosowanych do modelowania szlaków metabolicznych.*
- *Wstępne przetwarzanie wektorów parametryzujących pozwala na wykorzystanie algorytmów sztucznej inteligencji do modelowania, w oparciu o teorię kolejek a następnie symulację komputerową, komórek biologicznych.*

- *Wstępne przetwarzanie wektorów uczących w procesie klasyfikacji pozwala na wydobycie z nich ważnych informacji poprawiających dokładność procesu rozpoznania wybranych chorób onkologicznych i kardiologicznych.*
- *Wstępne przetwarzanie wektorów uczących pozwala na użycie modeli wykorzystujących mniejszą liczbę parametrów bez utraty dokładności wykonywanego zadania.*
- *Wykorzystanie hybrydowych metod uczenia maszynowego pozwala na poprawę procesu klasyfikacji.*

Rozpatrując przedstawione przez Doktoranta tezy, w kontekście wcześniej przytoczonego celu pracy, uważam, że tezy rozprawy zostały przez Autora dostatecznie jasno sformułowane.

Wykazaniem tez pracy powierzono monotematycznemu cyklowi dziesięciu publikacji. Uzasadnienie celowości cyklu zamieszczone w rozprawie jest dla mnie przekonujące.

- **Jaka jest przydatność rozprawy z punktu widzenia nauk technicznych, czy założenia przyjęte przez autora są uzasadnione?**

Badania związane z zastosowaniem metod sztucznej inteligencji w medycynie prowadzone są już od wielu lat w szerokim zakresie. Opracowanie odpowiednich procesów i mechanizmów automatyzacji analiz danych medycznych, które zapewnią szybką i skuteczną diagnostykę, jest ważnym nurtem współczesnych badań medycznych. W tym kontekście wykorzystanie metod sztucznej inteligencji do zwiększenia efektywności diagnostyki bardzo dobrze wpisuje się również w wyzwania stawiane przed badaczami w dziedzinie nauk technicznych, których celem jest zwiększenie zakresu aplikacyjności, opracowanie nowych, lub modyfikacja istniejących algorytmów sztucznej inteligencji.

Na uwagę zasługuje również wykorzystanie metod sztucznej inteligencji do modelowania komórek biologicznych jako metody alternatywnej dla równań różniczkowych wykorzystywanych w szlakach metabolicznych. W tym obszarze można mówić również o ogromnym potencjale zastosowań nie tylko diagnostycznych.

Można zatem stwierdzić, że tematyka badań podejmowanych w rozprawie bardzo dobrze wpisuje w międzynarodowy nurt badań teoretycznych i aplikacyjnych.

- **Czy autor rozwiązał postawione zagadnienie i czy użył właściwej do tego metody?**

Głównym celem badań prezentowanych w rozprawie było opracowanie odpowiednich modyfikacji algorytmów sztucznej inteligencji, które będą umożliwiały modelowanie komórek biologicznych oraz wspomagały proces wykrywania wybranych chorób onkologicznych i kardiologicznych. Wyniki potwierdzające realizację tego celu Autor przedstawił w monotematycznym zbiorze publikacji naukowych.

Opierając się na tekście pracy i na publikacjach Doktoranta, które stanowiących integralną część rozprawy, przedstawię krótką dyskusję realizacji przez Autora celu pracy.

- Pierwsza teza pracy zakłada, że w procesie modelowania cykli metabolicznych możliwe jest zastąpienie równań różniczkowych teorią kolejek. Tezę tę Doktorant wykazał zakładając, że równania Michaelisa-Menten można interpretować jako prawdopodobieństwo przemiany substratów w produkty. Takie założenie pozwoliło na zastosowanie teorii kolejek do modelowania cyklu Krebsa, szlaku pentozofosforanowego, beta-oksydacji kwasów tłuszczowych i odpowiedzi komórkowej na insulinę. Zastąpienie równań różniczkowych teorią kolejek pozwoliło na rozwiązanie m.in. problemu ujemnych wartości stężenia substancji, które występują w analizie opartej na równaniach, a nie mają miejsca w rzeczywistości. Problem ten nie występuje w podejściu wykorzystującym teorię kolejek, a więc w tym przypadku można uzyskać stabilniejsze rezultaty symulacji. Można zatem przyjąć, że teza została udowodniona.
- Druga teza pracy zakłada, że wstępne przetwarzanie wektorów parametryzujących symulację pozwala na wykorzystanie sztucznej inteligencji do modelowania, w oparciu o teorię kolejek a następnie symulację komputerową, komórek biologicznych. Tezę tę Kandydat wykazał doświadczalnie dokonując grupowania genów w chromosomie ze względu na ich przynależność do parametryzowanego równania Michaelisa-Menten. W ten sposób zmodyfikowano pierwotny algorytm genetyczny, który nie przewidywał takiego mechanizmu. Powstał zatem nowy mechanizm oceniania prawdopodobieństwa zajścia reakcji chemicznych bez symulowania eksperymentu. Pozwoliło to na przyspieszenie poszukiwania wartości wykorzystywanych w procesie optymalizacji symulacji oraz uproszczono funkcję straty, którą służyła do oceny chromosomu. Jak wskazuje Doktorant korzystając z tej modyfikacji możliwe było przeniesienie z funkcji straty do mechanizmu reprodukcji wymogu uzyskania określonej wariancji w szeregach czasowych obrazujących stężenia substancji. Dzięki temu funkcja straty określa wynik dopasowania chromosomu korzystając z mniejszej liczby warunków. To upraszcza ją czyniąc funkcję straty łatwiejszą do minimalizacji jej wyniku. Można więc przyjąć, że druga teza została udowodniona.
- Trzecia teza pracy zakłada, że wstępne przetwarzanie wektorów uczących w procesie klasyfikacji pozwala na wydobycie z nich ważnych informacji poprawiających dokładność procesu rozpoznania wybranych chorób onkologicznych i kardiologicznych. Doktorant udowodnił ją oddzielnie w pracach prezentujących badania dotyczące ostrej białaczki limfoblastycznej oraz wykorzystanie sygnałów pochodzących z elektrokardiografu (EKG) do wykrywania i klasyfikacji wybranych chorób kardiologicznych.

Zgodnie z danymi przytoczonymi przez Kandydata system wspomagający proces wykrywania ostrej białaczki limfoblastycznej został poprawiony dzięki prawidłowo zrealizowanemu wstępnemu przetwarzaniu wektorów uczących. Doktorant wskazał, że uzyskano w ten sposób model, który był o 1.2 punktu procentowego dokładniejszy od sieci neuronowej przetwarzającej korzystających z tzw. surowych danych. Ponadto, model ten miał 500 razy mniej parametrów i podejmował decyzje na podstawie tensora danych wejściowych ponad 3000 razy mniejszego. Zatem uzyskano w ten sposób model dokładniejszy, szybszy i prostszy, co potwierdza trzecią tezę pracy w zakresie chorób onkologicznych.

Przytoczone przez Doktoranta dane dotyczące procesu wykrywania wybranych chorób kardiologicznych pokazują, że może być on bardziej dokładny dzięki zastosowaniu wstępnego przetwarzania wektorów uczących, co potwierdza trzecią tezę pracy w zakresie chorób kardiologicznych.

- Czwarta teza pracy zakłada, że wstępne przetworzenia wektorów uczących pozwala na użycie modeli wykorzystujących mniejszą liczbę parametrów bez utraty dokładności wykonywanego zadania. Swoimi badaniami Doktorant wykazał, że w przypadku ostrej białaczki limfoblastycznej dzięki wstępnemu przetworzeniu wektorów uczących, uzyskano model nieco dokładniejszy od wzorcowej sieci neuronowej. Model ten charakteryzował się mniejszą liczbą wymaganych parametrów i podejmował decyzję na podstawie znacznie mniejszego tensora, co potwierdza czwartą tezę pracy.
- Piąta teza pracy zakłada, że wykorzystanie hybrydowych metod uczenia maszynowego pozwala na poprawę procesu klasyfikacji. Kandydat wykazał w swoich pracach, że opracował i zaimplementował hybrydowe sieci neuronowe łączące możliwości interpretacyjne różnych algorytmów sztucznej inteligencji oraz hybrydowych sieci neuronowych, które łączą przetwarzanie różnych typów danych (sieci wielomodalne). Uzyskane wyniki pozwalają na dokładniejszą diagnozę. Doktorant stwierdził, że najlepsze modele do wykrywania wybranych chorób kardiologicznych dla klasyfikacji 2 klas, 5 klas i 20 klas uzyskały dokładność lepszą w porównaniu od tzw. surowego sygnału odpowiednio o 0,7, 3,7 i 4,9 procenta. Udowodniono w ten sposób piątą tezę pracy.

Przyjęte przez Autora metody wpisują się w szeroki nurt metod wykorzystywanych w zagadnieniach diagnostyki chorób metabolicznych, onkologicznych i kardiologicznych. Potwierdzona praktyczność zaproponowanych metod, technik i algorytmów oraz numeryczna ocena ich jakości zawarta w rozprawie, pozwalają na stwierdzenie, że Doktorant rozwiązał postawione zagadnienie i użył do jego rozwiązania właściwej metody.

- **Jaki charakter ma rozprawa (teoretyczny, doświadczalny, konstrukcyjny), jaka jest jej pozycja w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?**

Dominującym charakterem rozprawy jest charakter doświadczalny.

Do rozwiązania zdefiniowanych problemów i zadań badawczych Kandydat wykorzystał metody sztucznej inteligencji, sieci neuronowych oraz metody symulacyjne dobrane odpowiednio do rodzaju problemu i zaproponowanych metod jego rozwiązania. Sposób rozwiązania problemów zaproponowany przez Doktoranta nie odbiega od stanu wiedzy czy poziomu techniki, reprezentowanych przez literaturę światową. Praca dobrze wpisuje się w nurt badań poświęconych wykorzystaniu metod sztucznej inteligencji w medycynie.

- **Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora?**

Za najbardziej istotne i oryginalne wyniki Autora uważam:

- opracowanie nowatorskiej odmiany algorytmu genetycznego wykorzystującego wstępne przetwarzanie wektorów parametryzujących;
- opracowanie modelu pozwalającego na analizowanie procesów biochemicznych w komórce, który może zostać wykorzystany do projektowania leków oraz m.in. w leczeniu nowotworów, otyłości i cukrzycy;
- spostrzeżenie, że otoczenie limfocytów zawiera informacje umożliwiające diagnostykę ostrej białaczki limfoblastycznej;
- opracowanie dwumodalnej sieci neuronowej interpretującej tzw. surowy sygnał elektrokardiografu oraz miary entropii uzyskane z sygnału z elektrokardiografu;
- opracowanie hybrydowej sieci neuronowej łączącej konwolucyjne sieci neuronowe (CCN) z algorytmami uczenia maszynowego wykorzystywanych do analizy sygnału elektrokardiografu.

Analizując tekst rozprawy oraz oświadczenia autora i współautorów publikacji stanowiących oceniany cykl publikacji, można dojść do wniosku, że prezentowane w rozprawie wyniki badań stanowią samodzielny i oryginalny dorobek Autora.

- **Jakie są słabe strony rozprawy i jej główne wady?**

Doktorant od kilku lat z powodzeniem zajmuje się podejmowaną problematyką. Wiele wyników przedstawionych w rozprawie zostało również opublikowanych w publikacjach, w których podlegały analizie i ocenie.

Podczas lektury rozprawy pojawiały się jednak pewne pytania i wątpliwości:

- Czy, a jeśli tak, to na ile, sztuczne zwiększenie zbioru „surowych” danych treningowych dla sieci neuronowych wpłynęło na wyniki uzyskiwane z zastosowaniem wstępnego przetwarzania wektorów uczących?
- Czy, a jeśli tak, to na ile, jakość danych wejściowych uzyskiwanych z różnych źródeł i za pomocą różnorodnych przyrządów pomiarowych, mogła wpłynąć na spójność i jakość wyników. Czy mogło to wpłynąć na ocenę jakości proponowanych metod - czasami różnica była na poziomie pojedynczych procentów?

Nasunęły mi się również pewne plemiczne wątpliwości i pytania dotyczące rozszerzonego streszczenia badań stanowiących podstawę pracy.

Analizując wewnętrzną strukturę streszczenia można zauważyć pewną dysproporcję pomiędzy obszernością opisu poświęconego pierwszym pięciu publikacjom cyklu (30 stron) i następnym publikacjom (11 stron). Z czego wynika ta asymetria? Pewien niedosyt budzi również brak systematycznego opisu algorytmów lub metod (np. sposobu wykorzystania modyfikacji algorytmu genetycznego w różnych badaniach) z uwypukleniem różnic pomiędzy nimi w kolejnych zastosowaniach (np. w postaci pseudokodu). Czy Doktorant rozważał taki podejści, a jeśli tak, to dlaczego Kandydat ostatecznie z niego zrezygnował?

W rozprawie nie znalazłem natomiast błędów merytorycznych, a zastrzeżenia lub wątpliwości, o których pisałem powyżej dotyczą raczej strony redakcyjnej pracy, lub mają charakter polemiczny i nie mogą mieć wpływu na ostateczną pozytywną ocenę pracy. Uważam, że recenzowana rozprawa zawiera wiele oryginalnych wyników i wnosi wartościowy wkład w rozwój informatyki technicznej i telekomunikacji.

- **Czy rozprawa świadczy o dostatecznej wiedzy autora i znajomości współczesnej literatury z zakresu dyscypliny naukowej, jakiej rozprawa dotyczy?**

Literatura przytoczona w pracy (spis publikacji zawiera 87 pozycji) świadczy o wiedzy i dobrej orientacji Autora w dziedzinie, którą uprawia. Zamieszczone pozycje z ostatnich lat (ponad 25% cytowanych prac zostało opublikowane po 2020 roku) potwierdzają, że Kandydat nie zajmuje się tematyką wyczerpaną,

lecz przeciwnie, jest ona aktualna i inspirująca badawczo. O kompetencji Kandydata świadczą również współautorskie publikacje zawarte w monotematycznym cyklu 10 prac opublikowanych w czasopiśmie. Prace te zostały opublikowane w następujących międzynarodowych czasopiśmie: *Bioinformatics* (lista MNiSW 200 pkt.), *Scientific Reports* (lista MNiSW 140 pkt.), *Computational Biology and Chemistry* (lista MNiSW 70 pkt.), *PLoS One* (lista MNiSW 100 pkt.), *Sensors* i *Entropy* (lista MNiSW 100 pkt.).

- **Czy autor wskazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięźłość, jasność, poprawność redakcyjna rozprawy)?**

Przygotowanie pracy która opiera się na monotematycznym cyklu publikacji może rodzić pewne pytania i nasuwać wątpliwości, w szczególności mogą one dotyczyć struktury samej pracy. Jest to zapewne obszar polemiczny.

Dlatego chciałbym w tym miejscu wyrazić swoją opinię dotyczącą struktury samej pracy. W moim odczuciu w celu zwiększenia czytelności pracy i jej uspołnienienia można byłoby rozważyć nieco inny układ umieszczonych w niej treści. Proponowałbym rozpocząć od streszczenia, następnie wstęp, później cel i hipoteza, i kolejno uzasadnienie cyklu i sam cykl. Dopiero po tych elementach mogłyby się znaleźć rozszerzone streszczenie badań zawartych w cyklu publikacji, a następnie - w załącznikach - kopie artykułów stanowiących cykl i oświadczenia autora i współautorów.

Niezależnie od struktury samej pracy, warto byłoby w rozszerzonym streszczeniu zadbać o wprowadzenie wykazu skrótów. Pewnien niedosyt budzi także jakość rysunków i ich anglojęzyczne opisy. W moim odczuciu spójność i czytelność tej części pracy porawiłaby także numeracja rysunków i tabel nawiązująca do struktury rozdziałów (np. Rys. 5.1 zamiast Rys. 13).

Niezależnie od przytoczonych uwag, uważam, że opisane wyniki, jak i rozwiązania, zostały przedstawione w rozprawie w sposób akceptowalny.

- **Czy i jaka jest przydatność rozprawy dla gospodarki narodowej?**

Praca dotyczy możliwości wykorzystania algorytmów i metod sztucznej inteligencji w medycynie. Jest to obecnie dynamicznie rozwijający się obszar badań o ogromnym znaczeniu społecznym i ekonomicznym. Badania przedstawione w pracy mogą mieć zastosowanie w procesie opracowania nowych leków oraz w prowadzeniu terapii leczenia wielu chorób.

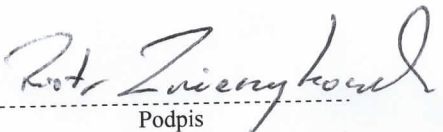
Można zatem stwierdzić, że rozwój innowacyjnych algorytmów genetycznych i metod sztucznej inteligencji, w tym rozwiązania opisane w pracy, z pewnością wpisują się w nurt, który może przyczynić

się do rozwoju i poprawy efektywności diagnostyki wielu chorób oraz do opracowywania nowych leków.

- **Czy rozprawa spełnia wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy?**

Biorąc pod uwagę wnioski zaprezentowane w poprzednich punktach i wymagania zawarte w Ustawie z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2023 r. poz. 742, z późn. zm.), uważam, że **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*” **zawiera oryginalne rozwiązania problemu naukowego** oraz dowodzi, że **Kandydat posiada ogólną wiedzę teoretyczną** w dyscyplinie informatyka techniczna i telekomunikacja i **posiada umiejętność samodzielnego prowadzenia pracy naukowej.**

Wnoszę o dopuszczenie rozprawy doktorskiej Pana 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” do publicznej obrony.



Podpis