

*Recenzja przyjeta
uzna'zenie*

Przewodnicząca
Rady Dyscypliny Nauki Medyczne
Katarzyna Sierakowska
dr hab. Katarzyna Sierakowska, prof. UMK

dr hab. inż. Robert Burduk
Politechnika Wroclawska
Wydział Informatyki i Telekomunikacji
Ul. Wybrzeże Stanisława Wyspiańskiego 27
50-370 Wrocław

Wrocław, dnia 27.07.2023 r.

RECENZJA

rozprawy doktorskiej mgr Anny Kloski

zatytułowanej: „**Use of Artificial Intelligence Method for Classification
of X-Ray Images of Patients with Lung Diseases**”

Recenzja rozprawy doktorskiej mgr Anny Kloski została sporządzona w związku z powołaniem przez Radę Naukową Dyscypliny Nauki Medyczne Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy, piszącego niniejszą recenzję, jako recenzenta rozprawy doktorskiej mgr Anny Kloski pismem z dnia 13 czerwca 2023 r.

Kryteria oceny dysertacji wynikają z przepisów zawartych w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2021 r. poz. 478). Rozprawę doktorską mgr Anny Kloski stanowi zbiór opublikowanych i powiązanych tematycznie artykułów naukowych, który spełnia wymagania zawarte w art. 187 punkt 3 oraz punkt 4 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2021 r. poz. 478).

Problem badawczy i jego znaczenie

Zakres recenzowanej rozprawy dotyczy szerokiej i dynamicznie rozwijającej się problematyki wykorzystania metod sztucznej inteligencji w diagnostyce medycznej, W szczególności Doktorantka koncentruje się na zagadnieniach dotyczących klasyfikacji jednostek chorobowych widocznych na zdjęciach rentgenowskich klatki piersiowej. Problem badawczy recenzowanej rozprawy obejmuje ocenę wpływu różnych technik rozszerzania danych, różnych technik przetwarzania wstępnego danych oraz ekstrakcji cech na wyniki klasyfikacji obrazów z prześwietleń klatki piersiowej. W badaniach eksperymentalnych

wykorzystano takie algorytmy uczenia maszynowego jak konwolucyjne sieci neuronowe oraz algorytmy bazujące na koncepcji drzew decyzyjnych. Zbiory danych zdjęć radiologicznych, które stanowiły podstawę do przeprowadzenia eksperymentów pochodzą z dwóch źródeł. Jeden zbiór danych został zgromadzony w Szpitalu Uniwersyteckim nr 1 im. dr Antoniego Jurasza w Bydgoszczy, dwa pozostałe pochodzą z publicznie dostępnego repozytorium danych (Kaggle).

Tematyka podjęta przez mgr Annę Kloskę jest interesująca, w pełni uzasadniona oraz odpowiada na wyzwania współczesnej medycyny, w szczególności wpisuje się w paradygmat Medycyny 4.0. Należy również zaznaczyć, że publikacje naukowe, które są podstawą o ubieganie się mgr Anny Kloski o stopień doktora są już zauważone przez światowe środowisko naukowe, czego dowodem są ich cytowania. Baza *Web of Science* indeksuje jedenaście cytowań, natomiast baza *Scopus* indeksuje siedemnaście cytowań prac Doktorantki bez autocytowań.

Struktura pracy oraz wiedza Autorki

Rozprawę doktorską mgr Anny Kloski stanowi zbiór trzech opublikowanych i powiązanych tematycznie artykułów naukowych, które ukazały się w następujących periodykach naukowych: *Journal of Clinical Medicine* (IF 4,96), *Plos One* (IF 3,75) oraz *Polish Journal of Radiology* (IF 1,2).

Zbiór omawianych publikacji opatrzony jest opracowaniem, które składa się z: listy stosowanych skrótów, wprowadzenia, zdefiniowanych celów badawczych, streszczenia prac wchodzących w cykl powiązanych tematycznie artykułów naukowych, kopii artykułów naukowych będących podstawą o ubieganie się Autorki dysertacji o nadanie stopnia doktora, wniosków końcowych, bibliografii odnoszącej się do treści przedstawionej we wprowadzeniu, kopii oświadczeń współautorów dotyczących wkładu do każdej publikacji wchodzącej w skład cyklu powiązanych publikacji, kopii opinii wydanej przez Komisję Bioetyczną, streszczenia w języku polskim oraz angielskim.

Rozdział *Introduction* zawiera wprowadzenie do problematyki obrazowania medycznego, ze szczególnym uwzględnieniem badań radiologicznych. Omawiany rozdział zawiera również opis wykorzystania metod uczenia maszynowego w diagnostyce obrazowej. W szczególności Doktorantka przedstawiła zagadnienia i problemy badawcze wykorzystywane w publikacjach naukowych wchodzących w skład cyklu powiązanych artykułów takie jak: augmentacja danych (wzbogacenie, rozszerzenie danych), metody przetwarzania wstępnego,

miary oraz metryki stosowane w ocenie trafności klasyfikacji, problemy etyczne dotyczące wykorzystania metod uczenia maszynowego w obrazowaniu medycznym oraz zganienia ułatwiające zrozumienie działania metod uczenia maszynowego, które wpisują się w nurt badań naukowych dotyczący wyjaśnialnej sztucznej inteligencji.

W rozdziale *Study aims* zostały przedstawione cztery szczegółowe cele badawcze, które zostały zrealizowane, a ich wyniki zostały przedstawione na łamach periodyków naukowych wchodzących w skład zbioru powiązanych tematycznie publikacji. Osiągnięte cele badawcze dotyczą:

- 1) oceny wpływu metod rozszerzania danych (augmentacji danych) na wyniki klasyfikacji z obrazowania radiologicznego,
- 2) oceny wpływu metod przetwarzania wstępnego obrazów na wyniki klasyfikacji z obrazowania radiologicznego,
- 3) porównania metod ekstrakcji cech wykonywanych dla różnych metod uczenia maszynowego w klasyfikacji obrazów klatki piersiowej,
- 4) dyskusję dotyczącą zagadnień etycznych wykorzystania metod uczenia maszynowego w komputerowo wspomaganiej diagnostyce obrazowej.

Zamieszczone oświadczenia współautorów publikacji naukowych wchodzących w skład zbioru powiązanych tematycznie artykułów naukowych pozwalają stwierdzić, iż wkład mgr Anny Kloski w ich przygotowanie jest znaczący i obejmuje najważniejsze aspekty merytoryczne oraz badawcze niezbędne w procesie tworzenia publikacji naukowej. Dodatkowo stosowny opis wykonanych prac badawczych, które wykonała mgr Anna Kloska, w każdej z publikacji znajduje się w streszczeniu opublikowanych artykułów (rozdział 5).

Spis literatury liczy 83 pozycje. Cytowane prace dobrane są prawidłowo, są aktualne i odnoszą się do omawianych w pracy problemów.

Wkład Autorki — oryginalne osiągnięcia

Oryginalne osiągnięcia mgr Anny Kloski przedstawione w dysertacji dotyczą analizy metod uczenia maszynowego zastosowanych w diagnostyce medycznej wykonywanej wykonanej z wykorzystaniem obrazowania medycznego. W szczególności prace mgr Anny Kloski dotyczą rozpoznawania jednostki chorobowej jaką jest COVID-19 oraz zapalenie płuc. Oryginalne osiągnięcia Autorki dysertacji dotyczą:

1. Zaplanowania oraz wykonania eksperymentów komputerowych mających na celu weryfikację celu badawczego dotyczącego oceny wpływu metod rozszerzania danych

(augmentacji danych) na wyniki klasyfikacji z obrazowania radiologicznego przy posiadaniu mało licznego zbioru danych uczących. Zbiór danych uczących został zgromadzony w Szpitalu Uniwersyteckim nr 1 im. dr Antoniego Jurasza w Bydgoszczy. Zgromadzony zbiór danych zawiera jedynie 60 opisów zdjęć radiologicznych wraz z etykietami klas, tak więc cel badawczy postawiony przez mgr Annę Kloskę jest jak najbardziej uzasadniony i trafny. Wyniki badań eksperymentalnych dotyczących niniejszego wątku badawczego ukazały się w następujących artykułach naukowych:

- Kloska, A., Tarczewska, M., Giełczyk, A., Kloska, S. M., Michalski, A., Serafin, Z., & Woźniak, M. (2023). Influence of augmentation on the performance of the double ResNet-based model for chest X-ray classification. *Polish Journal of Radiology*, 88, e244.
 - Giełczyk, A., Marciniak, A., Tarczewska, M., Kloska, S. M., Harmoza, A., Serafin, Z., & Woźniak, M. (2022). A novel lightweight approach to COVID-19 diagnostics based on chest X-ray images. *Journal of Clinical Medicine*, 11(19), 5501.
2. Zaplanowania oraz wykonania eksperymentów komputerowych mających na celu weryfikację celu badawczego dotyczącego oceny wpływu metod przetwarzania wstępnego obrazów radiologicznych na wyniki klasyfikacji COVID-19 oraz zapalenia płuc z wykorzystaniem publicznie dostępnego zbioru danych zawierającego blisko 7000 obrazów. Wyniki badań eksperymentalnych dotyczących niniejszego wątku badawczego ukazały się w:
- Giełczyk, A., Marciniak, A., Tarczewska, M., & Lutowski, Z. (2022). Pre-processing methods in chest X-ray image classification. *Plos one*, 17(4), e0265949.
3. Wykonania analizy porównawczej zaproponowanych schematów rozpoznawania jednostek chorobowych z innymi, znanymi z literatury wynikami badań dotyczącymi klasyfikacji zdjęć rentgenowskich klatki piersiowej. Stosowne porównania znajdują się we wszystkich pracach mgr Anny Kloski wchodzących w skład powiązanego tematycznie zbioru publikacji naukowych.

Wymienione osiągnięcia świadczą o umiejętności samodzielnego prowadzenia pracy naukowej przez mgr Annę Kloskę oraz o oryginalnym rozwiązaniu problemu naukowego jakim jest wykorzystanie metod uczenia maszynowego w diagnostyce obrazowej.

Uwagi krytyczne i dyskusje

W punkcie 3.4 przedstawiono metryki jakości klasyfikacji. Lakonicznie została przedstawiona krzywa ROC, która jest użyteczna w szukaniu optymalnego modelu uczenia maszynowego. Należy zauważyć, że pole pod krzywą ROC, dla binarnych zadań klasyfikacji, odpowiada metryce jaką jest dokładność zbalansowana BAC (*balanced accuracy*). Pole pod krzywą ROC to ogólnie metryka AUC (*area under ROC curve*). Powyższa uwaga wynika z faktu, iż Doktorantka na łamach rozprawy (str. 26) oraz publikacji odnosi się do metryki AUC. Dodatkowo na stronie 21 prowadzona jest dyskusja dotycząca tzw. zbiorów niezbalansowanych, dla których powinno stosować się metrykę dokładności zbalansowanej, o czym w tekście pracy nie wspomniano.

Cele badawcze dotyczące oceny wpływu metod rozszerzania oraz wstępnego przetwarzania danych na jakość klasyfikacji obrazów radiologicznych zostały osiągnięte. Aby rozwiązać wszelkie wątpliwości wynikające z podziału pierwotnego zbioru danych na zbiór uczący, walidacyjny oraz testowy eksperymenty powinny zostać powtórzone wielokrotnie. W artykule „*Pre-processing methods in chest X-ray image classification*” stosowny podział został wykonany trzykrotnie. Zwiększenie liczby losowego podziału pierwotnego zbioru danych pozwoliłoby na postawienie hipotezy badawczej o następującej treści: wybór metody przetwarzania wstępnego ma statystycznie istotny wpływ na wyniki klasyfikacji zdjęć rentgenowskich klatki piersiowej.

W artykule „*Pre-processing methods in chest X-ray image classification*” wyszczególniono trzy etykiety klas. W omawianym przypadku zastosowano miary jakości klasyfikacji wywodzące się z macierzy pomyłek w wersji „jeden kontra wszystkie pozostałe”. W tym przypadku zastosowanie miar jakości klasyfikacji w wersji mikro oraz makro uśredniania (np. *micro average precision*, *macro average precision*) pozwala na rozszerzenie dyskusji wyników.

Błędy językowe, kompozycji tekstu oraz typograficzne

- Odwołania do pozycji literaturowych podane są w nawiasach okrągłych.
- W przypadku cytowania wielu pozycji nie użyto znaku spacji np. str. 8: (10,11).
- W treści dysertacji w wielu miejscach separatorem dziesiętnym jest znak przecinka (,) np. str. 27. Na stronie 30 separatorem dziesiętnym jest znak kropki (.).

- Opis autorskiego wkładu dotyczącego każdej publikacji wchodzącej w zbiór opublikowanych i powiązanych tematycznie artykułów naukowych jest jednakowy (str. 28, 31 oraz 33).
- W przypadku pojęć odnoszących się do macierzy pomyłek (np. etykieta prawdziwie pozytywna – TP) w dysertacji stosowane są dwa pojęcia: miara i metryka (str. 30 oraz str. 33).
- Str. 9: examined. (14),
- Artykuł “*Influence of augmentation on the performance of the double ResNet-based model for chest X-ray classification*”, str. 3: in range \$ <-3, 3> \$,
- Artykuł “*Pre-processing methods in chest X-ray image classification*”, str. 7 kodowanie oznaczenia metod przetwarzania wstępnego A ... F, na str. 8 kodowanie 1...6.

Podsumowanie

Reasumując stwierdzam, iż mgr Anna Kloska posiada ogólną wiedzę z zakresu obrazowania medycznego (diagnostyki obrazowej), która mieści się w dyscyplinie naukowej nauki medyczne. Dodatkowo mgr Anna Kloska posiada ogólną wiedzę teoretyczną dotyczącą metod uczenia maszynowego, która mieści się w dyscyplinie informatyka techniczna i telekomunikacja, co świadczy o interdyscyplinarności wiedzy Doktorantki. Lektura dysertacji pozwala stwierdzić, że Autorka zaprezentowała na jej łamach umiejętność samodzielnego prowadzenia pracy naukowej, której efekty zostały opublikowane jako zbiór powiązanych tematycznie artykułów naukowych. Należy zaznaczyć, iż każda pozycja stanowiąca zbiór powiązanych tematycznie artykułów naukowych mieszczących się w zakresie dysertacji mgr Anny Kloski przypisana jest do dyscypliny nauki medyczne, zgodnie z załącznikiem do komunikatu Ministra Edukacji i Nauki z dnia 17 lipca 2023 r. w sprawie wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych.

Wobec powyższego, recenzowana praca spełnia wymagania zdefiniowane przez artykuł 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2021 r. poz. 478). Konkludując, wnoszę o przyjęcie rozprawy oraz dopuszczenie mgr Anny Kloski do publicznej obrony.

Jednocześnie, mając na uwadze aktualność podjętej tematyki badawczej, wysoki poziom naukowy oraz zauważalność prac Doktorantki przez środowisko naukowe, czego wyrazem są ich cytowania, przedkładam wniosek o wyróżnienie dysertacji mgr Anny Kloski.

R. Burduk