

Katedra Radiologii
Uniwersytet Jagielloński

Dr hab. med. Robert Chrzan prof. UJ
Katedra Radiologii
Uniwersytet Jagielloński
Collegium Medicum

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

Kraków, 18.07.2023

Ocena rozprawy doktorskiej mgr Anny Kloski

„Use Of Artificial Intelligence Methods for Classification Of X-Ray Images of Patients with Lung Diseases”

Przedstawiona do recenzji rozprawa doktorska jest oparta o spójny tematycznie cykl 3 publikacji, w recenzowanych czasopismach, o łącznym współczynniku Impact Factor 9,046 i łącznej punktacji Ministerstwa Edukacji i Nauki 310 punktów.

1. Giełczyk A, Marciniak A, Tarczewska M, Lutowski Z. Pre-processing methods in chest X-ray image classification. PLoS One. 2022 Apr 5;17(4):e0265949. doi: 10.1371/journal.pone.0265949. PMID: 35381050; PMCID: PMC8982897.

W pracy tej autorzy analizowali wpływ wstępnego przetwarzania obrazów rtg klatki piersiowej na skuteczność klasyfikacji tych obrazów przez sieć neuronową do jednej z trzech klas: obraz normalny, COVID-19, pneumonia.

Obrazy pochodziły z publicznie dostępnego zasobu, obejmującego łącznie 6939 przypadków, po 2313 obrazów dla każdej z powyższych klas, przy czym, zgodnie z opisem źródłowym zasobu, pneumonia obejmowała zarówno przypadki o etiologii bakteryjnej, jak i wirusowej.

Jako metody wstępnego przetwarzania obrazów testowano, niezależnie oraz w kilku kombinacjach: normalizację histogramu, rozmycie gaussowskie, adaptacyjne maskowanie oraz bilateralny filtr odszumiający.

Skuteczność klasyfikacji była oceniana za pomocą parametrów:

$$\text{accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{precision} = TP / (TP + FP) \text{ czyli wartość predykcyjna wyniku dodatniego}$$

$$\text{recall} = TP / (TP + FN) \text{ czyli czułość}$$

$$F1\text{-score} = 2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall})$$

Autorzy wykazali, że wstępne przetwarzanie obrazów może podwyższyć skuteczność ich klasyfikacji przez sieć neuronową, przy czym najbardziej obiecującą techniką okazało się połączenie normalizacji histogramu, rozmycia gaussowskiego i adaptacyjnego maskowania, gdzie dla poszczególnych klas: obraz normalny, COVID-19, pneumonia, wartości F1-score wynosiły odpowiednio: 0,97, 0,96, 0,99.

Jest to bardzo istotny wniosek, prezentujący jeden ze sposobów dalszego zwiększenia skuteczności oceny obrazów medycznych przez sieci neuronowe.

Klasa COVID-19 oznaczała oczywiście również zapalenie płuc, wywołane przez SARS-CoV-2, w związku z tym w tekście pracy przydatne byłoby podanie definicji klasy pneumonia, jako zapalenia płuc o innej etiologii, w tym zarówno bakteryjnej, jak i wirusowej.

Parametry oceny, wykorzystywane przez autorów, są typowo stosowane w analizie skuteczności sieci neuronowych, natomiast w publikacjach medycznych często używa się również specyficzności, obliczanej jako $TN / (FP + TN)$. Ewentualne dodanie tego parametru do tabeli wyników mogłoby ułatwić percepcję pracy przez czytelników przyzwyczajonych do typowo medycznych opracowań.

Na uzyskane przez autorów bardzo wysokie parametry skuteczności mógł mieć wpływ fakt operowania na wyselekcjonowanej grupie obrazów, zawierającej, oprócz przypadków prawidłowych, wyłącznie obrazy typowe dla COVID-19 albo dla pneumonii o innej etiologii. W przypadku testowania skuteczności detekcji COVID-19 na grupie obrazów rtg klatki piersiowej obejmującej znacznie większy zakres możliwych patologii, jak to się spotyka w codziennej praktyce, przypuszczalnie wartości tych parametrów byłyby niższe. Należy też zwrócić uwagę, że podczas pandemii u części pacjentów z COVID-19, występowało współistniejące zapalenie płuc o innej etiologii, zwykle bakteryjnej, jak też ostra zatorowość płucna.

2. Giełczyk A, Marciniak A, Tarczewska M, Kloska SM, Harmoza A, Serafin Z, Woźniak M. A Novel Lightweight Approach to COVID-19 Diagnostics Based on Chest X-ray Images. J Clin Med. 2022 Sep 20;11(19):5501. doi: 10.3390/jcm11195501. PMID: 36233368; PMCID: PMC9571927.

W powyższej pracy wykorzystano obrazy rtg klatki piersiowej 30 zdrowych osób i 30 pacjentów z COVID-19 potwierdzonym w teście PCR. Ze względu na małą liczebność zbioru, użyto techniki augmentacji - generowania sztucznych danych, opartej na obrotach, dodawaniu szumu oraz zmianie rozmiaru wyjściowych obrazów, co zwiększyło dziesięciokrotnie liczbę obrazów do dalszej analizy. 80% zestawu użyto do uczenia sieci, 10% do walidacji i 10% do testowania.

Po zamianie do skali szarości i normalizacji, obrazy zostały poddane maskowaniu, z segmentacją obszaru płuc, przy użyciu publicznie dostępnego wstępnie wyszkolonego modelu.

Następnie przy użyciu konwolucyjnej sieci neuronowej wykonano ekstrakcję cech oraz klasyfikację, jak też tylko ekstrakcję cech, natomiast klasyfikację niezależnie przy pomocy XGBoost, Random Forest, LightGBM i CatBoost, z porównaniem wyników pomiędzy klasyfikatorami.

Najwyższą skuteczność w różnicowaniu obrazów prawidłowych i obrazów pacjentów z COVID-19 uzyskano dla XGBoost i dla LightGBM - każdy z parametrów: accuracy, precision, recall, F1-score równy 1, przy czym LightGBM był istotnie szybszy.

Według autorów, powyższe wyniki były lepsze niż uzyskane przez innych badaczy, wykorzystujących dużo bardziej skomplikowane architektury, z ogromną liczbą parametrów podczas uczenia sieci, wymagających znacznie większej mocy obliczeniowej i długiego czasu uczenia.

Oznacza to, że wybór właściwej architektury jest kluczową decyzją w optymalizacji skuteczności sieci neuronowych w analizie obrazów medycznych, jak też, że implementacja takich rozwiązań jest już możliwa bezpośrednio na stacji roboczej końcowego użytkownika, zamiast na dedykowanych serwerach.

Podobnie jak w pierwszej pracy, na uzyskane maksymalne wartości skuteczności mogło mieć wpływ testowanie opracowanego rozwiązania na grupie obrazów, składającej się wyłącznie z dwóch klas – prawidłowych i COVID-19. Oczywiście, autorzy jasno definiują to w tekście pracy, jednak grupa czytelników, którzy z powodu ogromnej ilości interesujących publikacji

naukowych ograniczają się wyłącznie do analizy streszczeń, może nabrać przeświadczenia, że stuprocentowa trafność detekcji COVID-19 przez sieć neuronalną dotyczy wszelkich możliwych obrazów rtg klatki piersiowej.

Należy również zwrócić uwagę, że prace badaczy, z którymi porównywano wyniki, były oparte na analizie znacznie większej liczby obrazów, jak też w pracy Rajagopala, w tabeli 2 z najniższymi parametrami skuteczności, wykorzystywano trzy klasy obrazów (prawidłowe, COVID-19 i pneumonia), co oczywiście mogło wpływać na rezultaty analizy.

Interesująca byłaby próba sprawdzenia skuteczności rozwiązania autorów na zbiorze testowym jednego z innych badaczy.

Praca Laeli dotyczyła analizy obrazów rtg pacjentów z gruźlicą i prezentowała tylko wartości parametrów accuracy i AUC, związku z czym stosowniej byłoby ją pominąć w zestawieniu.

W tekście pracy podano, że obliczenia wykonywano w oparciu o karty Tesla, z przetwarzaniem przez GPU z wykorzystaniem architektury CUDA. Przydatne byłoby podanie, o jaki konkretnie model serii Tesla chodziło, co jest o tyle ważne, że seria ta została wprowadzona na rynek przez firmę Nvidia jeszcze w 2006 r. i najnowsze modele dysponują mocą obliczeniową ok. 10 razy większą niż pierwsze produkty.

We wstępie do pracy, autorzy stwierdzają, że będący złotym standardem w wykrywaniu SARS-CoV-2 test PCR może dawać fałszywe wyniki oraz niektóre osoby zarażone tym wirusem mogą nie prezentować objawów choroby i dlatego nie podlegają testom PCR. Według autorów, w takich przypadkach diagnoza jest możliwa na podstawie rtg klatki piersiowej. Jednak jest mało prawdopodobne, żeby pacjenci bez żadnych objawów klinicznych prezentowali zmiany zapalne płuc wywołane SARS-CoV-2, widoczne na obrazach rtg. W przypadku osób bezobjawowych, które miały kontakt z chorymi na COVID-19, testy PCR mogły wykryć zarażenie jeszcze przed wystąpieniem objawów, zarówno klinicznych jak i możliwych do uwidocznienia w badaniach obrazowych, co było podstawą dla zastosowania izolacji / kwarantanny, mającej na celu ograniczenie szerzenia się wirusa w populacji.

W numeracji odnośników do publikacji autorów w Tabeli 2 wkradły się drobne błędy edytorskie, m. in. dla Rajagopal powinno być [23] zamiast [27], a dla Laeli [24] zamiast [28], co oczywiście nie ma znaczenia dla samych prezentowanych wyników.

3. Kłoska A, Tarczewska M, Giełczyk A, Kłoska SM, Michalski A, Serafin Z, Woźniak M. Influence of augmentation on the performance of the double ResNet-based model for chest X-ray classification. Pol J Radiol. 2023 May 12;88:e244-e250. doi: 10.5114/pjr.2023.126717. PMID: 37346422; PMCID: PMC10280365.

W pracy tej badano wpływ zastosowania różnych technik augmentacji danych na skuteczność klasyfikacji przez sieć neuronową obrazów rtg klatki piersiowej osób zdrowych i pacjentów z COVID-19.

Klasyfikacja była wykonywana przy pomocy wstępnie wyszkolonej konwolucyjnej sieci neuronowej ResNet18, z uczeniem na zbiorze ponad 30000 obrazów rtg zawierającym przypadki prawidłowe oraz w przebiegu COVID-19.

Testowaniu podlegał wyjściowy zbiór 62 obrazów rtg przypadków prawidłowych oraz w przebiegu COVID-19, a następnie kolejne zbiory uzyskane w wyniku zastosowania różnych technik augmentacji danych (zmiana koloru, zmiana kontrastu i jasności, dodawanie szumu, zniekształcenia geometryczne, obrót oraz połączenie wszystkich powyższych metod).

Ponadto powyższa procedura została przeprowadzona zarówno bez, jak i z wykorzystaniem maskowania podczas wstępnego przetwarzania obrazów.

W każdym powyższym wariancie obliczono wartości parametrów: accuracy, precision, recall, F1-score, a następnie porównano wyniki.

Najwyższe wartości powyższych parametrów uzyskano dla wariantu augmentacji danych obejmującego połączenie wszystkich metod, z maskowaniem podczas wstępnego przetwarzania obrazów (accuracy 0,9839, precision 0,9697, recall 1,0000, F1-score 0,9846).

Wyniki te były istotnie wyższe w porównaniu z wartościami bez augmentacji, w szczególności dla wariantu bez maskowania (accuracy 0,8710, precision 0,8750, recall 0,8750, F1-score 0,8750).

Autorzy wykazali w ten sposób, że technika augmentacji danych, w szczególności łącząca różne metody, może istotnie podwyższyć skuteczność klasyfikacji przez sieć neuronową, co jest zwłaszcza istotne dla małych zbiorów danych. Ponadto zastosowanie maskowania dodatkowo podwyższa tę skuteczność.

Jest to kolejny, bardzo ważny wniosek autorów, mogący skutkować zwiększeniem skuteczności klasyfikacji obrazów medycznych przez sieci neuronowe.

Oprócz wyżej przedstawionych publikacji, rozprawa doktorska obejmuje również:

- wstęp, w którym wyczerpująco umówiono tematykę sieci neuronowych, w tym ich typy, z podkreśleniem zalet i niedogodności poszczególnych rozwiązań, parametry oceny skuteczności, jak też kwestie etyczne wykorzystania w medycynie,
- cele pracy ze zdefiniowanym celem głównym, którym było zbadanie wykonalności i skuteczności wykorzystania modeli AI do klasyfikacji diagnostycznej pacjentów z COVID-19 na podstawie ich zdjęć rentgenowskich klatki piersiowej oraz z celami pobocznymi,
- wnioski, w których potwierdzono realizację wszystkich zamierzonych celów pracy,
- listę piśmiennictwa, liczącą 83 pozycje.

We wstępie, na str. 20 do wzoru definiującego specyficzność wkraśl się błąd edytorski – zamiast $\text{Specificity} = \text{TN} / (\text{TN} + \text{FN})$ powinno być $\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$.

Recenzowana praca doktorska cechuje się innowacyjnością, dowodzi szerokiej ogólnej wiedzy teoretycznej kandydatki, jak też potwierdza umiejętność samodzielnego prowadzenia pracy naukowej.

Reasumując, przedstawiona do oceny rozprawa doktorska mgr Anny Kloski „Use Of Artificial Intelligence Methods for Classification Of X-Ray Images of Patients with Lung Diseases” spełnia wszystkie wymogi określone w ustawie Prawo o szkolnictwie wyższym i nauce, konieczne dla uzyskania stopnia doktora.

Dlatego też przedstawiam Radzie Dyscypliny Nauki Medyczne Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy, Uniwersytetu Mikołaja Kopernika w Toruniu, wniosek o dopuszczenie mgr Anny Kloski do dalszych etapów przewodu doktorskiego.

Jednocześnie uznaję, że praca zasługuje na wyróżnienie.

lekarz dr hab. n. med. prof. UJ
Robert Chrzan
specjalista radiodiagnostyki
30-051 Kraków, ul. Urzędnicza 25/6
nr 1212232 tel. 693271676

Kraków, 18.07.2023

Robert Chrzan