

Recenzja
Rozprawy Doktorskiej
mgr Weronika Hryniewska- Guzik
**pt.: A multi-level perspective and deep learning models and human-oriented
explanations with applications to medical images**

Dziedzina: nauki techniczne,

Dyscyplina: Informatyka Techniczna i Telekomunikacja

Recenzowana praca powstawała pod kierunkiem naukowym prof. dra hab. inż. Przemysława Biecka. Niniejszą recenzję przygotowano na podstawie uchwały z dn. 25.06.2024 Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej zgodnie z art. 187 ustawy z dnia 20 lipca 2018 r. – Prawo o Szkolnictwie Wyższym i Nauce (j.t. Dz.U. z 2020 r. poz. 85, z późn. zm.).

1. Cel i teza rozprawy

Rozprawa doktorska koncentruje się na wyjaśnialnej sztucznej inteligencji (Explainable AI, XAI) w zastosowaniach medycznych, stanowiąc istotny wkład w jedną z kluczowych i najszybciej rozwijających się dziedzin współczesnej nauki i technologii. Tematyka ta jest niezwykle aktualna, biorąc pod uwagę rosnące zapotrzebowanie na przejrzystość, interpretowalność i weryfikowalność decyzji podejmowanych przez algorytmy sztucznej inteligencji, szczególnie w obszarach o wysokim poziomie ryzyka, takich jak medycyna. Praca doktorskiego wpisuje się w ten nurt, uwzględniając aktualne regulacje i wytyczne Unii Europejskiej, w tym zapisy AI Act, które podkreślają konieczność rzetelnej, niezależnej i etycznej weryfikacji wyników systemów stosowanych w krytycznych zastosowaniach, takich jak analiza obrazów medycznych czy podejmowanie decyzji klinicznych.

Celem pracy było opracowanie nowatorskiego podejścia, które nie tylko dostarcza wyjaśnialnych wyników, ale również umożliwia ich praktyczne zastosowanie w medycynie. Rozwiązanie to wspiera lekarzy i personel medyczny w podejmowaniu bardziej świadomych i wiarygodnych decyzji diagnostycznych, jednocześnie zwiększając zaufanie do systemów sztucznej inteligencji oraz ich

akceptację w środowisku medycznym. Rozprawa uwzględnia także aspekty techniczne i etyczne wyjaśnialnej AI, co czyni ją kompleksowym wkładem w rozwój tej interdyscyplinarnej dziedziny.

W ramach rozprawy, doktorantka skupiła się na analizie wpływu na metryki jakości modeli wieloaspektowych oraz na rozwijaniu metod wyjaśnialności ukierunkowanych na użytkownika końcowego - zorientowanych na człowieka.

Celem badań było opracowanie rozwiązania umożliwiającego wykrywanie nawet najmniejszych zmian w obrazach medycznych oraz wprowadzenie pojęcia polireprezentacji, mającej na celu integrację informacji pochodzących z tej samej modalności (np. różne obrazy radiologiczne). Praca obejmuje również autorską metodę LIMEcraft, która wprowadza mechanizmy selekcji semantycznie spójnych fragmentów obrazu, co pozwala na precyzyjne wyjaśnienia zmienności obrazu w odniesieniu do predykcji modelu. **Powyższe zagadnienie stanowią fragment zaproponowanych rozwiązań, które zostały szczegółowo opisane w 6 artykułach stanowiących podstawę recenzowanej rozprawy doktorskiej.**

Istotnym aspektem badań było również zapewnienie zgodności z zasadami etycznymi, w tym z wymaganiami AI Act oraz RODO, co jest kluczowe dla implementacji rozwiązań wykorzystujących algorytmy sztucznej inteligencji, szczególnie głębokie, w środowisku medycznym. Autorka skupiła się na wypracowaniu i przetestowaniu metod wyjaśnialności oraz na procesie walidacji rozwiązań, zwracając szczególną uwagę na tworzenie etycznych zbiorów danych, które obejmują różnorodne scenariusze i zrównoważone podzbiory, dzięki czemu modele mogą być bardziej reprezentatywne i mniej podatne na błędy i stronniczość.

Podsumowując, **tezę pracy**, która została przedstawiona na stronie 18 można uznać za istotną z punktu widzenia zarówno naukowego (nowe podejścia do wyjaśnialności) jak też w aspekcie przyszłych zastosowań proponowanego rozwiązania w praktyce wdrożeniowej aplikacjach medycznych. Przedstawione wnioski i wyniki badawcze otwierają nowe możliwości w zakresie tworzenia bardziej przejrzystych i zrozumiałych modeli AI dla medycyny oraz ich implementacji w rzeczywistych zastosowaniach klinicznych, co stanowi istotny wkład w rozwój sztucznej inteligencji.

Teza rozprawy doktorskiej brzmi następująco:

Multifaceted models and human-oriented explanations have a positive impact on quality metrics.

W celu udowodnienia powyższej tezy pracy zdefiniowano, zaproponowano i rozwiązano następujące pytania badawcze, które określa jako Research Questions:

1. What are the weaknesses of explanations and models that classify patients with COVID-19?
2. How to combine models to achieve better predictive performance?
3. How to make explanations more human-friendly and aligned with domain knowledge?
4. How to combine explanations to achieve better faithfulness and localization?

Doktorantka na str. 18 udziela krótkiej odpowiedzi na każde z powyższych pytań wskazując publikację w której jest szczegółowa odpowiedź z opisem, implementacją oraz weryfikacją wyników.

Rysunek 1.1 ilustruje wkład mgr Weroniki Hryniewicz-Guzik w rozwój dyscypliny, uwzględniając podział osiągnięć na dwie kategorie: **reprezentacja płuc** oraz **wyjaśnialność**.

W ramach reprezentacji osiągnięto:

1. Opracowanie listy kontrolnej do tworzenia odpowiedzialnych rozwiązań opartych na głębokim uczeniu (**Checklist for Responsible Deep Learning**),
2. Opracowanie modeli wielozadaniowych do wykrywania niewielkich zmian na obrazach CT (**Multi-task Models on CT Scans**),
3. Wprowadzenie pojęcia polireprezentacji, którego celem jest łączenie reprezentacji uzyskanych na różne sposoby z danych tej samej modalności (**Polyrepresentation**).

W ramach wyjaśnialności zrealizowano:

1. Implementację interaktywnej metody **LIMEcraft**,
2. Opracowanie i implementację metod łączenia wyjaśnień **NormEnsembleXAI**,
3. Opracowanie i implementację metod łączenia wyjaśnień w podejściu **ensembling** dla sieci konwolucyjnych.

Na podstawie powyższych rozważań oraz uzyskanych wyników można stwierdzić, że hipoteza naukowa pracy, podana na początku niniejszej recenzji została potwierdzona.

2. Przegląd treści rozprawy

Tekst rozprawy liczy 212 stron i składa się z 5 rozdziałów oraz podsumowania. Rozprawa doktorska jest tzw. ‘zszywką’, choć poszczególne rozdziały, które są oparte na publikacjach zostały dodatkowo sformatowane co sprawia, że rozprawa doktorska jest jednorodna. Rozdział 3 i 4 bazują na 6 publikacjach, natomiast pozostałe stanowią wprowadzenie do tematyki rozprawy oraz jej

podsumowanie. Rozprawa doktorska zawiera również 295 pozycji literaturowych streszczenia, spis rysunków i spis tabel.

Przegląd treści rozprawy jest omówiony poniżej.

Rozdział 1 stanowi wprowadzenie do rozprawy doktorskiej przedstawiając zakres prowadzonych badań oraz konieczność ich podjęcia. Sformułowano problem badawczy, tezę pracy, pytania badawcze, określono wkład w dyscyplinę oraz przedstawiono artykuły, które są podstawą rozprawy doktorskiej.

W **rozdziale 2** Autorka wprowadza do tematyki badań z zakresu rozprawy. Rozdział zawiera podstawowe pojęcia dotyczące analizy obrazów medycznych, uczenia głębokiego, problemu i jednocześnie konieczności wyjaśnialności i interpretacji wyników oraz modeli.

Rozdział 3 i 4 stanowią główne części rozprawy doktorskiej i przedstawiają opublikowane artykuły dla zagadnień reprezentacji płuc oraz wyjaśnialnej AI.

W **rozdziale 3** przedstawiono wyniki badań, które zostały opisane w trzech artykułach i tworzą cykl publikacyjny dotyczący zagadnienia analizowania zdjęć CT oraz MRI płuc z uwzględnieniem reprezentacji oraz wyjaśnialności:

Pierwsza praca (55% udział procentowy, **Hryniewska, W.**, Bombiński, P., Szatkowski, P., Tomaszewska, P., Przelaskowski, A., Biecek, P., “Checklist for responsible deep learning modeling of medical images based on COVID-19 detection studies”. *Pattern Recognition*(2021), 118, p. 108035) zawiera szczegółowy przegląd procesu analizy zdjęć medycznych z wykorzystaniem metod głębokich. Autorka przedstawiła przegląd literatury, analizę zbiorów danych oraz poszczególnych etapów w tym augmentacji, modele głębokie służące jako backbone przy transferze stylu oraz proces ewaluacji wyników. Praca stanowi przede wszystkim bardzo solidny i krytyczny przegląd literatury co pozwoliło na stworzenie przez Doktorantkę listy kontrolnej przedstawiającej poszczególne kroki przy implementacji rozwiązania. Doktorantka na stronie 86-87 opisała wnioski wynikające z dogłębnej analizy tematu, które wydają się być wręcz aktualne również dla innych rozwiązań wykorzystujących modele głębokie. Sama praca nie przedstawia autorskich rozwiązań implementacyjnych, ale stanowi podstawę dogłębnej analizy dla kolejnych publikacji. Doktorantka deklaruje: oordynację zespołu, wyszukiwanie i przegląd literatury, przygotowanie schematów graficznych, wnioski wymagające znajomości głębokiego uczenia, napisanie większości tekstu oraz większości odpowiedzi na recenzje.

Kolejna praca z tego cyklu (60% udział procentowy, **Hryniewska-Guzik, W.**, Kędzierska, M., Biecek, P., “Multi-task learning for classification, segmentation, reconstruction, and detection on chest CT

scans”(2023), in: *Progress in Polish Artificial Intelligence Research 4*, Lodz University of Technology Press, pp. 251–257) przedstawia porównanie podejścia wielozadaniowego do detekcji, segmentacji i rekonstrukcji na zdjęciach CT płuc. Schemat 3.7 przedstawia implementacyjne podejście do rozwiązania problemu dzielenia parametrów jak i łączenia funkcji kosztów dla wielu problemów. Uzyskane wyniki potwierdzają tezę rozdziału, że w tym przypadku modele wielozadaniowe dają wyższe rezultaty niż pojedyncze. Doktorantka deklaruje: sformułowanie problemu badawczego, zaproponowanie metody, zaprojektowanie architektury głębokiego uczenia, konsultacje dotyczące implementacji, przygotowanie części schematów graficznych, napisanie większości tekstu.

Ostatnia praca w cyklu (85% udział procentowy, **Hryniewska-Guzik, W.**, Biecek, P., “X-ray transferable polyrepresentation learning”. *preprint*(2024)) przedstawia pojęcie polireprezentacji oraz sposób jej tworzenia (Rysunek 3.11). Doktorantka pokazuje proces łączenia trzech reprezentacji: Siamese, Self-supervised oraz Radiomics, które tworzone są dla tych samych danych wejściowych z wykorzystaniem różnych zadań, które mają wykonywać. Doktorantka szczegółowo przedstawiła poszczególne architektury systemu oraz opisała wyniki, które okazują się większościowo najwyższe dla połączenia parametrów z części Siamese, Segmentation i Self-supervised. Doktorantka deklaruje: Zaproponowanie i zaimplementowanie metody, przegląd literatury, przygotowanie schematów graficznych, napisanie całości tekstu.

W **rozdziale 4** przedstawione zostały autorskie rozwiązania dotyczące metod wyjaśnialności i interpretowalności rozwiązań. Doktorantka w trzech publikacjach opisała implementację interaktywnej metody LIMEcraft, metody łączenia wyjaśnień NormEnsembleXAI oraz metody łączenia wyjaśnień ensembling dla sieci konwolucyjnych.

W pracy (60% udział procentowy, Hryniewska, W., Grudzień, A., Biecek, P., “LIMEcraft: handcrafted super-pixel selection and inspection for Visual eXplanations”. *Machine Learning*(2022), 0123456789, doi: 10.1007/s10994-022-06204-w) doktorantka deklaruje: formułowanie problemu badawczego, przegląd literatury, zaproponowanie metody, implementacja frontendu, konsultacje dotyczące implementacji backendu, przygotowanie i przeprowadzenie eksperymentów, przeprowadzenie badania użytkowników, przygotowanie większości schematów graficznych, napisanie większej części tekstu oraz większości odpowiedzi na recenzje. Artykuł dotyczy nowej metody wyjaśniania działania modeli uczących się w kontekście analizy obrazów. Proponowana metoda opiera się na ręcznym i precyzyjnym wyborze superpikseli, co pozwala na lepsze zrozumienie, które fragmenty obrazu wpływają na decyzje modelu. Rozwiązanie umożliwia bardziej kontrolowaną i wiarygodną interpretację wyników, szczególnie w przypadku sieci głębokiego uczenia. Metoda wspiera użytkowników w identyfikacji

kluczowych regionów obrazu, poprawiając jakość wizualnych wyjaśnień. Wpisuje się to w rozwój Explainable AI, zwiększając zaufanie do modeli i ułatwiając ich praktyczne zastosowanie.

W pracy (60% udział procentowy, Hryniewska-Guzik, W., Sawicki, B., Biecek, P., “NormEnsembleXAI: unveiling the strengths and weaknesses of XAI ensemble techniques”. preprint(2024)) autorka deklaruje: sformułowanie problemu badawczego, zaproponowanie metody, napisanie wstępu, przegląd literatury, napisanie większości artykułu, zasugerowanie struktury artykułu, konsultacje podczas implementacji, wskazanie kolejnych kroków do podjęcia oraz bibliotek do wykorzystania, przygotowanie wszystkich schematów graficznych, przygotowanie części analiz. Artykuł przedstawia NormEnsembleXAI, metodę oceny oraz porównywania różnych technik wyjaśnialności sztucznej inteligencji (XAI) w oparciu o ich wady i zalety. Proponowane podejście polega na wykorzystaniu podejścia zespołowego (ensemble), które łączy wyniki różnych metod XAI w celu uzyskania bardziej kompleksowych i zróżnicowanych wyjaśnień. Autorzy podkreślają znaczenie normalizacji wyników poszczególnych metod, co umożliwi ich porównywanie i integrację. W artykule zaprezentowano eksperymenty, które pokazują, jak NormEnsembleXAI pozwala identyfikować kluczowe obszary modelu, które wymagają uwagi lub poprawy. Rozwiązanie to przyczynia się do lepszego zrozumienia, jak różne techniki XAI współpracują ze sobą w celu zapewnienia wiarygodnych i dokładnych wyjaśnień.

W pracy (85% udział procentowy, Hryniewska-Guzik, W., Longo, L., Biecek, P., “CNN-based explanation ensembling for dataset, representation and explanations evaluation”(2024), in: 2nd World Conference on eXplainable Artificial Intelligence) doktorantka deklaruje zaproponowanie i zaimplementowanie metody, przegląd literatury, przygotowanie schematów graficznych, przeprowadzenie eksperymentów, napisanie całości tekstu oraz większości odpowiedzi na recenzje. Artykuł dotyczy zastosowania zespołowego podejścia (ensembling) w wyjaśnialności modeli opartych na sieciach konwolucyjnych (CNN). W pracy zaprezentowano metody pozwalające na identyfikację kluczowych cech danych i oceny ich wpływu na decyzje modelu. Wyniki eksperymentów pokazują, że zastosowanie ensemblingu w wyjaśnialności zwiększa dokładność i wiarygodność analiz oraz pomaga w lepszym zrozumieniu działania modeli.

W rozdziale 5 przedstawiono wnioski końcowe z badań, przedstawiono szczegółowy wkład naukowy i otwarte problemy badawcze z rozważanej tematyki.

Bibliografia pracy jest obszerna, liczy 295 pozycji, bardzo dobrze sformatowana, czytelna jak i jest poprawnie dobrana pod kątem prezentacji rozważanej w pracy problematyki.

3. Uwagi dyskusyjne i krytyczne

Podczas lektury recenzowanej rozprawy nasunęły się również pewne uwagi o charakterze dyskusyjnym. Część z nich ma naturę bardziej ogólną, a część jest szczegółowa.

Uwagi ogólne:

- Proszę o informację dlaczego w pracy zastosowano sieć VGG-13, a nie bardziej aktualne rozwiązania dla części backbone.
- Proszę szczegółowo wyjaśnić wnioski płynące ze wzrostu skuteczności dla modelu wielozadaniowego.
- Proszę o bardziej szczegółowo porównać zaproponowane metody z istniejącymi rozwiązaniami (część 2).

Niektóre zauważone błędy edytorskie i ew. uwagi szczegółowe:

- Str. 5 – medical imaging z małej litery;
- Str. 17 – niepoprawnie używane nazwy rozpoczynające się od wielkiej litery. Uwaga dotyczy całej pracy;
- Str. 17 - Brak konsekwencji – CT zostało wyjaśnione, natomiast X-ray nie;
- Str. 21 – to solve, w pracy -solving;
- Deep Learning w całej pracy z wielkie litery;
- Str. 24 – odwołanie do rysunku 2.31;

4. Ocena formalna

Cała rozprawa została napisana w języku angielskim i pod względem formalnym nie budzi większych zastrzeżeń. Praca zawiera jednak pewne błędy językowe, stylistyczne, interpunkcyjne oraz redakcyjne, które wymagają korekty. Niemniej, układ pracy jest poprawny, logicznie spójny, właściwie sformatowany i zachowuje jednolitość w całej treści.

Uwagi dyskusyjne przedstawione w punktach 3-4 w żadnym wypadku nie umniejszają wartości naukowej rozprawy. Uważam, że praca przedstawia bardzo szeroki i oryginalny dorobek naukowy Doktorantki w obszarze dyscypliny Informatyka Techniczna i Telekomunikacja.

5. Podsumowanie

Zakres i poziom naukowy uzyskanych wyników badawczych odpowiadają w pełni ustawowym i zwyczajowym wymaganiom stawianym rozprawom doktorskim określonym w ustawie z dn. 20 lipca 2018 r (Dz. U. Nr Dz.U.2022.574, z późn. zm.) Prawa o szkolnictwie wyższym i nauce. Wnioskuje zatem do Wysokiej Komisji powołanej przez Radę Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej o przyjęcie rozprawy i dopuszczenie jej Autorki, mgr Weroniki Hryniewskiej-Guzik do publicznej obrony.

