dr hab. inż. Paweł B. Myszkowski, prof. PWR
Politechnika Wrocławska 29.10.2026

Recenzja rozprawy doktorskiej

Michała Wieczorka

zatytułowanej:

Semi-Supervised Training with Fuzzy Logic for Transformer-Based Semantic Segmentation

1. Problem badawczy i jego znaczenie

Oceniana rozprawa doktorska „bada metody projektowania systemów sztucznej inteligencji do analizy obrazów(…)" (ze streszczenia), a dalej naszkicowano jako 3 tezy, sprowadzających się do zastosowania mechanizmów atencji do zadania przetwarzania obrazów, zastosowania rozmytego pół-nadzorowanego algorytmu uczenia do segmentacji obrazów medycznych oraz wprowadzenie technik logiki rozmytej do architektur segmentacji semantycznej (ze strony III). Zastosowane mechanizmy mają poprawić zdolność do generalizacji, poprawić klasyfikację oraz zredukować błędy klasyfikacji pixeli. Choć to dość ogólnie zdefiniowane cele badawcze, to w realizacji mogą być wartościowe.

Niestety, aspekt badawczy w pracy został potraktowany dość ogólnie – podano szkicowo tezy na wstępie, bez weryfikacji w treści, a odniesiono się o nich dopiero w rodz.11.1. A sama rozprawa *de facto* dotyczy zastosowań poszczególnych architektur sieci transformerowych do klasyfikacji obrazów, a sam problem badawczy wydaje się tylko tytułowy. Nie zdefiniowano tez badawczych, nie podano zakresu ich badań, nie wskazano pytań badawczych.

Sama praca doktorska jest amalgamatem 6 różnych zastosowań sieci uczenia głębokiego/transformerów do rozwiązywania różnych problemów. I to jest faktyczna zawartość pracy: zastosowania sieci głębokich. Do detekcji malarii (rodz.4), obrazowania medycznego (rodz.5), wykrywania broni na obrazie (rodz.6), uczenia federacyjnego (rodz.7), klasyfikacji tkanek (rodz.8) i semantycznej hematologicznej segmentacji (rodz.9). I to jest jednocześnie wada i zaleta ocenianej rozprawy doktorskiej, gdyż doktorant nie skupia się na zaproponowaniu swojego autorskiego modelu/algorytmu/mechanizmu w sensie informatycznym, a na zastosowaniach. Większość rozdziałów sprowadza się do technicznego pokazania '*pipeline*' i szczegółów implementacyjnych uruchomionych mechanizmów. Rozdziały są w większości dokumentacją projektową niż właściwą rozprawą doktorską. W treści rozdziałów brakuje wskazania celu badań, zakresu, wskazań pytań badawczych czy dokładnej analizy wyników. Można też mieć poważne wątpliwości co do metodyki badań i porównań do innych mechanizmów referencyjnych, a tym samym trudno wskazać wkład autora rozprawy w rozwój informatyki.

2. Wkład autora

Oceniana praca doktorska zawiera głównie opisy użycia znanych technik do nowych zastosowań. Niestety także jakość tych opisów jest daleka od naukowych standardów. Aby prześledzić zawartość kolejnych rozdziałów:

- rozdziały 1-3 stanowią bardzo ogólne wprowadzenie do tematyki, bez definicji zagadnień. Dość rażące są braki wzorów, np. F1 (str.13) czy brak definicji formalnej wzorów dot. zagadnień logiki rozmytej (str.12).
- rozdział 4 dotyczy zastosowania uczenia głębokiego do wykrywania malarii. Brak formalnego opisu problemu, podano tylko ogólnie pojęcie istniejącej luki badawczej (str. 20), następnie opisano dane medyczne i ich schemat przetwarzania (rodz.4.2-4.4), a następnie przedstawiono architekturę i parametry zastosowanej sieci (rodz.4.5), a następnie uzyskane wyniki (4.6) i podsumowanie.
- Rozdział 5 zawiera opis zastosowania sieci CNN do obrazowania medycznego. Ten rozdział ma charakter bardziej naukowy ze względu na przegląd literatury (5.2) i większego wstępu teoretycznego (5.3), w dalszej części opisano zbiór i metody przetwarzania (5.4), aby w rozdziale 5.5. opisać parametry proponowanego modelu TSSN i procedurę strojenia (5.6). W pozostałej części przedstawiono wyniki (5.7), wraz z pół stronicowym rozdziałem dyskusji (5.8) i podsumowaniem (5.9).
- Rozdział 6 zawiera opis zastosowania sieci uczenia głębokiego do wykrywania broni na obrazach. Zarysowy opis problematyki wykrywania broni zawarto w rozdz. 6.1., a następnie przegląd literatury (6.2.), a następnie motywacje do utworzenia sztucznego zbioru danych (6.3). W rodz.6.4 przedstawiono sieci FCN ze szczegółami technicznymi, metodą uczenia (6.5.) i wynikami (6.6) oraz wskazaniem kierunków dalszych badań (6.7.) i krótkiego podsumowania (6.8).
- Treść rozdziału 7 dotyczący zastosowania logiki rozmytej reputacji klienta w uczeniu federacyjnym i gdyby był dopracowany, mógłby być bardzo wartościowym naukowo. W rozdziale 7.1. przedstawiono ogólnie zarys tematyki, architektur uczenia federacyjnego (7.2), a następnie naszkicowano zastosowany mechanizm (7.4), aby przejść do opisu zbioru danych i wyników (7.5), a następnie podsumowania (7.6).
- Rozdział 8 zawiera opis zastosowania modeli uczenia głębokiego i augmentacji obrazów klasyfikacji tkanek przewodu pokarmowego. Dalej, wprowadzenie to tematyki (8.1) i przestawiono zbiór danych (8.2), jego przetwarzanie (8.3) oraz augmentacja (8.4). Rozdział 8.5 zawiera opis zastosowanej sieci ResNet wraz z parametrami metody uczenia (8.6). Protokoły ewaluacji wskazano w rodz.8.7 wraz z wynikami, a dalej dyskusja oraz uwagi końcowe (8.8 i 8.9).
- W rozdziale 9 opisano pół-nadzorowane transformery do hematologicznej segmentacji semantycznej. Wstęp do problematyki i przegląd literatury przedstawiono w rozdziałach 9.1-9.4. Opis (nie definicja) problemu stanowi rodz.9.5, a dalej przedstawiono szczegóły komplementacyjne (9.6), zarys protokołu ewaluacji / porównań (9.7 i 9.8) i podsumowanie (9.9).
- Dwa ostatnie rozdziały (10 i 11) stanowią podsumowanie pracy.

Praca nie jest dobrze rozplanowana merytorycznie i redakcyjnie. W mojej ocenie, z punktu widzenia informatyki **nie wskazano nowości**, pokazano tylko przegląd zastosowań, elementów technicznych użycia poszczególnych modeli sieci głębokich z różnymi mechanizmami preprocessingu. W mojej ocenie nie stanowi to znaczącego wkładu do informatyki.

Badania przeprowadzone w ramach doktoratu nie są dopracowane i metodologicznie można mieć zastrzeżenie dotyczący **reprodukowalności badań**. Opisy użytych modeli są ogólne i trudno odtworzyć wyniki tych prac. Dodatkowo, w żadnym z rozdziałów nie znalazłem informacji o udostępnienia kodu, wyników czy też zbiorów danych. Ten ostatni argument jest szczególnie ważny dla danych medycznych (choć rozumiem, że mogą być objęte osobnymi umowami) i wyjątkowo

pozwolę się zapytać o publikację utworzonego sztucznie zbioru z rozdziału 6, którego jak rozumiem, nie wiążą żadne prawa autorskie?

Cel naukowy opisywanych badań w rozdziałach niestety umyka, a główny nacisk w rozprawie kładziony jest na zastosowanie. Nie jest to wadą samo w sobie – informatyka jest nauką służebną – ale przy rozprawie doktorskiej w dziedzinie informatyki główny nacisk powinien być położony na rozwój dziedziny. Ten trudno mi tutaj wskazać.

3. Poprawność

W rozprawie doktorskiej trudno wskazać elementy naukowe, gdyż praca jest bardzo niespójna strukturalnie i redakcyjne. Każdy z rozdziałów nie tylko dotyczy innego zastosowania sieci głębokich, ale także niespójności metodologii badań i warsztatu naukowego. Warto wskazać, że stosując modele uczenia maszynowego, pracujemy z losowością na wielu poziomach: danych (tj. podział), trudnością inicjalizacji (tj. różne *seedy*) a także z samym modelem (tj. strojenie modelu). Stąd, aby zbadać model uczenia maszynowego i określić jakość nauczonego modelu, wszystkie te aspekty musimy zaadresować. Inaczej nie możemy mówić o jakości modelu i generalizacji wiedzy. Dlatego, w protokole badawczym powinniśmy zbadać wielokrotne podziały danych, wielokrotne uruchomienie modelu (dla różnych *seedów*) i jego strojenie. To tworzy wiele zestawów wyników, które później uśredniamy i porównujemy z modelami konkurencyjnymi czy referencyjnymi, w miarę możliwości wspomagając się testami statystycznymi dla potwierdzenia istotności wyników. Pozwolę sobie tutaj wskazać wytyczne, jakie stawia prestiżowa konferencja NeurIPS (A*, MeiN:200pkt) dla nadsyłanych artykułów naukowych[1] z dziedziny uczenia maszynowego. Szczególnie polecam lekturze punkty 5 (o otwartym dostępie i reprodukowalności kodu), 6 (plan eksperymentów z podziałem danych, hiperparametrów) oraz punkt 7 (istotność statystyczna). Także, podkreślając znaczenie wielu uruchomień metod oraz wpływu ziarna (ang. *seed*) na badania, pozwolę sobie wskazać publikację D.Picarda [1]. Ponadto, w publikacji [2] (publikowanej na NeurIPS w 2021) wskazywano jak mogą być różne wyniki modelu, nawet dla tych samych ziaren czy różnych narzędzi implementacji. Jako podsumowanie chciałbym wskazać publikację [3], która podsumowuje gdzie występuje losowość w aktywnym uczeniu głębokim, ale z powodzeniem można to uogólnić.

W recenzowanej pracy doktorskiej, jeśli chodzi o podział danych, w początkowym rozdziale 2.8 autor opisuje walidację krzyżową, jednak w każdym rozdziale inaczej ją stosuje (albo pomija ten fakt). Dla przykładu, w rozdziale 4.6 mamy bez uzasadnienia podany podział danych 70%-15%-15%, w rozdziale 5.7. mamy informacje o zbiorze 36 obrazów, następnie w rozdziale 6.6 mamy zbiór walidujący 1000 masek-obraz, w rozdziale 7.5 nie udało się znaleźć informacji o sposobie podziału danych w procesie uczenia. Dalej, w rozdziale 8.7 mamy podział 85%/15% danych ze stratyfikacją (i „*single random seed*"), a dalej wspomniano o użyciu 5-foldowej krzyżowej walidacji stratyfikowanej. W rozdziale 9.7 opisujących protokół ewaluacji trudno znaleźć precyzyjny opis podziału danych.

W przypadku inicjalizacji modeli użytych sieci głębokich autor wspomina o tym tylko w wybranych rozdziałach o zastosowanych metodach (np.8.6). Warto jako zaletę pracy wskazać element strojenia modeli – ten element jest obecny w wielu rozdziałach (np. 5.6 *bayesian optimisation* czy 6.5. *Adam*). Niestety, finalne wyniki we wszystkich rozdziałach są podawane w tabelach, niestety nie jako średnie i bez odchyleń standardowych. Szczególnie to zastanawia, jeśli prezentowane modele autora mają skuteczność na poziomie Acc=99,1% a następna metoda ma Acc=97,78% – patrz Tab.9 (str.59). Także przy okazji danych w tej tabeli, nie rozumiem jak modele zostały porównane, bo autor (3 akapit, str.59) wskazuje trudność w porównaniu z danymi z publikacji (ten sam protokół i zbiór badawczy?). Także, dla przykładu trudno mi zrozumieć wyniki w Tab.21 (str.87), gdzie autor

1 https://neurips.cc/public/guides/PaperChecklist

porównuje wyniki swojego modelu tylko dla Acc dla 4 gatunków zwierząt do modeli referencyjnych uczonych na danych dla ludzi (?).

Z punktu widzenia metodologicznego wyniki porównywane powinny być przy użyciu jednego zestawu miar (np. w pracy czasem jest to Acc, innym razem F1), wraz z podaniem średnich, odchyleń standardowych, ale przede wszystkim wielu powtórzeń i użyciem testów statystycznych. W niniejszej rozprawie wspomniano „*statistical results*" w rozdz.10.6, a użyto tylko raz (tj. test Kołmogorowa-Smirnowa) w rozdz. 8.4 przy okazji badania wpływu augmentacji danych.

Warto wskazać zaletę pracy, tj. analiza ablacji, co jest cenne. Niestety wskazywane różnice jakości modeli często są na poziomie 1-2% (np. Tab.16, str.73 czy Tab.20 str.85) i bez uśrednienia, analizy statystycznej i analizy wyników badań wnioski końcowe są pozbawione wartości naukowej.

Także rysunki są bardzo słabej jakości (np. 5, 16-20), bez pełniejszego opisu w tekście stanowią tylko dodatek. Rysunek 16 ze str. 116 ma bardzo duży potencjał, jednak pomimo usilnych starań nie udało się znaleźć jego opisu w treści pracy.

Literatura:

[1] Pickard David, Torch.manual_seed(3407) is all you need: On the influence of random seeds in deep learning architectures for computer vision, 2022, DOI: 10.48550/arXiv.2109.08203
[2] Qian, Shangshu and Pham, Viet Hung (i inni), Are My Deep Learning Systems Fair? An Empirical Study of Fixed-Seed Training, NEURIPS2021, pp.30211–30227.
[3] Yilin Ji, Daniel Kaestner, Oliver Wirth, Christian Wressnegger, Randomness Is the Root of All Evil: More Reliable Evaluation of Deep Active Learning; Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023, pp. 3943-3952

4. Wiedza kandydata

Autor pracy wydaje się obeznany z technicznym aspektem zastosowań głębokich sieci, o czym mogą świadczyć prezentowane w pracy przeglądy literatury i autorskie publikacje doktoranta. Niestety, poszczególne rozdziały doktoratu i publikacje są bezpośrednio skupione na zastosowaniach i trudno w rozprawie wskazać mocne elementy naukowe, które są znaczące w dziedzinie informatyki. Powodem może przyjęta forma rozprawy doktorskiej jako luźno powiązanych ze sobą rozdziałów, bez spójności na poziomie treści czy redakcji.

Każdy z rozdziałów sprowadza się zasadniczo do technicznego opisu „*pipeline*" lub „*frameworku*", luźnego pseudokodu bez opisów w tekście, z wieloma „magicznymi liczbami" (np. Alg.1., str.39 czy Alg.7. str.79). Nie są to opisy naukowe, nie są wyjaśnione ani naukowo umotywowane. Brak rygoru badań naukowych i liczne uchybienia warsztatowe na poziomie. Dla przykładu, proponując pewien algorytm, warto go opisać w treści pracy, oszacować złożoność, ale też i podać alternatywy dla porównań. Jest to ważne z punktu widzenia warsztatu informatyka, gdzie algorytmiczny opis zastosowanych metod powinien być bez zarzutu.

Wartość merytoryczna pracy też bardzo mocno cierpi przy przekazie poprzez niedokładną redakcję i brak spójności treści. Na to nakładają się wskazane błędy metodologiczne i poważne usterki warsztatowe. Sprawia to, że niestety trudno zaakceptować rozprawę doktorską w tej formie jako wystarczającą do nadania stopnia naukowego.

5. Podsumowanie

Biorąc pod uwagę argumenty przedstawione w powyższych punktach recenzji i wymagania zdefiniowane przez art.187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (z późniejszymi zmianami[2]) moja recenzja pod względem trzech podstawowych kryteriów dotyczy:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego? (wybierz jedną opcję stawiając znak *X*)

☐	☐	☒	☐	☐
Zdecydowanie TAK	*Raczej TAK*	*Trudno powiedzieć*	*Raczej NIE*	*Zdecydowanie NIE*

B. Czy po przeczytaniu rozprawy zgadzasz się, że kantydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka techniczna i telekomunikacja?

☐	☐	☒	☐	☐
Zdecydowanie TAK	*Raczej TAK*	*Trudno powiedzieć*	*Raczej NIE*	*Zdecydowanie NIE*

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

☐	☐	☐	☒	☐
Zdecydowanie TAK	*Raczej TAK*	*Trudno powiedzieć*	*Raczej NIE*	*Zdecydowanie NIE*

Uwzględniając powyższe argumenty i sformułowne oceny stwierdzam, że rozprawa doktorska mgra inż. Michała Wieczorka pt. „*Semi-Supervised Training with Fuzzy Logic for Transformer-Based Semantic Segmentation*” nie spełnia wymagań art.187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (z późniejszymi zmianami) i tym samym nie stanowi podstawy do nadania jej Autorowi stopnia naukowego doktora nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja. A tym samym przedstawiam <u>negatywny</u> wniosek w sprawie dopuszczenia do publicznej obrony.

Paweł B. Myszkowski
Politechnika Wrocławska

2 https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190000276