

Łódź, 30 stycznia 2024 r.

prof. dr hab. inż. Michał Strzelecki
Instytut Elektroniki Politechniki Łódzkiej
ul. Wólczańska 211/215, 90-924 Łódź
michal.strzelecki@p.lodz.pl

Recenzja rozprawy doktorskiej mgr inż. Aleksandry Osowskiej-Kurczab
**„Uczenie maszynowe w zastosowaniu do różnicowania nowotworów nerek
na podstawie obrazowania tomograficznego”**
promotor prof. dr hab. inż. Tomasz Markiewicz

Podstawą niniejszej recenzji jest uchwała Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej z 8 grudnia 2023 r. powołująca mnie na recenzenta w postępowaniu o nadanie stopnia doktora mgr inż. Aleksandrze Osowskiej-Kurczab, prowadzonym w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja.

Dynamiczny rozwój metod sztucznej inteligencji powoduje, że znajdują one zastosowania w coraz większej liczbie obszarów. Należy do nich również medycyna, gdzie metody te są od dawna używane m.in. do wspomagania diagnostyki, planowania terapii, monitorowania jej skutków czy predykcji rokowań procesu terapeutycznego. Szczególnie rośnie znaczenie algorytmów AI opracowywanych do analizy obrazów medycznych, których celem jest wsparcie lekarzy podczas przeprowadzania diagnostyki obrazowej. Udział takich metod jest nieunikniony ze względu na rosnącą, w ostatnich latach, liczbę badań obrazowych pochodzących z różnych modalności przy jednocześnie utrzymującej się na zbliżonym poziomie liczbie lekarzy radiologów. Recenzowana rozprawa wpisuje się w ten obszar badawczy, ponieważ jej celem jest opracowanie systemu komputerowego umożliwiającego automatyczną klasyfikację wybranych zmian patologicznych nerek. Tematykę rozprawy uważam zatem za bardzo ważną i aktualną, ponieważ opracowane metody analizy obrazów, po weryfikacji w warunkach klinicznych, mogą znaleźć zastosowanie w praktyce medycznej.

Wybrany temat badawczy jest ambitny, gdyż system opracowany przez Autorkę rozprawy analizuje aż 7 różnych zmian chorobowych (większość znanych w literaturze rozwiązań uwzględnia mniejszą liczbę takich zmian), ponadto dokonała Ona skutecznej fuzji dwóch podejść uczenia maszynowego: klasycznego, wykorzystującego analizę tekstury obrazu oraz uczenia głębokiego testując różne rodzaje architektur sieci konwolucyjnych. W rozprawie sformułowano trzy tezy:

Teza 1: Możliwe jest stworzenie zautomatyzowanej metody pozwalającej na różnicowanie 7 typów guzów nerek z wysoką dokładnością jedynie na podstawie obrazowania CT z kontrastem. Pojedyncza faza tego badania jest wystarczającym źródłem informacji do stworzenia systemu opisu i klasyfikacji guzów nerek.

Teza 2: Połączenie różnych metod reprezentacji obrazów radiologicznych pozwala na poprawę dokładności klasyfikacji. W szczególności, połączenie metod neuronowej reprezentacji z operatorami teksturalnymi pozwala na uzyskanie wyników na poziomie aktualnego stanu wiedzy (ang. SOTA) ale przy znacznie większej liczbie reprezentowanych klas.

Teza 3: Metody przetwarzania obrazów, takie jak np. zmiana rozdzielczości, normalizacja czy ekstrakcja ROI mają znaczący wpływ na zdolności generalizacyjne systemów predykcyjnych w zadaniu analizy obrazowania medycznego.

Przedstawione tezy są sformułowane poprawnie. Pewne uwagi mam do tezy 3: badania dotyczące sposobu wyznaczania obszarów zainteresowania oraz ich normalizacji są już prowadzone od dawna. Zostało również wykazane (w szczególności w odniesieniu do analizy tekstury), że te dwa aspekty mają istotny wpływ na wyniki analiz szerokiej klasy obrazów medycznych. Teza ta została już, przynajmniej w części, udowodniona.

Rozprawa została podzielona na 7 rozdziałów. Pierwszy rozdział ma charakter wprowadzający w tematykę rozprawy, zawiera szczegółowo przedstawiony stan wiedzy, cel oraz tezy pracy. W rozdziale 2 określono problem badawczy, przedstawiono wybrane zagadnienia dotyczące diagnostyki nerek oraz opisano bazę danych wykorzystywanych w badaniach wraz z dogłębną analizą eksploracyjną zbioru obrazów. Rozdział 3 omawia wybrane pojęcia oraz metody z zakresu przetwarzania i analizy obrazów, koncentrując się na metodach stosowanych w rozprawie, w tym metody interpolacji danych użyte do zmiany rozdzielczości analizowanych obrazów. Przedstawiono tu również wybrane metody analizy tekstury obrazu oraz zestaw parametrów tekstury wykorzystywany w badaniach. W rozdziale 4 zaprezentowano klasyczne metody uczenia maszynowego. Omówiono także kilka wybranych metod selekcji cech oraz zagadnienia związane z problematyką uczenia głębokiego, w tym właściwości wykorzystywanych przez Autorkę rozprawy kilku architektur sieci splotowych. Przedstawiono również metody klasyfikacji zespołowej oraz metody walidacji klasyfikatorów jak i miary jakości stosowane do oceny jakości klasyfikacji. Rozdział omawia też metody redukcji wymiarowości przestrzeni danych wejściowych (PCA, TSNE, UMAP). Metody te jednak błędnie zostały zaklasyfikowane jako metody uczenia nienadzorowanego. Jednak te podejścia nie reprezentują algorytmów jakiegokolwiek uczenia, zapewniają jedynie zestaw parametrów uzyskanych z odpowiedniej transformacji cech wejściowych opisujących dane zjawisko. Dopiero taki zestaw parametrów może stanowić wejście dla algorytmów uczących (nadzorowanych lub nie).

Rozdział 5 przedstawia zastosowane w analizie obrazów modele uczenia maszynowego a także omawia wszystkie niezbędne kroki potrzebne do poprawnego użycia tych modeli. Opisano metody tworzenia zbiorów uczących i testowych, w tym obszarów zainteresowania oraz ich normalizacji. Przedstawiono schematy analizy obrazów z wykorzystaniem analizy tekstury oraz z użyciem wybranych architektur sieci głębokich. Zaprezentowano metody wyboru klasyfikatorów w zestawach jak i zasady agregacji wyników wygenerowanych przez poszczególne klasyfikatory. Pokazano również sposoby walidacji stosowanych modeli uczenia maszynowego. Należy stwierdzić, że wszystkie opisane tu metody zostały wybrane i zastosowane poprawnie z punktu widzenia uzyskania poprawnego rozwiązania zdefiniowanego w rozprawie problemu klasyfikacji obrazów.

W rozdziale 6 opisano uzyskane wyniki klasyfikacji obrazów dla wszystkich przeprowadzonych eksperymentów. Przeprowadzono ich bardzo dużo: obejmowały problem klasyfikacji zdefiniowany dla dwóch, trzech oraz wszystkich 7. klas patologii, przy czym grupowanie poszczególnych zmian chorobowych zostało dobrze uzasadnione bądź naturą tych zmian lub poziomem trudności ich klasyfikacji. W badaniach wykorzystano również wszystkie opisane w rozdziale 5 modele uczenia maszynowego. Uzyskane wyniki przedstawiono bardzo starannie, uwzględniając poza typowo prezentowanymi elementami oceny jakości klasyfikacji takimi jak wartości miar czy macierze pomyłek, także rozkłady parametrów tekstury po transformacji UMAP, listę cech tekstury odgrywających największą rolę w procesie klasyfikacji czy przykłady błędnych diagnoz wygenerowanych przez algorytmy. Te dodatkowe informacje są bardzo przydatne dla lepszego zrozumienia uzyskanych wyników i oceny użytych modeli. Otrzymane wyniki zostały starannie przedyskutowane również w odniesieniu do podobnych wyników opublikowanych w literaturze. Przekonujące jest także wyjaśnienie przewagi algorytmów stosujących analizę tekstury w porównaniu do metod uczenia głębokiego. Choć obecnie w piśmiennictwie dominuje stosowanie (często bezkrytyczne) algorytmów uczenia głębokiego, to Doktorantka wykazała, że w określonych okolicznościach tekstura skutecznie odwzorowuje właściwości obrazowanych narządów. W tym rozdziale przedstawiono również wyniki uzyskane dla zespołu klasyfikatorów, analizując wpływ doboru członków zespołu oraz metody agregacji ich wyników. Stosując to podejście, uzyskano poprawę wartości miar jakości klasyfikacji w zakresie 2%-5%, w zależności od realizowanego zadania. Ostatnia część rozdziału szóstego zawiera bardzo szeroką analizę ilościową oraz jakościową zespołowego modelu 7-klasowego zapewniającego najlepsze wyniki klasyfikacji. Z użyciem metody GradCam pokazano mapy aktywacji dla wybranej warstwy sieci konwolucyjnej. Szczególnie interesująca jest identyfikacja istotnych z punktu widzenia jakości klasyfikacji parametrów medycznych i populacyjnych obrazowanych pacjentów.

W rozdziale 7 podsumowano uzyskane wyniki, potwierdzono udowodnienie postawionych w rozprawie tez oraz wskazano kierunki dalszych badań. Praca zawiera również 6 dodatków. W pierwszych trzech omówiono wpływ hiperparametrów modeli na wyniki klasyfikacji w zadaniu 3-klasowym, przedstawiono wyniki badań odporności klasyfikatorów dla nieklasyfikowalnych przypadków nowotworów nerek i zawarto dodatkowe elementy analizy eksploracyjnej. Nieco zbędne wydają się kolejne dodatki (D, E, F), które zawierają nadmiarowe informacje biorąc pod uwagę bardzo dużą objętość rozprawy.

Liczący 264 pozycje wykaz literatury obejmuje wszystkie najważniejsze pozycje literatury światowej dotyczące tematyki związanej z rozprawą. Proponuję, aby Doktorantka zapoznała się z publikacją M. Kociołek, et al., Does image normalization and intensity resolution impact texture classification?, CMIG, 2020, <https://doi.org/10.1016/j.compmedimag.2020.101716>,

gdzie omówiono wpływ wstępnego przetwarzania obrazów na wyniki analizy ich tekstury. Wykaz publikacji zawiera również 4 publikacje współautorskie, zamieszczone w materiałach konferencyjnych (w tym jeden referat w materiałach bardzo dobrej konferencji IJCNN) oraz w czasopiśmie Bulletin of the Polish Academy of Sciences i Expert Systems with Applications.

Podsumowując merytoryczną ocenę rozprawy stwierdzam, że Doktorantka osiągnęła założone cele badawcze oraz udowodniła postawione tezy dzięki opracowaniu i walidacji nowatorskiego systemu do klasyfikacji wybranych guzów nerek w obrazach TK. Uważam, że jest to znaczące osiągnięcie badawcze, gdyż system integruje wiele metod przetwarzania i analizy obrazów. Doktorantka wykonała bardzo dużo eksperymentów analizując wpływ na działanie systemu różnych metod wstępnego przetwarzania obrazów, wyboru obszarów zainteresowania, rodzajów klasycznych klasyfikatorów, architektur sieci głębokich oraz ich hiperparametrów. Opracowała także skuteczne metody selekcji różnych typów algorytmów tworząc zespoły klasyfikatorów. System stanowi udaną fuzję klasycznych i głębokich metod uczenia maszynowego. W efekcie zostało wykazane, że połączenie obydwu podejść, a w szczególności uwzględnienie metod analizy tekstury obrazu, prowadzi do poprawy skuteczności klasyfikacji. Należy podkreślić, że dla osiągnięcia celów pracy mgr inż. Aleksandra Osowska-Kurczab musiała się wykazać głęboką wiedzą z dyscypliny informatyka techniczna i telekomunikacja a także umiejętnościami inżynierskimi umożliwiającymi implementację opracowanego systemu.

Praca generalnie nie budzi również zastrzeżeń od strony redakcyjnej. Chciałbym tylko prosić o wyjaśnienie pojęć „krzepkie metody reprezentacji danych” i „krzepkie działania systemu” w kontekście użycia pojęcia „krzepkie”.

Lektura pracy nasuwa również kilka przedstawionych poniżej uwag dyskusyjnych oraz sugestii dotyczących przyszłych badań.

1. Jedynym zauważonym ograniczeniem rozprawy jest wykorzystanie w badaniach dość ograniczonego zbioru obrazów pochodzących z jednego tomografu komputerowego. Czy Doktorantka planuje weryfikację opracowanego systemu dla większego zakresu danych, w szczególności pochodzących z wielu ośrodków medycznych?
2. W przypadku przyszłych badań dotyczących wpływu ROI na wyniki klasyfikacji sugeruję:
 - uwzględnienie większych rozmiarów kwadratowych lub prostokątnych okien ROI wyznaczanych w obszarze zmiany patologicznej. Obecna wielkość kwadratu (21 pikseli) jest na granicy możliwości pozyskania efektywnej informacji o teksturze.
 - zbadanie wpływu redukcji liczby bitów kodujących jasność piksela obrazu na wartości parametrów tekstury. Redukcja do np. 5 bitów prowadzi często do skuteczniejszej klasyfikacji tekstury ze względu na częściową eliminację szumów, co szczególnie jest zauważalne dla parametrów wyznaczanych dla macierzy GLCM.
3. Czy analiza PCA była wykorzystywana w opracowanym systemie? Można rozważyć użycie tej metody do dalszej redukcji liczby parametrów wejściowych dla algorytmów ML, niezależnie od stosowanej selekcji cech.

4. Proszę o wyjaśnienie, w jaki sposób uzyskano rozkłady cech na rys. 2.26 i 2.27. W tekście jest mowa o redukcji wymiarowości zastosowanej dla „surowych wartościach pikseli”, jednak nie opisano, na czym ten proces polega.

Wszystkie moje uwagi nie wpływają na jednoznacznie pozytywną ocenę recenzowanej pracy. Stwierdzam, że praca „Uczenie maszynowe w zastosowaniu do różnicowania nowotworów nerek na podstawie obrazowania tomograficznego” spełnia wymagania stawiane rozprawom doktorskim zgodnie z ustawą Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 r. z późniejszymi zmianami. W związku z powyższym wnioskuję o przyjęcie tej rozprawy i dopuszczenie mgr inż. Aleksandry Osowskiej-Kurczab do publicznej obrony.

Jednocześnie uważam, że opracowany w pracy system klasyfikacji obrazów cechuje się wysokim poziomem merytorycznym oraz dużą oryginalnością. Osiągnięcie Doktorantki stanowi zatem istotny wkład do dyscypliny informatyka techniczna i telekomunikacja. Rozwiązanie to ma charakter ogólny i opracowana metodyka klasyfikacji guzów nerek może mieć zastosowania również do analizy innej klasy obrazów biomedycznych pochodzących z różnych modalności. Należy również podkreślić dojrzałość Doktorantki w prowadzeniu badań naukowych oraz przy analizie i ocenie uzyskanych wyników. W związku z powyższym wnioskuję o wyróżnienie rozprawy mgr inż. Aleksandry Osowskiej-Kurczab.

Michał Stabek