

dr hab. inż. Damian Mazur

Rzeszów, 23 sierpnia 2024

Wydział Elektrotechniki i Informatyki

Politechnika Rzeszowska

ul. Powstańców Warszawy 12, 35-959 Rzeszów

Recenzja rozprawy doktorskiej kpt. mgr. inż. Fabiana Gila pt. „Zespoły sieci głębokich w rozpoznawaniu wybranych klas obrazów medycznych”.

Recenzja napisana została na podstawie Uchwały Rady Dyscypliny Naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Wojskowej Akademii Technicznej z dnia 19 czerwca 2024 r. w sprawie wyznaczenia recenzentów w postępowaniu o nadanie stopnia doktora kpt. mgr. inż. Fabianowi Gilowi. Recenzja przygotowana została na podstawie Ustawy z dnia 20 lipca 2018r. Prawo o szkolnictwie wyższym i nauce (Dz.U. 2023r. poz. 742 z późn.zm.). Rozprawa kpt. mgra inż. Fabiana Gila została napisana pod kierunkiem promotora prof. dra. hab. inż. Stanisława Osowskiego.

2. Zakres tematyczny rozprawy

Praca koncentruje się na opracowaniu automatycznych metod analizy obrazów medycznych czerniaka oraz raka piersi. W obu przypadkach podstawą jest obraz medyczny, przy czym sposoby ich akwizycji są różne. W przypadku czerniaka dotyczy zmian skórnych i jest uzyskiwany przy pomocy urządzenia zwanego dermatoskopem. W przypadku raka piersi (sutka) analizie podlega obraz mammograficzny, uzyskany w technice rentgenowskiej (X-ray). W obu przypadkach szybkie wykrycie zmian chorobowych pozwala na wczesne rozpoczęcie leczenia i uzyskanie skuteczniejszych wyników kuracji. Ważne więc jest opracowanie efektywnych metod komputerowych wspomagających diagnostykę lekarską umożliwiających jak najwcześniejsze wykrycie zmian patologicznych.

Tematyka pracy aktualna i praktyczna. Wspomaganie diagnostyki medycznej metodami sztucznej inteligencji jest niezwykle interesującym zagadnieniem, odpowiadającym na potrzeby społeczne i wykorzystującym nowoczesne metody informatyczne. W pracy przedstawiono wykorzystanie kilku rodzajów klasyfikatorów oraz zespołów klasyfikatorów w celu rozpoznania na podstawie obrazu komputerowego stanów patologicznych wywołanych przez czerniaka oraz raka piersi.

Cel pracy sformułowany przez Autora ukierunkowany został na „*opracowanie efektywnych metod tworzenia zespołów klasyfikacyjnych oraz sposobu fuzji pojedynczych wyników jego członków w werdykt końcowy.*” Na podstawie wyników przeprowadzonych eksperymentów numerycznych Autor sformułował tezę badawczą: „*Zastosowanie odpowiednio zdefiniowanego zespołu klasyfikatorów*

głębokich CNN pozwala istotnie polepszyć zdolności generalizacji systemu klasyfikacyjnego i poprawić wynik rozpoznania klasy przypisanej różnym zmianom nowotworowym w stosunku do rezultatów osiągniętych przez indywidualne klasyfikatory.”

Sieci konwolucyjne (CNN) są powszechnie stosowane w analizie obrazów medycznych, w tym do diagnozowania czerniaka, jednego z najgroźniejszych nowotworów skóry. Dzięki swojej zdolności do automatycznego wykrywania i ekstrakcji cech z obrazów, CNN mogą odgrywać kluczową rolę w identyfikacji zmian skórnych, które mogą być złośliwe. Sieci CNN uczą się rozpoznawać charakterystyczne cechy czerniaka, takie jak asymetria, nierówne brzegi, różnorodność kolorów oraz średnicę zmiany. Proces trenowania polega na iteracyjnym dopasowywaniu wag sieci na podstawie porównania przewidywań z rzeczywistymi etykietami danych.

Uzyskane wyniki badań oraz wybrane narzędzia badawcze wykorzystane przez Doktoranta w rozprawie doktorskiej świadczą o istotnym wpływie w rozwój dyscypliny naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne oraz o interdyscyplinarności zagadnień poruszanych w rozprawie.

3. Struktura rozprawy doktorskiej

Praca składa się ze 109 stron, w tym streszczenia w języku polskim i angielskim, spisu treści, 7 rozdziałów, 72 pozycji z literatury, spisu tabel i rysunków. Doktorant opublikował 4 artykuły współautorskie z czego 2 są artykułami konferencyjnymi, a pozostałe 2 w czasopismach międzynarodowych.

Na treść pracy składa się siedem rozdziałów:

Rozdział 1 stanowi wstęp, zawierający wprowadzenie w tematykę rozprawy, przegląd literatury światowej związanej z tematyką oraz przedstawia główne cele pracy i proponowaną tezę.

Rozdział 2 zawiera opis oraz analizę statystyczną obrazów medycznych czerniaka i raka piersi, które zostały wykorzystane w eksperymentach numerycznych przeprowadzonych w ramach pracy. Analiza obejmuje dwie bazy danych czerniaka oraz jedną bazę DDSM raka piersi.

Rozdział 3 skupia się na opisie wykorzystanych w pracy metod klasyfikacji obrazów, takich jak konwolucyjne sieci neuronowe (ang. Convolutional Neural Networks – CNN), drzewa decyzyjne (ang. Random Forest – RF) i maszyny wektorów nośnych (ang. Support Vector Machine - SVM). Opisuje również różne metody selekcji cech diagnostycznych, zastosowane w analizie danych przy wykorzystaniu różnych klasycznych klasyfikatorów.

Rozdział 4 zawiera opis koncepcji tworzenia zespołów klasyfikatorów oraz metod ich integracji. Zaproponowano dwa podejścia do budowy zespołów klasyfikatorów: w pierwszym oprócz klasyfikatorów CNN zastosowano również klasyfikatory RF oraz SVM zasilane wyselekcjonowanymi cechami, w drugim zaś wykorzystano tylko klasyfikatory CNN. Przedstawia również różne metody oceny jakości wyników klasyfikacji, użyteczne w porównywaniu skuteczności różnych analizowanych rozwiązań.

Rozdział 5 koncentruje się na wynikach badań eksperymentalnych dotyczących rozpoznawania obrazów zmian skórnych związanych z czerniakiem na dwóch zbiorach danych: małym zbiorze NIO stworzonym w polskim Instytucie Onkologii oraz międzynarodowym zbiorze ISIC. Przedstawia wyniki klasyfikacji oraz analizę skuteczności różnych rozwiązań.

Rozdział 6 zawiera wyniki badań eksperymentalnych w rozpoznawaniu raka piersi na zbiorze DDSM. Przedstawiono rezultaty badań dotyczące trzech zadań klasyfikacyjnych: rozróżnienia raka złośliwego od łagodnego z pominięciem stanu normalnego, rozpoznanie zmian nowotworowych, w których rak złośliwy i łagodny stanowią jedną klasę a stan normalny drugą klasę oraz rozpoznania trzech klas: rak złośliwy, rak łagodny, stan normalny.

Ostatni rozdział zawiera podsumowanie wyników całej rozprawy, wnioski płynące z przeprowadzonych badań oraz sugestie dotyczące kierunków dalszego rozwoju w dziedzinie rozpoznawania obrazów medycznych przy użyciu metod uczenia maszynowego związanego z uczeniem głębokim.

Struktura pracy prawidłowa i czytelna.

Przedstawiony we wstępie przegląd literatury wskazuje, iż Autor dogłębnie zapoznał się z aktualnymi metodami i trendami badawczymi związanymi z rozpatrywanymi w pracy zagadnieniami. W rozprawie przytacza wyniki 17 prac z lat 2016-2023 omawiających badania analogiczne do przedstawionych w pracy, podając uzyskane przez Autorów rezultaty oraz w większości przypadków wykorzystane przez nich metody. Nieco brakuje odniesienia się do wyników uzyskanych w poszczególnych pracach, w szczególności, jeżeli zważymy na duży rozrzut uzyskiwanych wyników.

Do badań Autor wykorzystuje dane z trzech źródeł: do badań na czerniakiem, z bazy danych utworzonej przez Zespół Kliniki Nowotworów Tkanek Miękkich, Kości i Czerniaków z Narodowego Instytutu Onkologii w Warszawie (NIO) i z ogólnodostępnego repozytorium zbioru utworzonego w ramach The International Skin Imaging Collaboration (ISIC) oraz do badań nad rakiem piersi, z bazy danych Digital Database for Screening Mammography (DDSM). Na wstępie dane z repozytoriów Autor poddał analizie statystycznej wyznaczając: średnią, odchylenie standardowe, energię, kurtozę oraz współczynnik

skońności dla obrazów z poszczególnych klas. Zdaniem Autora parametry statystyczne poszczególnych klas zarówno w przypadku czerniaka jak i raka piersi są bardzo podobne, a różnice mało istotne. Oznacza to, iż zadanie klasyfikacji jest problemem nietrywialnym.

Badania przedstawione w pracy obejmowały:

- a. diagnostykę dwuwartościową czerniaka,
- b. diagnostykę dwuwartościową raka piersi,
- c. diagnostykę trójwartościową raka piersi (rak złośliwy, rak łagodny, brak)/

W celu wybrania optymalnego układu klasyfikacji Autor przeanalizował pracę kilkunastu klasyfikatorów jednorodnych oraz kilkunastu zespołów klasyfikatorów. Bardzo duży zakres wykonanych badań jest wart podkreślenia za względu na aspekt merytoryczny oraz czasowy.

Spośród klasyfikatorów jednorodnych Autor przeanalizował pracę: 19 typów spłotowych sieci głębokich (Convolutional Neural Network - CNN), metody wektorów nośnych (Support Vector Machine – SVM), lasów losowych (Random Forest, RF).

W celu wybrania optymalnego zespołu klasyfikatorów Autor zbudował kilkanaście zespołów klasyfikatorów. Po badaniach zrealizowanych na zbiorze NIO zrezygnował z zespołów w skład, których wchodziły RF oraz SVM pozostawiając jedynie zespoły złożone z sieci CNN.

- a. w przypadku diagnostyki czerniaka 11 zespołów
- b. w przypadku diagnostyki raka piersi 12 zespołów zbudowanych z sieci CNN.

Autor wnikliwie przeprowadził i przedstawił badania mających na celu wybranie najlepszego podejścia do dostrajania sieci CNN do postawionych zadań.

Uzyskane przez Autora wyniki są wręcz rewelacyjne. Zarówno w przypadku czerniaka jak i raka piersi przewyższają najlepsze wyniki podawane w dostępnej literaturze światowej dla badań wykonanych na tej danych zaczerpniętych z tego samego repozytorium.

- a) W przypadku czerniaka dla badań zrealizowanych w oparciu o dane z bazy ISIC Autor uzyskał: dokładność ACC = 96.54, czułość TPR = 94.71%, specyficzność TNR = 97.67%. Najlepsze znane wyniki dla tego zbioru to ACC = 90.4%, czułość TPR = 82% i specyficzność TNR = 92.5%.

- b) W przypadku raka piersi dla badań zrealizowanych w oparciu o dane z bazy DDSM Autor uzyskał: ACC = 98.77%, TPR = 97.89%, TNR = 99.07%, AUC = 0.9975. Najlepsze znane wyniki dla tego zbioru to ACC = 89.71%, TPR = 93.54%, TNR = 80.58 i AUC = 0.9410

Bibliografia została dobrana starannie, odzwierciedla aktualny stan i badań prezentowanych w literaturze światowej z zakresu objętego niniejszą rozprawą oraz odpowiada zagadnieniom poruszonym w poszczególnych rozdziałach rozprawy doktorskiej.

4. Uwagi edytorskie i merytoryczne

- Używanie skrótów powinno być poprzedzone wyjaśnieniem ich znaczenia. Autor niestety nie stosuje się do tej zasady – przykładowo str. 9: F1, AUC. Wykorzystanie wskaźników powinno być poprzedzone ich zdefiniowaniem. Autor niestety nie stosuje się do tej zasady – przykładowo str. 9: dokładność, F1, AUC, czułość i inne na kolejnych stronach.
- Str 44: Podejście rozmyte – klasyfikator generuje wartości prawdopodobieństwa przynależności do każdej klasy, przy czym wartości te są z przedziału [0, 1]. Sumowaniu w każdej klasie podlegają wartości tych prawdopodobieństw – Jakiego rzędu mogą być wyznaczone sumy? Czy określenie „prawdopodobieństwo” jest zgodne z definicją matematyczną?
- W tabelach 2.1 – 2.5 Autor wykorzystuje parametry statystyczne obrazów nigdzie nie definiując ich.
- Wzór 3.1 jest niepoprawny – nie zwraca wartości funkcji dla argumentu $x = 0$. Powinno być: $\text{ReLU}(x) = \max(0, x)$.
- Stosowanie niejednolitego zapisu np. e/\exp ; $\sum_k \text{wyr} / \sum_{j=1}^M \text{wyr}$
- Warto było zrobić wykaz oznaczeń stosowanych w pracy. Bo raz znaczenie symboli jest wyjaśniane (np.: wzory 3.18, 3.19) gdzie indziej nie (np.: we wzorach 3.1, 3.2). A niektóre symbole używane są dla oznaczenia różnych wielkości np.: x
- Rys. 3.3, rys. 3.4. Opisy na rysunku w języku angielskim. Brak odwołania do źródła.
- Rys. 3.7. Nie opisano osi rysunku. Co reprezentują osie wykresów?
- Wzór 3.21 jest nieporwany, powinno być: $x_i = \frac{x_i - x_{i_min}}{x_{i_max} - x_{i_min}}$
- Rys. 3.8 Nie opisano osi rzędnych. Co reprezentuje oś rzędnych? Brak ten występuje jeszcze na kilku rysunkach.
- W rozdziale 4.2 zdefiniowano wskaźniki służące do oceny klasyfikacji binarnej. Zdefiniowanymi wskaźnikami Autor posługuje się w rozdziale 5 (Wyniki badań eksperymentalnych w

rozpoznawaniu czerniaka) oraz w rozdziale 6 (Wyniki badań eksperymentalnych w rozpoznawaniu raka piersi). Jednak część badań przedstawionych w rozdziale 6 (podrozdział 6.4) ma charakter klasyfikacji wieloklasowej. Autor, bardzo enigmatycznie wyjaśnił różnice w wyznaczaniu wskaźników jakości klasyfikacji.

Wykonano bardzo dużo badań, co wskazuje, iż Autor sprawnie porusza się w zagadnieniach klasyfikacji metodami sztucznej inteligencji oraz sprawnie operuje głębokimi sieciami neuronowymi.

Uwagi merytoryczne:

- a) Czy były próby wykorzystania systemu do wspomagania diagnostyki czerniaka bądź raka skóry w placówkach medycznych?
- b) We wstępie Autor pisze: „Celem pracy jest opracowanie efektywnych metod tworzenia takich zespołów klasyfikacyjnych oraz sposobu fuzji pojedynczych wyników jego członków w werdykt końcowy.” Zaś w podsumowaniu „Celem niniejszej rozprawy doktorskiej było opracowanie systemu zespołowego klasyfikatorów oraz zbadanie jego skuteczności w procesie rozpoznawania obrazów medycznych czerniaka oraz raka piersi”. Co więc było celem opracowanie metod tworzenia zespołów klasyfikacyjnych czy opracowanie zespołów klasyfikatorów.
- c) Czy było przeprowadzone wstępne przetwarzanie danych np.: w celu eliminacji danych wadliwych.
- d) Dlaczego wśród głównych osiągnięć pracy Autor nie wymienia uzyskanych rezultatów? Konkretnie: skonstruowania systemu klasyfikacyjnego pozwalającego skutecznie diagnozować czerniaka i raka skóry.
- e) Praca charakteryzuje się nierównomiernym rozłożeniem akcentów służących charakterystyce omawianych zagadnień. Niektóre z nich Autor omawia bardzo dokładnie, dogłębnie analizując wszelkie zawiłości np.: sieci konwolucyjne i ich dostrajanie, inne traktuje w sposób krótki np.: drzewa losowe, maszyny wektorów wspierających, a inne tylko sygnalizuje np.: metody testowania. Warto było w pracy metodom testowania poświęcić nieco więcej miejsca, gdyż sposób testowania ma bardzo duży wpływ na uzyskiwane wyniki, i jest elementem niezbędnym aby można było porównywać rezultaty przedstawiane w różnych pracach.
- f) W podsumowaniu brakuje wyraźnego odniesienia się do celu pracy i tezy postawianych we Wstępie (podrozdział 1.3). Pozostaje wrażenie, że celem nadrzędnym jest uzyskanie jak najlepszych wyników diagnostycznych, ale w kontekście zapisów we Wstępie jest to cel pośredni. Zasadniczym celem było „opracowanie efektywnych metod tworzenia (...) zespołów

klasyfikacyjnych oraz sposobu fuzji pojedynczych wyników jego członków w werdykt końcowy” zaś teza wskazywała, że zastosowanie „(...) zespołu klasyfikatorów głębokich CNN pozwala istotnie polepszyć zdolności generalizacji systemu klasyfikacyjnego i poprawić wynik rozpoznania klasy (...).

5. Wnioski końcowe

Doktorant wykazał się dużą wiedzą z zakresu opisanych w rozprawie prac badawczych, postawioną prawidłową tezę oraz cele udowodnił.

Doktorant kpt. mgr. inż. Fabiana Gila w rozprawie doktorskiej zaprezentował ważne wyniki badań z zakresu zastosowania sieci głębokich w rozpoznawaniu wybranych klas obrazów medycznych. Jest to oryginalne rozwiązanie problemu naukowego i stanowi istotny wkład w dyscyplinę naukową Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne oraz spełnia wymogi stawiane rozprawom doktorskim w Ustawie z dnia 20 lipca 2018r. - Prawo o szkolnictwie wyższym i nauce (Dz. U. 2023r.poz. 742 z póź. zm.).

Uważam, że cele założone w pracy zostały spełnione a hipoteza potwierdzona.

Sformułowane w recenzji uwagi mają w większości charakter dyskusyjny i nie umniejszają w żaden sposób wartości opracowanej metody ani wysokiemu poziomowi badawczemu rozprawy.

Wniosuję do Rady Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne o dopuszczenie Pana mgra inż. Fabiana Gila do dalszych etapów przewodu doktorskiego, w tym publicznej obrony.

Damian Maru

