

dr hab. Jan Bazan, prof. UR  
Instytut Informatyki  
Uniwersytet Rzeszowski,  
ul. Pigonia 1, 35-310 Rzeszów  
jbazan@ur.edu.pl

20 marca 2024 r.

Recenzja rozprawy doktorskiej  
*mgr. inż. Tomasza Krzywickiego*  
zatytułowanej:

*Metody rozpoznawania obrazów w badaniu dna oka dla celów diagnostyki medycznej*

**Promotor: prof. dr hab. Lech Polkowski.**

**Dziedzina:** nauki techniczne, **Dyscyplina:** informatyka techniczna i telekomunikacja

## **Cel i zakres rozprawy**

Rozprawa doktorska dotyczy metod lokalizowania położenia dołeczka oraz tarczy nerwu wzrokowego na zdjęciach dna oka, a zatem wpisuje się w nurt badań związany z metodami eksploracji danych obrazowych, który już od kilkudziesięciu lat znajduje się w centrum badań wielu silnych ośrodków badawczych na świecie. Tak duże zainteresowanie tym nurtem wynika z dużej dostępności coraz większych zbiorów danych obrazowych oraz pilnej potrzeby eksploracji tych danych. Postęp w tym zakresie jest niezwykle istotny m.in. dla rozwoju systemów inteligentnych bazujących na wspomaganiu podejmowania decyzji. Dlatego opracowywanie metod eksploracji danych obrazowych i wykazanie skuteczności tych metod w różnych zastosowaniach ma wciąż kluczowe znaczenie dla dalszego rozwoju systemów inteligentnych w wielu dziedzinach, jak choćby w medycynie, przemyśle, transporcie i innych.

Fundoskopia to badanie okulistyczne, którego celem jest obrazowanie dna oka. Dzięki temu obrazowaniu, można zobaczyć m.in. plamkę żółtą, której zadaniem jest zbieranie nadmiaru bodźców świetlnych. Leży ona w centrum siatkówki składającej się z wielu warstw komórek światłoczułych i neuronów przekazujących sygnały wizualne do mózgu. W centrum plamki żółtej znajduje się dołeczek, a w jej okolicach – tarcza nerwu wzrokowego, czyli miejsce, w którym nerw wzrokowy wychodzi z siatkówki. Na zdjęciu dna oka można również zobaczyć struktury naczyniowe, takie jak żyły i tętnice, które odpowiedzialne są za unaczynienie siatkówki. Posługując się dedykowanymi narzędziami, specjalista może ocenić stan krytycznych składowych tylnego odcinka oka: plamki żółtej, tarczy nerwu wzrokowego, siatkówki oraz naczyń krwionośnych.

Dzięki wykorzystaniu metod sztucznej inteligencji do analizy zdjęć dna oka, możliwe jest wykrywanie charakterystycznych zmian stanu oka, co pozwala na szybkie wdrożenie odpowiednich procedur klinicznych.

Jednym z kluczowych elementów analizy zdjęć dna oka jest odpowiednia lokalizacja dołeczka oraz tarczy nerwu wzrokowego, które są istotne w diagnostyce okulistycznych jednostek chorobowych. Lokalizacja dołeczka na zdjęciu dna oka umożliwia dokonanie analizy obszaru plamki żółtej i najbliższego otoczenia, której celem jest wykrywanie zmian, takich jak obrzęk czy uszkodzenia, które mogą być związane z różnymi chorobami siatkówki. Ponadto, lokalizacja struktur anatomicznych dna oka, takich jak tarcza nerwu wzrokowego oraz dołeczek, umożliwiają ocenę jakości i standardów wykonania zdjęć dna oka, które odgrywają istotną rolę w zapewnieniu dokładnej diagnostyki klinicznej.

Wyniki badań przedstawione w rozprawie mają zatem duże znaczenie dla okulistyki i problematyka tych badań jest niebanalna. Dlatego uważam, że podjęta tematyka może z powodzeniem stanowić przedmiot rozprawy doktorskiej.

## **Zawartość rozprawy**

Rozprawa doktorska została napisana w języku polskim i składa się z 7 rozdziałów i 4 dodatków. Rozdział 1 to wprowadzenie obejmujące ogólne przedstawienie motywacji, problemu badawczego, celu i tezy rozprawy, głównych wyników rozprawy, zastosowanych narzędzi i danych oraz opisu struktury rozprawy. Rozdział 2 opisuje rolę sztucznej inteligencji do wsparcia diagnostyki okulistycznej. Rozdział 3 omawia wybrane pojęcia, metody i algorytmy przetwarzania i rozpoznawania obrazów stosowane w dalszej części rozprawy. W szczególności, wprowadzono pojęcie złożoności obliczeniowej algorytmów, ideę obrazu cyfrowego i modeli barw, a także wybrane metody przetwarzania obrazów, takie jak filtrowanie obrazów, binaryzacja, wyrównywanie histogramów, wyznaczanie wypukłych powłok konturów wielokątów oraz analiza komponentów połączonych. Następnie przedstawiono algorytmy topologiczne i algorytmy inspirowane naturą, takie jak morfologia matematyczna oraz sieci neuronowe. Wreszcie, omówiono nowoczesne metody uczenia maszynowego, uczenia głębokiego i rozpoznawania obrazu, takie jak uczenie transferowe, uczenie zespołowe i detekcja obiektów oraz wybrane metody przyspieszania algorytmów. Rozdział 4 przedstawia dwa zbiory danych zawierających zdjęcia dna oka, które wykorzystano w procesie tworzenia proponowanych metod lokalizujących dołeczki i tarczę nerwu wzrokowego na zdjęciach dna oka. Rozdział 5 poświęcono zaprezentowaniu i szczegółowemu omówieniu proponowanych metod lokalizujących środki dołeczka oraz tarczy nerwu wzrokowego. Dla każdej z proponowanych metod przedstawiono motywację utworzenia oraz szczegółowy opis działania wraz z przykładami działania metod na przykładowych obrazach. Omówiono także aktualny stan wiedzy na ten temat, wspominając aktualne rozwiązania. W rozdziale 6 przedstawiono wyniki uzyskane przez proponowane metody lokalizujące środki dołeczka oraz tarczy nerwu wzrokowego na zdjęciach dna oka. Wyniki te poddano analizie oraz porównano je z wynikami innych istniejących rozwiązań. Rozdział 7 stanowi podsumowanie rozprawy. Cztery dodatki uzupełniają informacje związane z rozprawą. W szczególności, Dodatek A poświęcono przedstawieniu całościowych wyników, które z uwagi na swoją obszerność, nie były prezentowane w rozdziale 6. W dodatku B omówiono szczegóły implementacji zaproponowanych metod wraz z narzędziami dodatkowymi. Natomiast Dodatek C poświęcono przedstawieniu i omówieniu matematycznych rozwinięć technik stosowanych w opracowaniu proponowanych metod oraz w wynikach. Wreszcie, w dodatku D przedstawiono problemy, których nie rozwiązano w niniejszej rozprawie wraz ze wskazówkami, które mogą ułatwić ich rozwiązanie. Rozprawę kończy bibliografia zawierająca 208 pozycji.

## **Poprawność i oryginalność postawionej tezy (wkład Autora)**

Głównym celem badań zaprezentowanych w rozprawie było zaproponowanie nowych metod lokalizujących dołeczki (punkt centralny plamki żółtej) oraz tarczę nerwu wzrokowego na zdjęciu dna oka. Do opracowania proponowanych metod wykorzystano dwa odmienne podejścia:

1. podejście wykorzystujące techniki przetwarzania obrazów,
2. podejście wykorzystujące sieci neuronowe.

Przy opracowywaniu proponowanych metod uwzględniono aspekty takie jak dokładność i pokrycie lokalizacji cech anatomicznych będących przedmiotem zainteresowania, efektywność obliczeniową

w postaci zużytego czasu i pamięci oraz możliwości przyspieszania wykonania analiz za pomocą rozwiązań sprzętowych.

Na podstawie głównego celu pracy postawiono następujące cztery tezy.

1. Dokładność i pokrycie wyznaczonych lokalizacji cech anatomicznych na zdjęciach dna oka przy użyciu sieci neuronowych będzie wyższa, niż w przypadku zastosowania technik przetwarzania obrazów.
2. Czas wykonania obliczeń i zużycie pamięci operacyjnej będzie niższe w przypadku lokalizacji cech anatomicznych na zdjęciach dna oka przy użyciu technik przetwarzania obrazów, niż w przypadku zastosowania sieci neuronowych.
3. Zastosowanie przyspieszania sprzętowego zaproponowanych metod za pomocą równoleglenia obliczeń na procesorach graficznych skróci czas obliczeń potrzebny na wyznaczenie lokalizacji cech anatomicznych na zdjęciach dna oka.
4. Dokładność i pokrycie lokalizacji cech anatomicznych na zdjęciach dna oka przy użyciu zespołu sieci neuronowych będą wyższe, niż w przypadku pojedynczej sieci neuronowej.

Należy zauważyć, że postawienie powyższych tez było uzasadnione licznymi wykonanymi przez Autora eksperymentami związanymi z analiza obrazów i nie budzą zastrzeżeń.

Zgodnie z celem badawczym i tezami opracowano i zaimplementowano dwie metody lokalizujące dołeczek oraz tarczę nerwu wzrokowego na zdjęciach dna oka. Każdą z zaproponowanych metod poddano testom pod kątem: dokładności i pokrycia wyznaczonych lokalizacji za pomocą pojedynczych sieci neuronowych oraz za pomocą zespołu sieci neuronowych, czasu wyznaczania lokalizacji oraz zużycia pamięci operacyjnej w wariantach implementacji wykonywanych sekwencyjnie na procesorze centralnym oraz częściowo równolegle na procesorze graficznym.

Pierwszą z metod zaproponowanych w rozprawie to metoda oparta na algorytmie MSiRL, która w rozprawie została określona jako *metoda szybkiej i równoległej lokalizacji współrzędnych środków dołeczka i tarczy nerwu wzrokowego na zdjęciach dna oka przy użyciu segmentacji oraz analizy komponentów połączonych*. Podstawą działania tego algorytmu są metody przetwarzania obrazów, takie jak konwersja do innych modeli barw, transformacje morfologiczne, operacje splotu, wyrównywanie histogramu, binaryzacja, łączenie wierzchołków krawędzi regionów i ich etykietowanie oraz analiza struktury kształtu. Celem ich połączenia jest wydobycie ze zdjęcia dna oka jednoznacznego regionu plamki żółtej lub tarczy nerwu wzrokowego. Najlepszy region jest wybierany na podstawie wyznaczonego zakresu rozmiarów, zatem zdaniem Autora, metoda powinna być stosowana na próbie zdjęć o zbliżonym rozmiarze. Warto zwrócić także uwagę na zewnętrzne parametry MT1 i MT2 algorytmu MSiRL, które określają minimalną akceptowalną wartość współczynnika kolistości znalezionych regionów. W przypadku procesu segmentacji plamki żółtej proponowana wartość domyślna wynosi 0.1, a w przypadku tarczy nerwu wzrokowego 0.3. Kierując się własną metodą doboru wartości tych parametrów, Autor rozprawy sugeruje zachowanie zależności, gdzie wartość MT2 będzie większa od MT1 co wynika z widoczności oraz wyraźności tych obiektów na zdjęciach dna oka. Okazuje się bowiem, że tarcza nerwu wzrokowego jest znacznie wyraźniejsza i jaśniejsza od plamki żółtej na większości zdjęć dna oka dostępnych w publicznych bazach, co ma wpływ na kształt regionu powstałego w procesie segmentacji. Co więcej, z uwagi na mniej wyraźny szum, tarcza nerwu wzrokowego może zostać z powodzeniem poddana procesowi wyrównywania krawędzi struktury regionu, co może mieć pozytywny wpływ na jej finalną kolistość i uzyskaną wartość współczynnika kolistości.

Drugą z metod zaproponowanych w rozprawie to metoda oparta na algorytmie MSLW, która w rozprawie została określona jako *metoda symultanicznej lokalizacji współrzędnych środków dołączka i tarczy nerwu wzrokowego na zdjęciach dna oka przy użyciu zespołu sieci neuronowych*. Podstawą działania tego algorytmu stanowią sieci neuronowe o architekturze YOLOv8 oraz uczenie zespołowe za pomocą metody Pasting, która stanowi modyfikację popularnej metody Bagging. Celem połączenia modeli bazowych (sieci neuronowych o architekturze YOLOv8) w zespół za pomocą metody Pasting było poprawienie ogólnej jakości predykcji poprzez wykorzystanie wielu modeli uczonych na różnych podzbiorach danych treningowych. Metoda ta jest często stosowana w problemach uczenia maszynowego, lecz w połączeniu z zaproponowanym podejściem do agregacji lokalizacji, stanowi nowość. W metodzie MSLW udział w lokalizacji dołączka oraz tarczy nerwu wzrokowego na zdjęciu dna oka bierze 16 sieci neuronowych, z których każda jest wytrenowana na rozłącznym podzbiore zdjęć dna oka. Po wyznaczeniu lokalizacji każdej z cech anatomicznych, następuje ich agregacja w postaci operacji wyznaczenia ważonego środka masy otaczającej ramki. Mając na uwadze fakt, że zastosowane sieci neuronowe o architekturze YOLOv8, jako modele bazowe metody MSLW, dokonują transformacji zdjęcia do rozmiaru 640x640 pikseli, zgodność rozmiarów zdjęć w przetwarzanej bazie nie ma istotnego znaczenia. Zdjęcia nie powinny być jednak mniejsze niż wejście sieci neuronowej. Warto zwrócić także uwagę na zewnętrzny parametr metody BoxThe , który określa minimalną wartość współczynnika ufności, który musi posiadać ramka otaczająca wybraną cechę anatomiczną, aby była brana pod uwagę podczas agregacji lokalizacji. Wartość parametru może wpływać na wyniki dokładności lokalizacji oraz pokrycia. Wysoka wartość może pozostawić jedynie te ramki, których sieć neuronowa jest bardziej pewna, co może przyczynić się do mniejszej liczby wyznaczonych lokalizacji cech anatomicznych oraz do wyznaczenia dokładniejszych środków. W przypadku małej wartości podczas agregacji lokalizacji mogą zostać wzięte pod uwagę ramki, których sieć jest mniej pewna, co może przyczynić się do mniej dokładnych lokalizacji cech anatomicznych, lecz w większej liczbie.

W celu weryfikacji dokładności i wydajności zaproponowanych w rozprawie metod, dla każdego zdjęcia dna oka zawartego w próbie testowej bazy obrazów wyznaczono współrzędne środków dołączka oraz tarczy nerwu wzrokowego i porównano je ze środkami wyznaczonymi przez twórców bazy za pomocą odległości Minkowskiego. Dla przedstawionych rozwiązań wyznaczono także miarę pokrycia określającą odsetek zdjęć, które otrzymały lokalizację środka cechy anatomicznej dna oka za pomocą testowanej metody. Wydajność działania zaproponowanych rozwiązań określono za pomocą czasu wyznaczania pojedynczej lokalizacji oraz rozmiaru zużytych zasobów pamięciowych w systemie komputerowym.

Analizując uzyskane wyniki testów można zauważyć, że dokładność i pokrycie wyznaczonych lokalizacji cech anatomicznych na zdjęciach dna oka są znacznie wyższe w przypadku wykorzystania zespołu sieci neuronowych, niż w przypadku wykorzystania technik przetwarzania obrazów. Potwierdza to tezę numer 1.

Zauważono, że metoda lokalizująca cechy anatomiczne na zdjęciu dna oka z wykorzystaniem technik przetwarzania obrazów zużywa mniej pamięci operacyjnej na potrzeby obliczeń oraz wykonuje się w krótszym czasie, niż metoda wykorzystująca zespół sieci neuronowych. Potwierdza to tezę numer 2.

Zastosowanie podejścia przyspieszania sprzętowego z wykorzystaniem zrównoleglania obliczeń na procesorze graficznym przyniosło efekt w postaci skrócenia czasu obliczeń w obydwu metodach. Zauważono jednak, że wykorzystanie procesora graficznego wpłynęło na nieznaczne zwiększenie zużycia pamięci operacyjnej. Potwierdza to tezę numer 3.

Podczas analizy wyników testów dostrzeżono także, że zastosowanie zespołu sieci neuronowych przyniosło poprawę zarówno dokładności, jak i pokrycia wyznaczonych lokalizacji cech anatomicznych na zdjęciach dna oka w porównaniu do zastosowania pojedynczej sieci neuronowej. Potwierdza to tezę numer 4.

Do implementacji proponowanych metod wykorzystano nowoczesne i ogólnie uznane narzędzia eksploracji danych obrazowych, jak: język programowania Python wraz z bibliotekami zewnętrznymi, takimi jak OpenCV, Ultralytics, PyTorch. Warto dodać, że zaproponowane w rozprawie metody zaimplementowano w dwóch wariantach: sekwencyjnie, gdzie operacje wykonywane są synchronicznie i jednowątkowo na CPU oraz równoległe, gdzie operacje wykonywane są asynchronicznie i wielowątkowo na GPU, co umożliwia redukcję czasu potrzebną na dokonanie lokalizacji cech anatomicznych na zdjęciu dna oka. W badaniach wykorzystano dwie bazy kolorowych zdjęć dna oka: baza fundacji Okulistyka 21 (BFO21, zawierająca 3626 zdjęć) oraz baza zdjęć dna oka Indian Diabetic Retinopathy Image Dataset (IDRID, zawierająca 103 zdjęcia),

Podsumowując, wkład badawczy Autora rozprawy w rozwój dyscypliny naukowej przedstawiony w niniejszej rozprawie można określić następująco.

1. Zaproponowano dwie nowe metody lokalizujące na zdjęciu dna oka środek dołączka i środek tarczy nerwu wzrokowego.
2. Sprawdzone dokładność i pokrycie wyznaczonych lokalizacji tych obiektów na zdjęciach dna oka za pomocą obydwu zaproponowanych metod.
3. Sprawdzone czas wykonania obliczeń i zużycie pamięci operacyjnej za pomocą obydwu zaproponowanych metod.
4. Zastosowano przyspieszenie sprzętowe w implementacjach obydwu metod za pomocą zrównoleglania obliczeń na procesorze graficznym.
5. Sprawdzone dokładność i pokrycie lokalizacji tych obiektów na zdjęciach dna oka przy użyciu zespołu sieci neuronowych oraz przy użyciu pojedynczych sieci neuronowych.

## **Wiedza i umiejętności Autora do poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników**

W rozprawie Autor zamieścił przegląd aktualnego stanu wiedzy w zakresie metod rozpoznawania obrazów (rozdział 3). Opisy te zostały wykonane z wysoką starannością i z odwołaniami do literatury, co pozwoliło umiejscowić w literaturze przedmiotu prezentowane badania. W kolejnych rozdziałach także zamieszczono odwołania do literatury związane z używanymi pojęciami, metodami lub używanymi do eksperymentów bibliotekami oprogramowania. Bez wątpienia świadczy to o dużej wiedzy Kandydata. Ponadto, należy mocno podkreślić, że napisanie rozprawy wymagało wcześniejszego skonstruowania, w tym zaprogramowania, środowiska eksperymentalnego.

Odnosząc się do umiejętności Autora rozprawy w zakresie poprawnego i przekonującego przedstawiania uzyskanych wyników należy stwierdzić, że w tym zakresie Autor rozprawy wykazał się dobrymi umiejętnościami. Rozprawa napisana jest dość przejrzysto i bardzo dobrze zaprojektowano jej strukturę, która została uwidoczniła w spisie treści.

## Uwagi na temat rozprawy

Z formalnego i matematycznego punktu widzenia rozprawa nie budzi zastrzeżeń recenzenta.

Podczas lektury rozprawy nasuwają się jednak pewne uwagi, które można traktować jako sformułowanie pewnych mankamentów lub sugestii co do dalszych prac. Dobrze by było, aby Kandydat odniósł się do nich podczas obrony rozprawy.

1. Zdaniem recenzenta, w rozprawie brakuje solidnego podrozdziału, opisującego pokrewne prace znane z literatury. W prawdzie w podrozdziale 5.2 zatytułowanym „Istniejące rozwiązania” zamieszczono wiele odwołań do literatury, ale przedstawione opisy są zbyt mało szczegółowe. Np. brak jest bardziej szczegółowych informacji jakie stosują podejścia inni autorzy, jakich używają architektur w metodach DL, jakich dokonują augmentacji obrazów itd. W tym kontekście dobrze by było, aby w rozprawie pojawił się fragment opisujący szczegółowo innowacyjność proponowanych w rozprawie nowych algorytmów w kontekście istniejących. Informacje o tym są rozrzucone po całej pracy, ale mogłyby być zestawione razem.
2. *Bagging* i *pasting* to dość powszechne metody przy tworzeniu zespołów klasyfikatorów, autor wykorzystał *pasting*, ale nie przetestował wyników dla *baggingu*. Dlaczego? Jak wiadomo, metoda *pasting* została raczej wymyślona dla dużych zbiorów. Czy zatem jedynym powodem wykorzystania jedynie metody *pasting* jest duży rozmiar danych, a co za tym idzie potencjalnie długi czas obliczeń dla metody *bagging*, czy też są jakieś inne powody?
3. Co prawda metody YOLO w różnych wersjach to gotowe rozwiązania ale Autor przygotował zespół tego typu klasyfikatorów. Dlaczego? Czy ten fakt wynika np. z tego, że Autor przeprowadził pewną liczbę prób, celem dobrania metod i parametrów i prezentuje tylko te najlepsze?

## Działalność publikacyjna

Autor rozprawy opublikował łącznie 7 publikacji, w tym 5 w punktowanych czasopismach naukowych, 1 w recenzowanych materiałach pokonferencyjnych oraz 1 monografię. Zdaniem recenzenta, jest to znaczący dorobek jak na osobę, która ukończyła studia magisterskie w 2020 roku. Wątpliwości może budzić fakt, że żadne z tych czasopism nie jest przypisane do dyscypliny *Informatyka techniczna i telekomunikacja*.

Aktualne główne wskaźniki bibliometryczne Autora rozprawy to:

- według źródła Scopus: index Hirscha=2, liczba cytowań=7,
- według źródła Google Scholar: index Hirscha=2, liczba cytowań=9.

Są to zatem wskaźniki na niezbyt wysokim poziomie, ale wystarczającym zważając na młody wiek Autora rozprawy.

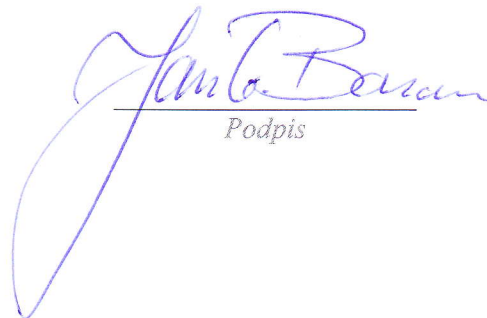
## Podsumowanie

Uzyskane wyniki są interesujące zarówno z teoretycznego, jak i praktycznego punktu widzenia. Dlatego niezależnie od wspomnianych wyżej mankamentów, uważam pracę za wartościową. Autor wykazał dobre opanowanie wielu różnorodnych technik matematycznych i informatycznych. Sposób wykorzystania tych technik wskazuje na opanowanie przez Niego warsztatu naukowego.

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach, moja ocena rozprawy pod względem trzech zwyczajowo rozpatrywanych kryteriów jest następująca:

1. rozprawa doktorska zawiera oryginalne rozwiązanie problemu naukowego
2. Kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie,
3. Kandydat ma umiejętność samodzielnego prowadzenia pracy naukowej.

W związku z powyższym, **wniosuję o dopuszczenie rozprawy doktorskiej do publicznej obrony.**



Podpis