

RECENZJA

rozprawy doktorskiej mgr. inż. Łukasza ROSZKOWIAKA
pt. *Nowa metoda komputerowego wspomagania ilościowej oceny komórek w
wielkoobszarowych obrazach preparatów tkankowych raka sutka
barwionych immunohistochemicznie*

1. Podstawa recenzji

Przedmiotem oceny jest rozprawa doktorska w formie opublikowanych i powiązanych tematycznie 6 artykułów naukowych wraz z pisemnym opracowaniem zawierającym 77 stron maszynopisu (Art. 187 ust.3 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. 2018 poz. 1889)).

Podstawą opracowania recenzji jest pismo Zastępcy Dyrektora Instytutu Biocybernetyki i Inżynierii Biomedycznej PAN, informujące o powierzeniu mi przez Radę Naukową IBIB PAN na posiedzeniu w dniu 27 września 2022 r. roli recenzenta rozprawy doktorskiej mgr. inż. Łukasza Roszkowiaka.

Opinia została opracowana na podstawie udostępnionej rozprawy w formie opracowania pisemnego wraz z 6 publikacjami oraz oświadczeń współautorów publikacji.

2. Zakres i cel rozprawy

Opiniowana praca doktorska jest rozprawą z zakresu komputerowego wspomagania diagnozy medycznej poprzez przetwarzanie, analizę i rozpoznawanie informacji obrazowej, ukierunkowanej na diagnozowanie raka sutka na podstawie obrazów barwionych preparatów tkankowych. Obejmuje złożoną problematykę analizy zadania diagnostycznego w celu zdefiniowania cech różnicujących normę od nowotworu, pozyskiwania preparatów mikroskopowych, ich przetwarzania w celu poprawy jakości obrazu, wydobycia obiektów i wyznaczenia cech ilościowych, a także metody i algorytmy klasyfikacji zmierzające do podjęcia stosownej decyzji diagnostycznej. Zawiera również wyniki badań eksperymentalnych weryfikujących praktyczną przydatność proponowanych koncepcji oraz dyskusję rezultatów opracowanych metod.

Medyczna diagnostyka obrazowa znajduje – w miarę rozwoju technik obrazowania – coraz większe zastosowanie w codziennej praktyce klinicznej. Mimo dużej na ogół zawartości informacyjnej obrazu, która ułatwia określenie stanu pacjenta i podjęcie stosownej decyzji,

diagnostyka obrazowa dostarcza także wielu problemów. Wynikają one przede wszystkim z trudności w oddzieleniu ważnego obiektu od tła, odróżnieniu informacji bardziej od mniej istotnej, czy na wyodrębnieniu określonych klas obiektów istniejących w obrazie. Trudności przysparza także liczbowo ocena niektórych cech charakteryzujących analizowane obiekty, która często jest wręcz niemożliwa do przeprowadzenia na drodze wzrokowego oglądu. Zastosowanie w tym zakresie metod cyfrowego (komputerowego) przetwarzania i analizy obrazów stanowi niewątpliwie ułatwienie w ilościowej ocenie obiektów zawartych na obrazie, która może być dalej wykorzystana w automatycznym procesie komputerowego wspomagania podejmowania decyzji.

Niniejsza praca mieści się w tym właśnie bardzo aktualnym nurcie informatyki medycznej, dlatego tematykę podjętą w rozprawie uważam za oryginalną i w pełni uzasadnioną, a samą pracę za wartościowy i istotny wkład we współczesną problematykę automatycznego przetwarzania i analizy medycznej informacji obrazowej oraz wspomagania procesów decyzyjnych w medycynie. Pozytywna ocena tematu rozprawy wynika również z faktu, iż każda praca z tego zakresu, której wynikiem są konstruktywne algorytmy przetwarzania i klasyfikacji informacji obrazowej ukierunkowane na konkretny medyczny problem diagnostyczny, jest celowa i ważna z praktycznego punktu widzenia i za taką właśnie uważam opiniowaną rozprawę.

3. Zawartość i forma rozprawy

Recenzowana rozprawa składa się z dwóch części: pisemnego syntetycznego opracowania oraz zestawu 6 artykułów naukowych.

Pisemne opracowanie, będące klamrą spinającą publikacje i nadającą im wspólny tematyczny mianownik, składa się z 7 rozdziałów poprzedzonych streszczeniami w języku polskim i angielskim i spisem stosowanych skrótów oraz bibliografii zawierającej 110 pozycji i obejmuje 77 stron maszynopisu. We *Wstępie*, w pierwszej kolejności przedstawiono medyczny kontekst problemu diagnostycznego rozpatrywanego w pracy, przedstawiono informacje o raku sutka i jego rodzajach (podtypy), dane statystyczne dotyczące choroby, charakterystyczne objawy oraz metody i badania diagnostyczne ze szczególnym uwzględnieniem badań histopatologicznych z wybarwieniem preparatów ukierunkowanych na ocenę statusu receptorów hormonalnych i innych biomarkerów prognostycznych. Następnie, dość szczegółowo został rozwinięty wątek technicznych i metodologicznych aspektów oceny preparatu histopatologicznego. Zaprezentowano tu procedurę oceny preparatu przez patomorfologa i jej liczne wady (np. czasochłonność badań, zgrubne i subiektywne wyniki liczbowe, brak standaryzacji) oraz przetwarzanie obrazów w patologii cyfrowej (automatyczna klasyfikacja pojedynczych jąder komórkowych) jako wspomaganie pracy lekarza i wynikające stąd korzyści. W dalszej części *Wstępu* Autor omawia proces przygotowania preparatów histopatologicznych oraz techniki digitalizacji ich obrazów z wykorzystaniem tzw. skanerów wirtualnych (patologia cyfrowa). Efektem działania skanera jest specyficzna hierarchiczna struktura zwana wirtualnym slajdem (*Whole Slide Image*). Obraz jest tu zapisywany w wielu rozdzielczościach, co umożliwia jego różnorodną wizualizację od całościowej miniatury do powiększonego fragmentu preparatu. Na koniec, Doktorant skupił się na osiągnięciach literaturowych w zakresie tzw. patologii obliczeniowej

związanej z algorytmami i metodami przetwarzania i analizy wirtualnych slajdów, przede wszystkim w diagnostyce raka sutka.

W kolejnym rozdziale zatytułowanym *Tezy pracy* zostały przedstawione w sposób jednoznaczny 4 tezy rozprawy związane z opracowaniem metody ilościowej oceny komórek w barwionych obrazach histopatologicznych pochodzących od pacjentów z rakiem sutka.

Następny rozdział (*Osiągnięcia*) zawiera zasadnicze wyniki Doktoranta, które zostały ujęte w 4 podrozdziały: (1) Od miniatury do pełnej rozdzielczości wirtualnego slajdu, (2) Przetwarzanie wstępne obrazu mikroskopowego, (3) Przetwarzanie końcowe obrazu i (4) System przetwarzania obrazów. Doktorant ograniczył się w prezentacji do tych wyników zawartych w publikacjach, które są wyłącznie jego autorstwa.

W rozdziale *Zbiory danych* opisane zostały dwa zbiory IISPV oraz WBCD, które posłużyły do weryfikacji opracowanego systemu komputerowego CHISEL oraz eksperymentalnej oceny zaproponowanych metod analizy i przetwarzania obrazów cyfrowych preparatów histopatologicznych. W rozdziale *Dyskusja* ujęto analizę wyników przeprowadzonych badań. Dwa ostatnie rozdziały *Podsumowanie* i *Wnioski* zawierają uwagi konkludujące pracę.

Na wysoką ocenę zasługuje przyjęta forma rozprawy, w której oprócz jej zasadniczej zawartości w postaci 6 publikacji naukowych, przedstawiono dodatkowe opracowanie, choć odpowiedni zapis ustawowy tego nie wymaga. Pozwoliło to z jednej strony – tak jak napisano już wcześniej – ująć publikacje w spójną całość o wspólnym tytule, a z drugiej strony umożliwiło przedstawienie ważnych treści, których ze zrozumiałych względów zabrakło w publikacjach. Dotyczy to w szczególności sformułowania tez pracy, bogatego kontekstu medycznego rozpatrywanego problemu komputerowego wspomagania decyzji, szczegółowego opisu wykorzystywanych w badaniach eksperymentalnych zbiorów rzeczywistych danych oraz wspólnych wniosków końcowych i podsumowania rozprawy. Innymi słowy, pierwsza część rozprawy prezentuje szerokie tło problemu badawczego rozwiązanego w rozprawie oraz rezultaty objęte jej tematyką w logicznym i przejrzystym układzie ujmującym osiągnięcia Doktoranta szczegółowo przedstawione w drugiej części, czyli w zestawie publikacji naukowych. Taka konstrukcja rozprawy umożliwiła zaprezentowanie uzyskanych wyników w sposób jasny, szczegółowy i kompletny, co zostało w ocenianym tekście starannie wykorzystane. Ogólnie zatem sposób ujęcia materiału w rozprawie oraz redakcję całości oceniam pozytywnie.

W 2 części opracowania przedstawione zostały odbitki 6 prac stanowiących zasadniczą treść rozprawy doktorskiej. Choć wszystkie prace są współautorskie, to na podstawie złożonych oświadczeń współautorów można jednoznacznie i dość szczegółowo określić wkład Doktoranta, zarówno w wymiarze procentowym, jak i w układzie merytorycznym. W każdym przypadku ten wkład jest dominujący i dotyczy kluczowych elementów publikacji (koncepcja metody, opracowanie algorytmów i ich implementacja, badania, wnioski i redakcja pracy).

Poniżej przedstawiono dane bibliograficzne publikacji, wkład Autora oraz krótką informację o zawartych wynikach.

1. Łukasz Roszkowiak, Anna Korzyńska, Jakub Żak, Dorota Pijanowska, Żaneta Swiderska-Chadaj, Tomasz Markiewicz; “Survey: interpolation methods for whole slide image processing” Journal of Microscopy, 2016; (IF2016=1,692);

Wkład doktoranta (według oświadczeń współautorów):

Conceptualization – 50 %	Data curation – 0 %
Methodology – 100 %	Writing (original draft preparation) – 100 %
Software – 100 %	Writing (review and editing) – 50 %
Validation – 33 %	Visualisation – 50 %
Formal analysis – 50 %	Supervision – 0 %
Investigation – 100 %	Project administration – 0%
Resources – 0 %	Funding acquisition – 0 %

W pracy dokonano analizy porównawczej 9 metod interpolacji związanych ze zmianą rozmiaru obrazu pod względem jakości przeskalowania oraz czasu trwania procedury. Celem badań było określenie, która metoda interpolacji jest najlepsza do zmiany rozmiaru całych obrazów slajdów.

2. Łukasz Roszkowiak, Carlos Lopez; “PATMA: Parser of archival tissue microarray”; Peer J, Vol. 4:e2741, 2016; (IF2016=2,177);

Wkład doktoranta (według oświadczeń współautorów):

Conceptualization – 50 %	Data curation – 0 %
Methodology – 100 %	Writing (original draft preparation) – 100 %
Software – 100 %	Writing (review and editing) – 50 %
Validation – 100 %	Visualisation – 100 %
Formal analysis – 50 %	Supervision – 0 %
Investigation – 100 %	Project administration – 0%
Resources – 0 %	Funding acquisition – 100 %

Praca przedstawia metodę PATMA realizującą półautomatyczną procedurę podziału wirtualnego slajdu mikromacierzy tkankowej na obrazy pojedynczych próbek wraz z odpowiednim powiązaniem lokalizacji pojedynczej próbki w mikromacierzy z danymi pacjenta.

3. Łukasz Roszkowiak, Jakub Żak, Krzysztof Siemion, Dorota Pijanowska, Anna Korzyńska; “Nuclei Detection with Local Threshold Processing in DAB&H Stained Breast Cancer Biopsy Images”, International Conference on Computer Vision and Graphics. ICCVG 2020. Lecture Notes in Computer Science, vol 12334, Springer, Cham;

Wkład doktoranta (według oświadczeń współautorów):

Conceptualization – 50 %	Data curation – 33 %
Methodology – 100 %	Writing (original draft preparation) – 100 %
Software – 100 %	Writing (review and editing) – 50 %
Validation – 33 %	Visualisation – 50 %
Formal analysis – 25 %	Supervision – 0 %
Investigation – 100 %	Project administration – 100%
Resources – 0 %	Funding acquisition – 33 %

W pracy została przeprowadzona analiza porównawcza 18 różnych metod wyznaczania progów adaptacyjnych segmentacji w celu wykrywania komórek. Badania przeprowadzono na zestawie obrazów histopatologicznych raka piersi z barwieniem immunohistochemicznym.

4. Łukasz Roszkowiak, Anna Korzyńska, Krzysztof Siemion, Dorota Pijanowska; “The Influence of Object Refining in Digital Pathology”; Image Processing and Communications Challenges 10. IP&C 2018. Advances in Intelligent Systems and Computing, Vol 892, Springer, Cham;

Wkład doktoranta (według oświadczeń współautorów):

Conceptualization – 50 %	Data curation – 50 %
Methodology – 100 %	Writing (original draft preparation) – 100 %
Software – 100 %	Writing (review and editing) – 50 %
Validation – 50 %	Visualisation – 100 %
Formal analysis – 25 %	Supervision – 0 %
Investigation – 100 %	Project administration – 100%
Resources – 0 %	Funding acquisition – 33 %

W pracy przeanalizowano skuteczność 5 algorytmów udoskonalania granic ROI (bazujących na metodach rozrostu obszarów, aktywnego konturu i k-means) w określaniu kształtu jąder komórkowych wysegmentowanych wcześniej metodą progowania w badaniu histopatologicznym różnych typów nowotworów.

5. Łukasz Roszkowiak, Anna Korzyńska, Dorota Pijanowska, Ramon Bosch, Marylene Lejeune, Carlos Lopez; “Clustered nuclei splitting based on recurrent distance transform in digital pathology images”; J Image Video Proc., Vol. 26, 2020; (IF2020=1,789);

Wkład doktoranta (według oświadczeń współautorów):

Conceptualization – 50 %	Data curation – 0 %
Methodology – 100 %	Writing (original draft preparation) – 100 %
Software – 100 %	Writing (review and editing) – 33 %
Validation – 100 %	Visualisation – 100 %
Formal analysis – 50 %	Supervision – 0 %
Investigation – 100 %	Project administration – 50%
Resources – 0 %	Funding acquisition – 33 %

Praca przedstawia oryginalną metodę podziału skupisk jąder komórkowych, które w wyniku procesu segmentacji utworzyły jeden klaster złożony z wielu jąder. Zaproponowana metoda wykorzystująca algorytm wododziałowy dzieli skupisko na mniejsze podgrupy i – finalnie – na oddzielne jądra w oparciu o cechy kształtu.

6. Łukasz Roszkowiak, Anna Korzyńska, Krzysztof Siemion, Jakub Żak, Dorota Pijanowska, Ramon Bosch, Marylene Lejeune, Carlos Lopez; “System for quantitative evaluation of DAB&H-stained breast cancer biopsy digital images (CHISEL)” Scientific Reports 11, Vol. 9291, 2021 (IF2020=4,380).

Wkład doktoranta (według oświadczeń współautorów):

Conceptualization – 50 %	Data curation – 0 %
--------------------------	---------------------

Methodology – 100 %
Software – 100 %
Validation – 50 %
Formal analysis – 33 %
Investigation – 100 %
Resources – 0 %

Writing (original draft preparation) – 100 %
Writing (review and editing) – 25 %
Visualisation – 50 %
Supervision – 0 %
Project administration – 50 %
Funding acquisition – 33 %

Praca przedstawia system komputerowy CHISEL, którego funkcjonalności pozwalają na kompleksową ilościową ocenę łagodnych i złośliwych (nowotworowych) preparatów histopatologicznych z immunohistochemicznym barwieniem jądrowym o różnej intensywności i zróżnicowanej zawartości.

4. Uzyskane wyniki

Celem pracy było wykazanie następujących tez (cyt.):

Teza I: Metoda interpolacji obrazów, zastosowana do skalowania wielkoobszarowych obrazów preparatów tkankowych, nie ma wpływu na analizę komputerową liczby obiektów.

Teza II: Zbieranie danych z wielu przylegających do siebie fragmentów obrazów wielkoobszarowych i łączenie ich w jeden wspólny wynik, bez utraty obiektów znajdujących się na ich krawędziach, jest możliwe z wykorzystaniem zaproponowanego schematu wieloprzesunięciowego.

Teza III: Adaptacyjne progowanie obrazu jest metodą wystarczającą do wyodrębnienia jąder komórkowych z obrazów wielkoobszarowych preparatów tkankowych barwionych immunohistochemicznie w celu ich zliczania.

Teza IV: Rozdzielanie stykających się i nakładających jąder komórkowych w obrazach barwionych tkanek jest możliwe z wykorzystaniem zaproponowanej metody opartej na rekurencyjnym progowaniu transformaty odległościowej.

Realizacja tak nakreślonego celu doprowadziła do szeregu szczegółowych i konkretnych osiągnięć. Do szczególnie ważnych i wartościowych wyników pracy – zarówno w aspekcie teoretycznym, jak i praktycznym – zaliczam:

I. W zakresie badań i analiz slajdów wirtualnych:

1. Eksperymentalna analiza porównawcza metod interpolacji do przeskalowania wirtualnych slajdów

W obrazowaniu medycznym, do wspomagania diagnostyki i prognozowania, często zachodzi potrzeba oceny całych obrazów histologicznych czy onkologicznych. Gdy są to obrazy wielkoskalowe, wiąże się to z koniecznością ich przeskalowania. Zmiana rozmiaru obrazu jest jednym z najczęstszych zastosowań interpolacji. W pracy dokonano analizy porównawczej 9 metod interpolacji związanych ze zmianą rozmiaru obrazu pod względem jakości przeskalowania oraz trwania procedury. Porównywano następujące metody: Nearest neighbor, Linear/bilinear, Bicubic, Cosine, Sins, B-spline, Hermite, Lagrange, Catmull-Rom splines.

Do badań wykorzystano cztery ogólne obrazy testowe (obraz Lena, wzór testowy Głowa Indianina, wzorze testowy kolorów PAL, wzorze testowy TV) i 12 specjalistycznych

obrazów próbek tkanek biologicznych raka piersi wykorzystywanych w diagnostyce patologicznej, klasyfikacji i rokowaniu i pochodzących ze zbioru IISPV. Tkanka przedstawiona na zdjęciach jest różna: od bardzo ciemnej do jasnej, z rzadką i zwartą architekturą komórek. Wszystkie obrazy są skupione, a jakość obrazów jest typowa dla tego rodzaju danych biologicznych.

Aby porównać efektywność badanych metod interpolacji, obrazy testowe zostały przeskalowane, a następnie ponownie przeskalowane do rozmiaru oryginalnego przy użyciu tego samego algorytmu. Zmodyfikowany obraz porównano z obrazem oryginalnym pod względem wskaźników MSE (Mean squared error), PSNR (Peak signal-to-noise ratio) i SSIM (Structural similarity). Porównano również czasy potrzebne do wykonania obliczeń.

2. Opracowanie komputerowej metody wsparcia analizy mikromacierzy tkankowych (TMA)

Mikromacierz tkankowa, czyli regularna struktura złożona z tzw. rdzeni tkankowych, jest coraz częściej stosowana w komputerowo wspomaganej analizie histopatologicznej preparatów tkankowych, gdyż pozwala na szybkie badanie wielu próbek jednocześnie. Niestety rozróżnienie właściwej kolejności rdzeni nie jest zadaniem trywialnym, ponieważ nie są one oznaczone bezpośrednio na skanie.

Głównym celem podjętego tematu było opracowanie procedury automatycznego wyszukiwania i wyodrębniania rdzeni z archiwalnych obrazów mikromacierzy tkankowych (system PATMA).

Opracowana metoda obejmuje 5 etapów:

- (1) Odczyt pliku. Załadowane pliki są sprawdzane pod kątem dostępnych rozdzielczości. Wybór rozdzielczości może być automatyczny (wtedy wybierana jest najmniejsza rozdzielczość) lub ręczny (użytkownik ma wybór z wszystkich dostępnych rozdzielczości).
- (2) Przetwarzanie wstępne - obejmuje konwersję z przestrzeni barw RGB do przestrzeni Lab, binaryzację (utworzona maska lokalizuje poszczególne rdzenie), usuwanie zbyt małych obiektów, operacje morfologiczne. Dodatkowo, jest możliwość ręcznego poprawiania niezadawalającego wyniku preprocessingu w tzw. trybie eksperckim.
- (3) Analiza rdzeni – wyznaczanie cech zidentyfikowanych obiektów, określenie odległości pomiędzy rdzeniami, przypisanie obiektów do kolumn i wierszy macierzy, tworzenie obszaru zainteresowania (ROI) wokół każdego obiektu (ROI nie zachodzą na siebie), dla każdego ROI naniesienie prawidłowych etykiet numerycznych, odpowiadających numeracji rdzeni TMA.
- (4) Ręczna korekta przez użytkownika – wyznaczone ROI są prezentowane w formie nakładki na przetworzonym obrazie – użytkownik ma możliwość korekty ROI (przesunięcie, zmiana rozmiaru, zmiana etykiety).
- (5) Ekstrakcja ROI – wybrane ROI zawierające rdzenie są wyodrębniane z obrazu o maksymalnej rozdzielczości jako osobne pliki (TIFF).

Przeprowadzono badania eksperymentalne oceniające skuteczność opracowanej metody. Analizie poddano 26 wirtualnych slajdów TMA pochodzących ze zbioru IISVP. W 5 przypadkach wybrano rozdzielczość ręcznie, gdyż dała lepsze wyniki niż wybór automatyczny. W 11 przypadkach użyto trybu eksperckiego, bowiem preprocessing nie dał zadowalającego rezultatu. Odsetek prawidłowo dobranych rdzeni w badanych slajdach TMA

wyniósł 89%. Liczbę obiektów, które zostały nieprawidłowo przypisane, można zdecydowanie zmniejszyć dzięki interakcji człowieka (tryb eksperta). Półautomatyczne przetwarzanie PATMA zajmuje mniej niż 2 minuty, aby w pełni przygotować preparat TMA, jeśli jakość TMA i zeskanowanego obrazu jest zadowalająca. W przypadku obrazów wymagających większej interakcji z człowiekiem zajmuje to około 5 minut.

II. W zakresie przetwarzania wstępnego obrazu mikroskopowego:

1. Analiza porównawcza metod segmentacji z progiem adaptacyjnym

Ponieważ charakterystyki obrazów barwionych tkanek mają mocno lokalny charakter (duża zmienność barwy, niejednorodność oświetlenia), wśród stosowanych metod segmentacji takich obrazów szczególne znaczenie mają algorytmy segmentacji progowej z adaptacyjnym wyznaczaniem wartości progu. Polegają one na lokalnym (w otoczeniu analizowanego punktu) wyznaczeniu progu z oknem zależnym od rozmiaru segmentowanego obiektu, które definiuje otoczenie. Z tego względu metody segmentacji z adaptacyjnym progowaniem są dziedzinowo zorientowane i algorytmu stosowanego z powodzeniem w jednym obszarze nie można wprost przenieść do innego obszaru zastosowań. Doktorant przeprowadził eksperymentalną analizę porównawczą 18 literaturowych metod segmentacji z progowaniem adaptacyjnym, które działają w oparciu o: lokalną wariancję (3 metody), lokalny kontrast (6 metod), schemat centrum-otoczenie (1 metoda), lokalną entropię (4 metody), grupowanie, także rozmyte (3 metody) i minimalny błąd (1 metoda).

Badania przeprowadzono na zestawie obrazów histopatologicznych raka piersi, na których wykonano immunohistochemiczne barwienie FOXP3. Jako ilościowe wskaźniki jakości ocenianej procedury segmentacji przyjęto miary precyzji, czułości i F-score względem oznaczeń dokonanych ręcznie (*ground truth*). W wyniku badań stwierdzono, że metoda Bradleya jest najskuteczniejszą metodą wykrywania jąder w tego typu wybarwionych próbkach tkanek i została ona zarekomendowana do praktycznego zastosowania.

Doktorant jest autorem dwóch oryginalnych metod segmentacji z adaptacyjnym progowaniem, które działają w oparciu o hybrydowy schemat wykorzystujący jednocześnie lokalną wariancję i lokalny kontrast. Metody te były przedmiotem wcześniejszych publikacji, nieujętych w zestawie składającym się na rozprawę doktorską. Nie były one także przedmiotem omówionych wyżej badań eksperymentalnych.

III. W zakresie przetwarzania końcowego obrazu:

1. Analiza porównawcza metod poprawy jakości segmentacji jąder komórkowych

Ocena kształtu określonych typów jąder komórkowych odgrywa ważną rolę w badaniu histopatologicznym różnych typów nowotworów. Kluczowy wpływ na uzyskanie zadowalających krawędzi (granic) obiektu ma dokładność zastosowanej techniki automatycznej segmentacji komórek. Po zastosowaniu dość prostej metody segmentacji progowej z odpowiednio dobranym progiem można uzyskany rezultat udoskonalać z wykorzystaniem jednej z wielu metod literaturowych. Po wstępnym określeniu progu granica może być dalej udoskonalana, aby lepiej pasowała do oryginalnych jąder. Doktorant przeprowadził eksperymentalne badania porównawcze w celu oceny skuteczności poprawy jakości segmentacji progowej. W pierwszej kolejności uzyskano obrazy monochromatyczne i

poddano je progowaniu adaptacyjnemu 4 metodami wyznaczania progów. Następnie na wysegmentowanych obiektach zastosowano operacje morfologiczne erozji i dylatacji i tak spreparowane obiekty poddano działaniu 5 następujących porównywanych metod udoskonalania segmentacji progowej:

- metoda AC wykorzystująca algorytm zlokalizowanego aktywnego konturu, w którym wyniki progowania traktowane są jako obrys początkowy;
- metoda CV stosująca algorytm aktywnego konturu bez krawędzi Chana-Vese'a – tu podobnie, wyniki progowania traktowane są jako obrys początkowy;
- metoda RG implementująca algorytm rozrostu regionu od punktu początkowego przy użyciu średniej intensywności uzyskanego w wyniku progowania po erozji;
- metoda RRG stosująca rekurencyjny algorytm rozrostu regionu; punktem wyjścia są wyniki progowania po erozji;
- metoda KM stosująca segmentację obrazu metodą k-średnich na podstawie histogramu.

Badania przeprowadzono na obrazach rzeczywistych pochodzących ze zbioru IISPV – losowo wybrano 13 ROI, na których ręcznie zostało wysegmentowanych 600 jąder komórkowych traktowanych w badaniach jako *ground truth*. Wyniki pokazały skuteczność metod opartych na algorytmie aktywnego konturu (metody AC i CV) pod względem wskaźników na poziomie pikselowym (czułość, specyficzność, odsetek błędnie zaindeksowanych pikseli), stąd zostały zarekomendowane do stosowania w rozpatrywanym problemie diagnostycznym oraz zastosowane w przedstawionym dalej systemie komputerowym CHISEL.

2. Metoda podziału skupisk jąder komórkowych

Dokładność zastosowanej techniki automatycznej segmentacji jąder ma kluczowe znaczenie dla uzyskania wysokiej jakości i skutecznych wniosków diagnostycznych. Niestety, wiele obiektów na obrazach histopatologicznych jest połączonych (skupionych), co znakomicie utrudnia interpretację komórek położonych w takich skupiskach (klastrach). Stosowane algorytmy rozdziału klastrow korzystające głównie z metody wododziałowej, charakteryzują się nienajlepszymi wskaźnikami jakościowymi. Doktorant opracował nową metodę podziału klastrow na pojedyncze komórki, która opiera się na rozróżnieniu rodzaju klastrow, rekurencyjnym progowaniu transformaty odległościowej i na algorytmie wododziałowym. Dane wejściowe opracowanego algorytmu stanowi obraz binarny obejmujący analizowany klaster. Następnie wyznaczana jest transformata odległościowa klastru oraz wartość progową wyliczana rekurencyjnie, dokonując odcięcia transformaty odległościowej i pozwalającą w ten sposób dokonać podziału klastra. Parametry algorytmu zależą od typu analizowanego klastra (wyróżniono 4 typy: (1) dwa stykające się obiekty, (2) 3 lub więcej obiektów o różnym poziomie nakładania, (3) mocno upakowany klaster z bardzo dużym poziomem nakładania, (4) bardzo duży klaster zawierający wiele obiektów). Typ klastra można zidentyfikować stosując zaproponowane parametryczne kryterium liczbowe. Na koniec, stosowana jest procedura wododziałowa, która znajdując linie podziału pogrubia ją i wstawia do obrazu wejściowego w celu końcowego rozdzielenia obiektów. Powstałe obiekty są weryfikowane pod względem rozmiaru (pole powierzchni) i regularności kształtu (mimośrodowość i indeks kształtu), aby zdecydować, czy można uznać procedurę za zakończoną. Jeśli nie, powtarza się ją ze zwiększonym licznikiem rekurencji.

Opracowana metoda została eksperymentalnie porównana z 5 metodami literaturowymi z wykorzystaniem zbioru rzeczywistych obrazów próbek histopatologicznych raka sutka (zbiór IISPV) oraz zbioru obrazów syntetycznych (zbiór BBBC0004). Do porównania skuteczności podziału klastrow jader komórkowych zastosowano znane miary (czułość, precyzja, F-score) oraz dodatkowo miary specyficzne dla rozpatrywanego problemu, oceniające podział klastrow na poziomie pikseli (współczynniki pominiętych i fałszywych detekcji oraz współczynniki niedoszacowania i przeszacowania podziału). Uzyskane wyniki jednoznacznie wskazują na lepszą skuteczność opracowanej metody w procesie podziału klastrow na pojedyncze obiekty (jądra).

IV. System przetwarzania obrazów

Opracowany system o nazwie CHISEL (*Computer-assisted Histopathological Image Segmentation and EvaLuation*) stanowi kompleksowe software'owe narzędzie do komputerowego wspomaganie diagnostyki raka sutka na podstawie ilościowej oceny komórek w wielkoobszarowych obrazach preparatów tkankowych barwionych immunohistochemicznie.

Oprócz typowych procedur przetwarzania obrazów, w systemie została zaimplementowana metoda poprawy jakości segmentacji progowej przebadana eksperymentalnie przez Doktoranta oraz autorski algorytm podziału skupiska obiektów. Po podziale klastra przeprowadzane jest proste przetwarzanie w celu pozbycia się obiektów o powierzchni mniejszej niż 50 pikseli. Takie obiekty są traktowane jako artefakty, co wynika z przeprowadzonej analizy statystycznej i są wykluczone z dalszego przetwarzania.

Dodatkowo, w systemie realizowana jest autorska procedura przetwarzania obrazów wielkoobszarowych (tzw. metoda wieloprzesunięciowa) oraz etap klasyfikacji poprzedzony porównawczymi badaniami eksperymentalnymi.

Metoda wieloprzesunięciowa. Metoda jest związana z koniecznością podziału obrazu wielkopowierzchniowego, czyli np. całej mikromacierzy tkankowej (TMA), czy części wirtualnego preparatu (WSI) na mniejsze fragmenty. Oryginalny pomysł Autora polega tu na zastosowaniu przesunięcia maski podziału na regiony zainteresowania zarówno w pionie, jak i w poziomie, w taki sposób, że wydzielone fragmenty na siebie nachodzą. Rozwiązuje to problem obiektów znajdujących się na granicy różnych obszarów zainteresowania. Kolejnym krokiem metody jest połączenie wielu przetworzonych ROI w celu usunięcia niejednoznaczności z obiektów znajdujących się w nakładających się regionach. Proces ten zwiększa dokładność w przypadku obiektów znajdujących się w pobliżu granicy pojedynczych ROI i częściowo widocznych jąder. W przypadku składania stosuje się mechanizm głosowania większościowego w celu sklasyfikowania każdego piksela. Maski binarne wynikające z każdego ROI są dodawane, aby utworzyć scalony obraz, na którym wyodrębnia się cechy dyskryminacyjne obiektów potrzebne w zadaniu klasyfikacji.

Klasyfikacja. Realizowane w systemie zadanie klasyfikacji składa się z dwóch podzadań. W pierwszym (klasyfikacja wstępna), chodzi o dyskryminację obszarów na zdjęciu, które nie nadają się do dalszej analizy (np. obszary o słabej jakości, obszary rozmyte o niewyraźnych granicach pomiędzy obiektami, czy puste, nie zawierające obrazu tkanki). Rozpoznanie takich obszarów na etapie wstępnym bardzo przyspiesza działanie systemu.

W drugim zadaniu (klasyfikacja obiektów) mamy do czynienia z problemem dychotomii, w którym klasyfikacji podlegają wysegmentowane obiekty ze względu na wybarwienie jako immunonegatywne lub immunopoztywne. Z każdego obiektu zostały wyekstrahowane 103 cechy, których liczba w wyniku zastosowania procedury PCA została finalnie zredukowana do 10 cech. Następnie porównano szereg klasycznych klasyfikatorów (drzewa decyzyjne, SVM, regresja liniowa, k-NN, jednokierunkowe sieci neuronowe) pod względem jakości klasyfikacji z zastosowaniem 5-krotnej walidacji krzyżowej. Do implementacji w systemie wybrano klasyfikator najlepszy którym okazała się SNN z jedną warstwą ukrytą złożoną z neuronów sigmoidalnych i z warstwą wyjściową typu softmax. Miara *accuracy* dla wybranego klasyfikatora uzyskana dla 600 próbek ze zbioru IISPV wyniosła 99,98%,

5. Uwagi krytyczne

Obok niewątpliwie wartościowych wyników, które wymieniałem i które przesądzą o pozytywnej ocenie całości, dostrzegam w recenzowanej rozprawie również niedostatki. Można je ująć w następujące punkty:

1. W badaniach eksperymentalnych zabrakło analizy statystycznej uzyskanych wyników porównawczych. Sama różnica w wartości kryterium jakości metod nie oznacza, że jest to różnica istotna statystycznie, a tylko taka uzasadnia wnioski, że jedna metoda jest lepsza od drugiej. Można tutaj skorzystać z gotowych procedur (sprowadzają się one na ogół do odpowiednich testów statystycznych dla weryfikowanych hipotez) – np. dla klasyfikacji: T.G.Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, *Neural Computation* 10: 1895-1923, 1998.
2. W zadaniu poprawy wyniku segmentacji uzyskanej metodą segmentacji progowej z progiem adaptacyjnym zastosowano procedury bazujące na metodzie aktywnego konturu, rozrostu obszaru oraz algorytm k-means. Wszystkie przytoczone algorytmy są znanymi metodami segmentacji. Interesujący byłby eksperyment porównujący wyniki segmentacji przy użyciu tych metod z poprzedzającą je segmentacją progową (co zrobiono) z wynikami działania tych metod jako samodzielnych algorytmów segmentacji (czego nie zrobiono). Gdyby rezultaty były porównywalne, uzasadnione byłoby pytanie po co segmentować dwukrotnie?
3. Z określeniem „wielkoobszarowy obraz” niewątpliwie związane są wyniki dotyczące interpolacji obrazów i schematu wieloprzesunięciowego (ujęte w tezach I i II). Jak się wydaje, wyniki związane z segmentacją (metody adaptacyjnego progowania, poprawa segmentacji, rozdzielanie skupisk obiektów) (tezy III i IV) mają charakter uniwersalny i nie dotyczą tylko obrazów wielkoobszarowych. Czy w kontekście tego spostrzeżenia jest uzasadnione umieszczenie zwrotu „obraz wielkoobszarowy” w tytule rozprawy?
4. W związku z tezą III można sformułować następujący komentarz. Adaptacyjna segmentacja progowa nie musi być wystarczająca do zliczania jąder komórkowych, jeśli pozostawi nierozdzielone skupiska jąder. W zadaniu klasyfikacji są potrzebne różne cechy obiektów (Autor wykorzystuje 103 cechy pierwotne do klasyfikacji). Do ich wydobycia segmentacja progowa zapewne niewystarcza, stąd procedury polepszania segmentacji. Recenzent nie znalazł w pracy zadania, które kończyłoby się jedynie wyznaczeniem liczby jąder komórkowych. Czy zatem ta teza ma praktyczne znaczenie?

5. Niejasny ostatni akapit na str. 42. Na etapie segmentacji, wynik klasyfikacji obiektów jest nieznan (schemat z rys. 5). Jak zatem można segmentować obiekty w zależności od klasy?
6. W ostatnim akapicie na str. 61 jest zdanie (cyt.) *Opracowana przeze mnie metoda komputerowego wspomagania ilościowej oceny komórek w wielkoobszarowych obrazach preparatów tkankowych raka sutka łączy cechy zarówno podejścia klasycznego, jak i opartego na sztucznej inteligencji. Może być zatem uznana jako metoda hybrydowa. Gdzie Doktorant stosuje opracowane metody AI?*
7. Brak informacji o przyjętej metryce definiującej odległość dla transformaty odległościowej w metodzie dzielenia klastrów.
8. Niezręczny zapis formuł matematycznych, gdyż zostały one wyrzucone poza strukturę tekstu. Wzory powinny być fragmentem zdań, w których występują. Dotyczą je zatem także znaki interpunkcyjne: np. jeśli trzeba, to przecinek po wzorze lub kropka, gdy wzor kończy zdanie.
9. Autor nie podaje, jak wyznaczono optymalną wartość ROI 1000x1000 (str. 50), optymalny klasyfikator (str. 53) i optymalną sieć neuronową (str. 53).
10. Jak klasyfikatory (drzewa decyzyjne, regresja liniowa, SVM, k-NN, SNN) działają jako klasyfikator wstępny? W jednej klasie nie ma wysegmentowanych obiektów – jak wtedy można wyznaczyć np. cechy geometryczne?
11. Metoda PCA nie jest metodą selekcji cech (str. 61).

6. Podsumowanie recenzji - ocena rozprawy

Reasumując stwierdzam, iż mgr inż. Łukasz Roszkowiak wykazał się dużą wiedzą z zakresu obrazowania medycznego oraz cyfrowego przetwarzania informacji obrazowej, a także opanowaniem i właściwym posługiwaniem się odpowiednim warsztatem badawczym. Przedstawiona praca zawiera poprawnie sformułowany i rozwiązany problem badawczy oraz stanowi istotny i wartościowy wkład w dziedzinę komputerowego wspomagania medycznej diagnostyki obrazowej. Zawarte w niej rezultaty obejmujące wszechstronną analizę problemu diagnozowania raka sutka na podstawie obrazów barwionych preparatów tkankowych, algorytmy segmentacji obszarów zainteresowania, komputerowe metody analizy mikromacierzy tkankowych, opracowanie systemu CHISEL do kompleksowej diagnostyki nowotworowej a także przeprowadzenie bogatych badań eksperymentalnych i analiz porównawczych na obrazach rzeczywistych, są oryginalne i zostały opublikowane w 6 pracach naukowych. Uważam, że praca doktorska Pana mgr. inż. Łukasza Roszkowiaka lokująca się w dyscyplinie Biocybernetyka i Inżynieria Biomedyczna spełnia wymogi stawiane pracom doktorskim przez odpowiednią Ustawę i wnioskuję o dopuszczenie jej do publicznej obrony.

Dodatkowo, z uwagi na:

1. wyjątkowo kompleksowe podejście do rozwiązywanego problemu, w którym poprzez zgłębienie wiedzy medycznej w zakresie diagnostyki, prognostyki i oceny leczenia raka sutka, opracowany został zestaw oryginalnych algorytmów do komputerowego wspomagania ilościowej oceny komórek w preparatach tkankowych barwionych immunohistochemicznie (opracowane i przebadane algorytmy obejmujące cały schemat

komputerowej diagnostyki obrazowej, mają w dużej mierze charakter uniwersalny, choć w sferze zastosowań, są dziedzinowo-zorientowane),

2. przeprowadzenie bogatych i różnorodnych badań eksperymentalnych na obrazach rzeczywistych we współpracy międzynarodowej,
3. opublikowanie wyników rozprawy w 6 pracach, w tym w 4 artykułach w czasopismach z listy JCR o łącznym IF=10,038 (we wszystkich pracach Doktorant jest pierwszym autorem i – zgodnie z oświadczeniami współautorów – odgrywał w ich powstaniu rolę dominującą),

wnioskuję o wyróżnienie pracy doktorskiej Pana mgr. inż. Łukasza Roszkowiaka.

M. Kuryś