

POLITECHNIKA WARSZAWSKA

Autoreferat rozprawy doktorskiej

Informatyka

Wydział Elektryczny

Rozprawa Doktorska

Metody analizy kontekstowej wirtualnych preparatów mikroskopowych do wspomagania diagnostyki patomorfologicznej

mgr inż. Żaneta Sylwia Świdarska-Chadaj

Promotor: dr hab. inż. Tomasz Markiewicz, prof. PW

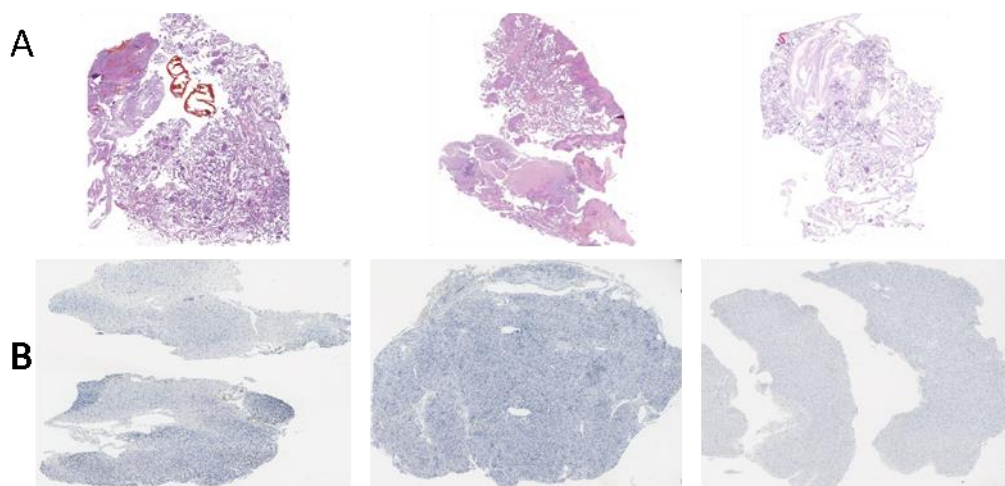
1. Wstęp

Patomorfologia jest dziedziną o kluczowym znaczeniu w diagnozowaniu wielu chorób, w szczególności chorób nowotworowych. Mała liczba czynnych patomorfologów (w Polsce około 300-500 osób), długi czas kształcenia przyszłych ekspertów oraz powiększająca się z roku na rok liczba zlecanych badań histopatologicznych powodują znaczące wydłużenie oczekiwania na wyniki oraz wpływają na przepracowanie ekspertów. Cyfrowa diagnostyka patomorfologiczna stanowi dynamicznie rozwijającą się dziedzinę związaną z analizą i przetwarzaniem zdigitalizowanej postaci preparatów patomorfologicznych.

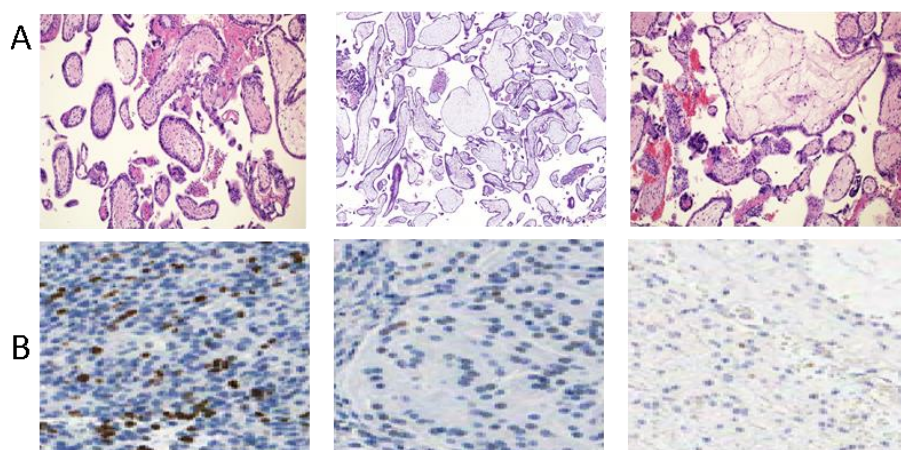
Preparat histopatologiczny stanowi fragment tkanki pobrany przez lekarza podczas zabiegu medycznego. Przygotowanie preparatu histopatologicznego stanowi wieloetapowy, złożony proces prowadzący do uzyskania szkiełka ze skrawkiem zabarwionej tkanki. Następnie uzyskane szkiełko z preparatem histopatologicznym może zostać poddane ocenie mikroskopowej, skanowaniu w celu uzyskania wirtualnego preparatu mikroskopowego lub archiwizacji.

Wirtualne preparaty (WSI) stanowią obrazy preparatów mikroskopowych uzyskane za pomocą skanera szkiełek mikroskopowych lub mikroskopu wyposażonego w kamerę oraz zmotoryzowany stolik. Sposób zapisu obrazu uzależniony jest od producenta sprzętu oraz oprogramowania związanego z rejestracją obrazu. Na rysunku 1 przedstawiono przykłady wirtualnych preparatów. Obrazy WSI cechują się zmiennym charakterem, związanym z różną intensywnością preparatów wynikająca z procesu barwienia, liczby poszczególnych struktur oraz rozmiarów tkanki.

Systemy do wspomagania diagnostyki patomorfologicznej związane są z analizą kontekstową WSI, która ma na celu wydzielenie określonych struktur oraz oszacowanie wskaźników związanych z ich oceną. Analizie poddawane są kolejne pola widzenia, które odpowiadają obszarowi widzianemu w okularze mikroskopu po zastosowaniu odpowiedniego powiększenia. Na rysunku 2 przedstawiono przykładowe pola widzenia.



Rys.1. Przykłady wirtualnych preparatów: A- fragmenty łożysk pochodzących z samoistnych poronień (barwienie H&E) , B- guzy mózgu (barwienie Ki-67).



Rys.2. Ilustracja zmienności pól widzenia: A-preparaty przedstawiające fragmenty łożysk, B- preparaty przedstawiające guzy mózgu.

2. Przegląd zawartości pracy

W pracy przedstawiono dwa kompletne algorytmy do analizy preparatów patomorfologicznych: system wspomagający diagnostykę guzów mózgu barwionych ki-67 oraz system wspomagający diagnostykę fragmentów łożysk pochodzących z samoistnych poronień w pierwszym tryestrze ciąży. Prowadzone badania były konsultowane z doświadczonym zespołem patomorfologów z Zakładu Patomorfologii Wojskowego Instytutu Medycznego w Warszawie.

Praca złożona jest z dziewięciu rozdziałów obejmujących cztery zasadnicze części. Pierwsza część pracy przedstawia złożoność problemu badawczego, tło medyczne oraz proces oceny manualnej preparatów. Zwrócono uwagę na problemy występujące podczas oceny preparatów przez patomorfologów. W drugiej części przedstawione zostały metody analizy i klasyfikacji obrazów, pozwalające na opracowanie złożonych algorytmów.

W trzeciej części pracy przedstawiono zaproponowane, autorskie algorytmy pozwalające na rozwiązanie postawionych problemów. Szczegółowo opisano rozwiązania pozwalające na detekcję i eliminację artefaktów, wpływ zaproponowanej funkcji kary na wybór pól hot-spot

w preparatach przedstawiających guzy mózgu oraz złożony algorytm pozwalający na detekcję i ocenę poszczególnych struktur biologicznych w preparatach przedstawiających fragmenty kosmków. Ponadto, przedstawione zostały metody optymalizacji opracowanych algorytmów poprzez zastosowanie zrównoleglenia obliczeń i obliczeń z wykorzystaniem architektury GPU-CPU. Innowacyjny algorytm umożliwiający komputerową analizę preparatów histopatologicznych zabarwionych Ki-67 przedstawiających guzy mózgu (takie jak skąpodrzewiaki oraz oponiaki) został zaprezentowany w rozdziale piątym. Natomiast w rozdziale szóstym omówiono złożoność problemu badawczego oraz zaproponowano algorytm do detekcji i oceny kosmków oraz metody oceny poprawności detekcji struktur.

W czwartej części pracy zawarto wyniki przeprowadzonych badań oraz wnioski. Rozdział ósmy zawiera informacje na temat zastosowanego materiału badawczego, szczegółowe wyniki przeprowadzonych badań oraz możliwości zastosowania opracowanych metod. Wykazano w nim, które z metod są najbardziej efektywne oraz przedstawiono wyniki weryfikacji algorytmów w odniesieniu do wyników ekspertów. W rozdziale dziewiątym zawarto podsumowanie pracy oraz wnioski wynikające z badań, w szczególności uwypuklono najważniejsze oryginalne osiągnięcia autora rozprawy.

3. Cel i teza pracy

Głównym celem pracy jest opracowanie komputerowych metod wspomagania diagnostyki patomorfologicznej preparatów przedstawiających guzy mózgu oraz fragmenty łożysk pochodzące z samoistnych poronień. Manualna ocena tego typu preparatów przez patomorfologów jest zadaniem złożonym oraz czasochłonnym. Wyniki oceny manualnej obciążone są błędem związanym z warunkami pracy, zmęczeniem, samopoczuciem. Zaletą systemów komputerowych jest odporność na tego typu problemy oraz powtarzalność wyników. Wyniki uzyskane na drodze automatycznej analizy preparatów są w pełni deterministyczne i ustandaryzowane. Zastosowanie automatycznych systemów analizy preparatów pozwoli na zmniejszenie obciążenia patomorfologów oraz zmniejszenie czasochłonności badań.

W pracy przedstawiono nowatorskie rozwiązania wykorzystujące wiedzę z zakresu analizy obrazów, sztucznej inteligencji oraz obliczeń równoległych. Badania związane z opracowaniem systemów realizujących przedstawiony cel pracy zostały podzielone na dwa zadania:

- I. Opracowanie systemu wspomagającego diagnostykę guzów mózgu
 - opracowanie metod wczytywania preparatów zapisanych w różnych formatach (.mrxs, .svs, .vsi);
 - opracowanie metod wstępnego przetwarzania danych (detekcja mapy preparatu);
 - opracowanie metod detekcji i eliminacji artefaktów;
 - opracowanie metod wyboru pól hot-spot (zastosowanie funkcji kary);
 - opracowanie metod detekcji i oszacowania liczby komórek proliferujących;
 - optymalizacja algorytmów z wykorzystaniem obliczeń równoległych;
 - przygotowanie bazy wirtualnych preparatów przedstawiających guzy mózgu barwione Ki-67 (oponiaki i skąpodrzewiaki);
 - ocena poprawności detekcji pól hot-spot przez automatyczny system w odniesieniu do wyników eksperta.
- II. Opracowanie systemu wspomagającego diagnostykę łożysk
 - opracowanie metod wstępnego przetwarzania danych;

- opracowanie metod detekcji struktur takich jak: trofoblast, mezenchyma, kolagen, naczynia krwionośne;
- opracowanie metod oceny poszczególnych kosmków pod względem stopnia unaczynienia oraz ich obrzęku;
- optymalizacja algorytmów z wykorzystaniem obliczeń równoległych;
- przygotowanie bazy wirtualnych preparatów przedstawiających fragmenty łożysk pochodzące z samoistnych poronień barwione H&E;
- ocena poprawności detekcji oraz ewaluacji kosmków przez automatyczny system w odniesieniu do wyników eksperta.

Przedstawione systemy oraz wyniki przeprowadzonych badań pozwalają udowodnić następującą tezę pracy: ***Dedykowane algorytmy przetwarzania obrazów w połączeniu ze strategią obliczeń równoległych pozwalają na opracowanie efektywnych systemów analizy kontekstowej wirtualnych preparatów mikroskopowych wspomagających diagnostykę histopatologiczną.***

4. Podstawowe metody przetwarzania obrazów

Obraz w formie cyfrowej przechowywany jest w pamięci komputera jako tablica. W pracy wykorzystano wiele złożonych zagadnień analizy obrazów obejmujących: różne reprezentacje barw, opisy teksturalne, metodę CORF, metody morfologii matematycznej i statystyczne oraz metody klasyfikacji.

Podstawową różnicą systemów reprezentacji barw jest sposób tworzenia barw. Ze względu na zalety poszczególnych modeli posiadają one różne zastosowania. Analiza spektrum barw, tekstury i kontekstu stanowi podstawę do oceny obrazów przez ludzi. Spektrum barw opisuje średnie odchylenia barwy w różnych pasmach widzialnego i/lub podczerwonego widma elektromagnetycznego, podczas gdy tekstura zawiera informacje o przestrzennym rozmieszczeniu odchyleń barw w paśmie.

Wirtualne preparaty mikroskopowe zapisane są w oparciu o przestrzeń barw RGB. Jest to popularny format zapisu obrazów cyfrowych, w którym najczęściej stosowany jest 24-bitowy zapis kolorów. W celu uzyskania maksymalnego rozróżnienia poszczególnych struktur w pracy zastosowano następujące przestrzenie barw: RGB, CMYK, Luv, YCbCr.

Obrazy cyfrowe można opisać za pomocą parametrów takich jak: wymiary obrazu, rozdzielczość oraz rozmiar. Wymiary obrazu informują o jego długości i szerokości wyrażonej w pikselach. Rozdzielczość to liczba pikseli jaką zawiera linia obrazu o długości jednego cala. Parametr ten jest bezpośrednio związany z jakością obrazu - im większa rozdzielczość tym lepsza jakość, ponieważ więcej pikseli reprezentuje dany element na obrazie. Przykładowy wirtualny preparat mikroskopowy posiada wymiary 61824x93961 pikseli oraz rozmiar 320391497 bajtów. Zastosowanie pełnej rozdzielczości obrazu nie jest konieczne we wszystkich etapach analizy. Zmniejszenie rozdzielczości obrazu powoduje zmniejszenie liczby pikseli, które przedstawiają dany obiekt. Ze względu na znaczący rozmiar wirtualnych preparatów, zastosowanie pomniejszenia obrazu jest istotne. Bardzo ważny jest odpowiedni dobór pomniejszenia tak, aby poszczególne struktury były widoczne na obrazie oraz reprezentowane przez więcej niż jeden piksel. W przypadku preparatów histopatologicznych nowotworów mózgu, pomniejszenie ośmiokrotne pozwala na uzyskanie odpowiedniej widoczności analizowanych komórek przy jednoczesnej, znaczącej redukcji danych, która pozwala na znaczące skrócenie czasu przetwarzania obrazu.

Metody morfologii matematycznej stanowią jedno z podstawowych narzędzi przetwarzania obrazów cyfrowych, ponieważ pozwalają na analizę kształtów, detekcję struktur, konturów, analizę tekstur, rekonstrukcję obiektów oraz filtrację obrazów. Jest to rozbudowana grupa funkcji, których połączenie umożliwia budowę złożonych systemów do analizy obrazów. Ich zaletami są: prostota, szybkość oraz skuteczność.

Operacje morfologiczne opierają się na zastosowaniu elementów strukturalnych (masek wypełnionych zerami oraz jedynkami) o dowolnym kształcie oraz wielkości. Wielkość elementu strukturalnego jest związana z analizą wnętrza obiektu, natomiast kształt powiązany jest z detekcją kształtu obiektu (konturu obiektu). Element strukturalny stanowi rodzaj „ruchomego okna”, które jest przesuwane po powierzchni obrazu tak, aby analizowany piksel był punktem centralnym elementu strukturalnego. Następnie wykonywane są testy logiczne pomiędzy elementem strukturalnym i odpowiednią częścią obrazu, wynikiem czego jest obliczenie nowej wartości dla danego punktu. Istotnym parametrem jest sąsiedztwo określające liczbę punktów wokół danego piksela branych pod uwagę podczas analizy. W pracy zastosowano następujące metody:

- dylację, która prowadzi do zwiększenia rozmiarów obiektów widocznych na obrazie;
- erozję, która prowadzi do zmniejszenia rozmiarów obiektów widocznych na obrazie;
- otwarcie i domknięcie związane są z sekwencyjnym zastosowaniem operacji erozji i dylacji, prowadzące do wygładzenia krawędzi obiektów poprzez usunięcia drobnych połączeń pomiędzy obiektami/ usunięcie dziur i wklęsłości.
- rekonstrukcję morfologiczną, która pozwala na odtworzenie pierwotnych obszarów tylko tych obiektów, które są oznaczone na obrazie maski. Operacja ta związana jest z iteracyjnym wykonywaniem erozji/dylacji obrazu wejściowego;
- metodę rozszerzonych maksimów/minimów lokalnych związaną z detekcją obszarów maksimów/minimów regionalnych (metodą h-maksimów), a następnie rozszerzeniu wykrytych obszarów o punkty sąsiadujące, które spełniają wartość progową.

Metoda CORF (Combination of Receptive Fields) została zastosowana do wykrywania krawędzi obiektów na obrazach. Jej działanie zostało zainspirowane sposobem przetwarzania informacji wizualnych w mózgu, a dokładnie funkcjonowaniem biologicznym komórki, która jest pobudzana/dezaktywowana odpowiednimi bodźcami. Model CORF pozwala na poprawę stosunku sygnału do szumu oraz uzyskanie separacji częstotliwości i orientacji przestrzennej.

Opis numeryczny tekstur stanowi jedną z metod analizy obrazów ponieważ tekstury określają właściwości specyficznego obszaru w obrazie. Lokalny wzorzec jest określany poprzez parametry takie jak: jasność, kierunkowość, porowatość. Jednoznaczny i uniwersalny matematyczny opis tekstur jest trudny do realizacji, dlatego też, do ich opisu stosowane są wektory cech. Cechy obrazu to opisujące go parametry, wśród których wyróżniamy parametry strukturalne, statystyczne, parametry wykorzystujące modele matematyczne oraz parametry związane z filtracją. Do popularnych metod analizy teksturalnej zaliczamy: LBP, cechy Markowa, zmodyfikowane cechy Unsera oraz filtry Gabora.

Local Binary Pattern (LBP) to metoda opisu tekstur wykorzystująca widmo analizowanych jednostek tekstury. Jednostka tekstury stanowi obszar złożony z n -elementów pochodzących z obszaru określonego sąsiedztwa względem analizowanego piksela. Analizie poddawany jest piksel oraz n -punktów znajdujących się w zadanym sąsiedztwie. Następnie obliczana jest różnica pomiędzy wartością piksela środkowego oraz każdego z analizowanych punktów. Uzyskany wynik poddawany jest progowaniu oraz przemnożeniu przez mapę wag. Suma uzyskanych wartości stanowi numer danej jednostki teksturalnej. Istotne znaczenie ma

rozmiar analizowanego sąsiedztwa, liczby analizowanych punktów oraz rozdzielczość analizowanego obrazu. Rozwinięciem metody LBP jest jej połączenie z histogramem ukierunkowanego gradientu (HOG), co znacząco poprawia wykrywanie określonych tekstur.

Opis tekstur za pomocą cech Unsera stanowi rozwinięcie metody zaproponowanej przez M. Haralica (opis tekstur w oparciu o znormalizowaną macierz zdarzeń dla obrazów w skali szarości), poprzez zastosowanie dwóch jednowymiarowych histogramów (histogram sum oraz histogram różnic) jako przybliżenie macierzy koincydencji. Zmodyfikowane cechy Unsera bazują na znormalizowanym prawdopodobieństwie intensywności w histogramach sum i różnic obrazów. Obrazy te są tworzone poprzez zastosowanie transformaty względnej (relative translation). Ostatecznie możliwe jest wyznaczenie ośmiu cech dla danego piksela. Istotnym zagadnieniem jest właściwy dobór rozdzielczości analizowanych obrazów oraz rozmiaru sąsiedztwa tak, aby możliwe było uzyskanie lokalnej charakterystyki analizowanych struktur. W przypadku analiz struktur przedstawiających komórki, powinny być one reprezentowane przez co najmniej kilka pikseli.

Metody statystyczne pozwalają na rozbudowę systemów analizy obrazów poprzez wyznaczenie licznych charakterystyk oraz parametrów statystycznych obrazów i obiektów bazując na histogramie obrazu, macierzy przejść, macierzy gradientu oraz macierzy długości pasm. Parametry statystyczne opisujące dany obiekt mogą być stosowane do jego klasyfikacji oraz oceny. Połączenie operacji morfologicznych, teksturalnych i metod analizy statystycznej pozwala na złożoną analizę obrazów cyfrowych. Do określenia przydatności danych cech do klasyfikacji (różnicowania poszczególnych klas) zastosowano miarę Fishera. Cechy o wysokiej wartości miary Fishera reprezentują duże zróżnicowanie pomiędzy klasami, przez co mogą być stosowane w klasyfikacji obszarów, natomiast cechy o niskiej wartości nie wnoszą istotnych informacji do procesu klasyfikacji.

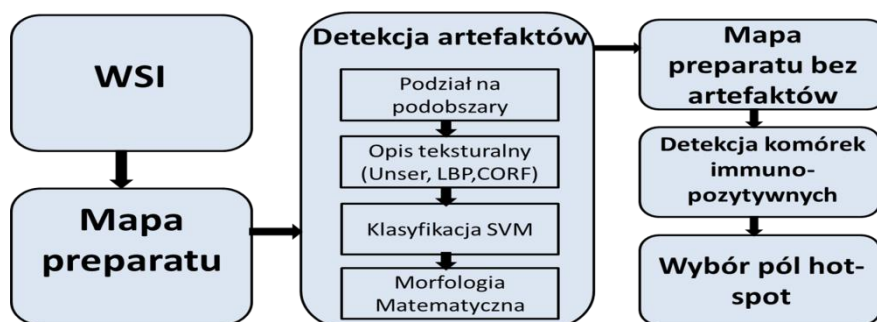
Istotnym aspektem analizy obrazów jest klasyfikacja danych pozwalająca na detekcję, wyodrębnienie oraz ocenę określonych struktur. Wśród sposobów klasyfikacji możemy wyróżnić metody: statystyczne, neuronowe, liniowe, nieliniowe oraz drzewa decyzyjne. W pracy zastosowano metodę wektorów nośnych (SVM), która jest z powodzeniem stosowana w problemach klasyfikacji i aproksymacji. Do zalet metody SVM możemy zaliczyć silne podstawy matematyczne, prawidłowe działanie w przypadku nieliniowych granic decyzyjnych, jak również możliwość uniezależnienia się od wymiaru danych.

Maszyna wektorów nośnych to sieć jednokierunkowa wielowarstwowa, w której możliwe jest zastosowanie różnych typów funkcji aktywacji, jednakże ze względu na jeden neuron w warstwie wyjściowej, pojedyncza sieć umożliwia separację danych tylko na dwie klasy. Celem uczenia sieci jest wyznaczenie lokalizacji hiperpłaszczyzny rozdzielającej dwie klasy danych oraz maksymalizacja marginesu separacji pomiędzy nimi poprzez odpowiedni dobór wartości wag oraz centrów rzutowania (wybrane wektory uczące). Dobór szerokości marginesów ma istotne znaczenie w procesie klasyfikacji. Szerszy margines zapewnia lepsze własności generalizacji oraz mniejszą podatność na przeuczenie, podczas gdy węższy margines może powodować radykalne zmiany klasy podczas klasyfikacji.

4. Automatyczna detekcja pól hot-spot w preparatach nowotworów mózgu

Pierwszy z zaproponowanych systemów to innowacyjny algorytm umożliwiający komputerową analizę preparatów mikroskopowych przedstawiających guzy mózgu (takie jak skąpodrzewiaki oraz oponiaki). Ocena preparatów histopatologicznych przedstawiających

oponiaki oraz skąpodrzewiaki polega na analizie ilościowej (oszacowaniu liczby komórek nowotworowych z uwzględnieniem ich immunoreaktywności) wybranych pól nazywanych hot-spotami. Obszary te cechuje wysoki odczyn immunohistochemiczny oraz zróżnicowana lokalizacja. Liczba pól wybieranych do analizy wynosi od kilku do dwudziestu. Wybierane pola powinny reprezentować zróżnicowane obszary guza o wysokim odczynie immunohistochemicznym tak, aby zapewnić kompleksową ocenę preparatu. Zaproponowany algorytm (rys.3) umożliwia w pełni automatyczny wybór pól do analizy, oszacowanie liczby komórek nowotworowych, jak również określenie indeksu proliferacji.



Rys.3. Schemat algorytmu do wykrywania pól hot-spot.

Liczne problemy występujące podczas przetwarzania obrazów, takie jak znaczny rozmiar danych, występowanie różnego rodzaju artefaktów, zmienność odcieni i różnice w intensywności barwienia preparatów oraz potrzeba uzyskania krótkiego czasu obliczeń zostały rozwiązane w zaproponowanym algorytmie.

Proces detekcji artefaktów takich jak wylewy krwi, zagięcia preparatów oraz zabarwione ściany naczyń wykonywany jest niezależnie dla każdego z wymienionych przypadków. We wszystkich przypadkach zastosowane są metody analizy teksturalnej oraz klasyfikacji.

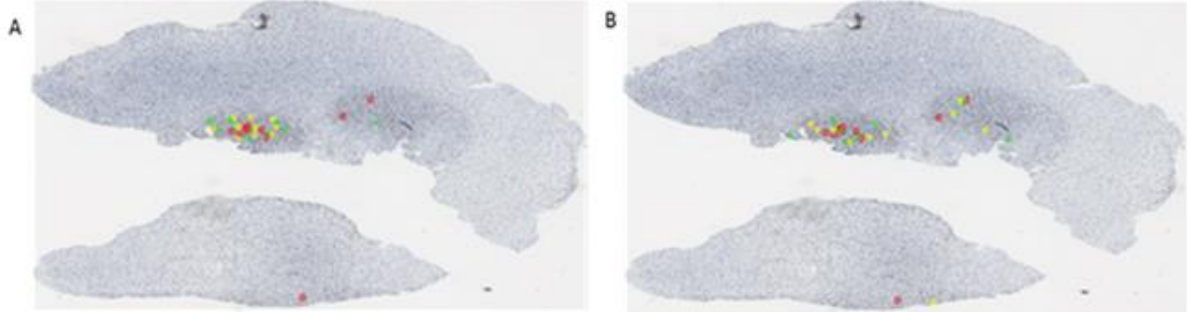
Kolejnym etapem algorytmu jest wykrywanie pól hot-spot związane z detekcją obszarów o wysokim odczynie proliferacji guza oraz zróżnicowanej lokalizacji. Obszary te wybierane są w oparciu o okienkowanie mapy komórek immunopozytywnych. W każdym z okienek zliczana jest liczba wykrytych komórek, a następnie zgodnie z opracowanym kryterium kary wybierane są kolejne pola hot-spot. Ze względu na konieczność uzyskania krótkich czasów obliczeń, jądra komórek wykrywane są na obrazie pomniejszonym ośmiokrotnie, przez co są one reprezentowane zaledwie przez kilka pikseli (4-6) oraz cechują się znacznym rozmyciem granic. Uniemożliwia to zastosowania analizy w oparciu o kształt obiektów lub ich teksturę. Na potrzeby pracy zaproponowano metodę wstępnej detekcji komórek immunopozytywnych w oparciu o zastosowanie metody rozszerzonych ekstremów lokalnych.

Wybór pól zainteresowania związany jest z wartością indeksu proliferacji w poszczególnych polach oraz ich lokalizacją. Algorytm stopniowego wygaszania pól został opracowany w celu wymuszenia automatycznego wyboru pól hot-spot o wysokiej wartości indeksu proliferacji oraz różnorodnej lokalizacji przestrzennej. Pierwszym wybranym polem hot-spot jest obszar o najwyższej liczbie komórek immunopozytywnych. W kolejnych krokach zastosowana zostaje funkcja kary, która modyfikuje wartości oceny poszczególnych pól, powodując zmniejszenie wartości dla pól sąsiadujących z polem już wybranym. Wybór kolejnych pól hot-spot związany jest z wyborem obszarów o najwyższym wskaźniku proliferacji zmodyfikowanym zgodnie z wartością współczynnika kary obliczanego zgodnie z:

$$kara = 1 - \rho \sum_i \frac{1}{\sqrt[4]{(x - x_i)^2 + (y - y_i)^2}}$$

gdzie: x_i, y_i to współrzędne dotychczas wybranych pól hot-spot, ρ - współczynnik kary (wybrany eksperymentalnie i wynoszący 0.2)

Na rys. 4 przedstawiony jest wpływ funkcji kary na lokalizację wybranych pól hot-spot. Celem stosowania funkcji kary jest zamodelowanie sposobu wyboru pól hot-spot przez ekspertów medycznych.



Rys. 4. Wpływ funkcji kary oraz parametru ρ na lokalizację pól hot-spot, gdzie A- wynik dla $\rho=0$, B - wynik dla $\rho=0.2$, czerwony- wyniki eksperta, zielony- algorytm I, żółty- algorytm II.

Istotnym etapem oceny poprawności wyboru pól hot-spot jest porównanie z wynikami ekspertów medycznych. W celu oceny rozmieszczenia pól wybranych automatycznie, w odniesieniu do pól wybieranych przez eksperta medycznego, wprowadzona została miara oceny zgodności lokalizacji (LCM), bazująca na różnicach odległości lokalizacji pól wybranych przez eksperta i algorytm oraz wartości indeksu proliferacji Ki-67 w poszczególnych polach. Wartość miary LCM obliczana jest zgodnie ze wzorem:

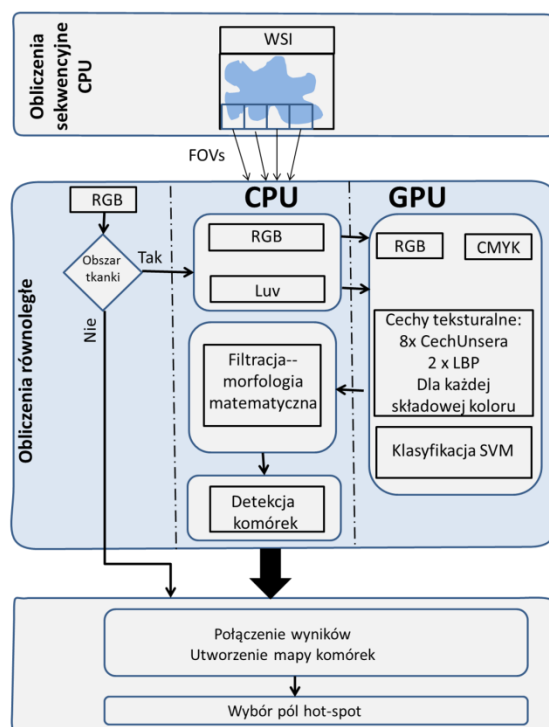
$$LCM = \sum_i \left(w_i * \text{sigm} \left(\min \left| \frac{\text{dist}[(x_j, y_j), (x_i, y_i)]}{4FOV_{SIZE}} \right| \right) \right)$$

gdzie: $w_i = \frac{L_{Ei}}{\bar{L}_E}$

gdzie: L_{Ei} - wartość indeksu proliferacji w polach wybranych przez eksperta, $\bar{L}_E$ - średnia wartość indeksu proliferacji w polach wybranych przez eksperta, FOV_{SIZE} - rozmiar pojedynczego pola widzenia.

Niska wartość LCM świadczy o podobnej lokalizacji obszarów hot-spot wybranych przez algorytm oraz eksperta. Oznacza to, że pola wybrane zarówno przez eksperta jak i algorytm reprezentują te same obszary guza. Wysoka wartość indeksu LCM oznacza, że pola wybrane przez eksperta oraz algorytm reprezentują różne obszary guza. Jeżeli pole ma niższą wartość indeksu Ki-67, to jego pominięcie ma mniejszy wpływ na wartość miary LCM. Proponowana metoda pozwala na ocenę poprawności wyboru pól hot-spot oraz dobór odpowiednich parametrów funkcji kary.

W celu uzyskania krótkich czasów obliczeń opracowany algorytm został zrównoleglony (rys.5). Zastosowanie obliczeń w oparciu o architekturę CPU-GPU pozwoliło na zredukowanie czasu analizy jednego preparatu do około 2 minut.

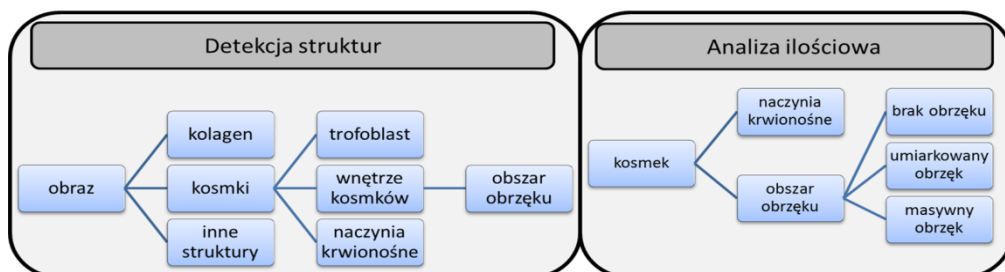


Rys. 5. Schemat zrównoleglenia opracowanego algorytmu w oparciu o architekturę CPU-GPU.

5. Automatyczna analiza obszarów zainteresowania w preparatach przedstawiających fragmenty łożysk pochodzące z samoistnych poronień w pierwszym trymestrze ciąży

Drugi z zaproponowanych systemów związany jest z kompleksową oceną preparatów patomorfologicznych przedstawiających fragmenty łożysk pochodzących z samoistnych poronień w pierwszym trymestrze ciąży. Opracowany algorytm umożliwia detekcję kosmków, trofoblastu, mezenchymy, kolagenu oraz naczyń krwionośnych, jak również analizę poszczególnych kosmków pod kątem unaczynienia oraz stopnia obrzęku. Na rys. 6 przedstawiono opracowany algorytm, który składa się z dwóch głównych kroków:

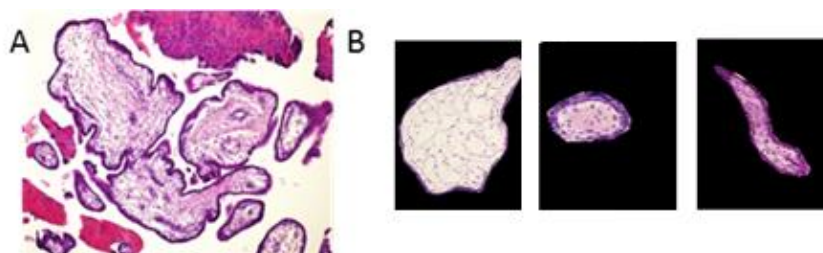
- detekcji struktur związanej z wykrywaniem oraz klasyfikacją poszczególnych obiektów;
- analizy ilościowej związanej z oszacowaniem liczby naczyń krwionośnych oraz określeniem stopnia obrzęku w poszczególnych kosmkach.



Rys.6 Schemat algorytmu do detekcji i oceny kosmków.

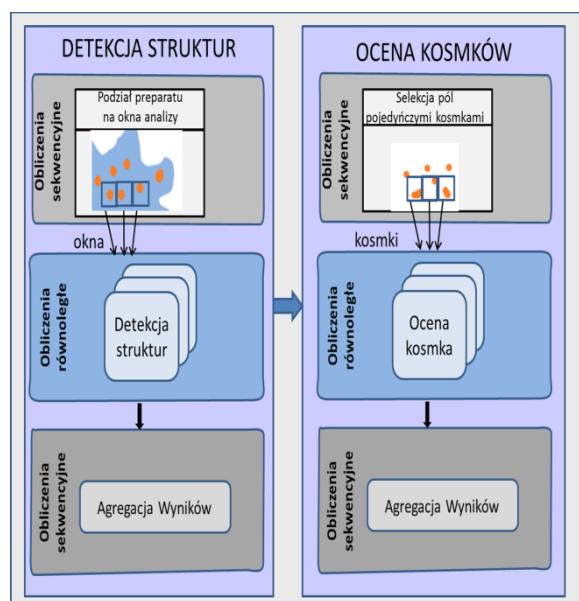
Każdy z głównych kroków algorytmu przedstawionego na rys. 6 związany jest ze złożonym algorytmem opartym o operacje morfologii matematycznej oraz analizę teksturalną. Automatyczna analiza preparatów histopatologicznych przedstawiających fragmenty

kosmków związana jest z licznymi trudnościami, do których można zaliczyć: różnorodność i niejednorodność struktur biologicznych, zmienną intensywność zabarwienia poszczególnych elementów związaną z procesem przygotowywania preparatu mikroskopowego, możliwość występowania artefaktów oraz podobieństwo różnych struktur biologicznych. Należy podkreślić, iż znacznie ograniczona jest możliwość zastosowania parametrów związanych z rozmiarem oraz geometrią np. do opisu i detekcji kosmków, gdyż ich rozmiar oraz kształt mogą być różnorodne (rys.7). Parametry zastosowane w algorytmie zostały dobrane eksperymentalnie podczas licznych badań.



Rys.7. Przykład fragmentu preparatu zawierającego kosmki (A) oraz kosmków o różnych kształtach i rozmiarach (B).

Kolejnym istotnym ograniczeniem jest czas analizy, który powinien być możliwie krótki. Aby sprostać temu wymaganiu niezbędne jest zastosowanie pomniejszania obrazów oraz zrównoleglenia obliczeń. Na rys.8 przedstawiono zastosowany schemat zrównoleglenia, w którym dwa główne kroki algorytmu zostały zrównoległone niezależnie.



Rys.8. Schemat zrównoleglenia algorytmu.

6. Wyniki

W celu zapewnienia wysokiej jakości wyników były one konsultowane z zespołem doświadczonych patomorfologów. Do badań zastosowano preparaty histopatologiczne pochodzące z archiwum Wojskowego Instytutu Medycznego w Warszawie i reprezentują przypadki z lat 2011-2016. Komisja Biotyczna działająca przy Wojskowym Instytucie Medycznym w Warszawie wyraziła zgodę na gromadzenie, przetwarzanie oraz analizowanie niniejszych preparatów.

6.1. Wyniki automatycznej detekcji pól hot-spot w preparatach nowotworów mózgu

Opracowany algorytm automatycznej detekcji pól hot-spot w preparatach nowotworów mózgu został poddany wielu testom, mającym na celu ustalenie skuteczności detekcji i eliminacji artefaktów, ocenę skuteczności detekcji pól hot-spot, jak również porównanie statystyczne wyników działania algorytmu z wynikami otrzymanymi podczas manualnej oceny preparatów w mikroskopie świetlnym oraz wskazaniemi ekspertów na wirtualnych preparatach. Do badań zastosowano preparaty histopatologiczne przedstawiające oponiaki i skąpodrzewiaki wytwarzania rozmazu poddane odczynom immunohistochemicznym Ki-67/MIB-1.

Istotnym etapem analizy preparatów są operacje pre-procesingu związane z eliminacją artefaktów. Najczęściej występującymi typami artefaktów w wirtualnych preparatach guzów mózgu są zagięcia tkanki, wylewy krwi oraz podbarwione ściany naczyń. Do oceny poprawności detekcji zagięć zastosowano jedenaście wirtualnych preparatów przedstawiających guzy mózgu. Ze względu na znaczący rozmiar preparatów detekcja zagięć została przeprowadzona na obrazach pomniejszonych ośmiokrotnie, co pozwoliło na znaczące skrócenie czasu obliczeń przy jednoczesnym zachowaniu ich użyteczności. W celu określenia poprawności detekcji wyniki zostały porównane z oceną manualną przeprowadzoną przez eksperta. Ocena poprawności detekcji została przeprowadzona w oparciu o porównanie map zagięć określonych w sposób automatyczny oraz manualny. Do oceny ilościowej zostały zastosowane następujące wskaźniki:

- procent pokrycia map *TPR*, jest to stosunek obszaru prawdziwie pozytywnego do obszaru wskazanego przez eksperta;
- procent zawartości konturów wskazanych przez eksperta *CI* odpowiadający procentowej zawartości konturów zagięć wskazanych przez eksperta w mapie zagięcia wykrytej przez algorytm.

Szczegółowe wyniki detekcji zagięć na obrazach znajdują się w tabeli 1.

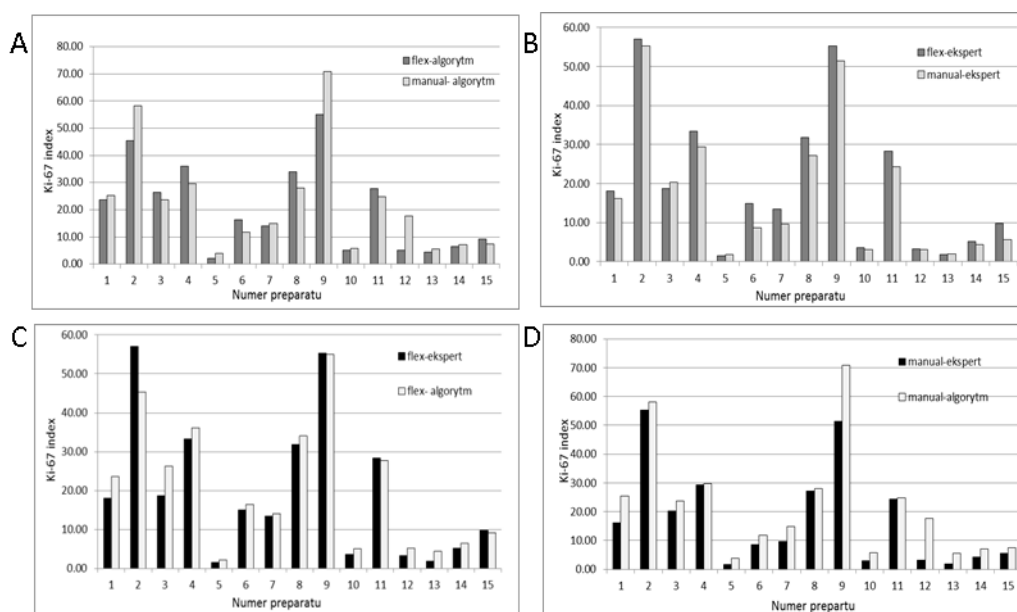
Tabela 1. Wyniki detekcji zagięć

lp.	TPR [%]	CI[%]
1	91.13	68.73
2	100.00	100.00
3	100.00	100.00
4	77.94	63.84
5	69.56	61.14
6	81.89	75.07
7	89.58	83.11
8	100.00	100.00
9	100.00	100.00
10	96.68	96.89
11	89.47	89.81
Max	100.00	100.00
Min	69.56	61.14
Średnia	90.57	85.33
Mediana	91.13	89.81

Uzyskane wyniki świadczą o prawidłowym działaniu procesu automatycznej detekcji zagięć. Średnia wartość pokrycia map zagięć *TPR* wynosi 91%, natomiast średnia wartość pokrycia

konturów CI wynosi 85%. Widoczna różnica pomiędzy parametrami TPR oraz CI związana jest z brakiem spójnych granic pomiędzy zagięciami a prawidłową tkanką. Przedstawione rozwiązanie cechuje krótki czas obliczeń oraz powtarzalność wyników, dzięki czemu może zostać zastosowane jako etap pre-procesingu podczas analizy wirtualnych preparatów przedstawiających guzy mózgu.

Preparaty histopatologiczne mogą być poddawane odczynom w procesie automatycznym oraz metodą manualną. Metoda automatyczna umożliwia uzyskanie lepszej jakości preparatów oraz jednorodnego zabarwienia. W celu oceny wpływu sposobu przygotowania na ocenę preparatów (ocenę indeksu proliferacji) przeprowadzono następujące badanie: piętnaście przypadków oponiaków oraz skąpodrzewiaków zostało oznaczonych metodą manualną (zastosowano odczynnik clone MIB-1, Dako, kod M7240) oraz automatycznie w Dako Autostainer Link (zastosowano odczynnik ready-to-use FLEX Ki-67 antigen, Dako, kod IR626). Uzyskane preparaty zostały zeskanowane oraz zapisane w postaci WSI. Skany zostały ocenione pod kątem wartości indeksu proliferacji przez doświadczonego patomorfologa oraz w sposób automatyczny, przez opracowany algorytm. Na rysunku 9 przedstawiono szczegółowe wyniki w formie graficznej.



Rys.9. Analiza wpływu procedury przygotowania preparatów na ocenę indeksu proliferacji, gdzie: *flex*- metoda automatyczna, *manual*- metoda ręczna, A- wyniki oceny indeksu proliferacji w sposób automatyczny dla preparatów przygotowanych manualnie oraz automatycznie, B- wyniki oceny indeksu proliferacji przez eksperta dla preparatów przygotowanych manualnie oraz automatycznie, C- wyniki oceny indeksu proliferacji w sposób automatyczny oraz przez eksperta dla preparatów przygotowanych automatycznie, D- wyniki oceny indeksu proliferacji w sposób automatyczny oraz przez eksperta dla preparatów przygotowanych manualnie.

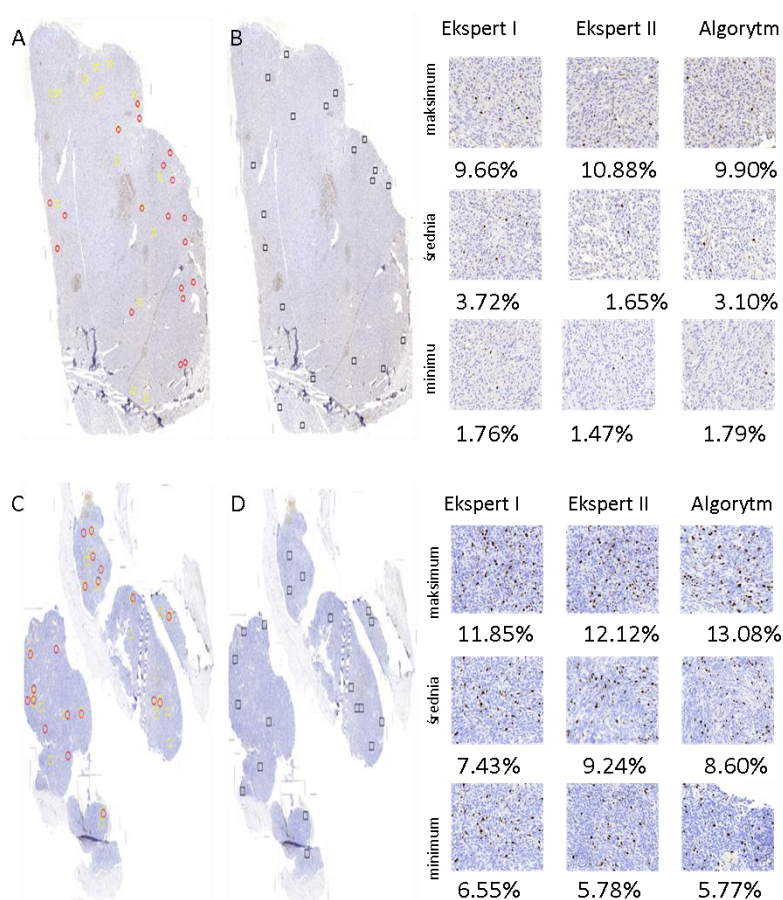
Uzyskane wyniki wykazały, że w przypadku oceny manualnej wyższe wartości indeksu proliferacji (średnio o 2.3 %) uzyskano dla odczynów wykonanych automatycznie, natomiast w przypadku oceny automatycznej wyższe wyniki (średnio o 3.5%) uzyskano dla odczynów wykonanych manualnie. Zarówno dla przygotowania manualnego, jak i automatycznego preparatów, wyższe wartości wskaźnika proliferacji uzyskiwano opracowanym algorytmem. Świadczy to o znaczącym wpływie procesu przygotowania odczynu na manualną ocenę preparatów, w której bardzo duże znacznie ma ich jakość oraz intensywność reakcji. Opracowany algorytm wykazał się stabilnym działaniem niezależnie od sposobu

przygotowania preparatów, co świadczy o jego uniwersalności i możliwości zastosowania niezależnie od metody wykonania odczynu.

W celu oceny poprawności detekcji pól hot-spot i określenia wskaźnika proliferacji Ki-67 w preparacie przeprowadzone zostały szczegółowe badania obejmujące porównanie rozmieszczenia oraz wartości indeksu proliferacji w polach wybieranych w sposób:

- automatyczny przez opracowane algorytmy,
- semiautomatyczny polegający na manualnym wyborze pól hot-spot przez patomorfologa w wirtualnym preparacie i automatycznym zliczeniu liczby komórek,
- manualny, przeprowadzony przez doświadczonego histopatologa przy użyciu mikroskopu.

Na rysunku 10 przedstawiono przykłady rozmieszczenia pól hot-spot wybranych przez opracowany algorytm oraz dwóch patomorfologów wraz z wartością indeksu proliferacji dla specyficznych pól.



Rys.10. Wyniki detekcji pól hot-spot dla dwóch preparatów, gdzie: A i C- pola hot-spot wybrane przez dwóch ekspertów medycznych zaznaczone przez czerwone i żółte kwadraty, B i D- pola hot-spot wybrane przez algorytm. Dla każdego z przypadków przedstawiono trzy pola hot-spot odpowiadające maksymalnej, średniej oraz minimalnej wartości indeksu proliferacji.

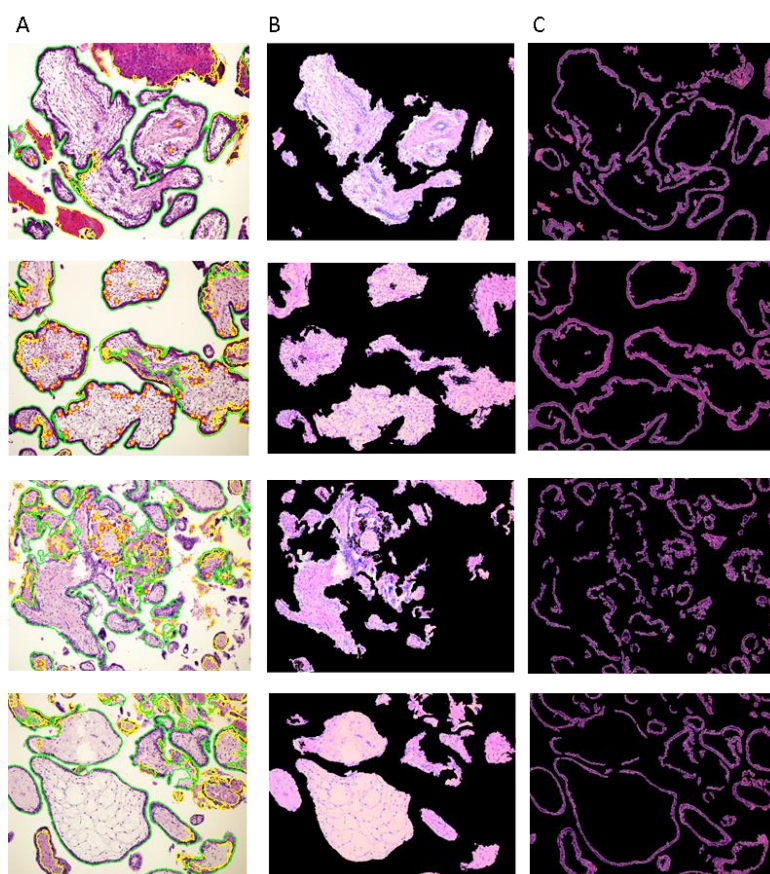
W pierwszym analizowanym przypadku (rys. 10 A i B) widoczne są duże rozbieżności w lokalizacji wybranych pól zarówno pomiędzy ekspertami, jak i ekspertem a algorytmem. Wynika to z niskiego indeksu proliferacji guza, rzędu 1-5 % w większości obszarów oraz incydentalnie występujących pól o proliferacji na poziomie 10%. Pola wybrane zarówno przez ekspertów jak i algorytm mają podobne wartości indeksów proliferacji. W drugim przypadku (rys. 10 C i D) widoczna jest wysoka zgodność dystrybucji pól hot-spot

wskazanych przez ekspertów oraz wybranych przez algorytm. W preparacie występują lokalne maksima proliferacji guza, które zostały wykryte zarówno przez ekspertów jak i opracowany algorytm.

6.2. Wyniki analizy kontekstowej fragmentów łożysk

Opracowany algorytm automatycznej analizy obszarów zainteresowania w preparatach przedstawiających fragmenty łożysk pochodzące z samoistnych poronień w pierwszym trymestrze ciąży został poddany wielu testom, mającym na celu ustalenie skuteczności detekcji określonych struktur oraz porównanie oceny automatycznej z oceną manualną. Preparaty stosowane do badań zostały zabarwione w sposób standardowy, przy użyciu Hematoksyliny i Eozyny. Wybrane przez patomorfologa fragmenty preparatów zawierające kosmki zostały poddane przetwarzaniu i analizie ilościowej.

Ze względu na diagnostykę medyczną najistotniejsza jest prawidłowa detekcja poszczególnych kosmków, które są poddawane dalszej analizie (ocenie unaczynienia oraz stopnia obrzęku). Do porównania obszarów wykrytych w sposób automatyczny oraz wskazanych przez histopatologa zastosowano następujące parametry: TPR- czułość, POA- procentowy obszar przeszacowania oraz PUA- procentowy obszar niedoszacowania. Wizualizację detekcji określonych struktur przedstawiono na rysunku 11 podczas gdy szczegółowe wyniki detekcji kosmków zawarto w tabeli 2.



Rys.11. Wyniki detekcji struktur występujących na obrazach zawierających fragmenty łożysk: A) oryginalne obrazy z zaznaczonymi obszarami: kosmków na zielono, kolagenu na żółto, naczyń krwionośnych przy użyciu symbolu gwiazdki, B) wynik segmentacji obszaru mezenchymy kosmków, C) wynik detekcji obszaru trofoblastu.

Tabela 2. Wartości parametrów opisujących dokładność automatycznej detekcji całych kosmków w odniesieniu do wyników eksperta.

lp.	TRP [%]	POA [%]	PUA [%]
1	98.19	6.94	1.81
2	99.51	10.41	0.49
3	79.75	7.85	20.25
4	99.28	32.81	0.72
5	100.0	18.13	0.00
6	100.0	25.17	0.00
7	99.98	3.96	0.02
8	99.62	13.17	0.38
9	99.98	16.07	0.02
10	99.96	5.01	0.04
11	99.94	7.38	0.06
12	99.98	29.31	0.02
13	99.70	8.40	0.30
14	99.44	11.35	0.56
15	99.59	10.01	0.41
16	99.78	6.73	0.22
17	99.37	7.02	0.63
18	98.63	15.28	1.37
19	99.89	13.47	0.11
20	98.72	14.03	1.28
Średnia	98.57	13.13	1.43
Odchylenie standardowe	4.35	7.76	4.35

Uzyskane wyniki liczbowe potwierdziły prawidłową detekcję struktur takich jak kosmki oraz mezenchyma na poziomie dokładności akceptowanej w diagnostyce histopatologicznej.

7. Podsumowanie i wnioski końcowe

Problem analizy kontekstowej wirtualnych preparatów mikroskopowych do wspomagania diagnostyki patomorfologicznej jest zadaniem złożonym wymagającym interdyscyplinarnych badań oraz ścisłej współpracy z doświadczonym zespołem patomorfologów. Ze względu na różnorodność preparatów histopatologicznych oraz procedur medycznych, każdy z przypadków chorobowych wymaga opracowania dedykowanych, innowacyjnych algorytmów. Istotnym problemem jest znaczący rozmiar preparatów oraz konieczność uzyskania możliwie krótkich czasów obliczeń. Wymaga to kreatywnego połączenia najnowszych technik analizy obrazów, autorskich rozwiązań, zrównoleglenia obliczeń oraz ścisłej współpracy z ekspertami medycznymi. Opracowane algorytmy wykorzystują metody cyfrowego przetwarzania obrazów, analizy teksturalnej oraz metody sztucznej inteligencji.

W pracy przedstawiono innowacyjny algorytm umożliwiający komputerową analizę preparatów histopatologicznych zabarwionych Ki-67 przedstawiających guzy mózgu (takie jak skąpodrzewiaki oraz oponiaki). Algorytm ten umożliwia w pełni automatyczny wybór pól do analizy, oszacowanie liczby komórek nowotworowych, jak również określenie indeksu proliferacji. Jest on odporny na występowanie artefaktów takich jak: zagięcia preparatu, zabarwione ściany naczyń krwionośnych oraz wylewy krwi. Na oryginalność zaproponowanych rozwiązań składają się następujące czynniki:

- Możliwość wczytania preparatów zapisanych w różnych formatach (np. mrsx, svs, vsi, ndpi). Jest to istotna cecha rozwiązania, ponieważ brakuje jednolitych standardów zapisu skanów preparatów mikroskopowych, co powoduje dowolność i różnorodność formatu danych stosowanych w różnych urządzeniach.

- Znacząca odporność na różnice w barwieniu preparatów, które są powszechnym zjawiskiem, mającym istotny wpływ na wynik analizy obrazów mikroskopowych.
- Eliminacja artefaktów występujących w preparatach.
- Zastosowanie innowacyjnej funkcji kary (metoda stopniowego wygaszania pól) w celu wymuszenia detekcji pól do analizy w całym preparacie. Funkcja ta stanowi nowatorskie rozwiązanie, mające na celu wybór pól o wysokim odczynie immunohistochemicznym oraz zróżnicowanej lokalizacji.
- Zastosowanie pomniejszenia obrazów podczas jego przetwarzania w celu skrócenia czasu analizy pod kątem wyboru pól zainteresowań bez jednoczesnego zmniejszenia jej dokładności, jednocześnie wykonując ostatni etap przetwarzania w postaci segmentacji i różnicowania poszczególnych komórek nowotworowych na obrazach w pełnej rozdzielczości.
- Implementacja nowatorskich metod przetwarzania obrazów oraz rozpoznawania i klasyfikacji wzorców w połączeniu ze zrównolegleniem obliczeń oraz wykonywaniem części obliczeń w oparciu o procesory graficzne (GPU) umożliwiła uzyskanie wyników analizy całych preparatów w czasie około 90 sekund.

Drugim obszarem prowadzonych badań było opracowanie algorytmów pozwalających na ocenę preparatów patomorfologicznych (barwionych H&E) przedstawiających fragmenty łożysk, pochodzących z samoistnych poronień w pierwszym tryestrze ciąży. Obrazy te są niezwykle trudne w analizie ze względu na różnorodność struktur biologicznych znajdujących się w preparatach. Badanie przeprowadzane jest w celu określenia potencjalnych przyczyn poronienia, takich jak: zaburzenia angiogenezy, wymiany osmotycznej, wady rozwojowe etc. Do analizy tychże preparatów niezbędne było opracowanie złożonych algorytmów wykorzystujących najnowsze metody analizy obrazów oraz tekstur. Opracowane rozwiązania obejmują sposoby detekcji kosmków, trofoblastu oraz naczyń krwionośnych. Istotnym problemem było utworzenie algorytmu odpornego na zmiany w kolorystyce obrazów, wynikające z różnej intensywności zabarwienia preparatów. Prawidłowa detekcja naczyń nastręczała wiele trudności ze względu na ich różnorodną wielkość oraz kształt. Z tego względu niezbędne było zastosowanie wielostopniowego wyszukiwania naczyń. Istotne znaczenie ma występowanie kolagenu oraz wylewów krwi, które mogą zostać błędnie sklasyfikowane jako naczynia (dokładniej jako krew wewnątrz naczyń). Dlatego też wyszukiwanie naczyń musi zostać zawężone do obszaru kosmków z wyeliminowanymi wylewami krwi. Nowatorskie połączenie różnorodnych metod analizy obrazów z wielkstopniową klasyfikacją umożliwia prawidłową detekcję szukanych struktur.

Liczne przeprowadzone badania wykazały skuteczność działania opracowanych algorytmów, co dowodzi tezie przedstawionej w pracy. Wysokie wskaźniki zgodności ocen ilościowych świadczą o możliwości i zasadności zastosowania zaproponowanego rozwiązania, jako narzędzia wsparcia w diagnostyce patomorfologicznej. Zastosowanie metod zrównoleglenia obliczeń oraz przeprowadzania obliczeń w oparciu o architekturę GPU-CPU pozwoliło na znaczą redukcję czasu obliczeń. Zaletą opracowanych algorytmów jest nie tylko krótki czas działania, ale również zagwarantowanie oceny całego preparatu oraz powtarzalność wyników. Opracowane algorytmy zostały udostępnione w internetowej platformie (<https://miap.wim.mil.pl/>), której celem jest wspieranie pracy patomorfologów oraz wymiana wiedzy i doświadczeń. Platforma umożliwia zdalną analizę wirtualnych preparatów mikroskopowych przedstawiających określone przypadki medyczne.