Uniwersytet Zielonogórski Wydział Informatyki, Elektrotechniki i Automatyki Instytut Sterowania i Systemów Informatycznych



Rozprawa doktorska

# Głębokie sieci neuronowe w klasyfikacji obrazów medycznych

Mgr inż. Marcin Skobel

Promotor: Dr hab. inż. Marek Kowal, prof. UZ

2023

Zielona Góra, Polska

Chciałbym wyrazić szczególne wyrazy podziękowania dla promotora mojej rozprawy, Dr hab. inż. Marka Kowala, prof. UZ, za niezastąpioną pomoc w procesie realizacji mojej rozprawy, przekazywanie wiedzy, zaangażowanie oraz za owocną współpracę naukową.

Chciałbym złożyć serdeczne podziękowania Prof. dr hab. inż. Józefowi Korbiczowi za cenne wsparcie naukowe oraz wskazanie właściwych ścieżek w mojej pracy oraz rozwoju, a także za opiekę naukową.

Chciałbym serdecznie podziękować Prof. dr hab. inż. Dariuszowi Ucińskiemu za umożliwienie współpracy dydaktycznej oraz rozwijania moich umiejętności w obszarze uczenia maszynowego.

Dziękuje wszystkim, z którymi miałem przyjemność prowadzić badania i publikować ich wyniki: Dr hab. inż. Arturowi Gramackiemu, prof UZ, Dr Michałowi Żejmo, Dr n. med. Romanowi Monczakowi, Prof. dr hab. inż. Andrzejowi Obuchowiczowi oraz Mgr inż. Norbertowi Nowickiemu. Dziękuję Prof. dr hab. inż. Krzysztofowi Patanowi za cenne uwagi podczas prezentacji wyników badań.

Chciałbym również złożyć podziękowania mojej rodzinie, zwłaszcza mojej żonie Basi, rodzicom Janinie i Janowi, teściowej Joli oraz rodzeństwu Justynie, Dawidowi i Joli. Wasze wsparcie i cierpliwość były nieocenione, a bez waszej obecności i wsparcia nie byłoby mnie tam, gdzie jestem.

## Spis treści

1	Wstęp			11	
	1.1	Motyw	racja	11	
	1.2	Cele p	racy	13	
	1.3	Teza .		16	
	1.4	Analiza	a stanu wiedzy	16	
		1.4.1	Wstęp	16	
		1.4.2	Segmentacja	16	
		1.4.3	Klasyfikacja	18	
		1.4.4	Klasyfikacja na wielu zbiorach danych $\hfill \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	20	
		1.4.5	Łączenie cech $\hdots$	21	
	1.5	Wykaz	najważniejszych osiągnięć	22	
		1.5.1	Ogólny schemat procesu klasyfikacji	22	
		1.5.2	Wykorzystanie segmentacji do normalizacji obrazów	22	
		1.5.3	Pozyskiwanie cech głębokich oraz fuzja cech $\ .\ .\ .\ .\ .$ .	22	
		1.5.4	Wybór cech istotnych diagnostycznie $\ . \ . \ . \ . \ . \ . \ . \ .$	23	
	1.6	Strukt	ura pracy	24	
<b>2</b>	Obrazowanie medyczne				
	2.1	Obraz	cyfrowy	25	
	2.2	Zbiory	$\operatorname{danych}\nolimits . \ . \ . \ . \ . \ . \ . \ . \ . \ . $	27	
		2.2.1	Zestaw obrazów cytologicznych ze Szpitala w Zielonej Górze $\ .\ .\ .$	27	
		2.2.2	Zestaw obrazów histopatologicznych Break His $\hfill\hfill$	28	
		2.2.3	Połączenie danych $\hfill \ldots \hfill \hfill \ldots \hfill \ldots \hfill \hfill \ldots \hfill \$	29	
	2.3	Wstępi	nie przetwarzanie obrazów cyfrowych	32	
	2.4	Podsur	nowanie	33	
3	Normalizacja obrazów przez segmentację				
	3.1	Wprow	vadzenie	34	
	3.2	Segme	ntacja przy użyciu spłotowej sieci neuronowej	34	
	3.3	Hybryd	dowy system segmentacji	36	
	3.4	Metod	y ewaluacji	38	

	3.5	Weryf	ikacja dokładności hybrydowej metody segmentacji	. 41
		3.5.1	Implementacja hybrydowej metody segmentacji $\ .\ .\ .\ .$ .	. 41
		3.5.2	Wyniki segmentacji jąder komórkowych	. 45
	3.6	Podsu	mowanie	. 51
4	Kon	npleks	owy system klasyfikacji	57
	4.1	Wprov	vadzenie	. 57
	4.2	Ekstra	akcja cech manualnych	. 57
	4.3	Ekstra	ukcja cech głębokich	. 60
		4.3.1	Uczenie maszynowe	. 60
		4.3.2	Elementy sztucznych sieci neuronowych	. 61
		4.3.3	Splotowe sieci neuronowe	. 62
		4.3.4	Metody uczenia zespołowego	. 66
	4.4	Reduk	zcja wymiarowości	. 68
		4.4.1	Ekstrakcja oraz konstrukcja nowych cech	. 68
		4.4.2	Selekcja cech	. 69
		4.4.3	Stochastyczna selekcja cech	. 72
	4.5	Klasyf	ikatory nadrzędne	. 76
	4.6	Klasyf	îkacja sztuczną splotową siecią neuronową	. 76
	4.7	Metod	ly ewaluacji	. 76
	4.8	Wynik	i	. 77
		4.8.1	Schemat przeprowadzonych badań empirycznych	. 77
		4.8.2	Sieci głębokie	. 78
		4.8.3	Opracowany system klasyfikacji	. 92
	4.9	Dysku	sja	. 111
	4.10	Podsu	mowanie	. 113
<b>5</b>	Pod	sumov	vanie	114
	5.1	Wnios	ki	. 115
	5.2	Analiz	a wyników oraz wkład w rozwój dyscypliny	. 115
	5.3	Dalsze	e prace	. 116
$\mathbf{A}$	Stru	ıktura	sieci U-Net do segmentacji	134
В	Kon	figura	cja sprzętowa oraz oprogramowanie	136
С	Wyı	niki do	odatkowych badań	137

## Wykaz skrótów

ACC	Dokładność (ang. Accuracy)
AUC	Pole pod krzywą charakterystyczno-operacyjną ( <i>ang. Area Under The ROC Curve</i> )
CAD	Diagnostyka wspomagana komputerowo (ang. Computer Aided Diagnosis)
CCD	Matryca światłoczuła CCD (ang. Charge-Coupled Device)
CMOS	Matryca światłoczuła CMOS (ang. Complementary Metal-Oxide-Semiconductor)
CNB	Biopsja gruboigłowa (ang. Core Needle Biopsy)
CNN	Splotowa sieć neuronowa (ang. Convolutional Neural Network)
DH	Dystans Hausdorffa
DT	Drzewo decyzyjne (ang. Decision Tree)
DW	Drzewo Wzmacniane
EM	ang. Expectation-Maximization
FNB	Biopsja cienkoigłowa (ang. Fine Needle Biopsy)
GLCM	Macierz współwystępowania poziomów szarości (ang. Gray-Level Co-Occurrence Matrix)
GLRLM	ang. Gray Level Run Length Matrix
GZG	Gruczolakowłókniaki (Szpital w Zielonej Górze)
H&E	Hematoksylina i eozyna
JI	Indeks Jaccarda
k-NN	k - najbliższych sąsiadów (ang. k-Nearest Neighbors)

#### WYKAZ SKRÓTÓW

KN	Klasyfikator Neuronowy
$\mathbf{L}\mathbf{L}$	Las Losowy (klasyfikator)
LOSS	Średni błąd predykcji danych przy użyciu modelu sieci
NB	Klasyfikator Naiwny Bayesa
PCA	Analiza głównych składowych (ang. Principal Component Analysis)
PNG	Format pliku PNG (ang. Portable Network Graphics)
$\operatorname{RL}$	Regresja Logistyczna
ROC	Krzywa charakterystyczno-operacyjna odbiornika ( <i>ang. ang. Receiver</i> Operating Curve)
ROI	ang. Region Of Intrest
SOB	Biopsja wycinająca (ang. Site Of Biopsy)
SSE	Błąd sumy kwadratów reszt (ang. Sum of Squares Error)
$\mathbf{SVM}$	Maszyna wektorów nośnych (ang. Support Vector Machine)
SzUZG	Zbiór danych ze Szpitala Uniwersyteckiego w Zielonej Górze
TIFF	Format pliku TIFF (ang. Tagged Image File Format)
TN	Klasa prawdziwie negatywna (ang. True Negative)
TNR	Swoistość (ang. True Negative Ratio)
ТР	Klasa prawdziwie pozytywna (ang. True Positive)
$\operatorname{TPR}$	Czułość (ang. True Positive Ratio)
VSI	Format pliku wirtualnych slajdów (ang. Virtual Slide Image)
WDS	Współczynnik Dice Sørensena

## Streszczenie

Diagnostyka obrazowa jest jednym z najważniejszych zastosowań sztucznej inteligencji w obszarze medycyny. Mimo obserwowanych obecnie spektakularnych sukcesów sztucznych sieci neuronowych w dziedzinie przetwarzania języka naturalnego i analizy obrazów, istnieją nadal trudne wyzwania, które muszą zostać rozwiązane, aby możliwe było wdrożenie sztucznej inteligencji w rutynową diagnostykę medyczną.

W ramach rozprawy podjęto problem klasyfikacji nowotworów piersi na podstawie obrazów histopatologicznych i cytologicznych. Literatura naukowa dostarcza nam bogata bazę rozwiązań dla tego problemu wskazując na głębokie sieci neuronowe jako aktualnie najlepsze rozwiązanie. Niestety, jeśli przyjrzymy się bliżej proponowanym modelom to spostrzeżemy, że są one najczęściej uczone i testowane na zbiorze obrazów pochodzących z jednego ośrodka medycznego. Przeprowadzone testy wykazały, że wyuczone w ten sposób modele głębokich sieci neuronowych nie są w stanie nabyć odpowiednich zdolności uogólniających, aby mogły być stosowane do klasyfikacji obrazów podchodzących z innych ośrodków medycznych. Przyczyną takiego stanu jest zróżnicowanie pomiędzy obrazami pochodzącymi z różnych ośrodków medycznych. Mimo, że procedury pozyskiwania obrazów są w pewien sposób standaryzowane, to jednak czynnik ludzki oraz techniczny powoduje, że w praktyce obrazy różnią się wieloma istotnymi cechami. Niestety nie jest również możliwe zbudowanie odpowiednio bogatego i różnorodnego zbioru obrazów histopatologicznych lub cytologicznych nowotworu piersi, który zapewniłby uogólnianie wiedzy modeli na poziomie pozwalającym stosować je dla obrazów pochodzących z różnych ośrodków medycznych. Publicznie dostępne zbiory obrazów dla nowotworu piersi są niewielkie, zbyt jednorodne (reprezentują zwykle tylko kilkunastu lub kilkudziesięciu pacjentów) i jest ich stanowczo za mało.

Badania podjęte w rozprawie skupiły się na opracowaniu modelu pozwalającego klasyfikować nowotwory piersi na podstawie obrazów histopatologicznych lub cytologicznych z ośrodków medycznych, których obrazy nie zostały uwzględnione w zbiorze uczącym. Do realizacji tego projektu wykorzystano obrazy medyczne ze zbioru BreakHis z Brazylii oraz obrazy ze Szpitala Uniwersyteckiego w Zielonej Górze (SzUZG). Eksperymenty wykazały, że na pewnym poziomie normalizacji obrazów występuje transfer wiedzy umożliwiający zbudowanie uogólnionego skutecznego systemu klasyfikacji. Wypracowane podejście składa się z czterech głównych rozwiązań. Po pierwsze obrazy poddano standaryzacji z wykorzystaniem zaprojektowanej w tym celu hybrydowej metody segmentacji. Do realizacji tego kroku zaproponowano dwie sieci neuronowe typu U-Net oraz metodę wododziałową. Pierwsza z sieci neuronowych była odpowiedzialna ze segmentację semantyczną obrazów a druga za lokalizację środków poszczególnych jąder komórkowych. Metoda wododziałowa dokonywała fuzji informacji pozyskanych z sieci neuronowych aby dokonać segmentacji poszczególnych instancji jąder komórkowych.

Kolejnym rozwiązaniem, które zaproponowano w ramach rozprawy była fuzja cech manualnych z cechami głębokimi aby zwiększyć odporność opisu obrazów na niejednorodność wewnątrzklasową obrazów. Miało to na celu nabycie przez budowany model zdolności uogólniających wykraczających poza obrazy pochodzące z jednego ośrodka medycznego. W tym celu konieczne było opracowanie zautomatyzowanego systemu do ekstrakcji cech manualnych w oparciu o segmentację jąder komórkowych. Wykorzystano do tego zadania hybrydową metodę segmentacji opracowaną na potrzeby standaryzacji obrazów. System do ekstrakcji cech głębokich powstał w oparciu o heterogeniczny zespół głębokich sieci neuronowych. Członkowie tego zespołu byli strojeni indywidualnie z wykorzystaniem tego samego zbioru obrazów. Ostatecznie z warstw pośrednich głębokich sieci neuronowych wyodrębniono bogate zestawy cech głębokich, które połączono w jeden zestaw wraz z cechami manualnymi.

W efekcie powstał bardzo liczny zestaw cech do opisu obrazów, który należało zredukować do cech istotnych przy klasyfikacji nowotworów piersi. Dlatego w ramach kolejnego kroku badawczego przetestowano wiele znanych metod selekcji cech. Niestety ze względu na duży rozmiar wektora cech i stosunkowo dużą liczbę próbek standardowe metody redukcji wymiarowości okazały się niezmiernie kosztowne czasowo i obliczeniowo. Z tego powodu opracowano nowe rozwiązanie do selekcji cech. Opracowana metoda bazuje na przeszukiwaniu stochastycznym. Na wstępie poszukiwany jest rozkład optymalnej liczby cech na podstawie reprezentatywnego podzbioru obrazów. Pozwala to na etapie właściwej selekcji cech znacząco ograniczyć przestrzeń poszukiwań i w efekcie przyśpieszyć i polepszyć wyniki selekcji cech.

Ostatnim elementem opracowanego systemu jest nadrzędny klasyfikator, który na wejście otrzymuje zestaw cech ustalony przez opracowaną metodę selekcji cech. Rolę tego klasyfikatora pełni model regresji z regularyzacją L2. Model ten został wybrany w toku wielu eksperymentów porównujących efektywność różnych technik uczenia maszynowego.

Zważywszy, że jednym z kluczowych zagadnień poruszonych w rozprawie była weryfikacja skuteczności działania zaproponowanego systemu zbudowanego na danych medycznych pochodzących z innego ośrodka badawczego niż dane testowe, przeprowadzone zostały kompleksowe badania. W wyniku przeprowadzonych eksperymentów wykazano, iż zaproponowana metoda uzyskuje średnią dokładność klasyfikacji obrazów na poziomie 79% co przekłada się na wynik średniej precyzji dla pojedynczego pacjenta na poziomie

#### STRESZCZENIE

przekraczającym 90% dokładności.

Główne osiągnięcia rozprawy doktorskiej obejmują:

- Przygotowanie hybrydowej metody segmentacji obrazów cytologicznych i histopatologicznych na potrzeby standaryzacji obrazów i ekstrakcji cech manualnych jąder komórkowych.
- Przygotowanie systemu ekstrakcji cech do opisu obrazów z wykorzystaniem fuzji cech manualnych z cechami głębokimi uzyskanymi z heterogenicznego zespołu głębokich sieci neuronowych.
- Opracowanie metody stochastycznej selekcji cech z zestawu zawierającego cechy głębokie oraz wyekstrahowane manualnie.
- Przeprowadzanie kompleksowych badań weryfikujących efektywność opracowanych metod dla rzeczywistych obrazów pochodzących ze zbioru BreakHis oraz zbioru obrazów SzUZG.

### Abstract

Medical imaging is one of the most important applications of artificial intelligence in the field of medicine. Despite the currently observed spectacular successes of artificial neural networks in natural language processing and image analysis, there are still difficult challenges that need to be addressed for the implementation of artificial intelligence in routine medical diagnostics.

The aim of the study was to address the problem of classifying breast cancer based on histopathological and cytological images. Scientific literature provides a rich base of solutions to this problem, indicating deep neural networks as the current best approach. However, if we take a closer look at the proposed models, we will notice that they are often trained and tested on a dataset of images from a single medical center. The conducted tests have shown that models trained in this manner, using deep neural networks, are unable to acquire the necessary generalization capabilities to be applied to classify images from different medical centers. Despite the procedures for image acquisition being somewhat standardized, human and technical factors result in significant differences among images in practice. These differences involve several crucial characteristics. Unfortunately, it is also not possible to construct a sufficiently rich and diverse dataset of histopathological or cytological images of breast tumors that would enable the models to generalize knowledge to a level allowing them to be applied to images from different medical centers. Publicly available image datasets for breast tumors are small, too homogeneous (usually representing only a few dozen patients), and insufficient in quantity.

The research conducted in the dissertation focused on developing a model capable of classifying breast tumors based on histopathological and cytological images from medical centers that were not included in the training dataset. To accomplish this project, medical images from the BreakHis dataset from Brazil and images from the University Hospital in Zielona Góra were used. The experiments demonstrated that, to a certain level of image normalization, knowledge transfer occurs, enabling the construction of a generalized and effective classification system.

The developed approach consists of four main solutions. Firstly, the images were standardized using a hybrid segmentation method designed specifically for this purpose. Two U-Net neural networks were proposed for this step, with one responsible for semantic

#### ABSTRACT

segmentation of the images and the other for the localization of individual cell nuclei centroids. The watershed method merged information obtained from neural networks to perform the segmentation of individual instances of cell nuclei.

Another solution proposed in the dissertation was the fusion of manual features with deep features to increase the robustness of image description against intra-class heterogeneity. This aimed to enable the constructed model to acquire generalization capabilities beyond images from a single medical center. For this purpose, an automated system was developed for extracting manual features based on cell nuclei segmentation. A hybrid segmentation method, developed for image standardization, was used for this task. The deep feature extraction system was based on a heterogeneous ensemble of deep neural networks. The members of this ensemble were individually fine-tuned using the same set of images. Ultimately, rich sets of deep features were extracted from intermediate layers of the deep neural networks, which were combined into a single set along with the manual features.

As a result, a very large set of features for describing images was created. This set needed to be reduced to relevant features for breast cancer classification. Therefore, in the next research step, multiple well-known feature selection methods were tested. Unfortunately, due to the large size of the feature vector and a relatively large number of samples, standard dimensionality reduction methods proved to be extremely time and computationally expensive. As a result, a new solution for feature selection was developed. The developed method is based on stochastic search. Initially, the optimal number of features distribution is sought based on a representative subset of images. This allows for a significant reduction of the search space during the actual feature selection phase, resulting in faster and improved feature selection results.

The final component of the developed system is the top-level classifier, which receives the set of features determined by the developed feature selection method as input. This classifier is implemented using a L2 regularized regression model. This model was chosen after conducting numerous experiments to compare the effectiveness of different machine learning techniques.

Considering that one of the key issues addressed in the dissertation was the verification of the effectiveness of the proposed system built on medical data from a different research center than the test data, comprehensive studies were conducted. As a result of the experiments, it was demonstrated that the proposed method achieves an average image classification accuracy of 79%, which translates to an average precision score for individual patients exceeding 90%.

The main achievements of the doctoral dissertation include:

• The development of a hybrid segmentation method for cytological and histopathological image segmentation, aimed at image standardization and manual feature extraction of cell nuclei.

- The development of a feature extraction system for image description using the fusion of manual features with deep features obtained from a heterogeneous ensemble of deep neural networks.
- The development of a stochastic feature selection method from a feature set that includes both deep features and manually extracted features.
- Conducting comprehensive research to verify the effectiveness of the developed methods on real images, sourced from the BreakHis dataset and the University Hospital in Zielona Góra (SzUZG).

## Rozdział 1

## Wstęp

#### 1.1 Motywacja

Sztuczna inteligencja i uczenie maszynowe stanowią dwa hasła, które od lat przykuwają uwagę badaczy i inżynierów oprogramowania na całym świecie. Wraz z rozwojem sprzętu pojawiają się nowe pomysły związane z dziedziną sztucznej inteligencji. Pierwsze systemy inteligentne pojawiły się w połowie XX wieku za sprawą podejścia symbolicznego. Rozwiązania oparte o podejście symboliczne stanowiły algorytmy zbudowane ze zbioru reguł. Początkowo wiązano spore nadzieje z podejściem symbolicznym [10] jednakże współcześnie odchodzi się od metod symbolicznych na rzecz subsymbolicznych takich jak sztuczne sieci neuronowe. Całkowite odejście od metod symbolicznych raczej jest mało prawdopodobne, gdyż wciąż jest rozwijane i przede wszystkim skuteczne, ponadto zastosowanie sieci neuronowych nie zawsze jest w pełni uzasadnione. Do podejścia symbolicznego zaliczmy również tak zaawansowane metody jak algorytmy genetyczne oraz wnioskowanie rozmyte. Współcześnie sztuczne sieci neuronowe stanowią krajobraz codziennego życia.

Sztuczne sieci neuronowe nie są nowym wynalazkiem, co więcej w historii nauki już kilkukrotnie przeżywały swoisty renesans po czym były masowo porzucane na rzecz innych podejść. Obecna fala ponownego zainteresowania sieciami neuronowymi pojawiła się około 2012 roku za sprawą przełomowego wyniku sieci AlexNet [49] w konkursie Image-Net. Współcześnie miano głębokich sieci neuronowych noszą modele zbudowane co najmniej z kilku warstw ukrytych (głębokich). Uczenie głębokie posiada dalsze perspektywy rozwoju, pozwala na to zarówno intensyfikacja prac nad algorytmami jak i gwałtowny postęp w technologii budowy procesorów graficznych, które są dostosowywane do potrzeb uczenia maszynowego. Obecnie możemy już nie tylko mówić o zastosowaniu jednostek cieniujących w kartach graficznych do zrównoleglania obliczeń sieci neuronowych, a wręcz o projektowaniu procesorów tensorowych idealnie przystosowanych do potrzeb uczenia głębokiego. Rozwój metod uczenia głębokiego pozwala również na zastosowanie go w coraz większej liczbie zagadnień i problemów, począwszy od przetwarzania pojedynczych sygnałów, skończywszy na pojazdach autonomicznych. Uczenie głębokie coraz powszechniej używane jest w medycynie. W niniejszej pracy zostanie zaprezentowane praktyczne zastosowanie uczenia głębokiego do wspomagania diagnostyki medycznej, a konkretnie do segmentacji, normalizacji i klasyfikacji obrazów medycznych.

Poprawna diagnostyka jest kluczowym aspektem działalności medycznej. Jednym z najważniejszych działów diagnostyki medycznej jest obrazowanie. Przez lata rozwoju obrazowania medycznego opracowano wiele różnych metod, wśród których można wymienić: radiografie, ultrasonografie, rezonans magnetyczny czy też tomografie. Większość z najpopularniejszych metod obrazowania medycznego jest nieinwazyjna lub mało inwazyjna. W przypadku obrazowania nowotworów piersi w powszechnym użyciu występuje metoda mammografii, która należy do grupy obrazowań przy użyciu promieni rentgenowskich. Mammografia jest badaniem profilaktycznym i przesiewowym, dlatego w przypadku wstępnego wykrycia nowotworu tą metodą należy rozszerzyć zakres diagnostyczny do badań klinicznych obejmujących biopsję. Współcześnie najcześciej stosowanym rodzajem biopsji jest biopsja gruboigłowa (CNB), jednakże powołując się na wytyczne Polskiego Towarzystwa Onkologii Klinicznej [41] w uzasadnionych przypadkach można również stosować biopsję cienkoigłową (FNB). Materiały komórkowe użyte podczas eksperymentów zawartych w niniejszej pracy pochodzą właśnie z FNB oraz CNB. Zaletami procedury FNB jest niska inwazyjność badania, niewielkie koszty wykonania zabiegu, dopracowana technologia. Do wad należy zaliczyć nieco niższą skuteczność diagnostyczną oraz wynikajacy z niej fakt, że obecnie nie jest już złotym standardem klinicznego postępowania diagnostycznego. Pomimo tych faktów, metoda ze względu na swoje zalety jest stosunkowo często wykorzystywana nadal w krajach rozwijających się.

Wspomaganie diagnostyki medycznej jest istotne ponieważ na rynku pracy brakuje wyszkolonych specjalistów, natomiast z roku na rok odnotowuje się wzrost liczby zachorowań na nowotwory na świecie i przewidywany jest ciągły wzrost zachorowań w ciągu najbliższych 30 lat [70]. Choroby nowotworowe stanowią drugą najczęstszą przyczynę zgonów w Polsce po chorobach układu krążenia. Statystyki światowe szacują, że w roku 2020 wykryto 2 261 419 nowych przypadków nowotworu piersi, wśród których 684 996 zakończyło się śmiercią [97]. Jest to najczęstszy rodzaj nowotworu wśród kobiet i stanowi około 24.5% wszystkich nowotworów wykrytych u kobiet w 2020 roku [97]. Wykrywanie nowotworu piersi przebiega w sposób trójetapowy. Pierwszy etap stanowi palpacja. Po wstępnej weryfikacji palpacyjnej następuje badanie mammograficzne lub ultrasonograficzne. Po stwierdzeniu obecności guza następuje skierowanie na biopsję. W tej pracy eksperymenty zostały przeprowadzone na próbkach pochodzących z FNB, dokonywanej przy zastosowaniu igieł o średnicy 0,4 lub 0,5 mm w Szpitalu Uniwersyteckim w Zielonej Górze oraz na próbkach z danych BreakHis [94] pochodzących z Brazylii pobranych w wyniku lumpektomii.

Dodatkową motywacją do podjęcia tego tematu jest również weryfikacja skuteczności klasyfikacji na cyfrowych obrazach pochodzących z różnych ośrodków badawczych, przygotowanych na bazie różnych narzędzi. Potrzeba ta wynika z krytycznych doniesień [58] dotyczących badań biomedycznych. Krytyka dotyczy głównie takich zagadnień jak: brak możliwości wzajemnego porównywania wyników, stosowanie rożnych metryk i technik ich liczenia w publikacjach oraz prawdopodobnie najpoważniejszy zarzut dotyczący osiągania lepszych wydajności poprzez odpowiedni dobór zestawów uczących i testowych [58]. Ponadto wraz z polepszaniem wyników może następować nadmierne dopasowanie do danych uczących, które zwykle skutkuje obniżeniem wydajności modelu dla danych testowych. Wyzwanie, które tu się pojawia polega na trudności w odszukaniu i pozyskaniu cyfrowych obrazów biopsji nowotworu piersi o zbliżonej rozdzielczości oraz jakości.

Klasyfikacja pacjentów na bazie obrazów pobranych w wyniku biopsji na przypadki złośliwe i łagodne to duże wyzwanie dla lekarzy patomorfologów. Praktyka związana z klasyfikowaniem tych obrazów wymaga przynajmniej kilku lat doświadczenia i specjalizacji. W trudnych przypadkach nawet wieloletni staż nie jest wystarczające do postawienia pewnej diagnozy, wówczas niezbędne może być zwołanie konsylium kilku specjalistów i dyskusja nad przypadkiem. Ponadto lekarz oprócz skanu dysponuje preparatem mikroskopowym, ale przede wszystkim możliwością wykonania dodatkowych badań onkologicznych.

Pocieszającym faktem jednak jest to, że liczne badania potwierdzają podobną wartość diagnostyczną wirtualnych slajdów oraz szkiełek pod mikroskopem [14, 17, 19, 24, 29, 107]. Dokładna analiza porównawcza wyników diagnostycznych na bazie cytopatologii cyfrowej oraz diagnostyki na bazie szkiełek mikroskopowych wykazała, że uzyskane diagnozy pokrywają się ze sobą w 97,6% [73]. W badaniach jednak ustalono [107], że diagnostyka na wirtualnym slajdzie zajmuje więcej czasu, niż diagnostyka mikroskopowa na szkiełku, pomijając czas skanowania. Jako zalety cytopatologii cyfrowej wymienia się: możliwość przechowywania próbek na komputerach, możliwość dzielenia się materiałami pomiędzy ośrodkami, szeroki dostęp do materiałów w celach dydaktycznych dla studentów oraz lekarzy w trakcie specjalizacji oraz możliwość wykorzystania cyfrowych obrazów do zastosowania splotowych sieci neuronowych (CNN) w celu wspomagania, rozpoznawania i klasyfikacji nowotworów.

#### 1.2 Cele pracy

CNB jest obecnie standardem w diagnostyce nowotworów piersi, lecz FNB nadal jest stosowana, a ponadto charakteryzuje się mniejszą inwazyjnością. Przewaga diagnostyczna CNB została wykazana eksperymentalnie. Na podstawie danych zebranych z różnych eksperymentów [111] obliczono, że średnia czułość diagnostyki opartej o biopsję cienkoigłową (FNB) wynosi 74% (w przedziale od 53% do 94%), natomiast średnią czułość biopsji (CNB) oszacowano na 87% (w przedziale od 51% do 96%). Bez względu na metodę pobierania materiału kluczowym elementem wykorzystywanym podczas diagnozy są jądra komórkowe. Niestety przygotowanie danych uczących do segmentacji wymaga ogromnych nakładów czasu wynikających z potrzeby ręcznego oznaczania jąder komórkowych na obrazach. Należy się liczyć z koniecznością pracy na ograniczonym zestawie danych uczących. Co więcej przy małych zbiorach danych medycznych konieczne staje się ponowne wykorzystanie danych uczących również do testowania. Aby zatem nie dopuścić do przecieku danych uczących do testowych w trakcie eksperymentów wykonywanych w ramach rozprawy, wykorzystano metodę k-krotnego sprawdzianu krzyżowego w odmianie jednoelementowej (ang. leave-one-out). Zbiory jednoelementowe ograniczały się do obrazów w obrębie jednego pacjenta. Przyjęta metoda walidacji prowadzi do konieczności wielokrotnego powtarzania eksperymentu. W konsekwencji przyjętych założeń sformułowano jeden z celów pracy polegający na zbudowaniu skutecznej i szybkiej metody segmentacji jader komórkowych.

Kolejnym z ważnych problemów poruszanych w pracy jest zagadnienie zdolności uogólniających zbudowanych modeli klasyfikatorów. Praktycznie zawsze w literaturze badanie polega na wykonywaniu eksperymentów na danych pochodzących z jednego ośrodka medycznego. Takie postępowanie może doprowadzić do odnalezienia skutecznej metody klasyfikacji w obrębie obrazów z jednego ośrodka medycznego. Metoda ta jednak na danych pochodzących z innego ośrodka badawczego jest nieskuteczna ze względu na nadmierne dopasowanie danych do modelu. Kolejnym celem postawionym w rozprawie jest **zbadanie skuteczności klasyfikacji na danych pochodzących z różnych ośrodków medycznych**.

Najczęściej klasyfikatory zbudowane na cechach wyodrębnionych manualnie stanowią konkurencję dla metod uczenia głębokiego. Sieć głęboka otrzymuje na wejściu zbiory obrazów uczących w postaci próbek łagodnych oraz złośliwych, następnie na podstawie wyuczonych sztucznie cech nabiera umiejętności klasyfikacji obrazów. Niestety tego typu podejścia w literaturze medycznej określanych mianem "czarno-skrzynkowych" [8] nie wzbudzają pełnego zaufania w środowisku lekarzy [18]. Biorąc jednak pod uwagę w jaki sposób podejście klasyczne oraz sieci głębokie realizują zadanie klasyfikacji, można przypuszczać, że połączenie ich w jeden model poprawi wyniki, choć nie ma co do tego pewności. Kolejnym, więc celem przeprowadzonych badań była **weryfikacja czy fuzja cech manualnych z głębokimi poprawi ogólny wynik klasyfikacji**.

Połączenie cech manualnych z głębokimi wymaga doprowadzenia obu rodzajów cech do wspólnego wymiaru. Połączenie cech manualnych oraz głębokich przynosi jeszcze skutek w postaci powiększenia łącznej liczby cech. Niekiedy bogaty zestaw cech będzie wymagał dalszej redukcji wymiarowości. Z tej przyczyny kolejnym celem pracy była **weryfikacja** 

## działania metod redukcji wymiarowości oraz przygotowanie metody selekcji cech dostosowanej do charakteru zbioru cech powstałych w wyniku fuzji.

Lekarz w trakcie praktyki diagnostycznej w ramach praktyki patomorfologicznej nabywa unikatowych umiejętności klasyfikacji obrazu cytologicznego, niestety potrafi podać jedynie kilka kluczowych cech, które są brane pod uwagę przy klasyfikacji nowotworu. Zatem niezwykle istotnym zadaniem jest rozpoznanie pozostałych cech, które wpływają na wynik klasyfikacji. Posiadane dane wejściowe w postaci obrazów przyporządkowanych do zbioru łagodnego oraz złośliwego oraz brak jednoznacznie określonych cech obrazów wpływających na klasyfikację diagnostyczną, determinuje zastosowanie sztucznych głębokich sieci neuronowych jako głównego narzędzia w procesie ekstrakcji cech. W pracy przedstawiono dwa podejścia do problemu klasyfikacji obrazów cytologicznych. Pierwsze podejście bazuje na ekstrakcji cech głębokich, powstałych w wyniku uczenia sieci poprawnej klasyfikacji zbioru obrazów cytologicznych i histopatologicznych. Drugie podejście rozpoczyna się od segmentacji semantycznej obiektów znajdujących się na obrazie. Następnie na podstawie posegmentowanych jąder komórkowych tworzony jest różnorodny zestaw cech. Jednym z podstawowych podejść do budowania zestawu cech jest ekstrakcja, wchodząca w skład metod redukcji wymiarowości. Dzięki ekstrakcji zmniejszeniu ulega pierwotna przestrzeń obrazu wejściowego do zestawu cech opisujących obiekty znajdujące się w przestrzeni tego obrazu. Cechy mogą zostać wygenerowane w sposób tradycyjny czyli poprzez określenie morfometrycznych, kolorymetrycznych i teksturowych własności jąder komórkowych. Zestaw czynności wykonywanych w tym podejściu stanowi zaprezentowany w tej pracy kompleksowy system klasyfikacji. Do wykonania zaplanowanego celu niezbędne było wykonanie szczegółowych badań weryfikujących najlepsze klasyfikatory, metody redukcji wymiarowości oraz zagadnienie fuzji cech głębokich z manualnymi. W rozprawie przedstawiono wyniki powstałe po wykonaniu kompleksowych badań w celu wyboru najlepszego modelu klasyfikacji.

Należy jednak wspomnieć, że istnieje zbiór cech łatwych do zdefiniowania którymi kieruje się lekarz podczas diagnozowania złośliwości nowotworu. Są to m.in.: wielkość jąder komórkowych, nieuporządkowane rozmieszczenie komórek, nachodzenie na siebie sąsiednich jąder, czy też stosunek wielkości jądra komórkowego do obszaru cytoplazmy go otaczającej. Te określone cechy jąder komórkowych oraz zdobyte przez lekarza doświadczenie warunkują poprawną diagnozę. Kolejnym celem pracy było uzyskanie dokładnej klasyfikacji nowotworu piersi oraz odnalezienie tych cech, które istotnie wpływają na poprawną diagnostykę nowotworu.

#### 1.3 Teza

Cechy jąder komórkowych pozyskiwane są z obrazu cyfrowego, zatem stanowi on pełną przestrzeń cech. Ekstrakcja cech obrazu jest zatem jedną z metod redukcji wymiarowości. Kolejnym etapem może być dalsza redukcja wymiarowości przestrzeni cech przy użyciu metod selekcji lub projekcji cech. W klasyfikacji przy użyciu uczenia głębokiego również mamy do czynienia z redukcją wymiarowości odbywającej się wewnątrz struktury sieci. Na bazie tych informacji można sformułować następującą tezę:

#### Poprawa wyników klasyfikacji cyfrowych obrazów cytologicznych może zostać osiągnięta dzięki utworzeniu przestrzeni cech jąder komórkowych oraz redukcji wymiarów tej przestrzeni do zbioru cech istotnych dla klasyfikacji.

W związku z postawioną tezą zostaną rozpatrzone trzy główne ścieżki klasyfikacji. Pierwsza klasyczna ekstrakcja cech morfometrycznych, kolorymetrycznych i teksturowych w połączeniu z zaawansowanymi metodami redukcji wymiarowości. W tym przypadku sieci głębokie znajdują zastosowanie w dokładnej segmentacji obszarów jąder komórkowych. Drugie podejście będzie bazować na klasyfikacji obrazów nowotworowych z użyciem metod uczenia głębokiego. Zaproponowane zostanie również nowe podejście bazujące na fuzji obu wymienionych ścieżek klasyfikacji.

#### 1.4 Analiza stanu wiedzy

#### 1.4.1 Wstęp

Współczesny gwałtowny wzrost zainteresowania klasyfikacją przy użyciu głębokiego uczenia rozpoczyna się w 2012 roku, gdy grupa badaczy uzyskała najwyższy wynik w konkursie ImageNet używając do tego zadania sieci neuronowej AlexNet [49]. W 2014 roku najlepszy wynik na zbiorze ImageNet uzyskała sieć GoogLeNet [99]. Niestety dalsze zwiększanie głębokości sieci doprowadziło do wystąpienia zjawiska zanikania gradientu. W roku 2015 problem ten został rozwiązany przy zastosowaniu sieci typu ResNet [33] wykorzystującej technologię warstw rezydualnych. Zainteresowanie technologią uczenia głębokiego spowodowało publikację bibliotek programistycznych do uczenia głębokiego takich jak: Keras [10], PyTorch [66], Tensorflow [119] czy Theano [103].

#### 1.4.2 Segmentacja

Wraz z rozwojem technologii skanowania preparatów cytologicznych, umożliwiającej przeglądanie oraz przechowywanie ich w formie cyfrowej pojawiły się wątpliwości dotyczące

#### ROZDZIAŁ 1. WSTĘP

skuteczności diagnostyki na cyfrowych slajdach względem badania mikroskopowego. Przeprowadzone obserwacje potwierdziły jednak podobną skuteczność diagnostyczną obu podejść [107]. Możliwość skutecznego diagnozowania przypadków nowotworów na obrazach cyfrowych otwiera możliwości wykorzystania metod CAD (diagnostyki wspomaganej komputerowo) oraz patomorfologii cyfrowej.

W przypadku przetwarzania obrazów cytologicznych i histopatologicznych, uczenie głębokie jest przydatne w dwóch kluczowych problemach. Pierwszym z nich jest segmentacja obrazów cytologicznych, ze szczególnym uwzględnieniem detekcji jąder komórkowych. Istnieją jednak alternatywne metody segmentacji polegające na przekształceniach morfologicznych bądź algorytmach stochastycznych. Najprostszą metodą segmentacji jąder komórkowych jest progowanie, które oferuje sporo podejść [31] między innymi: metodami progowania adaptacyjnego [57], progowania adaptacyjnego wraz z modelowaniem kształtu jądra komórkowego [69], progowanie oraz wygładzanie [22], algorytm centroidów z progowaniem [5], progowanie z algorytmem EM [1], optymalne progowanie z hierarchicznym przesunięciem średniej [36]. Przy zastosowaniu odpowiedniej obróbki wstępnej skuteczność segmentacji na bazie progowania może uzyskiwać wysokie rezultaty.

Częstym problemem w przypadku segmentacji przy użyciu metod progowania jest niewystarczająca separacja pojedynczych obiektów. Z kolei wykrycie pojedynczych obiektów nazywane jest segmentacją instancji. Rozwiązaniem problemu zarówno niewystarczającej jak i nadmiernej segmentacji jest segmentacja wododziałowa sterowana markerami [108, 109, 12]. Sam proces generowania markerów może być poprzedzony operacjami morfologicznymi [108, 81] jak również innymi metodami np. ze wstępną binaryzacją obszaru jąder komórkowych i wyznaczania punktów erodowania (np. Ultimate Erode Points) [86]. Inne podejścia to segmentacja wododziałowa z progowaniem Otsu oraz transformatą Hougha [106] oraz wododział połączony z hierarchicznym algorytmem centroidów [85]. Segmentacja wododziałowa kontrolowana markerami nadal jest popularna i w sprzyjających warunkach może być niezwykle skuteczna przekraczając poziom 91% dokładności F-score [81]. Dużą popularnością cieszą się również metody segmentacji bazujące na operacjach morfologicznych w połączeniu z innymi technikami takimi jak: progowanie [68, 113], algorytm centoridów [63, 118], PCA [102] czy też transformacja Hougha [72].

Kolejną grupą podejść jakie w ostatnich latach były popularne to metody aktywnych konturów. Idea metody polega na budowaniu granicy obiektu przy użyciu funkcji sklejanych a następnie na przekształcaniu ich w taki sposób, aby zminimalizować wartość funkcji energii zdefiniowanej przez informację o gradiencie [31]. Z przykładów zastosowania aktywnych konturów można wymienić geodezyjny aktywny kontur sterowany algorytmem EM [20] (czułość 0.80, precyzja 0.86) oraz aktywne kontury z zestawem poziomów opartych na interaktywnym modelu [71] (czułość 0.78, precyzja 0.90).

W zadaniu segmentacji obrazów medycznych zastosowanie znalazły również podejścia

bazujące na analizie skupień przy użyciu metody centroidów. Klasteryzacja centroidów wykazała skuteczność segmentacji obrazów histopatologicznych ze zbioru TCGA (*ang. The Cancer Genome Atlas*) na poziomie 97,6% dokładności [64]. Spośród proponowanych rozwiązań rozmyta klasteryzacja centroidów [79] wykazała skuteczność w segmentacji obrazów cytologicznych ze zbioru ISBI 2014 osiągając wynik precyzji na poziomie 0,933 oraz czułości 0,929.

Współcześnie "złotym standardem" zadaniach segmentacji cyfrowych obrazów medycznych są metody bazujące na uczeniu głębokim. Zastosowanie sieci neuronowych może prowadzić do skutecznej segmentacji semantycznej obiektów na obrazie. Niestety sama segmentacja semantyczna może nie być wystarczająca jeśli kluczowym zadaniem jest pozyskanie pojedynczych obiektów czyli segmentacja instancji. Jedno z możliwych rozwiązań polega na wprowadzeniu trzeciej klasy obiektów w postaci granic obiektów [50], co pozwoliło na uzyskanie wyników segmentacji jąder komórkowych nowotworu piersi od 0,7478 do 0,9149. W przypadku segmentacji przy użyciu sztucznych sieci neuronowych szczególnie dobrze spisują się sieci typu U-Net [4, 13, 62, 75]. Rozważane były także połączenia z użyciem wstępnej segmentacji przy użyciu sieci CNN wspomaganej algorytmem wododziałowym [115] oraz z wykrywaniem markerów wododziałów z użyciem sieci neuronowej typu U-Net [48].

W literaturze najnowsze prace badawcze skupiały się na rozbudowywaniu struktury sieci U-Net o znane architektury sieci głębokich [53]. U-Net jest siecią dwuczęściową w układzie enkoder-dekoder, gdzie rolę enkodera może pełnić model bardzo głębokiej sieci (np.: VGG, ResNet, Xception). Własności te zostały zbadane pod kątem jakości segmentacji obrazów medycznych [53, 122, 123]. Równolegle rozwijana jest technologia segmentacji przy użyciu sieci CNN opartych na regionach (R-CNN). Szczególnie przydatne z punktu widzenia obrazów cytologicznych mogą być sieci w wersji *Mask R-CNN* [32], ponieważ ich zadaniem jest segmentacja instancji obiektów. W artykule [53] zaproponowano również zespół klasyfikatorów segmentujących typu U-Net zbudowanych na bazie sieci VGG-19 [87], DenseNet-121 [37] oraz ResNet-101 [33]. Przeprowadzone na świecie badania [53] wykazują porównywalną skuteczność klasyfikatorów U-Net, Mask RCNN oraz przy użyciu zespołu klasyfikatorów na wejściu do U-Net. Uzyskana średnia dokładność Indeksu Jaccarda wykrytych po segmentacji jąder komórkowych wyniosła od 0.53 do 0.54.

#### 1.4.3 Klasyfikacja

Rozwój obrazowania medycznego przyczynił się do pozyskania wysokiej jakości cyfrowych obrazów cytologicznych już na początku lat 90 XX wieku oraz wykonania skutecznej klasyfikacji [96]. W konsekwencji przeprowadzonego wówczas eksperymentu [96] wykorzystano schemat podejścia składającego się z segmentacji jąder komórkowych, pozyskaniu danych morfometrycznych, kolorymetrycznych i teksturowych a następnie wykonaniu na wyekstrahowanych danych klasyfikacji. Dalsze prace [59, 114] doprowadziły do uzyskania zestawu danych zawierających kilkadziesiąt wyekstrahowanych cech. Zestaw ten znany w literaturze jako Breast Cancer Wisconsin nadal stanowi podstawę aktualnych badań naukowych [51, 61, 80, 90, 117].

Kolejne lata doprowadziły do powstania ogromnej liczby różnorodnych rozwiązań w zakresie przetwarzania obrazów cytologicznych i histopatologicznych z użyciem sieci neuronowych, w tym zagadnień związanych z klasyfikacją. Z tego względu podjęto temat systematyzacji istniejących rozwiązań. Jedna z zaproponowanych w literaturze [95] systematyzacji modeli uczenia głębokiego zalicza różne metody do 4 głównych grup: uczenia nadzorowanego, uczenia częściowo nadzorowanego, uczenie nienadzorowanego oraz uczenia transferowego. Przykładami uczenia nienadzorowanego są typowe modele klasyfikacji przy użyciu sieci CNN [33, 109]. Spośród metod częściowo nadzorowanych można ich zastosowanie odnaleźć w literaturze dotyczącej zagadnienia klasyfikacji obrazów histopatologicznych raka piersi i raka okrężnicy [38] oraz w klasyfikacji obrazów histopatologicznych raka płuc [112]. Z kolei przykładem wykorzystania uczenia nienadzorowanego w procesie klasyfikacji obrazów histopatologicznych, a dokładnie wirtualnych slajdów jest artykuł [7] badajacy przypadki guzów nowotworowych gruczołu krokowego. W literaturze [9] można odnaleźć również przykład klasyfikacji nowotworu piersi na obrazach histopatologicznych z użyciem uczenia transferowego z wykorzystaniem metody głosowania zespołu klasyfikatorów bazujących na CNN. Artykuł raportuje uzyskanie wyniku na poziomie 87% dokładności. Innym przykładem wykorzystania uczenia transferowego na obrazach histopatologicznych nowotworu piersi jest artykuł [52] wykorzystujący do tego celu głęboką sieć InceptionResNetV2 raportujący uzyskana dokładność wynoszaca 87%.

Metody uczenia nadzorowanego można podzielić na kolejne dwa główne podejścia. Pierwsze bazujące na ekstrakcji cech obrazów na podstawie, których klasyfikator podejmuje decyzję diagnostyczną [21, 42, 45, 47, 46, 61, 80]. Drugie podejście wykorzystuje sieć CNN do klasyfikacji obrazów [33, 93, 109]. W rozprawie podjęto temat uczenia nadzorowanego w obu podejściach oraz zbadano efekt połączenia tych podjeść.

W podejściach bazujących na uczeniu nadzorowanym często wybieranym skutecznym klasyfikatorem jest maszyna wektorów nośnych (SVM). Nieznacznie gorsze wyniki uzyskiwane są dla innych klasyfikatorów: drzew decyzyjnych (DT), k-najbliższych sąsiadków (k-NN), klasyfikator naiwny Bayesa (NB) [21, 45, 47]. Publikacje [21, 47] wykazały średnią dokładność klasyfikacji na poziomie 75-76%, lecz przy sprzyjających warunkach klasyfikatory przekraczały 80% dokładności. Można zatem przypuszczać, że niemal niewykonalne jest przekroczenie 90% dokładności przy użyciu standardowych klasyfikatorów. Nieco lepiej wyglądają wyniki klasyfikacji opartej o cechy głębokie (dokładność 81-86%) [93], a także klasyfikacja przy użyciu sieci CNN (85-90%) [109].

Niemniej niezwykle interesujące byłoby zbudowanie klasyfikatora, który podobną skutecznością będzie charakteryzował się nie tylko w obrębie danych pochodzących z jednego ośrodka badawczego, ale również z innych placówek. Innymi słowy, w pracy zostanie podjęte badanie polegające na odnalezieniu cech na obrazach medycznych, które staną się podstawą do budowy systemu skutecznie klasyfikującego zarówno dane z tej samej placówki medycznej, ale również dla zewnętrznych danych. W eksperymentach zostaną zatem użyte obrazy pochodzące ze Szpitala Uniwersyteckiego w Zielonej Górze oraz z Laboratorium Anatomii Patologii oraz Cytopatologii w mieście Parana w Brazylii. Zbiory te różni niemal wszystko (metoda pobierania próbki, technologia, sprzęt, położenie geograficzne). Podobieństwa to z kolei fakt, iż prezentują wizualnie obrazy komórek nowotworu piersi oraz zostały poddane działaniu hematoksyliny i eozyny (H&E).

#### 1.4.4 Klasyfikacja na wielu zbiorach danych

Najczęściej w literaturze można spotkać się z badaniami eksperymentalnymi na bazie jednego zbioru danych (z tego samego ośrodka badawczego). Niestety, aby system wspomagania diagnostyki medycznej był w pełni skuteczny powinien wykazywać się dokładnością klasyfikacji dla różnych danych. Jedno z badań nad skutecznością klasyfikacji tej samej metody na różnych zbiorach danych (Herlev [40], HEMLBC [120]) polegało na zastosowaniu głebokiej sieci neuronowej z uczeniem transferowym, w której pierwotne wagi pochodziły z zestawu ImageNet, natomiast sieć była douczana na badanych zbiorach [121]. Badanie wykonano na obrazach cytologicznych raka szyjki macicy pochodzących z różnych metod pobierania wymazu. Klasyfikacja polegała na wykryciu normalnych oraz nieprawidłowych jąder komórkowych. Uzyskany wynik klasyfikacji na obu zbiorach wyniósł 98.3%. W innym badaniu [92] wykorzystano dwa różne zbiory danych (własne dane oraz zbiór Herlev [40]). Po wstępnym przetworzeniu danych utworzono jeden wspólny zbiór, z którego wydzielono dane uczące, walidacyjne i testowe. W kolejnym kroku wykonano weryfikacje skuteczności klasyfikacji 4 różnych sieci głebokich z których najlepszy wynik uzyskano na VGG19 [87]. Dokładność uzyskanego modelu wyniosła 0.8871. W literaturze można również odnaleźć prace [6] dotyczącą klasyfikacji nowotworu piersi nawet na 3 zbiorach danych. Każdy z pozyskanych zbiorów posłużył do utworzenia trzech oddzielnych zestawów danych uczących oraz testowych. Zestawy te wykorzystano do zbudowania klasyfikatorów bazujących na sieci AlexNet [49] z wykorzystaniem uczenia transferowego. Proponowana metoda w zależności od zestawu uzyskała dokładność od 96.7% do 100%. Można więc zauważyć, iż w literaturze autorzy skupiają się na sprawdzeniu zaproponowanej metody na różnych zestawach danych w obrębie tych zestawów lub łączą zbiory w jeden zestaw oraz uczą i testują model na tych danych. Pozostaje zatem nadal otwarte pytanie co się stanie, gdy dane z jednego ośrodka medycznego zostaną wykorzystane do uczenia modelu, natomiast dane z innego ośrodka do jego testowania.

#### 1.4.5 Łączenie cech

Jednym z najważniejszych powodów przemawiających za stosowaniem głębokich sieci CNN jest ich zdolność do samodzielnego wyboru istotnych diagnostycznie elementów obrazu. Cecha ta sprawia, iż pomijany jest czasochłonny proces inżynierii cech. Sieci CNN stanowią zatem alternatywę dla metod manualnej ekstrakcji cech. Nie oznacza to jednak, że generowane przez sieć głęboką cechy obrazu mają swoje odpowiedniki wśród cech manualnych. Co więcej, można stwierdzić, iż zazwyczaj cechy głębokie są abstrakcyjne i istnieje nikłe prawdopodobieństwo ich zbieżności do cech manualnych. Istnieje zatem podejrzenie, że fuzja cech manualnych z głębokimi przyniesie poprawę wyników klasyfikacji.

Idea połączenia cech głębokich z manualnymi została z powodzeniem wykorzystana w zagadnieniu klasyfikacji roślin po liściach [28]. W tym badaniu wykorzystano generator cech głębokich w postaci sieci ConvNet oraz grupę cech manualnych bazujących na kształcie oraz kolorystyce liści. Wyniki klasyfikacji z obu grup cech generowane były przy użyciu lasu losowego (LL). Ostateczny wynik wykazał poprawę wyniku klasyfikacji na utworzonym zespole klasyfikatorów o 6,1% względem klasyfikatora opartego na cechach manualnych oraz o 2,8% względem wyniku uzyskanego na klasyfikatorze bazującym na cechach głębokich.

Badania nad skutecznością klasyfikatorów bazujących na zestawach cech manualnych oraz głębokich w zagadnieniach histopatologicznych wskazują, że najwyższe wyniki klasyfikacji uzyskuje się dla klasyfikatorów bazujących na cechach głębokich oraz porównywalne z nimi wyniki dla klasyfikatorów opartych o fuzję cech głębokich z manualnymi [105]. W badaniu tym zastosowano pozyskiwanie cech głębokich z warstw gęstych różnych sieci neuronowych. Warstwy te posiadały liczbę neuronów od 1024 to 4096, łącznie uzyskując liczbę 17408 cech głębokich. Liczba ta została zredukowana przy użyciu metody PCA. Klasyfikacja czterech typów nowotworów została wykonana na zbiorze CRCHistoPhenotypes [88], który posiada 100 przypadków nowotworów. Autorzy badania [105] raportują uzyskanie wyniku klasyfikacji na poziomie 0,9981 dla zestawu cech głębokich oraz 0,9978 dla fuzji cech głębokich oraz manualnych. Na podstawie istniejących badań można, więc stwierdzić, że fuzja cech głębokich z manualnymi może poprawić wyniki klasyfikacji lub w najgorszym wypadku utrzymać poziom lepszej z dwóch grup cech.

Łącznie cech manualnych z sieciami CNN zostało wykorzystanie w zadaniu wykrywania jąder komórkowych nowotworów na obrazach histopatologicznych [44]. W badaniu tym wykazano poprawę dokładności (F-score) wykrywania jąder komórkowych z 0.709 dla sieci CNN do 0.748 dla metody łączącej cechy manualne z siecią CNN. Innym przykładem łączenia cech manualnych z sieciami CNN jest wykrywania obiektów (jąder komórkowych w czasie mitozy) na obrazach histopatologicznych [110]. W tym badaniu dokładność (Fscore) wykrywania zmian mitotycznych wyniosła 0.5730 dla sieci CNN, 0.6864 dla cech manualnych oraz 0.7345 dla połączonych podejść.

#### 1.5 Wykaz najważniejszych osiągnięć

#### 1.5.1 Ogólny schemat procesu klasyfikacji

Zastosowane podejście stanowi złożony zestaw czynności prowadzących do uzyskania najskuteczniejszego modelu klasyfikacji obrazów medycznych raka piersi. Cały proces (rys. 1.1) można podzielić na trzy główne części: segmentację obrazów, utworzenie przestrzeni cech i selekcję najistotniejszych oraz klasyfikację obrazów.



Rysunek 1.1: Schemat proponowanego podejścia

#### 1.5.2 Wykorzystanie segmentacji do normalizacji obrazów

Pierwsza z nich to segmentacja obrazów, której zadaniem oprócz samej segmentacji semantycznej jest jeszcze normalizacja obrazów przed podaniem ich na wejście modeli sieci klasyfikujących typu CNN. Potrzeba opracowania takiej normalizacji wynika z faktu braku zdolności uogólniających sieci trenowanych na obrazach w przestrzeni RGB. Ostatnim istotnym efektem zastosowania segmentacji jest odrzucenie nieistotnych obrazów, na których nie występują jądra komórkowe wcale lub występują bardzo nielicznie. Opracowana w ramach pracy metoda segmentacji składa się z hybrydowego systemu bazującego na sieciach CNN oraz algorytmie wododziałowym. Segmentacji poświęcono rozdział 3 niniejszej rozprawy.

#### 1.5.3 Pozyskiwanie cech głębokich oraz fuzja cech

Druga część procesu polega na ekstrakcji cech manualnych oraz generowaniu cech głębokich. Ekstrakcja cech wymaga uprzedniego wykonania skutecznej segmentacji instancji jąder komórkowych. Pozyskane w ten sposób pojedyncze jądra komórkowe w postaci binarnej stanowią podstawę do obliczania cech morfometrycznych obiektów a także jako maski wycinające interesujące fragmenty bardziej złożonych obrazów. Generowanie cech głębokich wynika natomiast z faktu, iż podczas eksperymentów dowiedziono, że modele uczone na bazie obrazów po segmentacji mają wyższe zdolności uogólniające niż modele uczone na obrazach w przestrzeni RGB. Do wygenerowania zestawu cech cech głębokich (rys. 1.2) użyto zmodyfikowanej koncepcji zespołu klasyfikatorów. Wykonane modyfika-



Rysunek 1.2: Schemat pozyskiwania cech głębokich

cje względem pierwotnego algorytmu polegają na zbudowaniu homogenicznych próbek wejściowych do klasyfikatorów. Próbki te składają się z zestawu danych uczących i walidacyjnych podzielonych w taki sam sposób, różnią się natomiast modele klasyfikatorów. Zabieg ten pozwala na uzyskanie różnych wartości cech głębokich z przygotowanych zestawów danych z różnych architektur sieci neuronowych. Ostatnim krokiem algorytmu jest usunięcie ostatniej warstwy gęstej w zbudowanych modelach sieci neuronowych oraz wyodrębnienie wartości wyjść neuronów z przedostatniej warstwy. Wartości te następnie są wykorzystywane do zbudowania wektorów cech głębokich a następnie fuzji cech manualnych i głębokich. Szczegółowy opis zagadnienia znajduje się w podrozdziale 4.3

#### 1.5.4 Wybór cech istotnych diagnostycznie

Po otrzymaniu zestawu cech głębokich, manualnych a także wykonaniu ich fuzji łączna liczba cech wyniosła 276. W wyniku przeprowadzonych badań najwyższy wynik klasyfikacji fuzji cech uzyskano dla metod regresyjnych. Aby zbadać jakie liczby cech generują najlepsze wyniki klasyfikacji przeprowadzono eksperyment polegający na losowaniu liczby oraz losowego zestawu cech ograniczonego tą liczbą. Po wielokrotnym powtórzeniu eksperymentu wybrane zostało 100 najlepszych wyników wraz z informacją jaka liczba cech była potrzebna do uzyskania najlepszego wyniku klasyfikacji. Losowanie to stanowi pierwszy etap zaproponowanego algorytmu selekcji cech. Eksperyment dowodzi, iż rozkład empiryczny jest zbliżony kształtem do rozkładu Gamma (rys. 1.3). Po estymacji wartości parametrów rozkładu Gamma rozpoczyna się działanie algorytmu stochastycznego. Wewnątrz pętli wykonywany jest wybór długości wektora cech oraz losowany jest zestaw cech ograniczony wylosowaną z rozkładu liczbą. Wybrany wektor cech stanowi zestaw danych



Rysunek 1.3: Rozkład empiryczny oraz dopasowany rozkład Gamma

wejściowych do zbudowania modelu regresji logistycznej. W celu uniknięcia nadmiernego dopasowania modelu do danych uczących zastosowano walidację krzyżową danych uczących oraz regularyzację L2. Szczegółowy opis metody znajduje się w podrozdziale 4.4.3.

#### **1.6** Struktura pracy

Praca składa się z pięciu rozdziałów. Pierwszy rozdział stanowi wstęp do omawianego zagadnienia, czyli wprowadzenie do tematyki poruszanej w pracy oraz motywacja do podjęcia problemu, następnie jest omawiany cel pracy. Wstęp zawiera również tezę pracy, aktualny stan wiedzy, główne osiągnięcia pracy oraz omówienie zawartości rozprawy. Drugi rozdział porusza temat obrazowania medycznego, sprzętu użytego do pozyskania obrazów cytologicznych. Rozdział ten zawiera również opis zbioru danych oraz źródło obrazów cytologicznych, a także podejmuje zagadnienia związane z ich wstępnym przetwarzaniem. Trzeci rozdział dotyczy zagadnienia segmentacji obrazu, ze szczególnym uwzględnieniem metod uczenia głębokiego. Rozdział ten opisuje opracowaną metodę segmentacji oraz uzyskane wyniki. Rozdział czwarty z kolei podejmuje zagadnienie klasyfikacji obrazów cytologicznych i histopatologicznych zarówno w ujęciu klasycznym jak i przy zastosowaniu technik uczenia głębokiego. Ponadto w rozdziale tym przedstawiono kompleksową metodę skutecznej klasyfikacji obrazów raka piersi pochodzących z różnych ośrodków medycznych. Po wstępie teoretycznym następuje sekcja z wynikami badań empirycznych wraz z dyskusją i podsumowaniem. Ostatni rozdział to ogólne podsumowanie zawartości rozprawy oraz analiza wyników pod kątem przyszłych prac.

## Rozdział 2

### Obrazowanie medyczne

#### 2.1 Obraz cyfrowy

Rastrowy obraz cyfrowy zbudowany jest z pikseli przyjmujących wartości liczb z określonego zakresu. Piksele są definiowane w przestrzeni obrazu przy użyciu współrzędnych. Obraz cyfrowy można pozyskać utrwalając go przy użyciu przetwornika fotoelektrycznego CCD lub CMOS. Przetworniki CMOS są popularne w urządzeniach multimedialnych i ręcznych aparatach cyfrowych, natomiast w fotogrametrii czy obrazowaniu medycznym częściej wykorzystywane są matryce CCD. Odpowiednio przechowywany obraz cyfrowy jest trwałym nośnikiem zawartych na nim danych w przeciwieństwie do próbek medycznych, które z czasem tracą swoje właściwości biochemiczne. Uzyskanie cyfrowych obrazów próbek pobranych w wyniku biopsji możliwe jest dzięki zastosowaniu skanera wirtualnych slajdów posiadającego matrycę CCD. Wirtualny slajd stanowi cyfrowy skan próbki znajdującej się na szkiełku. Materiał, który jest potrzeby do wykonania eksperymentu składa się zatem z fragmentu ludzkiej tkanki. W obrazowaniu cytologicznym niezbędne jest stosowanie obrazów wysokiej jakości. Kolor, tekstura czy kształt obiektów na obrazie cyfrowym może mieć istotne znaczenie w diagnostyce. Standardowym formatem działania na obrazach cyfrowych w prezentowanych eksperymentach jest format TIFF wraz z nieskompresowanymi obrazami w formacie PNG.

Obrazy stanowiące temat badań zawartych w niniejszej pracy pozyskane zostały dzięki urządzeniu Olympus VS120. Urządzenie skanuje obrazy do formatu zamkniętego VSI, którego obsługę umożliwia przeglądarka OlyVia oraz alternatywna metoda w postaci biblioteki Bio-Formats [56]. Plik VSI składa się z wielu warstw o różnej szczegółowości (rozdzielczości) przechowywanych obrazów. Format ten stanowi skuteczne rozwiązanie problemu ograniczonych zasobów pamięci operacyjnej w porównaniu z obrazem o maksymalnej rozdzielczości. Biorąc pod uwagę, że rozdzielczość obrazu w pełnym powiększeniu wynosi 100 000 na 200 000 pikseli to można obliczyć że obraz zawiera  $2 * 10^{10}$  pikseli,

#### ROZDZIAŁ 2. OBRAZOWANIE MEDYCZNE

a po uwzględnieniu przestrzeni barw RGB otrzymujemy łącznie  $2 * 10^{10}$  pikseli. Wartość ta w przeliczeniu na jednostki pamięci wynosi 55,88 GB, czyli powyżej standardowych wielkości pamięci RAM współczesnych komputerów klasy PC. Należy też zauważyć, że do potencjalnego zapisu takiego pliku bez kompresji istnieje znikoma liczba formatów zapisu (np. BigTIFF). Skaner zastosowany w badaniach posiada czterdziestokrotne powiększenie optyczne oraz bardzo czułą matrycę CCD o wielkości piksela wynoszącego 3,45 µm. Dzięki tym właściwościom można bardzo dokładnie obejrzeć strukturę komórek zawartych w preparacie (rys. 2.1). Maksymalne powiększenie pozwala wyodrębnić charaktery-



Rysunek 2.1: Opis preparatu w maksymalnym powiększeniu - przypadek łagodny

styczne fragmenty jąder komórkowych takich jak: jąderko, błona komórkowa, cytoplazma oraz inne organella. Dodatkowo na preparatach można dostrzec czerwone krwinki, które stanowią cenne źródło wiedzy diagnostycznej dla lekarzy. Czerwone krwinki są zawsze podobnej wielkości, to z kolei pozwala oceniającemu lekarzowi określić wielkość jądra komórkowego w stosunku do wielkości czerwonej krwinki. Komputerowa weryfikacja wielkości jądra komórkowego nie wymaga obecności czerwonej krwinki na obrazie, a jedynie poprawnej segmentacji pojedynczych komórek. Wielkość jądra komórkowego jest zatem jedną z pierwszych cech, na które diagnosta zwraca uwagę.

Istnieją również inne czynniki warunkujące ostateczny efekt w postaci cyfrowego obrazu cytologicznego. Procedura biopsji rozpoczyna się od prawidłowej lokalizacji miejsca nakłucia. Zadanie wprowadzenia igły w guz odbywa się ze wspomaganiem w postaci obrazowania ultrasonograficznego, należy jednak zaznaczyć, iż najlepiej jest użyć obrazowania wspomagającego co najmniej z dwóch stron prostopadle do siebie. To znacznie zwiększa prawdopodobieństwo trafienie w guz. Kolejne czynności polegają na wybarwieniu próbki, prawidłowym położeniu oraz wykonaniu odpowiedniego rozmazu na szkiełku. To co najbardziej interesujące z punktu widzenia produktu końcowego to fakt wybarwiania preparatu z użyciem kwasochłonnej (wybarwia zasadowe fragmenty) eozyny oraz hematoksyliny, która wybarwia komórkowe struktury kwasowe (zasadochłonne). Eozyna nadaje czerwoną barwę fragmentom próbki natomiast hematoksylina jest barwnikiem niebieskim. Efekt działania tych dwóch substancji wybarwiających możemy zaobserwować na rys. 2.1. Wybarwienie próbki jest istotne z punktu widzenia dalszego przetwarzania obrazu cyfrowego.

#### 2.2 Zbiory danych

#### 2.2.1 Zestaw obrazów cytologicznych ze Szpitala w Zielonej Górze

Zbiór obrazów cytologicznych nowotworu piersi powstał w ramach współpracy Instytutu Sterowania i Systemów Informatycznych na Wydziale Informatyki, Elektrotechniki i Automatyki Uniwersytetu Zielonogórskiego oraz Zakładu Patomorfologii Szpitalu Uniwersyteckiego w Zielonej Górze. Materiał badawczy zgromadzony w ramach projektu składa się bazy zorganizowanej według złośliwści przypadków.

Dane użyte do eksperymentów zostały pozyskane przez lekarzy patomorfologów z Zakładu Patomorfologii Uniwersyteckiego Szpitala Klinicznego w Zielonej Górze. W wyniku zabiegów biopsji cienkoigłowej pozyskano 25 wirtualnych slajdów zawierających przypadki łagodne oraz 25 preparatów zawierających przypadki złośliwe. Istotne diagnostycznie fragmenty wirtualnych slajdów zostały zaznaczone przez lekarzy na 11 obrazach dla każdego pacjenta co daje łącznie 550 obrazów. Część obrazów została poddana procedurze ręcznej segmentacji (na jądra komórkowe, cytoplazmę z czerwonymi krwinkami oraz tło). Oprócz wymienionych dwóch zbiorów pozyskano również zbiór 25 wirtualnych slajdów przypadków gruczolakowłókniaków. Podobnie jak w poprzednich zbiorach dla każdego slajdu lekarze wybrali po 11 obrazów istotnych diagnostycznie Pozyskane dane posłużyły do zbudowania i wyuczenia sieci CNN w celu automatycznej segmentacji obrazów cytologicznych.

Warto podkreślić, że liczba pacjentów jest kwestią nieprzewidywalną i trudną do uzyskania w znacznej liczbie. Z tej przyczyny do eksperymentu pozyskano 50 próbek od pacjentów. Jak podaje literatura[10] zastosowanie sztucznej sieci neuronowej wymaga dużej ilości danych uczących. Liczba pacjentów wynosząca 50, spośród której ponadto należy wyodrębnić zbiór uczący i testowy wydaje się być zatem niewielka. Deficyt ten jest jednak częściowo rekompensowany przez rozmiar wirtualnych slajdów. Obszar pojedynczego wirtualnego slajdu to mniej więcej 200 000 na 100 000 pikseli. Dodatkowo wirtualny slajd posiada kilka mniej szczegółowych warstw (serii) służących do płynnego przeglądania zawartości wirtualnego slajdu. Obraz o rozmiarze 800 na 1500 pikseli zajmuje niemal niezauważalną część wirtualnego slajdu (rys. 2.2).



Rysunek 2.2: Maksymalne powiększenie fragmentu obszaru wirtualnego slajdu

#### 2.2.2 Zestaw obrazów histopatologicznych BreakHis

Powołując się na informacje pozyskane z witryny Uniwersytetu Federalnego Stanu Parana w Kurtybie [93], zbiór BreakHis 9109 cyfrowych obrazów pobranych z grupy 82 pacjentów. Niestety większość obrazów w bazie stanowią zdjęcia o niskim powiększeniu. Jedynie zbiór oznaczony jako "400X", zawiera obrazy odpowiadające skalą obrazom w zbiorze SzUZG. Wynika to z faktu, że do pozyskania obrazów w obu przypadkach zastosowano mikroskop z obiektywem o czterdziestokrotnym powiększeniu optycznym. Należy tu odróżnić powiększenie optyczne od wizualnego, gdyż zbiór BreakHis posługuje się nazwą "400X", ze względu na wizualne 400-krtotne powiększenie obiektów. Powyższe wyjaśnienie prowadzi

do końcowego stwierdzenia, iż zbiór BreakHis "400X" zawiera 1820 obrazów podzielonych na dwie główne grupy, czyli: łagodne (Benign) w liczbie 588 obrazów oraz złośliwe (ang. Malignant), których liczba wynosi 1232. Dane te zostały utworzone na bazie 82 pacjentów. Zbiór BreakHis charakteryzuje się dodatkowo podziałem szczegółowym zbioru łagodnego na grupy: gruczolak (*łac. adenosis*), gruczolakowłókniak (*łac. fibroadenoma*), guz liściasty (*łac. phyllodes tumor*) oraz gruczolak rurkowy (*łac. tubular adenona*). Z kolei przypadki złośliwe zostały podzielone na grupy: rak (*łac. carcinoma*), rak zrazikowy (*łac. lobular carcinoma*), rak śluzowy (*łac. mucinous carcinoma*) oraz rak brodawkowaty (*łac. papillary carcinoma*). Obrazy charakteryzują się rozdzielczością wynoszącą 700 na 460 pikseli, posiadają trzy kanały (RGB) o 8-bitowej głębi w każdym kanale oraz zapisane się w formacie PNG. Próbki pozyskane zostały przy użyciu biopsji wycinającej (SOB).

#### 2.2.3 Połączenie danych

Najistotniejszym problemem związanym z przetwarzaniem cyfrowych obrazów medycznych tkanek nowotworowych jest niedobór liczby pacjentów. Przykładowo zbiór SzUZG posiada 550 obrazów ale jedynie 50 pacjentów. Konieczne zatem było pozyskanie dodatkowej bazy w celu uzupełnienia posiadanych zasobów oraz weryfikacji skuteczności proponowanych metod na danych pochodzących z rożnych ośrodków.

Pierwszym problemem z jakim należało się zmierzyć w przypadku baz pochodzących z różnych ośrodków jest rozbieżność rozdzielczości obrazów oraz różne formaty (PNG oraz TIFF). Oba zastosowane formaty są odpowiednie do obrazów medycznych, gdyż TIFF jest formatem bezstratnym podobnie jak PNG. Ostatecznie do eksperymentów wybrano format PNG ze względu na mniejsze rozmiary plików oraz nieco lepsze wsparcie bibliotek Pythona.

Podczas analizy zawartości zbioru BreakHis X400, został dostrzeżony problem znacznej dysproporcji w liczbie obrazów łagodnych (588) do złośliwych (1232). Problem ten został rozwiązany poprzez dołączenie do danych BreakHis zbioru gruczolakowłókniaków z bazy Szpitala Uniwersyteckiego w Zielonej Górze. Nowy zbiór przyjął nazwę BreakHis + GZG, gdzie GZG będzie oznaczać właśnie gruczolakowłókniaki. Kolejny problem kompatybilności plików polegał na różnych rozdzielczościach obrazów (SzUZG - 1583x828, BreakHis - 700x460). Problem rozwiązano w taki sposób, że z każdego obrazu o rozmiarze 1583x828 wykadrowano 5 obrazów w rozmiarze 700x460 z czego jeden był wycinany z samego środka oraz cztery w narożnikach (rys. 2.3). Ponadto wykadrowane obrazy zostały poddane wstępnej selekcji polegającej na odrzuceniu z bazy obrazów, na których dominowało tło. Kryterium odrzucenia obrazu to pokrycie pikselami należącymi do jąder komórkowych wynoszące 20%. Pokrycie pikselami natomiast zostało wykonane na bazie szybkiej binaryzacji z użyciem rozplotu obrazu[77] a następnie binaryzacji metodą Otsu[65]. Rozwiązanie takie nie jest tak dokładne jak binaryzacja przy użyciu sztucznych



Rysunek 2.3: Kadrowanie obrazu do rozmiaru 700 x 460

sieci neuronowych, lecz jest ekstremalnie szybka oraz nie wymaga stosowania zestawu uczącego. To z kolei dwa najważniejsze kryteria jakie musi spełniać wstępna binaryzacja przeprowadzana na znacznej liczbie obrazów. Ostatecznie dołączenie zbioru gruczolakowłókniaków do BreakHis spowodowało uzyskanie stosunku liczby obrazów łagodnych (1408 - 53%) do złośliwych (1232 - 47%).

Druga baza obrazów SzUZG, w ramach dopasowania do zbioru BreakHis została wykadrowana w identyczny sposób jak zbiór GZG. Podobnie również i w tym przypadku zastosowano rozplot i binaryzacje w celu wykrycia obszarów zajmowanych przez piksele należące do jąder komórkowych oraz odrzucono obrazy zawierające znaczną ilość tła. W tym jednak przypadku kryterium odrzucenia było nieco mniej rygorystyczne i wynosiło 10% pokrycia. Dzięki temu uzyskano porównywalną liczbę obrazów w obu bazach a także znakomity stosunek liczby obrazów łagodnych (1318) do złośliwych (1330) wynoszący w zaokrągleniu 50 na 50.

Ostatecznie przygotowane obrazy posłużyły do wykonania segmentacji ręcznej. Następnie obrazy ręcznie posegmentowane posłużyły do budowy algorytmu segmentacji automatycznej przy użyciu CNN. Szczegóły dotyczące metod segmentacji zostaną poruszone w kolejnych rozdziałach. Na bazie obrazów w rozmiarze 700x460 zostały wykadrowane obrazy o rozmiarze 230x230, które z kolei zostały użyte do uczenia klasyfikujących sztucznych sieci neuronowych. Ze względu na duże skomplikowanie procesu przygotowania obrazów, cała procedura została dodatkowo uzupełniona rysunkiem schematycznym (rys. 2.4). Ostatecznie można zauważyć iż podstawową jednostką klasyfikacyjną będzie obraz o rozmiarze 230 x 230 pikseli. Jest to o tyle istotne, że badania[47, 89] wykazały, że klasyfikacja pojedynczych jąder komórkowych osiąga realną dokładność na poziomie 80%, stąd wniosek, dalsze badania należy oprzeć o większe jednostki danych obejmujące



obrazy zawierające większą ilość jąder komórkowych.

Rysunek 2.4: Schemat przygotowania baz danych z obrazami

Kadrowanie obrazów do rozmiaru 230x230 zostało podyktowane rozmiarem obrazów w zbiorze BreakHis, który wynosi 700x460. Przy kadrowaniu z jednego obrazu można wygenerować 6 niepowtarzalnych kadrów rozmiaru 230 z odcięciem dwóch pasków o sze-rokości 5 pikseli z prawej i lewej stronie. Po raz kolejny zastosowano kryterium odrzucenia obrazów zawierających zbyt dużo tła. W przypadku zbioru SzUZG kryterium wynosiło 7,5% pokrycia pikselami jąder komórkowych natomiast w przypadku zbioru BreakHis kryterium wyniosło 19%. Progi zostały dobrane eksperymentalnie w taki sposób żeby w miarę optymalnie rozłożyć stosunek liczby przypadków złośliwych do łagodnych w poszczególnych zbiorach a jednocześnie zrównoważyć łączną liczbę obrazów w obu zbiorach. Ostatecznie uzyskano w zbiorze SzUZG 10756 (45% z wszystkich) obrazów w tym łagodnych 5451 (51%) i złośliwych 5305 (49%), natomiast w zbiorze BreakHis + GZG 13292 (55% z wszystkich), w tym 6940 (52%) łagodnych i 6352 (48%) złośliwych.

#### 2.3 Wstępnie przetwarzanie obrazów cyfrowych

Niezwykle ważną rolę w pracy z obrazami cyfrowymi pełni wspomniane wyżej, wstępne przetwarzanie obrazów. Jedną z najpopularniejszych metod wstępnego przetwarzania jest rozplot (dekonwolucja) obrazu. W przypadku cyfrowych obrazów próbek pobranych w wyniku biopsji dekonwolucja pomaga w wstępnym odseparowaniu jąder komórkowych od reszty obiektów na obrazach. Możliwe jest to dzięki zastosowaniu koncepcji opartej o prawo absorpcji (Lamberta-Beera) [77]. Do przeprowadzenia operacji rozplotu wykorzystano bibliotekę HistomicsTK (*https://pypi.org/project/histomicstk/*). Próbki zostały utrwalone przy użyciu kwasochłonnej eozyny oraz zasadochłonnej hematoksyliny, w związku z powyższym zastosowano macierz dekonwolucji w poniższej formie:

$$\begin{vmatrix} eosin : \\ hematoxylin : \\ null : \end{vmatrix} \begin{vmatrix} 0.07, 0.99, 0.11 \\ 0.65, 0.70, 0.29 \\ 0.00, 0.00, 0.00 \end{vmatrix}$$
(2.1)

Po wykonaniu rozplotu otrzymany obraz można znormalizować lub rozpocząć dalsze przetwarzanie. Kolejne etapy, w zależności od obranej ścieżki przetwarzania, mogą obejmować zastosowanie operacji morfologicznych lub binaryzację. Docelowo obraz jest poddawany segmentacji semantycznej. Efekt zastosowania dekonwolucji został przedstawiony na rys. 2.5. Dekonwolucja stanowi wstęp do dalszego przetwarzania obrazów cytologicznych.



Rysunek 2.5: Efekt rozplotu obrazu metodą H&E z użyciem w oprogramowania Fiji [83]

W przypadku wstępnego przetwarzania obrazów zastosowanego w opisywanym eksperymencie wykorzystano metodę binaryzacji Otsu[65]. Metoda ta jest klasyfikowana jako globalna, co oznacza, iż do wyznaczenia progu binaryzacji używany jest obraz w całości. Idea metody polega na analizie histogramu oraz minimalizacji sumy ważonej węwnątrzklasowej obu klas:

$$V_w(b) = w_0(b)V_0(b) + w_1(b)V_1(b)$$
(2.2)

gdzie  $w_0$  oraz  $w_1$  są prawdopodobieństwami obu klas, natomiast *b* oznacza określoną wartość progu intensywności pikseli.  $V_0$  oraz  $V_1$  to wartości wariancji. Minimalizacja sumy ważonej węwnątrzklasowej jest równoznaczna z maksymalizacją wariancji międzyklasowej. Największą wadą metody Otsu w przypadku przetwarzania obrazów z biopsji jest błędna klasyfikacja erytrocytów jako jąder komórkowych, szczególnie w przypadku obrazów pozbawionych jąder komórkowych. Takie obrazy się niestety zdarzają w przypadku kadrowania większych obrazów.

#### 2.4 Podsumowanie

W wyniku przeprowadzonych prac uzyskano 2 zestawy uczące oraz 2 zestawy testowe. Pierwszy zestaw uczący składa się z obrazów ze zbioru BreakHis z laboratorium w Paranie oraz gruczolakowłókniaków ze Szpitala Uniwersyteckiego w Zielonej Górze. Modele utworzone na bazie tego zestawu uczącego będą testowane na danych ze Szpitala w Zielonej Górze bez gruczolakowłókniaków. Drugi zestaw danych uczących składa się wyłącznie z danych pochodzących ze Szpitala w Zielonej Górze, który będzie testowany na dwóch zestawach danych uczących. Jeden pochodzi ze zbioru BreakHis połączonego z przypadkami gruczolakowłókniaków z Zielonej Góry, drugi zestaw danych testowych zawiera wyłącznie dane z Brazylii. Połączenie danych BreakHis z gruczolakowłókniakami miało na celu zbadać jaki jest wpływ przecieku danych pomiędzy ośrodkami na wynik najlepszych metod klasyfikacji.

Obrazy wykorzystane do segmentacji różnią się od obrazów wykorzystanych z zadaniu klasyfikacji. Obrazy do segmentacji mają rozmiar 700x460 pikseli, wynika on z faktu rozmiaru obrazów w zbiorze BreakHis. Obrazy do klasyfikacji mają rozmiar 230x230. Mniejszy rozmiar obrazów do klasyfikacji wynika z wykorzystywania głębokich sieci neuronowych. Zmiana podyktowana jest koniecznością zastosowania różnych architektur uczenia głębokiego.

Zastosowanie szybkiej metody przetwarzania wstępnego pozwoliło oszczędzić czas przetwarzania obrazów podczas pracy głównego algorytmu segmentacji. Połączenie rozplotu H&E wraz z binaryzacją metodą Otsu jest wystarczająco skuteczne w wykrywaniu diagnostycznie istotnych obrazów, jednakże segmentacja jąder komórkowych jest niewystarczająca. Konieczne jest zastosowanie zaawansowanej metody segmentacji, którą opisano w kolejnym rozdziale.

## Rozdział 3

## Normalizacja obrazów przez segmentację

#### 3.1 Wprowadzenie

Segmentacja obrazu stanowi jedną z podstawowych technik przetwarzania obrazów. Zadaniem segmentacji jest podział obrazu na jednorodne części składowe. W przypadku obrazów cytologicznych segmentacja obrazu skupia się wokół oddzielenia jąder komórkowych od tła oraz na rozłączeniu jąder komórkowych względem siebie. Oznacza to, że oprócz punktowych metod segmentacji niezbędne wydaje się użycie metod wykrywania krawędzi. Połączenie tych dwóch zadań może mieć kluczowe znaczenie w procesie segmentacji obrazów cytologicznych. Istnieją różne podejścia związane z segmentacją obrazów polegające na wykonywaniu przekształcenia morfologicznych lub zastosowaniu sieci CNN. Wadą podejść opartych o przekształcenia morfologiczne jest konieczność ręcznego doboru parametrów oraz mniejsza skuteczność niż metody bazujące na sztucznych sieciach neuronowych. Natomiast podejście bazujące na CNN wymaga większej ilości czasu na przygotowanie danych jak również wymaga sporej ilości danych w zbiorze uczącym.

# 3.2 Segmentacja przy użyciu splotowej sieci neuronowej

Sieci neuronowe znajdują zastosowanie w zadaniu wykrywania obiektów na obrazie oraz segmentacji. Wśród tych metod możemy wyróżnić trzy podobne zagadnienia tj.: detekcję obiektów, segmentację semantyczną oraz segmentację instancji. Detekcja obiektów polega na wykryciu obiektu oraz jego położenia na obrazie, zazwyczaj ze wspomaganiem w postaci ramki ograniczającej (ang. *bounding box*). Segmentacja semantyczna polega na wyodrębnieniu obiektów należących do tej samej klasy na poziomie pikseli. W przy-
padku przeprowadzonych eksperymentów klasy obiektów podlegających segmentacji są trzy: wnętrze jądra komórkowego, krawedź jądra oraz tło. Segmentacja instancji polega natomiast na wyodrębnieniu pojedynczych obiektów z wybranej klasy obiektów. W przeprowadzonych badaniach całościowy algorytm segmentacji składa się z dwóch głównych etapów w postaci segmentacji semantycznej oraz segmentacji instancji. Różnice pomiędzy podstawowymi pojęciami zdefiniowanymi w powyższym akapicie ilustruje rysunek 3.1.



Rysunek 3.1: Różnice pomiędzy detekcją, segmentją semantyczną oraz segmentacją instancji

W tej pracy detekcja obiektów nie będzie wykorzystywana, więc pierwszą metodą jaka zostanie szerzej omówiona jest segmentacja semantyczna, a dokładniej sieć CNN typu U-Net, która realizuje tą segmentację. Aby przetworzyć obrazy o dowolnych rozmiarach, zastosowano strategię nakładania się obrazów zapewniająca płynną segmentację, zgodnie z opisem w oryginalnym artykule [75]. Oryginalne podejście zastosowane w publikacji [75] zakłada przetworzenie obrazu wejściowego o rozmiarze 572×572×1 do obrazu o rozmiarze 388×388×2. Oznacza to, że w wyniku wielokrotnych operacji splotu dokonała się znaczna redukcja wymiarów obrazu wyjściowego względem pierwotnego. Jedynie centrala część obrazu będzie podlegać segmentacji. Problem brzegowy można jednak rozwiązać przy użyciu dopełnienia. Dodatkowo można zauważyć, iż obraz wejściowego trójkanałowego. Dalsza analiza podejścia U-Net zostanie uzupełniona o schemat (rys. 3.2) sieci wzorowany na oryginalnym rysunku z artykułu [75].

Na schemacie można wyróżnić charakterystyczny kształt sieci przypominający literę U oraz wyraźne dwie główne części sieci w postaci fragmentu enkodującego oraz dekodującego. Obrazek wejściowy widoczny na skrajnie lewej części schematu przetwarzany jest przy użyciu splotu oraz filtra o wymiarze  $3\times3\times64$ . Wykonanie dwóch takich operacji powoduje zmianę wymiarów tensora z  $572\times572\times1$  do  $568\times568\times64$ . Dwie operacje splotu kończą się maksymalizującą warstwą łączącą, która redukuje wymiary szerokości i wysokości o połowę. Dalsze warstwy analogicznie poddawane są dwóm operacjom splotu, po których następuje maksymalizującą warstwą łączącą, a dokładnie schemat ten jest powtarzany jeszcze trzy razy. Po jego zakończeniu otrzymujemy tensor o wymiarze  $32\times32\times512$ ,



Rysunek 3.2: Schemat sieci U-Net na podstawie artykułu [75]

który jest poddawany kolejnym dwóm operacjom splotu, które w rezultacie generują tensor o wymiarze  $28 \times 28 \times 1024$ . Od tego tensora rozpoczyna się odtwarzanie obrazu z użyciem warstwy transponowanego splotu, która służy do ekspansji (*ang. upsampling*). W wyniku splotu transponowanego tensor o wymiarach  $28 \times 28 \times 1024$  jest przetwarzany do tensora o wymiarach  $56 \times 56 \times 512$ . Dodatkowo w tym miejscu następuje połączenie uzyskanego tensora z wycinkiem odpowiadającej mu warstwy z zestawu warstw w pierwszej części sieci. Zabieg ten ma na celu uzupełnić drugą część sieci o informacje z pierwszej w celu poprawienia wyników segmentacji. Następnie uzyskany tensor jest poddawany dwóm operacjom splotu. Czynności te począwszy od splotu transponowanego powtarzane są jeszcze trzykrotnie, aż do uzyskania tensora o rozmiarach  $392 \times 392 \times 64$  który jest uzupełniony wycinkiem o tych samych wymiarach pozyskanym z pierwszej warstwy po dwóch splotach. Uzyskany w ten sposób tensor jest poddawany podwójnej operacji splotu ( $3 \times 3 \times 64$ ) po której następuje ostatni splot filtrem o wymiarze  $1 \times 1 \times 1$ . Ostatni splot powoduje redukcję tensora do obrazu wynikowego, na którym znajdują się obiekty w dwóch klasach.

### 3.3 Hybrydowy system segmentacji

Segmentacja wododziałowa jest to algorytm inspirowany topograficzną charakterystyką zagłębień terenu oraz zalewania ich wodą. W literaturze [27] można odnaleźć definicję trzech podstawowych grup punktów na obrazie t.j.: (I) lokalne minima, (II) punkty po umieszczeniu, w których wody będzie ona zmierzać do jednego z minimów oraz (III)

37

punkty, w których woda z równym prawdopodobieństwem będzie zmierzać do przynajmniej dwóch minimów. Punkty wymienione w podpunkcie (III) zwane są działami wód lub wododziałami, stąd nazwa algorytmu. Obraz wejściowy zazwyczaj otrzymuje binarną reprezentację (np. obszar jąder komórkowych oraz tło). Obraz binarny można uzyskać w wyniku progowania lub podczas segmentacji semantycznej z użyciem sieci CNN. Po uzyskaniu obrazu binarnego obszary jąder komórkowych otrzymują wartość 1 natomiast tło 0, obraz ten można potraktować jako maskę binarną. W kolejnym kroku powierzchnia obrazów binarnych jest przekształcona przy użyciu transformacji euklidesowej w taki sposób, iż piksele otrzymują wartość oznaczającą odległość badanego piksela od krawędzi obiektów. Uzyskane wartości można sobie wyobrazić jako głębokości względem powierzchni (tła). Punkty najbardziej oddalone od krawędzi obiektów stają się minimami lokalnymi, od których rozpoczyna się proces zalewania basenów. W ostatnim korku algorytmu następuje zalewanie obszarów począwszy od minimów, do czasu aż obszary przypisane różnym minimom spotkają się na działach wodnych. Niestety algorytm w tej formie ma tendencję do tworzenia nadliczbowego zbioru segmentów [27, 116]. Jako główną przyczynę nadmiernej segmentacji wymienia się znaczna liczbę potencjalnych minimów [27]. Aby poradzić sobie z tym problemem, można użyć zmodyfikowanej wersji algorytmu wododziałowego, która wykorzystuje dodatkowe markery oznaczające punkty centralne jąder komórkowych. Niestety wykrywanie punktów środkowych jąder stanowi wyzwanie, jednak poprawa jakości segmentacji semantycznej jest niezwykle istotna. Zmodyfikowane podejście nosi nazwe algorytmu wododziałowego sterowanego markerami. Algorytm ten różni się od klasycznego podejścia tym, że w zmodyfikowanej wersji punktami startowymi, od których rozpoczyna się zatapianie powierzchni topograficznej, nie są lokalne minima lecz wyznaczone markery.

Aby poradzić sobie z przedstawionym problemem zaproponowano hybrydową metodę segmentacji (rys. 3.3), która wykorzystuje dwie sieci U-Net metodą wododziałową z markerami. Pierwsza sieć została przeszkolona do przewidywania, czy piksele należą do wnętrza jądra (klasa 1), czy do krawędzi jądra i tła (klasa 2). Na podstawie segmentacji semantycznej zapewnianej przez tę sieć zostały wyodrębnione wnętrza jąder w postaci obrazów binarnych. Obrazy te zostały użyte do określenia powierzchni topograficznej dla algorytmu wododziałowego przy użyciu transformacji odległości. Druga sieć U-Net została przeszkolona do wykrywania centrów jąder. Sieć przewidywała, czy piksele należą do centrum jądra (klasa 1), czy do wnętrza jądra, krawędzi i tła (klasa 2). Centra (markery jąder) wyodrębniono z wyników segmentacji semantycznej w postaci obrazów binarnych.

W ostatnim kroku powierzchnię topograficzną połączono z centrami (markerami jąder) przy użyciu rekonstrukcji morfologicznej. Zmodyfikowana powierzchnia topograficzna została przetworzona przez klasyczną transformację wododziałową w celu wykrycia jąder.



Rysunek 3.3: Hybrydowa metoda segmentacji wododziałowej z użyciem wstępnej segmentacji za pomocą sieci CNN oraz wykrywanie markerów wododziałów z użyciem drugiej sieci CNN

# 3.4 Metody ewaluacji

Aby poddać ocenie jakość segmentacji należy zastosować odpowiednie metryki. Ocena wyników rozpoczyna się od zaznaczenia jąder komórkowych na obrazie testowym. Czynność ta najczęściej sprowadza się do precyzyjnego zaznaczenia krawędzi jądra komórkowego przy użyciu urządzenia wskazującego na komputerze. Jądra można zaznaczyć przy użyciu obiektów ROI (*ang. Region Of Intrest*) i jest to standardowa technika zaznaczania obszarów na obrazach medycznych. W praktyce obiekt ROI stanowi obwiednie jądra komórkowego, który można następnie przekształcać do formy maski binarnej obiektu lub do maski z etykietami (Rys. 3.4).



Rysunek 3.4: Etap przejścia z obiektu rzeczywistego na ROI oraz maskę binarną

#### Indeks Jaccarda

Maska binarna jest przydatna w przypadku jednej z najbardziej intuicyjnych metryk zwanej Indeksem Jaccarda (JI) [39]. Metryka ta opiera się na algebrze zbiorów i mierzy podobieństwo pomiędzy dwoma zbiorami:

$$JI(X,Y) = \frac{X \cap Y}{X \cup Y} \tag{3.1}$$

Indeks Jaccarda oblicza się stosując maski jąder komórkowych uzyskane z segmentacji z maskami powstałymi z oznaczonych ręcznie ROI. Metoda jest skuteczna ponieważ Maski binarne stanowią zbiory pikseli. Kwestią istotną z praktycznego punktu widzenia jest określenie wielkości prostokąta (*ang. bounding box*), w którym zmieści się maska największego jądra komórkowego. Jest to o tyle istotne, że w algorytm przetwarza pojedyncze prostokąty wewnątrz, których znajduje się jądro komórkowe. Wielkość prostokąta jest zatem uzależniona od obszaru największego obiektu i wpływa na szybkość procesu przetwarzania obliczeń. Druga kwestia, która ma wpływ na uzyskany wynik to kryteria wyznaczania punktu centralnego maski binarnej. Lekkie przesunięcie punktu centralnego, w zależności od przyjętej metody wyznaczania środka, może nieco zmienić wynik końcowy Indeksu Jaccarda. Wartość JI waha się w przedziale od 0 do 1, im wyższa wartość tym obiekty są lepiej dopasowane do siebie. Czasem zamiast Indeksu Jaccarda można użyć Dystansu Jaccarda, który stanowi dopełnienie Indeksu Jaccarda do wartości 1:

$$JD(X,Y) = 1 - \frac{|X \cap Y|}{|X \cup Y|}$$
(3.2)

Zaletą JD jest uzyskanie bezpośredniej informacji o wartości błędu podobieństwa pomiędzy dwoma porównywanymi maskami obiektów.

#### Współczynnik Dice-Sørensena

Kolejną miarą często stosowaną w procesie ewaluacji wyników segmentacji jest Współczynnik Dice-Sørensena (WDS) [16, 91]. Wielkość ta również bazuje na algebrze zbiorów i przyjmuje postać:

$$WDS(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|}$$
(3.3)

Podobnie jak w przypadku JI, współczynnik Sørensena do obliczeń wymaga masek binarnych obiektów lub ich współrzędnych. Współczynnik Dice-Sørensena przyjmuje także m. in. nazwy: F-score, Indeks Czekanowskiego, Indeks Steinhausa, Indeks Sørensena. WDS przyjmuje również wartości od 0 do 1, gdzie wartość 1 oznacza, że obiekty są identyczne. Ze względu na matematyczne podobieństwo Indeksu Jaccarda i Współczynnika Dice-Sørensena często używane są zamiennie.

#### Dystans Hausdorffa

Istnieje jednak miara całkowicie różna w założeniach od JI i WDS. Miara która często wykorzystywana jest do oceny skuteczności segmentacji jąder komórkowych zwana dystansem lub metryką Hausdorffa (DH). Jest miarą odległości między dwoma zbiorami w przestrzeni metrycznej (rys. 3.5). Matematyczny zapis DH definiowany jest następująco:

$$DH(A,B) = \max\left\{\sup_{a\in A} \inf_{b\in B} d(a,b), \sup_{b\in B} \inf_{a\in A} d(a,B)\right\},\tag{3.4}$$

gdzie sup oznacza supremum, natomiast inf oznacza infimum.

W przypadku jąder komórkowych mamy do czynienia z danymi w przestrzeni dwuwymiarowej. Jądro pojedynczej komórki to zbiór pikseli:  $A = \{a_1, a_2, a_3, ..., a_m\}$  w przestrzeni



Rysunek 3.5: Graficzna interpretacja DH

obrazu. Odległość euklidesową pomiędzy poszczególnymi elementami zbioru A i zbioru  $B = \{b_1, b_2, b_3, ..., b_n\}$  w dwuwymiarowej płaszczyźnie Euklidesowej możemy zdefiniować jako:

$$d_E(a,b) = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2}$$
(3.5)

Następnie możemy zdefiniować odległości między poszczególnymi pikselami a innym zestawem pikseli:

$$d(a, B) = \min_{b \in B} d_E(a, b); \quad d(b, A) = \min_{a \in A} d_E(b, a)$$
(3.6)

Ostatecznie możemy obliczyć DH jako (rys. 3.5):

$$d_H(A, B) = \max\{\max_{a \in A} d(a, B), \max_{b \in B} d(b, A)\}$$
(3.7)

Jednak DH podczas przetwarzania binarnych masek jądra ogranicza się do badania krawędzi między maskami. Stosując to podejście, można znacznie przyspieszyć obliczenie DH. Maski jąder komórkowych powodują, że jednostką DH jest piksel. DH należy interpretować w ten sposób, iż im wynik jest wyższy tym niższe jest dopasowanie między dwoma zbiorami.

# 3.5 Weryfikacja dokładności hybrydowej metody segmentacji

#### 3.5.1 Implementacja hybrydowej metody segmentacji

Wyniki uzyskane w publikacji[47] pozwoliły ustalić skuteczną metodę segmentacji jąder komórkowych, bazującą na sieci CNN typu U-Net wspomaganej przez algorytm wododziałowy sterowany markerami. Dodatkowa obróbka obrazów posegmentowanych sztuczną siecią neuronową jest niezbędna do końcowego odseparowania pojedynczych jąder komórkowych. Wynika to z faktu, że sieć neuronowa generuje wynik prawdopodobieństwa, który nie definiuje krawędzi pomiędzy zlepionymi obiektami, innymi słowy wynik jest uzależniony od początkowej separacji jąder komórkowych na obrazach. Sieć zastosowana do segmentacji została zaczerpnięta z dokumentacji Keras (*https://keras.io/*) [10], która z kolei jest interpretacją sieci U-Net[75]. Jedyna różnica polega na wielkości obrazów wejściowych (700x460) oraz dodatkowym uzupełnieniu segmentacji U-Net o algorytm wododziałowy. Zbudowany model sztucznej sieci typu U-Net w formie opisowej przedstawiony został w załączniku A.

Analiza struktury sieci wskazuje, że na wejściu obraz o rozmiarze 700 x 460 został przetransformowany do rozmiaru 704 x 464. Taki zabieg umożliwił przejście kolejnych etapów uczenia bez deformacji kształtu obrazu wyjściowego. Z kolei tensor najmniejszego obrazu ma wymiary 88x58x256 co oznacza, że w najniższej warstwie modelowaniu podlegało 256 obrazów o rozmiarze 88x58, uzyskanych z trójkanałowego obrazu o rozmiarze 704 x 464. Łączna liczba parametrów modelu wyniosła 2 058 979 z czego 2 055 203 parametry podlegały uczeniu. W modelu zastosowano gradientowego algorytmu największego spadku RMSprop. Biorąc pod uwagę zasoby sprzętowe (zob. Dodatek B) sieć nie jest specjalnie wymagająca, zaledwie po kilkudziesięciu minutach można uzyskać ostateczny wynik uczenia.

Przykładowe wyniki modelowania zaprezentowano na rysunku 3.6. Zarówno na wykresie błędu jak i na wykresie dokładności widać, że krzywa uczenia dla danych walidacyjnych przestaje się poprawiać w okolicach 20 epoki uczącej. Poprawie ulegają natomiast parametry krzywych dla danych uczących, oznacza to zatem, że zaczyna występować zjawisko przeuczenia modelu. Szczegółowe dane wskazują iż najmniejsza strata modelu dla danych walidacyjnych nastąpiła w 21 epoce uczącej i wyniosła 0.1947, natomiast dokładność modelu w tej epoce wyniosła 0,9207. Dla danych uczących w 21 epoce wartość straty wyniosła 0,1796 natomiast dokładność 0,9264. W związku z powyższym 21 epoka stanowiła optimum jakie udało się uzyskać z tych danych.

Kolejnym istotnym aspektem było przygotowanie danych uczących oraz testowych. Zestaw danych uczących zawierał 844 ręcznie posegmentowane obrazy o rozmiarze 700x460.



(a) Dokładność modelu w kolejnych epokach (b) Strata modelu w kolejnych epokach

Rysunek 3.6: Przebieg uczenia modelu U-Net

Segmentacja odbywała się przy użyciu oprogramowania ImageJ (Fiji) za pomocą narzędzia do tworzenia plików ROI. Oznaczanie jąder komórkowych wykonywane było przy zastosowaniu myszy komputerowej oraz częściowo przy użyciu tabletu graficznego, który okazał się wygodniejszym narzędziem i wyraźnie przyspieszył proces oznaczania obrazów. Ich łączna liczba posegmentowanych ręcznie wyniosła 844, w tym: ze zbioru BreakHis 89 obrazów łagodnych (od 3 pacjentów), 62 obrazy złośliwe (od dwóch pacjentów); ze zbioru GZG 29 obrazów (od 1 pacjenta); ze zbioru SzUZG 330 obrazów łagodnych (od 25 pacjentów), 334 obrazy złośliwych (od 25 pacjentów). Kolejnym etapem przygotowań obrazów było utworzenie maski składającej się z 3 kategorii, do których należały: wnętrza jąder komórkowych, krawędź jądra oraz tło. Pierwszym etapem przygotowania danych było wczytanie pliku ROI oraz wykonanie masek binarnych dla każdego jądra komórkowego. W trakcie generowania maski pojedynczego jądra wykonywane było również wyznaczanie krawędzi jądra komórkowego. Do utworzenia krawędzi wykorzystano maskę powstałego jądra komórkowego, na której wykonano podwójną erozję morfologiczną ze zmiennymi elementami strukturalnymi (wzór 3.8) oraz pojedynczą dylatację.

$$\begin{bmatrix} 1, 1, 1\\ 1, 1, 1\\ 1, 1, 1 \end{bmatrix} \begin{bmatrix} 0, 1, 0\\ 1, 1, 1\\ 0, 1, 0 \end{bmatrix}$$
(3.8)

Obwiednie powstałe po erozji oraz dylatacji połączono w jeden obiekt w postaci krawędzi jądra komórkowego (Rys. 3.7). W zbiorze testowym zatem łącznie znalazło się 5288 obrazów. Część pacjentów ze zbioru uczącego powtarzała się w zbiorze testowym, szczególnie w przypadku danych pochodzących ze zbioru SzUZG. W związku z potrzebą posiadania wszystkich obrazów w późniejszej klasyfikacji oraz koniecznością zapewnienia braku przecieku danych testowych do zbioru uczącego zastosowano metodę walidacji krzyżowej w wersji *ang. Leave-One-Out.* Innymi słowy, chcąc uzyskać dane testowe dla konkretnego



Rysunek 3.7: Schemat tworzenia krawędzi jąder komórkowych obrazów w zbiorze uczącym

pacjenta, wykluczano obrazy do niej przynależne w zbiorze uczącym. Eksperyment segmentacji powtórzono zatem łącznie 55 razy, a łączny czas trwania serii eksperymentów wyniósł kilka tygodni. Ostatecznie wszystkie obrazy zostały posegmentowane przy użyciu sieci CNN. Co ciekawe we wszystkich 55 eksperymentach uzyskano porównywalne wyniki dokładności  $(0.92 \pm 0.01)$  oraz straty  $(0.19 \pm 0.01)$ , wyniki te również występowały w okolicach 20 epoki uczącej. Efektem końcowym działania sieci CNN było uzyskanie kanałów zawierających prawdopodobieństwa przynależności pikseli do poszczególnych klas (klasa wnętrza jądra, klasa krawędzi oraz klasa tła). Obrazy zostały przetworzone w taki sposób, że z końcowej mapy jąder komórkowych wykluczono krawędzie. Innymi słowy odjęcie klasy krawędzi od wnętrz jąder wzmocniło końcowy wynik separacji pojedynczych jąder komórkowych.

Na końcowy czas trwania eksperymentów dodatkowo znaczny wpływ miało przeplatanie uczenia sieci segmentującej jądra komórkowe z uczeniem sieci wykrywającej centra jąder komórkowych. Eksperyment ten polegał na detekcji punktów centralnych jąder komórkowych. Punkty miały powierzchnię kilkunastu pikseli a ich docelowym zadaniem było zastosowanie jako punkty startowe (markery) do algorytmu segmentacji wododziałowej. Detekcja punktów jest dość istotna, gdyż pozwala na oszacowanie liczby pojedynczych obiektów występujących na obrazie, a tym samym detekcję pojedynczych jąder komórkowych. Struktura sieci jest niemal identyczna jak zastosowana do segmentacji jąder komórkowych, różni się jedynie strukturą obrazu wynikowego, który składa się z dwóch kanałów (tła oraz centra jąder komórkowych). Podobnie jak w pierwszej sieci U-Net użyto w ostatniej warstwie funkcji aktywacji softmax. Jest to możliwe ponieważ w obu modelach wynikowa liczba kanałów jest większa od 1.

Przykładowe rezultaty uczenia U-Net wykorzystanej do detekcji punktów centralnych zostały zaprezentowane na rysunku 3.8. Widać wyraźnie, że w pierwszych epokach uczących dochodzi do osiągnięcia bardzo wysokiej wartości dokładności oraz bardzo małej



Rysunek 3.8: Przebieg uczenia modelu U-Net dla punktów centralnych

straty. Dopiero w dużym powiększeniu widać, iż w pobliżu 20 epoki uczącej sieć osiąga optymalne wyniki a potem następuje przeuczenie. Ekstremalnie wysokie wyniki dokładności oraz szybkie osiągnięcie bardzo niskiej straty jest podyktowane charakterystyką danych wejściowych. Obrazy, które zostały podane na wejściu do sieci posiadają bowiem znaczną dysproporcję w liczbie pikseli należących do poszczególnych klas (klasa centrum jądra komórkowego oraz klasa tła). Stąd nasuwa się wniosek, iż sieć neuronowa nie popełni dużego błędu, jeżeli sklasyfikuje piksele należące do centrów jąder komórkowych do dominującej liczbowo klasy tła. W takim przypadku można wymienić kilka potencjalnych metod zapobiegania takiej sytuacji np. wprowadzenie kary w sieci za klasyfikowanie pikseli należących do centrów jako tła. Z drugiej strony można też rozwiązać problem przetwarzając uzyskane obrazy przy użyciu progowania. Zastosowanie binaryzacji jest z kolei możliwe ponieważ wynikiem uzyskiwanym na wyjściu sieci neuronowej jest w tym przypadku prawdopodobieństwo przynależności piksela do poszczególnej klasy. Stąd można bardzo łatwo eksperymentalnie dobrać próg binaryzacji, który został ostatecznie ustalony na poziomie 0,10. Dodatkowo na rysunku 3.9 widnieje ogólny schemat przetwarzania obrazu jakie



Rysunek 3.9: Schemat segmentacji jąder komórkowych

zostało wykonane w eksperymencie. Z prawej strony widoczny jest efekt końcowy przetwarzania obrazu, po wyodrębnieniu pojedynczych jąder komórkowych. W górnej części obrazu widać jak sieć neuronowa skutecznie wyodrębniła obiekty należące do klasy jąder komórkowych, jednak nie była w stanie rozłączyć istniejących obiektów, zadanie to wymaga dodatkowej metody przetwarzania końcowego.

#### 3.5.2 Wyniki segmentacji jąder komórkowych

Uzyskane rezultaty w dużej mierze zależą od jakości segmentowanego obrazu. Zauważalna jest tendencja do spadku dokładności segmentacji dla obrazów w których jądra komórkowe nie zaabsorbowały wystarczającej ilości hematoksyliny. Z drugiej jednak strony metoda unika detekcji czerwonych krwinek jako jąder komórkowych oraz wyjątkowo skutecznie wykrywa skupiska zlepionych jąder komórkowych. Przykładowe rezultaty zostały zaprezentowane na rysunku 3.10. Wyniki dla przypadku łagodnego ze zbioru SzUZG obrazują, że sieć wykrywa jądra komórkowe w miejscach gdzie trudności pojawiły się nawet w czasie segmentacji manualnej.



Rysunek 3.10: Przykładowe wyniki segmentacji

Dalsze wyniki obejmują oceny ilościowe, które liczbowo charakteryzują dokładność segmentacji. Metoda ewaluacji ilościowej zastosowana w eksperymencie składa się z kilku etapów. Pierwszy etap polega na porównaniu pojedynczych jąder komórkowych z obrazów oznaczonych manualnie z jądrami komórkowymi wykrytymi przy zastosowaniu sieci neuronowej. Porównanie polegało na zbadaniu wartości JI. W przypadku gdy wartość indeksu przekroczyła wartość 0.5 czyli przynajmniej 50% pikseli wykrytego jądra komórkowego pokrywało się z pikselami jądra komórkowego z obrazu manualnie oznaczanego, wówczas jądro takie trafia do grupy wykrytych. Jeżeli jądro komórkowe na obrazie manualnie oznaczonym nie miało swojego odpowiednika (poniżej wartości 0,5) na obrazie segmentowanym siecią neuronową wówczas trafiało do zbioru niewykrytych. Trzecia grupa obiektów to nad-segmentowane, czyli jądra komórkowe wykryte siecią neuronową, które nie znalazły swoich odpowiedników na obrazach manualnie posegmentowanych. Dla wszystkich wykrytych obiektów uzyskano wartości JI oraz DH. Dodatkowo przeliczono procentowe wartości jąder wykrytych, niewykrytych oraz nad-segmentowane:

$$W[\%] = \frac{W}{W+N} 100 \tag{3.9}$$

$$N[\%] = \frac{N}{W+N} 100 \tag{3.10}$$

$$S[\%] = \frac{S}{W+S} 100, \tag{3.11}$$

gdzie w - Wykryte, N - Niewykryte, S- Nad-segmentowane. Pierwszą cechą poddaną analizie jest procentowa liczba wykrywanych jąder komórkowych. Kluczowa kwestią wydaje się być procentowy udział jąder komórkowych wykrytych, czyli mających swój odpowiednik w zbiorze manualnie posegmentowanym. Średni procent wykrytych poprawnie jąder komórkowych wynosi 62,7%, natomiast mediana wynosi 65,0%. Dodatkowo można zauważyć, że rozkład wykrytych jąder komórkowych jest mocno asymetryczny lewostronnie, co oznacza, że dominują wysokie wartości detekcji. Rozkłady obiektów niewykrytych oraz nad-segmentowanych są asymetryczne prawostronnie, co z kolei oznacza dominację niskich wartości. Średnia procentowa liczba jąder komórkowych niewykrytych wyniosła 37,3, natomiast nad-segmentowanych 32,3, mediany zaś wynoszą odpowiednio 35,0 oraz 30,0. Są to wartości znacznie niższe niż średnia wartość jąder komórkowych wykrytych. Na rozkładzie empirycznym jąder komórkowych wykrytych można również zauważyć niewielką liczbę niepokojących wyników, czyli takich gdzie liczba wykrytych obiektów nie przekroczyła 20%. W całym zbiorze wynoszącym 844 obserwacje znajduje sie dokładnie 36 takich przypadków co oznacza, że sieć neuronowa ma szczególny problem z segmentacją 4,3% obrazów. Analiza losowego obrazu, który trafił do grupy obrazów o niskiego wyniku segmentacji powinna nieco wyjaśnić potencjalne przyczyny słabszych wyników. Na przykładzie (rys. 3.12) zaprezentowano obraz, którego wynik ewaluacji określony zo-



Rysunek 3.11: Rozkłady wartości oraz statystyki

100

0.5%

0.0%

minimum

0

0

Ó

20

40

60

80



Rysunek 3.12: Przykładowy obraz o niskim wyniku ewaluacji segmentacji

stał na poziomie 0% wykrytych, 100% niewykrytych oraz 100% nad-segmentowanych. Obraz wejściowy w tym przykładzie charakteryzuje się bardzo mocno zlepioną strukturą

jąder komórkowych. Dodatkowo obszar zajmowany przez jądra komórkowe jest nasycony i lekko przysłonięty strukturami kwasochłonnymi (eozyna) oraz erytrocytami. Trudność w manualnej ocenie krawędzi jąder komórkowych odzwierciedla środkowa część obrazu, gdzie zaznaczone zostały jedynie 4 jądra komórkowe. Sieć neuronowa natomiast wykryła obszary jąder komórkowych o wyraźnej separacji i dobrym nasyceniu hematoksyliną. Niestety wykryte jądra komórkowe nie znalazły się w zbiorze manualnie oznaczonych. Zatem kolejny wniosek dotyczy samej metody oraz dokładności oznaczania, która spada wraz ze wzrostem skomplikowania struktur komórkowych. Jest to też przykład rozbieżności pomiędzy widzeniem ludzkim a komputerowym, a także niedoskonałości metod ewaluacji ilościowej oraz potrzebę uzupełniania jej metodami ewaluacji jakościowej.

Drugim przykładem może być losowo wybrany obraz (rys. 3.13) o przeciętnej procentowej zawartości poprawnie wykrytych jąder komórkowych (58,3%). Jądra komórkowe



Obraz wejściowy

Obraz posegmentowany manualnie

Obraz posegmentowany automatycznie

Rysunek 3.13: Przykładowy obraz o przeciętnym wyniku ewaluacji segmentacji

nad-segmentowane stanowią 18,1%, natomiast niewykryte osiągnęły poziom 41,3%. Uzyskane wyniki ilościowe nie odzwierciedlają rzeczywistej jakości segmentacji i należy ten przykład podać jako bardzo dobry wynik segmentacji, zważywszy na wizualny efekt jakościowy uzyskanego rezultatu.

Dalszą analizę rezultatów rozpoczyna przedstawienie rozkładów dla kategorii na jakie można podzielić zbiory (rys. 3.14). Pierwsza kategoria (Osrodek) dotyczy placówki



Rysunek 3.14: Rozkłady liczebności cech w grupach

medycznej, z której pochodzi próbka. Druga kategoria (Zbior) dotyczy przygotowanego w badaniach zbioru danych czyli dane BreakHis + GZG, a drugi zbiór to przypadki złośliwe i łagodne z SzUZG. Trzecia kategoria to rodzaj nowotworu na obrazach, czyli łagodne (B) lub złośliwe (M). Pierwszym problemem jaki można zauważyć jest duża dysproporcja w liczbie oznaczonych obrazów pomiędzy ośrodkami, która może wpłynąć na wynik analiz. Znacznie lepiej prezentuje się stosunek liczby przypadków złośliwych (47%) do łagodnych (53%). Interesujące byłoby ponadto przeanalizowanie czy jakość segmentacji będzie zależeć od ośrodków, z których obrazy pochodzą oraz czy rodzaje nowotworów mają wpływ na dokładność segmentacji. Kwestia dotycząca wpływu ośrodka pochodzenia obrazów, rozpoczyna się analizą wykresów pudełkowych (rys. 3.15) Pierwsze wykresy



Rysunek 3.15: Wykresy pudełkowe wyników segmentacji do ośrodka pochodzenia próbek

z lewej wskazują przewagę skuteczności detekcji na obrazach pochodzących ze Szpitala w Zielonej Górze (Zgora). Wszystkie parametry statystyczne są wyższe dla grupy obrazów z Zielonej Góry, w przypadku których kreska maksimum osiąga 100%. Dla porównania maksimum dla zbioru BreakHis zatrzymuje się na wartości 75%, natomiast jedna obserwacja wynosząca 100% jest traktowana jako odstająca. Środkowe wykresy dotyczące nad-segmentacji wykazują nieco inne wnioski. Średnia wartość procentowej liczby nad-segmentowanych jąder komórkowych w przypadku zbioru z Zielonej Góry jest niższa niż dla zbioru BreakHis. Z drugiej jednak strony rozrzut wartości procentowej liczby obiek-tów nad-segmentowanych niższy jest dla zbioru BreakHis. Ponadto wykres dla zbioru BreakHis wykazuje się większą symetrią. Ostatnia grupa wykresów (dla obiektów niewy-

krytych) stanowi lustrzane odbicie wykresów dla wykrytych obiektów, zatem wnioski są lustrzanym odzwierciedleniem konkluzji wyprowadzonych na pierwszych wykresach.

Po przeprowadzonej analizie wykresów pudełkowych, nasuwają się pierwsze spostrzeżenia dotyczące wpływu jakie poszczególne laboratoria medyczne, w których pozyskiwano obrazy, maja na detekcję jąder komórkowych. Dalsze analizy będa dotyczyć krzywych dopasowania ośrodka pozyskania obrazów do procentowej liczby wykrytych, nadsegmentowanych oraz niewykrytych obiektów (rys. 3.16). Na pierwszym wykresie od lewej widoczna jest krzywa logistyczna dopasowania procentowej wartości wykrytych obiektów do poszczególnych ośrodków, z których pochodzą dane. Z wykresu wynika, że wraz ze wzrostem wartości procentowej liczby wykrytych jąder rośnie procentowy udział obrazów ze Szpitala w Zielonej Górze. Statystyka testowa chi-kwadrat w tym modelu sprawdza hipotezę, że czynnik (wykryte [%]) nie ma wpływu na zmienną odpowiedzi (zmienna ta została nazwana Osrodek). P-wartość znajdująca się obok statystyki testowej została oszacowana na < 0,001 co oznacza, że mamy mocne uzasadnienie wskazujące na odrzucenie hipotezy zerowej i przyjęcie jako prawidłowej hipotezy mówiącej, że ośrodek pozyskiwania obrazów jest zależny od wyniku detekcji jąder komórkowych. Podobnie sytuacja kształtuje się dla ostatniego wykresu, z tą różnicą, że zwiększanie się liczby niewykrytych obiektów powoduje zwiększanie się procentowego udziału obrazów pochodzących ze zbioru BreakHis. Biorąc pod uwagę p-wartość również i w tym przypadku hipotezę o braku wpływu procentowej liczby niewykrytych obiektów względem ośrodka pochodzenia należy odrzucić. Srodkowy wykres wykazuje nieco mniejsze pochylenie, więc można wnioskować, iż wpływ tych cech względem siebie jest nieco niższy niż w poprzednich przypadkach. Nadal jest jednak zauważalna tendencja do spadku procentowego udziału obrazów ze zbioru z Zielonej Góry wraz ze wzrostem procentowej liczby obiektów nad-segmentowanych. Hipoteza zerowa mówiąca o braku zależności pomiędzy liczbą nad-segmentowanych obiektów oraz ośrodka pochodzenia obrazów powinna zostać odrzucona.

Przechodząc do kolejnej analizy warto krótko przypomnieć iż, dalsza część eksperymentu będzie bazować na dwóch zbiorach. Pierwszy zbiór to obrazy ze zbioru SzUZG natomiast drugi zbiór zawiera obrazy BreakHis + GZG. Zbiór GZG został połączony ze zbiorem BreakHis w celu zrównoważenia liczby przypadków łagodnych i złośliwych. Na koniec, można krótko przeanalizować wpływ badanego zbioru na skuteczność segmentacji jader komórkowych (rys. 3.17). Niestety uzyskane wyniki są zbliżone do wyników uzyskanych na podstawie ośrodka pochodzenia obrazów. Nadal widoczny jest mocny wpływ zbioru danych na detekcję oraz nad-segmentację. Innymi słowy najprawdopodobniej obrazy ze zbioru BreakHis + GZG będą gorzej posegmentowane niż obrazy pochodzące ze zbioru SzUZG. Z jednej strony może to wynikać z faktu, iż ze zbioru z Zielonej Góry każdy pacjent miał oznaczone przynajmniej kilka obrazów. W zbiorze BreakHis dotyczy to jedynie kilku pacjentów. Mimo tej rozbieżności ogólne wyniki segmentacji zbioru BreakHis są dobre, a to z kolei udowadnia, że sieć typu U-Net znakomicie radzi sobie nie tylko na danych pochodzących z jednego ośrodka badawczego, ale również ma też znakomite właściwości uogólniające.

Kolejna analiza dotyczyć będzie weryfikacji wpływu rodzaju nowotworu na obrazie względem detekcji. Na rysunku 3.18 wykres dla wykrytych oraz niewykrytych jest niemal płaski. Oznaczać to może niewielki wpływ tych cech na rodzaj nowotworu na obrazie. Aby potwierdzić wizualne spostrzeżenie można sprawdzić p-wartość dotyczącą hipotezy mówiącej, że czynniki (wykryte [%] oraz niewykryte [%]) nie mają wpływu na zmienną odpowiedzi (Przypadek). P-wartość wynosi dla obu czynników 0,6524 a to oznacza że nie mamy mocnych podstaw do odrzucenia hipotezy zerowej, zatem można stwierdzić, że czynniki te nie mają wpływu na rodzaj nowotworu. Innymi słowy detekcja obiektów jest podobna w przypadkach złośliwych oraz łagodnych. Nieco inaczej kształtuje się wykres dla obiektów nad-segmentowanych. Tu już widać tendencję, że większa liczba obiektów nad-segmentowanych dotyczy przypadków złośliwych. Uzyskana p-wartość wynosi 0,0132 zatem jest niższa od przyjętego poziomu ufności 0,05, a to oznacza że mamy przesłanki do odrzucenia hipotezy zerowej oraz przyjęcia hipotezy alternatywnej mówiącej, że czynnik w postaci procentowej liczby obiektów nad-segmentowanych ma wpływ na rodzaj nowotworu na obrazie. Ogólnie można zauważyć znacznie mniejszy wpływ rodzaju nowotworu na wynik detekcji.

Ostatnia analiza wyników segmentacji będzie dotyczyć dopasowania zbiorów danych oraz przypadków nowotworów do metryk ewaluacyjnych (JI, DH). Wyniki analizy widoczne na rysunku 3.19 sugerują, że rezultaty z obu metryk ewaluacyjnych mają wpływ zarówno na zbiór, z którego pochodzi obraz jak i rodzaj nowotworu. Generalnie można zauważyć, że przypadki łagodne uzyskały lepsze wyniki ewaluacyjne niż złośliwe, choć w przypadku JI nie jest to tak oczywiste, gdyż przy wyższych wartościach dopasowania obiektów przewagę w ewaluacji uzyskują przypadki złośliwe. Bardziej jednoznaczne wyniki ewaluacji dotyczą zbiorów, z których pochodzą próbki. W tym przypadku lepsze wartości uzyskują próbki ze zbioru z Zielonej Góry, zarówno dla metryki DH jak i JI.

# 3.6 Podsumowanie

W rozdziale tym opisano proces segmentacji obrazów cytologicznych i histopatologicznych przy użyciu hybrydowej metody bazującej na sieciach CNN oraz algorytmie wododziałowym sterowanym markerami. Głównym narzędziem segmentacji była opracowana metoda, która wykorzystuje dwie sieci CNN i algorytm wododziałowy z markerami. Pierwsza sieć dokonywała segmentacji obrazu, druga natomiast wykonywała zadanie detekcji punktów centralnych jąder komórkowych. Dane przygotowane w ramach badania pobrane zostały z dwóch różnych ośrodków medycznych. Pomimo niewielkiej liczby oznaczonych przykładów, sieć bardzo dobrze poradziła sobie z problemem segmentacji osiągając rezultaty porównywalne z wynikami w literaturze [53]. Zaletą proponowanego podejścia względem innych podejść [53] jest znacznie krótszy czas wykonania pojedynczego eksperymentu. Aby przeprowadzone w rozprawie badanie było rzetelne, wymagało ono zastosowania walidacji krzyżowej w wersji *ang. Leave-One-Out.* Podejście to znacznie wydłuża łączny czas trwania eksperymentu, lecz pozwala uniknąć sytuacji, w której obrazy od jednego pacjenta znajdą się zarówno w zestawie danych uczących jak i testowych. Czas trwania eksperymentów z segmentacją wyniósł łącznie kilka tygodni w przypadku modeli segmentacji z artykułu [53], czas trwania badań uległby wielokrotnemu wydłużeniu.

Pomimo uzyskania podobnych rezultatów oceny ilościowej w proponowanym podejściu względem literatury [53], ciężko jest w pełni porównać uzyskane wyniki. Problem ten wynika z braku standardowej bazy danych oraz wglądu w podział wykorzystanych baz w trakcie tworzenia danych uczących oraz testowych. Wykorzystane dane różnią się również jakością oraz rozdzielczością. Ogólnie można jednak zauważyć, że wyniki segmentacji uzyskiwane na bazie opracowanej metody są wysokiej jakości oraz metoda ta posiada duże zdolności uogólniające.



Rysunek 3.16: Dopasowanie wyników segmentacji do ośrodka pochodzenia próbek



Rysunek 3.17: Dopasowanie wyników segmentacji do utworzonych zbiorów



Rysunek 3.18: Dopasowanie wyników segmentacji do przypadków złośliwych i łagodnych



Rysunek 3.19: Dopasowanie wyników metryk segmentacji do pochodzenia ze zbiorów (a, b) oraz przypadków nowotworów (c, d)

# Rozdział 4

# Kompleksowy system klasyfikacji

# 4.1 Wprowadzenie

Klasyfikacja stanowi jedno z głównych zagadnień uczenia maszynowego. Klasyfikator jest natomiast rodzajem algorytmu statystycznego, którego zadaniem jest przyporządkowanie obserwacji do odpowiednich klas bazując na cechach badanej obserwacji. Klasyfikacja należy do grupy technik uczenia nadzorowanego. Zatem aby dokonać procesu klasyfikacji należy posiadać zbiór uczący z poprawnie zidentyfikowanymi przynależności poszczególnych obserwacji do klasy.

# 4.2 Ekstrakcja cech manualnych

Ekstrakcja cech (rys. 4.1) stanowi metodę redukcji wymiarowości stosowaną często do generowania cech opisujących obiekty na obrazie cyfrowym. Zbudowany zestaw cech może następnie posłużyć do klasyfikacji obrazów. Ekstrakcja zbioru cech może bazować na wiedzy dziedzinowej. W przypadku tworzenia cech na bazie obrazów cyfrowych niektóre środowiska programistyczne (m.in. funkcja regionprops w oprogramowaniu Matlab oraz Python) posiadają zaimplementowane narzędzia określające podstawowe właściwości obiektów. Jedną z współcześnie opracowanych metod gradingu inwazyjnego raka piersi jest



Rysunek 4.1: Ekstrakcja cech

skala Blooma-Richardsona w modyfikacji Nottingham. Metoda została opublikowana w roku 1957, a jeden ze stopni omawianej skali bazuje na polimorfizmie jąder komórkowych. Ocenie podlega kształt i wielkość jąder komórkowych:

- jądra komórkowe małe, o jednorodnym kształcie i jednakowej wielkości
- jądra komórkowe średnie lub duże, przeważnie jednak podobnych kształtów i rozmiarów
- jądra komórkowe duże z wyraźną rozbieżnością kształtów i rozmiarów

Ponadto lekarze wymieniają takie cechy jak:

- odległości pomiędzy sąsiednimi jądrami,

- stopień nakładania się na siebie jąder komórkowych.

Granica pomiędzy przypadkiem złośliwym i łagodnym jest bardzo nieprecyzyjna. Jedynie doświadczenie nabierane w ramach praktyki lekarskiej oraz nierozpoznane, trudne do zdefiniowania cechy jąder komórkowych prowadzą do skutecznej diagnozy.

Segmentacja przy użyciu sieci CNN daje możliwość obliczenia podstawowych parametrów wykrytych jąder komórkowych. Każde jądro komórkowe zostało scharakteryzowane jednym z 31 parametrów. Następnie dla każdego obrazu o rozmiarze 230x230 zostały policzone statystyki znajdujących się na nim jąder komórkowych. Statystyki te obejmowały wartość średnią, medianę, odchylenie standardowe, kurtozę, skośność, entropię, statystykęk, oraz rozstęp między kwartylowy. 31 parametrów pomnożone przez 8 statystyk dało w rezultacie 248 cech obiektów dla każdego obrazu plus 3 cechy gradientowe, co w rezultacie daje łącznie 251 cech manualnych. Na bazie tych cech zbudowany zostanie model klasyfikujący obrazy.

Zestaw cech jąder komórkowych można podzielić na trzy główne podgrupy. Pierwsza podgrupa definiuje fizyczne parametry jąder komórkowych bazując na binarnej masce obiektów. Do cech tych należy Obszar, ObszarBB (*ang. Bunding Box*), Obwód, Oś Wielka, Oś Mała, Ekscentr, Średnica odpowiadająca okręgowi wokół obiektu, Kolistość, Proporcje obiektu (Aspect), Eliptyczność, Szkielet.

Druga grupa cech związana jest z rozmieszczeniem organelli i chromatyny wewnątrz jądra komórkowego. Cechy te możemy określić jako teksturowe i badane są na podstawie macierzy współwystępowania poziomów szarości (GLCM) [30] i macierzy długości przebiegu na poziomie szarości (GLRLM) [25, 101].

Macierz GLCM powstaje w wyniku badania wartości sąsiednich pikseli. Weryfikacja polega na sprawdzeniu jakie wartości posiadają sąsiadujące ze sobą piksele. Zakładając najprostszy przypadek, w którym sprawdzamy dwa wertykalnie sąsiadujące ze sobą piksele, możemy sprawdzić ile w całej przestrzeni obrazu jest przypadków, gdy przykładowo piksel o wartości 2 znajduje się nad pikselem o wartości 152. Zliczając liczbę takich przypadków możemy ją zapisać w 3 wierszu oraz 153 kolumnie (przy założeniu adresowania od wartości 0). Stąd łatwo wywnioskować, iż przestrzeń dostępnych par porównań ograniczona jest przez wartości pikseli czyli w przestrzeni obrazu ośmiobitowego jednokanałowego jest to przestrzeń od 0 do 255. To w konsekwencji dało by macierz rozmiaru 256 na 256 na której sumowana jest liczba wystąpień poszczególnych par. Dlatego tuż przed wykonaniem macierzy GLCM obraz jest poddawany konwersji tak, aby wartości na obrazie oscylowały w granicach od 0 do 32. W ten sposób łatwiej będzie odnaleźć sąsiadujące piksele o zbliżonej intensywności. Kolejny etap tworzenia macierzy GLCM to normalizacja wartości, która sprowadza się do podzielenia wartości każdej pojedynczej komórki powstałej macierzy przez sumę wartości wszystkich komórek macierzy. Macierz następnie przeliczana jest do ostatecznej wersji na podstawie uśrednienia wartości dla powstałych macierzy w kierunkach 0°, 45°, 90° i 135°. Ostatni parametr to odległość, w której badane są pary i wynosi ona 5 pikseli. Na podstawie powstałej macierzy GLCM obliczone a następnie uśrednione zostały jej standardowe statystyki takie jak: kontrast, niepodobieństwo, jednorodność, energia, korelacja oraz drugi moment kątowy. Z uśrednionych statystyk następnie obliczone statystyki w postaci: średniej, mediany, odchylenia standardowego, kurtozy, skośności, entropii, statystykę-k, oraz rozstęp międzykwartylowy. Końcowa liczba cech z macierzy GLCM wyniosła 48.

Kolejny zestaw cech teksturowych bazował na GLRLM (ang. Gray Level Run Length Matrix) [25], niestety nazwa macierzy nie figuruje w polskim tłumaczeniu. GLRLM można zdefiniować jako macierz  $N \times M$  p, gdzie N to liczba poziomów szarości, a M to maksymalna długość przebiegu, definiowana dla danego obrazu jako liczba przebiegów z pikselami poziomu szarości i oraz długość j. Podobnie jak w przypadku GLCM, obliczamy macierze długości przebiegu (ang. run level) dla 0°, 45°, 90° i 135° przy użyciu pięciu poziomów szarości: SRE (ang. Short Run Emphasis) - nacisk na krótki okres GLRLM, LRE (ang. Long Run Emphasis) - nacisk na długi okres GLRLM, GLU (ang. Gray Level Uniformity) - niejednorodność poziomu szarości GLRLM, RLU (ang. Gray Level Uniformity) - niejednorodność długości serii GLRLM, RPC (ang. Run Percentage) - procent przebiegu GLRLM.

Ostatnia grupa cech związana jest z cechami kolorymetrycznymi jąder, są to: średnia wartość kanału czerwonego (MR), średnia wartość kanału zielonego (MG), średnia wartość kanału niebieskiego (MB), średnia wartość jasności (ML), wariancja wartości kanału czerwonego (VR), wariancja wartości kanału zielonego (VG), wariancja wartości kanału niebieskiego (VB), wariancja wartości jasności (VGR), entropia wartości jasności (ENGR).

Przygotowano także niewielki zestaw cech gradientowych bazujących na tensorze strukturalnym (opisują one rozkład gradientu w sąsiedztwie punktu). Cechy te trudno sklasyfikować do jednej z głównych kategorii cech. Do zestawu cech dołączane są maksymalne wartości składowych wektora strukturalnego biorąc pod uwagę cały obrazek i wszystkie obliczone tensory. Składowe wektora są 4 z czego 2 są jednakowej wartości. Ostatecznie otrzymano 3 nowe cechy.

Cechy morfometryczne	
Obszar, ObszarBB (Bounding Box), Obwód, Oś wielka, Oś mała,	
Ekscentr, Średnica (równoważna), Kolistość, Współczynnik proporcji (Aspect), Eliptyczność, Szkielet	
Cechy kolorymetryczne	
Średnia wartość kanału czerwonego (MR), Średnia wartość kanału zielonego (MG),	
Średnia wartość kanału niebieskiego (MB), Średnia wartość intensywności w odcieniach szarości (MGR),	
Wariancja kanału czerwonego (VR), Wariancja kanału zielonego (VG),	
Wariancja kanału niebieskiego (VB), Wariancja intensywności odcieni szarości (VGR),	
Entropia intensywności odcieni szarości (ENGR)	
Cechy teksturowe	
GLCM	GLRLM
Kontrast (GLCMContrast),	<ul> <li>SRE (Short Run Emphasis) - nacisk na krótki okres,</li> <li>LRE (Long Run Emphasis) - nacisk na długi okres,</li> <li>GLU (Gray Level Uniformity) - niejednorodność poziomu szarości,</li> <li>RLU (Gray Level Uniformity) - niejednorodność długości serii,</li> <li>RPC (Run Percentage) - procent przebiegu</li> </ul>
Korelacja (GLCMCorrelation),	
Energia (GLCMEnergy),	
Jednorodność (GLCMHomogenity),	
Niepodobieństwo (GLCMDissimilitary),	
Drugi moment kątowy (GLCMASM)	
Cechy gradientowe	
Składowe tensora strukturalnego: Tensor str arr, Tensor str acr, Tensor str acc	

Tabela 4.1: Lista wyodrębnionych cech jąder komórkowych

### 4.3 Ekstrakcja cech głębokich

Podejściem o wysokim stopniu skuteczności jest wyodrębnienie cech głębokich na podstawie wag uzyskanych dla neuronów w pośrednich warstwach. Sieć neuronowa w trakcie uczenia się dostosowuje wagi neuronów do odpowiedzi generującej najmniejszą stratę. W przypadku klasyfikacji binarnej ostatnia warstwa może przyjmować strukturę jednego neuronu, do którego napływają informacje z warstw wyższych. W zaproponowanym podejściu można wyodrębnić wartości wyjść neuronów z wytrenowanej sieci podczas przejścia nowego obrazu przez sieć. Pozyskane w ten sposób wartości stanowią zestawy cech głębokich.

#### 4.3.1 Uczenie maszynowe

Głównym tematem poruszanym w pracy jest zagadnienie głębokich sieci neuronowych, które stanowią część rozbudowanej tematyki znanej pod nazwą uczenia maszynowego, które z kolei stanowi część działu informatyki zwanego sztuczną inteligencją (Rys. 4.2).

Podejście symboliczne polega na tworzeniu logicznych modeli badanego problemu w formie zbioru reguł sformułowanych i rozumianych przez człowieka. Zbiór reguł z kolei pozwala na wykonywanie inteligentnych zadań w zakresie reguł zawartych w implementacji. Podejście symboliczne dominowało na świecie od lat 50 do niemal 90 XX wieku, lecz jest ono nadal powszechnie stosowane, ponieważ w wielu prostych zadaniach zastosowanie uczenia maszynowego jest nieuzasadnione. Wśród najciekawszych przykładów podejścia symbolicznego można wymienić: logikę rozmytą oraz algorytmy genetyczne.

Uczenie maszynowe natomiast tworzy reguły na podstawie dostarczanych danych.



Rysunek 4.2: Schemat działu sztucznej inteligencji [10]

Oczywiście ciężko w pełni zdefiniować czym jest uczenie maszynowe w kilku zdaniach, więc warto wymienić kilka podstawowych zagadnień jakie pojawiają się w dziele uczenia maszynowego. Jednym z podstawowych zadań uczenia maszynowego jest klasyfikacja. Zadaniem procesu klasyfikacji jest przyporządkowanie elementów z zadanego zbioru do odpowiednich klas po uwzględnieniu cech badanych obiektów. Z kolei cechy obiektów pozyskiwane są w procesie ekstrakcji cech, a następnie redukcji wymiarowości do zbioru istotnych cech. Proces ekstrakcji i redukcji wymiarowości może się odbywać w pełnej kontroli przez człowieka lub w przypadku sztucznych sieci neuronowych w postaci "czarnej skrzynki", do której trafiają przygotowane uprzednio zbiory uczące na podstawie, których sieć interpretuje najistotniejsze cechy badanych obiektów w kolejnych epokach uczących.

#### 4.3.2 Elementy sztucznych sieci neuronowych

Jako początek zagadnienia sztucznych sieci neuronowych powszechnie uważa się zbudowanie perceptronu [76]. Nieco ponad dekadę później zdefiniowano ograniczenia perceptronu polegające na możliwości rozwiązywania problemów jedynie separowanych liniowo [60]. Krytyka perceptronu spowodowała spadek zainteresowania technologią. Kolejną dekadę później pojawiła się koncepcja sieci CNN oraz zastosowania komórek złożonych, czyli idei podobnej do współczesnych warstw łączących (*ang. pooling*) [23]. Uczenie złożonych sieci neuronowych nie mogło zostać przeprowadzone bez odpowiedniego procesu uczenia, więc aby rozwiązać ten problem opracowany został algorytm wstecznej propagacji błędu [78]. Oprócz nowego algorytmu niezbędne było opracowanie różniczkowalnej funkcji aktywacji neuronów. Opracowano znaczną ilość funkcji aktywacji spełniających kryterium ciągłości. Przykładami różniczkowalnych funkcji aktywacją są:

- funkcja sigmoidalna unipolarna:

$$y(x) = \frac{1}{1 + e^{-\beta x}}$$
(4.1)

- jednostronnie obcięta funkcja liniowa (ReLU)

$$f(x) = \begin{cases} 0 & \text{for } x \leq 0. \\ x & \text{for } x > 0. \end{cases}$$

$$(4.2)$$

- tangens hiperboliczny:

$$y(x) = tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
(4.3)

Prace nad rozwojem sieci CNN [54] zaowocowały połączeniem idei sieci CNN oraz algorytmu wstecznej propagacji błędu oraz uzyskaniu sieci neuronowej zdolnej do skutecznego uczenia [55]. Można zatem stwierdzić, iż potencjał sieci CNN w procesie przetwarzania obrazów cyfrowych został w praktyce wykorzystany już w latach 90 XX wieku [54], jednak niewielka grupa naukowców skupiona była wokół rozwoju tej technologii. W tym czasie dominowało podejście klasyczne polegające na budowaniu zestawu cech i klasyfikacji obiektów według zredukowanego zbioru cech. W latach 90 wysnuto wniosek, że CNN mogą pomóc pozbyć się problemu wstępnego przetwarzania obrazów i ekstrakcji cech obiektów[54]. Trwający obecnie poziom zainteresowania głębokimi sieciami neuronowym nastąpił po sukcesie sieci AlexNet w roku 2012 [49] w konkursie ImageNet.

#### 4.3.3 Splotowe sieci neuronowe

#### Wstęp

CNN to przykład sieci głębokiej zorganizowanej w sposób hierarchiczny. Kolejne warstwy dokonują przetwarzania danych z niższego poziomu. Dane uzyskane na niższym poziomie przekazywane są do kolejnej warstwy, dzięki temu sieć stopniowo uzyskuje bardziej złożone informacje. Sieci CNN znajdują przede wszystkim zastosowanie w segmentacji obrazów oraz ich klasyfikacji. Ponadto sieci CNN używane są w zagadnieniu uczenia transferowego, czyli problemie badawczym polegającym na zastosowaniu sieci wyspecjalizowanych w rozwiązywaniu jednego zagadnienia do rozwiązywanie innego problemu. Przykładowa sieć CNN została schematycznie przedstawiona na rysunku 4.3. Załączony przykład obrazuje sieć CNN przeznaczoną do klasyfikacji obrazu. Jest to standardowe zadanie, do którego przeznaczone są sieci CNN. Wejście do sieci stanowi cyfrowy obraz o określonym wymiarze. Zaprezentowana na rysunku 4.3 sieć w pierwszym kroku przetwarzania wykonuje operację splotu. Obraz wejściowy jest przetwarzany przez 32 filtry o rozmiarze jądra 3 na 3 wartości przesunięcia równej 1 oraz dopełnienia wynoszącej również 1. W zamieszczonym przykładzie została wybrana funkcja aktywacji neuronu ReLU. Po tej operacji następuje kolejny splot a następnie wykonywana jest operacja maksymalizującą warstwą łączącą (ang. maxpooling), po których następują dwa kolejne sploty a potem kolejny maksymalizującą warstwą łączącą. Następnie pojawia się połączenie z neuronami w warstwach



Rysunek 4.3: Przykładowa sieć CNN

ukrytych. Warstwa wyjściowa składa się z 4 neuronów, co oznacza, że sieć dokonywała klasyfikacja czy też segmentacji 4 rodzajów obiektów na obrazie.

#### AlexNet

Jedną z najprostszych architektur sieci CNN stanowi AlexNet, jednocześnie będąc kamieniem milowym w rozpowszechnianiu zagadnienia uczenia głębokiego. Struktura sieci AlexNet przedstawiona została na rysunku 4.4. Sieć ta została zaprojektowana do pracy

Wejście: obraz 224x224x3	
Splot 11x11 + przesunięcie 4: 54x54x96 -> ReLu	
Łącząca 3x3 + przesunięcie 2: 26x26x96	
Splot 5x5 + dopełnienie 2: 26x26x256 -> ReLu	
Łącząca 3x3 + przesunięcie 2: 12x12x256	
Splot 3x3 + dopełnienie 1: 12x12x384 -> ReLu	
Splot 3x3 + dopełnienie 1: 12x12x384 -> ReLu	
Splot 3x3 + dopełnienie 1: 12x12x256 -> ReLu	
Łącząca 3x3 + przesunięcie 2: 5x5x256 -> spłaszczenie	
Warstwa gęsta: 4096 neuronów FC -> ReLu + dropout (0,5)	
Warstwa gęsta: 4096 neuronów FC -> ReLu + dropout (0,5)	
Warstwa gęsta: 1000 neuronów FC	
Wyjście: 1 z 1000 klas	

Rysunek 4.4: Sieć AlexNet

z obrazami o wymiarze 224 x 224 oraz 3 kanałach. Głębokość sieci wynosi 8 warstw, w tym 5 warstw splotowych oraz 3 warstwy gęste (z neuronami w pełni połączonymi). Pomimo stosunkowo niewielkiej głębokości sieć ta posiada około 62 miliony parametrów. Obecnie dzięki rozwojowi technologii kart graficznych sieć może stanowić skuteczne i szybkie narzę-

dzie rozwiązywania problemów klasyfikacyjnych. Ponadto wciąż pomimo upływu dekady od debiutu sieć AlexNet nadal jest czasami wykorzystywana [124].

#### VGG

Kolejna grupa głębokich sieci neuronowych w zależności od liczby warstw głębokich (16 lub 19) powszechnie określane są nazwą VGG16 lub VGG19 [87]. Liczba warstw głębokich w tych sieciach zdecydowanie odbiega od większości popularnych głębokich sieci neuronowych, jednocześnie dorównując lub przewyższając pozostałe modele liczbą parametrów. Sieci VGG charakteryzuje zastosowanie stałego jądra splotu o rozmiarze 3x3 z przesunięciem oraz dopełnieniem o wartości 1. Ponadto każdy zestaw warstw splotowych kończy się maksymalizującą warstwą łączącą z przesunięciem oraz dopełnieniem o wartości 2. Trzecia zastosowana zasada polega na podwajaniu głębokości filtrów po każdej warstwie łączącej. Różnica pomiędzy siecią VGG16 umieszczoną na rysunku 4.5 a siecią VGG19 polega na



Rysunek 4.5: Sieć VGG16

zwiększeniu liczby dwóch ostatnich bloków warstw splotowych z 3 do 4. Głównym zastosowaniem sieci VGG jest rozpoznawanie oraz klasyfikacja obrazów i do tego celu została zaprojektowana. Sieć VGG16 charakteryzuje się głębokością 16 warstw oraz w pierwotnej wersji posiada 138,4 milionów parametrów. Sieć VGG19 natomiast charakteryzuje 19 warstw oraz 143,7 milionów parametrów.

#### GoogLeNet

Założenia projektowe sieci GoogLeNet [99] skupiły się na uzyskaniu wysokiej wydajności, przy jednoczesnym zachowaniu dużej głębokości sieci. Dzięki temu sieć została wyposażona w kilka innowacyjnych rozwiązań. Jedną z tych nowości było wprowadzenie nowej podsieci zwanej modułem incepcyjnym (rys. 4.6). Dodatkowo sieć GoogLeNet została wyposażona w dwie mniejsze sieci klasyfikujące, składające się z kilku warstw: łączącej uśredniającej, pojedynczej splotowej, dwóch w pełni połączonych, warstwę aktywacji softmax. Sieci te zostały podpięte przy trzecim oraz szóstym module incepcyjnym, a ich zadaniem było zwiększenie zdolności uogólniających sieci, a przede wszystkim redukcja problemu zanikania gradientu [26].



Rysunek 4.6: Moduł incepcyjny

#### ResNet

Sieć ResNet a właściwie rodzina sieci rezydualnych ResNet została zaproponowana w latach 2015-2016 w kilku artykułach dotyczących rozpoznawania obrazów [33, 34]. Sieci te charakteryzują się dużą głębokością oznaczoną w ilości warstw (18-151) oraz występowaniem połączeń pomijających pomiędzy warstwami. Celem, który przyświeca technice pomijania połączeń jest próba unikania zjawiska zanikania gradientu. Sieci ResNet wykazały wysoką skuteczność działania oraz wygrały kilka konkursów klasyfikacji obrazów. Można więc przypuszczać, że w badaniach dotyczących klasyfikacji obrazów medycznych sieci te osiągną dużą dokładność.

#### DenseNet

Sieć DenseNet [37] została zaproponowana w 2016 roku. Model ten charakteryzuje się bezpośrednimi połączeniami pomiędzy wszystkimi warstwami w celu poprawy przepływu informacji [37]. Sieci DenseNet należą do grupy sieci rezydualnych, więc w dużej mierze przypominają sieci ResNet. Kluczowa różnica polega na gęstszej koncentracji warstw w sieci DenseNet (stąd jej nazwa) oraz na sposób łączenia obiektów (w ResNet stosuje się sumowanie natomiast w DenseNet konkatenacja) [37]. Podobnie jak w przypadku wszystkich sieci rezydualnych pomijanie warstw ma na celu zapobieganie zjawisku zanikania gradientu.

#### Xception

Sieć Xception została zaproponowana przez Chollet'a [11] jako alternatywa dla dużej sieci Inception V3 (rozwinięcie GoogLeNet). Sieć Xception posiada taką samą liczbę parametrów jak InceptionV3, jednak efektywniej wykorzystuje parametry modelu [11]. Sieć Xception może byc zatem wykorzystana jako alternatywa sieci InceptionV3. To co łączy z sobą te sieci to zastosowanie modułu incepcyjnego. Moduł ten służy do redukcji wymiarowości danych w celu efektywniejszego działania i składa się z grupy filtrów konwolucyjnych zazwyczaj o rożnym rozmiarze. W sieci Xception moduł incepcyjny został zredukowany do "ekstremalnej"formy [11], w której występuje jedna warstwa splotowa wielkości 1x1, która następnie podlega podziałowi na zbiór elementów o rozmiarze 3x3. Ograniczenie liczby warstw splotowych wpływa pozytywnie na wydajność sieci Xception.

#### 4.3.4 Metody uczenia zespołowego

Jednym ze sposobów poprawy przewidywania przy użyciu klasyfikatorów jest łączenie ich w zespoły. Idea zakłada, że większa liczba słabszych modeli przewidujących ten sam problem może zwiększyć dokładność tej prognozy. W literaturze występuje kilka podstawowych metod klasyfikacji zespołowej takich jak: agregacja (*ang. bagging*), wzmacnianie (*ang. boosting*), las losowe (*ang. random forest*), głosowanie (*ang. voting*) oraz kontaminacja (*ang. stacking*).

#### Głosowanie

Jednym z podstawowym przykładowych budowania zespołów klasyfikatorów jest głosowanie. Metoda ta polega na podejmowaniu decyzji klasyfikacyjnej na podstawie głosowania większościowego. Wynik głosowania  $\hat{y}$  pozyskiwany na bazie danych uczących (x) wykorzystanych podczas budowania modelu klasyfikatora (C):

$$\hat{y} = moda\{C_1(x), C_2(x), \dots C_n(x)\}$$
(4.4)

Najczęściej zespoły składają się z heterogenicznych modeli klasyfikatorów z użyciem zestawu danych uczących. Celem stosowanej różnorodności modeli jest eliminacja błędów systematycznych. Modyfikacją głosowania większościowego jest metoda głosowania ważonego, w której klasyfikatory mają różne wagi, natomiast uzyskany wynik powstaje na bazie średniej ważonej.

#### Agregacja

Innym podejściem do rozwiązania problemu zespołu klasyfikatorów jest metoda agregacji. Zadanie polega na losowaniu ze zwracaniem wielu próbek danych uczących. Wylosowane próbki biorą udział w strojeniu podobnych modeli klasyfikatorów. Podobnie jak w przypadku metody głosowania ostateczny wynik klasyfikacji otrzymywany jest przy zastosowaniu dominanty wyniku modeli składowych.

#### Lasy losowe

Rozwinięciem metody agregacji są lasy losowe. W przypadku tej metody bazowymi klasyfikatorami są drzewa decyzyjne. Każde drzewo modelowane jest na losowym zestawie uczącym. Końcowa decyzja jest podejmowana przy zastosowaniu głosowania większościowego.

#### Wzmacnianie

Idea wzmacniania polega na budowaniu sekwencyjnego zespołowego klasyfikatora mocnego na bazie słabych klasyfikatorów w taki sposób, że kolejny model stara się poprawić wynik klasyfikatora poprzedzającego. [26]. Wzmacnianie może przybrać formę jednego z podstawowych algorytmów np.: wzmacnianie adaptacyjne (AdaBoost), wzmacnianie gradientowe.

#### Kontaminacja

Koncepcja kontaminacji różni się od pozostałych metod zespołowych tym, że do podejmowania decyzji nie wykorzystuje się prostej funkcji głosującej lecz kolejny klasyfikator nadrzędny.

#### Generator cech głębokich

W ramach pracy zaproponowano procedurę budowania generatorów cech głębokich inspirowaną zagadnieniem zespołów klasyfikatorów. Ogólna idea polega na wzmacnianiu modeli klasyfikatorów w wyniku ich łączenia. Proponowana metoda bazuje w głównej mierze



Rysunek 4.7: Generatory cech głębokich

na metodach zespołowych, gdzie występuje losowy zestaw danych oraz heterogeniczny zespół klasyfikatorów. W tym przypadku zestaw klasyfikatorów zastępowany jest przez generatory cech głębokich. Decyzja klasyfikacyjna nie jest podejmowana przez wynik poszczególnych klasyfikatorów, lecz na podstawie cech głębokich uzyskanych z przedostatniej warstwy neuronów. W celu urozmaicenia zespołu wybrano sieci neuronowe reprezentujące różne grupy sieci takie jak: ResNet, DenseNet, VGG, InceptionResNetV2 oraz Xception. Heterogeniczność sieci ma na celu zapewnić różnorodność cech podobnie jak ma to miejsce w metodzie zespołu klasyfikatorów głosujących.

Każdy z modeli sieci neuronowych wchodzących w skład zespołu klasyfikatorów w przedostatniej warstwie otrzymał po 5 neuronów z funkcją aktywacji ReLU. Ograniczenie to nie wpłynęło negatywnie na jakość uzyskanych wyników dokładności. Ponadto, zastosowanie jedynie 5 neuronów w przedostatniej warstwie spowodowało uzyskanie 25 cech głębokich z 5 badanych sieci. Łączna liczba cech głębokich nie zmienia zatem drastycznie łącznej przestrzeni po fuzji cech (251 cech manualnych oraz 25 głębokich). Wygenerowany zestaw cech głębokich charakteryzuje się wysoką korelacją cech w obrębie poszczególnych klasyfikatorów oraz znacznie mniejszą pomiędzy klasyfikatorami. Niestety duży wpływ na korelację ma charakter uzyskanych rozkładów wartości cech głębokich, gdyż neurony maja tendencję do uzyskiwania zerowych wartości. Rozkłady cech głębokich charakteryzują się, więc inflacją wartości zero. Wykorzystanie innych funkcji aktywacji nie rozwiązuje problemu, ponadto uzyskiwane wyniki klasyfikacji ulegają nieznacznemu pogorszeniu.

# 4.4 Redukcja wymiarowości

Redukcja wymiarowości stanowi ważne zadanie w procesie klasyfikacji. Duża liczba cech stanowi podstawę do konieczności jej redukcji. Powodem takiego stanu rzeczy jest występowanie zjawiska przekleństwa wymiarowości. Zjawisko to związane jest z wzrostem złożoności klasyfikatora spowodowanym wzrostem liczby parametrów, których liczba z kolei wynika z wielkości wektora cech badanych obiektów. Zbyt duża liczba cech może również wywołać efekt przeuczenia, czyli nadmiernego dopasowania klasyfikatora do zbioru uczącego. Zjawisko przeuczenia powoduje obniżenie wyników uzyskiwanych na zbiorze testowym. Metody redukcji wymiarowości możemy podzielić na dwie zasadnicze grupy: metody selekcji cech oraz metody projekcji cech.

#### 4.4.1 Ekstrakcja oraz konstrukcja nowych cech

Ekstrakcja cech również zaliczana jest do metod redukcji wymiarowości (np. obliczanie powierzchni lub obwodu jądra komórkowego stanowi redukcję z przestrzeni obrazu na konkretną cechę).

$$\begin{vmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ c_n \end{vmatrix} \rightarrow \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} = f \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ c_n \end{bmatrix} \end{pmatrix}$$

Również konwersja obrazu do odcieni szarości czy też binaryzacja obrazu stanowią metody redukcji wymiarowości. Są one tak powszechne, że czasem wręcz niezauważane, stanowiąc jedynie element wstępnego przetwarzania obrazu. Ekstrakcja cech również często, stanowi wstęp do dalszej redukcji przy użyciu np. metod selekcji cech. Oprócz ekstrakcji można też wymienić inne metody projekcji cech zaliczane do metod redukcji wymiarowości natki jak np. Analiza głównych składowych (PCA). Metoda ta polega na utworzeniu w×r wymiarowej macierzy transformacji, która umożliwi rzutowanie w-wymiarowej przestrzeni na mniejszą r-wymiarową [74]. Zadaniem analizy PCA jest przekształcenie wygenerowanej macierzy w taki sposób aby zmaksymalizować wariancję pierwszej oraz kolejnych składowych. PCA można przeprowadzić na zmiennych wejściowych poddanych normalizacji. Wówczas możliwe jest zastosowanie PCA bazującego na macierzy kowariancji. Problem różnoskalowych danych wejściowych można obejść stosując model PCA oparty o macierz korelacji. Wówczas model nie jest wrażliwy na różnoskalowość danych. Do konstrukcji nowych cech można również wykorzystać wiedzę dziedzinową, bazując na doświadczeniu diagnostów oraz implementując rozwiązania, na które diagnosta zwraca uwagę podczas badania próbki.

#### 4.4.2 Selekcja cech

Selekcja cech stanowi jedną z metod redukcji wymiarowości. Zakres jej zastosowania jest związana z budowaniem modeli regresyjnych, które wykazują się wrażliwością na nadmierną liczbę cech wejściowych. Fakt ten nie stoi w sprzeczności z możliwością zastosowania selekcji cech w innych modelach klasyfikacji. Selekcję cech można podzielić na 3 główne grupy metod: opakowane, fltrujące oraz wbudowane.



#### Metody opakowane

Rysunek 4.8: Selekcja cech - metody opakowane (wrappery)

Metody opakowane (Rys. 4.8) do uzyskania odpowiedniego podzbioru wymagają kolejnych kroków uczących. Problem w zasadzie ograniczony jest do zagadnienia wyszukiwania podzbioru cech. Wadą tych metod jest koszt obliczeniowy. Selekcja w przód - jest to metoda selekcji polegającej na dodawaniu kolejnych cech. Algorytm rozpoczyna się od pustego zestawu cech, następnie za pomocą wybranego kryterium (np.: Akaikego [2], bayesowskiego Shwarza [84]) sprawdza się czy dodawana cech poprawia wartość kryterium. Można również badać istotność cech w modelu i dodawać te cechy, które spełniają wymagany poziom ufności. Dodawanie kolejnych cech odbywa się w sposób iteracyjny, więc jako punkt zatrzymania działania algorytmu można wybrać np. wybraną liczbę kolejnych iteracji, które nie poprawiają wyniku walidacji.

Selekcja wstecz - jest to metoda selekcji cech analogiczna do selekcji w przód z tą różnicą, iż algorytm rozpoczyna działanie z modelem zawierającym wszystkie cechy. Iteracja natomiast polega na odrzucaniu najmniej istotnej cechy, aż do momentu w którym model przestanie się poprawiać.

Selekcja mieszana - jest to metoda selekcji łącząca podejście w przód oraz wstecz. Algorytm może odrzucać cechę lub ją dodawać w zależności od tego czy taka czynność poprawi jakość modelu.



#### Metody filtrujące

Rysunek 4.9: Selekcja cech - metody filtrujące

Metody filtrowania (Rys. 4.9) są zwykle używane jako etap wstępnego przetwarzania. Wybór funkcji jest niezależny od algorytmów uczenia maszynowego. Zamiast tego, cechy są wybierane na podstawie ich wyników w różnych testach statystycznych.

Zazwyczaj przy niewielkich zestawach cech wystarczającym filtrem służącym do eliminacji cech jest zastosowanie liniowej korelacji Pearsona [67], która wyraża się wzorem:

$$r_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{(\Sigma(x_i - \bar{x})^2)(\Sigma(y_i - \bar{y})^2)},\tag{4.5}$$

gdzie  $r_{xy}$  to współczynnik korelacji zmiennych,  $x_i$  to pojedyncza obserwacja ze zbioru x,  $\bar{x}$  to średnia wartość obserwacji w zbiorze x,  $x_i$  to pojedyncza obserwacja ze zbioru x,  $\bar{x}$  to średnia wartość obserwacji w zbiorze x. Jeśli zatem zmienna x jest mocno skorelowana ze zmienną y to można jedną z tych cech wyeliminować z zestawu. Analiza ta jest skuteczna dla mniejszych zestawów cech. Jednak w przypadku posiadania większego zestawu cech
warto zbadać je pod kątem związania z większą liczbą cech jednocześnie. Do tego celu idealnym narzędziem jest współczynnik inflacji wariancji (VIF), definiowany wzorem:

$$VIF_m = \frac{1}{1 - R_m^2},$$
(4.6)

gdzie  $R^2$  oznacza współczynnik determinacji dla czynnika m. Współczynnik determinacji służy do określania dopasowania modelu do danych i określany jest w zakresie od 0 do 1 (im wyższa wartość tym zmienna jest lepiej objaśniania). Innymi słowy  $R^2$  wskazuje jaki bardzo zmienna odpowiedzi jest objaśniania przez model. Jeżeli zatem obliczony zostanie współczynnik  $R^2$  dla konkretnej zmiennej to można zauważyć, że im bardziej zmienna ta jest objaśniania przez pozostałe zmienne tym wyższa wartość  $R^2$ , a co za tym idzie VIF. Przyjmuje się, że z modelu należy usuwać takie zmienne, dla których współczynnik VIF przekracza wartość 10. Zmienna, która uzyska wartość 10 dla współczynnika VIF jest objaśniania w 0.90 przez pozostałe zmienne. Współczynniki te mogą być wykorzystywane w przypadku zmiennych ciągłych. Jeżeli jednak dane do modelu mają charakter nominalny wówczas odpowiednim filtrem może okazać się test niezależności  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^m \sum_{i=j}^n \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}},$$
(4.7)

gdzie  $n_i j$  oznacza liczebności empiryczne, natomiast  $\hat{n_i j}$  teoretyczne.

## Metody wbudowane



Rysunek 4.10: Selekcja cech - metody wbudowane

Metody wbudowane (Rys. 4.10) to grupa technik, w których selekcja cech jest efektem procesu budowy modelu. Działanie metod wbudowanych można wyjaśnić na podstawie regresji liniowej. Model regresji liniowej jest uczony aż do momentu minimalizacji sumy kwadratów reszt pomiędzy wartościami prawdziwymi a prognozowanymi (zwanej w literaturze RSS lub SSE) [3]. Omawiana suma kwadratów wyraża się wzorem:

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \qquad (4.8)$$

gdzie  $y_i$  oznacza wartości rzeczywiste, natomiast  $\hat{y}_i$  oznacza wartości prognozowane. Modele regresyjne mogą podlegać regularyzacji. W takim przypadku do funkcji kosztu dodawany jest człon odpowiedzialny za regularyzację [26]. Zadaniem regularyzacji jest zmniejszenie wariancji modelu [3].

Popularną metodą regularyzacji modeli regresyjnych jest Lasso [104]. Nazwa regularyzatora pochodzi od angielskich słów *least absolute selection and shrinkage operator* co można przetłumaczyć jako najmniejszy bezwzględny operator selekcji i redukcji.

$$SSE_L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \qquad (4.9)$$

Metoda Lasso przypomina zatem metodę grzbietową z tą różnicą, iż zamiast sumy kwadratów wag bierze pod uwagę sumę wartości bezwzględnych wag modelu podczas przemnażania przez stopień regularyzacji.

Nieco nowsze podejście nazwane elastyczna siatka (*ang.* Elastic Net [125]) łączy w sobie obie wyżej wymienione metody regularyzacji otrzymując wzór:

$$SSE_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \qquad (4.10)$$

Według [26] metoda elastycznej siatki jest skuteczniejsza od metody Lasso w przypadku, gdy istnieją silne związki korelacyjne pomiędzy kilkoma cechami.

Oprócz metod z wbudowaną selekcją cech należy wspomnieć o metodzie regularyzacji L2. Metoda ta charakteryzuje się brakiem wbudowanej selekcji cech. Regularyzacja L2 zwana też regresją grzbietową [35].

$$SSE_G = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2, \qquad (4.11)$$

gdzie hiperparametr  $\lambda$  określa stopień regularyzacji [26]. Wartość  $\lambda$  równa 0 oznacza brak regularyzacji. Jeśli wartość hiperparametru  $\lambda$  jest wysoka oznacza to, że wagi ( $\beta_j$ ) modelu będą bliskie 0 [26].

## 4.4.3 Stochastyczna selekcja cech

W przeprowadzonych badaniach najwyższe średnie wyniki dokładności klasyfikacji obrazów cytologicznych i histopatologicznych uzyskano dla modeli regresji logistycznej przy użyciu metody selekcji cech w przód oraz metody regularyzacji elastycznej siatki. Praktyczne problemy jakie pojawiły się podczas testowania danych na zbudowanych modelach w przypadku selekcji cech w przód dotyczyły stosunkowo słabej klasyfikacji obrazów zawierających przypadki złośliwe, przy jednoczesnej ponadprzeciętnej poprawnej klasyfikacji przypadków łagodnych. Modele utworzone przy użyciu regularyzacji elastycznej siatki są z kolei trudne do interpretacji ze względu niewystarczającą redukcję wymiarowości modeli. Obie wymienione metody uzyskały zbliżone wyniki dokładności klasyfikacji sięgające nieco ponad 0.78. Alternatywne rozwiązanie problemu selekcji cech może bazować na wykorzystaniu metod stochastycznych. Dokładne przeszukiwanie wiąże się jednak z ogromnym kosztem obliczeniowym. Ten fakt przyczynił się do zaproponowania podejścia ograniczającego liczbę potencjalnych cech potrzebnych do uzyskania dokładnego modelu klasyfikacji.

Założenie polega na losowaniu długości wektora cech wykorzystanych w uczeniu modelu regresji logistycznej. Jednakże wektor może zawierać dowolny zestaw cech, zatem moc przestrzeni zdarzeń elementarnych  $|\Omega|$  wyraża się wzorem:

$$|\Omega| = \sum_{k=1}^{n} \frac{n!}{k!(n-k)!},\tag{4.12}$$

gdzie n oznacza łączną liczbę cech, natomiast k oznacza długość wektora cech. Ogromna liczba możliwości wiąże się z ogromną ilością czasu potrzebną do przeszukiwania takiej przestrzeni, zatem w kolejnym kroku zostaną zastosowane pewne ograniczenia.

Można założyć, że długość wektora cech ma wpływ na prawdopodobieństwo uzyskania wysokiego wyniku klasyfikacji. Aby zweryfikować postawione założenie wykonano doświadczenie losowania 100 zestawów cech o losowej długości z zakresu od 1 do n - 1, gdyż zestaw wszystkich n cech został przetestowany w ramach wcześniejszych eksperymentów. Ze wszystkich 100 zestawów wybierany jest ten, który uzyska najwyższy wynik klasyfikacji na losowo wydzielanych danych walidacyjnych. Całe doświadczenie zostało powtórzone 100 razy, w celu uzyskania empirycznego rozkładu długości wektorów cech najlepszych zestawów. Uzyskane rozkłady empiryczne charakteryzują się prawostronną



Rysunek 4.11: Przykładowe rozkłady empiryczne

asymetrią, przypominając kształtem rozkład gamma. Zastosowanie rozkładu teoretycznego gamma wymaga obliczenia parametrów tego rozkładu. Do estymacji wartości parametrów można zastosować metodę momentów. Zakładając, że  $x_i$  to pojedyncza obserwacja, gdzie i = 1, 2, ..., n; ponadto jeżeli zbiór  $X \sim Gamma(\alpha, \theta)$ , to wartość  $E(X) = \frac{\alpha}{\theta}$ oraz  $E(X^2) = \frac{\alpha(1+\alpha)}{\theta^2}$ . Metoda będzie zatem polegać na rozwiązaniu układu równań:

$$\begin{cases}
\frac{\hat{\alpha}}{\hat{\theta}} = \mu_I \\
\frac{\hat{\alpha}(1+\hat{\alpha})}{\hat{\theta}^2} = \mu_{II}
\end{cases}$$
(4.13)

Podstawiając  $\mu_I$  do drugiego równania otrzymujemy:

$$\left(\frac{1}{\hat{\alpha}} + 1\right)\mu_I^2 = \mu_{II},\tag{4.14}$$

następnie:

$$\frac{1}{\hat{\alpha}} = \frac{\mu_{II}}{\mu_I^2} - 1 \Rightarrow \hat{\alpha} = \frac{\mu_I^2}{\mu_{II} - \mu_I^2} = \frac{\bar{X}^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$
(4.15)

oraz podstawiając $\hat{\alpha}$ do drugiego równania:

$$\hat{\theta} = \frac{\hat{\alpha}}{\hat{\mu}_I} = \frac{\bar{X}}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2},$$
(4.16)

gdzie  $\bar{X}$  oznacza średnią, natomiast  $\frac{\bar{X}}{\frac{1}{\bar{n}}\sum_{i=1}^{n}(X_i-\bar{X})^2}$  wariancję. Bardzo często parametr $\theta$  przekształcany jest do parametru skali  $\beta$ :

$$\beta = \frac{1}{\theta} \tag{4.17}$$

stąd ostatecznie otrzymać można parametr k<br/>ształtu  $\alpha$ oraz parametr skali $\beta :$ 

$$\alpha = \frac{\bar{X}^2}{V[X]} \qquad \beta = \frac{V[X]}{\bar{X}}, \qquad (4.18)$$

gdzie V[X] oznacza wariancję próby. W praktyce zastosowanie rozkładu gamma pozwala na ograniczenie zbioru poszukiwanych długości wektora cech. Cały proces selekcji cech składa się z dwóch głównych części. Pierwsza cześć została opisana za pomocą pseudokodu oznaczonego jako Algorytm 1. W przedstawionym algorytmie kluczowe znaczenie ma wymiar wektora cech oznaczony jako dim(x).

Po obliczeniu parametrów rozkładu gamma następuje przejście do części odpowiedzialnej za stochastyczną selekcję cech (Algorytm 2). Aby zwiększyć zdolności uogólniające losowanych modeli zastosowano 5-krotną walidację krzyżową oraz regularyzację L2, która w przeciwieństwie do L1, nie posiada wbudowanej redukcji wymiarowości. Liczba iteracji algorytmu może być ustawiona ręcznie, warto było by rozważyć również zakończenie pracy algorytmu po braku poprawy wyniku dokładności (ACC) po ustalonej liczbie iteracji.

Algorytm 2 Stochastyczna selekcja cech

1: Inicjalizuj:  $x\_Best = 0$ ; ACC = 0;  $ACC\_Best = 0$ 2: Wpisz: Plik z danymi do klasyfikacji,  $\alpha$ ,  $\beta$ 3: Wypisz: ACC, x\_Best 4: for i = 1 to k do Wylosuj liczbę x z przedziału od 1 do n-15:if x > n-1 then 6:  $n-1 \leftarrow x$ 7: end if 8: 9: Wylosuj zestaw x liczb oznaczających cechy z przedziału 1 and n-1Zbuduj model regresji logistycznej z 5-krotną walidacją krzyżową i regularyzacją 10: L2Oblicz ACC dla danych walidacyjnych 11: if  $ACC > ACC_Best$  then 12:  $ACC\_Best \leftarrow ACC$ 13: $x\_Best \leftarrow x$ 14: end if 15:16: **end for** 

# 4.5 Klasyfikatory nadrzędne

W tym podrozdziale zostaną zwięźle przedstawione klasyfikatory nadrzędne wykorzystane w badaniach empirycznych nad dokładnością klasyfikacji przy użyciu zestawów cech manualnych głębokich oraz fuzji cech. Klasyfikator nadrzędny pełni rolę podsystemu wyjściowego w opracowanym systemie klasyfikacji obrazów raka piersi (rys. 1.1). Na jego wejście trafiają w zależności od wersji systemu cechy manualne, cechy głębokie lub zestaw połączonych cech manualnych i głębokich. Na wyjściu klasyfikatora nadrzędnego otrzymujemy odpowiedź czy przypadek nowotworu jest łagodny czy złośliwy. W przeprowadzonym badaniu wykonano kompleksowe eksperymenty weryfikujące skuteczność klasyfikatorów: DT, RL, k-NN, SVM, SN, LL, drzewo wzmacniane(DW) , NB oraz możliwości klasyfikacyjne nieujętych w wynikach metod bazujących na analizie dyskryminacyjnej. Większość zastosowanych klasyfikatorów podlegała 5-krotnej walidacji krzyżowej. W przypadku metod regresyjnych lepsze wyniki uzyskiwano podczas stosowania walidacji z użyciem kryteriów informacyjnych.

# 4.6 Klasyfikacja sztuczną splotową siecią neuronową

W wyniku upowszechnienia się gotowych pakietów do budowania sztucznych sieci neuronowych, można uznać, że najprostszą i jedną z najskuteczniejszych metod klasyfikacji jest głęboka sieć CNN. W zastosowanym pakiecie Keras przygotowane zostały najpopularniejsze modele sieci neuronowych do klasyfikacji obrazów m. in.: ResNet50 [33], VGG16 [87], Xception [11]. Zasadniczo przygotowanie modelu do uczenia polega na zdefiniowaniu kilku podstawowych własności takich jak: wielkość obrazu wejściowego, dołączenie lub usunięcie górnych warstw sieci, użycie pre-trenowanych wag lub nie, zastosowanie warstwy łączącej, itp. Wadą podejścia jest zapotrzebowanie na zasoby obliczeniowe oraz czasochłonność zarówno w przygotowaniu odpowiedniej bazy jak i samych obliczeń.

# 4.7 Metody ewaluacji

Wśród metod ewaluacji klasyfikacji binarnej istnieje jedno standardowe narzędzie. Tym podstawowym narzędziem oceny jakości jest macierz pomyłek Tab. 4.2. Tablica zawiera dwie kolumny oraz dwa wiersze na klasy predykowanej oraz klasy rzeczywistej. W tabeli został pokazany jedynie ułamek z liczby współczynników jakie można z niej uzyskać, jednak z punktu widzenia bieżących eksperymentów najbardziej istotne. Kluczowe znaczenie dla eksperymentów będzie miał współczynnik dokładności (ACC), który określa poziom prawidłowego dopasowania klasyfikatora do danych. Z tego współczynnika można również wyznaczyć błąd klasyfikacji jako dopełnienie dokładności do wartości 1. Alternatywny

		Klas		
		Pozytywna	Negatywna	
Klasa	a Pozytywna TP - Prawdziwie pozytywna I		FN - Falezywie negatywna	TPR - Czułość
rzoczywieta			riv - raiszywie negatywia	TP / (TP + FN)
1 Zeczy wista	Negatywna	FP - Falezuwie pozytywna	TN - Prawdziwa nagatywna	TNR - Swoistość
	Negatywna 11 - Faiszywie pozytywna		110 - 1 Tawaziwe negaty wha	TN / (FP + TN)
		F-score	ACC - Dokładność	PPV - Precyzja
		2TP / (2TP + FP + FN)	(TP + TN) / (TP + TN + FP + FN)	TP / (TP + FP)

Tabela 4.2: Macierz pomyłek

współczynnik wskazujący na ogólną jakość modelu i segmentacji to współczynnik F (Fscore). Kolejne istotne wskaźniki to czułość (TPR) oraz swoistość (TNR). Współczynnik TPR będzie wskazywać jak dobrze zostały sklasyfikowane przypadki klasy pozytywnej, natomiast TNR generuje informację o skuteczności klasyfikacji przypadków prawdziwie ujemnych.

# 4.8 Wyniki

## 4.8.1 Schemat przeprowadzonych badań empirycznych

Zaprezentowane w tym podrozdziale wyniki dotyczą dwóch głównych podejść (rys. 4.12). Pierwsza część wyników (rozdz. 4.8.2) polega na zbadaniu dokładności klasyfikacji pojedynczych obrazów przy użyciu głębokich sieci CNN. Eksperyment rozpoczyna badanie na obrazach w przestrzeni RGB wraz z opisem zastosowanych modyfikacji modeli. Dalsza część eksperymentu zawiera również wyniki klasyfikacji wybranych sieci CNN na obrazach normalizowanych przy użyciu metody rozplotu H&E. Następnie wykorzystano nową metodę normalizacji opartą na opracowanym hybrydowym algorytmie segmentacji jąder komórkowych.



Rysunek 4.12: Schemat głównych eksperymentów

Druga część badań (rozdz. 4.8.3) polegała na weryfikacji dokładności klasyfikacji na

zestawie cech manualnych, głębokich oraz fuzji cech z wykorzystaniem opracowanego w ramach pracy kompleksowego systemu klasyfikacji (rys. 4.12). Badanie wykonano dla różnych klasyfikatorów nadrzędnych, które zdefiniowano w rozdziale 4.5.

Klasyfikacja pojedynczych małych obrazów dostarcza cennych informacji na temat dokładności klasyfikacji. Jednak pełny obraz wyników klasyfikacji zostanie uzupełniony poprzez zbadanie najlepszych klasyfikatorów na podstawie końcowego sprawdzianu weryfikujący klasyfikację na poziomie pacjentów.

## 4.8.2 Sieci głębokie

### 4.8.2.1 Plan eksperymentów

Eksperymenty na sieciach CNN składają się z trzech głównych części. Pierwsza część dotyczy uczenia sieci na bazie danych w przestrzeni RGB. W tej części opisano również strukturę modeli, wykonane modyfikacje oraz przedstawiono proces uczenia. Eksperymenty wykonano dla 7 różnych sieci głębokich (ResNet50, ResNet152, VGG16, VGG19, DenseNet121, Xception oraz InceptionResNetV2).

Druga część eksperymentów zawiera wyniki uczenia i testowania sieci na obrazach po normalizacji przy użyciu rozplotu H&E. Liczbę sieci w eksperymencie zredukowano do pięciu (ResNet50, VGG19, DenseNet121, Xception oraz InceptionResNetV2). Eksperymenty zostały zaprezentowane w formie zwięzłego opisu oraz tabeli zawierającej wyniki końcowe.

Trzecia część przedstawia wyniki uzyskane w eksperymentach przeprowadzonych na bazie obrazów po normalizacji z użyciem zaproponowanej metody segmentacji. W tej części wykonano eksperymenty dla pięciu modeli sieci głębokich. Eksperymenty zostały podsumowane w formie tabeli.

Schemat planowanych eksperymentów z użyciem sieci CNN umieszczony został na rysunku 4.13.

#### 4.8.2.2 Wprowadzenie

W zastosowanych modelach nie skorzystano z wstępnie uczonych wartości wag oraz usunięto górne warstwy w celu zbudowania własnej konfiguracji. Dodatkowo dodano do modelu pierwszą warstwę, która modyfikuje zakres wartości pikseli w kanałach obrazu z (0... 255) do (-1... 1). Zbiory danych składają się z obrazów o rozmiarze 230 x 230 pikseli. Utworzono dwa zbiory: 1. BreakHis + GZG - 12255 obrazów; 2. SzUZG - 10071 obrazów. Dzięki temu, że w obu zbiorach liczba danych uczących była stosunkowo duża oba zbiory posłużyły do zbudowania odrębnych modeli, do których testowania posłużyły dane z drugiego zbioru obrazów. Takie podejście pokaże na ile model zbudowany na danych z jednego ośrodka medycznego może być zastosowany w klasyfikacji danych z innych



Rysunek 4.13: Schemat eksperymentów na sieciach CNN

ośrodków. Innymi słowy czy uzyskane rozwiązanie będzie uniwersalne i będzie posiadać wysokie właściwości uogólniające.

## 4.8.2.3 Eksperymenty

## Obrazy w przestrzeni RGB

**ResNet** W badaniu rozpoczęto pracę od Sieci ResNet50 składającej się z 50 warstw głębokich. Wszystkie warstwy uczono od nowa na danych uczących bez użycia wstępnych wag z innych zbiorów danych. Wielkość obrazu wejściowego została zmieniona na odpowiadającej obrazom uczącym (230 x 230 x 3). Usunięto warstwę górną (top) ponieważ składała się z jednej warstwy Dense zawierającej 1000 neuronów i zastąpiono ją zestawem warstw Flatten, Dense(512), Dropout(0.5), BatchNormalization oraz Dense(1) ponieważ klasyfikator wybiera jedną z dwóch klas obrazów (łagodny, złośliwy). Dodatkowo na samym początku dodano warstwę przeskalowującą obrazy wejściowe z wartości pikseli (0..255) na (od -1 do 1). Wykonywana była również na bieżąco augmentacja w postaci obrotów oraz odbić lustrzanych. Nie zastosowano przeskalowania wielkości obiektów, ponieważ wielkość jąder komórkowych jest istotnym czynnikiem diagnostycznym.

Pierwszy zestaw uczący zawierał dane ze zbioru BreakHis + GZG. Ze zbioru wyodrębniono 20% danych (2451) i przekazano do do zbioru walidacyjnego. Do uczenia użyto model ResNet50 natomiast górny zestaw warstw zawiera: warstwę Flatten, Warstwę Dense z 512 neuronami z funkcją aktywacji ReLU, Warstwę Dropout (0.5), warstwę Batch\_Normalization oraz końcową warstwę Dense z jednym neuronem z sigmoidalną funkcją aktywacji. Wybór najlepszego modelu nastąpił w 255 epoce uczącej na podstawie uzyskania najniższego błędu na danych walidacyjnych wynoszącego 0.0264. Pozostałe parametry uzyskane w epoce 255 to dla danych uczących: dokładność 0.9944, błąd 0.0175 natomiast na danych walidacyjnych dokładność 0.9914. Uzyskany wynik na danych testowych wyniósł 0.5242, czyli niemal całkowity brak klasyfikacji danych testowych, przy błędzie wynoszącym 3,2591. Przebieg uczenia modelu (rys. 4.14) charakteryzuje się stabilnym przebiegiem w kolejnych epokach uczących. Dane walidacyjne w procesie uczenia modelu wykazują się większymi wahaniami wartości w kolejnych epokach jednak wskazują, że w kolejnych epokach fluktuacje ulegają redukcji.





Rysunek 4.14: Przebieg uczenia modelu ResNet50 dla zbioru uczącego BreakHis + GZG

Kolejny eksperyment polegał na zamianie miejscami zbioru testowego z uczącym. W wyniku tej operacji zbiór uczący oraz walidacyjny powstał na podstawie danych wyłącznie ze zbioru SzUZG, natomiast zbiór testowy powstał na bazie zestawu BreakHis + GZG. Zbiór walidacyjny zawierał 2014 (20%) obrazów wyodrębnionych ze zbioru uczącego. Wybór najlepszego modelu nastąpił w 230 epoce uczącej na podstawie uzyskania najniższego błędu na danych walidacyjnych wynoszącego 0.1123. Pozostałe parametry uzyskane w epoce 230 to dla danych uczących: dokładność 0.9767, błąd 0.0612 natomiast na danych walidacyjnych dokładność 0.9568. Uzyskany wynik na danych testowych wyniósł 0.6003, natomiast błąd 1,8699. Przebieg uczenia modelu (rys. 4.15) wykazuje się znacznie większymi fluktuacjami w kolejnych epokach dla zbioru walidacyjnego przy jednocześnie bardzo stabilnym procesie uczenia na danych uczących. Wyraźniejszy jest tu również efekt przeuczenia, gdy po przekroczeniu 230 epoki uczącej błąd dla danych walidacyjnych zaczyna rosnąć przy dalszej redukcji błędu dla danych uczących.



Rysunek 4.15: Przebieg uczenia modelu Resnet50 dla danych uczących ze zbioru SzUZG

Największa sieć z grupy oznaczona została jako ResNet152V2. Nazwa wskazuje, że sieć posiada 152 warstwy głębokie. Podobnie jak w przypadku mniejszej sieci ResNet górny zestaw warstw zawiera: warstwę Flatten, Warstwę Dense z 512 neuronami z funkcją aktywacji ReLU, warstwę Dropout (0.5), warstwę Batch\_Normalization oraz końcową warstwę Dense z jednym neuronem z sigmoidalną funkcją aktywacji. Dla danych ze zbioru BreakHis + GZG najniższy wynik straty na danych walidacyjnych wyniósł 0.0275 w 230 epoce uczącej, dokładność dla zbioru walidacyjnego 0.9898. Dla zbioru uczącego w 230 epoce uzyskano stratę na poziomie 0.0241 przy dokładności 0.9916. Niestety strata dla danych testowych wyniosła 3,6754 oraz dokładność 0.5151. Przebieg krzywych uczenia (rys. 4.16) jest bardzo podobny jak w przypadku sieci ResNet50. Występują jedynie nieco większe wahania wartości dla poszczególnych epok uczących. W początkowej fazie uczenia wyniki straty oraz dokładności na danych walidacyjnych nie odbiegają znacząco od wyników dla danych uczących, ale mniej więcej od 200 epoki uczącej zaczyna pojawiać się rozdzielenie krzywych.

Sieć ResNet152V2 ucząc się na zbiorze SzUZG uzyskała najniższy wynik błędu dla danych walidacyjnych w 286 epoce uczącej wynoszący 0.1034 przy dokładności oszacowanej na 0.9657. Dla danych uczących wartość błędu w 286 epoce wyniosła 0.0351 oraz dokładność 0.9881. Błąd dla danych testowych wyniósł 1,6730 przy dokładności na poziomie 0.6562. Uzyskana krzywa uczenia (rys. 4.17) dla zbioru SzUZG na danych walidacyjnych charakteryzuje się znacznymi fluktuacjami podczas, gdy krzywą uczenia dla danych uczących cechuje się wzorcowym przebiegiem. W okolicach 150 epoki uczącej krzywe zaczynają odstawać od siebie, co może wskazywać na postępujące nadmierne dopasowanie do danych uczących.



(a) Dokładność modelu w kolejnych epokach

(b) Strata modelu w kolejnych epokach





(a) Dokładność modelu w kolejnych epokach (b) Strata modelu w kolejnych epokach

Rysunek 4.17: Przebieg uczenia modelu Resnet152V2 dla danych uczących ze zbioru SzUZG

VGG16 oraz VGG19 Przygotowanie modelów VGG16 oraz VGG19 polegało na usunięciu górnych warstw oraz zastąpieniu ich własnym układem. Nie zostały zastosowane również żadne wstępnie nauczone wagi lecz wstępnie inicjalizowane losowo. Dodatkowo przed główną częścią modelu poprzedzono warstwami z augmentacją danych. Górne warstwy zostały zbudowane od nowa i zawierały kolejno Flatten dwie warstwy Dense z aktywacją ReLU o rozmiarze 4096 oraz (2048 w przypadku VGG16 lub 1024 w przypadku VGG19)<sup>1</sup> neuronów, następnie Dropout o wartości 0.5 dalej znalazła się warstwa Normalizacji, a na koniec warstwa Dense z jednym neuronem z sigmoidalną funkcją aktywacji. Łączna liczba parametrów dla sieci VGG16 wyniosła 125 876 033 oraz dla VGG19 126 987 329.

Wybór najlepszego modelu nastąpił w 289 epoce uczącej na podstawie uzyskania najniższego błędu na danych walidacyjnych wynoszącego 0.0168. Pozostałe parametry uzyskane w epoce 289 to dla danych uczących: dokładność 0.9982, błąd 0.0063, natomiast na

 $<sup>^1 {\</sup>rm Sieć}$  VGG19 otrzymała o połowę mniej neuronów w przedostatniej warstwie Dense ze względu na przekroczenie limitu pamięci karty graficznej.

danych walidacyjnych dokładność 0.9943. Uzyskany wynik na danych testowych wyniósł 0.5020 a błąd 6,0449.



Rysunek 4.18: Przebieg uczenia modelu VGG16 dla zbioru uczącego BreakHis + GZG

W przypadku uczenia modelu VGG16 na danych uczących ze zbioru SzUZG najniższy błąd dla danych walidacyjnych wyniósł 0.1035 w 144 epoce uczącej. Dokładność modelu dla danych walidacyjnych osiągnęła poziom 0.9623, natomiast dla danych uczących błąd wyniósł 0.0618 przy dokładności 0.9746. Dane testowe wskazują na błąd o wartości 1,6542 oraz dokładność wynoszącą 0.7359.



Rysunek 4.19: Przebieg uczenia modelu VGG16 dla danych uczących ze zbioru SzUZG

Model VGG19 najniższą wartość błędu na danych walidacyjnych zbioru BreakHis + GZG osiągnął w 252 epoce uczącej i wyniósł 0.0241 przy dokładności oszacowanej na 0.9918. Dane uczące z kolei scharakteryzowane zostały przez błąd o wrtości 0.0112 przy dokładności 0.9968. Niestety dane testowe znów ujawniły niskie wyniki dokładności 0.5032 podczas gdy błąd oszacowano na 6,4768.

Dla danych uczących ze zbioru SzUZG, najlepszy model został odnotowany w 168 epoce uczącej na podstawie błędu danych walidacyjnych wynoszącego 0.0874 przy dokład-



Rysunek 4.20: Przebieg uczenia modelu VGG19 dla zbioru uczącego BreakHis + GZG

ności na poziomie 0.9682. Błąd dla danych uczących osiągnął wartość 0.0451, natomiast dokładność 0.9831. Dane testowe wykazały błąd o wartości 1,8078 a dokładność 0.7579.



Rysunek 4.21: Przebieg uczenia modelu VGG19 dla danych uczących ze zbioru SzUZG

**DenseNet121** Sieć została pozbawiona jakichkolwiek wstępnie przetrenowanych wag w celu uczenia od podstaw na przygotowanych danych. Usunięto również warstwy górne w celu utworzenia nowego układu warstw składającego się z warstwy Flatten, następnie gęstej warstwy (Dense) zawierającej 512 neuronów z funkcją aktywacji ReLU. Kolejna warstwa to Dropout na poziomie 0.5 oraz warstwa Batch Normalization. Na końcu wyniki generuje warstwa Dense z jednym neuronem z sigmoidalną funkcją aktywacji. Łączna liczba parametrów przedstawionej sieci wyniosła 32 656 017 parametrów, dzięki czemu możliwe było zastosowanie stosunkowo dużej wielkości wejściowej paczki obrazów (batch -64). Ponadto model został poprzedzony warstwami wykonującymi augmentację z losowym obrotem obrazów oraz przerzucaniem wertykalnym i horyzontalnym.

Na uczenie modelu przeznaczono 300 epok uczących. Najniższy wynik funkcji straty dla

danych walidacyjnych ze zbioru BreakHis + GZG wystąpił w 271 epoce uczącej i wyniósł 0.0271 przy dokładności wynoszącej 0.9894. W tej samej epoce funkcja straty na danych uczących osiągnęła wynik 0.0020. A dokładność 0.9999, niestety na danych testowych wynik funkcji straty to 4,7317 oraz dokładność wynosząca 0.5123.



Rysunek 4.22: Przebieg uczenia modelu DenseNet121 dla danych uczących ze zbioru BreakHis + GZG

Uczenie na danych ze zbioru SzUZG trwał 300 epok, jednak najlepszy wynik na danych walidacyjnych osiągnięty został w 120 epoce uczącej. Wartość funkcji straty dla danych walidacyjnych osiągnęła najniższy poziom wynoszący 0.1513 oraz dokładność 0.9394. W tej samej epoce równolegle dla danych walidacyjnych osiągnięty został wynik 0.0200 funkcji straty oraz 0.9940 dokładności. Na danych testowych z kolei osiągnięty został wynik funkcji straty wynoszący 1,6325 oraz dokładność wynosząca 0.6601.



(a) Dokładność modelu w kolejnych epokach
 (b) Strata modelu w kolejnych epokach
 Rysunek 4.23: Przebieg uczenia modelu DenseNet121 dla danych uczących ze zbioru SzUZG

Zaobserwowany na rys. 4.22 proces uczenia sieci na zbiorze BreakHis + GZG ma bardzo łagodny przebieg, co oznacza iż wartość współczynnika uczenia (*ang. learning rate*) jest dobrze dopasowana. Nieco gorzej wygląda krzywa uczenia na rys. 4.23 niemniej dzięki zastosowaniu tej samej wartości współczynnika uczenia, można wyraźnie zaobserwować moment przeuczenia sieci i charakterystyczny trend na krzywej walidacji po 120 epoce uczącej.

**Xception** W przygotowanej bazowej sieci nie wykorzystano żadnych wstępnie przetrenowanych wag, usunięto również warstwy górne (top). Zastąpiono je sekwencją następujących elementów: warstwa Flatten, następnie warstwa Dense z 512 neuronami z funkcją aktywacji ReLU, warstwa Dropuot (0.5), warstwa normalizacyjna Batch Normalization i na koniec warstwa Dense zawierająca jeden neuron z sigmoidalną funkcją aktywacji. Dodatkowo przeprowadzono augmentację danych w postaci losowych przekształceń horyzontalnych i wertykalnych oraz rotacji. Sieć również poprzedzona jest normalizacją wartości z (0..255) do (-1..1). Łącza liczba parametrów zbudowanego modelu osiągnęła wartość 72 189 225, z czego 55 552 nie podlegało uczeniu.

Najlepszy wynik dla danych walidacyjnych z zestawu BreakHis + GZG uzyskała w 283 epoce uczącej. Błąd w tej epoce osiągnął wartość 0.0198 a dokładność 0.9918, natomiast dla danyh uczących błąd w 283 epoce wyniósł 0.0056 oraz dokładność wynosząca 0.9981. Dla danych testowych wynik błędu to 3,7645 przy dokłdności 0.5158



(a) Dokładność modelu w kolejnych epokach

(b) Strata modelu w kolejnych epokach

Rysunek 4.24: Przebieg uczenia modelu Xception dla zbioru uczącego BreakHis + GZG

Na danych ze zbioru SzUZG najmniejszy błąd na zbiorze walidacyjnym wystąpił w 227 epoce uczącej i wyniósł 0.0869, natomiast dokładność osiągnęła rezultat 0.9707. Zbiór uczący w tej samej epoce uzyskał odpowiednio błąd na poziomie 0.0195 oraz dokładność 0.9945. Dane testowe z kolei wykazały błąd o wartości 1,8638 a dokładność klasyfikatora na 0.6888.

**InceptionResNetV2** Sieć InceptionResNetV2 [98] należy do grupy sieci rezydualnych. W tym jednak przypadku cechy sieci rezydualnych łączone są z poprawioną wersją sieci





Rysunek 4.25: Przebieg uczenia modelu Xception dla danych uczących ze zbioru SzUZG

Inception [100]. Powodem dla którego autorzy modelu podjęli eksperymenty było sprawdzenie czy zachowana zostanie wydajność sieci głębokiej podczas zastąpienia etapów konkatenacji filtrów za pomocą połączeń resztowych [98]. Zastosowanie połączeń resztowych wynika z kolei z faktu bardzo dużej głębokości sieci typu Inception, a co za tym idzie zwiększeniem szansy na zanikanie gradientów podczas uczenia.

Sieć została przygotowana do uczenia od podstaw czyli wagi otrzymały losową inicjalizację. Górne warstwy zostały usunięte i zastąpione zestawem warstw począwszy od warstwy Flatten, następnie warstwy gęstej (Dense) z 512 neuronami z funkcja aktywacji ReLU. Po tej warstwie umieszczona została warstwa Dropout o wartości 0.5. Przedostatnia warstwa to Batch Normalization po której następuje warstwa z jednym neuronem z aktywacją sigmoidalną. Ten model został również poprzedzony warstwami dokonującymi augmentacji danych z losowym obrotem obrazów oraz przerzucaniem wertykalnym i horyzontalnym. Łączna liczba parametrów modelu wyniosła 74 000 609 w tym 61 568 parametrów nie podlegających uczeniu.

Ze względu na bardzo dużą liczbę warstw sieci cały proces uczenia został przedłużony do 500 epok. Najlepszy wynik na danych walidacyjnych dla zbioru BreakHis + GZG nastąpił w 458 epoce uczącej. Zaobserwowana wartość funkcji straty na danych uczących wyniosła 0.0948 oraz dokładność 0.9665. W tej samej epoce uczącej na zbiorze uczącym uzyskano o wynik dla funkcji straty o wartości 0.3528 oraz dokładność wynoszącą 0.7758. Co ciekawe po raz pierwszy dane uczące uzyskały gorsze wyniki niż dane walidacyjne. Pomimo tej anomalii dane testowe nie odbiegają od wyników uzyskiwanych w innych modelach sieci a jest to: wartość funkcji straty 3,5477 oraz dokładność 0.5158.

Na zbiorze uczącym z obrazów ze zbioru SzUZG najlepszy wynik został uzyskany już w 184 epoce uczącej. Wartość straty dla danych walidacyjnych wyniosła w tej epoce 0.2522 przy dokładności wynoszącej 0.8932. W tej samej epoce uczącej wartość straty na danych uczących wyniosła 0.4552 oraz dokładność 0.7186. Również w tym przypadku



Rysunek 4.26: Przebieg uczenia modelu InceptionResNetV2 dla danych uczących ze zbioru BreakHis + GZG

sieć InceptionResNetV2 uzyskała lepsze wyniki na zbiorze walidacyjnym niż na uczącym. Paradoksalnie uzyskany wynik dokładności (0.7031) na danych testowych jest zbliżony do wyniku uzyskanego na danych uczących. Wartość funkcji straty na danych testowych osiągnęła rezultat 0.8288.



(a) Dominances modeling in constant (b) Diffusion modeling in constant (b) Diffusion modeling in constant (b)

 $Rysunek \ 4.27: Przebieg \ uczenia \ modelu \ Inception Res Net V2 \ dla \ danych \ uczącego \ ze \ zbioru \ SzUZG$ 

Zarówno na rysunku 4.26 jak i na 4.27 wyniki na danych walidacyjnych osiągają lepsze rezultaty niż na danych uczących. Ponadto dla danych na rysunku 4.27 przeuczenie występuje bardzo szybko natomiast na rysunku 4.26 wydaje się, że wyraźnego efektu przeuczenia nie ma, ale krzywe dla danych uczących i walidacyjnych są znacznie odseparowane od siebie. Ogólnie rzecz biorąc przebieg uczenia na sieci InceptionResNetV2 różni się od krzywych zaobserwowanych na innych modelach sieci.

Podsumowanie wyników Przetestowane modele głębokich sieci neuronowych dotychczas z powodzeniem stosowane były w zadaniach związanych z klasyfikacją wieloklasową (np.: 1000 klas) oraz dużych zbiorów danych. W zaprezentowanym eksperymencie mamy natomiast do czynienia z klasyfikacją danych dychotomicznych na stosunkowo niewielkim zestawie danych. O ile sama liczba obrazów oscyluje w granicach 10 tys. - 12 tys. w zależności od zbioru, to należy podkreślić, iż obrazy te pochodzą od kilkudziesięciu pacjentów ze zbioru SzUZG oraz kilkudziesięciu ze zbioru BreakHis. Kolejne problemy dotyczą jakości danych, sprzętu użytego do skanowania, a także procedur pobierania materiału. Co ciekawe wszystkie modele za wyjątkiem InceptionResNetV2 osiągnęły wysokie rezultaty na danych uczących oraz walidacyjnych. Oznacza to, że dopasowanie modeli do danych jest równie ekstremalne, a to z kolei przekłada się na niezadowalające wyniki na danych testowych. Przyczyn można się doszukiwać liczbie dostępnych danych medycznych, a co za tym idzie uboga różnorodnością przypadków pomimo zastosowania augmentacji. Paradoksalnie problemem może również być nadmiar informacji wejściowych w postaci obrazów trójkanałowych o rozmiarze 230x230 pikseli. W kolejnych eksperymentach zostanie sprawdzony efekt redukcji wymiarów obrazów wejściowych na elastyczność wybranych modeli głębokich sieci neuronowych.

Dane uczące BreakHis + GZG; Dane testowe SzUZG							
Madalata CNN	Uczenie		Walida	icja	Test		
	ACC	LOSS	ACC	LOSS	ACC	LOSS	
Resnet50	0.9944	0.0175	0.9914	0.0264	0.5242	3.2591	
ResNet151V2	0.9916	0.0241	0.9898	0.0275	0.5151	3.6754	
VGG16	0.9982	0.0063	0.9943	0.0168	0.5020	6.0449	
VGG19	0.9968	0.0112	0.9918	0.0241	0.5032	6.4768	
DenseNet121	0.9999	0.0020	0.9894	0.0271	0.5123	4.7317	
Xception	0.9981	0.0056	0.9918	0.0198	0.5158	3.7645	
InceptionResNetV2	0.7758	0.3528	0.9665	0.0948	0.5158	3.5477	

Tabela 4.3: Podsumowanie wyników klasyfikacji obrazów w przestrzeni RGB

L L	,			•		
Model size CNN	Uczenie		Walida	icja	Test	
	ACC	LOSS	ACC	LOSS	ACC	LOSS
Resnet50	0.9767	0.0612	0.9568	0.1123	0.6003	1.8699
ResNet151V2	0.9881	0.0351	0.9657	0.1034	0.6562	1.6730
VGG16	0.9746	0.0618	0.9623	0.1035	0.7359	1.6542
VGG19	0.9831	0.0451	0.9682	0.0874	0.7579	1.8078
DenseNet121	0.9940	0.0200	0.9394	0.1513	0.6601	1.6325
Xception	0.9945	0.0195	0.9707	0.0869	0.6888	1.8638
InceptionResNetV2	0.7186	0.4252	0.8932	0.2522	0.7031	0.8288

Dane pochodzące ze zbioru BreakHis + GZG testowane były na zbiorze SzUZG. Uzy-

skane wyniki na zbiorze testowym osiągnęły wartości od 0.5020 do 0.5242 (tab. 4.3). Wyniki te oznaczają, iż uzyskane modele nie mają praktycznie żadnych zdolności uogólniających pomiędzy danymi z różnych ośrodków medycznych. Uzyskane modele dopasowały się jedynie do danych uczących. Nieco lepiej wyglądają wyniki dla modeli uczonych na zbiorze SzUZG. Te z kolei były testowane na zbiorze BreakHis + GZG. Wyniki oscylowały w zakresie wartości od 0.6003 do 0.7579. Najniższą wartość uzyskał model ResNet50 natomiast najwyższą VGG19. Wartość uzyskana przez model VGG19 można uznać za dobry wynik pokazujący zdolności uogólniające modelu. Ogólnie można podsumować, że ten eksperyment dowiódł skłonność metod głębokiego uczenia do nadmiernego dopasowywania się do danych uczących bazujących na obrazach w przestrzeni RGB.

## Obrazy po normalizacja przy użyciu rozplotu H&E

**Podsumowanie wyników** Modyfikacją powyższego eksperymentu była redukcja wymiarowości obrazów w przestrzeni trójkanałowej do jednokanałowej. Procedura ta została wykonana przy użyciu rozpłotu obrazu. Rozpłot został wykonany w oprogramowaniu Fiji [83] przy użyciu macierzy rozpłotu (H&E). Procedura rozpłotu została opisana w rozdziale 2.3. Do eksperymentu wykorzystano obrazy, na których rozpłot wyodrębnił miejsca ekspozycji hematoksyliny. Obiekty w których odkłada się hematoksylina to przede wszystkim jądra komórkowe, czyli kluczowe diagnostycznie elementy obrazów histopatologicznych. Do eksperymentu wybrano większość modeli wykorzystanych w poprzednim badaniu dla obrazów w przestrzeni trójkanałowej.

Dane uczące BreakHis + GZG; Dane testowe SzUZG							
Model siegi CNN	Uczenie		Walidacja		Test		
Wodel sleer Civit	ACC	LOSS	ACC	LOSS	ACC	LOSS	
Resnet50	0.9149	0.1989	0.8654	0.3273	0.5306	2.2535	
VGG19	0.9197	0.1845	0.8662	0.2973	0.5183	2.5664	
DenseNet121	0.9778	0.0701	0.8658	0.3408	0.5257	2.8245	
Xception	0.7587	0.4709	0.7071	0.5179	0.4848	17.9988	
InceptionResNetV2	0.8709	0.2887	0.8490	0.3289	0.5385	1.8650	
Dane uczące SzUZ	G; Dane	e testow	e Break	His + G	ZG		
Dane uczące SzUZ	G; Dane Uczeni	e testow e	e Break Walida	His + G Icja	ZG Test		
Dane uczące SzUZ Model sieci CNN	G; Dane Uczeni ACC	e testow e LOSS	e Break Walida ACC	His + G Icja LOSS	ZG Test ACC	LOSS	
Dane uczące SzUZ Model sieci CNN Resnet50	G; Dane Uczeni ACC 0.9106	e testowe e LOSS 0.2156	e Break Walida ACC 0.8868	$\frac{\text{His} + \mathbf{G}}{\text{LOSS}}$ $\frac{0.2644}{0.2644}$	<b>ZG</b> <b>Test</b> ACC 0.6617	LOSS 1.0817	
Dane uczące SzUZ Model sieci CNN Resnet50 VGG19	G; Dane Uczeni ACC 0.9106 0.9135	e testow e LOSS 0.2156 0.1995	e Break Walida ACC 0.8868 0.9086	His + G Icja LOSS 0.2644 0.2270	ZG Test ACC 0.6617 0.7019	LOSS 1.0817 1.3525	
Dane uczące SzUZ Model sieci CNN Resnet50 VGG19 DenseNet121	G; Dane Uczeni ACC 0.9106 0.9135 0.9419	e testowe e LOSS 0.2156 0.1995 0.1494	e Break Walida ACC 0.8868 0.9086 0.9076	His + G Icja LOSS 0.2644 0.2270 0.2441	ZG Test ACC 0.6617 0.7019 0.6627	LOSS 1.0817 1.3525 1.1087	
Dane uczące SzUZ Model sieci CNN Resnet50 VGG19 DenseNet121 Xception	G; Dano Uczeni ACC 0.9106 0.9135 0.9419 0.8107	e testow e LOSS 0.2156 0.1995 0.1494 0.3942	e Break Walida ACC 0.8868 0.9086 0.9076 0.7974	His + G Icja LOSS 0.2644 0.2270 0.2441 0.4246	ZG Test ACC 0.6617 0.7019 0.6627 0.4474	LOSS 1.0817 1.3525 1.1087 11.0942	

Tabela 4.4: Podsumowanie wyników klasyfikacji obrazów po rozplocie

Uzyskane wyniki (tab. 4.4) nie wykazują poprawy względem modeli zbudowanych na obrazach trójkanałowych, a w przypadku sieci Xception wręcz pogarszają wyraźnie rezultaty. Podsumowując eksperyment można stwierdzić, że redukcja wymiarowości obrazu wejściowego do jednokanałowych obrazów po rozplocie nie przyczynia się do poprawy zdolności uogólniających sieci CNN. Uzyskane wyniki są zbliżone do tych, które uzyskano na obrazach w przestrzeni RGB.

## Obrazy po normalizacji przy użyciu segmentacji

**Podsumowanie wyników** Ostatni eksperyment polegał na klasyfikacji obrazów wejściowych do sieci znormalizowanych przy użyciu zaproponowanej hybrydowej metody segmentacji. Obrazy te charakteryzują się posiadaniem jednego kanału oraz binarną reprezentacją pikseli w tym kanale. Podstawą do podjęcia eksperymentu było założenie, iż normalizacja do przestrzeni binarnej obrazów, sprowadzi zestawy uczące do przejrzystej i prostej formy oraz jedynie kształt obiektów będzie wówczas wpływał na ostateczne uczenie modelu.

Dane uczące BreakHis + GZG; Dane testowe SzUZG								
Model size: CNN	Uczenie		Walida	acja	Test			
Model sleci Cinin	Acc	Loss	Acc	Loss	Acc	Loss		
Resnet50	0.7220	0.5393	0.7095	0.5382	0.7813	0.4539		
VGG19	0.7283	0.5129	0.7536	0.4774	0.7897	0.4596		
DenseNet121	0.7373	0.5202	0.7364	0.4991	0.7695	0.4977		
Xception	0.7191	0.5579	0.7254	0.5269	0.7811	0.4703		
InceptionResNetV2	0.7666	0.5042	0.7417	0.5077	0.7806	0.4691		

Tabela 4.5: Podsumowanie wyników klasyfikacji obrazów po segmentacji

Dane uczące SzUZG; Dane testowe BreakHis + GZG	

Madal sizei CNN	Uczenie		Walida	ıcja	Test		
Model Slect Civin	Acc	Loss	Acc	Loss	Acc	Loss	
Resnet50	0.8022	0.4711	0.7939	0.4788	0.7004	0.5539	
VGG19	0.7761	0.4786	0.8287	0.4244	0.7239	0.5275	
DenseNet121	0.8267	0.4049	0.8093	0.4378	0.7121	0.5605	
Xception	0.7724	0.4746	0.8183	0.4173	0.7122	0.5242	
InceptionResNetV2	0.7383	0.5159	0.8019	0.4420	0.7049	0.5813	

Uzyskane na danych testowych wyniki (tab. 4.5) jednoznacznie pokazują, iż dane na obrazach binarnych zbliżyły się do siebie na tyle, że możliwe stało się przeprowadzenie skutecznej klasyfikacji. Na zbiorze uczącym BreakHis + GZG uzyskany poziom dokładności jest niższy niz na zbiorze testowym, jak również większości walidacyjnych. Może to oznaczać, iż faktycznie na poziomie danych binarnych zdolności uogólniające modelu uległy znacznej poprawie w stosunku to obrazów z większą ilością odcieni. Gdyby uśrednić

wszystkie wyniki dokładności ze zbioru uczącego, walidacyjnego i testowego z obu zestawów danych to najwyższy wynik uzyskał model VGG19, jednakże różnice te są bardzo małe. Kluczowym wskaźnikiem będzie osiągnięty wynik dla danych testowych i w tym przypadku również nieznacznie lepszy od pozostałych okazała się model VGG19. Różnice te nie są jednak znaczne i można stwierdzić że wszystkie modele z podobną dokładnością radzą sobie z tymi danymi.

## 4.8.3 Opracowany system klasyfikacji

## 4.8.3.1 Plan eksperymentów

Druga część eksperymentów dotyczy zbadania wpływu poszczególnych klasyfikatorów nadrzędnych w opracowanym systemie oraz wykazania czy zaproponowana metoda fuzji cech manualnych i głębokich poprawi wyniki uzyskane dla modeli zbudowanych na oddzielnych zestawach cech manualnych oraz cech głębokich. W tej części zbadane zostanie również zagadnienie dotyczące poprawy wyników klasyfikacji przy użyciu zespoły generatorów cech głębokich względem uzyskanych wyników indywidualnych dla badanych sieci CNN. W badaniu zastosowano 8 różnych modeli klasyfikacji, które były badane w różnych konfiguracjach. Wyniki zawierają jedynie wybrane, najlepsze konfiguracje. Schemat przeprowadzonych eksperymentów z udziałem klasyfikatorów nadrzędnych na bazie cech manualnych, głębokich oraz fuzji cech przedstawiony został na rysunku 4.28. Kolejność prezentowania kolejnych eksperymentów opiera się na paragrafach nazwanych na podstawie badanego klasyfikatora. Pod każdym paragrafem umieszczono komentarz oraz 2 tabele z wynikami. Jedna z nich dotyczy modeli zbudowanych na danych ze zbioru BreakHis + GZG, które testowano na zbiorze SzUZG. Druga tabela dotyczy modeli zbudowanych na danych ze zbioru SzUZG, które były testowane na danych BreakHis + GZG.

#### 4.8.3.2 Wprowadzenie

Zbiory obrazów wykorzystane do wykonania eksperymentów posłużyły do wyekstrahowania dwóch grup cech: manualnych oraz głębokich. Cechy manualne zostały pozyskane na podstawie kształtów, kolorów czy też tekstur jąder komórkowych wykrytych na obrazach przy użyciu sieci segmentacyjnej. Cechy głębokie stanowią wartości wyjść pojedynczych neuronów występujących w przedostatniej warstwie głębokich sieci neuronowych. W obu przypadkach wyodrębnione cechy charakteryzują się jedno-wymiarowością i dzięki temu mogą być łączone ze sobą we wspólny zestaw cech. Własność ta zostanie wykorzystana w celu zbadania czy łączenie cech manualnych i głębokich w jeden zestaw poprawia wyniki klasyfikacji. Eksperymenty klasyfikacyjne z bieżącego rozdziału zostały zrealizowane przy użyciu oprogramowanie JMP Pro [82].

Przyjęto za klasę pozytywną (TP) przypadki złośliwe oznaczone wartością 1, natomiast

Zestaw cech manualnych	Zestaw cech głębokich	Fuzja cech głębokich oraz manualnych		
Regresja	Regresja	Regresja		
I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG		
Dane testowe SzUZG	Dane testowe SzUZG	Dane testowe SzUZG		
II) Dane uczące SzUZG	II) Dane uczące SzUZG	II) Dane uczące SzUZG		
Dane testowe BreakHis+GZG	Dane testowe BreakHis+GZG	Dane testowe BreakHis+GZG		
Drzewo decyzyjne	Drzewo decyzyjne	Drzewo decyzyjne		
I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG		
Dane testowe SzUZG	Dane testowe SzUZG	Dane testowe SzUZG		
II) Dane uczące SzUZG	II) Dane uczące SzUZG	II) Dane uczące SzUZG		
Dane testowe BreakHis+GZG	Dane testowe BreakHis+GZG	Dane testowe BreakHis+GZG		
Las losowe	Las losowe	Las losowe		
I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG		
Dane testowe SzUZG	Dane testowe SzUZG	Dane testowe SzUZG		
II) Dane uczące SzUZG	II) Dane uczące SzUZG	II) Dane uczące SzUZG		
Dane testowe BreakHis+GZG     Wzmacnianie gradientowe     Dane uczące BreakHis+GZG     Dane testowe SzUZG     Dane uczące SzUZG	Dane testowe BreakHis+GZG Wzmacnianie gradientowe I) Dane uczące BreakHis+GZG Dane testowe SzUZG II) Dane uczące SzUZG	Dane testowe BreakHis+GZG     Wzmacnianie gradientowe     Dane uczące BreakHis+GZG     Dane testowe SzUZG     Dane uczące SzUZG		
<ul> <li>Dane testowe BreakHis+GZG</li> <li>Klasyfikator naiwny Bayesa</li> <li>Dane uczące BreakHis+GZG</li> <li>Dane testowe SzUZG</li> <li>Dane testowe BreakHis+GZG</li> </ul>	<ul> <li>Dane testowe BreakHis+GZG</li> <li>Klasyfikator naiwny Bayesa</li> <li>Dane uczące BreakHis+GZG</li> <li>Dane testowe SzUZG</li> <li>Dane testowe BreakHis+GZG</li> </ul>	<ul> <li>Dane testowe BreakHis+GZG</li> <li>Klasyfikator naiwny Bayesa</li> <li>Dane uczące BreakHis+GZG</li> <li>Dane testowe SzUZG</li> <li>Dane testowe BreakHis+GZG</li> </ul>		
Klasyfikator k-NN I) Dane uczące BreakHis+GZG Dane testowe SzUZG II) Dane uczące SzUZG Dane testowe BreakHis+GZG	Klasyfikator k-NN       I)     Dane uczące BreakHis+GZG Dane testowe SzUZG       II)     Dane uczące SzUZG Dane testowe BreakHis+GZG	Klasyfikator k-NN I) Dane uczące BreakHis+GZG Dane testowe SzUZG II) Dane uczące SzUZG Dane testowe BreakHis+GZG		
Maszyna wektorów nośnych	Maszyna wektorów nośnych	Maszyna wektorów nośnych		
I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG		
Dane testowe SzUZG	Dane testowe SzUZG	Dane testowe SzUZG		
II) Dane uczące SzUZG	II) Dane uczące SzUZG	II) Dane uczące SzUZG		
Dane testowe BreakHis+GZG	Dane testowe BreakHis+GZG	Dane testowe BreakHis+GZG		
Klasyfikator neuronowy	Klasyfikator neuronowy	Klasyfikator neuronowy		
I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG	I) Dane uczące BreakHis+GZG		
Dane testowe SzUZG	Dane testowe SzUZG	Dane testowe SzUZG		
II) Dane uczące SzUZG	II) Dane uczące SzUZG	II) Dane uczące SzUZG		
Dane testowe BreakHis+GZG	Dane testowe BreakHis+GZG	Dane testowe BreakHis+GZG		

Rysunek 4.28: Schemat eksperymentów na klasyfikatorach nadrzędnych

za klasę negatywną (TN) oznaczone wartością 0 przypadki łagodne.

## 4.8.3.3 Imputacji brakujących danych

Na etapie ekstrakcji cech oraz obliczania statystyk, część danych nie została obliczona ze względu na problemy numeryczne. W przypadku kilku cech liczba brakujących danych nie przekraczała 2. Jedna cecha, a dokładnie entropia korelacji macierzy GLCM, wykazała się znaczną liczbą brakujących danych a dokładnie 7154 spośród wszystkich 22326 wierszy pozyskanych ze Zbioru BreakHis + GZG oraz zbioru SzUZG. Liczba ta niestety mocno rzutuje na wiarygodność uzyskanej cechy po imputacji danych. Zastosowana metoda imputacji bazowała na zastępowaniu brakujących danych przy użyciu metody najmniejszych kwadratów w oparciu o istniejące dane oraz model rozkładu normalnego.

### 4.8.3.4 Eksperymenty

**Regresja** Regresja oferuje szeroki wachlarz dostępnych metod, redukcji wymiarowości, regularyzacji oraz walidacji. W eksperymentach zbadano różne metody walidacji: Bayesowskie kryterium informacyjne, kryterium informacyjne Akaikego, walidację krzyżową (KFold, Holdout). Przeprowadzone zostały również testy modeli regresji logistycznej oraz regularyzowanych modeli regresji w tym: regresji grzbietowej, regresjo Lasso oraz regresji metodą elastycznej siatki. Zbadano również wyniki dla modelu regresji logistycznej z krokową eliminacją cech. Zazwyczaj przyjmuje się, że walidacja krzyżowa jest lepszą metodą walidacji niż Bayesowskie kryterium informacyjne. Niemniej zaletą tego kryterium informacyjnego jest skuteczniejsza eliminacja cech i ostateczna redukcja wymiarowości modelu. Różnice pomiędzy badanymi modelami regresji były niewielkie, stąd w podsumowaniu znajdą się wyniki najlepszych dwóch metod. W tej sekcji umieszczono jedynie pełne wyniki dla modelu regresji z regularyzacją siatki elastycznej z walidacjcją przy użyciu Bayesowskiego kryterium informacyjnego, gdyż metoda regresji krokowej nie była w stanie dokończyć zadania selekcji cech spowodowanej złożonością obliczeniową dla zestawu cech manualnych.

Tabela 4.6: Regresja z regularyzacją Elastycznej siatki oraz walidacją przy użyciu kryterium informacyjnym Bayesa, uczona na zbiorze BreakHis + GZG, natomiast testowana na zbiorze SzUZG

	Manualne		Głębokie			Fuzja			
	TPR	TNR	ACC	TPR	TNR	ACC	$\mathrm{TPR}$	TNR	ACC
Uczenie	0.8100	0.8636	0.8396	0.8612	0.7293	0.7883	0.8627	0.8574	0.8597
Walidacja	-	-	-	-	-	-	-	-	-
Test	0.4594	0.9198	0.6967	0.7030	0.8620	0.7849	0.6606	0.8961	0.7819

W pierwszej kolejności obliczono model na podstawie cech manualnych ze zbioru BreakHis + GZG. Z łącznej liczby 251 cech do modelu po redukcji trafiło 173 cechy plus wyraz wolny czyli zbiór uległ redukcji o około 30% stanu pierwotnego. Pole AUC osiągnęło wartość 0.9177 dla danych uczących a wartość uogólnionego  $R^2$  0.6358. Budowanie modelu na bazie cech głębokich zakończyło się dodaniem wyrazu wolnego oraz 14 spośród 25 cech, czyli równe 60% pierwotnej liczby. Uzyskana wartość uogólnionego współczynnika determinacji wyniosła 0.5567, natomiast pole AUC zostało oszacowane na 0.8762. Największa jednak redukcja wymiarowości dotyczyła modelu opartego na fuzji cech manualnych oraz głębokich. Model potrzebował 96 cech plus wyraz wolny czyli niecałe 40% wszystkich cech uzyskując wartość uogólnionego współczynnika  $R^2$  na poziomie 0.6939, natomiast pole AUC 0.9360.

Podsumowanie wyników (tab. 4.6) dla klasyfikatora trenowanego na danych BreakHis + GZG wskazuje, że dla danych testowych najgorszy wynik klasyfikacji uzyskał klasyfikator zbudowany z zestawie cech manualnych, natomiast najlepszy na zestawie cech głębokich. Klasyfikator uzyskany na fuzji cech uzyskał wynik niemal identyczny jak zestaw

#### cech głębokich.

Tabela 4.7: Regresja z regularyzacją Elastycznej siatki oraz walidacją przy użyciu kryterium informacyjnym Bayesa, uczona na zbiorze SzUZG, natomiast testowana na zbiorze BreakHis + GZG

	Manualne		Głębokie			Fuzja			
	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR	ACC
Uczenie	0.8365	0.8751	0.8564	0.8568	0.8601	0.8585	0.8779	0.8849	0.8815
Walidacja	-	-	-	-	-	-	-	-	-
Test	0.6418	0.8733	0.7697	0.7968	0.6400	0.7102	0.7124	0.8333	0.7792

Zestaw cech manualnych ze zbioru SzUZG w wyniku przeprowadzonej regularyzacji został zredukowany do 125 cech oraz wyrazu wolnego. Uzyskany wynik uogólnionego współczynnika determinacji wyniósł 0.6862. Wartość AUC dla klasyfikatora zbudowanego na cechach manualnych wyniosła 0.9328. Zestaw cech głębokich w wyniku regularyzacji nie uległ redukcji, czyli wszystkie cechy okazały się być istotne w modelu. Wartość uogólnionego uogólnionego współczynnika  $R^2$  wyniosła 0.7020, natomiast współczynnika AUC 0.9371. Fuzja cech z kolei charakteryzowała się modelem zbudowanym na bazie 70 cech oraz wyrazu wolnego, czyli regularyzacja przyczyniła się do znacznej redukcji liczby cech. Jednocześnie wartość uogólnionego współczynnika determinacji osiągnęła wynik 0.7556, natomiast pole AUC wartość 0.9545.

W przypadku klasyfikatorów zbudowanych na zbiorze obrazów uczących SzUZG najgorzej wypadł model zbudowany na zestawie cech głębokich, najlepiej natomiast model bazujący na fuzji cech, nieznacznie poprawiając model zbudowany na zestawie cech manualnych.

Biorąc pod uwagę wartości średnie dokładności (tab. 4.7) uzyskanych na podstawie obu zbiorów danych (BreakHis + GZG oraz SzUZG) najlepszy wynik uzyskały modele bazujące na fuzji cech. Wartości średnie to 0.7805 dla fuzji, 0.7475 dla cech głębokich oraz 0.7332 dla manualnych. Fuzja cech poprawia zatem dokładność modelu na danych testowych średnio przynajmniej o 3%. Biorąc jednak pod uwagę minimalne wartości uzyskane dla modeli: fuzja (0.7792), głębokie (0.7102) oraz manualne (0.6967), można zauważyć jak bardzo uległy poprawie właściwości uogólniające modelu opartego na fuzji cech względem pozostałych.

**Drzewo decyzyjne** Zbiory użyte do uczenia oraz walidacji modelu zostały podzielone w stosunku 0.8: 0.2. Oznacza to, że przygotowane zbiory walidacyjne osiągają ilość około 2000 obserwacji. Wybór najlepszego podziału polega na odnalezieniu maksymalnej wartości współczynnika determinacji  $R^2$ , natomiast kryterium zatrzymania następuje po 10 kolejnych podziałach, jeśli nie wystąpi w nich poprawa rezultatu. Zaletą modelu drzewa decyzyjnego (DT) jest przejrzysta struktura modelu, natomiast wadą zawężanie zbioru obserwacji w kolejnych podziałach. Wyniki z przeprowadzonych eksperymentów zostały zaprezentowane w tabelach: 4.8 oraz 4.9. W przypadku zbioru BreakHis+GZG podział

	Manua	Manualne		Głębol	Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	$\mathrm{TPR}$	TNR	ACC	
Uczenie	0.8718	0.8955	0.8850	0.9166	0.6779	0.7849	0.9072	0.8626	0.8825	
Walidacja	0.8307	0.8777	0.8559	0.9104	0.6862	0.7857	0.8827	0.8296	0.8537	
Test	0.6565	0.7799	0.7201	0.7485	0.8156	0.7830	0.7104	0.8266	0.7702	

Tabela 4.8: DT, uczone na zbiorze BreakHis + GZG, natomiast testowane na zbiorze SzUZG

DT dla danych z cech manualnych został zakończony w 57 iteracji. Wówczas została osiągnięta najwyższa wartość uogólnionego współczynnika  $R^2$  dla zbioru walidacyjnego i wyniosła 0.6673. W tej samej iteracji zbioru uczącego wartość uogólnionego współczynnika determinacji osiągnęła wartość 0.7473. Pod wykresem krzywej ROC wyznaczona została wielkość pola (AUC) i wyniosła 0.9262 dla danych walidacyjnych oraz 0.9495 dla danych uczących. Następnie wykonano analogicznie model DT dla cech głębokich. Ze względu na znacznie mniejszą łączną liczbę cech głębokich, można było przypuszczać, iż liczba iteracji potrzebnych do wygenerowania najlepszego modelu ulegnie redukcji. Nie jest zatem zaskoczeniem, iż do uzyskania najlepszego modelu wykonano zaledwie 6 podziałów drzewa. Wyniki dla tego podziału prezentują się następująco: uogólnionego  $R^2$  dla danych walidacyjnych 0.5489, dla uczących 0.5467. Pole AUC dla danych walidacyjnych 0.8674 oraz dla danych uczących 0.8648. Ostatni test został przeprowadzony na połączonych w jeden zestaw cech manualnych oraz głębokich, zestaw ten będzie nazywany fuzją. Zestaw pomimo swojego znacznego rozmiaru, najwyższy wynik uogólnionego współczynnika determinacji dla zbioru walidacyjnego uzyskał w 49 podziale i wyniósł 0.6880. W tej samej iteracji, uogólnionego współczynnika  $R^2$  dla danych uczących osiągnał wartość 0.7607. Pola AUC uzyskały odpowiednio dla zbioru walidacyjnego oraz uczącego takie rezultaty: 0.9327 i 0.9542.

Najlepszy wynik klasyfikacji dla danych testowych uzyskał model zbudowany na zestawie cech głębokich. Najgorszy natomiast powstał na bazie cech manualnych. Model zbudowany na fuzji cech manualnych oraz głębokich okazał się nieznacznie gorszy od modelu powstałego wyłącznie na bazie cech głębokich.

Tabela 4.9: DT, uczone na zbiorze SzUZG, natomiast testowane na zbiorze BreakHis + GZG

	Manualne			Głębol	Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR	ACC	
Uczenie	0.8198	0.8441	0.8323	0.8261	0.8802	0.8541	0.8622	0.8722	0.8674	
Walidacja	0.7891	0.8215	0.8059	0.8200	0.8466	0.8334	0.8370	0.8693	0.8535	
Test	0.5461	0.7482	0.6578	0.7345	0.6797	0.7042	0.6372	0.7132	0.6792	

Najwyższa wartość (0.5579) współczynnika determinacji  $\mathbb{R}^2$  dla zbioru walidacyjnego

dla cech manualnych został osiągnięty w 24 podziale drzewa. Podział ten również charakteryzował się wartością współczynnika  $R^2$  wynoszącą 0.6351 dla zbioru uczącego. Współczynnik AUC na zbiorze uczącym osiągnął wartość 0.9123, natomiast na zbiorze walidacyjnym 0.8893. Dla zestawu cech głębokich najwyższa wartość  $R^2$  dla danych walidacyjnych wyniosła 0.6488 w 15 podziale, w którym wartość  $R^2$  dla danych uczących wyniosła 0.6998. Dla cech głębokich współczynnik AUC na zbiorze uczącym wyniósł 0.9332, natomiast na walidacyjnym 0.9165. Jako ostatnia do drzewa decyzyjnego trafiają połączone zestawy cech manualnych i głębokich. Model drzewa został wybrany w 16 podziale. Co ciekawe znów pierwsze dwa podziały zostały zrealizowane w obrębie cech głębokich, a kolejne głównie wśród manualnych. Współczynnik determinacji dla danych walidacyjnych wyniósł 0.6584, natomiast dla uczących 0.7225. Współczynnik AUC dla danych walidacyjnych osiągnął wartość 0.9233 natomiast dla uczących 0.9407.

Modele drzew decyzyjnych zbudowane na danych uczących pochodzących ze zbioru SzUZG wykazały znów najlepszy wynik dla zestawu cech głębokich (0.7042). Najgorszy wynik osiągnął model zbudowany na zestawie cech manualnych (0.6578). Fuzja cech poprawiła wyniki względem zestawu cech manualnych, jednocześnie wypadając wyraźnie gorzej od zestawu cech głębokich.

Ogólnie wyniki klasyfikacji na danych testowych wykazały, że średnio najlepszy model drzewa decyzyjnego bazował na cechach głębokich osiągając wynik 0.7436. Pozostałe średnie wyniki to 0.7247 dla fuzji cech oraz 0.6889 dla cech manualnych. Ranking minimalnych wyników również wypada podobnie. Wniosek wynikający z eksperymentu jest taki, iż dla modeli drzew decyzyjnych najlepszy okazał się zestaw cech głębokich.

Las losowy Klasyfikator LL jak sama nazwa wskazuje wymaga na samym początku zdefiniowania liczby drzew, które w tym lesie się znajdą, a właściwie zostaną użyte. Do eksperymentu wybrano domyślna liczbę drzew wynoszącą 100. Kolejny ważny parametr to liczba kolumn (cech), które zostaną rozważone do podziału w kolejnej iteracji podziałowej. Wartość ta domyślnie została ustawiona jako 3/4 łącznej liczby kolumn w zestawie danych. Można również wprowadzić współczynnik próby samowspornej<sup>2</sup>. Definiowana jest również minimalna liczba (10) oraz maksymalna (2000) liczba podziałów w pojedynczym drzewie. Dostępna jest również opcja definiowania minimalnej liczby próbek dla której może zostać wykonany podział. Liczba ta wynosi 12. Opcja wczesnego zatrzymania jest dostępna w przypadku używania zbioru walidacyjnego. Wczesne zatrzymanie oznacza zakończenie procesu uczenia gdy dodawanie kolejnych drzew nie poprawia wyniku walidacji.

Wyniki dla klasyfikatora LL charakteryzują się znacznym dopasowaniem do danych uczącego (tab. 4.10). Klasyfikator uczony na cechach manualnych ze zbioru Breakhis + GZG uzyskał dla danych uczących wartość uogólnionego współczynnika  $R^2$  wynoszącą 0.8811, a dla danych walidacyjnych 0.7846. Krzywa ROC dla danych uczących była wzor-

 $<sup>^2 \</sup>mathrm{Slowo}$  bootstrap pojawia się w literaturze pod hasłem metody samow<br/>spornej

	Manualne		Głębol	Głębokie			Fuzja		
	TPR	TNR	ACC	$\mathrm{TPR}$	TNR	ACC	$\mathrm{TPR}$	TNR	ACC
Uczenie	0.9805	0.9761	0.9780	0.9440	0.8649	0.9000	0.9787	0.9744	0.9763
Walidacja	0.8930	0.8969	0.8951	0.8565	0.7677	0.8088	0.9109	0.9000	0.9050
Test	0.6411	0.8227	0.7347	0.6962	0.8562	0.7787	0.6968	0.8601	0.7810

Tabela 4.10: LL, uczony na zbiorze BreakHis + GZG, natomiast testowany na zbiorze SzUZG

cowa, a pole AUC osiągnęło wartość 0.9985. Dla danych walidacyjnych pole AUC uzyskuje również bardzo wysoki wynik: 0.9697. Kolejne zadanie polegało na zbudowaniu modelu w oparciu o zestaw cech głębokich. Wartość uogólnionego współczynnika  $R^2$  dla danych uczących wyniosła 0.7444, natomiast dla danych walidacyjnych 0.5840. Pole AUC dla danych uczących osiągnęło rozmiar 0.9749, a dla walidacyjnych 0.8868. Ostatni model dla zbioru BreakHis + GZG został zbudowany na fuzji cech i dla danych uczących uzyskał wartość uogólnionego  $R^2$  równą 0.8892 oraz dla danych walidacyjnych 0.8041. Z kolei pole AUC dla danych uczących charakteryzuje się wielkością 0.9981, natomiast dla danych walidacyjnych 0.9731.

Model klasyfikatora zbudowany na zestawie cech manualnych uzyskał najsłabszy rezultat wynoszący 0.7347. Modele zbudowany na cechach głębokich oraz fuzji cech uzyskały podobne rezultaty w okolicach 0.7800 dokładności.

	Manualne			Głębol	Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR	ACC	
Uczenie	0.9646	0.9753	0.9701	0.9338	0.9263	0.9299	0.9720	0.9671	0.9695	
Walidacja	0.8279	0.8683	0.8488	0.8604	0.8595	0.8600	0.8804	0.8820	0.8812	
Test	0.5375	0.8104	0.6883	0.7992	0.6329	0.7073	0.7073	0.7064	0.7068	

Tabela 4.11: LL, uczony na zbiorze SzUZG, natomiast testowany na zbiorze BreakHis + GZG

Zestaw cech manualnych pozyskany ze zbioru SzUZG posłużył do zbudowania modelu, który dla danych uczących uzyskał wartość uogólnionego współczynnika determinacji na poziomie 0.8566. Na danych walidacyjnych generalizowany współczynnik determinacji osiągnął 0.6868. Pole AUC dla danych uczących uzyskało wielkość wynoszącą 0.9968 a na danych walidacyjnych 0.9336. Drugi etap polegał na zbudowaniu modelu na zestawie cech głębokich. Zestaw ten uzyskał wartość uogólnionego współczynnika  $R^2$  na poziomie 0.8395 dla danych uczących oraz 0.7082 dla danych walidacyjnych. Pole AUC pod krzywą ROC osiągnęło wartość 0.9859 dla danych uczących oraz 0.9385 dla danych walidacyjnych. Ostatni eksperyment polegał na zbudowaniu modelu na bazie fuzji cech. Generalizowany współczynnik  $R^2$  dla danych uczących osiągnęł wynik 0.8895 , natomiast dla danych walidacyjnych 0.7690. Pole AUC osiągnęło wartość 0.9970 na danych uczących oraz 0.9595 na danych walidacyjnych.

Wśród modeli zbudowanych na zbiorze SzUZG najgorszy wynik (tab. 4.11) przypadł

w udziale klasyfikatorowi zbudowanemu na zestawie cech manualnych (0.6883). Najlepszy wynik uzyskały modele zbudowane na zestawie cech głębokich (0.7073) oraz fuzji cech (0.7068), różnica pomiędzy nimi jest znikoma. Na uwagę zasługuje fakt zrównoważenia poziomów TPR i TNR w modelu zbudowanym na fuzji cech.

Średnie wyniki klasyfikacji na zbiorach testowych dla par modeli wyniosły 0.7439 dla fuzji, 0.7430 dla cech głębokich oraz 0.7115 dla cech manualnych. Minimalne wartości dla poszczególnych par modeli są zbieżne z wynikami uzyskanymi na zbiorze SzUZG. Klasyfikatory bazujące na LL pomimo bardzo wysokich wyników klasyfikacji na danych uczących oraz walidacyjnych nie potwierdziły równie wysokiej skuteczności na danych testowych. Modele te prawdopodobnie mają tendencję do nadmiernego dopasowywania się do danych uczących.

Wzmacnianie gradientowe (drzewa wzmacniane) Pierwszym parametrem jaki można ustalić przed przystąpieniem do budowania modelu DW jest ustalenie maksymalnej liczby warstw (drzew), które mogą zostać dodane do drzewa. We wszystkich eksperymentach wartość ta została ustalona na 200. Kolejny czynnik, który został ustalony to liczba podziałów w każdej warstwie. Parametr ten został ustalony na 18 podziałów. Kolejny parametr to współczynnik uczenia. Jego wartość ustala się w granicach od 0 do 1, z uwzględnieniem faktu, że im wyższa jest jego wartość tym szybciej osiągnięta zostanie zbieżność modelu, ale jednocześnie podniesione zostanie ryzyko nadmiernego dopasowania modelu [82]. W eksperymentach przyjęto wartość 0.05. Ostatni istotny do ustalenia parametr to liczba minimalna liczba obserwacji na której można wykonać podział. Została ona ustalona na 5. W celu zachowania powtarzalności wybranych zbiorów dokonano resetowania losowego wyboru podziału zbioru na dane walidacyjne i uczące. Budowa modelu w oparciu o zbiór

	Manualne			Głębol	Głębokie			Fuzja		
	TPR	TNR	ACC	$\mathrm{TPR}$	TNR	ACC	$\mathrm{TPR}$	TNR	ACC	
Uczenie	0.9165	0.9309	0.9245	0.8578	0.7633	0.8052	0.9213	0.8989	0.9089	
Walidacja	0.8476	0.8523	0.8501	0.8601	0.7585	0.8055	0.8850	0.8615	0.8724	
Test	0.6846	0.7932	0.7405	0.7061	0.8576	0.7841	0.7366	0.8439	0.7919	

Tabela 4.12: DW, uczone na zbiorze BreakHis + GZG, natomiast testowane na zbiorze SzUZG

danych BreakHis+GZG na podstawie zestawu cech manualnych zakończyła się wynikiem dla uogólnionego współczynnika  $R^2$  na poziomie 0.7399 dla danych uczących oraz 0.6532 dla walidacyjnych. Współczynnik AUC dla danych uczących uzyskał wartość 0.9824 natomiast dla walidacyjnych 0.9390. Model do uzyskania najlepszego wyniku potrzebował 135 warstw. Kolejny model został zbudowany w oparciu o zestaw cech głębokich. Tym razem liczba warstw wykorzystanych w drzewie wyniosła 91. Wartość uogólnionego współczynnika determinacji wyniosła 0.6039 dla danych uczących oraz 0.5786 dla walidacyjnych. Pole AUC dla danych uczących uzyskało wynik 0.8995 a dla walidacyjnych 0.8827. Mo-

del zbudowany na fuzji cech charakteryzował się 145 warstwami. Wartość uogólnionego współczynnika  $R^2$  wyniosła 0.7798 dla danych uczących oraz 0.7166 dla walidacyjnych. Pola AUC zostały oszacowane na 0.9771 oraz 0.9491 odpowiednio dla danych uczących oraz walidacyjnych.

W wyniku przeprowadzonych testów najgorszą dokładność klasyfikacji (tab. 4.12) uzyskał model zbudowany na bazie cech manualnych (0.7405). Wyraźnie lepszy rezultat na danych testowych uzyskał model zbudowany na zestawie cech głębokich (0.7841), natomiast najlepiej wypadł model zbudowany na fuzji cech (0.7919).

	Manualne			Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR	ACC
Uczenie	0.8374	0.9099	0.8747	0.8758	0.8756	0.8757	0.8939	0.8927	0.8932
Walidacja	0.7733	0.8507	0.8134	0.8478	0.8624	0.8554	0.8678	0.8527	0.8600
Test	0.4218	0.8342	0.6497	0.8118	0.6341	0.7136	0.7558	0.6757	0.7115

Tabela 4.13: DW, uczone na zbiorze SzUZG, natomiast testowane na zbiorze BreakHis + GZG

Cechy manualne (ze zbioru SzUZG) posłużyły do zbudowania modelu, którego wartość uogólnionego współczynnika determinacji dla zbioru uczącego wyniosła 0.6960, natomiast dla walidacyjnego 0.6043. Pole pod krzywą ROC dla danych uczących uzyskało wynik 0.9525, natomiast dla walidacyjnych 0.9055. Drzewo uzyskało najlepszy wynik po 104 warstwach. W przypadku zestawu cech głębokich liczba warstw potrzebnych do uzyskania najlepszego modelu wyniosła 115 czyli nieco więcej niż dla manualnych. Jest to zaskakujące zważywszy znacznie mniejszą liczbę cech głębokich w modelu. Wartość uogólnionego współczynnika  $R^2$  dla danych uczących to 0.7561, natomiast dla walidacyjnych 0.7017. Pole AUC osiągnęło natomiast wynik 0.9556 dla danych uczących oraz 0.9363 dla walidacyjnych. Fuzja cech manualnych i głębokich potrzebowała 90 warstw do uzyskania najlepszego modelu. Dane uczące uzyskały wartość współczynnika determinacji wynoszącą 0.7691 a dane walidacyjne 0.7119. Pole AUC charakteryzowało się dla danych uczących wynikiem 0.9640, natomiast dla walidacyjnych 0.9420.

Klasyfikator zbudowany na zestawie cech manualnych uzyskał bardzo słaby wynik wykrywania przypadków złośliwych (TPR = 0.4218). Wynik dokładności dla tego modelu jest podnoszony do wartości 0.6497 dzięki sprawnemu wykrywaniu obrazów łagodnych (TNR = 0.0.8342). Pozostałe dwa klasyfikatory uzyskały porównywalne rezultaty dla danych uczących: cechy głębokie (ACC = 0.7136) oraz fuzja cech (ACC = 0.7115).

Końcowe podsumowanie klasyfikatorów oraz wyników uzyskanych na obu zbiorach danych (BreakHis + GZG oraz SzUZG) wskazują, że średni najniższy wynik 0.6951 uzyskano dla zestawu cech manualnych (tab. 4.13). Wyraźnie lepszy wynik średni uzyskano dla zestawu cech głębokich (0.7488) oraz dla zestawu składającego się z fuzji cech manualnych oraz głębokich (0.7517). **Klasyfikator naiwny Bayesa** Klasyfikator ten charakteryzuje się brakiem wstępnych parametrów modelu do ustalenia. Oznacza to, że jedyna opcja jaka może wpływać na wynik klasyfikacji to podział danych na zbiór danych walidacyjnych oraz uczących. Istotne jest również uzyskanie powtarzalności modelu dlatego zresetowano losowość podziału danych oraz ustawiono podobnie jak w pozostałych klasyfikatorach wartość wydzielonej części walidacyjnej na 0.2 całości zbioru.

	Manualne			Głębol	Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR	ACC	
Uczenie	0.6504	0.7370	0.6983	0.9016	0.6617	0.7692	0.7925	0.7093	0.7465	
Walidacja	0.6387	0.7391	0.6942	0.8993	0.6519	0.7618	0.8246	0.6874	0.7489	
Test	0.7523	0.7776	0.7651	0.7483	0.8377	0.7944	0.7661	0.8111	0.7893	

Tabela 4.14: NB uczony na zbiorze BreakHis + GZG, natomiast testowany na zbiorze SzUZG

Klasyfikator NB zbudowany na cechach manualych pozyskanych ze zbioru BreakHis + GZG uzyskał wynik AUC dla danych uczących wynoszący 0.7264 (łagodne) oraz 0.7265 (złośliwe), natomiast dla danych walidacyjnych 0.7231 (łagodne) oraz 0.7224 (złośliwe). Użycie cech głębokich do budowy klasyfikatora spowodowało uzyskanie wartości współczynnika AUC dla danych uczących wynoszącej 0.8162 (łagodne i złośliwe), natomiast dla walidacyjnych 0.8163 (łagodne) i 0.8170 (złośliwe). Klasyfikator zbudowany na fuzji cech charakteryzował się wartością współczynnika AUC dla danych uczących na poziomie 0.7777 (łagodne) oraz 0.7786 (złośliwe). Wartość współczynnika AUC dla danych walidacynych z kolei uzyskała wartość 0.7775 zarówno dla przypadków złośliwych jak i lagodnych.

Klasyfikator NB zbudowany na zestawie cech manualnych uzyskał najniższy wynik ACC (0.7523). Pozostałe zestawy cech uzyskały odpowiednio: 0.7944 dla cech głębokich oraz 0.7893 dla fuzji cech (tab. 4.14). Interesujący wydaje się być fakt, iż współczynnik dokładności ACC uzyskał wyższe wartości na danych testowych, niż dla danych uczących oraz walidacyjnych. Można zatem przypuszczać, że klasyfikator NB wykazuje się dużą odpornością na nadmierne dopasowanie do badanych danych.

Tabela 4.15: NB, uczony na zbiorze SzUZG, natomiast testowane na zbiorze BreakHis + v	GZ	Z	7	(	5
---	----	---	---	---	---

	Manualne		Głęboł	Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	$\mathrm{TPR}$	TNR	ACC
Uczenie	0.7659	0.7823	0.7743	0.8193	0.8531	0.8366	0.8148	0.8210	0.8180
Walidacja	0.7881	0.7794	0.7835	0.8069	0.8571	0.8333	0.7805	0.8207	0.8011
Test	0.6028	0.7550	0.6869	0.8421	0.6351	0.7277	0.7531	0.7029	0.7253

Klasyfikator budowany na zestawie cech manualnych pozyskanych ze zbioru SzUZG charakteryzuje się wartością pola AUC dla danych uczących wynoszącą 0.8013 (łagodne) oraz 0.7997 (złośliwe). Dla danych walidacyjnych wielkości pól AUC wynoszą 0.8063 dla

złośliwych oraz 0.8067 dla łagodnych przypadków. Kolejny klasyfikator NB tym razem zbudowany na cechach głębokich charakteryzował się polem AUC o wielkości 0.8756 (łagodne) oraz 0.8757 (złośliwe) dla zestawu uczącego. Dla danych walidacyjnych, natomiast AUC wyniosło 0.8681 w stosunku do łagodnych przypadków oraz 0.8680 dla złośliwych. Fuzja cech nie przyniosła polepszenia wartości pól AUC. Dla danych uczących wartość pola AUC wynosi 0.8337 oraz 0.8334 odpowiednio dla przypadków łagodnych oraz złośliwych. Dla danych walidacyjnych natomiast wartość pola AUC wyniosła 0.8174 zarówno dla przypadków łagodnych oraz złośliwych.

Klasyfikatory zbudowane na zbiorze SzUZG charakteryzują się wyższymi wartościami współczynnika ACC dla danych uczących i walidacyjnych względem danych testowych. Najniższy wynik klasyfikacji uzyskano dla modelu zbudowanego na bazie cech manualnych (0.6869). Lepszy wynik uzyskały klasyfikatory trenowane na zestawie cech głębokich (0.7277) oraz fuzji cech (0.7253). Można również zauważyć, że fuzja cech powoduje wyrównanie poziomów TPR i TNR względem ich poziomów uzyskanych dla zestawu cech manualnych i głębokich.

Średni najwyższy wynik dokładności dla obu zbiorów danych uzyskały klasyfikatory zbudowany na zestawie cech głębokich (0.7610). Minimalnie niższy wynik średni uzyskały modele zbudowane na fuzji cech (0.7573). Najniższy wynik po raz kolejny odnotowano dla cech manualnych (0.7260). Specyficzną cechą NB są stosunkowo niskie wartości dokładności uzyskiwane na zbiorach uczących oraz walidacyjnych (tab. 4.15).

**Klasyfikator k-NN** Klasyfikator ten w porównaniu z innymi klasyfikatorami posiada stosunkowo niewielki zestaw parametrów do ustalenia. W zasadzie jedynym parametrem dla modeli zbudowanych na zmiennych ciągłych jest maksymalna liczba k-najbliższych sąsiadów jaka będzie testowana w modelu. Spośród tej liczby jest wybierana taka, której model uzyskuje najmniejszy błąd klasyfikacji dla danych walidacyjnych.

Tabela 4.16: Klasyfikator k-NN, uczony na zbiorze Break<br/>His+ GZG, natomiast testowany na zbiorze<br/> SzUZG

	Manualne			Głębol	Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	$\mathrm{TPR}$	TNR	ACC	
Uczenie	0.7741	0.8295	0.8049	0.8714	0.7069	0.7798	0.8179	0.8224	0.8204	
Walidacja	0.7807	0.8546	0.8204	0.8824	0.7362	0.8039	0.8209	0.8531	0.8382	
Test	0.7026	0.8318	0.7691	0.7333	0.8551	0.7960	0.6925	0.8638	0.7808	

Pierwszy model klasyfikatora k-NN został zbudowany na cechach manualnych. Model ten został wybrany dla liczby k równej 19 sąsiadów ponieważ uzyskał najniższy błąd klasyfikacji dla danych walidacyjnych. Wartość współczynnika  $R^2$  dla danych uczących wyniósł 0.3900, natomiast dla walidacyjnych 0.4102. Zestaw cech głębokich najlepszy model uzyskał dla liczby k o wartości 94. Uzyskana wartość współczynnika  $R^2$  dla danych uczących to 0.3693, a dla walidacyjnych 0.4093. Model zbudowany na fuzji cech najlepszy wynik klasyfikacji dla danych walidacyjnych został osiągnięty dla 11 sąsiadów. Wartość współczynnika determinacji dla danych uczących wyniosła 0.4455 natomiast dla walidacyjnych 0.4670.

Klasyfikatory k-NN najwyższy wynik (tab. 4.16) uzyskały na zestawie cech głębokich (0.7960), następnie na bazie fuzji cech (0.7808) oraz cech manualnych (0.7691). Ogólnie można zatem powiedzieć, że na wszystkich zestawach cech klasyfikator uzyskał wysokie wyniki. Klasyfikator k-NN uzyskał na zestawie cech manualnych najwyższy wynik klasyfikacji z dotychczas zbadanych modeli.

Tabela 4.17: Klasyfikator k-NN, trenowany na zbiorze SzUZG, natomiast testowana na zbiorze BreakHis + GZG

	Manualne			Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR	ACC
Uczenie	0.7865	0.8646	0.8266	0.8455	0.8564	0.8511	0.8425	0.8797	0.8616
Walidacja	0.7775	0.8839	0.8327	0.8489	0.8654	0.8574	0.8384	0.8868	0.8635
Test	0.4539	0.8165	0.6543	0.8080	0.6360	0.7129	0.6794	0.7166	0.7000

Zestaw cech manualnych najlepszy wynik klasyfikacji dla danych walidacyjnych uzyskał dla klasyfikatora o wartość k wynoszącej 19. Wartość współczynnika  $R^2$  dla danych uczących wyniosła 0.4576, natomiast dla danych walidacyjnych 0.4669. Dla zestawu cech głębokich klasyfikator dla danych walidacyjnych najniższy błąd klasyfikacji uzyskał dla modelu zbudowanego na 47 najbliższych sąsiadach. Wartości współczynnika determinacji dla danych walidacyjnych wyniosła 0.5359, natomiast dla uczących 0.5155. Ostatni test dotyczył wyników dla fuzji cech. W tym przypadku również optymalna liczba najbliższych sąsiadów otrzymała wartość 23. Wartość współczynnika determinacji dla danych uczących wyniosła 0.5399 natomiast dla walidacyjnych 0.5555.

W odróżnieniu od klasyfikatorów k-NN zbudowanych na zbiorze BreakHis + GZG, modele zbudowane na zbiorze SzUZG charakteryzują się znacznie niższymi rezultatami klasyfikacji na danych testowych (tab. 4.17). Dla zestawu cech manualnych dokładność osiągnęła wynik wynoszący jedynie 0.6543. Wynik ten pośrednio wynika z bardzo niskiej wartości współczynnika TPR, którego wartość wyniosła 0.4539. Pozostałe dwa modele uzyskały odpowiednio wyniki: dla zestawu cech głębokich: 0.7129 oraz dla fuzji cech 0.7000.

Najlepszy wynik średni uzyskał model zbudowany na bazie cech głębokich (0.7544), nieco niższy wynik uzyskał model bazujący na fuzji cech 0.7404, najniższy średni wynik osiągnęły modele zbudowane na zestawie cech manualnych (0.7117).

Maszyna wektorów nośnych z radialną funkcją aktywacji Klasyfikator SVM w oprogramowaniu JMP występuje w wersji liniowej oraz radialnej. Po krótkiej analizie wyników testowych nieco lepsze rezultaty otrzymano w modelach klasyfikatorów SVM w wersji radialnej. Ponadto wersje radialne okazały się mniej złożone obliczeniowo dla komputera. Wersja radialna klasyfikatora SVM posiada dwa główne parametry w oprogramowaniu JMP. Pierwszy z nich to Cost, który odpowiada za szerokość marginesu (im niższa wartość parametru tym margines szerszy). Parametr Cost znajduje się w przedziale od 0 do 1, a wartość domyślna jest ustawiona 1 i taką pozostawiono. Wartość 1 oznacza, że algorytm jest skłonny do uzyskania niższego błędu klasyfikacji [82]. Parametr Cost występuje również w liniowej wersji algorytmu, w przeciwieństwie do parametru Gamma, występującego jedynie w wersji radialnej. Parametr Gamma odpowiada za krzywiznę lini decyzyjnej, im wyższa tym większa krzywizna [82]. Duża wartość krzywizny pozwala na bardziej elastyczne dopasowanie, to z kolei może prowadzić do łatwego przeuczenia [82]. Ze względu trudność związaną z ustaleniem idealnego parametru, pozostawiono wartość domyślną wynoszącą jeden dzielone na liczbę parametrów.

Tabela 4.18: SVM z radialną funkcją aktywacji,	uczona na zbiorze	BreakHis + GZG	, natomiast
$testowana\ na\ zbiorze\ SzUZG$			

	Manualne			Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR	ACC
Uczenie	0.9074	0.9299	0.9198	0.8860	0.7043	0.7856	0.9298	0.9273	0.9284
Walidacja	0.8706	0.8936	0.8833	0.8978	0.7262	0.8029	0.9006	0.8877	0.8935
Test	0.6618	0.8697	0.7689	0.7321	0.8458	0.7907	0.6925	0.8769	0.7875

Model klasyfikatora maszyny wektorów nośnych zbudowany na zbiorze BreakHis + GZG (zestaw cech manualnych) uzsykał wartość uogólnionego współczynnika  $R^2$  na poziomie 0.7531 dla danych walidacyjnych oraz 0.8211 dla uczących. Model ten został zbudowany w oparciu o zestaw cech manualnych. Łączna licczba wektorów nośnych w tym modelu to 4535. Pole AUC dla tego modelu uzyskało wartość 0.9717 dla danych uczących oraz 0.9547 dla walidacyjnych. Kolejny model został zbudowany na zestawie cech głębokich. Model ten wykorzystał 4599 wektorów nośnych, natomiast wartość uogólnionego współczynnika  $R^2$  wyniosła 0.4945 dla danych uczących oraz 0.5222 dla walidacyjnych. Pola AUC uzyskały niższe wartości niż w poprzednim modelu i wynosiły 0.8666 dla danych uczących oraz 0.8702 dla walidacyjnych. Trzeci model został zbudowany na zestawie fuzji cech i charakteryzował się 4088 wektorami nośnymi. Wartość uogólnionego współczynnika  $R^2$  wyniosła 0.8463 dla danych uczących oraz 0.7804 dla walidacyjnych. Pole pod krzywą ROC uzyskało wynik 0.9780 dla danych uczących oraz 0.9623 dla walidacyjnych.

Spośród klasyfikatorów zbudowanych na zbiorze BreakHis + GZG a testowanych na SzUZG najlepszy wynik klasyfikacji (tab. 4.18) uzyskano dla zestawu cech głębokich (0.7907), nieznacznie gorszy wynik odnotowano dla fuzji cech (0.7875). Najniższy wynik uzyskano natomiast dla zestawu cech manualnych, jednakże jest to jeden z najwyższych uzyskanych dla tego zestawu wyników (0.7689).

	Manualne			Głębokie			Fuzja		
	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR	ACC
Uczenie	0.8664	0.9205	0.8943	0.8541	0.8591	0.8566	0.9132	0.9162	0.9147
Walidacja	0.8371	0.8815	0.8600	0.8514	0.8642	0.8580	0.8975	0.8940	0.8957
Test	0.5667	0.8504	0.7235	0.8163	0.6350	0.7161	0.7638	0.7631	0.7634

Tabela 4.19: SVM z radialną funkcją aktywacji, uczona na zbiorze SzUZG, natomiast testowana na zbiorze BreakHis+GZG

Kolejny zestaw modeli budowany będzie na zbiorze obrazów ze zbioru SzUZG. Zestaw cech manualnych został wykorzystany do zbudowania pierwszego modelu, który charakteryzował się liczbą 3534 wektorów nośnych. Dla tego modelu wartość uogólnionego współczynnika  $R^2$  na danych uczących wyniosła 0.7795, natomiast na walidacyjnych 0.6915. Pole AUC dla danych uczących uzyskało wartość 0.9610, a dla walidacyjnych 0.9361. Kolejny model bazował na zestawie cech głębokich i uzyksał 2651 wektorów nośnych. Ten model z kolei charakteryzował się wartością uogólnionego współczynnika  $R^2$  na poziomie 0.6329 dla danych uczących oraz 0.6426 dla walidacyjnych. Pole pod krzywą ROC uzyskało wartość 0.9146 (dane uczące) oraz 0.9177 (dane walidacyjne). Ostatni model został zbudowany na zestawie fuzji cech, a do jego zbudowania wykorzystano 2897 wektorów nośnych. Generalizowany współczynnik  $R^2$  uzyskał wartość 0.8260 dla danych uczących oraz 0.7800 dla walidacyjnych. Wartość pola AUC dla danych uczących to 0.9731, natomiast dla walidacyjnych 0.9616.

Dla klasyfikatorów SVM budowanych na zbiorze SzUZG najwyższy wynik dokładności (tab. 4.19) uzyskano dla fuzji cech (0.7634). Drugi najwyższy wynik wyjątkowo uzyskany został dla zestawu cech manualnych (0.7235), natomiast najniższy wynik dokładności tym razem uzyskano dla zestawu cech głębokich (0.7161).

Po uśrednieniu uzyskanych wartości dokładności uzyskanych na danych testowych ranking zestawów danych jest następujący: najlepszy wynik (0.7754) dla fuzji cech, kolejny wynik to 0.7534 dla zestawu cech głębokich oraz 0.7462 dla zestawu cech manualnych. Jest to najwyższy średni wynik klasyfikacji uzyskany dla zestawów cech manualnych.

Klasyfikator neuronowy Klasyfikator neuronowy (KN) dostępny w oprogramowaniu JMP Pro [82] umożliwia wykorzystanie perceptronu wielowarstwowego z maksymalną liczbą dwóch warstw ukrytych z możliwością definiowania neuronów z funkcjami aktywacji: tangensa hiperbolicznego, liniową oraz Gaussa. Po wielu eksperymentach ustalono, że rozbudowa warstw sieci ukrytych poprawia jakość klasyfikatora dla danych uczących oraz walidacyjnych, ale niestety pogarsza wyniki dla danych testowych. Za taki stan rzeczy najprawdopodobniej odpowiada wzrost skłonności bardziej rozbudowanych sieci neuronowych do przeuczenia. W związku z powyższym podjęto decyzję o wprowadzeniu najprostszego modelu z jednym neuronem z funkcją aktywacji w postaci tangensa hiperbolicznego. Oprócz wspomnianych możliwości moduł KN umożliwia wykonanie modelu wzmacnianego (boosting), działającego na podobnej zasadzie jak DW. Istnieje również możliwość transformacji zmiennych ciągłych do stanu zbliżonego do rozkładu normalnego przy użyciu rozkładu  $S_U$  lub  $S_B$  Johnsona. Niestety żadna z wymienionych modyfikacji nie poprawia, a wręcz pogarsza wyniki dla danych testowych, niemniej wyniki dla danych uczących oraz walidacyjnych ulegają poprawie. W modelach wybrano metodę kwadratowej funkcji kary oraz liczbę modeli do porównania równą 100. Oznacza to że wybierany jest najlepszy model spośród 100 wygenerowanych. Z kolei najlepszy model oznacza taki, w którym uzyskano najniższy błąd dla danych walidacyjnych. Jako metodę walidacji wybrano sprawdzian krzyżowy pięciokrotny (Kfold), który zwiększa czas uczenia modelu względem prostej metody podziału (Holdout), jednakże w podczas uczenia bardzo prostego modelu KN czas ten w obu przypadkach jest niski.

Tabela 4.20: KN z pojedynczym neuronem z aktywacją Tan<br/>H, uczony na zbiorze z Break His+GZG, natomiast testowany na zbiorze<br/> SzUZG

	Manualne		Głębokie			Fuzja			
	TPR	TNR	ACC	TPR	TNR	ACC	TPR	TNR	ACC
Uczenie	0.8409	0.8869	0.8663	0.8623	0.7346	0.7917	0.8833	0.8780	0.8804
Walidacja	0.8387	0.8870	0.8654	0.8596	0.7371	0.7919	0.8961	0.8811	0.8878
Test	0.4326	0.9289	0.6883	0.7100	0.8558	0.7851	0.5793	0.9002	0.7446

Pierwszy model KN został zbudowany w oparciu o cechy manualne wyodrębnione ze zbioru BreakHis + GZG. Model ten charakteryzował się wartością uogólnionego współczynnika  $R^2$  wynoszącego 0.6909 dla danych uczących oraz 0.6992 dla danych walidacyjnych. Wartość współczynnika AUC została oszacowana na 0.9374 dla danych uczących oraz 0.9380 dla walidacyjnych. Kolejny model KN został zbudowany na bazie pozyskanych cech głębokich. Osiągnięta przez ten model wartość uogólnionego współczynnika  $R^2$ wyniosła 0.5667 dla danych uczących oraz 0.5416 dla walidacyjnych. Pole pod krzywą ROC dla danych uczących uzyskało wartość 0.8800, natomiast dla walidacyjnych 0.8700. Ostatni model został zbudowany na fuzji cech osiągnął wartość współczynnika determinacji  $R^2$  wynoszącą 0.7490 dla danych uczących oraz 0.7668 dla danych walidacyjnych. Pola AUC uzyskały wyniki 0.9522 oraz 0.9575 odpowiednio dla danych uczących oraz walidacyjnych.

Biorąc pod uwagę zestaw cech manualnych osiągnięty został przeciętny wynik dokładności wynoszący 0.6883, uwagę zwraca natomiast stosunek wielkości współczynników TPR do TNR wynoszący 0.4326 do 0.9289 (tab. 4.20). Taki rezultat wskazuje na znakomitą klasyfikację przypadków łagodnych oraz słabą klasyfikację przypadków złośliwych. Najlepszy wynik dokładności klasyfikacji został uzyskany dla zestawu cech głębokich (0.7851), natomiast wyraźnie gorszy wynik uzyskano dla fuzji cech (0.7446)

KN zbudowany na cechach manualnych ze zbioru SzUZG charakteryzował się warto-
	Manualne		Głębokie			Fuzja			
	TPR	TNR	ACC	$\mathrm{TPR}$	TNR	ACC	$\mathrm{TPR}$	TNR	ACC
Uczenie	0.8581	0.8728	0.8657	0.8571	0.8579	0.8575	0.8925	0.9017	0.8972
Walidacja	0.8567	0.8872	0.8724	0.8648	0.8651	0.8649	0.8934	0.9085	0.9012
Test	0.6890	0.8296	0.7667	0.7963	0.6390	0.7093	0.7618	0.7862	0.7753

Tabela 4.21: KN z pojedynczym neuronem z aktywacją Tan<br/>H, uczony na zbiorze SzUZG, natomiast testowana na zbiorze Break<br/>His+ GZG

ścią uogólnionego współczynnika  $R^2$  wynoszącą 0.7197 dla danych uczących oraz 0.7291 dla danych walidacyjnych. Pole AUC dla danych uczących uzyskało wielkość 0.9431, natomiast dla walidacyjych 0.9458. Model zbudowany na zestawie cech głębokich uzyskał dla danych uczących wartość uogólnionego współczynnika determinacji wynoszącą 0.7014, natomiast dla walidacyjnych 0.7043. Pole pod krzywą ROC uzyskało wielkość równą 0.9369 dla danych uczących oraz 0.9376 dla walidacyjnych. Ostatni model został zbudowany na bazie fuzji cech manualnych oraz głębokich. Model ten charakteryzował się wartością uogólnionego współczynnika  $R^2$  wynoszącą 0.7934 dla danych uczących oraz 0.8056 dla walidacyjnych. Wartość pola AUC dla danych uczących to 0.9647, natomiast dla walidacyjnych 0.9680.

Dla KN zbudowanych na zbiorze SzUZG najwyższy wynik dokładności klasyfikacji uzyskano dla fuzji cech (0.7753), drugi wynik uzyskano dla zestawu cech manualnych (0.7667), natomiast najniższy wynik dla zestawu cech głębokich (0.7093). Co ciekawe kolejność wyników w tym przypadku jest identyczna jak dla klasyfikatorów SVM modelowanych na zbiorze SzUZG.

Uśrednione na podstawie badań wartości dokładności (tab. 4.21) kształtują się następująco: 0.7275 (cechy manualne), 0.7472 (cechy głębokie) oraz 0.7599 (fuzja cech manualnych i głębokich). Fuzja cech w tym przypadku zwiększyła średnią dokładność modelu zbudowanego na zestawie cech głębokich o ponad 1%. Zastanawiający jest fakt bardzo niskiej dokładności klasyfikacji (0.6883) modelu uczonego dla zestawu cech manualnych na danych BreakHis + GZG, a testowanego na danych SzUZG. Biorąc pod uwagę wartość współczynnika TPR (0.4326), można wnioskować, że głównym problemem tego klasyfikatora bazującego na danych manualnych była niska wykrywalność obrazów z przypadkami złośliwymi.

**Podsumowanie wyników klasyfikacji** Najlepszy wynik klasyfikacji na zbiorach testowych uzyskano dla uogólnionego modelu regresji z regularyzacją siatki elastycznej. Średnie wartości dokładności klasyfikacji obrazów z dwóch różnych zestawów danych dla najlepszych modeli bazujących na fuzji cech wyniosły: 0.7805 dla regresji z regularyzacji metodą siatki elastycznej (Fuzja\_se) oraz 0.7826 dla modelu regresji krokowej (Fuzja\_rkw). Modele zbudowane na zaproponowanej fuzji cech wyprzedziły mniej więcej o 2 % najlepszy wynik uzyskany dla cech głębokich (0.7610) bazujący na klasyfikatorze NB. Jednocześnie uzyskany dla cech głębokich wynik jest nieco lepszy niż wynik uzyskany z klasyfikacji siecią neuronową VGG19 (0.7568). Najlepszy średni wynik klasyfikacji (0.7462) uzyskany dla cech manualnych uzyskano dla modelu SVM i wynik ten jest niższy o około 3,5% względem najlepszego wyniku dla fuzji cech. Do oceny klasyfikacji można również zbadać i porównać najniższe wyniki wchodzące w skład najlepszych średnich. Wówczas najniższy wynik dla fuzji cech to 0.7792 dla cech głębokich 0.7277 czyli o nieco ponad 5% gorzej, natomiast dla cech manualnych 0.7235 czyli nieco ponad 5,5% gorzej aniżeli wynik dla fuzji cech. Pełny obraz najlepszych wyników uzupełnia tabela 4.22. Na końcu tabeli umieszczono również

Tabela 4.22: Podsumowanie najlepszych klasyfikatorów na podstawie dokładności uzyskanych dla danych testowych z podziałem na zestawy cech (Zapis BreakHis + GZG skrócono w tabeli do BreakHis)

Dane	BreakHis	SzUZG	Średnia				
Cechy Manualne (SVM)							
Uczenie	0.9198	0.8943	0.9070				
Walidacja	0.8833	0.8600	0.8716				
Test	0.7689	0.7235	0.7462				
(	Cechy Głębo	okie (NB)					
Uczenie	0.7692	0.8366	0.8029				
Walidacja	0.7618	0.8333	0.7975				
Test	0.7944	0.7277	0.7610				
$\mathbf{F}$	uzja_se (sele	ekcja cech	)				
Uczenie	0.8579	0.8815	0.8697				
Walidacja	-	-	-				
Test	0.7819	0.7792	0.7805				
Fu	zja_rkw (sel	lekcja cecl	n)				
Uczenie	0.8574	0.8853	0.8714				
Walidacja	-	-	-				
Test	0.7726	0.7927	0.7826				
Fuzja_ssg (stochastyczna selekcja cech)							
Uczenie	0.8319	0.8696	0.8508				
Walidacja	0.8295	0.8729	0.8512				
Test	0.8049	0.7791	0.7920				

wyniki klasyfikacji obrazów przy użyciu zaproponowanej metody selekcji cech z regulatyzacją L2 (Fuzja\_ssg). Zaproponowana metoda selekcji cech pozwoliła uzyskanie wyniku średniej dokładności klasyfikacji obrazów na dwóch zbiorach danych wyższej o 1% względem pozostałych metod. W tabeli można również zauważyć dość ciekawą prawidłowość związaną z klasyfikatorem wybranym dla zestawu cech głębokich. Mianowicie klasyfikator ten (NB) poradził sobie najgorzej z klasyfikacją na danych uczących. Można by zatem alternatywnie wykorzystać klasyfikator kNN lub SVM, które uzyskały nieznacznie gorsze średnie wyniki klasyfikacji dla danych testowych jednocześnie posiadając wyraźnie lepsze wyniki klasyfikacji na danych uczących. Dodatkowo można zauważyć, iż najlepszy model klasyfikatora zbudowanego na zestawie cech głębokich uzyskuje wyższe wyniki klasyfikacji niż najlepsza pojedyncza sieć neuronowa (VGG19). Wynik ten (0.7610) zatem potwierdza nieco wyższą skuteczność zaproponowanego zespołu generatorów cech względem pojedynczych modeli (0.7408-0.7568).

# 4.8.3.5 Wyniki diagnostyki nowotworów na podstawie sklasyfikowanych obrazów testowych w odniesieniu do pojedynczego pacjenta

Wstęp Pojedynczy obraz wykorzystany do klasyfikacji ma rozmiar 230x230 pikseli, co może się okazać zbyt małym zasobem do postawienia prawidłowej diagnozy. Te niewielkie obszary diagnostyczne stanowią jednak fragmenty większych obrazów, których zbiór ostatecznie przypisany jest do pojedynczego pacjenta. Finalnym testem dla grupy najlepszych klasyfikatorów było badanie krzywej Czułość-Precyzja oraz obliczenie na jej podstawie wartości średniej precyzji. Współczynnik ten wskaże, który klasyfikator poradził sobie najlepiej w zadaniu klasyfikacji przypadków nowotworu na poziomie pacjentów.

Wyniki klasyfikacji dla modelu trenowanego na zbiorze BreakHis + GZG, natomiast testowanego na zbiorze SzUZG W przypadku skuteczności klasyfikatora zbudowanego na zbiorze BreakHis + GZG można zauważyć, że bez względu na zastosowany zestaw cech należy się spodziewać bardzo wysokiego wyniku dokładności klasyfikacji (rys. 4.29) na zabiorze testowym SzUZG. Wysoki wynik klasyfikacji może wynikać



Rysunek 4.29: Wykres czułości i precyzji dla najlepszych modeli trenowanych na zbiorze BreakHis+ GZG, natomiast testowanych na SzUZG

z większego zróżnicowania przypadków dostępnych w zbiorze uczącym, jak również z faktu uzupełnienia tego zbioru przez zbiór GZG pochodzących z tego samego ośrodka co zbiór

testowy. Z drugiej jednak strony zbiór uczący (BreakHis + GZG) jest znacznie bogatszy pod względem indywidualnych przypadków niż zbiór testowy SzUZG.

Wyniki klasyfikacji dla modelu trenowanego na zbiorze SzUZG, natomiast testowanego na zbiorze BreakHis + GZG Jak ustalono powyżej zbiór SzUZG jest zbiorem uboższym pod kątem indywidualnych przypadków, stąd wniosek, iż trudniejszym zadaniem klasyfikacyjnym dla przygotowanych modeli było uczenie na zbiorze SzUZG, natomiast testowanie na zbiorze BreakHis + GZG. W zadaniu tym najwyższa skuteczność klasyfikacji (rys. 4.30) została uzyskana dla modelu klasyfikatora zbudowanego na zestawie zawierającym fuzję cech manualnych oraz głębokich. Przebieg krzywej czułość-precyzja również wskazuje jako najlepszy oraz odznaczający się największymi zdolnościami uogólniającymi model zbudowany na zestawie fuzji cech. Krzywa ta wyka-



Rysunek 4.30: Wykres czułości i precyzji dla najlepszych modeli trenowanych na zbiorze SzUZG, natomiast testowanych na BreakHis + GZG

zuje lepsze proporcje czułości do precyzji w zakresie ich wysokich wartości. Najwyższa wartość średniej precyzji została uzyskana dla klasyfikatora zbudowanego w oparciu o zaproponowana metodę selekcji cech, jednocześnie przekraczając wartość 0.9. Najwyższy osiągnięty rezultat klasyfikacji pacjentów przekroczył wartość 84%.

Wyniki klasyfikacji dla modelu trenowanego na zbiorze SzUZG, natomiast testowanego na zbiorze BreakHis Ostatni wynik (rys. 4.31) jest ciekawy z punktu widzenia ogólnych zdolności uogólniających proponowanych modeli klasyfikacji. Dane uczące pochodzą bowiem w całości z innego ośrodka medycznego niż dane testowe. Co ciekawe maksymalne uzyskane wyniki dla wszystkich wybranych klasyfikatorów przekraczają 80%. Przebieg krzywej precyzja-czułość oraz obliczony współczynnik średniej precyzji wskazują,



Rysunek 4.31: Wykres czułości i precyzji dla najlepszych modeli trenowanych na zbiorze SzUZG, natomiast testowanych na BreakHis

iż najlepszy klasyfikator powstał na bazie fuzji cech wybranych przy zastosowaniu proponowanej metody selekcji cech. Spośród badanych klasyfikatorów trzy wyniki przekroczyły wartość średniej precyzji 0.9. Model regresji logistycznej z krokową selekcją cech (0.917) oraz zaproponowany model regresji logistycznej z stochastyczną selekcją cech z najwyższym wynikiem wynoszącym **0.922**.

#### 4.9 Dyskusja

W literaturze często można spotkać wyniki klasyfikacji obrazów medycznych sięgających niemal 100% w obrębie materiałów pobranych z jednego laboratorium medycznego. Biorąc pod uwagę wyniki dokładności klasyfikacji uzyskane na danych uczących oraz walidacyjnych pochodzących z jednego ośrodka (SzUZG) lub nawet w niewielkim stopniu przemieszane (BreakHis + GZG) wskazują na niemal idealną klasyfikację przypadków nowotworów. Przykładowo dla zbioru BreakHis +GZG dokładność klasyfikacji siecią VGG16 na danych uczących wyniosła 0.9982, natomiast na danych walidacyjnych 0.9943. Taki zestaw danych medycznych pomimo pewnej różnorodności przypadków okazuje się być niewystarczający do zbudowania modelu o zdolnościach uogólniających, gdyż model ten na danych testowych SzUZG osiągnął wartość 0.5020. Problem ten z dużym prawdopodobieństwem mógłby być zredukowany gdyby dane medyczne posiadały ogromne zbiory rozmaitych przypadków. Niestety w praktyce zbiory medyczne najczęściej posiadają kilkadziesiąt przypadków. Kolejnym problemem jest różnorodność biologiczna występujących przypadków, która powoduje, że nawet lekarz z wieloletnim doświadczeniem często nie jest pewny swoich diagnoz stawianych na podstawie obrazów histopatologicznych lub cytologicznych, przez co jest zmuszony do wykonania dodatkowych badań oraz zbierania konsyliów w celu ustalenia diagnozy. Problem głębokich sieci neuronowych polega na dużej liczbie parametrów. Modele te pomimo stosowania różnych technik unikania nadmiernego dopasowania takich jak: porzucanie (*ang. dropout*), normalizacja, warstwy łączące, regularyzacja, stosowanie danych walidacyjnych, nadal cechują się dużą skłonnością do nadmiernego dopasowania. Nowym podejściem, które pozwoliło zredukować problem nadmiernego dopasowania była normalizacja obrazów na wejściu przy użyciu hybrydowej metody segmentacji. Obrazy zredukowane do masek binarnych okazały się odpowiednim narzędziem do transferu wiedzy pomiędzy danymi z różnych laboratoriów medycznych. Co ciekawe, nawet normalizacja obrazów przy użyciu rozplotu H&E nie pozwoliła na osiągnięcie przez głębokie modele sieci CNN wyższych zdolności uogólniających, można nawet zauważyć lekki spadek. Jest to niespodziewane zważywszy, że obrazy po rozplocie można uznać za etap pośredni pomiędzy przestrzenią RGB a obrazami binarnymi.

Klasyfikatory nadrzędne charakteryzują się różnymi właściwościami klasyfikacyjnymi, zazwyczaj jednak im wyższą wartość dokładności klasyfikacji osiągały na danych uczących tym niższa wartość uzyskiwały dla danych testowych. Bardzo dobrym przykładem tego zjawiska jest model klasyfikatora nadrzędnego LL uczonego na zbiorze SzUZG, dla którego model zbudowany na zestawie cech manualnych osiągnął wynik dokładności klasyfikacji wynoszący 0.9701, podczas gdy wynik na danych testowych ze zbioru BreakHis + GZG wyniósł 0.6883. Odwrotną sytuację można odnotować w przypadku klasyfikatora NB uczonego na zbiorze BreakHis + GZG, gdzie dokładność klasyfikacji modelu klasyfikatora nadrzędnego zbudowanego na cechach manualnych wyniosła 0.6983, podczas gdy dokładność klasyfikacji na zbiorze testowym wyniosła 0.7651. Można też zauważyć, że w przypadku gdy klasyfikator nadrzędny (np. k-NN) osiąga wysokie wyniki klasyfikacji na zbiorze BreakHis + GZG dla danych testowych to osiąga znacznie gorsze rezultaty klasyfikacji modeli uczonych na zbiorze SzUZG. Odwrotna sytuacja ma miejsce dla metod bazujących na drzewach decyzyjnych. Sytuacja wygląda podobnie dla modeli regresyjnych zbudowanych na zestawie cech manualnych oraz głębokich, z wyjatkiem modeli zbudowanych na fuzji cech. Wnioskiem wynikającym z tej analizy jest fakt, iż połączenie zaproponowanej metody fuzji cech w połączeniu z modelami regresyjnymi jest w stanie wygenerować klasyfikatory nadrzędne o najwyższych zdolnościach uogólniających. W żadnej innej konfiguracji nie udało się osiągnąć podobnych rezultatów.

Model regresyjny bazujący na opracowanej metodzie stochastycznej selekcji cech uzyskał najwyższy średni wynik dokładności klasyfikacji. Jednakże w rozbiciu na poszczególne składowe (tab. 4.22) okazuje się, iż najlepszy wynik modelu uczonego na danych SzUZG a testowanego na danych BreakHis + GZG osiągnięty został dla regresji krokowej. Dlaczego zatem wynik diagnostyki na poziomie pacjentów (rys. 4.30 oraz rys. 4.31)wskazuje jako najlepszy model zaproponowanej metody regresji ze stochastyczną selekcją cech? Okazuje się, że model regresji krokowej słabo rozpoznaje obrazy złośliwe, lecz mocno braki nadrabia w rozpoznawaniu przypadków łagodnych. To z kolei przekłada się na gorsze wyniki na poziomie pacjentów. Model regresji ze stochastyczną selekcją cech równomierniej rozkłada błędy, dzięki temu zyskuje przewagę diagnostyczną nad pozostałymi modelami na poziomie pacjentów.

#### 4.10 Podsumowanie

Przeprowadzone w tym eksperymencie badania pozwoliły na wykrycie problemu nadmiernego dopasowania modeli głębokich sieci CNN do danych uczących w postaci obrazów RGB. Rozwiązaniem napotkanego problemu było wykorzystanie jako danych uczących obrazów znormalizowanych przy użyciu metody segmentacji. Ponadto rozwiązanie problemu nadmiernego dopasowania pozwoliło na zbudowanie generatorów cech głębokich oraz dokonanie fuzji cech manualnych z głębokimi. Modele zbudowane na zaproponowanej fuzji cech posiadają wyższe wyniki klasyfikacji oraz większe zdolności uogólniające niż modele zbudowane wyłącznie na cechach manualnych lub głębokich. Do budowy najlepszego modelu zastosowano regresję z zaproponowaną stochastyczną selekcją cech bazującą na regularyzacji L2 oraz losowaniu liczby cech w oparciu o rozkład gamma. Model regresji ze stochastyczną selekcją cech wykazał się nie tylko najlepszą średnią dokładnością klasyfikacji na poziomie obrazów, lecz także najwyższym wynikiem średniej precyzji na poziomie pacjentów. W ramach eksperymentów wykonano dodatkowe badania związane z redukcją wymiarowości oraz uczenia transferowego, badania te jednak nie wykazały się wyższą dokładnością. Wyniki tych badań zostały krótko scharakteryzowane w załączniku C.

### Rozdział 5

### Podsumowanie

Podczas przeprowadzonych eksperymentów wykonano weryfikację segmentacji jąder komórkowych na obrazach cytologicznych z użyciem metod morfologicznych oraz sztucznych sieci CNN. W wyniku przeprowadzonych badań dla zbiorów pochodzących z jednego ośrodka badawczego, okazało się, że metody CNN wykrywają jądra komórkowe z wyższą dokładnością niż pozostałe podejścia do problemu segmentacji. W najnowszej literaturze zdecydowanie najpopularniejszą grupą metod segmentacji są głębokie sieci neuronowe. Niemniej jednak zastosowanie prostszych modeli segmentacji w niektórych przykładach obrazowania medycznego może być wystarczająco skuteczne.

Wyniki klasyfikacji bazujące na procedurze ekstrakcji cech, selekcji cech oraz zastosowaniu najpopularniejszych klasyfikatorów pozwalają osiągać wyniki na poziomie 78%-80% dokładności dla danych pochodzących z różnych ośrodków medycznych. Eksperymenty z klasyfikacją obrazów przy użyciu sieci CNN wskazują na skuteczność przekraczającą 90% jedynie w obrębie danych pochodzących z jednego ośrodka medycznego. W przypadku realizacji zadania klasyfikacji obrazów z różnych ośrodków medycznych niezbędne okazało się opracowanie kompleksowego systemu klasyfikacyjnego obejmującego proces segmentacji, generacji cech głębokich, ekstrakcji cech manualnych, fuzji cech głębokich z manualnymi, opracowania metody selekcji cech oraz wybór klasyfikatora nadrzędnego realizującego końcowy proces modelowania.

Skuteczna segmentacja jest istotna dla dokładnej klasyfikacji. Obrazy po segmentacji zostały zastosowane do wyodrębnienia cech manualnych, a także przetestowane jako wejścia do sieci klasyfikacyjnych. Na podstawie otrzymanych rezultatów najlepsze klasyfikatory osiągały wartość 20-21% błędnie sklasyfikowanych pojedynczych obrazów co przekładało się na 86% - 96% poprawnie zdiagnozowanych pacjentów.

### 5.1 Wnioski

Najlepsze wyniki osiąga zaproponowany system (rys. 1.1) w konfiguracji z modelem regresyjnym jako klasyfikator nadrzędny ze stochastyczną selekcji cech opracowaną w ramach rozprawy. Podobne wyniki dokładności generuje model regresji logistycznej jako klasyfikator nadrzędny z krokową selekcja cech w przód, jednakże model ten wyraźnie gorzej klasyfikuje obserwacje złośliwe, nadrabia natomiast lepszą klasyfikacją obserwacji łagodnych. Pomimo zastosowania różnych metod selekcji cech oraz klasyfikatorów nadrzędnych, wyniki oscylują w granicach 70-80 % dokładności klasyfikacji. Najsłabiej w całym zestawieniu klasyfikatorów nadrzędnych prezentują się metody bazujące na drzewach decyzyjnych.

Za pomocą zaproponowanego podejścia uzyskano wysoką dokładność klasyfikacji raka piersi (96%) na podstawie obrazów cytologicznych oraz histopatologicznych dla obrazów pochodzących z różnych laboratoriów medycznych. Uzyskana dokładność dorównuje podejściom, w którym uczenie i testy wykonywano dla obrazów pochodzących z tego samego laboratorium medycznego. Wypracowana metoda segmentacji pozwala dokładnie segmentować jądra komórkowe na obrazach cytologicznych oraz histopatologicznych. Wykazano, że fuzja cech głębokich oraz manualnych/eksperckich pozwala zwiększyć dokładność klasyfikacji nowotworu piersi. Wykazano, że możliwy jest transfer wiedzy w zadaniach klasyfikacji pomiędzy obrazami cytologicznymi i histopatologicznymi.

### 5.2 Analiza wyników oraz wkład w rozwój dyscypliny

Niewielki zbiór danych uczących jest wystarczający do skutecznej segmentacji obrazów cytologicznych i histopatologicznych przy użyciu opracowanej hybrydowej metody nawet dla obrazów pochodzących z innych ośrodków medycznych jak i różnych metod pobierania materiału. Zastosowane podejście udowodniło, że przy odpowiedniej redukcji wymiarowości danych można uzyskać wysoki wynik klasyfikacji oraz diagnostyki na danych pozyskiwanych przy użyciu rożnych metod, pochodzących z różnych ośrodków medycznych. Modele spłotowych klasyfikatorów neuronowych zbudowanych na podstawie binarnych obrazów uzyskanych po segmentacji jąder komórkowych wykazały się znacznie wyższym stopniem uogólnienia, przez to są mniej wrażliwe na przeuczenie niż modele trenowane na obrazach w przestrzeni RGB oraz po normalizacji H&E. Fuzja cech głębokich oraz manualnych przyczynia się do ogólnej poprawy wyników klasyfikacji pojedynczych obrazów o 3 %. Cechy głębokie pozyskane z jednej sieci klasyfikacyjnej nieznacznie ustępują zespołowi generatorów cech głębokich. Uzyskane wyniki wskazują, iż warto w przyszłości zbadać dogłębniej temat zespołów generatorów cech głębokich. Lista najważniejszych zadań to:

Opracowanie kompleksowej strategii poprawy dokładności klasyfikacji obrazów cytolo-

#### ROZDZIAŁ 5. PODSUMOWANIE

gicznych i histopatologicznych pochodzących z różnych laboratoriów medycznych.

• Opracowanie metody segmentacji instancji jąder komórkowych na obrazach cytologicznych oraz histopatologicznych z wykorzystaniem głębokiej sieci neuronowej U-Net oraz metody wododziałowej.

• Opracowanie metody fuzji wyuczonych cech głębokich oraz cech eksperckich/manualnych na potrzeby klasyfikacji nowotworu piersi z wykorzystaniem obrazów cytologicznych oraz histopatologicznych.

• Opracowanie metody ekstrakcji cech eksperckich/manualnych jąder komórkowych z obrazów cytologicznych oraz histopatologicznych.

• Wykonanie kompleksowych badań weryfikujących efektywność sieci głębokich o różnych architekturach w zadaniu generacji cech głębokich.

• Wykonanie kompleksowych badań weryfikujących skuteczność różnych metod redukcji wymiarowości dla problemu klasyfikacji nowotworu piersi z wykorzystaniem obrazów cy-tologicznych oraz histopatologicznych.

• Opracowanie nowej metody selekcji cech dla zbiorów składających się z fuzji cech manualnych i głębokich.

### 5.3 Dalsze prace

Zagadnienie semantycznej segmentacji obrazów medycznych, pomimo wysokiej skuteczności sieci CNN, wymaga dalszych dopracowań. Do poprawy wyników segmentacji ważny wydaje się problem regulacji kształtów jąder komórkowych w zakresie wyznaczania granic obszarów w miejscach nakładania się obiektów, jak również najlepsza normalizacja danych wejściowych. Przyszłe prace powinny się również skupić na zagadnieniach takich jak:

• Konstruowanie nowych cech na bazie wiedzy dziedzinowej oraz z wykorzystaniem metod heurystycznych.

• Wydobywanie nowych cech obiektów ze sztucznych sieci neuronowych.

• Pełna automatyzacja segmentacji i klasyfikacji obrazów cytologicznych.

• Dokładniejsze zbadanie zjawiska poprawy zdolności uogólniających sieci trenowanych na obrazach po segmentacji.

• Pozyskanie nowych danych w celu zbadania skuteczności klasyfikacji zaproponowanego systemu.

• Zastosowanie połączonych szeregowo modeli U-Net w celu poprawy segmentacji jąder komórkowych.

• Opracowanie nowych metod standaryzacji obrazów cytologicznych oraz histopatologicznych.

• Rozszerzenie zbioru danych z obrazami cytologicznymi i histopatologicznymi o obrazy pochodzące z innych laboratoriów medycznych/ośrodków medycznych.

## Spis rysunków

1.1	Schemat proponowanego podejścia	22
1.2	Schemat pozyskiwania cech głębokich	23
1.3	Rozkład empiryczny oraz dopasowany rozkład Gamma	24
2.1	Opis preparatu w maksymalnym powiększeniu - przypadek łagodny	26
2.2	Maksymalne powiększenie fragmentu obszaru wirtualnego slajdu $\ .\ .\ .$	28
2.3	Kadrowanie obrazu do rozmiaru 700 x 460	30
2.4	Schemat przygotowania baz danych z obrazami	31
2.5	Efekt rozplotu obrazu metodą H&E z użyciem w oprogramowania Fiji $\left[83\right]$	32
3.1	Różnice pomiędzy detekcją, segmentją semantyczną oraz segmentacją in-	
	stancji	35
3.2	Schemat sieci U-Net na podstawie artykułu [75]	36
3.3	Hybrydowa metoda segmentacji wododziałowej z użyciem wstępnej seg-	
	mentacji za pomocą sieci CNN oraz wykrywanie markerów wododziałów	
	z użyciem drugiej sieci CNN	38
3.4	Etap przejścia z obiektu rzeczywistego na ROI oraz maskę binarną	38
3.5	Graficzna interpretacja DH	40
3.6	Przebieg uczenia modelu U-Net	42
3.7	Schemat tworzenia krawędzi jąder komórkowych obrazów w zbiorze uczącym	43
3.8	Przebieg uczenia modelu U-Net dla punktów centralnych $\ . \ . \ . \ .$	44
3.9	Schemat segmentacji jąder komórkowych	44
3.10	Przykładowe wyniki segmentacji	45
3.11	Rozkłady wartości oraz statystyki	47
3.12	Przykładowy obraz o niskim wyniku ewaluacji segmentacji $\ .\ .\ .\ .$	47
3.13	Przykładowy obraz o przeciętnym wyniku ewaluacji segmentacji $\ .\ .\ .$	48
3.14	Rozkłady liczebności cech w grupach	48
3.15	Wykresy pudełkowe wyników segmentacji do ośrodka pochodzenia próbek $% \mathcal{A}$ .	49
3.16	Dopasowanie wyników segmentacji do ośrodka pochodzenia próbek $\ .\ .\ .$	53
3.17	Dopasowanie wyników segmentacji do utworzonych zbiorów	54
3.18	Dopasowanie wyników segmentacji do przypadków złośliwych i łagodnych .	55

3.19	Dopasowanie wyników metryk segmentacji do pochodzenia ze zbiorów (a,	
	b) oraz przypadków nowotworów (c, d) $\hdots$	56
4.1	Ekstrakcja cech	57
4.2	Schemat działu sztucznej inteligencji [10]	61
4.3	Przykładowa sieć CNN	63
4.4	Sieć AlexNet	63
4.5	Sieć VGG16	64
4.6	Moduł incepcyjny	65
4.7	Generatory cech głębokich	67
4.8	Selekcja cech - metody opakowane (wrappery)	69
4.9	Selekcja cech - metody filtrujące	70
4.10	Selekcja cech - metody wbudowane	71
4.11	Przykładowe rozkłady empiryczne	73
4.12	Schemat głównych eksperymentów	77
4.13	Schemat eksperymentów na sieciach CNN	79
4.14	Przebieg uczenia modelu Res Net50 dla zbioru uczącego Break His $+~{\rm GZG}~$ .	80
4.15	Przebieg uczenia modelu Resnet 50 dla danych uczących ze zbioru SzUZG $% \left( {{\rm SzUZG}} \right)$ .	81
4.16	Przebieg uczenia modelu Resnet 152 V2 dla zbioru uczącego Break His $+~{\rm GZG}$	82
4.17	Przebieg uczenia modelu Resnet 152 V2 dla danych uczących ze zbioru SzUZG	82
4.18	Przebieg uczenia modelu VGG16 dla zbioru uczącego Break His $+~{\rm GZG}~$	83
4.19	Przebieg uczenia modelu VGG16 dla danych uczących ze zbioru SzUZG	83
4.20	Przebieg uczenia modelu VGG19 dla zbioru uczącego Break His $+~{\rm GZG}~$	84
4.21	Przebieg uczenia modelu VGG19 dla danych uczących ze zbioru SzUZG	84
4.22	Przebieg uczenia modelu DenseNet121 dla danych uczących ze zbioru Bre-	
	$akHis + GZG \dots \dots$	85
4.23	Przebieg uczenia modelu DenseNet 121 dla danych uczących ze zbioru SzUZG	85
4.24	Przebieg uczenia modelu X ception dla zbioru uczącego Break His $+~{\rm GZG}~$ .	86
4.25	Przebieg uczenia modelu X ception dla danych uczących ze zbioru SzUZG $% \left( {{\rm Sz}} \right)$ .	87
4.26	$\label{eq:przebieg} Przebieg uczenia modelu Inception ResNetV2 dla danych uczących ze zbioru$	
	BreakHis + GZG	88
4.27	Przebieg uczenia modelu Inception ResNetV2 dla danych uczącego ze zbioru $$	
	SzUZG	88
4.28	Schemat eksperymentów na klasyfikatorach nadrzędnych $\ . \ . \ . \ .$ .	93
4.29	Wykres czułości i precyzji dla najlepszych modeli trenowanych na zbiorze	
	BreakHis+ GZG, natomiast testowanych na SzUZG	109
4.30	Wykres czułości i precyzji dla najlepszych modeli trenowanych na zbiorze	
	SzUZG, natomiast testowanych na Break Hi s $+$ GZG $\ $	110

4.31	Wykres czułości i precyzji dla najlepszych modeli trenowanych na zbiorze
	SzUZG, natomiast testowanych na BreakHis
C.1	Przebieg uczenia modelu InceptionResNetV2 na danych ze zbioru BreakHis
	$+ GZG \dots \dots$
C.2	Przebieg uczenia modelu Inception ResNetV2 na danych ze zbioru SzUZG $% = 1.012$ . $1.012$

## Spis tabel

4.1	Lista wyodrębnionych cech jąder komórkowych	60
4.2	Macierz pomyłek	77
4.3	Podsumowanie wyników klasyfikacji obrazów w przestrzeni RGB	89
4.4	Podsumowanie wyników klasyfikacji obrazów po rozplocie	90
4.5	Podsumowanie wyników klasyfikacji obrazów po segmentacji	91
4.6	Regresja z regularyzacją Elastycznej siatki oraz walidacją przy użyciu kry-	
	terium informacyjnym Bayesa, uczona na zbiorze BreakHis + GZG, nato-	
	miast testowana na zbiorze SzUZG	94
4.7	Regresja z regularyzacją Elastycznej siatki oraz walidacją przy użyciu kry-	
	terium informacyjnym Bayesa, uczona na zbiorze SzUZG, natomiast testo-	
	wana na zbiorze BreakHis + GZG	95
4.8	DT, uczone na zbiorze Break His $+$ GZG, natomiast testowane na zbiorze	
	SzUZG	96
4.9	DT, uczone na zbiorze SzUZG, natomiast testowane na zbiorze BreakHis	
	$+ GZG \dots \dots$	96
4.10	LL, uczony na zbiorze BreakHis + GZG, natomiast testowany na zbiorze	
	SzUZG	98
4.11	LL, uczony na zbiorze SzUZG, natomiast testowany na zbiorze BreakHis	
	$+ GZG \dots \dots$	98
4.12	DW, uczone na zbiorze BreakHis + GZG, natomiast testowane na zbiorze	
	SzUZG	99
4.13	DW, uczone na zbiorze SzUZG, natomiast testowane na zbiorze BreakHis	
	$+ GZG \dots \dots$	100
4.14	NB uczony na zbiorze BreakHis + GZG, natomiast testowany na zbiorze	
	SzUZG	101
4.15	NB, uczony na zbiorze SzUZG, natomiast testowane na zbiorze BreakHis	
	+ GZG	101
4.16	Klasyfikator k-NN, uczony na zbiorze BreakHis + GZG, natomiast testo-	
	wany na zbiorze SzUZG	102

4.17	Klasyfikator k-NN, trenowany na zbiorze SzUZG, natomiast testowana na
	zbiorze BreakHis + GZG
4.18	SVM z radialną funkcją aktywacji, uczona na zbiorze Break His $+$ GZG,
	natomiast testowana na zbiorze SzUZG
4.19	SVM z radialną funkcją aktywacji, uczona na zbiorze SzUZG, natomiast
	testowana na zbiorze Break Hi s $+$ GZG $\hfill .$
4.20	${\rm KN}$ z pojedynczym neuronem z aktywacją Tan H, uczony na zbiorze z Bre-
	ak His $+$ GZG, natomiast testowany na zbiorze SzUZG $\ \ldots \ \ldots \ \ldots \ \ldots \ 106$
4.21	${\rm KN}$ z pojedynczym neuronem z aktywacją Tan H, uczony na zbiorze SzUZG,
	natomiast testowana na zbiorze Break Hi s $+$ GZG $\hfill .$ 107
4.22	Podsumowanie najlepszych klasyfikatorów na podstawie dokładności uzy-
	skanych dla danych testowych z podziałem na zestawy cech (Zapis Break His $$
	+ GZG skrócono w tabeli do BreakHis)
C.1	Wyniki klasyfikacji dla danych testowych dla modeli trenowanych na zbio-
	rze Break His +GZG po redukcji wymiarowości zestawu cech metodą $\rm PCA$ . 139
C.2	Wyniki klasyfikacji dla danych testowych dla modeli trenowanych na zbio-
	rze SzUZG po redukcji wymiarowości zestawu cech metodą PCA 140
C.3	Wyniki klasyfikacji dla danych testowych dla modeli trenowanych na zbio-
	rze Break His $+$ GZG po redukcji wymiarowości zestawu cech przy użyciu
	analizy VIF, rozkładów oraz wartości odsatających $\hfill \ldots \ldots \ldots \ldots \ldots 147$
C.4	Wyniki klasyfikacji dla danych testowych dla modeli trenowanych na zbio-
	rze SzUZG po redukcji wymiarowości zestawu cech przy użyciu analizy
	VIF, rozkładów oraz wartości odsatających

### Bibliografia

- H. Ahmady Phoulady, D. Goldgof, L. Hall, P. Mouton. A new approach to detect and segment overlapping cells in multi-layer cervical cell volume images. *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, strony 201–204, 2016.
- [2] H. Akaike. Information Theory and an Extension of the Maximum Likelihood Principle, strony 199–213. Springer, Nowy Jork, 1998.
- [3] C. Albon. Uczenie maszynowe w Pythonie. Helion, Gliwice, 2019.
- [4] M. Z. Alom, C. Yakopcic, T. M. Taha, V. K. Asari. Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net). *IEEE National Aerospace and Electronics Conference (NAECON)*, strony 228–233, 2018.
- [5] P. Anishiya, M. Sasikala. Segmentation and localization of epithelial cells in the histopathological images of stomach adenocarcinoma. 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), strony 486–489, 2016.
- [6] S. Arooj, Atta-Ur-Rahman, M. Zubair, M. F. Khan, K. Alissa, M. A. Khan, A. Mosavi. Breast cancer detection and classification empowered with transfer learning. *Frontiers in Public Health*, 10, 2022.
- [7] W. Bulten, G. Litjens. Unsupervised prostate cancer detection on h&e using convolutional adversarial autoencoders. *Medical Imaging with Deep Learning*, 2018.
- [8] D. Castelvecchi. Can we open the black box of ai? Nature, 538:20–23, 10 2016.
- [9] S. S. Chennamsetty, M. Safwan, V. Alex. Classification of breast cancer histology image using ensemble of pre-trained neural networks. A. Campilho, F. Karray, B. ter Haar Romeny, redaktorzy, *Image Analysis and Recognition*, strony 804–811, Cham, 2018. Springer International Publishing.
- [10] F. Chollet. Deep Learning with Python. Manning Publications Co., Nowy Jork, 2017.

- [11] F. Chollet. Xception: Deep learning with depthwise separable convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), strony 1800– 1807, Los Alamitos, 2017. IEEE Computer Society.
- [12] Y. Cui, J. Hu. Self-adjusting nuclei segmentation (SANS) of hematoxylin-eosin stained histopathological breast cancer images. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, strony 956–963, 2016.
- [13] Y. Cui, G. Zhang, Z. Liu, Z. Xiong, J. Hu. A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images. *Medical & Biological Engineering & Computing*, 57(9):2027–2043, 2019.
- [14] F. R. Dee, A. Donnelly, S. Radio, T. Leaven, M. S. Zaleski, C. Kreiter. Utility of 2-D and 3-D virtual microscopy in cervical cytology education and testing. Acta Cytologica, 51(4):523–529, 2007.
- [15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei. Imagenet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, strony 248–255, 2009.
- [16] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [17] A. D. Donnelly, M. S. Mukherjee, E. R. Lyden, J. A. Bridge, S. M. Lele, N. Wright, M. F. McGaughey, A. M. Culberson, A. J. Horn, W. R. Wedel, S. J. Radio. Optimal z-axis scanning parameters for gynecologic cytology specimens. *Journal of Pathology Informatics*, 4(1):38, 2013.
- [18] T. Evans, C. O. Retzlaff, C. Geißler, M. Kargl, M. Plass, H. Müller, T.-R. Kiehl, N. Zerbe, A. Holzinger. The explainability paradox: Challenges for xai in digital pathology. *Future Generation Computer Systems*, 133:281–296, 2022.
- [19] A. Evered, N. Dudding. Accuracy and perceptions of virtual microscopy compared with glass slide microscopy in cervical cytology. *Cytopathology*, 22(2):82–87, 2011.
- [20] H. Fatakdawala, J. Xu, A. Basavanhally, G. Bhanot, S. Ganesan, M. Feldman, J. E. Tomaszewski, A. Madabhushi. Expectation-maximization-driven geodesic active contour with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Transactions on Biomedical Engineering*, 57(7):1676–1689, 2010.
- [21] I. Fondón, A. Sarmiento, A. I. García, M. Silvestre, C. Eloy, A. Polónia, P. Aguiar. Automatic classification of tissue malignancy for breast carcinoma diagnosis. *Computers in Biology and Medicine*, 96:41 – 51, 2018.

- [22] K. Fukuma, V. B. S. Prasath, H. Kawanaka, B. J. Aronow, H. Takase. A study on feature extraction and disease stage classification for glioma pathology images. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, strona 2150–2156. IEEE Press, 2016.
- [23] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.
- [24] M. Gagnon, S. Inhorn, J. Hancock, B. Keller, D. Carpenter, T. Merlin, T. Hearn, P. Thompson, R. Whalen. Comparison of cytology proficiency testing. *Acta Cytologica*, 48(6):788–794, 2004.
- [25] M. M. Galloway. Texture analysis using gray level run lengths. Computer Graphics and Image Processing, 4(2):172–179, 1975.
- [26] A. Géron. Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow. Helion, Gliwice, 2020.
- [27] R. J. Gonzalez, R. E. Woods. Digital Image Processing, 4th edition. Pearson, Nowy Jork, 2017.
- [28] D. Hall, C. McCool, F. Dayoub, N. Sunderhauf, B. Upcroft. Evaluation of features for leaf classification in challenging conditions. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, strony 797–804, 2015.
- [29] M. G. Hanna, S. E. Monaco, J. Cuda, J. Xing, I. Ahmed, L. Pantanowitz. Comparison of glass slides and various digital-slide modalities for cytopathology screening and interpretation. *Cancer Cytopathology*, 125(9):701–709, 2017.
- [30] R. Haralick, K. Shanmugam, I. Dinstein. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, 3(6):610–621, 1973.
- [31] T. Hayakawa, V. B. S. Prasath, H. Kawanaka, B. J. Aronow, S. Tsuruoka. Computational nuclei segmentation methods in digital pathology: A survey. Archives of Computational Methods in Engineering, 28(1):1–13, 2021.
- [32] K. He, G. Gkioxari, P. Dollár, R. Girshick. Mask r-cnn. IEEE International Conference on Computer Vision (ICCV), strony 2980–2988, 2017.
- [33] K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, strony 770–778, 2016.

- [34] K. He, X. Zhang, S. Ren, J. Sun. Identity mappings in deep residual networks.
  B. Leibe, J. Matas, N. Sebe, M. Welling, redaktorzy, *Computer Vision ECCV 2016*, strony 630–645, Cham, 2016. Springer International Publishing.
- [35] A. E. Hoerl. Application of ridge analysis to regression problems. Chemical Engineering Progress, 58(3):54–59, 1962.
- [36] L. Hou, K. Singh, D. Samaras, T. M. Kurc, Y. Gao, R. J. Seidman, J. H. Saltz. Automatic histopathology image analysis with CNNs. *New York Scientific Data Summit (NYSDS)*, strony 1–6, 2016.
- [37] G. Huang, Z. Liu, L. V. D. Maaten, K. Q. Weinberger. Densely connected convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), strony 2261–2269, Los Alamitos, 2017. IEEE Computer Society.
- [38] M. Ilse, J. Tomczak, M. Welling. Attention-based deep multiple instance learning. J. Dy, A. Krause, redaktorzy, *Proceedings of the 35th International Conference on Machine Learning*, wolumen 80 serii *Proceedings of Machine Learning Research*, strony 2127–2136. PMLR, 2018.
- [39] P. Jaccard. The distribution of the flora in the alpine zone.1. New Phytologist, 11(2):37–50, 1912.
- [40] J. Jantzen, J. Norup, G. Dounias, B. Bjerregaard. Pap-smear benchmark data for pattern classification. *Nature inspired Smart Information Systems (NiSIS)*, strony 1–9, 2005.
- [41] J. Jassem, M. Krzakowski. Breast cancer. Oncology in Clinical Practice, 14(4):171– 215, 2018.
- [42] L. Jeleń, T. Fevens, A. Krzyżak. Classification of breast cancer malignancy using cytological images of fine needle aspiration biopsies. *International Journal of Applied Mathematics and Computer Science*, 18(1):75–83, 2008.
- [43] H. F. Kaiser. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20(1):141–151, 1960.
- [44] M. N. Kashif, S. E. Ahmed Raza, K. Sirinukunwattana, M. Arif, N. Rajpoot. Handcrafted features with convolutional neural networks for detection of tumor cells in histology images. *IEEE 13th International Symposium on Biomedical Imaging* (ISBI), strony 1029–1032, 2016.
- [45] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, R. Monczak. Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images. *Compu*ters in Biology and Medicine, 43(10):1563–1572, 2013.

- [46] M. Kowal, M. Skobel, A. Gramacki, J. Korbicz. Breast cancer nuclei segmentation and classification based on a deep learning approach. *International Journal of Applied Mathematics and Computer Science*, 31(1):85–106, 2021.
- [47] M. Kowal, M. Skobel, N. Nowicki. The feature selection problem in computer-assisted cytology. International Journal of Applied Mathematics and Computer Science, 28:759–770, 12 2018.
- [48] M. Kowal, M. Żejmo, M. Skobel, J. Korbicz, R. Monczak. Cell nuclei segmentation in cytological images using convolutional neural network and seeded watershed algorithm. *Journal of Digital Imaging*, 33(1):231–242, 2020.
- [49] A. Krizhevsky, I. Sutskever, G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, strona 1097–1105, Red Hook, 2012. Curran Associates Inc.
- [50] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, A. Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, 36(7):1550–1560, 2017.
- [51] V. Kumar, B. K. Mishra, M. Mazzara, D. N. H. Thanh, A. Verma. Prediction of malignant and benign breast cancer: A data mining approach in healthcare applications. S. Borah, V. Emilia Balas, Z. Polkowski, redaktorzy, *Advances in Data Science and Management*, strony 435–442, Singapore, 2020. Springer Singapore.
- [52] S. Kwok. Multiclass classification of breast cancer in whole-slide images. A. Campilho, F. Karray, B. ter Haar Romeny, redaktorzy, *Image Analysis and Recognition*, strony 931–940, Cham, 2018. Springer International Publishing.
- [53] A. Lagree, M. Mohebpour, N. Meti, K. Saednia, F.-I. Lu, E. Slodkowska, S. Gandhi, E. Rakovitch, A. Shenfield, A. Sadeghi-Naini, W. T. Tran. A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks. *Scientific Reports*, 11(1):8025, Apr 2021.
- [54] Y. LeCun, Y. Bengio. Convolutional networks for images, speech, and time series.
   M. A. Arbib, redaktor, *Handbook of Brain Theory and Neural Networks*, strony 33–61. MIT Press, 1995.
- [55] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

- [56] M. Linkert, C. T. Rueden, C. Allan, J.-M. Burel, W. Moore, A. Patterson, B. Loranger, J. Moore, C. Neves, D. MacDonald, A. Tarkowska, C. Sticco, E. Hill, M. Rossner, K. W. Eliceiri, J. R. Swedlow. Metadata matters: access to image data in the real world. *Journal of Cell Biology*, 189(5):777–782, 2010.
- [57] C. Lu, M. Mahmood, N. Jha, M. Mandal. A robust automatic nuclei segmentation technique for quantitative histopathological image analysis. Anal Quant Cytopathol Histpathol, 34(6):296–308, 2012.
- [58] L. Maier-Hein, i inni. Why rankings of biomedical image analysis competitions should be interpreted with care. *Nature Communications*, 9(1):5217, 2018.
- [59] O. L. Mangasarian, W. N. Street, W. H. Wolberg. Breast cancer diagnosis and prognosis via linear programming. *Operations Research*, 43(4):570–577, 1995.
- [60] M. Minsky, S. Papert. Perceptrons: An Introduction to Computational Geometry. MIT Press, Cambridge, 1969.
- [61] M. A. Naji, S. E. Filali, K. Aarika, E. H. Benlahmar, R. A. Abdelouhahid, O. Debauche. Machine learning algorithms for breast cancer prediction and diagnosis. *Procedia Computer Science*, 191:487–492, 2021.
- [62] P. Naylor, M. Laé, F. Reyal, T. Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 38(2):448–459, 2019.
- [63] M. Neghina, C. Rasche, M. Ciuc, A. Sultana, C. Tiganesteanu. Automatic detection of cervical cells in pap-smear images using polar transform and k-means segmentation. Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), strony 1–6, 2016.
- [64] M. K. K. Niazi, K. Yao, D. L. Zynger, S. K. Clinton, J. Chen, M. Koyutürk, T. La-Framboise, M. Gurcan. Visually meaningful histopathological features for automatic grading of prostate cancer. *IEEE Journal of Biomedical and Health Informatics*, 21(4):1027–1038, 2017.
- [65] N. Otsu. A threshold selection method from gray-level histograms. IEEE Trans. Systems, Man, and Cybernetics, 9(1):62–66, 1979.
- [66] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems 32*, strony 8024–8035. Curran Associates, Inc., 2019.

- [67] K. Pearson. Note on Regression and Inheritance in the Case of Two Parents. Proceedings of the Royal Society of London Series I, 58:240–242, 1895.
- [68] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, A. Tozeren. Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC Medical Imaging*, 6(1):14, 2006.
- [69] H. A. Phoulady, M. Zhou, D. B. Goldgof, L. O. Hall, P. R. Mouton. Automatic quantification and classification of cervical cancer via adaptive nucleus shape modeling. *IEEE International Conference on Image Processing (ICIP)*, strony 2658–2662, 2016.
- S. Pilleron, E. Soto-Perez-de Celis, J. Vignat, J. Ferlay, I. Soerjomataram, F. Bray,
   D. Sarfati. Estimated global cancer incidence in the oldest adults in 2018 and
   projections to 2050. *International Journal of Cancer*, 148(3):601–608, 2021.
- [71] X. Qi, F. Xing, D. J. Foran, L. Yang. Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set. *IEEE Trans Biomed Eng*, 59(3):754–765, 2011.
- [72] S. Ragothaman, S. Narasimhan, M. G. Basavaraj, R. Dewar. Unsupervised segmentation of cervical cell images using gaussian mixture model. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, strony 1374–1379, 2016.
- [73] S. Rajaganesan, R. Kumar, V. Rao, T. Pai, N. Mittal, A. Sahay, S. Menon, S. Desai. Comparative assessment of digital pathology systems for primary diagnosis. *Journal of Pathology Informatics*, 12(1):25, 2021.
- [74] S. Raschka, V. Mirjalili. Python. Uczenie maszynowe. Wydanie II. Helion, Gliwice, 2019.
- [75] O. Ronneberger, P. Fischer, T. Brox. U-net: Convolutional networks for biomedical image segmentation. N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi, redaktorzy, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, strony 234–241, Cham, 2015. Springer International Publishing.
- [76] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65 6:386–408, 1958.
- [77] A. Ruifrok, D. Johnston. Quantification of histochemical staining by color deconvolution. Analytical and Quantitative Cytology and Histology, 23, 2001.

- [78] D. E. Rumelhart, G. E. Hinton, R. J. Williams. Learning representations by backpropagating errors. *Nature*, 323(6088):533–536, 1986.
- [79] R. Saha, M. Bajger, G. Lee. Spatial shape constrained fuzzy c-means (FCM) clustering for nucleus segmentation in pap smear images. *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, strony 1–8, 2016.
- [80] S. Sakib, N. Yasmin, A. K. Tanzeem, F. Shorna, K. Md. Hasib, S. B. Alam. Breast cancer detection and classification: A comparative analysis using machine learning algorithms. V. Bindhu, J. M. R. S. Tavares, K.-L. Du, redaktorzy, *Proceedings* of Third International Conference on Communication, Computing and Electronics Systems, strony 703–717, Singapur, 2022. Springer Singapore.
- [81] M. Salvi, F. Molinari. Multi-tissue and multi-scale approach for nuclei segmentation in h&e stained images. *BioMedical Engineering OnLine*, 17, 2018.
- [82] SAS Institute Inc. JMP®, Version 16.0.0. 1989-2022.
- [83] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, 2012.
- [84] G. Schwarz. Estimating the Dimension of a Model. The Annals of Statistics, 6(2):461 - 464, 1978.
- [85] P. Shi, J. Zhong, R. Huang, J.-J. Lin. Automated quantitative image analysis of hematoxylin-eosin staining slides in lymphoma based on hierarchical kmeans clustering. 2016 8th International Conference on Information Technology in Medicine and Education (ITME), strony 99–104, 2016.
- [86] J. Shu, H. Fu, G. Qiu, P. Kaye, M. Ilyas. Segmenting overlapping cell nuclei in digital histopathology images. 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), strony 5445–5448, 2013.
- [87] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
- [88] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, N. M. Rajpoot. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging*, 35(5):1196–1206, 2016.

- [89] M. Skobel, M. Kowal, J. Korbicz. Breast cancer computer-aided diagnosis system using k-NN algorithm based on hausdorff distance. J. Korbicz, R. Maniewski, K. Patan, M. Kowal, redaktorzy, *Current Trends in Biomedical Engineering and Bioimages Analysis*, strony 179–188, Cham, 2020. Springer International Publishing.
- [90] Y. S. Solanki, P. Chakrabarti, M. Jasinski, Z. Leonowicz, V. Bolshev, A. Vinogradov, E. Jasinska, R. Gono, M. Nami. A hybrid supervised machine learning classifier system for breast cancer prognosis using feature selection and data imbalance handling approaches. *Electronics*, 10(6), 2021.
- [91] T. Sørensen. A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons. Biologiske skrifter. I kommission hos E. Munksgaard, 1948.
- [92] S. Sornapudi, G. T. Brown, Z. Xue, R. Long, L. Allen, S. Antani. Comparing deep learning models for multi-cell classification in liquid- based cervical cytology image. *AMIA Annu. Symp. Proc.*, 2019:820–827, 2019.
- [93] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, L. Heutte. Deep features for breast cancer histopathological image classification. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, strony 1868–1873, 2017.
- [94] F. A. Spanhol, L. S. Oliveira, C. Petitjean, L. Heutte. A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462, 2016.
- [95] C. L. Srinidhi, O. Ciga, A. L. Martel. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, 67:1–30, 2021.
- [96] W. N. Street, W. H. Wolberg, O. L. Mangasarian. Nuclear feature extraction for breast tumor diagnosis. R. S. Acharya, D. B. Goldgof, redaktorzy, *Biomedical Image Processing and Biomedical Visualization*, wolumen 1905, strony 861 – 870. International Society for Optics and Photonics, SPIE, 1993.
- [97] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3):209–249, 2021.
- [98] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. 2016.

- [99] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, i in. Going deeper with convolutions. arxiv 2014. arXiv preprint arXiv:1409.4842, 1409, 2014.
- [100] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.
- [101] X. Tang. Texture information in run-length matrices. IEEE Transactions on Image Processing, 7(11):1602–1609, 1998.
- [102] A. Tareef, Y. Song, W. Cai, Y. Wang, D. D. Feng, M. Chen. Automatic nuclei and cytoplasm segmentation of leukocytes with color and texture-based image enhancement. *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, strony 935–938, 2016.
- [103] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. arXiv e-prints, abs/1605.02688, 2016.
- [104] R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1):267–288, 1996.
- [105] S. Tripathi, S. K. Singh. Ensembling handcrafted features with deep features: an analytical study for classification of routine colon cancer histopathological nuclei images. *Multimedia Tools and Applications*, 79(47):34931–34954, 2020.
- [106] R. Uppada, S. K. Rao, K. S. Prasad. Supervised classification of breast cancer malignancy using integrated modified marker controlled watershed approach. strony 584–589, 2017.
- [107] S. L. Van Es. Digital pathology: semper ad meliora. *Pathology*, 51(1):1–10, 2018.
- [108] M. Veta, A. Huisman, M. Viergever, P. van Diest, J. Pluim. Marker-controlled watershed segmentation of nuclei in H&E stained breast cancer biopsy images. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, strony 618– 621, 2011.
- [109] M. Veta, P. J. van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, J. P. W. Pluim. Automatic nuclei segmentation in H&E stained breast cancer histopathology images. *PLOS ONE*, 8(7), 07 2013.
- [110] H. Wang, A. C. Roa, A. N. Basavanhally, H. L. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, A. Madabhushi. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, 1(3):034003, 2014.

- [111] M. Wang, X. He, Y. Chang, G. Sun, L. Thabane. A sensitivity and specificity comparison of fine needle aspiration cytology and core needle biopsy in evaluation of suspicious breast lesions: A systematic review and meta-analysis. *The Breast*, 31:157 – 166, 2017.
- [112] X. Wang, H. Chen, C. Gan, H. Lin, Q. Dou, E. Tsougenis, Q. Huang, M. Cai, P.-A. Heng. Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Transactions on Cybernetics*, 50(9):3950–3962, 2020.
- [113] K. Y. Win, S. Choomchuay. Automated segmentation of cell nuclei in cytology pleural fluid images using otsu thresholding. *International Conference on Digital Arts, Media and Technology (ICDAMT)*, strony 14–18, 2017.
- [114] W. H. Wolberg, W. Street, O. Mangasarian. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer Letters*, 77(2):163–171, 1994.
- [115] L. Xie, J. Qi, L. Pan, S. Wali. Integrating deep convolutional neural networks with marker-controlled watershed for overlapping nuclei segmentation in histopathology images. *Neurocomputing*, 376:166 – 179, 2020.
- [116] X. Yang, H. Li, X. Zhou. Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy. *IEEE Trans*actions on Circuits and Systems I: Regular Papers, 53(11):2405–2414, 2006.
- [117] C. G. Yedjou, S. S. Tchounwou, R. A. Aló, R. Elhag, B. Mochona, L. Latinwo. Application of machine learning algorithms in breast cancer diagnosis and classification. *Int J Sci Acad Res*, 2(1):3081–3086, 2021.
- [118] N. Zarei, A. Bakhtiari, P. Gallagher, M. Keys, C. MacAulay. Automated prostate glandular and nuclei detection using hyperspectral imaging. *IEEE 14th Internatio*nal Symposium on Biomedical Imaging (ISBI 2017), strony 1028–1031, 2017.
- [119] Zespół Tensorflow. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [120] L. Zhang, H. Kong, C. T. Chin, S. Liu, Z. Chen, T. Wang, S. Chen. Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts. *Computerized Medical Imaging and Graphics*, 38(5):369–380, 2014.
- [121] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu, J. Yao. Deeppap: Deep convolutional networks for cervical cell classification. *IEEE Journal of Biomedical and Health Informatics*, 21(6):1633–1643, 2017.

- [122] R. Zhang, L. Du, Q. Xiao, J. Liu. Comparison of backbones for semantic segmentation network. *Journal of Physics: Conference Series*, 1544(1):012196, 2020.
- [123] Z. Zhang, C. Wu, S. Coleman, D. Kerr. Dense-inception u-net for medical image segmentation. *Computer Methods and Programs in Biomedicine*, 192:105395, 2020.
- [124] Z. Zhao, D. Zhao, S. Yang, L. Xu. Image-based malware classification method with the alexnet convolutional neural network model. *Security and Communication Networks*, 2023:6390023, 2023.
- [125] H. Zou, T. Hastie. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 67(2):301–320, 2005.

## Dodatek A

## Struktura sieci U-Net do segmentacji

Model: "model"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 464, 704, 3	0)]	0
conv2d (Conv2D)	(None, 232, 352, 32	896)	['input_1[0][0]']
<pre>batch_normalization (BatchNormalization)</pre>	(None, 232, 352, 32	128)	['conv2d[0][0]']
activation (Activation)	(None, 232, 352, 32	0)	['batch_normalization[0][0]']
activation_1 (Activation)	(None, 232, 352, 32	0)	['activation[0][0]']
<pre>separable_conv2d (SeparableConv2D)</pre>	(None, 232, 352, 64	2400)	['activation_1[0][0]']
activation_2 (Activation)	(None, 232, 352, 64	0)	['batch_normalization_1[0][0]']
<pre>separable_conv2d_1 (SeparableConv2D)</pre>	(None, 232, 352, 64	4736)	['activation_2[0][0]']
<pre>batch_normalization_2 (BatchNormalization)</pre>	(None, 232, 352, 64	256)	['separable_conv2d_1[0][0]']
<pre>max_pooling2d (MaxPooling2D)</pre>	(None, 116, 176, 64	0)	['batch_normalization_2[0][0]']
conv2d_1 (Conv2D)	(None, 116, 176, 64	2112)	['activation[0][0]']
add (Add)	(None, 116, 176, 64	0)	['max_pooling2d[0][0]','conv2d_1[0][0]']
activation_3 (Activation)	(None, 116, 176, 64	0)	['add[0][0]']
<pre>separable_conv2d_2 (SeparableConv2D)</pre>	(None, 116, 176, 12	88968)	['activation_3[0][0]']
<pre>batch_normalization_3 (BatchNormalization)</pre>	(None, 116, 176, 12	5128)	['separable_conv2d_2[0][0]']
activation_4 (Activation)	(None, 116, 176, 12	08)	['batch_normalization_3[0][0]']
<pre>separable_conv2d_3 (SeparableConv2D)</pre>	(None, 116, 176, 12	176648)	['activation_4[0][0]']
batch_normalization_4 (BatchNormalization)	(None, 116, 176, 12	5128)	['separable_conv2d_3[0][0]']
<pre>max_pooling2d_1 (MaxPooling2D)</pre>	(None, 58, 88, 128)	0	['batch_normalization_4[0][0]']
conv2d_2 (Conv2D)	(None, 58, 88, 128)	8320	['add[0][0]']
add_1 (Add)	(None, 58, 88, 128)	0	['max_pooling2d_1[0][0]','conv2d_2[0][0]']
activation_5 (Activation)	(None, 58, 88, 128)	0	['add_1[0][0]']
<pre>separable_conv2d_4 (SeparableConv2D)</pre>	(None, 58, 88, 256)	34176	['activation_5[0][0]']
<pre>batch_normalization_5 (BatchNormalization)</pre>	(None, 58, 88, 256)	1024	['separable_conv2d_4[0][0]']
activation_6 (Activation)	(None, 58, 88, 256)	0	['batch_normalization_5[0][0]']
<pre>separable_conv2d_5 (SeparableConv2d)</pre>	(None, 58, 88, 256)	68096	['activation_6[0][0]']
<pre>batch_normalization_6 (BatchNormalization)</pre>	(None, 58, 88, 256)	1024	['separable_conv2d_5[0][0]']
<pre>max_pooling2d_2 (MaxPooling2D)</pre>	(None, 29, 44, 256)	0	['batch_normalization_6[0][0]']
conv2d_3 (Conv2D)	(None, 29, 44, 256)	33024	['add_1[0][0]']
add_2 (Add)	(None, 29, 44, 256)	0	['max_pooling2d_2[0][0]','conv2d_3[0][0]']
activation_7 (Activation)	(None, 29, 44, 256)	0	['add_2[0][0]']
<pre>conv2d_transposeose) (Conv2DTransp)</pre>	(None, 29, 44, 256)	590080	['activation_7[0][0]']
<pre>batch_normalization_7 (BatchNormalization)</pre>	(None, 29, 44, 256)	1024	['conv2d_transpose[0][0]']
activation_8 (Activation)	(None, 29, 44, 256)	0	['batch_normalization_7[0][0]']
<pre>conv2d_transpose_1 (Conv2DTranspose)</pre>	(None, 29, 44, 256)	590080	['activation_8[0][0]']
<pre>batch_normalization_8 (BatchNormalization)</pre>	(None, 29, 44, 256)	1024	['conv2d_transpose_1[0][0]']
up_sampling2d_1 (UpSampling2D)	(None, 58, 88, 256)	0	['add_2[0][0]']
up_sampling2d (UpSampling2D)	(None, 58, 88, 256)	0	['batch_normalization_8[0][0]']
conv2d_4 (Conv2D)	(None, 58, 88, 256)	65792	['up_sampling2d_1[0][0]']
add_3 (Add)	(None, 58, 88, 256)	0	['up_sampling2d[0][0]','conv2d_4[0][0]']
activation_9 (Activation)	(None, 58, 88, 256)	0	['add_3[0][0]']
<pre>conv2d_transpose_2 (Conv2DTranspose)</pre>	(None, 58, 88, 128)	295040	['activation_9[0][0]']
<pre>batch_normalization_9 (BatchNormalization)</pre>	(None, 58, 88, 128)	512	['conv2d_transpose_2[0][0]']
activation_10 (Activation)	(None, 58, 88, 128)	0	['batch_normalization_9[0][0]']
<pre>conv2d_transpose_3 (Conv2DTranspose)</pre>	(None, 58, 88, 128)	147584	['activation_10[0][0]']
<pre>batch_normalization_10 (BatchNormalization)</pre>	(None, 58, 88, 128)	512	['conv2d_transpose_3[0][0]']
up_sampling2d_3 (UpSampling2D)	(None, 116, 176, 25	06)	['add_3[0][0]']
up_sampling2d_2 (UpSampling2D)	(None, 116, 176, 12	08)	['batch_normalization_10[0][0]']
conv2d_5 (Conv2D)	(None, 116, 176, 12	328968)	['up_sampling2d_3[0][0]']
add_4 (Add)	(None, 116, 176, 12	08	['up_sampling2d_2[0][0]','conv2d_5[0][0]']
activation_11 (Activation)	(None, 116, 176, 12	08)	['add_4[0][0]']

<pre>conv2d_transpose_4 (Conv2DTranspose)</pre>	(None,	116,	176,	64	73792)	['activation_11[0][0]']
<pre>batch_normalization_11 (BatchNormalization)</pre>	(None,	116,	176,	64	256)	['conv2d_transpose_4[0][0]']
activation_12 (Activation)	(None,	116,	176,	64	0)	['batch_normalization_11[0][0]']
<pre>conv2d_transpose_5 (Conv2DTranspose)</pre>	(None,	116,	176,	64	36928)	['activation_12[0][0]']
<pre>batch_normalization_12 (BatchNormalization)</pre>	(None,	116,	176,	64	256)	['conv2d_transpose_5[0][0]']
up_sampling2d_5 (UpSampling2D)	(None,	232,	352,	12	08)	['add_4[0][0]']
up_sampling2d_4 (UpSampling2D)	(None,	232,	352,	64	0)	['batch_normalization_12[0][0]']
conv2d_6 (Conv2D)	(None,	232,	352,	64	8256)	['up_sampling2d_5[0][0]']
add_5 (Add)	(None,	232,	352,	64	0)	['up_sampling2d_4[0][0]','conv2d_6[0][0]']
activation_13 (Activation)	(None,	232,	352,	64	0)	['add_5[0][0]']
<pre>conv2d_transpose_6 (Conv2DTranspose)</pre>	(None,	232,	352,	32	18464)	['activation_13[0][0]']
<pre>batch_normalization_13 (BatchNormalization)</pre>	(None,	232,	352,	32	128)	['conv2d_transpose_6[0][0]']
activation_14 (Activation)	(None,	232,	352,	32	0)	['batch_normalization_13[0][0]']
<pre>conv2d_transpose_7 (Conv2DTranspose)</pre>	(None,	232,	352,	32	9248)	['activation_14[0][0]']
<pre>batch_normalization_14 (BatchNormalization)</pre>	(None,	232,	352,	32	128)	['conv2d_transpose_7[0][0]']
up_sampling2d_7 (UpSampling2D)	(None,	464,	704,	64	0)	['add_5[0][0]']
up_sampling2d_6 (UpSampling2D)	(None,	464,	704,	32	0)	['batch_normalization_14[0][0]']
conv2d_7 (Conv2D)	(None,	464,	704,	32	2080)	['up_sampling2d_7[0][0]']
add_6 (Add)	(None,	464,	704,	32	0)	['up_sampling2d_6[0][0]','conv2d_7[0][0]']
conv2d_8 (Conv2D)	(None,	464,	704,	3)	867	['add_6[0][0]']

Total params: 2,058,979 Trainable params: 2,055,203 Non-trainable params: 3,776

## Dodatek B

# Konfiguracja sprzętowa oraz oprogramowanie

Do wykonania badań wykorzystano sprzęt oraz oprogramowanie o poniższej konfiguracji:

- Karta graficzna: Nvidia GeForce RTX 3060 12GB VRAM
- Procesor: Intel i5-10600K, 6 rdzeni, 12 wątków.
- Płyta główna: ASRock Z490M-ITX/ac
- Pamięć RAM: 32 GB
- Dysk: SSD 500 GB
- System: Ubuntu Linux, Windows 10

Do ręcznego zaznaczania jąder komórkowych:

• Tablet graficzny Wacom Cintiq Pro 24

Oprogramowanie:

- Język Python w tym: Anaconda, Keras, Tensorflow, ScikitLearn, Numpy, Matplotlib, OpenCV, Pillow
- JMP Pro [82]

### Dodatek C

### Wyniki dodatkowych badań

#### Uczenie transferowe

Duże modele głębokich sieci neuronowych potrzebują ogromnych zbiorów danych żeby w pełni pokazać swój potencjał klasyfikacyjny. Bardzo często jednak dane, którymi dysponujemy nie dostarczają tak dużych zbiorów informacji. Jedną z metod radzenia sobie z tym problemem, oprócz augmentacji danych stosowanych w zaprezentowanych powyżej eksperymentach, jest metoda zwana uczeniem transferowym (*ang.* transfer learning). Uczenie transferowe zostało podzielone na dwie części. Pierwsza polegała na uczeniu jedynie górnej warstwy sieci, natomiast wszystkie warstwy głównego modelu zostały zamrożone z wagami uzyskanymi na bazie zbioru Imagenet [15]. Druga część eksperymentu to dokładne strojenie modelu (*ang.* fine tuning), które polega na odmrożeniu wszystkich warstw sieci z jednoczesnym zmniejszeniem kroku uczenia.

Niestety po wykonaniu pierwszych eksperymentów okazało się, że uczenie transferowe nie przynosi spodziewanej poprawy wyników względem uczenia sieci od podstaw. Pozytywnym aspektem uczenia transferowego było przyspieszenie procesu uczenia ze względu na wstępnie przygotowane wagi modelu. Wyniki procesu uczenia transferowego umieszczone zostały na rysunkach C.1 oraz C.2

#### Dane po redukcji wymiarowości PCA

Jedną z najbardziej popularnych metod redukcji wymiarowości jest analiza głównych składowych (PCA). Aby dobrze przeprowadzić analizę PCA należy wstępnie zbadać dane. Pierwsza kwestia dotyczy liczby obserwacji, w obu bazach przekracza liczbę 10000. Liczba cech jest niższa niż 500. Te dwa parametry powodują, iż aplikacja JMP Pro [82] sugeruje metodę PCA porównań parami (*ang.* pairwise). Kolejny niezbędny parametr wymagający ustalenia to liczba głównych składowych. Liczba ta jest wybierana w sposób arbitralny



Rysunek C.1: Przebieg uczenia modelu InceptionResNetV2 na danych ze zbioru BreakHis + GZG

jednak istnieją podejścia ułatwiające podjęcie właściwej decyzji. Jedną z najpopularniejszych metod jest kryterium Kaisera [43]. Polega ona na wybraniu tylko tych głównych składowych, których wartości własne przekraczają liczbę 1. Oznacza to sytuację w której badana składowa wnosi do modelu przynajmniej taką sama informację jaką wniosła by pojedyncza zmienna. Program JMP Pro [82] umożliwia analizę PCA z użyciem macierzy korelacji, w związku z tym nie ma potrzeby normalizacji danych wejściowych.

Analiza głównych składowych na zbiorze BreakHis + GZG Pierwsza część analizy dotyczy zbioru BreakHis + GZG. Spośród 251 cech manualnych wybranych analiza PCA wykazała 29 istotnych głównych składowych biorąc pod uwagę kryterium Kaisera. Wybrane główne składowe objaśniają 89,413 % wariancji wszystkich pierwotnych cech. Kolejny etap analizy polegał na wykonaniu PCA dla zestawu cech głębokich. Zastosowanie kryterium Kaisera pozwoliło na uzyskanie 3 głównych składowych, które objaśniają 89,615 % wariancji pozostałych zmiennych. Zatem na bazie 25 cech głębokich metoda PCA wyznaczyła jedynie 3 główne składowe. W tabeli C.1 wykazano jak wypadły się



(c) Dokładność modelu w kolejnych epokach

(d) Strata modelu w kolejnych epokach

Rysunek C.2: Przebieg uczenia modelu InceptionResNetV2 na danych ze zbioru SzUZG

•1 •	11 C1	1 .	1/ 1	111 1	11 1 1	1 1
wvniki	klasvfikacii	na bazie	głownych	składowych	dla danvch	testowych
"J IIIII	masymacji	na bazio	810 mil 901	Shiladowyon	and daily on	000000 9011.

Tabela C.1: Wyniki klasyfikacji dla danych testowych dla modeli trenowanych na zbiorze BreakHis+GZG po redukcji wymiarowości zestawu cech metodą PCA

Klasyfikator	ACC [Manualne]	ACC [Głębokie]	ACC [Fuzja]
RL	0.7303	0.7926	0.7958
DT	0.7228	0.7871	0.7940
LL	0.7642	0.7857	0.7981
DW	0.7569	0.7918	0.7971
NB	0.7258	0.7950	0.7729
k-NN	0.7233	0.7958	0.7739
SVM	0.7571	0.7917	0.7903
KN	0.7342	0.7951	0.7937

Na podstawie uzyskanych wyników można zauważyć, że redukcja wymiarowości metodą PCA nieznacznie poprawiła niemal wszystkie wyniki klasyfikacji względem modeli zbudowanych na zbiorach przed redukcją PCA. Jedyny wyjątek stanowi klasyfikator k-NN, który lepsze rezultaty uzyskiwał dla zbiorów przed redukcją PCA. Zarówno poprawa jak i pogorszenie wyników po zastosowaniu metody PCA są bardzo niewielkie, zatem główną zaletą zastosowania redukcji wymiarowości metodą PCA jest znacznie niższy czas potrzebny do uzyskania wyników, wadą natomiast jest pogorszenie interpretacji wyników ze względu na zastąpienie pierwotnych cech głównymi składowymi.

Analiza głównych składowych na zbiorze SzUZG Eksperyment został przeprowadzony oddzielnie dla cech manualnych oraz cech głębokich, aby umożliwić porównanie wyników z danych z poprzednich podrozdziałów. W wyniku przetwarzania cech manualnych wybrana na podstawie kryterium Kaisera liczba głównych składowych wyniosła 29. Liczba ta objaśnia 89,95% wariancji wszystkich cech. Zatem metoda PCA doprowadziła do redukcji przestrzeni 251 cech do 29. W przypadku cech głębokich liczba głównych składowych które spełniają kryterium wynosi 2. Wybrane główne składowe objaśniają 90.48% wariancji wszystkich cech. W przypadku cech głębokich redukcja wymiarowości nastąpiła z 25 do 2 wymiarów. Redukcja wymiarowości z dużym prawdopodobieństwem nie polepszy rezultatów klasyfikacji na danych uczących, ale poprzez redukcję "klątwy wymiarowości"może przyczynić się do poprawy wyników na danych testowych. Wobec tego przedstawione zostały tu wyniki klasyfikacji dla danych testowych dla klasyfikatorów użytych w podrozdziale 4.8.3.3.

Tabela C.2	: Wyniki klası	yfikacji dla	danych	testowych	dla mod	leli trenov	vanych ne	a zbiorze	SzUZG
po redukcji	i wymiarowość	ci zestawu	cech me	todą PCA					

Klasyfikator	ACC [Manualne]	ACC [Głębokie]	ACC [Fuzja]
RL	0.6590	0.7271	0.7166
DT	0.6469	0.7191	0.7234
LL	0.6596	0.7202	0.7187
DW	0.6612	0.7209	0.7142
NB	0.6563	0.7271	0.7210
k-NN	0.6566	0.7209	0.6996
SVM	0.6563	0.7224	0.7057
KN	0.6580	0.7269	0.7192

Redukcja wymiarowości PCA na zbiorze SzUZG przyczyniła się do nieznacznej ogólnej poprawy modeli opartych o drzewa decyzyjne (tab. C.2). Uwagę również zwraca poprawa wyników klasyfikacji na zestawie cech głębokich. Z przeprowadzonych badań wynika, że dwie główne składowe są wystarczające do zastąpienia 25 cech pozyskanych z 5 różnych sieci neuronowych. Można zatem przypuszczać iż wszystkie sieci neuronowe posiadają zbliżoną charakterystykę uzyskanych cech głębokich. Wniosek ten prowadzi do potrzeby wykonania dodatkowego eksperymentu polegającego na wykorzystaniu jednej sieci neuronowej do uzyskania większej liczby cech i ewentualnej redukcji do cech istotnych. Sukces takiego eksperymentu jednocześnie pozwoliłby ograniczyć złożoność modelu jak również czas potrzebny na jego zbudowanie. Wśród negatywnych skutków analizy PCA należy nadmienić, iż znacznie pogorszyła ona maksymalne wyniki uzyskane dla modeli klasyfikatorów, a co za tym idzie nie przyczyniła się do uzyskania lepszych rozwiązań. Biorąc pod uwagę najlepsze średnie wyniki, to dla zestawu cech manualnych uległ on pogorszeniu, podobnie jak w przypadku klasyfikatorów zbudowanych na fuzji cech. Najciekawsze jest jednak to, że najwyższy średni wynik klasyfikacji dla zestawu cech głębokich pozostał na identycznym poziomie jak przed analizą PCA i osiągnięty został dla tego samego klasyfikatora.

#### Dane po wykonaniu analizy eksploracyjnej cech

W tym badaniu przed wykonaniem klasyfikacji dokonano eksploracji posiadanego zestawu zmiennych. Badanie polegało na weryfikacji występowania związków liniowych pomiędzy zmiennymi. Wykrycie wartości współczynnika inflacji wariancji VIF na poziomie przekraczającym wartość 10 powodowało wykluczenie takiej zmiennej z zestawu zmiennych. Kolejny krok analizy polegał na zbadaniu zmiennych pod kątem skośności rozkładów oraz przy wykryciu takiej skośności podjęcia próby redukcji skośności przy użyciu metod transformacji zmiennych. Ostatni etap polegał na wykryciu oraz eliminacji ze zbioru wszystkich danych odstających.

**Eksploracja danych na zbiorze BreakHis + GZG** Wykonana analiza doprowadziła do redukcji wymiarowości zestawu cech manualnych z 255 do 94. Zestaw cech głębokich zmniejszył się natomiast z 25 do 7. Łączna liczba cech spadła więc z 276 do 101.

Klasyfikator	ACC [Manualne]	ACC [Głębokie]	ACC [Fuzja]
RL	0.6942	0.7833	0.7775
DT	0.6910	0.7669	0.7728
LL	0.7194	0.7750	0.7842
DW	0.7165	0.7799	0.7699
NB	0.7409	0.7962	0.7897
k-NN	0.7046	0.7924	0.7715
SVM	0.7387	0.7886	0.7765
KN	0.7191	0.7831	0.7665

Tabela C.3: Wyniki klasyfikacji dla danych testowych dla modeli trenowanych na zbiorze BreakHis + GZG po redukcji wymiarowości zestawu cech przy użyciu analizy VIF, rozkładów oraz wartości odsatających

**Eksploracja danych na zbiorze SzUZG** W wyniku przeprowadzonej analizy spośród 251 cech głębokich w bazie pozostało 97, przy czym większość z pozostałych cech podlegała transformacji logarytmicznej ze względu na asymetrię rozkładów. Spośród cech głębokich w bazie pozostało jedynie 9 z 25. Łączna liczba cech zmalała po przeprowadzonej analizie z 276 do 106. Wykonane eksperymenty nie przyniosły ogólnej poprawy modeli. Szczególną uwagę może jedynie zwrócić model KN (tab. C.3 oraz tab. C.4), który dzięki zastosowaniu analizy eksploracyjnej uzyskał wyniki dokładności klasyfikacji dla obu zbioTabela C.4: Wyniki klasyfikacji dla danych testowych dla modeli trenowanych na zbiorze SzUZG po redukcji wymiarowości zestawu cech przy użyciu analizy VIF, rozkładów oraz wartości odsatających

Klasyfikator	ACC [Manualne]	ACC [Głębokie]	ACC [Fuzja]
RL	0.7199	0.7112	0.7278
DT	0.6468	0.7094	0.6610
LL	0.6823	0.7071	0.7148
DW	0.6613	0.7115	0.7142
NB	0.6823	0.7190	0.7214
k-NN	0.6444	0.7059	0.6920
SVM	0.7000	0.7146	0.7363
KN	0.7335	0.7118	0.7634

rów danych testowych powyżej 0.7600. Niestety jest to nadal wynik niemal o 3% gorszy od modeli zbudowanych na danych przed wykonaną analizą.