

WOJSKOWA AKADEMIA TECHNICZNA
im. Jarosława Dąbrowskiego



ROZPRAWA DOKTORSKA
„Zespoły sieci głębokich w rozpoznawaniu
wybranych klas obrazów medycznych”

kpt. mgr inż. Fabian GIL

Promotor: prof. dr hab. inż. Stanisław OSOWSKI

Zespoły sieci głębokich w rozpoznawaniu wybranych klas obrazów medycznych.

kpt. mgr inż. Fabian GIL

Słowa kluczowe: rozpoznawanie obrazów, głębokie uczenie, konwolucyjne sieci neuronowe, zespół klasyfikatorów, selekcja cech, czerniak, rak piersi.

Streszczenie

Tematyka pracy koncentruje się na opracowaniu automatycznych metod analizy obrazów medycznych czerniaka oraz raka piersi. Proponowane w pracy rozwiązania opierają się na zastosowaniu różnych neuronowych struktur sieci głębokich CNN oraz klasyfikatorów typu klasycznego (SVM, las losowy drzew decyzyjnych).

Celem rozprawy jest opracowanie efektywnych metod tworzenia zespołów klasyfikacyjnych oraz sposobu fuzji pojedynczych wyników jego członków w werdykt końcowy. Sprawdzeniu poddano hipotezę, że zastosowanie odpowiednio zdefiniowanego zespołu klasyfikatorów głębokich CNN pozwala istotnie polepszyć zdolności generalizacji systemu klasyfikacyjnego i poprawić wynik rozpoznania klasy przypisanej różnym zmianom nowotworowym w stosunku do rezultatów osiągniętych przez indywidualne klasyfikatory.

Badania eksperymentalne zostały przeprowadzone dla dwóch rodzajów obrazowania. W przypadku czerniaka jest to obraz dermoskopowy skóry, a w przypadku raka piersi – obraz mammograficzny. W pracy przedstawiono kilka rozwiązań budowy zespołu klasyfikatorów. Jednym z nich jest współpraca rozwiązań klasycznych klasyfikatorów z siecią głęboką CNN. Wyniki przeprowadzonych eksperymentów pokazują stosunkowo niewielką poprawę wyników takiego rozwiązania w stosunku do najlepszych indywidualnych rozwiązań sieci CNN.

Drugim proponowanym rozwiązaniem jest zespół złożony z wielu współpracujących ze sobą różnych sieci głębokich, których wyniki podlegają fuzji poprzez głosowanie większościowe, oparte na prawdopodobieństwie przynależności klasowej. Niezależnie od rodzaju obrazowania, opracowany system klasyfikacji okazał się bardzo wydajny, generując wyniki na poziomie najlepszych rozwiązań światowych.

Wnioski z przeprowadzonych badań potwierdzają, że zespołowe uczenie maszynowe, oparte na kombinacji różnych klasyfikatorów, może przynieść znaczące korzyści w procesie analizy obrazów medycznych. Jednocześnie, wyniki te podkreślają wagę odpowiedniego doboru członków zespołu oraz ich optymalnej konfiguracji, aby osiągnąć najlepsze rezultaty.

Ensemble of deep networks in recognition of selected classes of medical images.

kpt. mgr inż. Fabian GIL

Keywords: computer vision, deep learning, convolutional neural network, ensemble of classifiers, feature selection, melanoma, breast cancer.

Abstract

The subject of work focuses on the development of automatic methods of analysis of medical images of melanoma and breast cancer. The solutions proposed in this work are based on the use of different deep neural CNN structures and classic type classifiers (SVM, random forest of decision trees).

The main task is to develop effective methods of creating the classification ensembles and the methods of fusion of individual results of its members in the final verdict. The results of numerical experiments confirm the hypothesis that the use of a properly defined and integrated set of deep CNN classifiers allows for improvement in the generalization ability of the proposed classification system concerning the results achieved by the individual classifiers.

The experimental study was conducted for two types of images. In the case of melanoma, it is a dermatoscopic image of the skin and in the case of breast cancer, it is a mammographic image. Several solutions for building a team of classifiers are proposed and tested in the paper. One of them is the cooperation of the deep CNN network with classical classifiers. The results of the experiments showed a relatively slight improvement in the results of such a solution compared to the best individual CNN solutions.

The second proposed solution is a team consisting of many cooperating deep networks of different structures, the results of which are fused through a majority voting, based on the probability of class membership. Regardless of the type of imaging, the developed classification system turned out to be very efficient, generating results at the level of the best-world solutions.

Conclusions from the conducted research confirm that team machine learning, based on combinations of various classifiers, can bring significant benefits in the process of analyzing medical images. At the same time, these results emphasize the importance of proper selection of team members and their optimal configuration to achieve the best results.

Spis treści

| | |
|-----------------------------------------------------------------------------------------------------------------------|-----------|
| 1. Wstęp | 7 |
| 1.1. Wprowadzenie w tematykę rozprawy..... | 7 |
| 1.2. Stan wiedzy w tematyce pracy..... | 7 |
| 1.3. Cel i przegląd zawartości pracy | 13 |
| 2. Baza danych medycznych | 15 |
| 2.1. Czerniak – opis, charakterystyka i statystyka | 15 |
| 2.2. Rak piersi – opis, charakterystyka i statystyka | 19 |
| 3. Indywidualne systemy klasyfikacyjne | 22 |
| 3.1. Sieci głębokie CNN – warstwy lokalnie połączone..... | 22 |
| 3.2. Sieci głębokie CNN – klasyfikator końcowy | 28 |
| 3.2.1. Klasyfikator typu softmax | 29 |
| 3.2.2. Klasyczne klasyfikatory zastosowane w pracy..... | 32 |
| 3.3. Deterministyczne metody selekcji cech diagnostycznych..... | 36 |
| 4. Zespoły klasyfikatorów | 42 |
| 4.1. Zasada działania i sposoby tworzenia zespołów klasyfikatorów..... | 42 |
| 4.2. Metody oceny jakości klasyfikacji zastosowane w pracy..... | 50 |
| 5. Wyniki badań eksperymentalnych w rozpoznawaniu czerniaka | 53 |
| 5.1. Rozpoznanie czerniaka na zbiorze NIO..... | 53 |
| 5.1.1. Wyniki klasyfikacji indywidualnych modeli CNN – zbiór NIO..... | 55 |
| 5.1.2. Wyniki klasyfikacji zespołów zbudowanych na pojedynczej sieci CNN rozszerzonej o SVM i RF – zbiór NIO | 65 |
| 5.1.3. Wyniki klasyfikacji zespołów złożonych wyłącznie z sieci CNN – zbiór NIO | 70 |
| 5.2. Wyniki rozpoznania czerniaka na zbiorze ISIC..... | 72 |
| 5.2.1. Wyniki klasyfikacji indywidualnych modeli CNN – zbiór ISIC | 73 |
| 5.2.2. Wyniki klasyfikacji zespołów złożonych z sieci CNN – zbiór ISIC..... | 76 |
| 6. Wyniki badań eksperymentalnych w rozpoznawaniu raka piersi | 79 |
| 6.1. Organizacja badań eksperymentalnych | 80 |
| 6.2. Rozpoznanie raka złośliwego od łagodnego..... | 80 |
| 6.3. Rozpoznanie zmian nowotworowych od stanu normalnego | 88 |
| 6.4. Klasyfikacja trójklasowa..... | 90 |

| | |
|-----------------------------------------------|------------|
| 7. Podsumowanie i wnioski końcowe..... | 95 |
| 7.1. Główne osiągnięcia rozprawy..... | 97 |
| 7.2. Kierunki dalszych badań..... | 97 |
| Wykaz literatury | 99 |
| Spis rysunków | 105 |
| Spis tabel..... | 108 |

1. Wstęp

1.1. Wprowadzenie w tematykę rozprawy

Tematyka rozprawy dotyczy opracowania komputerowych metod analizy wybranej klasy obrazów medycznych ukierunkowanych na wspomaganie ekspertów medycznych w rozpoznaniu stanów patologicznych na podstawie obrazowania komputerowego określonych organów. Metody takie są rozwijane na świecie od wielu lat i ciągle udoskonalane wraz z rozwojem techniki obliczeniowej i stosowanych metod. Przy ogromnym obserwowanym postępie technologicznym w systemach komputerowych zainteresowanie takim podejściem do problemu jest coraz większe. Coraz doskonalsze komputery kreują szybki postęp w dziedzinie sztucznej inteligencji i uczenia maszynowego umożliwiając uzyskanie coraz lepszych wyników analizy na podstawie dostępnej ilości danych medycznych. Dzięki temu możliwe staje się skuteczne wspomaganie lekarzy w stawianiu diagnozy, co prowadzi do bardziej skutecznego leczenia.

Zastosowanie systemów komputerowych w analizie obrazów medycznych dotyczy wielu różnych problemów i metod przetwarzania związanych z konkretnym rodzajem obrazowania, dostosowanego do określonych celów, związanych z rodzajem diagnozy medycznej.

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

1.2. Stan wiedzy w tematyce pracy

Postęp w dziedzinie automatycznej analizy obrazów medycznych dokonywał się równoległe z rozwojem komputerowej techniki obliczeniowej, stosując coraz bardziej doskonałe narzędzia wykorzystujące inteligencję obliczeniową i uczenie maszynowe.

Podstawę tych zastosowań od wielu lat stanowią sztuczne sieci neuronowe, w tym sieci klasyczne typu perceptron wielowarstwowy MLP, sieć radialna RBF czy maszyna wektorów nośnych SVM, która również zaliczana jest do rodziny sieci neuronowych [43]. Struktury te stosują zwykle jedną (co najwyżej dwie) warstwy ukryte. Podstawą aplikacji tych rozwiązań jest ich zdolność generalizacji. Struktury sieciowe wytrenowane na zbiorze danych uczących uczą się mechanizmu tworzenia tych danych, co jest podstawą uogólnienia nabytej wiedzy na zbiór danych nie uczestniczących w uczeniu, ale generowanych według wykrytego mechanizmu. Wadą rozwiązań klasycznych jest konieczność eksperckiej znajomości analizowanych procesów przy generacji cech diagnostycznych procesu, stanowiących atrybuty wejściowe tych systemów.

Duży postęp w dziedzinie uczenia maszynowego zawdzięczamy opracowaniu struktur i mechanizmów uczenia sieci głębokich stosujących wiele warstw ukrytych odpowiedzialnych za automatyczną generację cech diagnostycznych procesu [4,20] na podstawie oryginalnych danych pomiarowych. Poczynając od prac LeCuna [4] powstało wiele oryginalnych rozwiązań sieci głębokich, z których najważniejszą pozycję w analizie obrazów stanowią sieci konwolucyjne (splotowe) CNN. Warstwy ukryte, których liczba może sięgać setek, są odpowiedzialne za generowanie deskryptorów numerycznych procesu, zwalniając w ten sposób użytkownika z manualnego podejścia do tego problemu. Stąd sieci głębokie stanowią dzisiaj podstawowe narzędzie najczęściej stosowane w analizie i rozpoznawaniu obrazów medycznych, w tym obrazów dotyczących czerniaka i raka piersi, stanowiących główne zainteresowanie autora rozprawy.

Niezależnie od rodzaju zastosowanych struktur klasyfikatora skuteczność systemu rozpoznania obrazów może być zwiększona poprzez fuzję wyników wielu klasyfikatorów pracujących równolegle i połączonych w zespół. Wyniki rozpoznania klas uzyskane przez poszczególnych członków zespołu podlegają integracji (agregacji), najczęściej przy zastosowaniu głosowania większościowego zwykłego lub ważonego [36,44,72]. Zakładając niezależność działania klasyfikatorów powstaje wówczas pole do poprawy skuteczności tak zorganizowanego systemu podejmowania decyzji.

W efekcie przeprowadzonych badań powstało wiele metod tworzenia zespołu i algorytmów jego integracji. Najbardziej typowe stosują rozwiązania bazujące na różnych strukturach sieciowych, trenowanych na losowo dobranych podzbiorach uczących [35,56]. Ważną rolę w tworzeniu zespołu odgrywa losowość, stosowana zarówno w doborze danych uczących jak i tworzeniu składu zespołu. W pracy [55,65] pokazano rozwiązania, w których losowość stosowana jest na poziomie zarówno warstw ukrytych jak

i tworzenia wyniku wyjściowego zespołu. Interesującym rozwiązaniem fuzji zespołu jest również integracja dynamiczna, w której o wyniku zespołu decyduje jeden (lub kilka równorzędnych członków) wykazujący najlepsze wyniki na danych testujących najbliższych odpowiednim danym uczącym [56].

Zastosowanie określonych metod analizy obrazów prowadzi do wyników, które mogą różnić się co do jakości w zależności od stopnia złożoności problemu, rodzaju obrazowania komputerowego, typu narzędzi zastosowanych w przetwarzaniu, jak również wielkości zbioru danych uczących dostępnych w analizie. Stąd wyniki uzyskane przez różne zespoły badawcze (nawet przy zastosowaniu podobnych metod) mogą różnić się co do wartości wskaźników jakościowych.

W przypadku czerniaka przedstawiony tutaj przegląd rozwiązań skoncentrowany będzie na zbiorze własnym NIO oraz na zbiorze ISIC powstałym przy współpracy różnych instytucji międzynarodowych. Zbiór NIO, stworzony przez Zespół Kliniki Nowotworów Tkanek Miękkich, Kości i Czerniaków z Narodowego Instytutu Onkologii w Warszawie podlegał badaniom przy użyciu różnych metod przetwarzania w pracach [42,55,56].

W pracy [55] pokazano zastosowanie zespołu klasyfikatorów CNN stosującego dwa rodzaje aktywacji ReLU oraz softplus połączonych losowo w każdej warstwie. Na zbiorze NIO uzyskano wyniki rozpoznania czerniaka lepsze niż każdego indywidualnego rozwiązania. W trybie 10-krotnej walidacji krzyżowej uzyskano dokładność równą $91.8\% \pm 5.40\%$, miarę $F1 = 92.3\% \pm 5.02\%$ oraz $AUC = 0.9710 \pm 0.0312$.

W pracy [42] zastosowano zespół klasyfikatorów bazujący na sieci SVM, lesie losowym RF i softmaxie zasilane poprzez specjalnie wyselekcjonowane deskryptory (5 różnych metod) uzyskując na zbiorze NIO dokładność równą $93.89\% \pm 2.14\%$.

Praca [56] pokazuje wyniki działania zespołu integrowanego dynamicznie z użyciem klasyfikatora najlepiej dostosowanego do aktualnie analizowanego obrazu czerniaka. Uzyskano istotną poprawę wyników w stosunku do najlepszego klasyfikatora indywidualnego. Deklarowana dokładność rozpoznania 2 klas (czerniak kontra zmiana nie będąca czerniakiem) to 98%, czułość 93% i $AUC = 0.9240$.

W przypadku międzynarodowego zbioru ISIC wyniki rozpoznania czerniaka od różnych zmian skórnych nie czerniakowych zasadniczo różnią się w zależności od wielkości zbioru (zbiór ten podlega ciągłemu powiększaniu w miarę dołączania coraz nowszych danych przez różne zespoły badawcze), przy czym wiele obrazów jest zdublowanych. W pracy [6] porównano zbiory z roku 2020 i 2017. Po usunięciu 14310

obrazów ze zbioru 2020 uzyskano wynik przedstawiony jedynie w postaci $AUC = 0.80$. Wyniki rozpoznania poszczególnych klas różnią się również w zależności od zastosowanego podejścia, choć najlepsze odpowiadają zastosowaniu sieci CNN zorganizowanych w zespole.

W pracy [13] zaprezentowano wyniki zespołu klasyfikatorów CNN dotyczące ogromnej liczby danych (129 450 klinicznych obrazów użytych w uczeniu), i trzech rozpoznawanych klas zmian skórnych bazujących na zbiorze ISIC i danych z uniwersytetu Stanforda. Uzyskany wynik to dokładność $72.1\% \pm 0.9\%$, lepszy niż średni wynik 21 ekspertów dermatologicznych dla tego zbioru.

W pracy [70] zaprezentowano wyniki rozpoznania 6 klas zmian skórnych na bazie ISIC przy wykorzystaniu zespołu klasyfikatorów CNN. Uzyskano dokładność 76%.

W pracy [69] badaniom poddano rozpoznanie dwóch klas czerniaka na zbiorze ISIC z roku 2017 prezentując wyniki jedynie w postaci $AUC = 0.880$.

Praca [8] prezentuje wyniki działania zespołu sieci CNN na zbiorze ISIC ograniczonym do 900 obrazów treningowych i 379 obrazów testowych. Deklarowane wartości to $AUC = 0.843$, dokładność 76%, precyzja 64.9%, czułość 82% i specyficzność 62%.

W pracy [42] przedstawiono wyniki rozpoznania dwóch klas (czerniak reprezentowany przez 945 obrazów kontra nie czerniak – 1543 obrazów) wszystkie ze zbioru ISIC z 2017 roku, przy wykorzystaniu zespołu utworzonego na podstawie trzech rodzajów klasyfikatorów (SVM, RF oraz softmax) zasilanych poprzez atrybuty wytworzone z deskryptorów CNN przy użyciu pięciu różnych metod generacji cech. W efekcie uzyskano dokładność równą 80.86% oraz $AUC = 0.853$.

W pracy [1] przeanalizowano różne formy preprocessingu i klasyfikacji obrazów czerniaka ze zbioru ISIC 2017. Najlepszy model eVida M6 stosujący sieć CNN typu Resnet152 pozwolił uzyskać dokładność $ACC = 90.4\%$, czułość 82% i specyficzność 92.5%.

Wyniki badań na zbiorach ISIC z trzech lat (2016, 2017, 2020) przy zastosowaniu stworzonej od podstaw sieci głębokiej zawierającej 11 warstw zwanej „lesion classification network” (LCNet) zostały przedstawione w pracy [32]. Przedstawiono jedynie wartości dokładności rozpoznania czerniaka na tych zbiorach: 81.41% (ISIC 2016), 88.23% (ISIC 2017) oraz 90.42% (ISIC 2020).

Praca [49] przedstawia wyniki zwycięzcy konkursu IIM-ISIC Melanoma Classification Challenge 2019, polegającego na rozpoznaniu 9 klas zmian skórnych z bazy ISIC. W rozwiązaniu zastosowano zespół klasyfikatorów bazujących głównie na sieciach

Efficientnet i Resnet (Efficientnet B3, B4, B5, B6, B7, SE-ResNeXt-101 oraz ResNeSt-101). Wyniki podano jedynie w odniesieniu do jednej miary AUC, która była równa 0.96.

W pracy [31] porównano różne rozwiązania sieci głębokiej (10 różnych struktur) w rozpoznaniu czerniaka na bazie SIIM-ISIC Melanoma Classification 2020 Kaggle dataset. Rozwiązania CNN są wspomagane przez dodatkowe zbiory danych wygenerowane przez zmodyfikowane wersje sieci GAN. Porównanie dotyczy wyłącznie miary AUC, dla której najlepsze wyniki dotyczą zastosowania sieci SGAN (Semi-Supervised GAN), uzyskując $AUC = 0.9431$ na zbiorze testowym.

Wiele prac dotyczy analizy obrazów mammograficznych. W pracy [41] przedstawiono historię obrazowania mammograficznego raka piersi oraz kierunki badań przyszłościowych. Praca [2] przedstawia metody przetwarzania obrazów mammograficznych ukierunkowanych na wykrywanie obszarów zmian nowotworowych piersi. Badania przeprowadzono na bazie "Mammographic Image Analysis Society" (MIAS) stworzonej w szpitalach Arabii Saudyjskiej zawierającej 2892 obrazów uzyskując dokładność rozpoznania 92%.

W pracy [63] zaprezentowano rozwiązanie zespołowe wykorzystujące sieć MLP, klasyfikator LDA oraz klasyfikator naiwny Bayesa. Sieci te zasilane są cechami diagnostycznymi wytworzonymi przy zastosowaniu transformacji falkowej i fourierowskiej i poddany selekcji typu entropijnego. Rozwiązanie zostało przetestowane na bazie obrazów mammograficznych utworzonej przez Instytucję „Eastern Health in Newfoundland and Labrador of Canada”, pozwalając uzyskać dokładność 96.06%, czułość 96.45% i specyficzność 97.45%.

Praca [46] przedstawia system klasyfikacyjny do rozpoznania raka złośliwego od łagodnego oparty na trzech sieciach głębokich (AlexNet, ResNet18, i ResNet34). Badania przeprowadzono na zbiorze 1986 mamogramów, z czego w testowaniu użyto 389 przypadków złośliwych i 611 łagodnych. Najlepsze wyniki działania odpowiadały sieci AlexNet generując czułość 98%, specyficzność 89% i $AUC = 0.98$.

W pracy [47] zbadano zastosowanie modeli głębokich bazujących na ResNet, EfficientNet oraz MaxViT. Badania eksperymentalne porównujące różne rozwiązania sieciowe przeprowadzono na bardzo dużym zbiorze ADMANI [14] z którego wybrano 11913 mamogramów zawierających jedynie 486 przypadków raka piersi. Najlepsze wyniki osiągnięto dla sieci ResNet34 uzyskując średnią dokładność 97%, ale przy bardzo niskiej (nieakceptowalnej) czułości 9%, precyzji 16% i mierze $F1 = 0.12$.

Badania przedstawione w pracy opierać się będą na bazie DDSM [26], dla której przeprowadzono badania w pracach [10,37,54]. Praca [10] prezentuje rozwiązanie w którym istotną rolę odgrywa metoda generacji cech diagnostycznych oparta na teorii momentów statystycznych zastosowanych do transformacji krzywiznowej typu curvelet. Klasyfikator KNN zasilany jest przez zbiór cech wyselekcjonowanych przy użyciu statystycznego testu t-test. Badania przeprowadzone na zbiorze DDSM przy rozpoznaniu przypadku zmian rakotwórczych od zbioru zawierającego przypadki normalne wykazały średnią dokładność zmieniającą się od 81.3% do 86.4% w zależności od zastosowanego zbioru atrybutów wejściowych. Wobec ogromnej przewagi przypadków normalnych w zbiorze, brak wartości odnoszącej się do czułości nie pozwala jednoznacznie określić jakość systemu klasyfikacyjnego.

W pracy [57] zaproponowano rozwiązanie zespołu klasyfikatorów składającego się z 10 jednostek w tym SVM, drzewo decyzyjne, las losowy drzew, klasyfikator KNN, autoencoder, klasyfikator logistyczny w różnych aranżacjach wynikających z zastosowania różnego zestawu atrybutów wejściowych. Badania przeprowadzono na bazie DDSM uzyskując najlepsze rezultaty w dwóch różnych zadaniach klasyfikacyjnych. Przy rozpoznaniu przypadków chorobowych (rak złośliwy i łagodny) od normalnych uzyskano: dokładność 84.5%, czułość 82.9%, specyficzność 84.9%, AUC = 0.929. Z kolei rozpoznanie przypadków raka złośliwego od reszty (rak łagodny łącznie z przypadkami normalnymi) dało wyniki: dokładność 80.2%, czułość 83.3%, specyficzność 79.8%, AUC = 0.890.

W pracy [37] dokonano porównania dwu podejść do rozpoznania obrazów mammograficznych: zastosowanie klasycznych rozwiązań sieciowych i sieci głębokich, wykazując przewagę tych ostatnich. Na bazie DDSM w przypadku podejścia klasycznego, wykorzystującego wiele różnorodnych deskryptorów statystycznych obrazu stanowiących (po selekcji) atrybuty wejściowe dla klasyfikatora SVM, uzyskano dokładność 81%, czułość 82.48% i specyficzność 79.63% w rozpoznaniu klasy nowotworowej od przypadków normalnych. Sieci głębokie pozwoliły uzyskać lepsze rezultaty rozpoznania: dokładność 85.83%, czułość 82.92% i specyficzność 86.59%.

Praca [54] pokazuje efekt wzbogacenia danych uczących poprzez dodatkowo wygenerowane obrazy klasy mniejszościowej przy użyciu sieci zmodyfikowanej GAN. Na zbiorze danych z bazy DDSM uzyskano bardzo dobre wyniki rozpoznania klasy reprezentującej zmiany nowotworowe od przypadków normalnych, w tym dokładność równą 89.71%, czułość 93.54%, specyficzność 80.58% i AUC = 0.9410.

1.3. Cel i przegląd zawartości pracy

Tematyka pracy koncentrować się będzie na opracowaniu automatycznych metod zespołowych rozpoznawania obrazów medycznych. Proponowane rozwiązania opierać się będą na zastosowaniu różnych neuronowych struktur sieci głębokich CNN oraz klasyfikatorów typu klasycznego (SVM, las losowy drzew decyzyjnych).

Celem pracy jest opracowanie efektywnych metod tworzenia takich zespołów klasyfikacyjnych oraz sposobu fuzji pojedynczych wyników jego członków w werdykt końcowy. Proponowane rozwiązania wykorzystywać będą różnorodne podejścia do tworzenia zespołowego systemu klasyfikacyjnego, w tym współpracę sieci klasycznych i głębokich.

Niezależnie od proponowanego rozwiązania kluczową rolę w zespole odgrywać będą sieci głębokie CNN, stosowane zarówno do generacji deskryptorów numerycznych stanowiących podstawę tworzenia atrybutów wejściowych dla klasyfikatorów klasycznych jak i stanowiących główne narzędzia do podejmowania ostatecznej decyzji przypisania danych wejściowych do określonej klasy.

Opracowane systemy zespołowe zostaną zastosowane do analizy dwóch rodzajów obrazów medycznych: obrazów dermoskopowych dotyczących czerniaka oraz obrazów mammograficznych raka piersi. Zostanie pokazane, że opracowane systemy pozwalają na istotną poprawę wyników rozpoznania klasy przypisanej różnym zmianom nowotworowym.

Na podstawie wyników wielu przeprowadzonych eksperymentów numerycznych sformułowano następującą tezę:

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

Na treść pracy składa się siedem rozdziałów:

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

Rozdział 2 zawiera opis oraz analizę statystyczną obrazów medycznych czerniaka i raka piersi, które zostały wykorzystane w eksperymentach numerycznych

przeprowadzonych w ramach pracy. Analiza obejmuje dwie bazy danych czerniaka oraz jedną bazę DDSM raka piersi.

Rozdział 3 skupia się na opisie wykorzystanych w pracy metod klasyfikacji obrazów, takich jak konwolucyjne sieci neuronowe, drzewa decyzyjne i Support Vector Machine. Opisuje również różne metody selekcji cech diagnostycznych, zastosowane w analizie danych przy wykorzystaniu różnych klasycznych klasyfikatorów.

Rozdział 4 zawiera opis koncepcji tworzenia zespołów klasyfikatorów oraz metod ich integracji. Przedstawia również różne metody oceny jakości wyników klasyfikacji, użyteczne w porównywaniu skuteczności różnych analizowanych rozwiązań.

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

Rozdział 6 zawiera wyniki badań eksperymentalnych w rozpoznawaniu raka piersi na zbiorze DDSM. Przedstawiono rezultaty badań dotyczące trzech zadań klasyfikacyjnych związanych z rozpoznaniem raka złośliwego, łagodnego oraz stanu normalnego (zadania dwuklasowe oraz trzyklasowe).

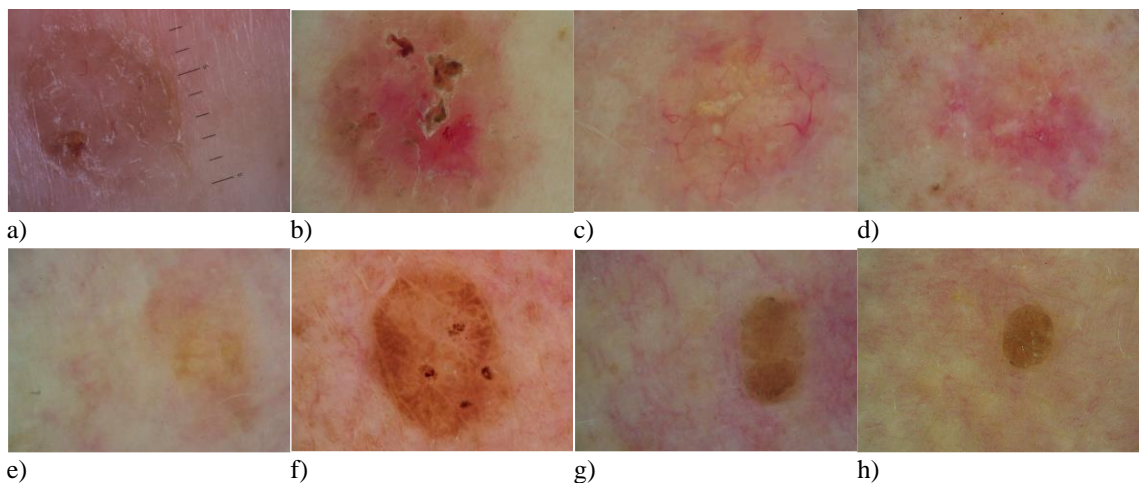
Ostatni rozdział zawiera podsumowanie wyników całej rozprawy, wnioski płynące z przeprowadzonych badań oraz sugestie dotyczące kierunków dalszego rozwoju w dziedzinie rozpoznawania obrazów medycznych przy użyciu metod uczenia maszynowego związanego z uczeniem głębokim.

2. Baza danych medycznych

Badania eksperymentalne przeprowadzone w niniejszej rozprawie dotyczą dwóch różnych klas problemów medycznych. Pierwszy związany jest z rozpoznaniem obrazów dermoskopowych, obejmujących przypadki zmian skórnych czerniaka i zmian łagodnych nie będących czerniakiem. Drugi problem związany jest z rozpoznaniem obrazów mamograficznych ilustrujących raka piersi w stadiach złośliwym i łagodnym oraz przypadki zdrowe. W przypadku czerniaka wykorzystano dwie bazy danych. Jedna z nich została stworzona przez Zespół Kliniki Nowotworów Tkanek Miękkich, Kości i Czerniaków z Narodowego Instytutu Onkologii w Warszawie, która będzie oznaczana skrótem NIO. Druga baza została zaczerpnięta z ogólnodostępnego zbioru ISIC (The International Skin Imaging Collaboration) [64,70]. Wybór drugiej bazy stworzył możliwość odniesienia i porównania wyników badań do istniejących publikacji światowych obejmujących problematykę rozpoznawania takich obrazów medycznych, przedstawionych w rozdziale pierwszym.

2.1. Czerniak – opis, charakterystyka i statystyka

Baza danych NIO została stworzona przez zespół pracowników pod kierunkiem dr Moniki Słowińskiej [55,35]. Zbiór NIO zawiera 134 obrazy czerniaka oraz 112 obrazów brodawki łojotokowej reprezentujących łagodną odmianę nowotworu skóry niebędącą czerniakiem. Jest to bardzo mała baza danych stwarzająca dodatkowe problemy w procesie uczenia systemu.



Rys. 2.1 – Przykładowe obrazy zbioru danych NIO. Obrazy w górnym wierszu przedstawiają czerniaka, a w dolnym wierszu przypadki niebędące czerniakiem.

Obrazy zmian skórnych z różnych części ciała zostały wykonane za pomocą dermatoskopu o powiększeniu 20× i zapisane jako obrazy RGB z wykorzystaniem kompresji JPEG. Baza danych została opisana przez ekspertów z dziedziny dermatologii przy zastosowaniu klasycznego kryterium ABCDE [19] (A - asymetria, B – brzegi obszaru zmienionego, C - kolor, D – rozmiar zmian, E – ewolucja zachodzących w czasie zmian) oraz potwierdzona dokładnym badaniem patomorfologicznym. Obrazy powstawały w różnym czasie i były zapisywane w różnych rozdzielczościach, w związku z tym rozmiar obrazów oscylował pomiędzy 767×576 a 4273×2848 pikseli.

Na rys. 2.1 przedstawiono kilka przykładowych obrazów ilustrujących przypadki czerniaka i innych zmian skórnych. Obrazy w górnym wierszu (a, b, c, d) reprezentują czerniaka, a w dolnym wierszu (e, f, g, h) brodawkę łojotokową. Zasadniczym problemem w rozpoznawaniu zmian skórnych jest występowanie znacznych różnic w kształcie, wielkości a także rozmieszczeniu kolorów w obszarach uszkodzeń w obrębie tej samej klasy. Co więcej występują także duże podobieństwa wymienionych cech pomiędzy obrazami należącymi do dwóch różnych klas.

Dla lepszego scharakteryzowania obu klas w pracy dokonano porównania pewnych wartości statystycznych analizowanych obrazów, w tym wartości średniej, odchylenia standardowego, energii oraz kurtozy i skośności na stopniu jasności pikseli w obrazach tworzących zbiory. Wartości średnie tych parametrów obliczono dla wszystkich obrazów w obu zbiorach. Wartości wyżej wymienionych parametrów dotyczące obu klas przedstawiono w tabeli 2.1. W obu klasach można zauważyć dużą wartość relacji odchylenia standardowego do analizowanej wartości średniej pikseli, co świadczy o dużej różnorodności obrazów reprezentujących tę samą klasę.

Tabela 2.1 - Parametry statystyczne obrazów ze zmianami czerniakowymi i nieczerniakowymi zbioru NIO dla wszystkich obrazów zbioru.

| Klasa | Średnia | Odchylenie standardowe | Energia | Kurtoza | Współczynnik skośności |
|--------------|----------------|------------------------|--------------|-------------|------------------------|
| Czerniak | 133.22 ± 26.57 | 36.42 ± 13.99 | 19969 ± 8387 | 2.48 ± 0.74 | -0.17 ± 0.39 |
| Nie-czerniak | 117.81 ± 21.61 | 27.43 ± 5.6 | 15125 ± 5703 | 3.12 ± 1.12 | -0.57 ± 0.5 |

Jednocześnie można zauważyć, że w większości przypadków średnie wartości charakteryzujące parametry statystyczne obrazów dwu różnych klas są do siebie zbliżone. Ten fakt potwierdza statystyczne podobieństwo między próbkami należącymi do obu klas. Znaczną różnicę obserwuje się jedynie w przypadku energii i współczynnika

skości. W pierwszym przypadku mogło to być spowodowane zróżnicowanym poziomem oświetlenia w dermatoskopach wykorzystanych do tworzenia obrazów, natomiast drugi parametr należy potraktować jako mało wiarygodny ze względu na jego bardzo duże odchylenie standardowe.

Ważną cechą pokazującą trudność rozpoznania klas jest stosunkowo duża różnorodność reprezentantów każdej z klas. Miarą tej zmienności jest indeks strukturalnego podobieństwa obrazów (ang. *structural similarity index measure – SSIM*). W tabeli 2.2 przedstawiono wartości średnie wraz z odchyleniem standardowym liczone oddzielnie dla każdego reprezentanta klasy oraz te same wartości (średnia \pm odchylenie standardowe) dla odchylenia standardowego poszczególnych obrazów. Wartości dotyczą obrazów wewnątrz pojedynczej klasy oraz podobieństwo obrazów reprezentujących przeciwne klasy.

Tabela 2.2 Wartości indeksu podobieństwa strukturalnego wewnątrzklasowego i międzyklasowego obrazów dla bazy NIO.

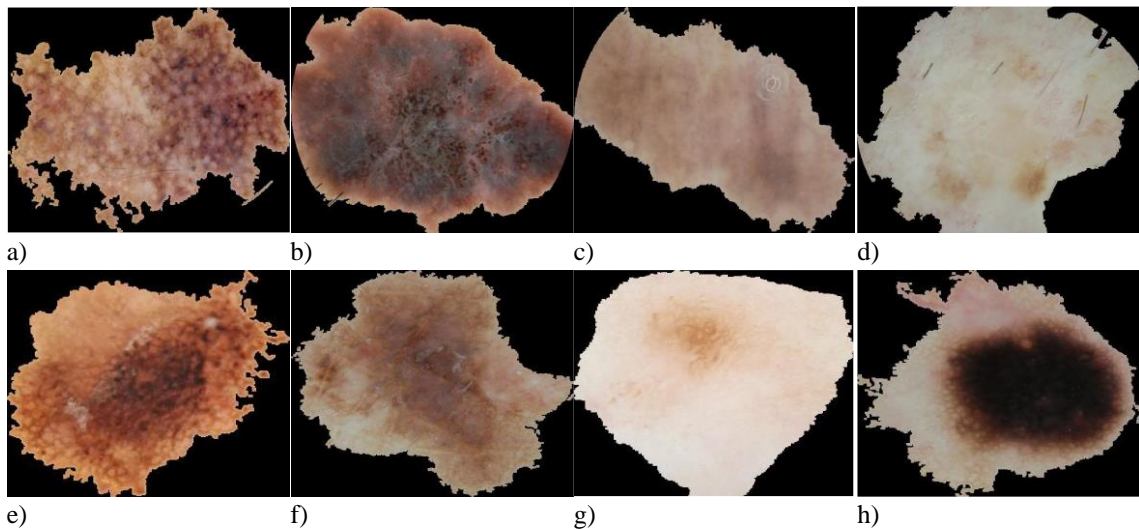
| Miara SSIM | Średnia | Odchylenie standardowe |
|--------------------------------|---------------------|------------------------|
| Wewnątrzklasowa - czerniak | 0.7514 \pm 0.0544 | 0.0987 \pm 0.0209 |
| Wewnątrzklasowa – nie czerniak | 0.7653 \pm 0.1225 | 0.1474 \pm 0.0213 |
| Międzyklasowa | 0.7216 \pm 0.1049 | 0.1111 \pm 0.0177 |

Drugi zbiór obrazów wykorzystany w eksperymencie dotyczy również zmian czerniakowych skóry, ale został stworzony przez międzynarodową organizację ISIC [64,70]. Celem grupy było dostarczenie zainteresowanym ośrodkom badawczym bazy danych do badań nad cyfrowym przetwarzaniem obrazów zmian skórnych, umożliwiając w ten sposób przyspieszenie procesu rozpoznawaniu raka we wczesnych stadiach, a tym samym zmniejszenie śmiertelności z powodu czerniaka.

Zbiór ISIC jest znacznie większy i ciągle podlega ciągłej modyfikacji. Baza użyta przez autora pochodzi sprzed kilku lat i zawiera 945 obrazów RGB czerniaka oraz 1543 obrazów znamion niebędących czerniakiem. Zmiany nie czerniakowe dotyczą różnych schorzeń skórnych. Obrazy pochodzą z różnych części ciała (plecy, brzuch, szyja, pierś, policzek, ramię, noga, udo, twarz) i zostały wykonane przy użyciu dermatografu metodą analogiczną jak w bazie NIO. Postaci obrazów odpowiadające tym zmianom są znacznie bardziej zróżnicowane niż w przypadku zbioru NIO.

Ważna różnica wynika również z dużego zróżnicowania zespołów wykonujących obrazy. W przypadku pierwszej bazy danych był to jeden zespół. Natomiast w przypadku

zestawu ISIC obrazy pochodziły z wielu różnych źródeł. Pomimo stosowania zbliżonych standardów przez wszystkie zespoły, różnice w obrazach widoczne wewnątrz zbioru danych są znaczne i mogą mieć również istotny wpływ na wyniki eksperymentów. Na rys. 2.2 przedstawiono kilka przykładowych obrazów zmian skórnych ze zbioru ISIC. Obrazy w górnym wierszu (a, b, c, d) przedstawiają czerniaka, a w dolnym wierszu (e, f, g, h) znamiona nie będące czerniakiem. Widoczne jest bardzo duże zróżnicowanie obrazów wewnątrz tej samej klasy jak i znaczne podobieństwo reprezentantów przeciwnych klas. Zastosowanie liczniejszej i bardziej różnorodnej bazy obrazów w stosunku do NIO powoduje dodatkowe utrudnienie zadania klasyfikacji związane ze zwiększonym podobieństwem przypadków zmian skórnych należących do różnych klas oraz ich różnorodnością wewnątrz każdej klasy.



Rys. 2.2 – Przykładowe obrazy zbioru danych ISIC. Obrazy w górnym szeregu przedstawiają czerniaka, a w dolnym szeregu przypadki niebędące czerniakiem.

W przypadku zbioru ISIC, wartości parametrów statystycznych reprezentujących obrazy czerniaka i znamion (Tabela 2.3) są bardzo bliskie sobie. Na przykład średnia jasność pikseli w całym zbiorze jest równa 88.04 dla czerniaka i 82.75 dla nie czerniaka, przy bardzo dużej wartości średniego odchylenia standardowego. Podobne relacje można zauważyć dla pozostałych parametrów statystycznych (odchylenie standardowe, energia czy kurtoza). Mało reprezentatywna jest wartość średnia współczynnika skośności ze względu na bardzo duże względne wartości odchylenia standardowego.

Jednocześnie można zauważyć dużą różnicę w stosunku do wartości parametrów statystycznych obrazów tworzących zbiór NIO. Jakkolwiek różnica w parametrach pomiędzy obrazami znamion ze zbioru ISIC i brodawki łojotokowej ze zbioru NIO nie budzi wątpliwości, tak rozbieżność między parametrami obrazów czerniaka ze zbioru ISIC

i NIO, pokazuje jak odrębne mogą być zbiory wykonane podobną metodą ale przez różne zespoły medyczne.

Tabela 2.3 - Parametry statystyczne obrazów ze zmianami czerniakowymi i nieczerniakowymi zbioru ISIC (średnia i odchylenie standardowe ze wszystkich obrazów zbiorów).

| Klasa | Średnia | Odchylenie standardowe | Energia | Kurtoza | Współczynnik skośności |
|--------------|---------------|------------------------|--------------|-------------|------------------------|
| Czerniak | 88.04 ± 22.82 | 66.21 ± 12.73 | 12818 ± 5018 | 2.17 ± 1.01 | -0.08 ± 0.68 |
| Nie-czerniak | 82.75 ± 21.08 | 67.65 ± 11.5 | 12000 ± 5159 | 1.91 ± 0.49 | 0.14 ± 0.46 |

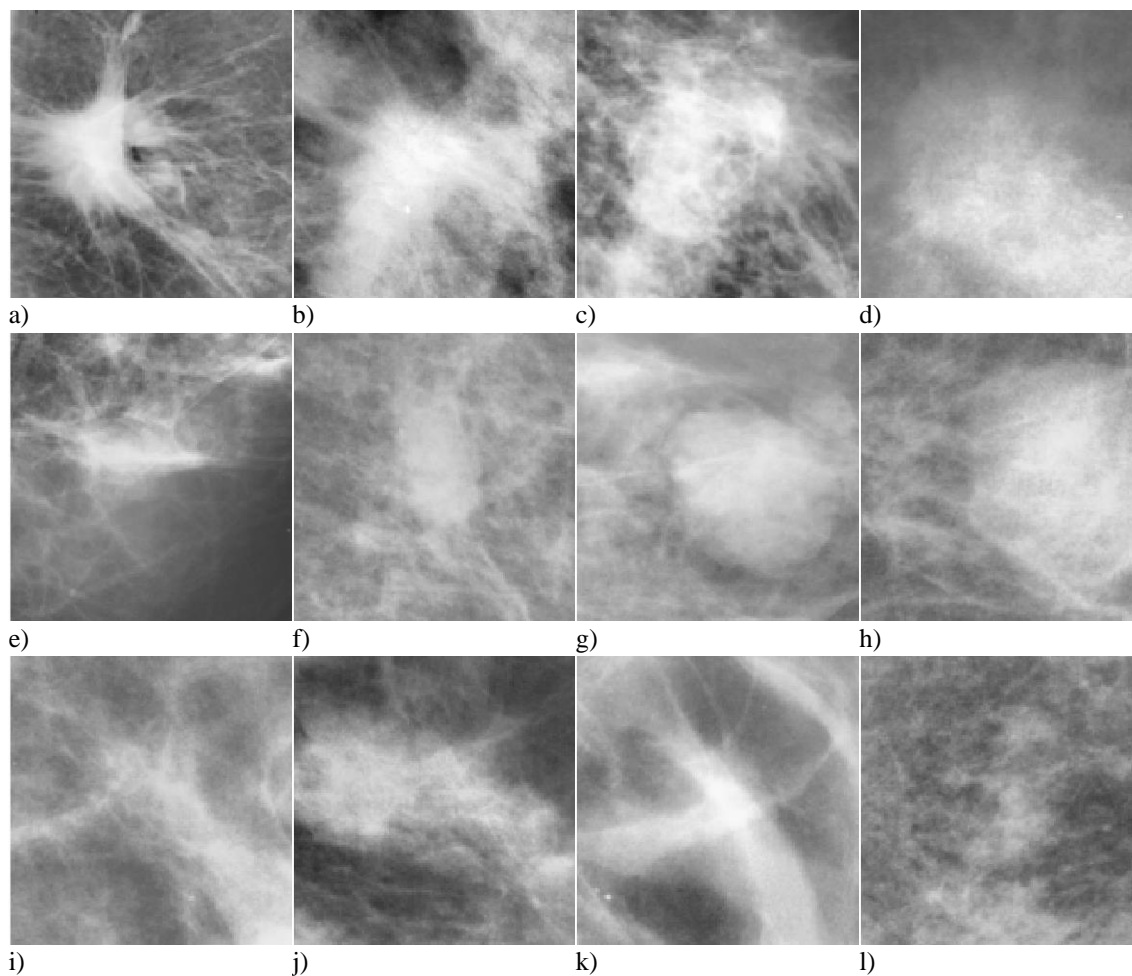
2.2. Rak piersi – opis, charakterystyka i statystyka

Drugim problemem poddanym badaniu w niniejszej rozprawie jest rozpoznawanie obrazów mammograficznych raka piersi uzyskanych w wyniku rentgenowskiego badania obrazowego, które obecnie jest traktowane jako standard w diagnostyce raka piersi. Badanie takie pozwala na uwidocznienie zmian nowotworowych na wczesnym etapie ich rozwoju.

W tym przypadku do eksperymentu wykorzystano ogólnodostępną bazę obrazów mammograficznych „Digital Database for Screening Mammography” (DDSM) stworzoną przez zespół z Uniwersytetu Południowej Florydy [26].

Zbiór obrazów zawiera 2802 przypadków, z których każda składa się z 4 mammogramów. Obserwacja obejmuje obrazy lewej i prawej piersi z góry, reprezentujące widok czaszkowo-ogonowy oraz skośne przedstawiając widok środkowo-boczny-skośny. Wszystkie obserwacje zostały podzielone na 3 klasy obejmujące przypadki zdrowe oraz raka piersi w stadiach łagodnym i złośliwym. Tkanki zdrowe traktowane jako normalne są reprezentowane przez 9215 obrazów. Przypadki z rozpoznaniem nowotworem obejmują 888 obrazów raka w stadium łagodnym oraz 1115 obrazów ze zmianami złośliwymi. Obrazy zostały przygotowane w formie ROI (ang. *Region of Interest*), wyodrębnionym przez ekspertów medycznych. Rozmiar obrazów ROI wynosił 128x128 pikseli.

Na rys. 2.3 przedstawiono kilka przykładowych obrazów ROI mammogramów. Obrazy w pierwszym wierszu reprezentują raka w stadium złośliwym, w drugim przypadku zmian łagodnych i w trzecim tkanki normalne bez zmian nowotworowych. Analogicznie do problemu zmian czerniakowych, w zbiorze mammogramów obserwujemy istotne różnice pomiędzy obrazami należącymi do tej samej klasy danych, a jednocześnie duże podobieństwo między próbkami reprezentującymi różne klasy.



Rys. 2.3 – Przykładowe obrazy mammograficzne. Obrazy w górnym szeregu przedstawiają raka w stadium złośliwym, w środkowym raka w stadium łagodnym, a w dolnym przypadki tkanek normalnych.

Dla ich scharakteryzowania w tabeli 2.4 przedstawiono wartości wybranych parametrów statystycznych dotyczące zbiorów wszystkich obrazów mammograficznych w całej bazie danych. Niezależnie od klasy (rak złośliwy, rak łagodny, przypadek normalny) wartości te są bardzo podobne, a różnice mało istotne statystycznie.

Tabela 2.4 - Parametry statystyczne obrazów mammograficznych raka w stadium złośliwym i łagodnym oraz tkanek normalnych dla wszystkich obrazów zbioru.

| Klasa | Średnia | Odchylenie standardowe | Energia | Kurtoza | Współczynnik skośności |
|--------------|----------------|------------------------|--------------|-------------|------------------------|
| Rak złośliwy | 160.79 ± 19 | 33.3 ± 9.33 | 27412 ± 6051 | 2.48 ± 0.69 | 0.04 ± 0.47 |
| Rak łagodny | 158.69 ± 19.64 | 31.34 ± 9.34 | 26638 ± 6200 | 2.39 ± 0.51 | 0.01 ± 0.41 |
| Normalne | 160.79 ± 16.61 | 27.44 ± 9.33 | 26968 ± 5185 | 2.79 ± 0.99 | 0.01 ± 0.46 |

Dodatkowym utrudnieniem w procesie rozpoznania klas jest poważne niezrównoważenie ich populacji. Przeważa tu klasa przypadków normalnych, które stanowią 81%

całej bazy. Oznacza to konieczność zastosowania specjalnego sposobu podejścia do uczenia systemu klasyfikacyjnego, które minimalizowałoby te różnice liczebności klasowej.

Podobnie jak w przypadku czerniaka przeprowadzono estymację podobieństwa strukturalnego obrazów w poszczególnych klasach i podobieństwa międzyklasowego. W tabeli 2.5 przedstawiono wartości uśrednione średnich tej miary każdego obrazu oddzielnie wraz z odchyleniem standardowym oraz te same wartości uśrednione dla odchylenia standardowego poszczególnych obrazów. Wartości dotyczą obrazów wewnątrz pojedynczej klasy oraz podobieństwo obrazów reprezentujących dwie różne klasy. Niezależnie od typu badania (statystyka wewnątrzklasowa czy międzyklasowa) wartości te są do siebie zbliżone, co świadczy o dużym podobieństwie obrazów reprezentujących różne klasy.

Tabela 2.5 - Wartości indeksu podobieństwa strukturalnego wewnątrzklasowego i międzyklasowego obrazów mammograficznych dla bazy DDSM.

| Miara SSIM | Średnia | Odchylenie standardowe |
|-----------------------------------|---------------------|------------------------|
| Rak złośliwy | 0.3896 ± 0.0609 | 0.0638 ± 0.0157 |
| Rak łagodny | 0.4193 ± 0.0554 | 0.0585 ± 0.0126 |
| Przypadek normalny | 0.4386 ± 0.0501 | 0.0527 ± 0.0089 |
| Rak złośliwy – rak łagodny | 0.4037 ± 0.0530 | 0.0664 ± 0.0140 |
| Rak złośliwy – przypadek normalny | 0.4085 ± 0.0641 | 0.0486 ± 0.0112 |
| Rak łagodny – przypadek normalny | 0.4250 ± 0.0556 | 0.0510 ± 0.0100 |

3. Indywidualne systemy klasyfikacyjne

Głównym zadaniem w pracy jest opracowanie systemów komputerowych do rozpoznawania klas chorobowych na podstawie obrazów charakteryzujących zmiany patologiczne w organizmie. Proponowane rozwiązanie bazować będzie na zastosowaniu zespołu klasyfikatorów, tworzonego w różny sposób. Podstawowym narzędziem analizy obrazowej będą głębokie sieci neuronowe typu konwolucyjnego (ang. *Convolutional Neural Networks – CNN*). Ich wykorzystanie będzie przybierać dwie różne formy:

- Bezpośrednie zastosowanie struktury sieciowej CNN do wskazania klasy, do której przypisany będzie obraz wejściowy przy użyciu klasyfikatora typu softmax.
- Wykorzystanie sieci CNN jedynie do generacji deskryptorów numerycznych opisujących obraz wejściowy i wykorzystaniu różnych rozwiązań klasyfikatorów połączonych w zespole wykorzystujących te deskryptory.

W pierwszym przypadku wybór cech diagnostycznych ze zbioru wygenerowanych deskryptorów odbywać się będzie losowo. W drugim przypadku zbiór deskryptorów numerycznych podlegać będzie dodatkowej analizie typu deterministycznego w celu wyłonienia tych cech, które następnie będą podane jako atrybuty wejściowe na klasyczne klasyfikatory, typu maszyna wektorów nośnych (ang. *Support Vector Machine – SVM*) i las losowy drzew decyzyjnych (ang. *random forest – RF*).

W tym rozdziale podane zostaną podstawowe informacje dotyczące różnych aspektów związanych z działaniem tych klasyfikatorów.

3.1. Sieci głębokie CNN – warstwy lokalnie połączone

Głębokie sieci neuronowe i związane z tym metody uczenia głębokiego stanowią obecnie główny trzon uczenia maszynowego prowadząc do istotnego przyspieszenia rozwoju metod sztucznej inteligencji, w tym rozpoznawanie obrazów i mowy, widzenie komputerowe, przetwarzanie danych biomedycznych czy generowanie zupełnie nowych produktów wizyjnych i dźwiękowych.

W pracy wykorzystane zostaną sieci jednokierunkowe CNN [4,20]. Struktura tych sieci jest wielowarstwowa i składa się z dwóch podstawowych części. Główna część to bardzo wiele warstw lokalnie połączonych zawierających ogromną liczbę neuronów o nieliniowości typu ReLU (ang. *Rectified Linear Unit*). Ich główne zadanie to automatyczna generacja deskryptorów numerycznych dobrze opisujących obraz wejściowy

poddany analizie. Dane w poszczególnych warstwach reprezentują kolejno przetwarzane obrazy uporządkowane w formie tensorów trójwymiarowych. Dwa pierwsze parametry tensora reprezentują wymiar x i y obrazu, trzeci liczbę obrazów w danej warstwie. W ogólności, im więcej warstw ma sieć, tym bardziej złożone i abstrakcyjne cechy mogą być wyodrębniane z danych wejściowych. Dzięki temu głębokie sieci neuronowe mają zdolność do uczenia się i odtworzenia bardziej skomplikowanych wzorców wejściowych, co przyczynia się do uzyskiwania lepszych wyników w zadaniach wykonywanych przez te sieci.

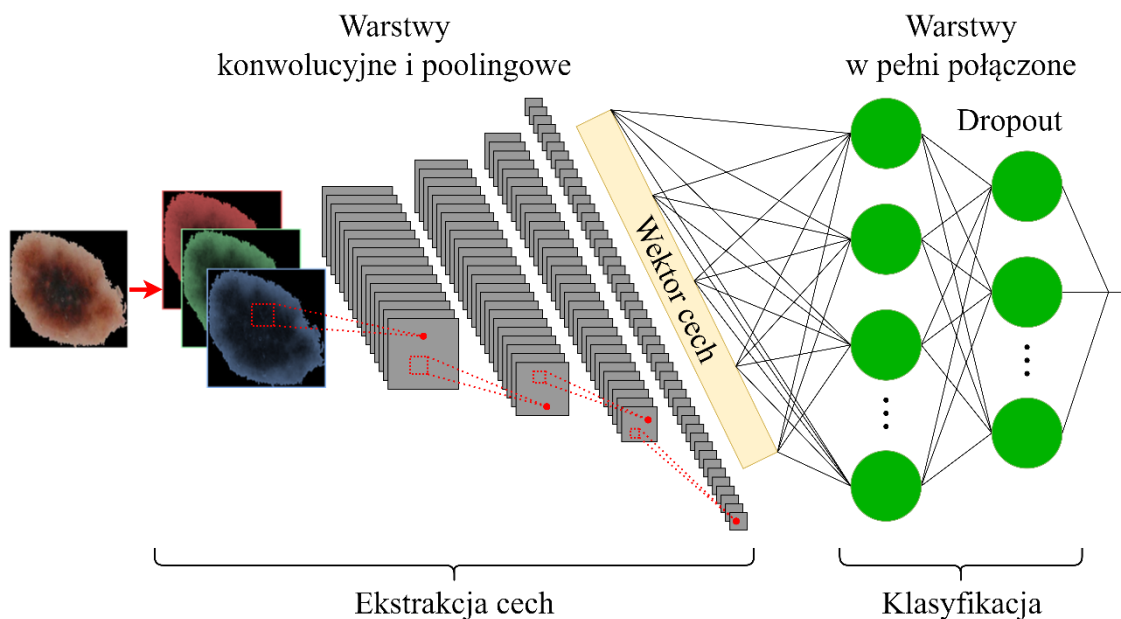
Tensor ostatniej warstwy lokalnie połączonej jest przetworzony w postać wektorową, poprzez operację wypłaszczenia (ang. *flattening*). Wektor ten zawiera zbiór deskryptorów numerycznych odpowiadający aktualnemu obrazowi wejściowemu sieci i stanowi początek drugiej części sieci CNN pełniąc funkcję końcowego klasyfikatora. Może być to sieć jednowarstwowa o pełnym połączeniu między neuronami i uproszczonym algorytmie uczenia lub bardziej złożona struktura oparta na klasycznej architekturze sieci neuronowej bądź innym rozwiązaniu maszynowym.

Na rysunku 3.1 przedstawiono typowy schemat ideowy konwolucyjnej sieci neuronowej. Warstwy konwolucyjne i poolingowe sieci odpowiadają za automatyczną ekstrakcję deskryptorów obrazów wejściowych. Wygenerowane deskryptory w formie obrazów ustawionych w postaci tensorowej wykorzystywane są jako sygnały wejściowe dla następnej warstwy.

Po przejściu sygnałów przez wiele warstw następuje etap wypłaszczenia, czyli zamiany formy tensorowej danych na wektorową, oznaczoną na rysunku jako wektor cech. Elementy tego wektora stanowią atrybuty wejściowe dla klasyfikatora. Liczba neuronów wyjściowych jest równa liczbie rozpoznawanych klas.

W ten sposób konwolucyjne sieci neuronowe pełnią jednocześnie rolę zintegrowanego ekstraktora i selektora cech diagnostycznych z końcowym klasyfikatorem (lub układem regresyjnym w przypadku zadania regresji).

W ogólnej strukturze lokalnie połączonej sieci CNN można wyróżnić warstwy liniowej konwolucji połączone z nieliniową funkcją aktywacji (zwykle typu ReLU) oraz warstwy poolingu których zadaniem jest redukcja wymiarów obrazów w tensorze wyjściowym. W niektórych strukturach sieci CNN (na przykład Alexnet [34]) w strukturze w pełni połączonej stosowana była dodatkowa warstwa ukryta. W obecnych implementacjach unika się tej dodatkowej warstwy łącząc elementy wektora cech bezpośrednio z softmaxem.

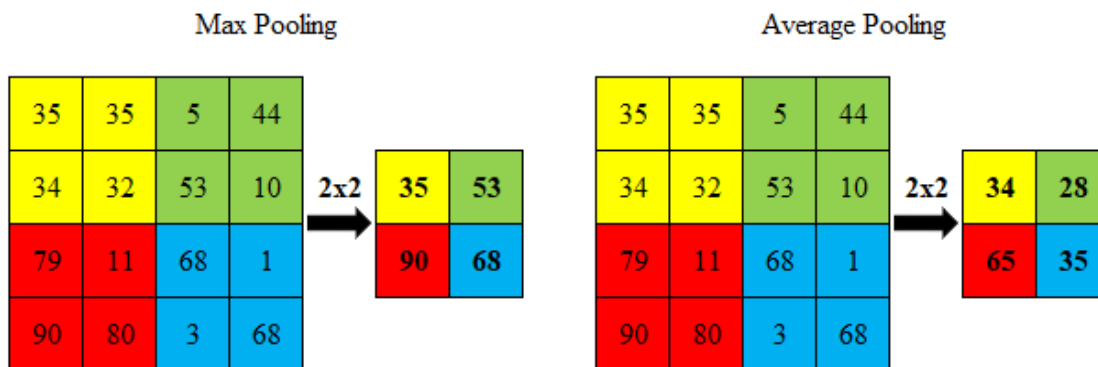


Rys. 3.1 – Schemat ideowy konwolucyjnej sieci neuronowej.

Warstwy konwolucyjne są kluczowym elementem struktury sieci głębokiej. Składają się z zestawu filtrów (tworzących tak zwane jądro przekształcenia), które przesuwają się po obrazie z krokiem (ang. *stride*), wykonując operację konwolucji. Wartości wagi filtra są parametrami i są poddane uczeniu podczas treningu sieci. Każdy filtr wyodrębnia pewne lokalne cechy statystyczne z danych wejściowych takie jak ułożenie krawędzi, zakrzywienia, tekstury itp. Operacja konwolucji z określonym krokiem polega na wykonaniu punktowego iloczynu między wagami filtra a odpowiadającymi mu fragmentami danych obrazu wejściowego. Operacja taka odbywa się równolegle na wielu obrazach poprzedniej warstwy. Następnie wyniki są sumowane, tworząc w efekcie nowy obraz wyjściowy (mapę cech). W wyniku przetworzenia danych przez wiele filtrów powstaje wiele map cech, uporządkowanych w formie tensora. Czasami, aby zachować odpowiednie wymiary przestrzenne danych, stosuje się dodawanie zer wokół brzegów danych wejściowych przed operacją konwolucji. Ten proces zwany jest z angielska „zero-padding”. W warstwach konwolucyjnych po operacji liniowej konwolucji stosuje się funkcję aktywacji, zwykle w postaci ReLU lub softplus, aby uniknąć ujemnych wartości elementów obrazu wyjściowego. Funkcje te definiowane są w postaci:

$$ReLU(x) = \begin{cases} x & \text{dla } x > 0 \\ 0 & \text{dla } x < 0 \end{cases} \quad (3.1)$$

$$softplus(x) = \ln(1 + e^x) \quad (3.2)$$



Rys. 3.2 – Ilustracja operacji Max Pooling i Average Pooling o rozmiarze 2x2.

Warstwa poolingu ma za zadanie zmniejszyć wymiary przestrzenne obrazów wyjściowych po konwolucji. Operacja taka polega na wybraniu informacji statystycznych z określonych regionów obrazu, co pomaga zmniejszyć ilość danych przetwarzanych przez sieć oraz wprowadza pewne odporności na translacje (przesunięcia) w danych wejściowych. Najczęściej stosowaną operacją poolingu jest max pooling (wybór wartości maksymalnej z analizowanego regionu obrazu), choć istnieją też inne rodzaje poolingu, takie jak average pooling (wartość średnia elementów w danym regionie obrazu) czy L2-norm pooling (zastąpienie danego regionu obrazu poprzez normę L2). Operacja poolingu stosuje filtry o dobranych przez użytkownika rozmiarach, na przykład 2x2 czy 3x3. Przykładową operację max pooling oraz average pooling zilustrowano na rysunku 3.2. Warstwy poolingowe są istotnym elementem architektury CNN, pozwalając na redukcję wymiarów obrazu przy zachowaniu cech statystycznych odpowiadających określonym regionom obrazu.

Rozmiary obrazów (mapy cech) w poszczególnych warstwach wpływają bezpośrednio na ilość informacji, która jest przetwarzana w kolejnych warstwach sieci. Ostateczny rozmiar mapy cech (tensora wyjściowego ostatniej warstwy lokalnie połączonej) jest wynikiem kombinacji wymienionych powyżej parametrów to jest wielkości filtrów i kroku ich przesuwania, zero-paddingu oraz parametrów warstw poolingowych a także architektury sieci i rozmiaru danych wejściowych.

Przez architekturę sieci należy rozumieć dobór liczby i układu warstw konwulcyjnych i poolingowych. Sieci o większej liczbie warstw konwulcyjnych mogą wydobywać bardziej złożone wzorce, jednocześnie doprowadzając do mniejszych wymiarów mapy cech. Ważne jest odpowiednie dostosowanie tych parametrów podczas projektowania sieci, aby uzyskać optymalne rozmiary map cech dla konkretnego zadania i danych

wejściowych. Pierwszą ogólnie dostępną strukturą głęboką sieci CNN był Alexnet opracowany przez zespół [34]. Jego pierwotna struktura przedstawiona jest poniżej.

```
1 'data' Image Input 227x227x3 images with 'zerocenter' normalization
2 'conv1' Convolution 96 11x11x3 convolutions with stride [4 4] and padding [0 0]
3 'relu1' ReLU
4 'norm1' Cross Channel Normalization with 5 channels per element
5 'pool1' Max Pooling 3x3 max pooling with stride [2 2] and padding [0 0]
6 'conv2' Convolution 256 5x5x48 convolutions with stride [1 1] and padding [2 2]
7 'relu2' ReLU
8 'norm2' Cross Channel Normalization with 5 channels per element
9 'pool2' Max Pooling 3x3 max pooling with stride [2 2] and padding [0 0]
10 'conv3' Convolution 384 3x3x256 convolutions with stride [1 1] and padding [1 1]
11 'relu3' ReLU
12 'conv4' Convolution 384 3x3x192 convolutions with stride [1 1] and padding [1 1]
13 'relu4' ReLU
14 'conv5' Convolution 256 3x3x192 convolutions with stride [1 1] and padding [1 1]
15 'relu5' ReLU
16 'pool5' Max Pooling 3x3 max pooling with stride [2 2] and padding [0 0]
17 'fc6' Fully Connected 4096 fully connected layer
18 'relu6' ReLU
19 'drop6' Dropout 50% dropout
20 'fc7' Fully Connected 4096 fully connected layer
21 'relu7' ReLU
22 'drop7' Dropout 50% dropout
23 'fc8' Fully Connected 1000 fully connected output layer
24 'prob' Softmax softmax
25 'output' Classification 1000 classes
```

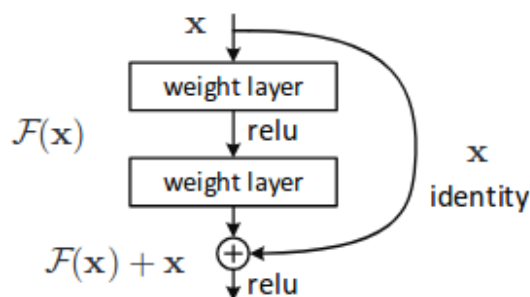
Sieć zawiera 5 warstw konwolucyjnych lokalnie połączonych zawierających podwarstwę konwolucji liniowej połączonej z nieliniową aktywacją ReLU, podwarstwę normalizacyjną oraz podwarstwę pooling. Wektor wypłaszczonego zawiera 4096 elementów (deskryptorów numerycznych obrazu) stanowiących sygnały wejściowe dla klasyfikatora typu softmax. W wersji oryginalnej klasyfikator zawiera jedną warstwę ukrytą (w pełni połączoną) i 1000 neuronów wyjściowych odpowiedzialnych za rozpoznanie maksimum 1000 klas.

Rozmiary mapy cech wpływają na kilka kluczowych aspektów związanych z wykorzystywaniem konwolucyjnych sieci neuronowych. Jednym z nich jest dostępne zaplecze technologiczne. W tej perspektywie wymiary mapy cech przekładają się na wydajność pamięciową i obliczeniową. Większa mapa cech naturalnie wymaga większej ilości zasobów obliczeniowych. Wraz ze wzrostem jej wymiarów, zwiększa się liczba parametrów sieci i operacji konwolucji do wykonania, co może powodować zwiększenie czasu

treningu i inferencji. Jednocześnie dochodzi do większego zużycia pamięci GPU, która wpływa na możliwość równoległego przetwarzania danych i wymusza ograniczenie wielkość danych, które można przetwarzać symultanicznie.

W aspekcie jakościowym kwestia negatywnego lub pozytywnego wpływu rozmiaru mapy cech zależy od problemu, z którym mierzy się sieć. W zbyt małych mapach sieć może utracić cenne informacje i nie nauczyć się odpowiednich cech do wykonywania zadania. Natomiast zbyt duża mapa cech może prowadzić do przeuczenia, gdyż sieć może nauczyć się niestandardowych lub zaszumionych wzorców. W niektórych zadaniach, takich jak klasyfikacja obrazów, większa mapa cech może pomóc w wykrywaniu bardziej złożonych wzorców i cech. Jednakże w innych zadaniach, takich jak segmentacja obrazu, mniejsza mapa cech może być wystarczająca do rozpoznawania podstawowych cech, a dokładność zostanie zrekompensowana przez czas uczenia bądź wykonywania zadania. Optymalny rozmiar mapy cech jest zatem zależny od konkretnej architektury sieci, rozmiarów danych wejściowych i charakterystyki zadania. Dobór odpowiedniego rozmiaru mapy cech jest jednym z kluczowych aspektów, które wpływają na skuteczność i wydajność sieci w konkretnym zadaniu.

Rozwiązanie bardziej złożonych problemów klasyfikacyjnych może wymagać znacznego zwiększenia liczby warstw lokalnie połączonych, co w sposób oczywisty zwiększa złożoność obliczeniową, prowadząc jednocześnie do zaburzenia procesu generacji gradientu, podstawowego elementu w uczeniu gradientowym sieci. Zaburzenie to może przejawiać się w postaci zanikania bądź eksplozji wartości elementów wektora gradientu w procesie uczenia. W efekcie przy dużej liczbie warstw ukrytych obserwuje się nawet wzrost błędu dopasowania w procesie uczenia. Problem ten został rozwiązany przez autorów pracy [52] poprzez wprowadzenie dodatkowego połączenia między-warstwowego o tym samym kierunku przepływu sygnałów, tak zwane połączenie rezydualne.

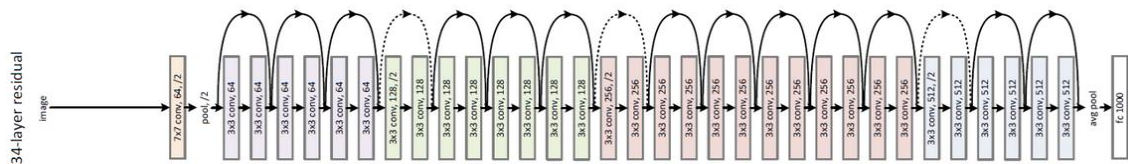


Rys. 3.3 – Struktura bloku rezydualnego w sieci Resnet. Połączenie skrótowe (rezydualne) występuje między dwoma warstwami konwolucyjnymi [52].

Rozwiązanie takie zwane przez autorów Resnet pozwoliło na budowanie sieci o bardzo dużej liczbie warstw sięgających nawet setek. Sposób proponowanego połączenia w sieci Resnet przedstawiony jest na rys. 3.3, w którym operacje sumowania dotyczą wszystkich elementów obrazów poszczególnych warstw [11].

Rozwiązanie polega na zastąpieniu oryginalnego odwzorowania $F(\mathbf{x})$ reprezentowanego przez 2 warstwy konwolucyjne poprzez odwzorowanie $F(\mathbf{x}) + \mathbf{x}$. Oznacza to, że w rzeczywistości wagi sieci adaptowane są w procesie uczenia do odwzorowania resztkowego (residualnego) $F(\mathbf{x}) - \mathbf{x}$.

Przykładowe rozwiązanie struktury sieci Resnet34 przedstawione jest na rys. 3.4. Bloki konwolucji składają się z operacji konwolucji, „batch normalization” oraz funkcji nieliniowej ReLU.



Rys. 3.4 – Przykładowa struktura sieci Resnet34. Połączenie kropkowane występuje, gdy rozmiary wymiary warstwy startowej i docelowej różnią się.

Linie ciągłe dotyczą połączeń międzywarstwowych o tych samych wymiarach mapy cech (obrazów). Połączenie kropkowane dotyczy przypadku, gdy wymiary warstwy startowej i docelowej różnią się (na przykład warstwa startowa 256×256 , warstwa docelowa 128×128). Zmniejszanie wymiaru między warstwami uzyskane jest dzięki parametrowi *stride* (krok przesunięcia filtra konwolucyjnego) równemu 2. W przypadku odwrotnym, gdy wymiary bloku startowego są mniejsze niż docelowego, wymiary \mathbf{x} uzupełnia się przez dodanie odpowiedniej liczby zer (tzw. *zero padding*). Dzięki takiej modyfikacji struktury sieci możliwe jest zbudowanie bardzo dużych sieci. Przykładem jest sieć Resnet152.

3.2. Sieci głębokie CNN – klasyfikator końcowy

Warstwy lokalnie połączone służą do generacji deskryptorów numerycznych obrazów wejściowych. Zorganizowane w formie tensorów zostają w ostatniej warstwie przekształcone w postać wektorową. Zamiana zbioru obrazów ostatniego tensora w jeden wektor reprezentujący cechy wejściowego obrazu oryginalnego może odbywać się na wiele sposobów.

Najprostszym, stosowanym między innymi w pierwszej sieci CNN Alexnet, jest zwykle wypłaszczenie macierzy w postać wektorową. W takim przypadku wymiar wektora cech jest równy iloczynowi wymiarów obrazów tworzących tensor wyjściowy przez ich liczbę. W przypadku sieci Alex net jest to $256 \times 4 \times 4 = 4096$ (256 obrazów o wymiarze 4×4).

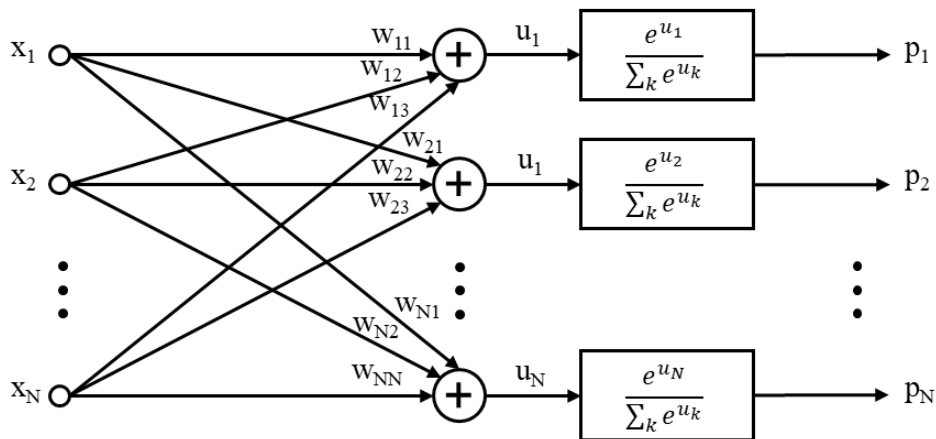
Inna metoda polega na poddaniu obrazów ostatniej warstwy procesowi redukcji poprzez operacje uśredniania filtracyjnego lokalnych rejonów obrazu bądź zastosowania odpowiednio zdefiniowanej operacji pooling i następnie wypłaszczeniu. Typowym rozwiązaniem aktualnie jest również zastosowanie „global average pooling” do wszystkich obrazów tensora wyjściowego. W tej operacji każdy obraz wyjściowy tensora zastępowany jest jedną uśrednioną wartością stopni jasności pikseli a wymiar wektora cech jest równy liczbie tych obrazów.

W efekcie wypłaszczenia dane wejściowe są reprezentowane przez wektor cech, który następnie służy jako wejście dla końcowego klasyfikatora (bądź regresora w przypadku zadania regresji). Rozwiązania klasyfikatora mogą przyjmować różne formy. Typowym rozwiązaniem w sieciach CNN jest tak zwany softnet, którego głównym elementem jest funkcja softmax.

3.2.1. Klasyfikator typu softmax

Softmax reprezentuje uproszczone podejście do problemu klasyfikacji dostosowane pod względem obliczeniowym do skali złożoności obliczeniowej. Z punktu widzenia architektury jest strukturą jednowarstwową neuronów w pełni połączonych (elementy wektora cech są połączone ze wszystkimi neuronami wyjściowymi, a liczba neuronów wyjściowych jest równa liczbie klas). Rys. 3.5 prezentuje strukturę softnetu w której główną rolę pełni softmax.

Dla polepszenia zdolności generalizacyjnych w połączeniach softnetu stosuje się tak zwaną operację „dropout”, czyli losowe zerowanie pewnych połączeń wagowych między wektorem cech i neuronami wyjściowymi w każdej iteracji. Jest to metoda selekcji wykorzystująca losowość doboru cech diagnostycznych. Daje to efekt zmniejszenia złożoności sieci i polepsza jej zdolności generalizacji. Kodowanie klas jest binarne. W procesie uczenia pozycja jedynek w wektorze zadany \mathbf{d} reprezentuje przynależność do danej klasy, zero – brak przynależności).



Rys. 3.5 – Struktura połączeń softnetu. Sygnały x_i reprezentują elementy wypłaszczonego wektora cech, liczba neuronów wyjściowych jest równa liczbie klas, sygnały p_i reprezentują wartości prawdopodobieństwa przynależności danych wejściowych do odpowiedniej klasy. Zwycięża klasa o największej wartości prawdopodobieństwa.

Ważnym elementem rozwiązania jest definicja funkcji celu (podlegającej minimalizacji) wykorzystująca tak zwaną entropię krzyżową (ang. *cross-entropy function*). W tym podejściu dobór wag softmaxu stanowi zintegrowaną część z doбором parametrów warstw lokalnie połączonych odpowiedzialnych za generację cech. Wektor podlegający adaptacji w procesie uczenia zawiera zarówno wszystkie parametry filtrów konwulucyjnych jak i wagi połączeń softnetu. Jest to więc bardzo długi wektor.

Uczenie sieci CNN odbywa się poprzez minimalizację gradientową funkcji celu. Funkcja ta bazuje na prawdopodobieństwie przynależności danych wejściowych do określonej klasy. Prawdopodobieństwo to definiowane jest w postaci [20]

$$f(u_i) = \frac{\exp(u_i)}{\sum_{j=1}^M \exp(u_j)} \quad (3.3)$$

gdzie zmienna u_i jest sumą wagową sygnałów dopływających do i -tego neuronu wyjściowego $u_i(\mathbf{x}) = \sum_j w_{ij}x_j + w_0$. Funkcja $f(u)$ nosi nazwę softmax. Największa wartość prawdopodobieństwa odpowiadająca neuronowi wyjściowemu wyznacza ostateczną przynależność do jednej spośród M klas.

Entropijna definicja funkcji celu dla jednego rekordu danych może być zapisana w postaci

$$E = -\sum_{i=1}^M d_i \ln(f(u_i)) \quad (3.4)$$

Przy M klasach tylko jedna z nich jest prawdziwa dla aktualnych danych wejściowych, a etykieta tej klasy $d = 1$. Pozostałe klasy mają etykietę $d = 0$. Stąd w procesie uczenia tylko jedna klasa wskazana przez sieć jako prawdziwa bierze bezpośredni udział w sumie (3.4) przy definiowaniu funkcji błędu ze względu na pojedynczą wartość $d = 1$. Funkcja

celu zdefiniowana powyżej upraszcza się do postaci, w której u_p odpowiada danej klasie uznanej za „pozytywną” ($d_i = 1$)

$$E = -\ln(f(u)_i) = -\ln\left(\frac{\exp(u_p)}{\sum_{j=1}^M \exp(u_j)}\right) \quad (3.5)$$

Pozostałe składniki funkcji celu we wzorze (3.4) względem klas “negatywnych” ($d_i = 0$) są zerowane. Wprawdzie tylko klasa o $d = 1$ bierze bezpośredni udział w tworzeniu funkcji celu, ale wektor gradientu funkcji celu jest uzależniony również od klas „negatywnych” ze względu na postać (3.4) funkcji celu. W obliczeniach pochodnej funkcji celu względem sygnałów sumacyjnych uzyskuje się następujące wyrażenia [20,44]

$$\frac{\partial}{\partial u_n} \left(-\ln\left(\frac{\exp(u_p)}{\sum_{j=1}^M \exp(u_j)}\right) \right) = \frac{\exp(u_n)}{\sum_{j=1}^M \exp(u_j)} \quad (3.6)$$

$$\frac{\partial}{\partial u_p} \left(-\ln\left(\frac{\exp(u_p)}{\sum_{j=1}^M \exp(u_j)}\right) \right) = \frac{\exp(u_p)}{\sum_{j=1}^M \exp(u_j)} - 1 \quad (3.7)$$

Składniki wektora gradientu względem poszczególnych wag w_i klasyfikatora są określane wyrażeniem (niezależnie do tego czy u dotyczy przypadku „pozytywnego” czy „negatywnego”

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial u} \frac{\partial u}{\partial w_i} = \frac{\partial E}{\partial u} x_i \quad (3.8)$$

Składniki gradientu odpowiadające wagom filtrów (w części sieci lokalnie połączonej) określa się klasyczną metodą propagacji błędu z warstwy wyjściowej na warstwy poprzedzające.

Stosowane są różne metody, z których do najczęściej używanych należy stochastyczny algorytm największego spadku z momentem rozpędowym (ang. *Stochastic Gradient Descent with Momentum – SGDM*) [20,44] oraz ADAM (skrót of ang. *ADaptive Momenet estimation*) [33] lub ADAMW (algorytm ADAM z regularyzacją wagową). W przypadku SGDM adaptacja elementów wektora \mathbf{w} odbywa się iteracyjnie według wzoru

$$\mathbf{w}(k) = \mathbf{w}(k-1) - \eta \mathbf{g}(k-1) + \alpha [\mathbf{w}(k-1) - \mathbf{w}(k-2)] \quad (3.9)$$

We wzorze tym ostatni człon reprezentuje moment rozpędowy, w którym α jest współczynnikiem tego momentu przyjmowanym z przedziału $[0, 1]$ natomiast η jest krokiem uczenia dobieranym zwykle w sposób adaptacyjny odpowiednio do zmian wartości minimalizowanej funkcji celu. W procesie uczenia używa się zwykle losowo wybranego podzbioru uczącego (tzw. *minibatch*).

W metodzie ADAM główna różnica polega na tym, że każdy element w wektora \mathbf{w} dobiera się niezależnie od innych, a adaptacja uzależniona jest od momentu statystycznego gradientu pierwszego i drugiego rzędu, definiowanych w kolejnych iteracjach według wzorów

- moment pierwszego rzędu

$$m_w(k) = \beta_1 m_w(k-1) + (1 - \beta_1) g_w(k) \quad (3.10)$$

- moment drugiego rzędu

$$v_w(k) = \beta_2 v_w(k-1) + (1 - \beta_2) g_w^2(k) \quad (3.11)$$

We wzorach tych g_w jest elementem gradientu odpowiadającym jednemu elementowi w , natomiast β_1 i β_2 są hiperparametrami dobieranymi przez użytkownika, zwykle w zakresie [0.9 – 0.99]. Definiuje się następnie estymaty nieobciążone obu momentów,

$$\hat{m}_w(k) = \frac{m_w(k)}{1 - \beta_1^k} \quad (3.12)$$

$$\hat{v}_w(k) = \frac{v_w(k)}{1 - \beta_2^k} \quad (3.13)$$

i przeprowadza adaptację wag według wzoru

$$w(k) = w(k-1) - \eta \frac{\hat{m}_w(k)}{\sqrt{\hat{v}_w(k) + \epsilon}} + \alpha [w(k-1) - w(k-2)] \quad (3.14)$$

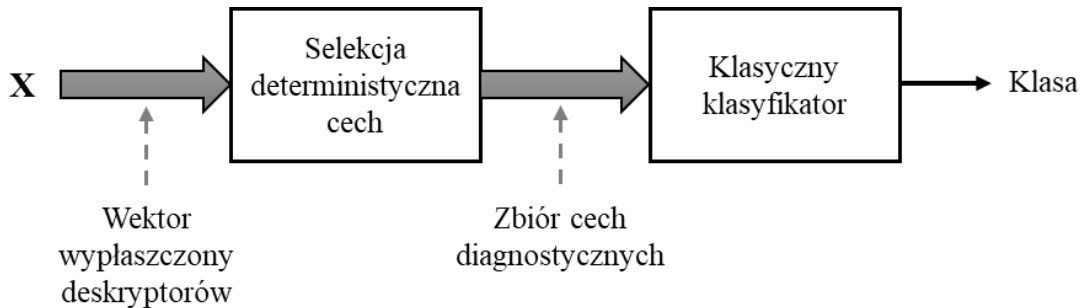
W wielu rozwiązaniach ostatni człon reprezentujący moment rozpędowy jest eliminowany. W pracy używany będzie algorytm ADAM jako nieco bardziej skuteczny w minimalizacji funkcji celu.

3.2.2. Klasyczne klasyfikatory zastosowane w pracy

W klasycznych rozwiązaniach sieci CNN w zadaniach klasyfikacji standardem jest zastosowanie funkcji softmax. W procesie uczenia softmaxu atrybuty wejściowe są dobierane losowo spośród deskryptorów reprezentowanych przez wektor wypłaszczony (typowa wartość to 50% dropout) w każdej iteracji. Jest to najprostsze, a jednocześnie skuteczne, rozwiązanie redukcji wymiarowości wektora wejściowego podawanego na softmax, poprawiające zdolność generalizacji sieci.

Można również zastosować inne podejście do selekcji tych atrybutów, wykorzystujące wiele znanych algorytmów łącząc je z klasycznymi rodzajami klasyfikatorów, na przykład SVM czy drzewa losowe. Tego typu klasyfikatory mogą być bardziej efektywne

pod względem generalizacji niż softmax, gdyż operują znacznie mniejszą liczbą atrybutów wejściowych wyselekcjonowanych przez deterministyczny algorytm selekcji. W związku z powyższym struktura części klasyfikacyjnej sieci CNN będzie miała postać jak na rysunku 3.6.



Rys. 3.6 – Alternatywne rozwiązanie klasyfikatora końcowego wykorzystujące deterministyczny algorytm selekcji cech i klasyczne rozwiązanie klasyfikatorów.

Spośród wielu rozwiązań klasyfikatorów klasycznych do najbardziej efektywnych należy maszyna wektorów nośnych (Support Vector Machine – SVM) oraz las losowy drzew decyzyjnych. Oba klasyfikatory działają na podstawie różnych mechanizmów stawiących niezależne rozwiązania, bardzo przydatne przy tworzeniu zespołu klasyfikatorów.

Klasyfikator SVM

Klasyfikator SVM stworzony przez prof. Vapnika [51,67] stanowi strukturę nieopodobną i stosuje specjalny sposób uczenia sprowadzający się do programowania kwadratowego. W procesie uczenia parametry sieci dobiera się maksymalizując margines separacji między dwoma klasami, przypisanymi danym (\mathbf{x}_i, d_i) , przy zachowaniu jak najmniejszego błędu klasyfikacji na danych uczących.

Sieci SVM są strukturami dwuwarstwowymi, w których warstwa ukryta stanowi nieliniową transformację danych wejściowych tworząc zbiór danych liniowo separowalnych, a warstwa wyjściowa składa się z pojedynczego neuronu, którego sygnał w klasyfikacji przybiera wartość jeden (rozpoznana klasa) lub zero (brak przynależności klasowej). Dodatkową zaletą SVM jest możliwość zastosowania różnorodnych funkcji jądra, w tym gaussowskiej czy wielomianowej dobieranej odpowiednio do aktualnego zadania. Proces uczenia jest bardzo szybki dzięki wprowadzeniu funkcji skalarnej jądra i uproszczeniu optymalizacji do programowania kwadratowego.

Funkcja celu w klasyfikatorze SVM jest definiowana całkowicie odmiennie w stosunku do klasycznej sieci CNN i koncentruje się na maksymalizacji szerokości

marginesu separacji między dwoma klasami przy zachowaniu jak najmniejszych błędów klasyfikacji danych uczących. Typowe podejście startuje z zadania pierwotnego definowanego w postaci [67]

$$\min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^p \xi_i \quad (3.15)$$

przy ograniczeniach funkcyjnych i kostkowych określonych dla $i = 1, 2, \dots, p$

$$\begin{aligned} d_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) &\geq 1 - \xi_i \\ \xi_i &\geq 0 \end{aligned} \quad (3.16)$$

gdzie \mathbf{w} reprezentuje wektor adaptowanych wag, b jest wartością polaryzacji, $\phi(\mathbf{x}_i)$ funkcją nieliniowego rzutowania wektora oryginalnego \mathbf{x}_i w inną rozszerzoną przestrzeń funkcyjną, C – stała regularyzacji, a ξ_i zmienną dopełniającą określającą jak dalece można zmniejszyć efektywnie szerokość marginesu separacji. Po zastosowaniu funkcjonału Lagrange’a zadanie pierwotne przekształcane jest w problem dualny względem mnożników α_i odpowiedzialnych za spełnienie ograniczeń klasowych dla danych uczących. Rozwiązanie to operuje funkcją jądra $K(\mathbf{x}, \mathbf{x}_i) = \phi^T(\mathbf{x})\phi(\mathbf{x}_i)$, spełniająca warunki Mercera [51,67]. W efekcie po rozwiązaniu zadania dualnego względem mnożników Lagrange’a sygnał wyjściowy $y(\mathbf{x})$ sieci SVM określony jest zależnością [67]

$$y(\mathbf{x}) = \sum_{i=1}^{N_{sv}} \alpha_i d_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (3.17)$$

w której tylko mnożniki α_i różne od zera mają wpływ na rozwiązanie. Definiują one zbiór N_{sv} wektorów \mathbf{x}_i zwanych wektorami nośnymi (ang. *support vectors*). Wartość $y(\mathbf{x}) > 0$ oznacza przynależność wektora \mathbf{x} do danej klasy, a $y(\mathbf{x}) < 0$ brak tej przynależności. Przy większej liczbie klas stosuje się wiele 2-klasowych sieci SVM a ostateczna decyzje przynależności klasowej podejmuje się poprzez głosowanie większościowe. Stosowanych jest przy tym wiele różnych strategii, choć najczęściej wykorzystywane są jeden-przeciw-jednemu lub jeden-przeciw-reszcie [67].

Ważną zaletą sieci SVM jest dobra zdolność generalizacji. Badania wykazały, że sieć SVM jest stosunkowo mało wrażliwa na liczebność danych uczących, co ma szczególne znaczenie w rozwiązaniu problemów medycznych poddanych analizie w rozprawie, gdzie liczba rekordów danych jest bardzo ograniczona.

Istotnym elementem jest sprowadzenie uczenia sieci do zadania programowania kwadratowego dzięki czemu punkt startu w uczeniu nie ma większego wpływu na wynik. Tym nie mniej jakość rozwiązania uzależniona jest od przyjętych arbitralnie wartości "hiperparametrów", w tym szerokości funkcji gaussowskiej σ (najczęściej stosowanej

w rozwiązaniu) i wartości parametru regularyzacyjnego C . Wartości te są dobierane zwykle metodą prób przy użyciu specjalnego zestawu danych weryfikujących, wydzielonych ze zbioru uczącego metodą walidacji krzyżowej.

Drzewa decyzyjne

Drzewa decyzyjne stanowią graficzny schemat podejmowania decyzji klasyfikacyjnej. Oryginalny zbiór danych jest dzielony na dwa podzbiory według progu wartości jednej (wybranej dla każdego węzła) zmiennej spośród zbioru zmiennych wejściowych x_i . Wszystkie rekordy danych dla których wartość tej zmiennej jest mniejsza niż przyjęty próg są kierowane na lewo, pozostałe na prawo. Istotną sprawą jest dobór zmiennej oraz wartości progu z nią związaną. W procesie uczenia drzewa analizowane są wszystkie możliwości i wybiera tę, która zapewnia najmniejszą wartość łącznego rozproszenia klasowego tych dwu powstałych podzbiorów.

Rozproszenie klasowe może być definiowane na różne sposoby, przy czym najczęściej używana jest miara Giniego $I(A)$ lub entropijna $E(A)$, definiowane odpowiednio

$$I(A) = 1 - \sum_{k=1}^M r_k^2 \quad (3.18)$$

$$E(A) = - \sum_{k=1}^M r_k \log_2 r_k \quad (3.19)$$

We wzorze tym A oznacza zbiór rekordów danych, M liczbę klas, a r_k proporcję wystąpień k -tej klasy w zbiorze A . Jeśli analizowany obszar zawiera jedynie dane należące do jednej klasy wówczas obie miary zarówno Giniego jak i entropijna, są równe zero.

Przy podziale danych ze zbioru A na dwa podzbiory A_1 i A_2 łączna wartość każdej z miar jest sumą wagową miar obu podzbiorów uwzględniającą populację obu podzbiorów A_1 oraz A_2 i obliczana według identycznego wzoru dla miary Giniego i entropijnej

$$I(A) = \frac{n_{A_1}}{n_A} I(A_1) + \frac{n_{A_2}}{n_A} I(A_2) \quad (3.20)$$

Zmienne n_{A_1} i n_{A_2} oznaczają liczbę elementów zawartych w obszarze odpowiednio A_1 i A_2 , natomiast n - całkowitą liczbę elementów w obu obszarach łącznie ($n = n_{A_1} + n_{A_2}$). Proces podziału zbiorów ulega zatrzymaniu, gdy aktualny zbiór rekordów jest pusty lub zawiera obiekty należące wyłącznie do jednej klasy decyzyjnej (łączna miara zanieczyszczeń klasowych równa zero).

Pojedyncze drzewo decyzyjne jest stosunkowo słabym klasyfikatorem (ang. *weak classifier*) o ograniczonej zdolności generalizacji. Poprawa jego działania jest możliwa na różne sposoby. Jednym z nich jest przycinanie drzewa. Mniejsze drzewo wprowadzie

proceeds to a deterioration of the accuracy of the action on the training set, but usually results in an increase in accuracy on the testing set not participating in the learning process.

The most effective method of improving generalization ability is the application of a large number of decision trees performing the same classification task. A collection of such weak classifiers is integrated through majority voting. Improving the generalization ability of the system.

On this basis, an original variant of the ensemble of decision trees called a random forest [5]. The ensemble consists of many parallel decision trees. Each tree is trained on a different randomly selected subset of the training data, so the structure of each tree is different and each solution is characterized by a certain degree of independence. Usually, about 2/3 of the data is used for training, and the remaining 1/3 forms a test set. K decision trees are created in parallel, and the available data is split into training and testing sets for each tree randomly.

In solving the ensemble problem, a random forest is used at each level of the decision. The data is split into subsets, and instead of one variable, many (m) variables are used, where $m < N$. The selection of these decision variables for each node is random. Usually, the final decision is based on the majority vote of the m variables. The decision is based on the quality of the attributes tested in the node. The procedure is repeated at each level of the decision. In the end, fully grown trees are formed. The final classification result is determined by the majority vote of all trees in the forest for the same input vector x .

Random forest is currently considered a very effective classification method, comparable to the best solutions of classifiers trained in machine learning. Many implementations of such an ensemble exist, for example Extra Random Forest (ERF), also known as Extra Trees (ET) [44].

3.3. Deterministic methods of feature selection for diagnostic purposes

Classical solutions for classifier feature selection require control of the complexity of the solution to achieve the best generalization ability of the system. Therefore, there is a need to select the most relevant features for such solutions. A set of descriptors is generated

przez sieć CNN obejmuje zwykle tysiące zmiennych, które powinny podlegać selekcji dla wyłonienia ich zredukowanej liczby, najlepiej powiązanych z decyzją klasyfikacyjną systemu. Wyselekcjonowane deskryptory tworzą zbiór, który nazywać będziemy cechami diagnostycznymi, lub inaczej cechami procesu.

Dobry zbiór cech powinien spełniać określone warunki, w tym w miarę jednoznacznie charakteryzować analizowany proces. Wartości określonej cechy obserwowane dla różnych rekordów danych należących do tej samej klasy powinny być podobne i różnić się istotnie dla rekordów należących do różnych klas. Powinny być przy tym dobrze skorelowane z klasą do której należy dany rekord, być niezależne od siebie, a liczba cech charakteryzujących zbiór danych powinna być jak najbardziej ograniczona dla uniknięcia nadmiernie rozbudowanej struktury klasyfikatora. Koniecznym warunkiem jest normalizacja (na przykład standaryzacja statystyczna czy min-max) wartości wyselekcjonowanych cech, aby uniknąć problemu dominacji cech o największych wartościach względem pozostałych.

Powstało bardzo wiele rozwiązań metod selekcji [22,44] uwzględniających różne mechanizmy powiązań między atrybutami wejściowymi i zmienną decyzyjną dotyczącą przynależności klasowej. W ogólności wyróżnić można metody dostosowane do oceny pojedynczej cechy odseparowanej od zbioru pozostałych bądź ocena jakości danej cechy działającej w otoczeniu innych. Należą do nich metody filtracyjne, opakowane oraz wbudowane [44].

W przypadku pierwszej z nich mierzy się znaczenie cech poprzez ich korelację ze zmienną wyjściową procesu (decyzja klasyfikacyjną). Przykładem jest test metodą dyskryminacyjną Fishera, statystyką chi kwadrat, miara informacyjna cechy, współczynnik korelacji między cechą i zmienną decyzyjną [44].

W metodach opakowanych uwzględnia się wartość diagnostyczną cech w powiązaniu z modelem. Do tej grupy metod należą między innymi: rekursywna eliminacja cech bazująca zwykle na liniowej sieci SVM, sekwencyjna eliminacja cech czy algorytm genetyczny [44].

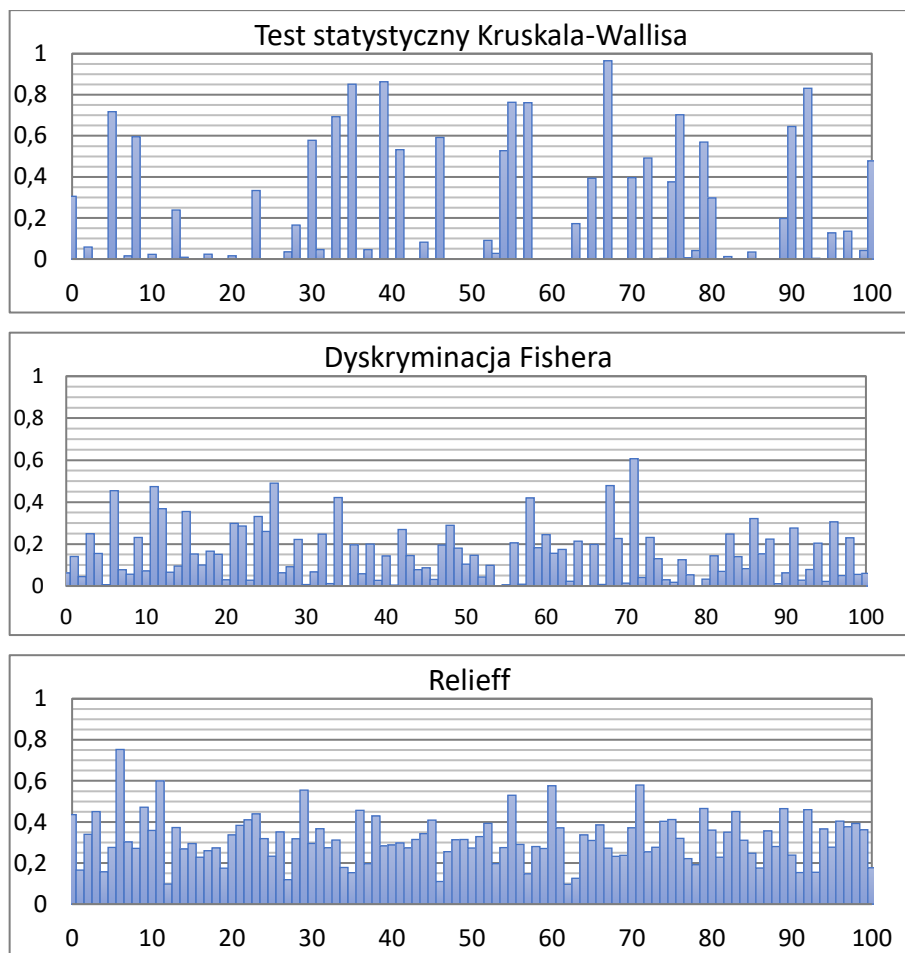
W metodach opakowanych, selekcja jest wbudowana bezpośrednio w algorytm uczenia modelu. Przykładem takiej metody jest drzewo decyzyjne, w którym w podziale danych na każdym poziomie decyzyjnym dokonuje się automatycznie wyboru zmiennej decyzyjnej spośród wielu stanowiących zbiór atrybutów wejściowych. Jest to więc automatycznie dokonująca się selekcja cech na poszczególnych etapach dzielenia podzbiorów.

Każda z opracowanych metod selekcji stosuje inne mechanizmy oceny wartości diagnostycznej cechy. Ich przydatność może być różna i jest uzależniona od rodzaju problemu podlegającego rozwiązaniu. W wyniku wstępnych eksperymentów numerycznych przeprowadzonych na bazie obrazów medycznych wyselekcjonowano ograniczoną liczbę metod selekcji. Przy wyborze cech uwzględniono ich skuteczność na zbiorach walidacyjnych jak również złożoność obliczeniową. W wyniku tych eksperymentów wyselekcjonowano następujące metody selekcji:

- Metoda dyskryminacyjna Fishera (FD) [12,44],
- T-test Studenta (T-test) [12,38,59],
- Test statystyczny Kolmogorova-Smirnova (KS) [38,59],
- Test statystyczny Kruskala-Wallisa (KW) [38,59],
- Metoda korelacji danych z klasą (CDC) [60],
- Selekcja krokowa (SWF) [23],
- Metoda analizy najbliższych sąsiadów (NNA) [66],
- Metoda Relieff (REL) [49].

Wybrane metody reprezentują różne mechanizmy podejmowania decyzji, w tym należące do klas filtracyjnych, opakowaniowych i wbudowanych. Metody FD, T-test, KS, KW, CDC oceniają indywidualnie każdą cechę nie uwzględniając współpracy między nimi. Inne, w tym SWF, NNA, REL oceniają cechy działające i współpracujące ze sobą w określonych podzbiorach tworzonych ad-hoc. Metody T-test, KS, KW bazują na wynikach hipotez statystycznych. Z kolei metody REL, NNA, FD, SWF reprezentują specjalne formuły definiujące powiązania zmiennych wejściowych z atrybutami wyjściowymi (przynależnością klasową). Cechy o największym powiązaniu uważane są za najbardziej wartościowe.

Zbiór wejściowy deskryptorów podlegających selekcji stanowią elementy wektora wypłaszczonego generowanego przez sieć głęboką CNN. Liczba takich deskryptorów zależy od rozwiązania struktury sieciowej. Na przykład w przypadku sieci Alexnet jest równa 4096. Przykładowy wynik działania trzech metod selekcji (test Kruskal-Wallisa, metoda Fishera i metoda Relieff) dla pierwszych 100 spośród 4096 deskryptorów wygenerowanych przez sieć Alexnet w zbiorze obrazów czerniaka przedstawiony jest na rys. 3.7. Widoczne jest duże zróżnicowanie wyników dotyczących poszczególnych cech zależnie od zastosowanej metody. Podobne zróżnicowanie składu wybranych cech występuje przy zastosowaniu innych metod selekcji.



Rys. 3.7 – Wyniki działania mechanizmów selekcji cech metodą testu statystycznego Kruskal-Wallisa, Fishera i Relieff dla pierwszych 100 deskryptorów numerycznych dla obrazów czerniaka.

Następnym krokiem jest ustawienie deskryptorów w kolejności uzależnionej od wartości cechy diagnostycznej (od największej do najmniejszej). W efekcie uzyskuje się uporządkowane zbiory cech dla każdej z wymienionych 8 metod selekcji. Analiza wartości diagnostycznych poszczególnych cech pozwoliła ustalić ich ograniczoną liczbę na 300 branych pod uwagę w dalszych badaniach. Wszystkie wartości wygenerowanych zbiorów podlegają normalizacji w ramach danego zbioru. Wykorzystano przy tym metodę min-max, w której każda wartość oryginalna podlegała transformacji zgodnie ze wzorem

$$x_i := \frac{x_i - x_{i_min}}{x_{i_min_max}} \quad (3.21)$$

w którym x_{i_min} oraz x_{i_max} oznaczają odpowiednio wielkości minimalne i maksymalne deskryptora x_i w zbiorze wyselekcjonowanym przez daną metodę selekcji.

Dla porównania wyników działania różnych mechanizmów selekcji wartości wyselekcjonowanych 300 najważniejszych cech poddano ocenie statystycznej przy zastosowaniu podstawowych miar, takich jak wartość średnia, mediana, odchylenie

standardowe, różnica między percentylem 75 i 25 (IQR), skośność i kurtoza. Wyniki przedstawiono w tabeli 3.1. W ramach poszczególnych elementów statystyki można zaobserwować dość istotne różnice. Dotyczy to zwłaszcza mediany, skośności i kurtozy.

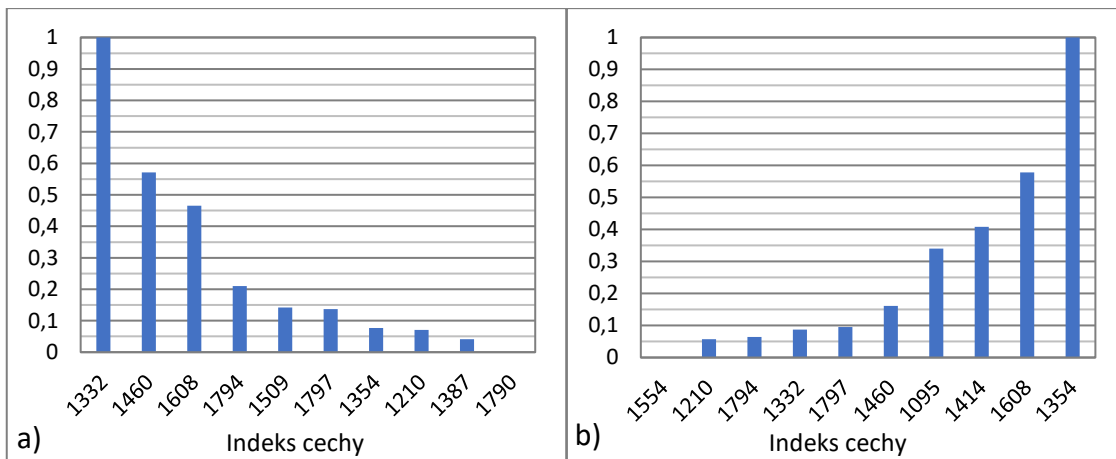
Tabela 3.1 Wielkości statystyczne charakteryzujące 300 najważniejszych cech diagnostycznych obrazów melanomy wyselekcjonowanych przez poszczególne metody selekcji.

| Metoda selekcji | Wartość średnia | Mediana | Odchylenie standardowe | IQR | Skośność | Kurtoza |
|-----------------|-----------------|---------|------------------------|--------|----------|---------|
| SWF | 0.3763 | 0.3196 | 0.2042 | 0.1422 | 1.0206 | 4.2161 |
| T-test | 0.2536 | 0.1879 | 0.2094 | 0.3322 | 0.9657 | 3.4561 |
| KS | 0.2067 | 0.09598 | 0.2598 | 0.2982 | 1.5615 | 4.7850 |
| KW | 0.2040 | 0.0802 | 0.2627 | 0.2821 | 1.3689 | 3.7128 |
| CDC | 0.2101 | 0.0876 | 0.2674 | 0.3553 | 1.4871 | 4.3775 |
| FD | 0.2612 | 0.1963 | 0.2105 | 0.3429 | 0.9461 | 3.3415 |
| NNA | 0.2601 | 0.2345 | 0.2462 | 0.4368 | 0.6830 | 2.7035 |
| REL | 0.1950 | 0.2173 | 0.2080 | 0.3751 | 0.5871 | 2.8812 |

Wszystkie metody selekcji pozwalają na ustawienie wyselekcjonowanych cech według ich ważności definiowanej odpowiednio dla danej metody. Typowy przykład takiego uszeregowania przedstawia rys. 3.8 dla dwóch wybranych metod (FD i KW). W metodzie dyskryminacyjna Fishera większe wartości parametrów oznaczają bardziej znaczące cechy natomiast test statystyczny Kruskala-Wallisa wraz ze wzrostem ważności przypisuje cechom mniejszą wartość parametru.

Niezależnie od sposobu sortowania, zmienność wartości diagnostycznej deskryptorów w obu metodach jest typu monotonicznego, choć niejednakowa. Stąd wynika dodatkowy problem wyboru miejsca cięcia przy wyborze tych cech ze zbioru deskryptorów. Zbyt mała liczba cech prowadzi do niedouczenia sieci, a zbyt duża do przeuczenia i pogorszenia zdolności generalizacyjnych. Problem wyboru punktu odcięcia został rozwiązany w pracy poprzez przeprowadzenie wstępnych eksperymentów na zbiorze danych walidacyjnych.

W efekcie takiej procedury zbiory cech wyselekcjonowane poprzez 8 zastosowanych metod selekcji stanowią atrybuty wejściowe dla dwu klasycznych klasyfikatorów: SVM i lasu losowego. Kojarząc każdy zbiór z tymi dwoma klasyfikatorami uzyskuje się 16 różniących się rozwiązań klasyfikatorów, które mogą być sprzężone w zespół.



Rys. 3.8 – Przykład uszeregowania ważności 10 najlepszych cech diagnostycznych dla selekcji metodą Fishera (a) i Kruskal-Wallisa (b).

4. Zespoły klasyfikatorów

Istnienie wielu rozwiązań sieciowych stwarza możliwość połączenia ich w zespół, odpowiedzialny za wypracowanie ostatecznego werdyktu klasyfikacyjnego. Agregacja (integracja, fuzja) wielu rozwiązań uzyskanych niezależnie stwarza możliwość uzyskania bardziej dokładnych wyników rozpoznania.

Niezależność działania poszczególnych jednostek tworzących zespół jest kluczowym warunkiem dobrego działania zespołu. Wobec trudności z teoretycznym uzasadnieniem niezależności wymaga się, aby poszczególne rozwiązania generowane przez członków zespołu były różniące się między sobą, co stwarza perspektywę ich efektywnego współdziałania w generowaniu decyzji ostatecznej zespołu.

4.1. Zasada działania i sposoby tworzenia zespołów klasyfikatorów

Niezależność działania klasyfikatorów można zwiększyć na wiele różnych sposobów. Podstawową (powszechnie stosowaną) jest technika zwana „bagging”, czyli losowy przydział danych uczących dla każdego członka zespołu. W efekcie każdy model klasyfikatora trenowany jest na różniących się podzbiorach danych, zwracając uwagę na nieco inne aspekty procesu reprezentowane przez te podzbiory [12]. Inna technika stosuje tak zwany boosting (podział wzmacniany danych uczących) [44]. Dane klas trudno rozpoznawalnych mają wyższą wagę (priorytet) przy losowaniu, w efekcie klasyfikator zasilany takimi danymi jest lepiej ukierunkowany w klasyfikacji trudno rozpoznawalnych klas. Są stosowane różne techniki agregacji zespołu w takich systemach, na przykład ADA boosting, czy gradient boosting [44].

Należy zauważyć, że fuzja wyników poszczególnych jednostek zespołu stwarza szansę kompensacji pewnych błędów i poprawy dokładności tak stworzonego systemu klasyfikacyjnego. Poszczególne rozwiązania członków zespołu nie mogą znacząco różnić się co do jakości. Najsłabsze ogniwo systemu może pogorszyć działanie całego zespołu względem najlepszego jej członka. Według pracy [36] prawdopodobieństwo poprawnej klasyfikacji (prognozowana dokładność) w zespole M niezależnych klasyfikatorów o jednakowej dokładności P można wyrazić wzorem [36]

$$Acc = \sum_{m=\lfloor M/2 \rfloor + 1}^M \binom{M}{m} P^m (1 - P)^{M-m} \quad (4.1)$$

Łatwo pokazać, że już przy $P = 0.7$ i 9 członkach dokładność zespołu wzrasta z wartości $P = 0.7$ do 0.901. Powyższy wzór przyjął idealistyczne założenie o pełnej niezależności

i jednakowej dokładności wszystkich członków zespołu, co w praktyce jest zwykle nieosiągalne. Tym nie mniej wskazuje potęgę działania zespołu.

Niezależnie od standardowej metody bagging czy boosting [30] istnieje wiele metod poprawiających niezależność działania członków zespołu. Jednym z nich jest zastosowanie rozwiązań poszczególnych jednostek bazujących na innej zasadzie działania. Do najbardziej typowych należy zastosowanie jednostek charakteryzujących się zasadniczo różnym sposobem przetwarzania danych. Niezależnie od tego można również stosować struktury różniące się pod względem złożoności i sposobu doboru parametrów, na przykład liczby neuronów ukrytych w sieciach neuronowych, doboru hiperparametrów sieci w procesie uczenia, liczby drzew decyzyjnych w lesie losowym, czy dobór liczby zmiennych w węzłach decyzyjnych. Dodatkowo każdy klasyfikator może być trenowany na nieco różniącym się zbiorze uczącym, generowanych losowo.

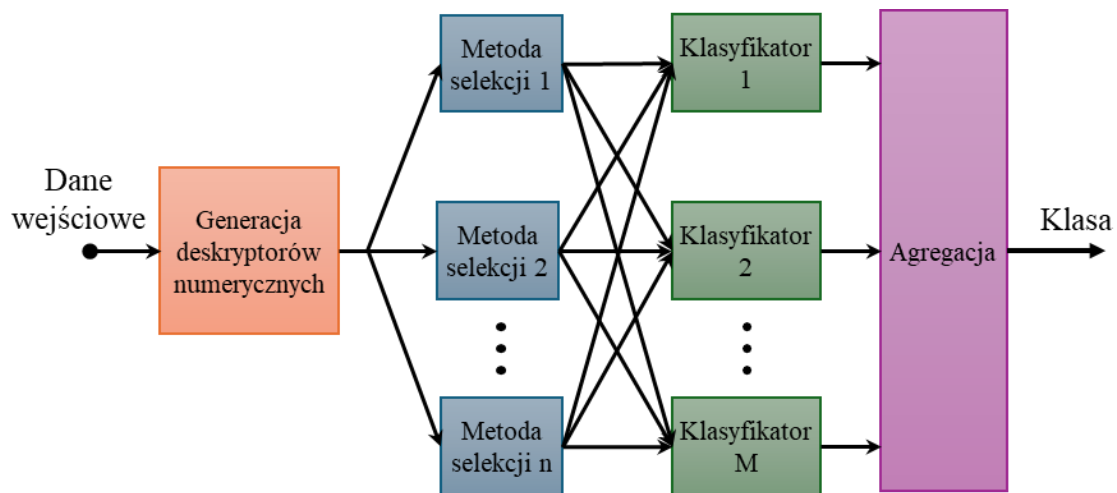
Przy zastosowaniu sieci głębokich jako członków zespołu można stosować różne struktury sieciowe, na przykład Alexnet, Resnet, Shufflenet, Efficientnet, Mobilenet, Inception, itp. Dodatkowo można stosować sieci tego samego rodzaju, ale przy innym doborze hiperparametrów wewnętrznych i parametrów uczenia, stosując przy tym zróżnicowane zbiory uczące. Zróżnicowanie zbiorów uczących jest ważnym elementem przy kształtowaniu niezależnych rozwiązań klasyfikatorów, niezależnie od ich struktury (płytką czy głęboką). Dzięki temu adaptowane parametry poszczególnych rozwiązań będą różne, a więc ich działanie w trybie odtworzeniowym również odmienne. Problemem może być ograniczony zasób danych przeznaczonych do uczenia.

Ciekawym rozwiązaniem zapewniającym poprawę niezależności działania poszczególnych jednostek zespołu jest wtrącenie losowości w samą strukturę jednostek klasyfikacyjnych jak i losowa fuzja wyników działania poszczególnych jednostek. Takie rozwiązanie zostało zaprezentowane między innymi w pracy [55].

Ważnym elementem niezależności działania jest zróżnicowanie sposobu generacji cech diagnostycznych (atrybutów wejściowych) dla każdego indywidualnego członka zespołu. Cechy te generowane są na podstawie tego samego zbioru oryginalnych danych pomiarowych, ale przetwarzane według innego mechanizmu oceny i selekcji. Przy takim podejściu indywidualni członkowie zespołu mają dostęp do pełnej informacji pomiarowej, ale inaczej przetworzonej [54,57].

Typowa struktura zespołu klasyfikatorów może być przedstawiona jak na rys. 4.1. Dane wejściowe są przetwarzane na zbiór deskryptorów numerycznych (w naszym przypadku obrazy wejściowe są przetwarzane przez sieć CNN na deskryptory zawarte

w wektorze wypłaszczonej), poddanych działaniu różnym metodom selekcji cech. Wynik selekcji każdej metody jest podawany na zbiór M klasyfikatorów końcowych. Wyniki ich działania podlegają agregacji (fuzji) w jeden uzgodniony werdykt klasyfikacyjny zespołu (klasę przynależności danych wejściowych).



Rys. 4.1 – Schemat typowej struktury zespołu klasyfikatorów [43].

Końcowy werdykt zespołu może być ustalony poprzez głosowanie jego członków, przy czym można zastosować różne strategie głosowania.

Najprostszy wybór werdyktu końcowego zespołu klasyfikacyjnego to zwykle głosowanie większościowego (ang. *majority voting*). Każdy rekord ze zbioru danych wejściowych jest przypisany do odpowiedniej klasy przez poszczególne klasyfikatory. Są tu możliwe dwa podejścia do wyniku klasyfikacji

- Podejście ostre, w którym przynależność do klasy wskazana przez klasyfikator przyjmuje tylko dwie wartości binarne: 1 (dane należą do określonej klasy) i 0 (brak przynależności do pozostałych klas)
- Podejście rozmyte – klasyfikator generuje wartości prawdopodobieństwa przynależności do każdej klasy, przy czym wartości te są z przedziału $[0, 1]$. Sumowaniu w każdej klasie podlegają wartości tych prawdopodobieństw.

W procesie agregacji następuje sumowanie wyników dla każdej klasy, niezależnie od tego czy jest to wynik w postaci binarnej czy rozmytej. Zwycięża klasa, która uzyska największą wartość. Tego typu głosowanie jest efektywne, przy założeniu, że wszystkie klasyfikatory charakteryzują się podobną skutecznością (niekoniecznie dokładnie taką samą).

W praktyce ten warunek nie zawsze jest możliwy do spełnienia. Najsłabszy klasyfikator może obniżyć skuteczność zespołu. W efekcie wynik końcowy może być

znacznie gorszy niż wynik najlepszego członka zespołu (choć zwykle jest lepszy niż średnia wszystkich członków). W przypadku bardzo nierównej jakości członków zespołu lepsze jest głosowanie większościowe ważone. W takim rozwiązaniu wkład wyniku klasyfikatora indywidualnego zależy od jego dokładności rozpoznania poszczególnych klas, przy czym ocenia się ją na zbiorze danych uczących. Jest to słuszne założenie zakładając, że dane uczące są reprezentatywne dla zbioru danych, w tym dla zbioru testującego. Przy M klasyfikatorach w zespole należy określić wielkość $y_i(\mathbf{x})$ proporcjonalną do prawdopodobieństwa przynależności wektora \mathbf{x} do i -tej klasy według wzoru [44]

$$y_i(\mathbf{x}) = \sum_{k=1}^M w_{ki} z_{ki}(\mathbf{x}) \quad (4.2)$$

w którym w_{ki} jest wagą z jaką k -ty klasyfikator jest uwzględniany we wzorze (3.20) dla i -tej klasy, natomiast z_{ki} jest sygnałem wyjściowym k -tego klasyfikatora odpowiedzialnym za rozpoznanie i -tej klasy.

Wartości wag w_{ki} mogą być wyznaczone w różny sposób. Często stosuje się w agregacji skalę logarytmiczną [36] w której wpływ słabego klasyfikatora jest jeszcze bardziej minimalizowany przy podejmowaniu ostatecznej decyzji.

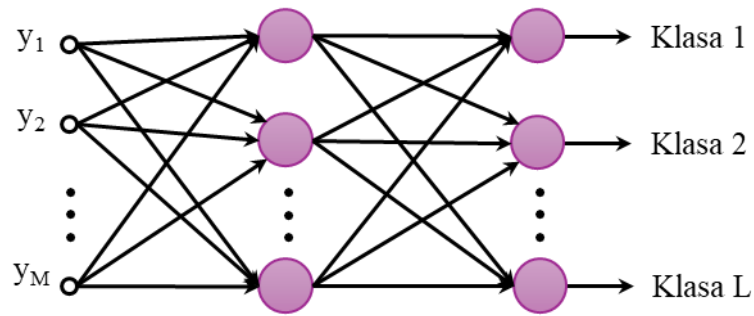
$$w_{ki} = \ln \left(\frac{\eta_{ki}}{1-\eta_{ki}} \right) \quad (4.3)$$

Stosowana jest również zależność normalizacyjna definiowana w postaci [27]

$$w_{ki} = \frac{\eta_{ki}^m}{\sum_{j=1}^M \eta_{kj}^m} \quad (4.4)$$

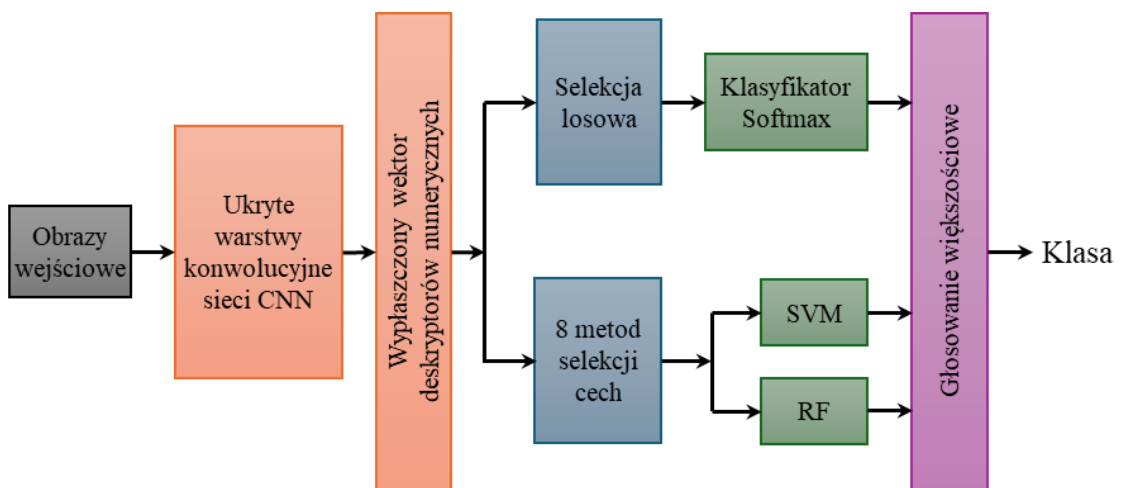
gdzie η_{ki}^m oznacza wskaźnik jakości (np. dokładność, czułość, precyzja) k -tego klasyfikatora przy rozpoznaniu i -tej klasy, m – wykładnik różnicujący wpływ poszczególnych jednostek zespołu na wynik działania zespołu (np. $m = 1, 2, 3$). Wyższa wartość m preferuje klasyfikatory o wyższej jakości, zmniejszając wpływ mniej dokładnych wyników na werdykt zespołu.

Innym sposobem jest zastosowanie dodatkowej sieci sumacyjnej (na przykład sieć neuronowa) przetwarzającej w sposób automatyczny wyniki poszczególnych klasyfikatorów na werdykt końcowy zespołu. Sieć taka zasilana jest poprzez zbiór wyników poszczególnych klasyfikatorów generując na wyjściu wynik zespołu w postaci prawdopodobieństwa przynależności do poszczególnych klas. Schemat takiego integratora wykorzystującego typową sieć neuronową MLP przedstawiony jest na rys. 4.2.



Rys. 4.2 – Schemat integratora neuronowego do agregacji ważonej wyników M członków zespołu. Zmienne y_1, y_2, \dots, y_M oznaczają wektory wynikowe M klasyfikatorów w zespole.

Wektor wejściowy sieci utworzony jest przez wektory wynikowe poszczególnych klasyfikatorów. Przy L klasach każdy taki wektor ma wymiar L , co przy M klasyfikatorach otrzymuje się rozmiar wektora wejściowego równy LM . Liczba neuronów wyjściowych sigmoidalnych jest równa liczbie klas. Sieć podlega uczeniu z nauczycielem w postaci par danych uczących składających się z wektora wejściowego $[y_1, y_2, \dots, y_M]$ generowanego przez poszczególne M klasyfikatory i stowarzyszonego z nim wektora właściwej przynależności danych uczących do jednej z L klas. Liczba neuronów ukrytych podlega doborowi w procesie uczenia. Po zakończeniu uczenia parametry sieci podlegają zamrożeniu i integrator jest gotowy do trybu testującego na danych nie uczestniczących w uczeniu.

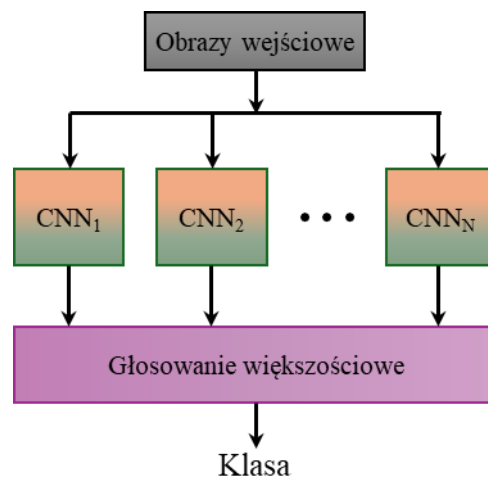


Rys. 4.3 – Schemat pierwszego rozwiązania zespołu utworzonego przez 16 klasyfikatorów klasycznych zbudowanych na bazie SVM i RF i 8 metodach deterministycznych selekcji oraz klasyfikatora typu softmax wykorzystującego losowy dobór cech dokonywany w każdej iteracji uczącej (zmiany).

W pracy zastosowano dwa różne podejścia do tworzenia zespołu klasyfikatorów. Rozwiązanie pierwsze stosuje połączenie zbioru klasyfikatorów klasycznych SVM i RF zasilanych przez sygnały wyselekcjonowane za pomocą przedstawionych wcześniej

ośmiu metod selekcji ze zbiorem klasyfikatorów głębokich CNN zasilanych poprzez losowo wybierane elementy wektora wypłaszczonego w każdej iteracji. Schemat tak utworzonego zespołu przedstawiony jest na rys. 4.3 [16,18].

Drugie rozwiązanie zespołu bazuje wyłącznie na strukturach głębokich i zastosowaniu różnych architektur sieci CNN (rys. 4.4). Poszczególne klasyfikatory zostały wytrenowane przy użyciu techniki „transfer learning” z zastosowaniem gotowych wstępnie przetrenowanych w Matlabie sieci CNN [39]. W podejściu „transfer learning” korzysta się ze wstępnie wytrenowanej sieci o strukturze i parametrach dobranych przez określone zespoły światowych specjalistów w tej dziedzinie. Zadanie użytkownika sprowadza się jedynie do adaptacji wstępnie wytrenowanej struktury sieci (zwykle kilka ostatnich warstw lokalnie połączonych i klasyfikatora finalnego) na własnych zbiorach danych. Parametry pierwszych warstw lokalnie połączonych odpowiedzialnych za mechanizm generacji deskryptorów numerycznych obrazu pozostają zwykle niezmiennione. W efekcie użytkownik dokonuje jedynie douczenia sieci do rozwiązania własnego problemu klasyfikacyjnego (lub regresyjnego w przypadku regresji). Odbywa się ono przy użyciu metod gradientowych z zastosowaniem propagacji wstecznej błędu. Proces douczenia trwa stosunkowo krótko. Dzięki zastosowaniu tej techniki douczania zbiór uczący może być znacząco mniejszy, co jest ważną zaletą zwłaszcza w problemach rozpoznawania obrazów medycznych, w których liczba danych jest zwykle bardzo ograniczona.



Rys. 4.4 – Schemat połączeń drugiego rozwiązania zespołu składającego się wyłącznie ze struktur głębokich CNN.

W tabeli 4.1 przedstawiono zestaw 19 głębokich struktur CNN dostępnych w Matlabie, które będą podlegać wykorzystaniu w dalszych eksperymentach numerycznych. Kolejne kolumny w tabeli przedstawiają aktualną nazwę sieci, głębokość rozumianą jako liczbę warstw według oznaczeń autorów (często warstwa jest traktowana jako zbiór kilku

podwarstw, na przykład w Alexnet), zajętość pamięci, liczbę adaptowanych parametrów (liczona w milionach) oraz wymagany wymiar obrazu wejściowego.

Tabela 4.1– Wykaz wstępnie przetrenowanych architektur sieci CNN biorących udział w badaniach symulacyjnych.
* Sieci neuronowe Nasnetmobile i Nasnetlarge nie składają się z liniowej sekwencji modułów.

| Nazwa sieci neuronowej | Głębokość | Rozmiar | Liczba parametrów (miliony) | Wymiar obrazu wejściowego |
|------------------------|-----------|---------|-----------------------------|---------------------------|
| Squeezenet [29] | 18 | 5.2 MB | 1.24 | 227x227 |
| Googlenet [59] | 22 | 27 MB | 7.0 | 224x224 |
| Inceptionv3 [60] | 48 | 89 MB | 23.9 | 299x299 |
| Densenet201 [28] | 201 | 77 MB | 20.0 | 224x224 |
| Mobilenetv2 [50] | 53 | 13 MB | 3.5 | 224x224 |
| Resnet18 [25] | 18 | 44 MB | 11.7 | 224x224 |
| Resnet50 [25] | 50 | 96 MB | 25.6 | 224x224 |
| Resnet101 [25] | 101 | 167 MB | 44.6 | 224x224 |
| Xception [7] | 71 | 85 MB | 22.9 | 299x299 |
| Inceptionresnet2 [58] | 164 | 209 MB | 55.9 | 299x299 |
| Shufflenet [71] | 50 | 5.4 MB | 1.4 | 224x224 |
| Nasnetmobile [3] | * | 20 MB | 5.3 | 224x224 |
| Nasnetlarge [3] | * | 332 MB | 88.9 | 331x331 |
| Darknet19 [48] | 19 | 78 MB | 20.8 | 256x256 |
| Darknet53 [48] | 53 | 155 MB | 41.6 | 256x256 |
| Efficientnetb0 [61] | 82 | 20 MB | 5.3 | 224x224 |
| Alexnet [34] | 8 | 227 MB | 61.0 | 227x227 |
| Vgg16 [53] | 16 | 515 MB | 138 | 224x224 |
| Vgg19 [53] | 19 | 535 MB | 144 | 224x224 |

Poszczególne sieci różnią się zasadniczo pod względem architektury i sposobu przetwarzania danych w warstwach lokalnie połączonych, stwarzając tym samym perspektywy niezależnego działania i podejmowania decyzji klasyfikacyjnej. Podstawowa (pierwotna) struktura sieci Alexnet [34] jako pierwowzór rozwiązania została przez licznych autorów zmodyfikowana na wiele różnych sposobów.

Na przykład struktury typu Resnet [25] wprowadzają połączenia skróśne (residualne) między dwoma warstwami konwolucyjnymi. Sieć Squeezenet [29] zastępuje duże filtry poprzez specjalne moduły typu Fire stosujące małe filtry 1x1 i 3x3 oraz konwolucję grupową na mniejszych podzbiorach obrazów. Sieć Shufflenet [71] wprowadza grupowanie filtrów 1x1 w grupy operujące na zredukowanej liczbie kanałów (obrazów) wejściowych tworząc równoległe wiele grup operujących na wybranych losowo podgrupach obrazów wejściowych i dokonując przetasowanie podgrup. Sieć Mobilenetv2 wprowadza 2 rodzaje konwolucji. Pierwsza z nich tzw. *DWConv* stosuje pojedynczy filtr do każdego kanału (obrazu). Następuje po niej drugi etap konwolucji punktowej *PConv*, stosującej filtr 1x1 działający na sumie obrazów wytworzonych w etapie poprzednim. Sieci typu

Inception [58] stosują równoległe działające filtry o różnych wymiarach 1×1 , 3×3 oraz 5×5 wprowadzając operacje rozszerzenia warstw wszerz. Dzięki temu każdy filtr jest w różnym stopniu wrażliwy na zróżnicowanie szczegółowości obrazu. Sieci Efficientnet [27,52,61] tworzone są poprzez dodatkowe wprowadzenie w procesie uczenia doboru odpowiednich proporcji między głębokością (*depth*) architektury (liczba warstw konwolucyjnych), szerokością (liczbą równoległe tworzonych obrazów w warstwach) oraz rozdzielczością obrazów tworzonych w poszczególnych warstwach. W odróżnieniu od większości rozwiązań, gdzie jest brak połączeń między odległymi warstwami sieć Densenet [28] operuje wieloma połączeniami między-warstwowymi w przód. W efekcie uzyskuje się wzbogacenie informacji przekazywanej do następnych warstw konwolucyjnych.

Liczba sieci CNN biorących udział w zespole klasyfikatorów podlega doborowi w procesie tworzenia zespołu. W wyniku wielu przeprowadzonych symulacji okazało się, że bardzo duża liczba członków zespołu nie wprowadza polepszenia końcowego wyniku, a w sposób zdecydowany wydłuża czas uczenia całego zespołu. Stąd dobór wielkości zespołu i składu jego członków wymaga przeprowadzenia wielu eksperymentów dla każdego zadania klasyfikacyjnego.

W pracy zbadano wykorzystanie różnych dostępnych w Matlabie struktur sieci CNN, biorąc pod uwagę ich efektywność, czas uczenia oraz różnorodność cech, na podstawie których sieć podejmuje decyzję klasyfikacyjną. Wyniki tych eksperymentów, przeprowadzonych przy wykorzystaniu wersji 2023 Matlaba będą przedstawione w następnych rozdziałach.

Zastosowanie techniki „transfer learning” nie zwalania użytkownika z zapewnienia jak największego możliwie zbioru danych uczących. Z teorii generalizacji Vapnika-Chervonenkisa wynika, że im większa populacja danych uczących tym lepsze będą wyniki działania systemu na danych testujących. Stosuje się różne metody zwiększenia zbioru uczącego. Klasyczne techniki augmentacji stosują takie operacje na obrazach jak rotacja, przesunięcia, skalowanie itp., choć ich skuteczność jest ograniczona. W aktualnym stopniu rozwoju sieci głębokich możliwe jest również generowanie sztucznych obrazów na bazie odwzorowania cech statystycznych oryginałów, np. techniki GAN [11,21,54].

4.2. Metody oceny jakości klasyfikacji zastosowane w pracy

W wyniku działania systemu klasyfikacyjnego otrzymuje się przypisanie wektorów wejściowych układu do odpowiedniej klasy, które następnie są porównywane z etykietą definiującą prawdziwą przynależność do klasy. Dokonuje się to poprzez zastosowanie macierzy rozkładu klas zwanej macierzą pomyłek [44,62]. Wiersze tej macierzy reprezentują liczbę rekordów danych należących do kolejnych klas, a kolumny – liczbę wzorców przypisanych przez klasyfikator do danej klasy. Elementy diagonalne macierzy pomyłek reprezentują liczbę poprawnie rozpoznanych danych, natomiast elementy poza-diagonalne błędne rozpoznanie. Element ij -ty tej macierzy określa liczbę przypadków klasy i -tej przypisanych przez klasyfikator do klasy j -tej.

Najczęściej w praktyce medycznej definiuje się macierz pomyłek dla dwu klas: tej podlegającej rozpoznaniu oraz pozostałych klas łącznie. Stąd macierz pomyłek przedstawić można w postaci jak w tabeli 4.2, w której elementy oznaczone poprzez + reprezentują klasę podlegającą rozpoznaniu (klasa pozytywna) a elementy oznaczone poprzez – klasę reprezentującą resztę zwykle klasę większościową (klasa negatywna). Element TP (ang. *True Positive*) macierzy reprezentuje liczbę poprawnie rozpoznanych przypadków podlegających rozpoznaniu, FN (ang. *False Negative*) liczbę przypadków rozpoznanych błędnie jako większościowe, FP (ang. *False Positive*) liczbę przypadków większościowych błędnie przypisanych do klasy rozpoznawanej, TN (ang. *True Negative*) liczbę poprawnie rozpoznanych przypadków klasy większościowej.

Tabela 4.2 – Oznaczenia macierzy pomyłek reprezentującej rozkładu 2 klas (+ oznacza klasę poszukiwaną, natomiast – klasę reprezentującą pozostałe przypadki (zwaną dalej klasą większościową))

| Predykowana Rzeczywista | Klasa + | Klasa – |
|----------------------------|---------|---------|
| Klasa + | TP | FN |
| Klasa – | FP | TN |

Na podstawie macierzy pomyłek tworzy się różne wskaźniki jakości klasyfikatora [44,62]. Podstawowym wskaźnikiem jest średnia dokładność względna (ACC) definiowana jako stosunek liczby rozpoznań poprawnych do liczby wszystkich przypadków poddanych badaniu. Przy oznaczeniu macierzy pomyłek jak wyżej wartość ACC określa się wzorem

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.5)$$

Ważnym wskaźnikiem z punktu widzenia medycznego jest czułość, oznaczana zwykle symbolem TPR (ang. *True Positive Rate*) i definiowana w postaci

$$TPR = \frac{TP}{TP+FN} \quad (4.6)$$

Odpowiadającym jej wskaźnikiem dotyczącym klasy większościowej jest specyficzność oznaczana symbolem TNR (ang. *True Negative Rate*), definiowana w postaci

$$TNR = \frac{TN}{TN+FP} \quad (4.7)$$

Dodatkowo definiuje się precyzję rozpoznania obu klas: mniejszościowej PPV (ang. *Positive Predictive Value*) i większościowej NPV (ang. *Negative Predictive Value*). Są one wyrażone wzorami

$$PPV = \frac{TP}{TP+FP} \quad (4.8)$$

$$NPV = \frac{TN}{TN+FN} \quad (4.9)$$

Połączeniem precyzji z czułością i specyficznością jest miara F1. Definiowana jest dla każdej klasy oddzielnie jako stosunek iloczynu precyzji i czułości lub precyzji i specyficzności) do ich średniej wartości. Dla klasy mniejszościowej miara F1(+) jest definiowana w postaci

$$F1(+) = \frac{TPR*PPV}{0.5(TPR+PPV)} \quad (4.10)$$

Podobnie dla klasy większościowej

$$F1(-) = \frac{TNR*NPV}{0.5(TNR+NPV)} \quad (4.11)$$

Istotnym elementem w ocenie rozwiązania klasyfikatora 2-klasowego jest porównanie z klasyfikatorem losowym. Dokonuje się to poprzez charakterystykę ROC (ang. *Receiver Operating Characteristic*), reprezentującą graficzną zależność $TPR = f(1 - TNR)$, gdzie wyrażenie $(1 - TNR)$ reprezentuje wielkość zwaną FPR (ang. *False Positive Rate*), $FPR = 1 - TNR$. Każdy punkt krzywej ROC odpowiada innemu wynikowi modelu klasyfikatora. Punkty położone na diagonalu odpowiadają wynikowi klasyfikacji losowej, powyżej diagonalu – klasyfikatorowi lepszym niż losowy, poniżej diagonalu – gorszym niż klasyfikator losowy.

Na podstawie krzywej ROC ocenia się jakość rozwiązania klasyfikatora w postaci pola pod krzywą, zwanego AUC (ang. *Area Under Curve*). Wartość $AUC = 1$ oznacza

klasyfikator idealny ($TPR = 1$, $FPR = 0$). Wartość $AUC = 0.5$ oznacza klasyfikator losowy. Im wartość AUC jest bliższe jedności, tym lepsze jest proponowane rozwiązanie klasyfikatora.

W przypadku problemu wieloklasowego definiuje się czułość i precyzję rozpoznania każdej klasy oddzielnie. W przypadku czułości i -tej klasy $Sens(i)$ bierze się pod uwagę i -ty wiersz macierzy pomyłek, a w przypadku precyzji $Prec(i)$ i -tą kolumnę tej macierzy, korzystając ze wzorów wprowadzonych wcześniej. Miara F1 w takim przypadku dotyczy każdej klasy oddzielnie i definiowana jest w postaci

$$F1(i) = \frac{Sens(i)*Prec(i)}{0.5(Sens(i)+Prec(i))} \quad (4.12)$$

5. Wyniki badań eksperymentalnych w rozpoznawaniu czerniaka

W niniejszym rozdziale przedstawione będą wyniki badań eksperymentalnych dotyczących skuteczności różnych systemów klasyfikacyjnych. Celem eksperymentów jest zbadanie efektywności różnych technik tworzenia zespołów klasyfikacyjnych opartych na głębokich sieciach neuronowych z wykorzystaniem również klasyfikatorów klasycznych (SVM, las losowy). Analizie poddane zostaną miary jakości, takie jak dokładność, precyzja, czułość, specyficzność oraz krzywe ROC (pole AUC), które pozwolą na rzetelną ocenę działania poszczególnych modeli. Koncepcja badań obejmowała przeprowadzenie w pierwszej kolejności testów indywidualnej klasyfikacji różnych modeli sieci CNN przy wykorzystaniu modeli wstępnie wytrenowanych w środowisku Matlab, a następnie zbudowanie zespołu bazującego na tych wynikach.

Po przeprowadzeniu oceny jakości poszczególnych rozwiązań indywidualnych wybrano ograniczoną liczbę sieci tworzących zespoły klasyfikacyjne. Przy wyborze konkretnych członków zespołu wzięto pod uwagę zarówno czynniki jakościowe związane z dokładnością klasyfikacji jak i wymagane zasoby pamięci komputerowej. Z punktu widzenia implementacji praktycznej istotnym czynnikiem warunkującym możliwość efektywnego uruchomienia systemu jest wielkość sieci, która przekładała się bezpośrednio na wymaganą do przeprowadzenia badań pamięć jednostki obliczeniowej. Z podsumowania wyników badań wynika ocena, na ile zastosowanie zespołów klasyfikatorów ma pozytywny wpływ na końcowy wynik w porównaniu z wynikami indywidualnych jednostek.

Badanie skuteczności rozpoznawania i klasyfikacji czerniaka zostało przeprowadzone na dwóch zbiorach danych: małym NIO (powstałym przy współpracy z lekarzami w Centrum Onkologii w Warszawie) i zbiorem dużym ISIC dostępnym w Internecie. W obu przypadkach obrazy poddane analizie zostały wytworzone przy użyciu dermatoskopu.

5.1. Rozpoznanie czerniaka na zbiorze NIO

Zbiór NIO obejmował 134 obrazy czerniaka oraz 112 obrazów brodawki łojotokowej reprezentującej łagodne zmiany nowotworowe skóry niebędące czerniakiem. Rozpoznanie rodzaju zmian skórnych sprowadza się do problemu klasyfikacji dwuklasowej: klasa 1 – czerniak, klasa 2 – brodawka. Rozwiązania przedstawione w tym rozdziale

wykorzystywać będą sieci głębokie CNN o różnorodnej strukturze (przedstawione w rozdziale trzecim).

Przedstawione w tym podrozdziale wyniki klasyfikacji obrazów czerniaka ze zbiorów NIO poprzedzone były wieloma próbami mającymi na celu dopasowanie modeli sieci CNN do rozpatrywanego problemu. Zastosowano podejście opierające się na transferze wiedzy z modelu wstępnie wytrenowanego do zadania właściwego pracy (transfer learning). Dostrojanie modelu klasyfikatora CNN w kontekście transfer learningu polega na dostosowaniu struktury i wartości parametrów tego modelu CNN w Matlabie, do właściwego zadania klasyfikacji. Należy zauważyć, że modele użyte w transfer learningu trenowane były na zupełnie innych danych, dalece odbiegających od rozwiązywanego w pracy problemu rozpoznania obrazów medycznych. Proces transferu umożliwia wykorzystanie wiedzy nabytej przez model na tych zbiorach danych i dostosowanie go do problemu klasyfikacji czerniaka.

W pierwszym etapie badań przeprowadzono uczenie 19 indywidualnych modeli sieci CNN na zbiorze danych NIO wykorzystując wstępnie wytrenowane w Matlabie modele sieci. Każdy model sieci będzie reprezentowany w dalszych eksperymentach numerycznie według poniższego przyporządkowania:

1. Squeezenet
2. Googlenet
3. Inceptionv3
4. Densenet201
5. Mobilenetv2
6. Resnet18
7. Resnet50
8. Resnet101
9. Xception
10. Inceptionresnetv2
11. Shufflenet
12. Nasnetmobile
13. Nasnetlarge
14. Darknet19
15. Darknet53
16. Efficientnetb0
17. Alexnet
18. Vgg16
19. Vgg19

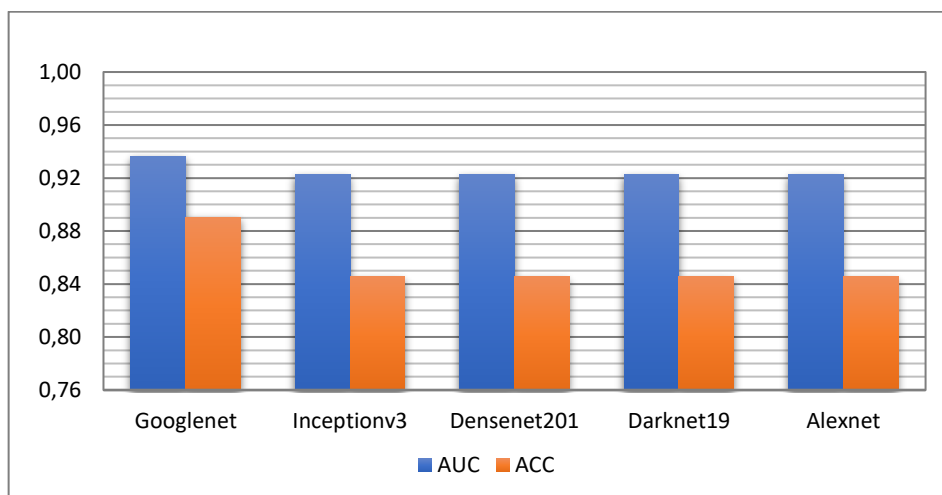
W adaptacji sieci powstaje problem wyboru liczby warstw istniejącej struktury podlegających docięciu na właściwym zbiorze danych. Przeprowadzono wiele prób „odmrażania” warstw wstępnie wytrenowanych.

W pierwszym podejściu tylko ostatnie warstwy klasyfikacyjne zostały docięte do zadania dwuklasowego bez zmiany parametrów warstw lokalnie połączonych. W następnym kroku sprawdzono wpływ zamrażania warstw lokalnie połączonych modelu jak również wartości hiperparametrów uczenia, takich jak liczba epok uczących, wartości startowe współczynników uczenia czy rozmiar minibatch.

Dla uzyskania jak najbardziej obiektywnej oceny działania systemu klasyfikacyjnego wyniki wszystkich przeprowadzonych badań są przedstawione w trybie 5-krotnej walidacji krzyżowej i dotyczą wyłącznie danych testujących nie uczestniczących w uczeniu.

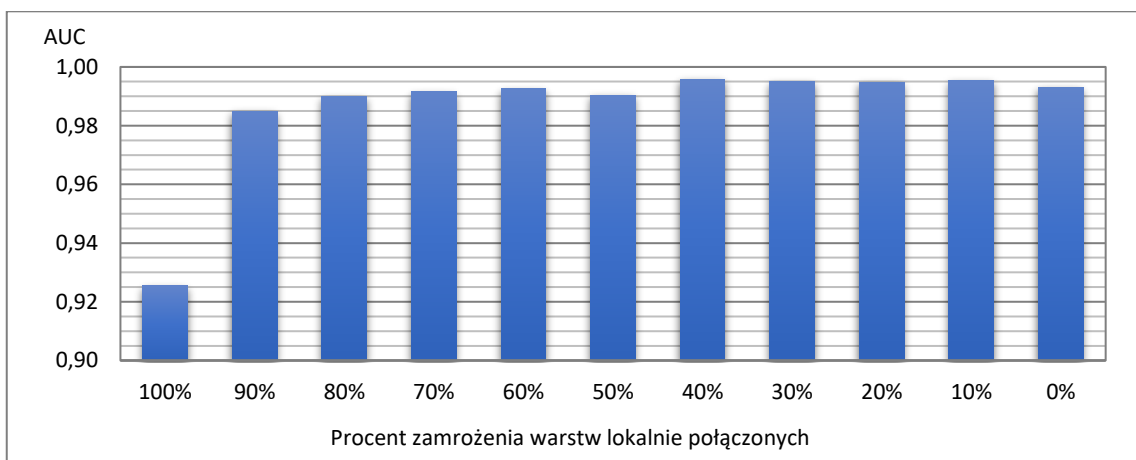
5.1.1. Wyniki klasyfikacji indywidualnych modeli CNN – zbiór NIO

Wstępne badania dotyczące wpływu zamrażania warstw zostały przeprowadzone na wybranych sieciach CNN. Po zamrożeniu wszystkich warstw lokalnie połączonych uczeniu na zbiorze NIO podlegała tylko podsieć globalnie połączona (softnet) stanowiąca końcowy klasyfikator softmaxu. Na rys. 5.1 przedstawiono wartości średnie dokładności rozpoznania czerniaka oraz AUC klasyfikatorów uzyskane w eksperymentach dla pięciu wybranych sieci. Najniższą wartość AUC (0.845) uzyskano dla sieci – Darknet19, natomiast najwyższą (0.892) dla sieci Googlenet.

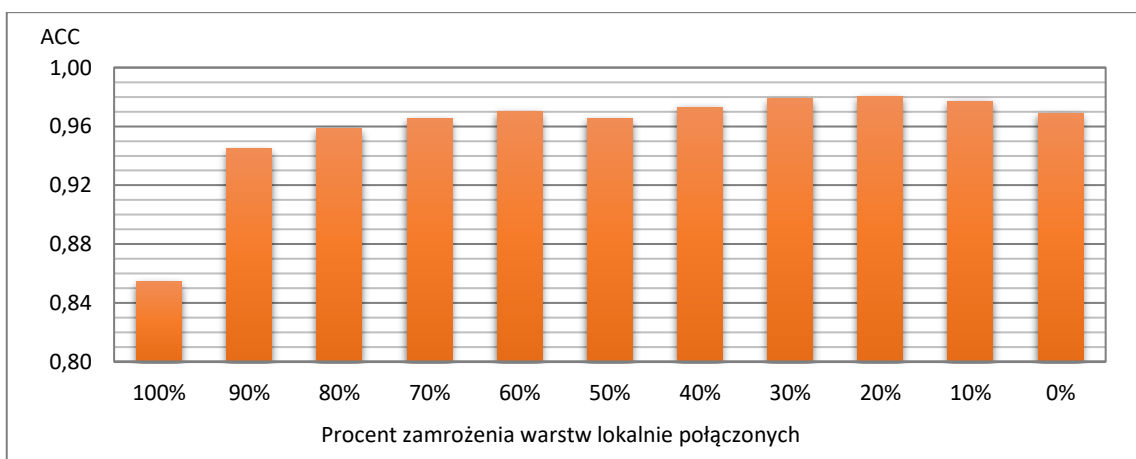


Rys. 5.1 – Wykres AUC i dokładności rozpoznania czerniaka w zbiorze NIO przy zastosowaniu 5 wybranych struktur sieci CNN przy zamrożeniu wszystkich warstw lokalnie połączonych.

Dalsze badania zostały przeprowadzone przy „odmrażaniu” warstw lokalnie połączonych. Proces douczenia parametrów sieci na zbiorze NIO przeprowadzono startując z wartości zapisanych w sieci wstępnie wytrenowanej. W związku z powyższym douczenie dotyczyło zarówno parametrów filtrów w określonych warstwach lokalnie połączonych jak i wag klasyfikatora softnetu. Przeprowadzono liczne próby zamrażania różnej liczby warstw ukrytych, poczynając zamrażanie od warstw pierwszych. Wpływ ilości douczanych warstw lokalnie połączonych na pole AUC pod krzywą ROC oraz dokładność klasyfikacji ACC przedstawiono kolejno na rysunkach 5.2 i 5.3. Wyniki dotyczą sieci Googlenet, Inceptionv3, Densenet201, Darknet19 oraz Alexnet (średnia dla wszystkich sieci).



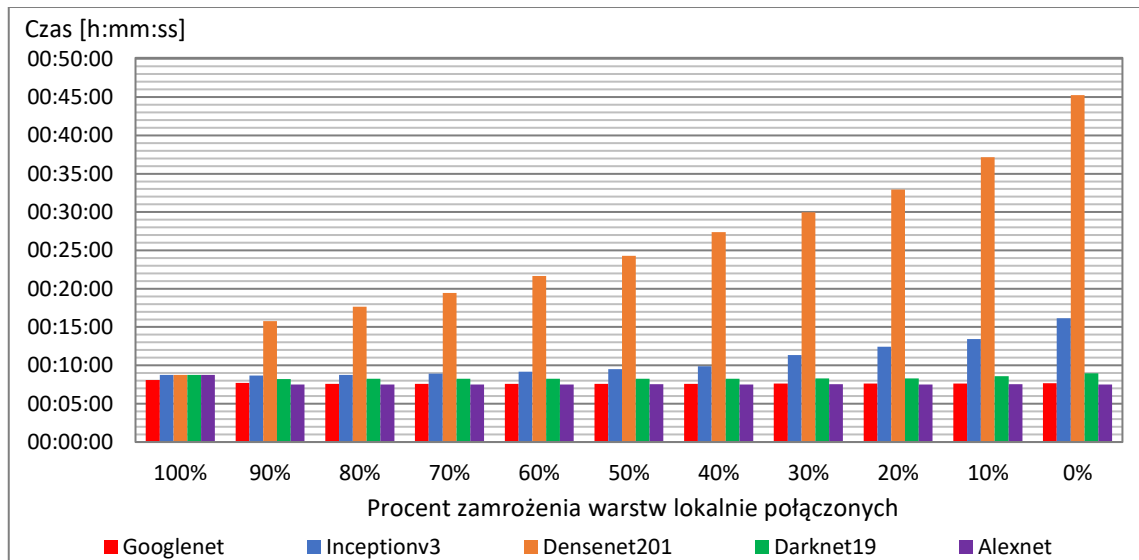
Rys. 5.2 – Średnie wartości pola AUC wybranych klasyfikatorów w zależności od ilości zamrożonych warstw lokalnie połączonych dla sieci Googlenet, Inceptionv3, Densenet201, Darknet19 oraz Alexnet.



Rys. 5.3 – Średnie wartości dokładności ACC wybranych klasyfikatorów w zależności od ilości zamrożonych warstw lokalnie połączonych dla sieci Googlenet, Inceptionv3, Densenet201, Darknet19 oraz Alexnet.

Minimalne zwiększenie liczby douczanych warstw lokalnie połączonych skutkowało znacznym polepszeniem jakości każdego z klasyfikatorów. Dalsze zmniejszanie ilości

zamrożonych warstw ukrytych powodowało zwiększanie zarówno dokładności jak i obszaru pod krzywą ROC, ale tylko do pewnego momentu. Najlepsze wyniki zarówno pod względem wartości AUC jak i ACC uzyskano przy zamrożeniu około 30% początkowych warstw.



Rys. 5.4 – Średnie wartości czasu uczenia wybranych klasyfikatorów w zależności od ilości zamrożonych warstw lokalnie połączonych.

W procesie douczania istotnym czynnikiem jest również czas potrzebny na przeprowadzenie tego procesu. Na rys. 5.4 przedstawiono porównanie czasu douczania 5 wybranych sieci (Googlenet, Inceptionv3, Densenet201, Darknet19 oraz Alexnet) przy różnej proporcji zamrożonych warstw lokalnie połączonych. Można zauważyć, że czas douczenia w każdym przypadku nieznacznie zwiększał się wraz z ilością warstw podlegających douczeniu dla większości struktur. Wyjątek stanowi sieć Densenet201, dla której każde zwiększenie liczby douczanych warstw znacznie wydłużało czas uczenia. Dla tej sieci obserwuje się najdłuższy czas douczenia, niezależnie od liczby zamrożonych wag. Wynika to z rozbudowanej struktury połączeń międzywarstwowych zastosowanych w tej sieci. W przypadku sieci małych, zawierających mniej niż 200 warstw, zmiany te były jednak tak niewielkie, że ich przyczyny należy dopatrywać bardziej w losowości wyboru wstępnych wartości wag niż w procesie uczenia sieci neuronowych. Wpływ ilości zamrożonych warstw na czas uczenia uwidacznia się dopiero przy sieciach dużych takich jak np. Densenet201 oraz Inceptionv3 zawierające odpowiednio 704 i 311 warstwy lokalnie połączone.

W tabeli 5.1 przedstawiono wyniki w postaci dokładności (ACC), pola (AUC) pod krzywą ROC oraz czasu uczenia (Czas) dla wybranego poziomu zamrożenia 30%

warstw i porównano je z wynikami działania sieci przy braku zamrożonych warstw (wszystkie parametry sieci podlegają douczeniu). Wyniki dotyczą wszystkich 19 struktur sieci CNN dostępnych w Matlabie. Dotyczą one danych testowych nie uczestniczących w uczeniu (tryb 5-krotnej walidacji krzyżowej). Istotnym elementem porównania jest również kolumna pokazująca łączny czas uczenia i testowania w trybie walidacji krzyżowej poszczególnych klasyfikatorów dla obu analizowanych przypadków. Ze względu na zastosowany tryb walidacji krzyżowej (testowanie w każdym „foldzie” następuje bezpośrednio po zakończeniu przebiegu uczenia) podane wartości liczbowe obejmują wprawdzie sumę obu czasów, ale dotyczą praktycznie czasu uczenia, który stanowi blisko 99% podanych wartości, stąd w dalszej części pracy będą traktowane jako czas uczenia.

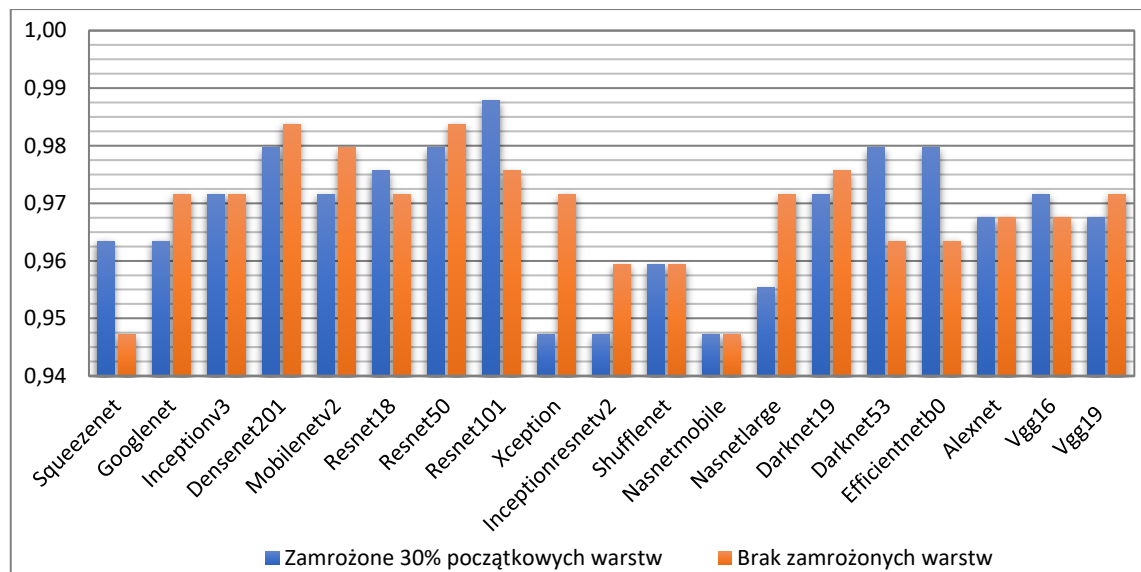
Tabela 5.1 – Porównanie wyników klasyfikacji przy użyciu różnych sieci CNN w zależności od ilości zamrożonych warstw

| Sieć CNN | Zamrożone 30% początkowych warstw | | | Brak zamrożonych warstw | | | Liczba warstw sieci |
|-------------------|-----------------------------------|---------------|-----------------|-------------------------|---------------|-----------------|---------------------|
| | AUC | ACC | Czas [h:mm:ss] | AUC | ACC | Czas [h:mm:ss] | |
| Squeezenet | 0.9894 | 0.9634 | 00:07:09 | 0.9894 | 0.9472 | 00:07:06 | 68 |
| Googlenet | 0.9886 | 0.9634 | 00:07:23 | 0.9915 | 0.9715 | 00:07:16 | 144 |
| Inceptionv3 | 0.9982 | 0.9715 | 00:11:18 | 0.9977 | 0.9715 | 00:15:43 | 315 |
| Densenet201 | 0.9981 | 0.9797 | 00:29:43 | 0.9967 | 0.9837 | 00:45:16 | 708 |
| Mobilenetv2 | 0.9968 | 0.9715 | 00:08:02 | 0.9955 | 0.9797 | 00:08:45 | 154 |
| Resnet18 | 0.9969 | 0.9756 | 00:07:52 | 0.9979 | 0.9715 | 00:07:49 | 71 |
| Resnet50 | 0.9979 | 0.9797 | 00:08:13 | 0.9985 | 0.9837 | 00:09:40 | 177 |
| Resnet101 | 0.9995 | 0.9878 | 00:11:55 | 0.9982 | 0.9756 | 00:17:36 | 347 |
| Xception | 0.9925 | 0.9472 | 00:12:09 | 0.9958 | 0.9715 | 00:17:24 | 170 |
| Inceptionresnetv2 | 0.9793 | 0.9472 | 00:31:52 | 0.9887 | 0.9593 | 00:43:11 | 824 |
| Shufflenet | 0.9918 | 0.9593 | 00:08:15 | 0.9909 | 0.9593 | 00:08:40 | 172 |
| Nasnetmobile | 0.9663 | 0.9472 | 00:29:03 | 0.9913 | 0.9472 | 00:42:25 | 913 |
| Nasnetlarge | 0.9940 | 0.9553 | 02:20:55 | 0.9968 | 0.9715 | 04:12:14 | 1243 |
| Darknet19 | 0.9963 | 0.9715 | 00:07:50 | 0.9899 | 0.9756 | 00:08:49 | 64 |
| Darknet53 | 0.9984 | 0.9797 | 00:14:53 | 0.9927 | 0.9634 | 00:26:03 | 184 |
| Efficientnetb0 | 0.9983 | 0.9797 | 00:13:02 | 0.9965 | 0.9634 | 00:19:18 | 290 |
| Alexnet | 0.9955 | 0.9675 | 00:07:07 | 0.9969 | 0.9675 | 00:07:09 | 25 |
| Vgg16 | 0.9887 | 0.9715 | 00:13:22 | 0.9928 | 0.9675 | 00:24:32 | 41 |
| Vgg19 | 0.9957 | 0.9675 | 00:15:35 | 0.9930 | 0.9715 | 00:29:42 | 47 |

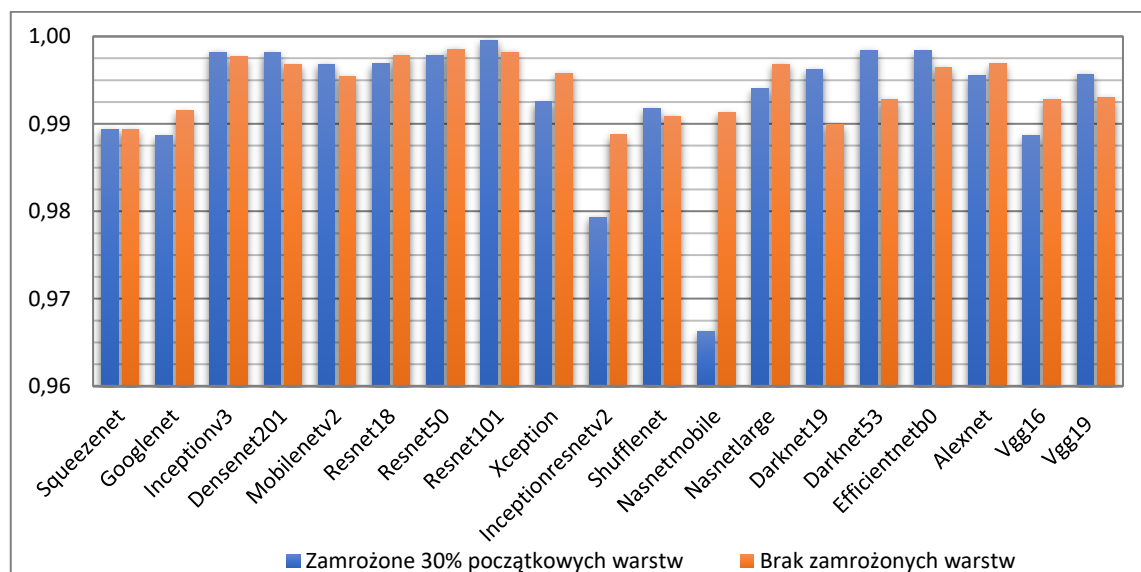
Wszystkie analizowane sieci wykazywały się dokładnością powyżej 94%. Najlepszy wynik uzyskano dla sieci Resnet101 (dokładność 98.78%, AUC = 0.9995), a najgorszy

odpowiada sieci SqueezeNet (dokładność 94.72%, AUC = 0.9894). Uzyskane wyniki (niezależnie od liczby zamrożonych warstw) są zdecydowanie lepsze niż te uzyskane wstępnie przy zamrożeniu wszystkich warstw lokalnie połączonych przedstawionych na rys. 5.1. Oznacza to, że w technice transfer learning istotne jest douczanie nie tylko warstwy globalnie połączonej, ale również warstw filtracyjnych, odpowiedzialnych za generację cech diagnostycznych procesu.

Poniżej na rysunkach przedstawiono graficzne zestawienie wyników testowania 19 sieci na zbiorze NIO w postaci uzyskanej dokładności ACC (rys. 5.5) oraz pola pod krzywą AUC (rys. 5.6).

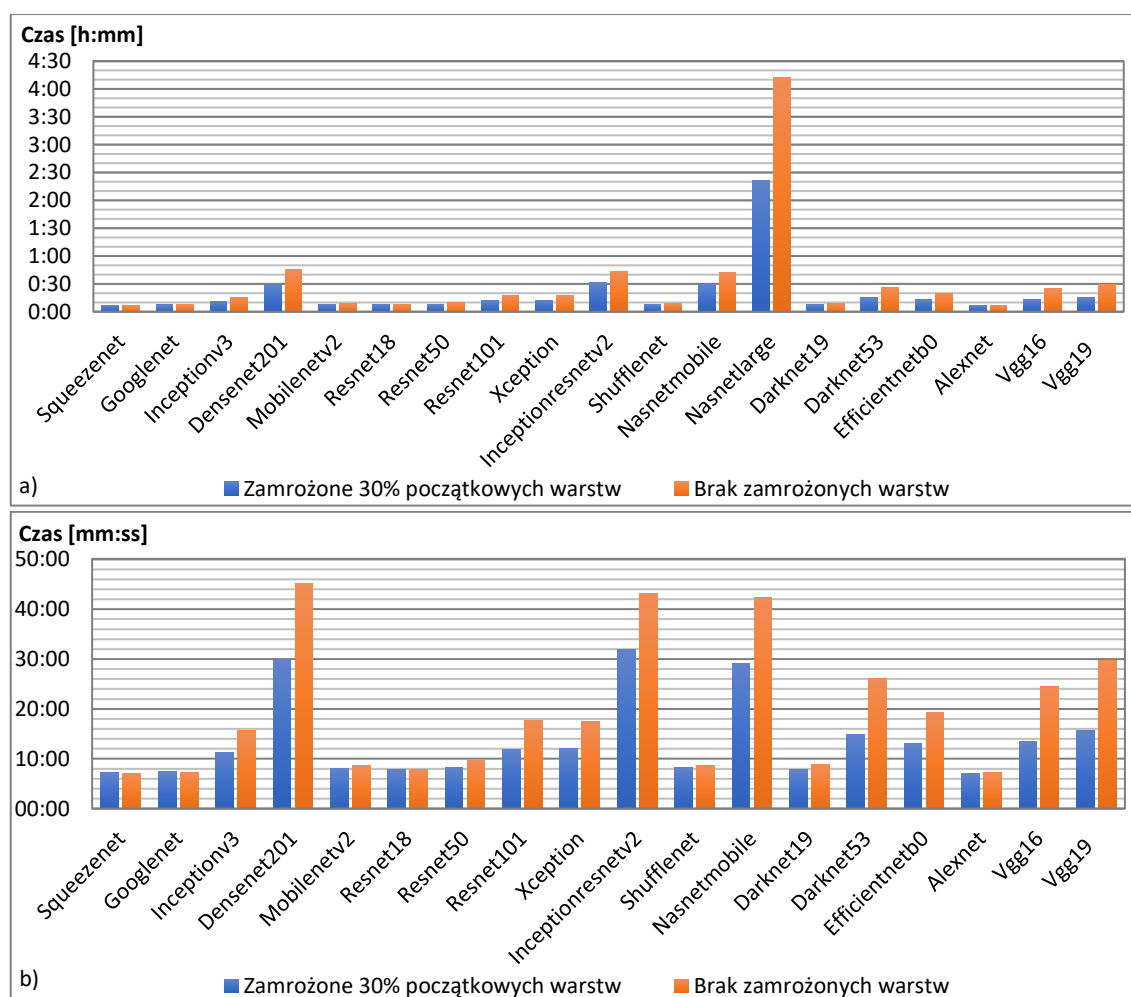


Rys. 5.5 – Wykres dokładności klasyfikacji przy zastosowaniu różnych modeli sieci CNN na zbiorze NIO.



Rys. 5.6 – Wykres powierzchni AUC pod krzywą ROC w klasyfikacji przy zastosowaniu różnych modeli sieci CNN na zbiorze NIO.

W zależności od modelu sieci douczanie parametrów filtrów wszystkich warstw lokalnie połączonych nie skutkowało jednoznaczną poprawą dokładności i pola pod krzywą ROC, co jest wyraźniej widoczne na rysunkach 5.5 i 5.6. Sieci o mniejszej liczbie warstw (na przykład od 150 do 170) wykazywały tendencję do lepszych wyników przy odblokowanych w procesie douczenia wszystkich warstwach sieci. Natomiast sieci większe zawierające kilkaset warstw sprawowały się lepiej przy zamrożonych 30% początkowych warstwach lokalnie połączonych. Przyczyna tego zjawiska może wynikać ze zwiększonego prawdopodobieństwa utknięcia procesu uczenia w niekorzystnym minimum lokalnym przy bardzo dużym rozmiarze wektora wagowego poddanego adaptacji.

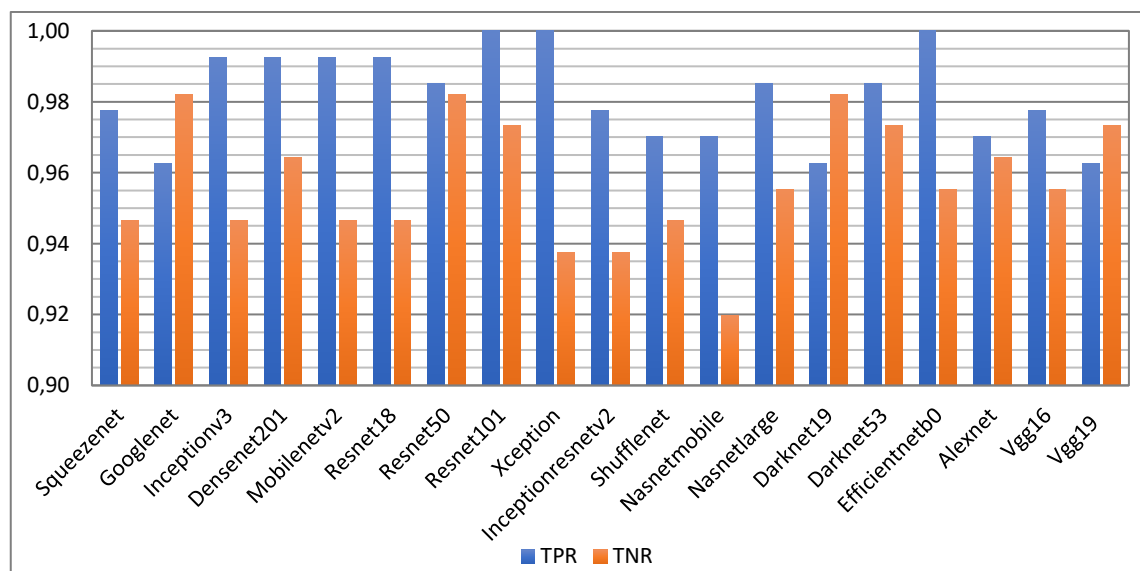


Rys. 5.7 – Wykres czasu trenowania i klasyfikacji w trybie 5-krotnej walidacji krzyżowej przy użyciu różnych modeli sieci CNN na zbiorze NIO. Górny wykres przedstawia wszystkie badane sieci CNN. Na dolnym wykresie wykluczono ze zbioru sieć Nasnetlarge dla uwypuklenia różnic czasowych dla pozostałych sieci.

Interesująca jest zależność czasu uczenia od wielkości sieci (liczba warstw w ostatniej kolumnie tabeli 5.1) i sposobu zamrażania. Zależności te przedstawiono na rys. 5.7. Dla stosunkowo niewielkich sieci (liczba warstw poniżej 200) czas uczenia sieci z zamrożonymi warstwami i bez zamrożonych wag jest zbliżony (różnice sekundowe), przy

czym w niektórych przypadkach przy braku zamrożonych warstw czas uczenia był nieznacznie krótszy (na przykład w przypadku Goglenet różnice dotyczą 7 sekund na korzyść sieci bez zamrożonych wag). Zasadnicze różnice w czasie uczenia obserwuje się dla sieci bardzo dużych. Na przykład sieć Nasnetlarge o 1243 warstwach wymagało 2 godzin i 21 minut przy 30% zamrożonych warstw i aż 4 godzin i 12 minut przy odmrożeniu wszystkich warstw, przy czym uzyskany wynik należy do najslabszych ze wszystkich sieci (dokładność 95.53% przy 30% zamrożonych wag i 97.15% przy wszystkich warstwach odmrożonych).

Powyższe testy umożliwiły wybranie ostatecznych ustawień dostosowujących wybrane sieci do realizowanego zadania. Na rys. 5.8 przedstawiono wykresy przedstawiające inne wskaźniki jakości w postaci czułości i specyficzności w rozpoznaniu obu klas: czerniaka (TPR) i brodawki (TNR) dla każdego analizowanego modelu klasyfikatora.

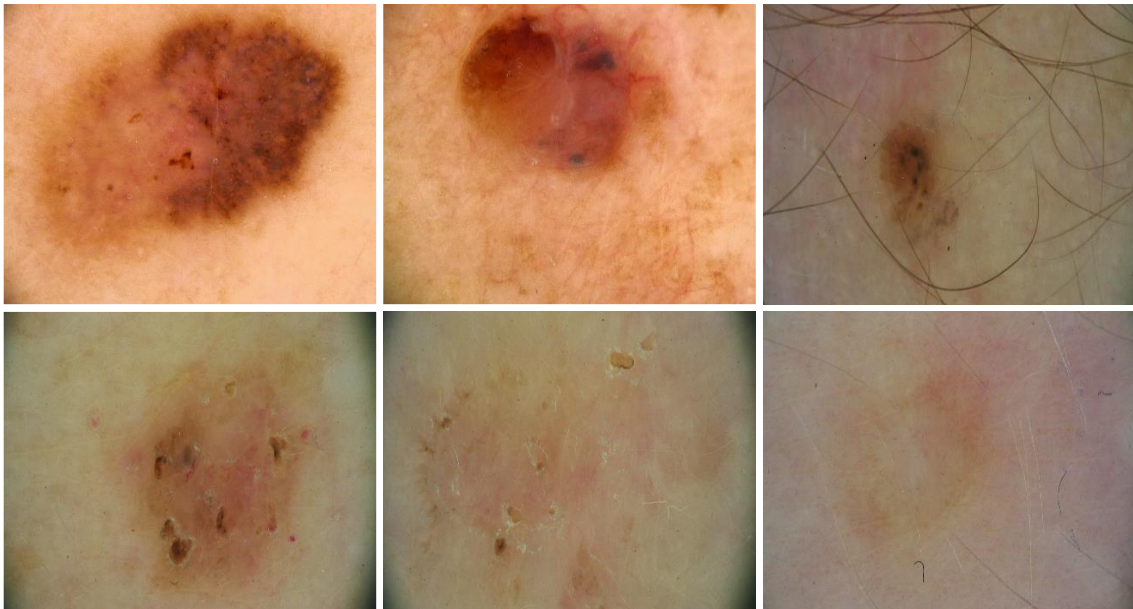


Rys. 5.8 – Wykres czułości (TPR) i specyficzności (TNR) badanych modeli klasyfikatorów.

Wymienione wyżej wskaźniki są ważnymi parametrami opisującymi jakość klasyfikatorów. W analizowanym problemie rozpoznawania obrazowego chorób nowotworowych bardzo istotny jest zwłaszcza TPR, gdyż oznacza rozpoznanie prawidłowe przypadku chorobowego (tak zwana czułość). Wskaźnik TNR oznacza prawidłowe zdiagnozowanie przypadku zdrowego (tak zwana specyficzność). Związane z nimi są: wskaźnik rozpoznania fałszywie pozytywnych (FPR) traktowanych jako błąd pierwszego rodzaju oraz wskaźnik rozpoznania fałszywie negatywnych (FNR) reprezentujący błąd drugiego rodzaju. Szczególnie niebezpieczny jest drugi rodzaj błędu, (zaliczenie pacjenta chorego do

klasy zdrowych), gdyż może spowodować błędne przeświadczenie o braku zagrożenia. Błędy obu rodzajów są ściśle powiązane z czułością i specyficznością. Stąd pożądana jest pewna równowaga między tymi wskaźnikami.

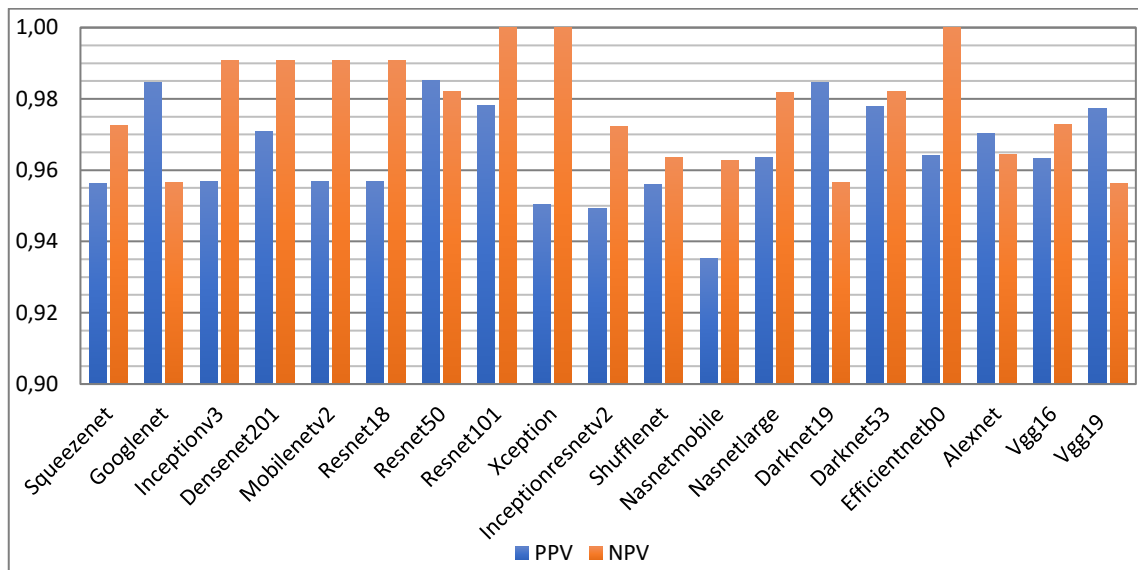
Wszystkie badane sieci CNN wykazały się czułością (wskaźnik TPR) powyżej 96%. Trzy sieci o najgorszych wynikach (Googlenet, Darknet19 i Vgg19) popełniły pięć razy błąd drugiego rodzaju, czyli zaliczenie pacjenta chorego do klasy zdrowych. Przykłady błędnie sklasyfikowanych obrazów tego typu zostały przedstawione na rys. 5.9. Występująca w badanym zbiorze brodawka łojotokowa, w opozycji do czerniaka, również należy do grupy zmian nowotworowych ale o charakterze łagodnym. W związku z powyższym nawet w tym przypadku zmiana skórna nie powinna zostać zbagatelizowana, gdyż istnieje pewne prawdopodobieństwo przekształcenia w zmianę czerniakową w kolejnym etapie leczenia.



Rys. 5.9 – Przykłady obrazów czerniaka sklasyfikowanych z błędem II rodzaju.

Trzy z badanych klasyfikatorów (Resnet101, Xception, Efficientnetb0) wykazały się 100-procentową czułością rozpoznania czerniaka, ale okupioną znacznym spadkiem specyficzności. Kolejne cztery sieci CNN osiągnęły poziom czułości powyżej 99% co oznacza popełnienie jednego błędu II rodzaju. Również w tym przypadku zostało to okupione zmniejszoną wartością specyficzności. Realizowana w dalszym etapie pracy klasyfikacja przy użyciu zespołów klasyfikatorów powinna zwiększyć czułość TPR rozpoznawania zmian czerniakowych przy jednoczesnym wzroście drugiego wskaźnika – TNR.

Wskaźniki precyzji PPV i NPV w rozpoznawaniu klas świadczą o tendencji klasyfikatora do przyznawania badanemu obiektowi etykiety danej klasy. Wyniki badań dotyczące obu rodzajów precyzji są przedstawione na rys. 5.10. Oba wskaźniki wykazały się wartością powyżej 93%. Wyraźna jest przewaga precyzji rozpoznawania brodawki względem czerniaka. Najlepsze zrównoważenie wartości obu wskaźników uzyskano dla sieci Resnet50 (obie wartości nieco powyżej 98%). Największe niezrównoważenie obu wartości precyzji obserwuje się dla sieci Xception i Efficientnetb0.



Rys. 5.10 – Wykres porównujący precyzję rozpoznawania obu klas wszystkich badanych modeli klasyfikatorów. Wskaźnik PPV dotyczy czerniaka, a NPV brodawki łojotokowej.

W tabeli 5.2 przedstawiono zestawienie wszystkich analizowanych wskaźników jakości dla każdego analizowanego modelu klasyfikatora CNN, wzbogacone o macierz pomyłek (Conf. Matrix). Wyniki dotyczą zbioru NIO przy założonym doborze procentowym ilości zamrożonych warstw w procesie douczania sieci. Zamrożenie 30% początkowych warstw zastosowano dla sieci: Squeezenet, Inceptionv3, Densenet201, Mobilenetv2, Resnet101, Shufflenet, Darknet19, Darknet53, Efficientnetb0 i Vgg19. Dla pozostałych dziewięciu sieci proces douczania obejmował wszystkie warstwy w strukturze CNN.

Kolumna dotycząca macierzy pomyłek prezentuje wyniki 5-krotnej walidacji krzyżowej i dotyczy wyłącznie danych testujących nie uczestniczących w uczeniu. Opisanie w niniejszym rozdziale wyniki klasyfikacji obrazów czerniaka ze zbioru NIO oraz ich szczegółowa analiza dały podstawę do zbadania efektywności różnych rozwiązań zespołu klasyfikatorów, dla polepszenia jakości końcowych wyników systemu klasyfikacyjnego.

Tabela 5.2 – Wyniki zbiorcze dotyczące wartości wskaźników jakości rozpoznania czerniaka od brodawki dla różnych indywidualnych rozwiązań sieciowych CNN na zbiorze NIO.

| Sieć CNN | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|-------------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|-----|
| Squeezenet | 0.9894 | 0.9634 | 0.9776 | 0.9464 | 0.9562 | 0.9725 | 131 | 3 |
| | | | | | | | 6 | 106 |
| Googlenet | 0.9915 | 0.9715 | 0.9627 | 0.9821 | 0.9847 | 0.9565 | 129 | 5 |
| | | | | | | | 2 | 110 |
| Inceptionv3 | 0.9982 | 0.9715 | 0.9925 | 0.9464 | 0.9568 | 0.9907 | 133 | 1 |
| | | | | | | | 6 | 106 |
| Densenet201 | 0.9981 | 0.9797 | 0.9925 | 0.9643 | 0.9708 | 0.9908 | 133 | 1 |
| | | | | | | | 4 | 108 |
| Mobilenetv2 | 0.9968 | 0.9715 | 0.9925 | 0.9464 | 0.9568 | 0.9907 | 133 | 1 |
| | | | | | | | 6 | 106 |
| Resnet18 | 0.9979 | 0.9715 | 0.9925 | 0.9464 | 0.9568 | 0.9907 | 133 | 1 |
| | | | | | | | 6 | 106 |
| Resnet50 | 0.9985 | 0.9837 | 0.9851 | 0.9821 | 0.9851 | 0.9821 | 132 | 2 |
| | | | | | | | 2 | 110 |
| Resnet101 | 0.9995 | 0.9878 | 1.0000 | 0.9732 | 0.9781 | 1.0000 | 134 | 0 |
| | | | | | | | 3 | 109 |
| Xception | 0.9958 | 0.9715 | 1.0000 | 0.9375 | 0.9504 | 1.0000 | 134 | 0 |
| | | | | | | | 7 | 105 |
| Inceptionresnetv2 | 0.9887 | 0.9593 | 0.9776 | 0.9375 | 0.9493 | 0.9722 | 131 | 3 |
| | | | | | | | 7 | 105 |
| Shufflenet | 0.9918 | 0.9593 | 0.9701 | 0.9464 | 0.9559 | 0.9636 | 130 | 4 |
| | | | | | | | 6 | 106 |
| Nasnetmobile | 0.9913 | 0.9472 | 0.9701 | 0.9196 | 0.9353 | 0.9626 | 130 | 4 |
| | | | | | | | 9 | 103 |
| Nasnetlarge | 0.9968 | 0.9715 | 0.9851 | 0.9554 | 0.9635 | 0.9817 | 132 | 2 |
| | | | | | | | 5 | 107 |
| Darknet19 | 0.9963 | 0.9715 | 0.9627 | 0.9821 | 0.9847 | 0.9565 | 129 | 5 |
| | | | | | | | 2 | 110 |
| Darknet53 | 0.9984 | 0.9797 | 0.9851 | 0.9732 | 0.9778 | 0.9820 | 132 | 2 |
| | | | | | | | 3 | 109 |
| Efficientnetb0 | 0.9983 | 0.9797 | 1.0000 | 0.9554 | 0.9640 | 1.0000 | 134 | 0 |
| | | | | | | | 5 | 107 |
| Alexnet | 0.9969 | 0.9675 | 0.9701 | 0.9643 | 0.9701 | 0.9643 | 130 | 4 |
| | | | | | | | 4 | 108 |
| Vgg16 | 0.9928 | 0.9675 | 0.9776 | 0.9554 | 0.9632 | 0.9727 | 131 | 3 |
| | | | | | | | 5 | 107 |
| Vgg19 | 0.9957 | 0.9675 | 0.9627 | 0.9732 | 0.9773 | 0.9561 | 129 | 5 |
| | | | | | | | 3 | 109 |

5.1.2. Wyniki klasyfikacji zespołów zbudowanych na pojedynczej sieci CNN rozszerzonej o SVM i RF – zbiór NIO

Zespół klasyfikatorów zbudowany na bazie struktur CNN może przybierać różne formy wykorzystując jedynie sieci głębokie jako członków zespołu lub łącząc struktury głębokie klasyfikatorów z rozwiązaniami klasycznymi. W pierwszym etapie przebadano możliwość połączenia sieci głębokich z klasyfikatorami klasycznymi. Ograniczono się przy tym do dwóch takich rozwiązań: sieć SVM i las losowy drzew decyzyjnych RF, uważanych powszechnie za najbardziej efektywne narzędzia w trybie klasyfikacyjnym.

Współpraca klasyfikatora klasycznego z siecią głęboką CNN odbywa się w kilku fazach. W pierwszej sieć CNN dokonuje ekstrakcji deskryptorów numerycznych z obrazów wejściowych przy zastosowaniu lokalnie połączonych warstw konwolucyjnych. Na końcu tego etapu, po wypłaszczeniu, dane przyjmują formę wektorową, a każdy jego element stanowi pojedynczy deskryptor analizowanego obrazu. Jest to wektor o dużej wymiarowości. W tabeli 5.3 przedstawiono wymiarowość takiego wektora w różnych rozwiązaniach sieci CNN.

Tabela 5.3 – Populacja deskryptorów generowanych przez różne struktury sieci CNN.

| Sieć CNN | Liczba deskryptorów |
|-------------------|---------------------|
| Squeezenet | 1000 |
| Googlenet | 1024 |
| Inceptionv3 | 2048 |
| Densenet201 | 1920 |
| Mobilenetv2 | 1280 |
| Resnet18 | 512 |
| Resnet50 | 2048 |
| Resnet101 | 2048 |
| Xception | 2048 |
| Inceptionresnetv2 | 1536 |
| Shufflenet | 544 |
| Nasnetmobile | 1056 |
| Nasnetlarge | 4032 |
| Darknet19 | 1000 |
| Darknet53 | 1024 |
| Efficientnetb0 | 1280 |
| Alexnet | 4096 |
| Vgg16 | 4096 |
| Vgg19 | 4096 |

Widoczne są duże różnice w liczebności tych deskryptorów. Na przykład w Alexnet wektor cech ma wymiar 4096, a w sieci Resnet18 tylko 512. W fazie drugiej, wektor zostaje

poddany selekcji cech dla ograniczenia jego wymiarowości. Przy małym wymiarze tego wektora możliwe jest również pominięcie tej fazy i użycie wektora w pełnym wymiarze. Utworzone w ten sposób cechy stanowią atrybuty wejściowe dla trzeciej fazy współpracy, gdzie do klasyfikacji wykorzystuje się rozwiązania inne niż softmax. Stąd powstała koncepcja włączenia dodatkowych klasycznych klasyfikatorów SVM i RF w skład zespołu.

Dla zwiększenia niezależności działania poszczególnych jednostek, wyselekcjonowany przez sieć CNN wektor deskryptorów (potencjalnych cech) poddano procesowi wyboru najważniejszych cech przy zastosowaniu wybranych deterministycznych metod selekcji. Celem tego procesu jest wyłonienie z całego zbioru deskryptorów tych które najsilniej wpływają na prawidłową decyzję klasyfikatora. Jednocześnie odrzucone zostają cechy wpływające na proces klasyfikacji w sposób negatywny lub obojętny.

Badania przeprowadzono dla kilku wybranych sieci CNN. Celem było określenie optymalnej liczby cech wybieranych dla każdej kombinacji klasycznego klasyfikatora (SVM, RF) i metody selekcji cech. W wyniku wstępnych eksperymentów wzięto pod uwagę następujące metody selekcji: metoda dyskryminacyjna Fishera (FD), T-test Studenta (T-test), test statystyczny Kolmogorova-Smirnova (KS), test statystyczny Kruskala-Wallisa (KW), metoda korelacji danych z klasą (CDC), selekcja krokowa (SWF), metoda analizy najbliższych sąsiadów (NNA) oraz metoda Relief (REL).

Przeprowadzono szereg eksperymentów, których celem było wyłonienie ze zbioru deskryptorów ograniczonej liczby cech diagnostycznych gwarantującej najlepsze wyniki działania klasyfikatora klasycznego. Wyniki tych eksperymentów prezentujące wybrane liczby cech dla sieci SVM i RF przedstawiono w tabeli 5.4. Dotyczą one sieci Inceptionv3. Badania przeprowadzono dla zbioru NIO, przy czym jakość rozwiązania dotyczyła średniej dokładności przy wykorzystaniu 5-krotnej walidacji krzyżowej.

Tabela 5.4 – Wybrana liczba cech dla każdej z badanych metod selekcji dla sieci SVM i RF (Inceptionv3).

| Metoda selekcji | Liczba cech dla sieci SVM | Liczba cech dla sieci RF |
|---------------------------------------------|---------------------------|--------------------------|
| Metoda dyskryminacyjna Fishera (FD) | 120 | 40 |
| T-test Studenta (T-test) | 100 | 30 |
| Test statystyczny Kolmogorova-Smirnova (KS) | 100 | 40 |
| Test statystyczny Kruskala-Wallisa (KW) | 100 | 40 |
| Metoda korelacji danych z klasą (CDC) | 100 | 40 |
| Selekcja krokowa (SWF) | 140 | 40 |
| Metoda analizy najbliższych sąsiadów (NNA) | 100 | 20 |
| Metoda Relief (REL) | 100 | 100 |

Połączenie rezultatów selekcji przy pomocy wybranej metody z klasyfikatorem SVM i RF tworzyło jednostkę klasyfikacyjną włączoną do zespołu. W skład zespołu wchodzi również sieć głęboka użyta do generacji cech, ale wykorzystująca softmax jako jednostkę klasyfikacyjną przy losowym wyborze 50% deskryptorów użytych jako cechy procesu w każdej iteracji uczącej. Stąd skład zespołu przyjmuje formę przedstawioną w rozdziale czwartym (rys. 4.3).

Istotność metod selekcji cech potwierdza fakt, że stosunkowo niewielka liczba cech była wystarczająca do uzyskania dokładności klasyfikacji powyżej 90%. W przypadku SVM uzyskano dokładność powyżej 96%, a więc zbliżoną do dokładności uzyskiwanej indywidualnie w sieciach głębokich stosujących selekcję losową oraz softmax. Wyniki uzyskane przy użyciu klasyfikatora RF okazały się słabsze (dokładność nie przekraczała 92%) w porównaniu do tych uzyskanych przy pomocy sieci bazowej CNN z klasyfikatorem SVM. Stąd zasadne wydaje się zbadać jaki jest wpływ pominięcia sieci RF w składzie zespołu. Ostatecznie zbadano trzy rozwiązania zespołu.

1. Zespół składający się ze zbioru 8 klasyfikatorów SVM, 8 klasyfikatorów RF (obie sieci zasilane różnym zbiorem cech diagnostycznych) oraz 1 klasyfikatora głębokiego typu softmax współpracującego w zespole z klasyfikatorami SVM i RF. Badania przeprowadzono w oparciu o różne struktury klasyfikatora głębokiego.
2. Zespół złożony ze zbioru 8 klasyfikatorów SVM (sieci zasilane różnym zbiorem cech diagnostycznych) oraz 1 klasyfikatora głębokiego typu softmax współpracującego z klasyfikatorami SVM. Badania przeprowadzono w oparciu o różne struktury klasyfikatora głębokiego.
3. Zespół złożony wyłącznie z sieci głębokich CNN stosujących klasyfikator typu softmax. W tym przypadku zbadano różne składy zespołu, wybierając jednostki o najlepszych wartościach wskaźników jakości dotyczących ACC, AUC i TPR.

W tabeli 5.5 przedstawiono zestawienie wskaźników jakości dla każdego analizowanego zespołu klasyfikatorów głębokich wykorzystujących jednocześnie SVM i RF. Wyniki dotyczą danych testowych uzyskanych na zbiorze NIO przy wykorzystaniu 5-krotnej walidacji krzyżowej. Uzyskane wyniki najlepszego zespołu na bazie sieci Darknet19 nie poprawiły wyników najlepszego pojedynczego klasyfikatora głębokiego (Resnet101 o ACC = 98.23%, AUC = 0.9985 i TPR = 1.0).

Tabela 5.5 – Wyniki zespołów stosujących współpracę sieci SVM i RF z różnymi strukturami głębokimi CNN (zbiór NIO).

| Sieć CNN | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|-------------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|-----|
| SqueezeNet | 0.9864 | 0.9472 | 0.9701 | 0.9196 | 0.9353 | 0.9626 | 130 | 4 |
| | | | | | | | 9 | 103 |
| GoogLeNet | 0.9774 | 0.9715 | 0.9701 | 0.9732 | 0.9774 | 0.9646 | 130 | 4 |
| | | | | | | | 3 | 109 |
| Inceptionv3 | 0.9938 | 0.9715 | 0.9925 | 0.9464 | 0.9568 | 0.9907 | 133 | 1 |
| | | | | | | | 6 | 106 |
| DenseNet201 | 0.9866 | 0.9756 | 0.9925 | 0.9554 | 0.9638 | 0.9907 | 133 | 1 |
| | | | | | | | 5 | 107 |
| MobileNetV2 | 0.9953 | 0.9756 | 1.0000 | 0.9464 | 0.9571 | 1.0000 | 134 | 0 |
| | | | | | | | 6 | 106 |
| ResNet18 | 0.9891 | 0.9797 | 0.9925 | 0.9643 | 0.9708 | 0.9908 | 133 | 1 |
| | | | | | | | 4 | 108 |
| ResNet50 | 0.9887 | 0.9675 | 0.9776 | 0.9554 | 0.9632 | 0.9727 | 131 | 3 |
| | | | | | | | 5 | 107 |
| ResNet101 | 0.9893 | 0.9593 | 0.9701 | 0.9464 | 0.9559 | 0.9636 | 130 | 4 |
| | | | | | | | 6 | 106 |
| Xception | 0.9834 | 0.9593 | 1.0000 | 0.9107 | 0.9306 | 1.0000 | 134 | 0 |
| | | | | | | | 10 | 102 |
| InceptionResNetV2 | 0.9853 | 0.9553 | 0.9776 | 0.9286 | 0.9424 | 0.9720 | 131 | 3 |
| | | | | | | | 8 | 104 |
| ShuffleNet | 0.9869 | 0.9675 | 0.9701 | 0.9643 | 0.9701 | 0.9643 | 130 | 4 |
| | | | | | | | 4 | 108 |
| NASNetMobile | 0.9831 | 0.9553 | 0.9701 | 0.9375 | 0.9489 | 0.9633 | 130 | 4 |
| | | | | | | | 7 | 105 |
| NASNetLarge | 0.9819 | 0.9634 | 0.9851 | 0.9375 | 0.9496 | 0.9813 | 132 | 2 |
| | | | | | | | 7 | 105 |
| DarkNet19 | 0.9903 | 0.9797 | 0.9701 | 0.9911 | 0.9924 | 0.9652 | 130 | 4 |
| | | | | | | | 1 | 111 |
| DarkNet53 | 0.9982 | 0.9837 | 1.0000 | 0.9643 | 0.9710 | 1.0000 | 134 | 0 |
| | | | | | | | 4 | 108 |
| EfficientNetB0 | 0.9926 | 0.9472 | 0.9925 | 0.8929 | 0.9172 | 0.9901 | 133 | 1 |
| | | | | | | | 12 | 100 |
| AlexNet | 0.9830 | 0.9756 | 0.9776 | 0.9732 | 0.9776 | 0.9732 | 131 | 3 |
| | | | | | | | 3 | 109 |
| VGG16 | 0.9886 | 0.9634 | 0.9701 | 0.9554 | 0.9630 | 0.9640 | 130 | 4 |
| | | | | | | | 5 | 107 |
| VGG19 | 0.9900 | 0.9715 | 0.9701 | 0.9732 | 0.9774 | 0.9646 | 130 | 4 |
| | | | | | | | 3 | 109 |

Tabela 5.6 prezentuje wyniki współpracy wyłącznie klasyfikatorów SVM z różnymi strukturami sieci głębokich CNN (wyłączono z zespołu klasyfikatory RF). Tym

razem najlepszymi wynikami charakteryzuje się współpraca sieci SVM z siecią CNN o strukturze Inceptionv3. Uzyskana wartość średniej dokładności ACC = 98.37% przewyższa nieco najlepsze osiągnięcie sieci indywidualnej Resnet101.

Tabela 5.6 – Wyniki zespołów stosujących współpracę sieci SVM z różnymi strukturami głębokimi CNN bez uwzględnienia sieci RF (zbiór NIO).

| Sieć CNN | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|--------------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|-----|
| Squeezenet | 0.9639 | 0.9553 | 0.9851 | 0.9196 | 0.9362 | 0.9810 | 132 | 2 |
| | | | | | | | 9 | 103 |
| Googlenet | 0.9756 | 0.9715 | 0.9701 | 0.9732 | 0.9774 | 0.9646 | 130 | 4 |
| | | | | | | | 3 | 109 |
| Inceptionv3 | 0.9914 | 0.9837 | 0.9925 | 0.9732 | 0.9779 | 0.9909 | 133 | 1 |
| | | | | | | | 3 | 109 |
| Densenet201 | 0.9816 | 0.9756 | 1.0000 | 0.9464 | 0.9571 | 1.0000 | 134 | 0 |
| | | | | | | | 6 | 106 |
| Mobilenetv2 | 0.9773 | 0.9756 | 1.0000 | 0.9464 | 0.9571 | 1.0000 | 134 | 0 |
| | | | | | | | 6 | 106 |
| Resnet18 | 0.9828 | 0.9797 | 0.9925 | 0.9643 | 0.9708 | 0.9908 | 133 | 1 |
| | | | | | | | 4 | 108 |
| Resnet50 | 0.9863 | 0.9756 | 0.9851 | 0.9643 | 0.9706 | 0.9818 | 132 | 2 |
| | | | | | | | 4 | 108 |
| Resnet101 | 0.9895 | 0.9675 | 0.9776 | 0.9554 | 0.9632 | 0.9727 | 131 | 3 |
| | | | | | | | 5 | 107 |
| Xception | 0.9807 | 0.9634 | 1.0000 | 0.9196 | 0.9371 | 1.0000 | 134 | 0 |
| | | | | | | | 9 | 103 |
| Inceptionresnetv2 | 0.9796 | 0.9512 | 0.9851 | 0.9107 | 0.9296 | 0.9808 | 132 | 2 |
| | | | | | | | 10 | 102 |
| Shufflenet | 0.9811 | 0.9675 | 0.9776 | 0.9554 | 0.9632 | 0.9727 | 131 | 3 |
| | | | | | | | 5 | 107 |
| Nasnetmobile | 0.9687 | 0.9512 | 0.9701 | 0.9286 | 0.9420 | 0.9630 | 130 | 4 |
| | | | | | | | 8 | 104 |
| Nasnetlarge | 0.9759 | 0.9634 | 0.9851 | 0.9375 | 0.9496 | 0.9813 | 132 | 2 |
| | | | | | | | 7 | 105 |
| Darknet19 | 0.9872 | 0.9715 | 0.9627 | 0.9821 | 0.9847 | 0.9565 | 129 | 5 |
| | | | | | | | 2 | 110 |
| Darknet53 | 0.9906 | 0.9756 | 0.9851 | 0.9643 | 0.9706 | 0.9818 | 132 | 2 |
| | | | | | | | 4 | 108 |
| Efficientnetb0 | 0.9789 | 0.9593 | 0.9776 | 0.9375 | 0.9493 | 0.9722 | 131 | 3 |
| | | | | | | | 7 | 105 |
| Alexnet | 0.9784 | 0.9756 | 0.9776 | 0.9732 | 0.9776 | 0.9732 | 131 | 3 |
| | | | | | | | 3 | 109 |
| Vgg16 | 0.9736 | 0.9715 | 0.9776 | 0.9643 | 0.9704 | 0.9730 | 131 | 3 |
| | | | | | | | 4 | 108 |
| Vgg19 | 0.9757 | 0.9675 | 0.9701 | 0.9643 | 0.9701 | 0.9643 | 130 | 4 |
| | | | | | | | 4 | 108 |

Analiza wartości wskaźników jakości odpowiadających obu rozwiązaniom zespołowym pokazuje minimalne różnice wartości wskaźników jakości. Usunięcie klasyfikatora RF z zespołu w minimalnym stopniu polepszyło jakość uzyskanych wskazań w zespole. Przykładowo wartość średnia dokładności ACC wszystkich 19 rozwiązań zmieniła się z $96.68\% \pm 1.08\%$ (zespół SVM + RF + CNN) na $96.85\% \pm 0.93\%$ (zespół SVM + CNN). Podobne relacje dotyczą pozostałych wskaźników jakości. Tym nie mniej również tak zorganizowany zespół nie pozwolił uzyskać wyników zdecydowanie lepszych od najlepszego indywidualnego klasyfikatora CNN (Resnet101 z tabeli 5.2).

5.1.3. Wyniki klasyfikacji zespołów złożonych wyłącznie z sieci CNN – zbiór NIO

W przypadku trzeciego rozwiązania zbadano różne możliwości łączenia struktur sieciowych CNN w jeden zespół klasyfikacyjny. Skład zespołu tworzyć będą kombinacje najlepszych klasyfikatorów indywidualnych pod względem wybranych wskaźników jakości, w szczególności AUC, ACC oraz TPR. W wyniku szeregu eksperymentów numerycznych wyłoniono 12 takich rozwiązań, przedstawionych w tabeli 5.7. Ostatnia kolumna w tabeli reprezentuje wskaźniki jakości na podstawie których wyselekcjonowano najlepsze struktury CNN tworzące zespół. Ostatni wiersz odpowiada włączeniu wszystkich 19 badanych struktur w skład zespołu.

Tabela 5.7 – Składy zespołów sieci CNN wyłonione w wyniku wstępnych eksperymentów (zbiór NIO).

| Nr zespołu | Sieci CNN wchodzące w skład zespołu | Liczba członków | Podstawa wyboru członków zespołu |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|----------------------------------|
| 1 | Resnet50, Resnet101, Darknet53 | 3 | ACC i AUC |
| 2 | Resnet101, Xception, Efficientnetb0 | 3 | TPR |
| 3 | Densenet201, Resnet50, Resnet101, Darknet53, Efficientnetb0 | 5 | ACC |
| 4 | Inceptionv3, Resnet50, Resnet101, Darknet53, Efficientnetb0 | 5 | AUC |
| 5 | Inceptionv3, Densenet201, Resnet101, Xception, Efficientnetb0 | 5 | TPR |
| 6 | Inceptionv3, Densenet201, Mobilenetv2, Resnet18, Resnet50, Resnet101, Nasnetlarge, Darknet19, Darknet53, Efficientnetb0 | 10 | ACC |
| 7 | Inceptionv3, Densenet201, Mobilenetv2, Resnet18, Resnet50, Resnet101, Nasnetlarge, Darknet53, Efficientnetb0, Alexnet | 10 | AUC |
| 8 | Inceptionv3, Densenet201, Mobilenetv2, Resnet18, Resnet50, Resnet101, Xception, Nasnetlarge, Darknet53, Efficientnetb0 | 10 | TPR |
| 9 | Googlenet, Inceptionv3, Densenet201, Mobilenetv2, Resnet18, Resnet50, Resnet101, Xception, Nasnetlarge, Darknet19, Darknet53, Efficientnetb0, Alexnet, Vgg16, Vgg19 | 15 | ACC |
| 10 | Inceptionv3, Densenet201, Mobilenetv2, Resnet18, Resnet50, Resnet101, Xception, Shufflenet, Nasnetlarge, Darknet19, Darknet53, Efficientnetb0, Alexnet, Vgg16, Vgg19 | 15 | AUC |

| Nr zespołu | Sieci CNN wchodzące w skład zespołu | Liczba członków | Podstawa wyboru członków zespołu |
|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|----------------------------------|
| 11 | Squeezenet, Inceptionv3, Densenet201, Mobilenetv2, Resnet18, Resnet50, Resnet101, Xception, Inceptionresnetv2, Shufflenet, Nasnetlarge, Darknet53, Efficientnetb0, Alexnet, Vgg16 | 15 | TPR |
| 12 | Wszystkie sieci CNN | 19 | - |

W tabeli 5.8 przedstawiono wyniki rozpoznania klas w zbiorze NIO uzyskane przez zespoły zbudowane wyłącznie na bazie różnych struktur sieci głębokich stosujących klasyfikatory typu softmax. W pierwszej kolumnie pokazano numer aktualnie badanego składu zespołu.

Tabela 5.8 – Wyniki zespołów zbudowanych wyłącznie z różnych struktur sieci głębokich (zbiór NIO).

| Zespół | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|--------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|-----|
| 1 | 0.9837 | 0.9908 | 0.9851 | 0.9821 | 0.9851 | 0.9821 | 132 | 2 |
| | | | | | | | 2 | 110 |
| 2 | 0.9797 | 0.9953 | 1.0000 | 0.9554 | 0.9640 | 1.0000 | 134 | 0 |
| | | | | | | | 5 | 107 |
| 3 | 0.9919 | 0.9953 | 1.0000 | 0.9821 | 0.9853 | 1.0000 | 134 | 0 |
| | | | | | | | 2 | 110 |
| 4 | 0.9919 | 0.9953 | 1.0000 | 0.9821 | 0.9853 | 1.0000 | 134 | 0 |
| | | | | | | | 2 | 110 |
| 5 | 0.9837 | 0.9953 | 1.0000 | 0.9643 | 0.9710 | 1.0000 | 134 | 0 |
| | | | | | | | 4 | 108 |
| 6 | 0.9878 | 0.9948 | 1.0000 | 0.9732 | 0.9781 | 1.0000 | 134 | 0 |
| | | | | | | | 3 | 109 |
| 7 | 0.9878 | 0.9948 | 1.0000 | 0.9732 | 0.9781 | 1.0000 | 134 | 0 |
| | | | | | | | 3 | 109 |
| 8 | 0.9878 | 0.9950 | 1.0000 | 0.9732 | 0.9781 | 1.0000 | 134 | 0 |
| | | | | | | | 3 | 109 |
| 9 | 0.9878 | 0.9948 | 1.0000 | 0.9732 | 0.9781 | 1.0000 | 134 | 0 |
| | | | | | | | 3 | 109 |
| 10 | 0.9878 | 0.9947 | 1.0000 | 0.9732 | 0.9781 | 1.0000 | 134 | 0 |
| | | | | | | | 3 | 109 |
| 11 | 0.9878 | 0.9945 | 1.0000 | 0.9732 | 0.9781 | 1.0000 | 134 | 0 |
| | | | | | | | 3 | 109 |
| 12 | 0.9837 | 0.9944 | 0.9925 | 0.9732 | 0.9779 | 0.9909 | 133 | 1 |
| | | | | | | | 3 | 109 |

Zastosowanie wyłącznie sieci CNN o różnych strukturach tworzących zespół poprawiło jakość rozwiązania w stosunku do najlepszego klasyfikatora indywidualnego (Resnet101). Najlepszy zespół nr 4 utworzony przez Inceptionv3, Resnet50, Resnet101, Darknet53, Efficientnetb0 zwiększył dokładność rozpoznania klas z wartości

ACC = 98.23% (Resnet101) do 99.53%, a średnia wszystkich rozwiązań zespołowych wzrosła do wartości $ACC_{sr} = 99.46\% \pm 0.12\%$, co zdecydowanie góruje nad najlepszym średnim wynikiem zespołu bazującego na sieci SVM ($ACC_{sr} = 96.85\% \pm 0.93\%$). Wynik ten również przewyższa średni rezultat klasyfikatorów indywidualnych CNN z tabeli 5.2 dla których $ACC_{sr} = 97.03\% \pm 0.88\%$.

Uzyskane wyniki badań wskazują na przewagę zintegrowanego zespołu zbudowanego wyłącznie z sieci głębokich. Ważnym elementem w tworzeniu takiego zespołu jest określenie optymalnej liczby wybranych członków gwarantujących najlepsze rozwiązanie pod względem zadeklarowanych wskaźników jakości. Stąd w dalszych badaniach przy wykorzystaniu zbioru ISIC skoncentrowano się na tworzeniu optymalnego składu zespołu bazującego wyłącznie na strukturach głębokich CNN.

Należy nadmienić, że rezultaty zespołów przedstawione powyżej zostały uzyskane przy fuzji wyników indywidualnych członków bazującej na wartości prawdopodobieństwa klasowego (wyniki przed binaryzacją). Taki system fuzji został wybrany w wyniku wstępnych eksperymentów, w których porównywano skuteczność zespołu integrowanego na bazie prawdopodobieństwa i wyników binarnych. Średnia przewaga systemu bazującego na prawdopodobieństwie oceniona została na poziomie od 1 do 2 punktów procentowych w zależności od wskaźnika jakościowego.

5.2. Wyniki rozpoznania czerniaka na zbiorze ISIC

Drugim analizowanym zbiorem danych czerniaka jest ISIC. Reprezentuje dużo większą bazę danych, zawierającą 945 obrazów czerniaka oraz 1543 obrazów różnorodnych znamion niebędących czerniakiem. Obrazy tworzące ten zbiór pochodzą z różnych części ciała i zawierają różne rodzaje zmian łagodnych różniących się od czerniaka. Zbiór obrazów został wytworzony przez wiele różnych instytucji badawczych na świecie. Sposób wstępnego przetworzenia tych obrazów przez te jednostki jest również zróżnicowany. Stąd obrazy tworzące tę bazę tworzą znacznie trudniejsze zadanie rozpoznania klasy.

Na podstawie doświadczeń ze zbiorem NIO, z których wynikała zdecydowana przewaga klasyfikatorów CNN nad rozwiązaniami klasycznymi (SVM + las losowy) badania nad tym zbiorem skoncentrowały się na zastosowaniu wyłącznie sieci klasyfikatorów CNN współpracujących w zespole. Dla stworzenia optymalnego składu zespołu klasyfikacyjnego w pierwszej kolejności przebadano indywidualną skuteczność działania różnych rozwiązań strukturalnych tych sieci. Na podstawie tych wyników został dobrany

odpowiedni skład zespołu. Wszystkie badania zostały przeprowadzone przy wykorzystaniu 5-krotnej walidacji krzyżowej, zarówno w trybie indywidualnych klasyfikatorów jak i w przypadku zespołu.

5.2.1. Wyniki klasyfikacji indywidualnych modeli CNN – zbiór ISIC

Wyniki działania indywidualnych sieci CNN z klasyfikatorem typu softmax o różnych strukturach na zbiorze ISIC przedstawiono w tabeli 5.9. Wykorzystano technikę transfer learning z douczaniem określonej procentowo liczby warstw sieci. Zamrożenie 30% początkowych warstw zastosowano dla sieci: Squeezenet, Inceptionv3, Densenet201, Mobilenetv2, Resnet101, Shufflenet, Darknet19, Darknet53, Efficientnetb0 i Vgg19. Dla pozostałych dziewięciu sieci proces douczania obejmował wszystkie warstwy struktury CNN. Wyniki dotyczą pola AUC pod krzywą ROC, dokładności, czułości, specyficzności i precyzji rozpoznania obu klas. Dodatkowo zamieszczono macierze pomyłek oraz czas działania poszczególnych rozwiązań (ze względu na zastosowaną technikę 5-krotnej walidacji krzyżowej czasy te są sumą czasu uczenia i testowania w tym trybie).

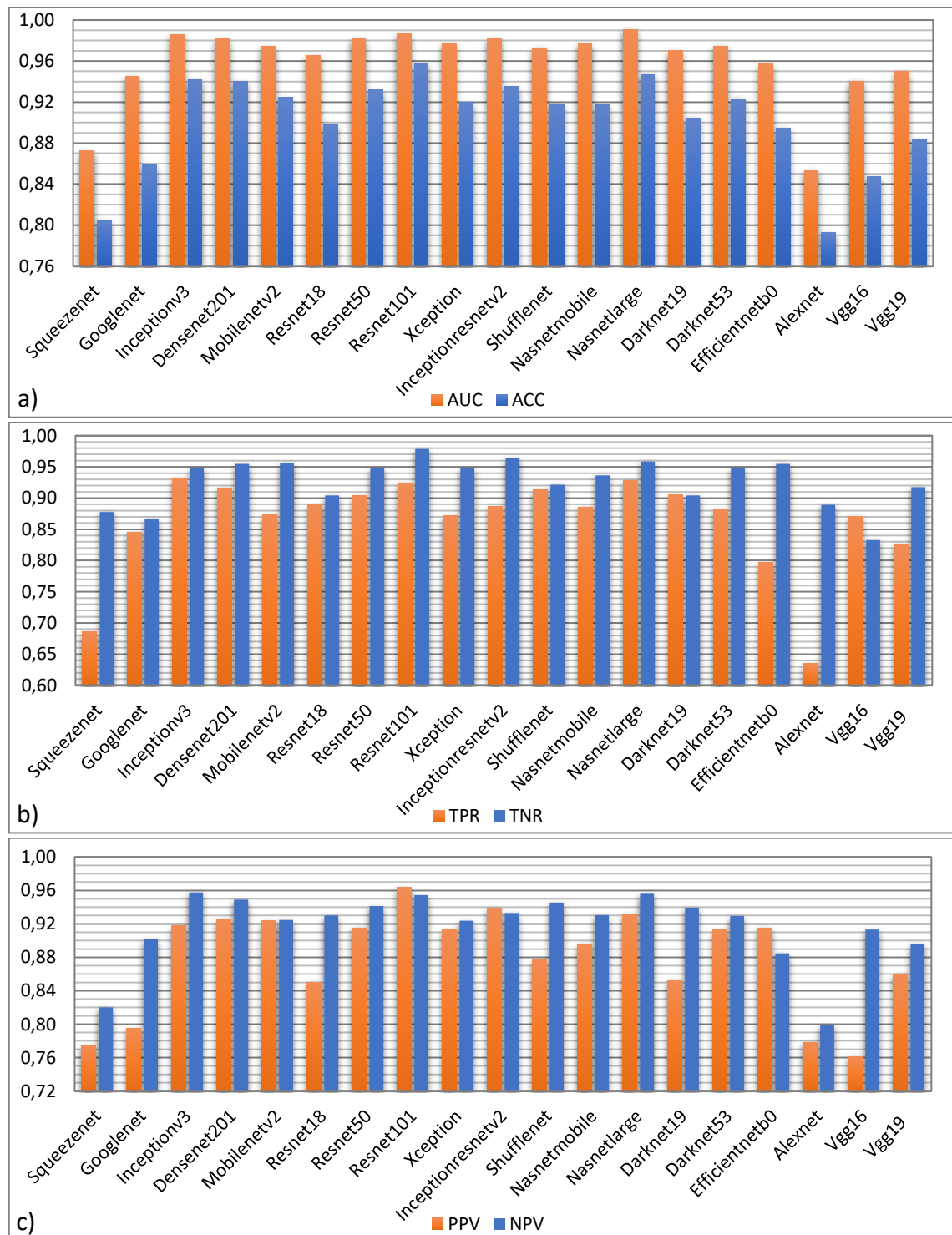
Wartości wymienionych miar pokazują znaczną rozpiętość poszczególnych rozwiązań. Najmniejszą dokładnością wykazała się sieć Alexnet (ACC = 79.26%) i Squeezenet (ACC = 80.47%). Takie wyniki indywidualne wykluczają ich zastosowanie w zespole. Najlepsze wyniki ACC dotyczą sieci Resnet101 (ACC = 95.78%). Również pozostałe wskaźniki jakości dla tej sieci (czułość = 92.38%, specyficzność = 97.86%, precyzje PPV = 96.36% i NPV = 95.45%) są na wysokim, dobrze zbalansowanym poziomie. Należy przy tym zauważyć, że bardzo dobry wynik sieci Resnet101 został okupiony długim czasem uczenia (1 godzina i 22 minuty), podczas gdy czas uczenia Alexnet to zaledwie niecałe 17 minut. Bardzo długi czas uczenia sieci Nasnetlarge (ponad 94 godziny) doprowadził do dobrych wskaźników jakości (ACC = 94.69%). Biorąc pod uwagę, że użytkowanie sieci w trybie testowym odbywa się z użyciem już wytrenowanej struktury długi czas uczenia nie eliminuje tego rozwiązania z udziału w zespole klasyfikatorów.

Tabela 5.9 – Wyniki zbiorcze dotyczące wartości wskaźników jakości rozpoznania czerniaka dla różnych indywidualnych rozwiązań sieciowych CNN na zbiorze ISIC.

| Sieć CNN | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | | Czas |
|--------------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|-------------|----------|
| SqueezeNet | 0.8726 | 0.8047 | 0.6857 | 0.8775 | 0.7742 | 0.8201 | 648 189 | 297 1354 | 00:12:16 |
| GoogLeNet | 0.9451 | 0.8585 | 0.8455 | 0.8665 | 0.7950 | 0.9016 | 799 206 | 146 1337 | 00:28:06 |
| Inceptionv3 | 0.9861 | 0.9421 | 0.9312 | 0.9488 | 0.9176 | 0.9575 | 880 79 | 65 1464 | 01:18:29 |
| DenseNet201 | 0.9813 | 0.9401 | 0.9164 | 0.9546 | 0.9252 | 0.9491 | 866 70 | 79 1473 | 04:01:39 |
| MobileNetV2 | 0.9743 | 0.9244 | 0.8730 | 0.9559 | 0.9239 | 0.9248 | 825 68 | 120 1475 | 00:40:27 |
| ResNet18 | 0.9655 | 0.8983 | 0.8889 | 0.9041 | 0.8502 | 0.9300 | 840 148 | 105 1395 | 00:20:07 |
| ResNet50 | 0.9817 | 0.9317 | 0.9037 | 0.9488 | 0.9153 | 0.9415 | 854 79 | 91 1464 | 00:56:18 |
| ResNet101 | 0.9865 | 0.9578 | 0.9238 | 0.9786 | 0.9636 | 0.9545 | 873 33 | 72 1510 | 01:22:52 |
| Xception | 0.9777 | 0.9196 | 0.8720 | 0.9488 | 0.9125 | 0.9237 | 824 79 | 121 1464 | 01:35:26 |
| InceptionResNetV2 | 0.9819 | 0.9349 | 0.8868 | 0.9644 | 0.9384 | 0.9329 | 838 55 | 107 1488 | 05:31:55 |
| ShuffleNet | 0.9731 | 0.9184 | 0.9132 | 0.9216 | 0.8770 | 0.9455 | 863 121 | 82 1422 | 00:44:49 |
| NasNetMobile | 0.9767 | 0.9172 | 0.8857 | 0.9365 | 0.8952 | 0.9305 | 837 98 | 108 1445 | 06:51:20 |
| NasNetLarge | 0.9909 | 0.9469 | 0.9280 | 0.9585 | 0.9320 | 0.9560 | 877 64 | 68 1479 | 94:05:01 |
| DarkNet19 | 0.9702 | 0.9043 | 0.9048 | 0.9041 | 0.8524 | 0.9394 | 855 148 | 90 1395 | 00:31:15 |
| DarkNet53 | 0.9745 | 0.9232 | 0.8825 | 0.9482 | 0.9125 | 0.9295 | 834 80 | 111 1463 | 01:00:09 |
| EfficientNetB0 | 0.9572 | 0.8947 | 0.7968 | 0.9546 | 0.9149 | 0.8847 | 753 70 | 192 1473 | 01:40:40 |
| AlexNet | 0.8539 | 0.7926 | 0.6349 | 0.8892 | 0.7782 | 0.7991 | 600 171 | 345 1372 | 00:16:46 |
| Vgg16 | 0.9403 | 0.8473 | 0.8709 | 0.8328 | 0.7613 | 0.9133 | 823 258 | 122 1285 | 01:35:47 |
| Vgg19 | 0.9498 | 0.8830 | 0.8265 | 0.9177 | 0.8601 | 0.8962 | 781 127 | 164 1416 | 01:58:31 |

Na rys. 5.11 przedstawiono wykresy graficzne wartości sześciu najważniejszych wskaźników jakości AUC, ACC, TPR, TNR oraz PPV, NPV. Przedstawione wykresy uwypuklają poziom zbalansowania wartości poszczególnych wskaźników. W praktyce

medycznej ważne jest zwłaszcza zbalansowanie czułości i specyficzności (TPR i TNR) oraz precyzji rozpoznania klasy pozytywnej i negatywnej (PPV i NPV). Pod względem zbalansowania TPR i TNR najlepszą okazała się sieć Shufflenet (odpowiednio 91.32% i 92.16%). Pod względem precyzji najlepsze zbalansowanie uzyskano dla sieci Inception-resnetv2 (odpowiednio PPV = 93.84%, NPV = 93.29%).



Rys. 5.11 – Wykresy wskaźników jakości odpowiadające różnym architekturom sieci CNN: a) AUC, ACC, b) TPR, TNR, c) PPV, NPV.

5.2.2. Wyniki klasyfikacji zespołów złożonych z sieci CNN – zbiór ISIC

W dalszej części pracy przebadano możliwości polepszenia wyników rozpoznania czerniaka przy zastosowaniu zespołu klasyfikatorów CNN. W wyniku wstępnych eksperymentów stwierdzono zdecydowanie lepsze wyniki uzyskane przy zastosowaniu zespołu zbudowanego wyłącznie z sieci CNN. Stąd prezentacja dalszych wyników będzie ograniczona do sieci CNN z klasyfikatorem typu softmax i wykorzystania techniki transfer learning z douczaniem określonej liczby warstw końcowych poszczególnych struktur sieci.

Ważnym elementem w tej części badawczej jest ustalenie optymalnego składu zespołu. Wzięto pod uwagę wyniki działania poszczególnych sieci CNN, preferując te o najlepszych osiągnięciach. Czas uczenia nie odgrywa istotnej roli w tworzeniu zespołu, gdyż poszczególne jednostki tworzące zespół są trenowane oddzielnie i w działaniu zespołu wykorzystuje się już wytrenowane struktury. Czas testowania gotowych sieci jest w praktyce niezależny od czasu uczenia i stanowi ułamek sekundy dla pojedynczego obrazu. W doborze najlepszych członków zespołu wzięto pod uwagę różne wskaźniki jakości, w tym ACC, AUC oraz TPR.

Ostateczny skład zespołu powstał na bazie wielu eksperymentów numerycznych. Wyniki tych eksperymentów wskazały, że zbyt mały zespół nie pozwala poprawić wyników indywidualnych w dostatecznym stopniu. Z kolei zbyt duży skład zespołu wydłużał czas działania nie prowadząc przy tym do poprawy wyników.

Odrębny problem stanowi sposób fuzji wyników działania poszczególnych członków zespołu. Wobec porównywalnej wartości wskaźników jakości poszczególnych członków zespołu zastosowano metodę głosowania większościowego. Są tu możliwe dwa rozwiązania. W pierwszym sumuje się głosy oddane na poszczególne klasy w postaci binarnej (jedynek głos na daną klasę, zero brak poparcia danej klasy). W drugim sumuje się wartości prawdopodobieństwa wystąpienia każdej klasy (bezpośredni wynik softmax przed binaryzacją). Ten drugi sposób fuzji bardziej precyzyjnie uwzględnia zachowanie się poszczególnych jednostek w procesie klasyfikacji i został zastosowany w eksperymentach numerycznych.

W tworzeniu zespołu istotny okazuje się wybór wskaźnika jakości na podstawie którego dokonuje się wybór składu zespołu. Przebadano szereg różnych kombinacji uwzględniając te sieci dla których określony wskaźnik jakości przybierał największe wartości. W tabeli 5.10 zestawiono jednaście wybranych (najlepszych) kombinacji sieci

tworzących zespoły. Kolumna druga pokazuje wskaźnik jakości na podstawie którego wybierano najlepsze struktury. Kolumny trzecia i czwarta prezentują liczbę członków zespołu oraz numery sieci CNN (stowarzyszone z typem sieci jak przedstawiono na stronie 54) wzięte pod uwagę przy tworzeniu zespołu. Ostatni zespół (jedenasty) składał się ze wszystkich badanych sieci (19). Najlepszy okazał się zespół numer 6 utworzony z 10 najlepszych sieci wybranych na podstawie dokładności.

Tabela 5.10 – Zestawienie kombinacji sieci CNN tworzących różne składy zespołów.

| Nr zespołu | Podstawa wyboru członków zespołu | Liczba członków zespołu | Sieci wybrane do poszczególnych zespołów |
|------------|----------------------------------|-------------------------|------------------------------------------|
| 1 | AUC | 3 | [3,8,13] |
| | ACC | 3 | [3,8,13] |
| | TPR | 3 | [3,8,13] |
| 2 | AUC | 5 | [3,8,13,10,7] |
| 3 | ACC | 5 | [3,8,13,4,10] |
| 4 | TPR | 5 | [3,8,13,4,11] |
| 5 | AUC | 10 | [3,7,8,10,13,4,9,12,15,5] |
| 6 | ACC | 10 | [3,8,13,4,10,7,5,15,9,11] |
| 7 | TPR | 10 | [3,8,13,4,11,14,7,6,10,12] |
| 8 | AUC | 15 | [3,7,8,10,13,4,9,12,15,5,11,14,6,16,19] |
| 9 | ACC | 15 | [3,8,13,4,10,7,5,15,9,11,12,14,6,16,19] |
| 10 | TPR | 15 | [3,8,13,4,11,14,7,6,10,12,15,5,9,18,2] |
| 11 | Wszystkie sieci | 19 | Wszystkie sieci |

Wyniki szczegółowe działania poszczególnych 11 zespołów obejmujące wartości wskaźników AUC, ACC, TPR, TNR, PPV, NPV oraz macierz pomyłek przedstawione są w tabeli 5.11. Największą dokładnością wykazał się zespół numer 6 tworzony na podstawie wskaźnika dokładnościowego zawierający następujące sieci: Inceptionv3, Resnet101, Nasnetlarge, Densenet201, Inceptionresnetv2, Resnet50, Mobilnetv2, Darknet53, Xception, Shufflenet. Uzyskana przez ten zespół wartość dokładności ACC = 96.54% przewyższa najlepszy indywidualny wynik uzyskany przez sieć Resnet101 (ACC = 95.78%). Wynik średni sieci tworzących ten zespół wynosił $ACC_{sr} = 93.25\% \pm 1.36\%$. Wszystkie rozwiązania zespołowe wykazały się bardzo dobrymi wskaźnikami jakości (powyżej wartości 93%). Najmniej efektywny okazał się zespół utworzony ze wszystkich 19 analizowanych struktur sieciowych, choć również on uzyskał dokładność ACC = 95.46% przewyższając zdecydowanie wartość średnią dokładności $ACC_{sr} = 90.08\% \pm 4.67\%$ wszystkich sieci tworzących ten zespół. Na podkreślenie zasługuje znacząca poprawa wykrywalności czerniaka uzyskana dzięki zastosowaniu zespołu. W przypadku najlepszego klasyfikatora indywidualnego (Resnet101) zostało

popelnionych 72 błędów drugiego rodzaju (nierozpoznanie czerniaka). Liczba błędnych rozpoznań czerniaka w przypadku zespołu numer 6 została zredukowana do 50.

Tabela 5.11 – Wartości wskaźników jakości dla zbioru ISIC uzyskane przez zespoły złożone z wybranych sieci CNN (11 kombinacji sieciowych wskazanych w tabeli 5.10).

| Zespół | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|----------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|-------------|
| 1 | 0.9811 | 0.9582 | 0.9312 | 0.9747 | 0.9576 | 0.9586 | 880 | 65 |
| | | | | | | | 39 | 1504 |
| 2 | 0.9866 | 0.9634 | 0.9376 | 0.9793 | 0.9651 | 0.9624 | 886 | 59 |
| | | | | | | | 32 | 1511 |
| 3 | 0.9866 | 0.9626 | 0.9354 | 0.9793 | 0.9651 | 0.9612 | 884 | 61 |
| | | | | | | | 32 | 1511 |
| 4 | 0.9867 | 0.9602 | 0.9397 | 0.9728 | 0.9548 | 0.9634 | 888 | 57 |
| | | | | | | | 42 | 1501 |
| 5 | 0.9926 | 0.9626 | 0.9429 | 0.9747 | 0.9581 | 0.9653 | 891 | 54 |
| | | | | | | | 39 | 1504 |
| 6 | 0.9906 | 0.9654 | 0.9471 | 0.9767 | 0.9613 | 0.9679 | 895 | 50 |
| | | | | | | | 36 | 1507 |
| 7 | 0.9915 | 0.9590 | 0.9460 | 0.9669 | 0.9460 | 0.9669 | 894 | 51 |
| | | | | | | | 51 | 1492 |
| 8 | 0.9924 | 0.9598 | 0.9323 | 0.9767 | 0.9607 | 0.9593 | 881 | 64 |
| | | | | | | | 36 | 1507 |
| 9 | 0.9924 | 0.9598 | 0.9323 | 0.9767 | 0.9607 | 0.9593 | 881 | 64 |
| | | | | | | | 36 | 1507 |
| 10 | 0.9922 | 0.9598 | 0.9386 | 0.9728 | 0.9548 | 0.9628 | 887 | 58 |
| | | | | | | | 42 | 1501 |
| 11 | 0.9908 | 0.9546 | 0.9249 | 0.9728 | 0.9541 | 0.9548 | 874 | 71 |
| | | | | | | | 42 | 1501 |

6. Wyniki badań eksperymentalnych w rozpoznawaniu raka piersi

Rak piersi jest najczęściej występującym nowotworem złośliwym u kobiet i stanowi około 20% wszystkich zachorowań na nowotwory złośliwe u Polek. Wczesne wykrycie zmian możliwe jest przy użyciu mammografii, uznawanej za jedną ze skutecznych metod obrazowych stosowanych w monitorowaniu choroby nowotworowej. Badanie mammograficzne piersi stanowi obecnie podstawową metodę stosowaną we wczesnej diagnostyce raka piersi. Umożliwia dobrą ocenę mikrozwapnień, które w badaniu rezonansem magnetycznym lub USG mogą pozostać niezauważone. Według powszechnej opinii lekarzy onkologów mammografia klasyczna cechuje się czułością 75-95%, przy 80-95% specyficzności i 99% ujemnej wartości prognostycznej.

Badania przesiewowe polegają na poszukiwaniu choroby w populacji bez objawów chorobowych w celu wykrycia schorzenia na bardzo wczesnym etapie rozwoju. Dzięki temu uzyskuje się redukcję umieralności z powodu nowotworu, jak również poprawę wyników leczenia oraz obniżenie kosztów terapii. Biorąc pod uwagę skalę badań przesiewowych istotnym czynnikiem przyspieszającym podjęcie właściwej diagnozy jest wspomaganie decyzji lekarskiej poprzez system komputerowy ukierunkowany na analizę obrazów mammograficznych.

W tym rozdziale zostanie pokazane, że metody stosowane w analizie obrazów zmian skórnych mogą być z powodzeniem zastosowane również do wykrycia zmian nowotworowych piersi na podstawie obrazów mammograficznych o zupełnie innym charakterze niż obrazy zmian skórnych. Badania zostaną przeprowadzone na ogólnodostępnej bazie danych mammograficznych DDSM [26], przedstawionej w rozdziale drugim. Zawiera ona obrazy mammograficzne reprezentujące nowotwór złośliwy (1115), nowotwór łagodny (888) oraz stan normalny (9215) w formie ROI. Z powyższej analizy populacji widoczne jest duże niezrównoważenie klas reprezentujących różne stany pacjentek. Stanowi to dodatkowy problem w zbudowaniu skutecznego narzędzia rozpoznania obrazów. Najważniejszym problemem w analizie jest duże podobieństwo obrazów reprezentujących różne stany pacjentek. Jak pokazano w rozdziale drugim obrazy reprezentujące różne klasy są bardzo zbliżone zarówno pod względem ogólnego wyglądu jak i pod względem wartości parametrów statystycznych charakteryzujących obrazy.

6.1. Organizacja badań eksperymentalnych

W badaniach eksperymentalnych przeprowadzone zostaną próby analizy i rozwiązania trzech zadań: dwa zadania 2-klasowe i jedno trójklasowe.

1. Rozpoznanie raka złośliwego od łagodnego z pominięciem stanu normalnego.
2. Rozpoznanie zmian nowotworowych, w których rak złośliwy i łagodny stanowią jedną klasę a stan normalny drugą klasę.
3. Rozpoznanie trzech klas: klasa 1 – rak złośliwy, klasa 2 – rak łagodny, klasa 3 – stan normalny.

W przypadku drugim i trzecim występuje zdecydowane niezrównoważenie liczebności poszczególnych klas, gdyż zbiór przypadków normalnych zawiera 9215 obrazów a przypadki chorobowe to 1115 - nowotwór złośliwy i 888 - nowotwór łagodny. Może to stanowić przeszkodę w uzyskaniu dobrych wyników rozpoznania. Niezależnie od zadania, ostateczny wynik rozpoznania danych testujących nie uczestniczących w uczeniu generuje się przy zastosowaniu głosowania większościowego wszystkich wykształconych klasyfikatorów. Zastosowano zmodyfikowaną metodę głosowania bazującą na prawdopodobieństwie przynależności klasowej (przed procesem binaryzacji).

Na podstawie doświadczeń z klasyfikacji obrazów zmian skórnych opisanych w poprzednim rozdziale, badania nad tym zbiorem skoncentrowały się na zastosowaniu wyłącznie sieci klasyfikatorów CNN w zespole. Podobnie jak przy analizie zbioru ISIC, dla stworzenia optymalnego składu zespołu klasyfikacyjnego w pierwszej kolejności przebadano indywidualną skuteczność działania różnych rozwiązań strukturalnych tych sieci, a następnie na podstawie tych wyników został dobrany odpowiedni skład zespołu. Wszystkie badania zostały przeprowadzone przy zastosowaniu 5-krotnej walidacji krzyżowej. W prezentacji wyników eksperymentów ograniczono się wyłącznie do rezultatów testowania na danych nie uczestniczących w uczeniu, prezentując wartości dotyczące dokładności, precyzji klasowej, czułości i pola AUC pod krzywą ROC.

6.2. Rozpoznanie raka złośliwego od łagodnego

Podobnie jak w przypadku czerniaka w pierwszym kroku zostały zbadane różne struktury sieci CNN w rozpoznaniu raka złośliwego (klasa 1) od postaci łagodnej (klasa 2). Baza danych dotycząca raka łagodnego obejmowała 888 obrazów ROI. W przypadku raka złośliwego zbiór obejmuje 1115 obrazów ROI. Obie klasy są więc reprezentowane przez zbalansowane liczebnie zbiory danych.

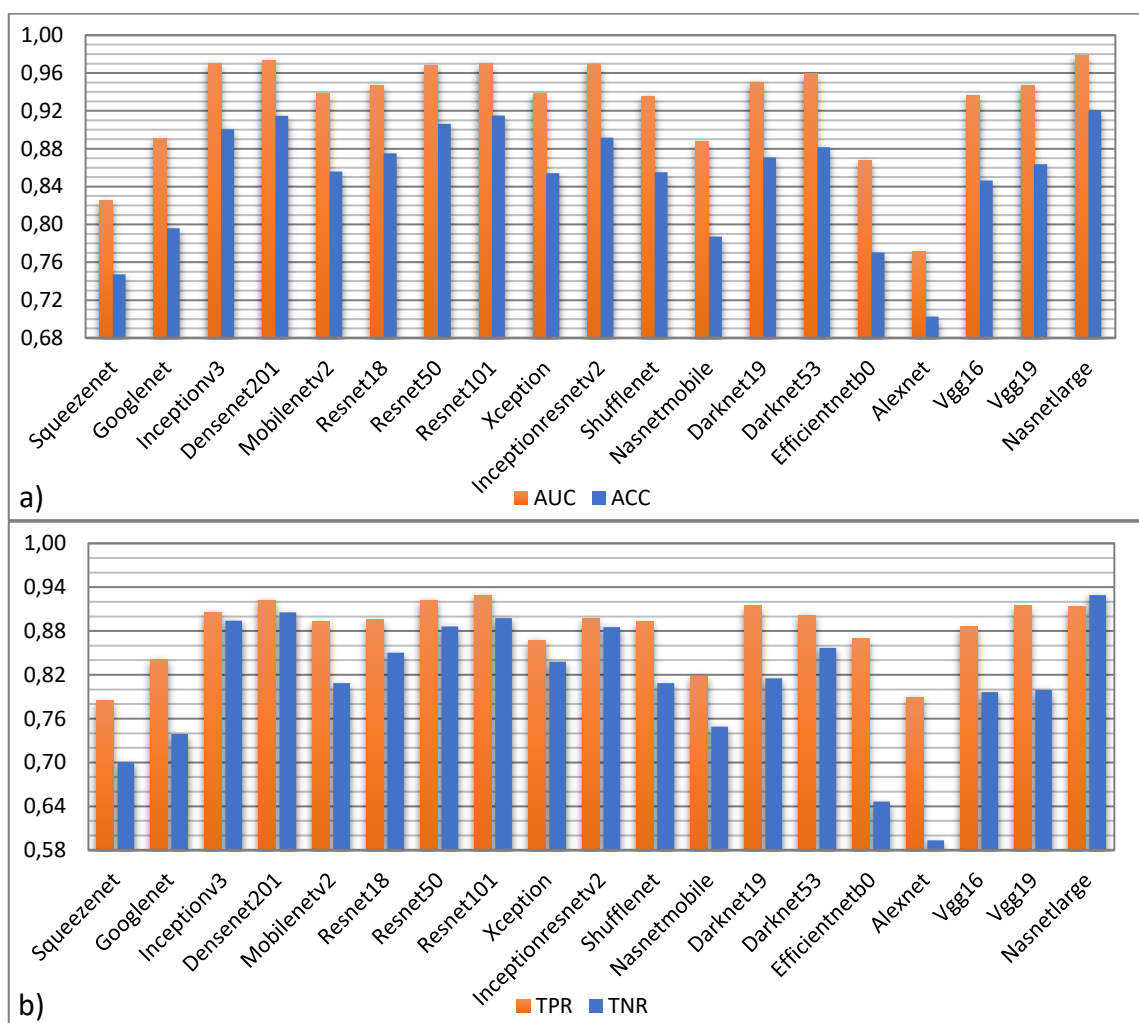
Wykorzystując strategię transfer learningu wstępnie wytrenowane sieci, dostępne w Matlabie [39], zostały douczone na właściwych danych obu klas stosując 5-krotną walidację krzyżową. W procesie adaptacji sieci zastosowano indywidualny dobór liczby warstw podlegających douczaniu. Zamrożenie 30% początkowych warstw zastosowano dla sieci: Squeezenet, Inceptionv3, Densenet201, Mobilenetv2, Resnet101, Shufflenet, Darknet19, Darknet53, Efficientnetb0 i Vgg19. Dla pozostałych dziewięciu sieci proces douczania obejmował wszystkie warstwy w strukturze CNN.

Wyniki statystyczne rozpoznania raka złośliwego od postaci łagodnej dla każdej badanej sieci przedstawione są w tabeli 6.1. Poniższe wartości podane są jedynie dla danych testowych nie uczestniczących w uczeniu.

Tabela 6.1 – Wyniki statystyczne rozpoznania raka złośliwego od postaci łagodnej dla poszczególnych architektur sieci CNN.

| Sieć CNN | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | | Czas [h] |
|--------------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|------------|---------------|
| Squeezenet | 0.8254 | 0.7474 | 0.7848 | 0.7005 | 0.7669 | 0.7216 | 875 | 240 | 0.34 |
| | | | | | | | 266 | 622 | |
| Googlenet | 0.8911 | 0.7958 | 0.8413 | 0.7387 | 0.8017 | 0.7875 | 938 | 177 | 0.67 |
| | | | | | | | 232 | 656 | |
| Inceptionv3 | 0.9696 | 0.9006 | 0.9058 | 0.8941 | 0.9149 | 0.8832 | 1010 | 105 | 4.11 |
| | | | | | | | 94 | 794 | |
| Densenet201 | 0.9735 | 0.9146 | 0.9220 | 0.9054 | 0.9245 | 0.9024 | 1028 | 87 | 7.86 |
| | | | | | | | 84 | 804 | |
| Mobilenetv2 | 0.9381 | 0.8557 | 0.8933 | 0.8086 | 0.8542 | 0.8578 | 996 | 119 | 1.62 |
| | | | | | | | 170 | 718 | |
| Resnet18 | 0.9473 | 0.8752 | 0.8951 | 0.8502 | 0.8824 | 0.8658 | 998 | 117 | 0.57 |
| | | | | | | | 133 | 755 | |
| Resnet50 | 0.9682 | 0.9061 | 0.9220 | 0.8863 | 0.9105 | 0.9005 | 1028 | 87 | 2.14 |
| | | | | | | | 101 | 787 | |
| Resnet101 | 0.9695 | 0.9151 | 0.9291 | 0.8975 | 0.9193 | 0.9098 | 1036 | 79 | 5.05 |
| | | | | | | | 91 | 797 | |
| Xception | 0.9380 | 0.8542 | 0.8673 | 0.8378 | 0.8704 | 0.8341 | 967 | 148 | 3.23 |
| | | | | | | | 144 | 744 | |
| Inceptionresnetv2 | 0.9688 | 0.8917 | 0.8969 | 0.8851 | 0.9074 | 0.8724 | 1000 | 115 | 8.17 |
| | | | | | | | 102 | 786 | |
| Shufflenet | 0.9356 | 0.8552 | 0.8924 | 0.8086 | 0.8541 | 0.8568 | 995 | 120 | 0.99 |
| | | | | | | | 170 | 718 | |
| Nasnetmobile | 0.8874 | 0.7873 | 0.8179 | 0.7489 | 0.8035 | 0.7661 | 912 | 203 | 8.08 |
| | | | | | | | 223 | 665 | |
| Nasnetlarge | 0.9782 | 0.9201 | 0.9130 | 0.9291 | 0.9417 | 0.8948 | 1018 | 97 | 223.02 |
| | | | | | | | 63 | 825 | |
| Darknet19 | 0.9504 | 0.8707 | 0.9148 | 0.8153 | 0.8615 | 0.8840 | 1020 | 95 | 0.81 |
| | | | | | | | 164 | 724 | |

| Sieć CNN | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | | Czas [h] |
|----------------|--------|--------|--------|--------|--------|--------|--------------|-----|----------|
| Darknet53 | 0.9591 | 0.8817 | 0.9013 | 0.8570 | 0.8878 | 0.8737 | 1005 | 110 | 2.64 |
| | | | | | | | 127 | 761 | |
| Efficientnetb0 | 0.8677 | 0.7703 | 0.8691 | 0.6464 | 0.7553 | 0.7972 | 969 | 146 | 4.09 |
| | | | | | | | 314 | 574 | |
| Alexnet | 0.7714 | 0.7024 | 0.7892 | 0.5935 | 0.7091 | 0.6916 | 880 | 235 | 0.40 |
| | | | | | | | 361 | 527 | |
| Vgg16 | 0.9361 | 0.8462 | 0.8861 | 0.7962 | 0.8452 | 0.8477 | 988 | 127 | 2.13 |
| | | | | | | | 181 | 707 | |
| Vgg19 | 0.9470 | 0.8637 | 0.9148 | 0.7995 | 0.8514 | 0.8820 | 1020 | 95 | 2.70 |
| | | | | | | | 178 | 710 | |



Rys. 6.1 – Wykresy przedstawiające wartości miar jakości poszczególnych rozwiązań sieci CNN: a) dokładność ACC i miara AUC, b) czułość TPR i specyficzność TNR.

Dla lepszego zobrazowania uzyskanych wyników na rys. 6.1 a i b przedstawiono wykresy miar jakości uzyskanych dla poszczególnych sieci. Najwyższą dokładnością wykazała się sieć Nasnetlarge, Resnet101 oraz Densenet201. Jednocześnie Resnet101 uzyskał najwyższą wartość czułości TPR, co oznacza najmniejszą liczbę przypadków raka złośliwego

zaklasyfikowanych jako rak łagodny (Resnet101 – 79, Nasnetlarge – 97). Należy przy tym zauważyć, że czas uczenia sieci Nasnetlarge był szczególnie wysoki (223 godziny pracy komputera w trybie 5-krotnej walidacji krzyżowej) i przewyższał sumę czasów uczenia wszystkich pozostałych sieci razem wziętych.

Biorąc pod uwagę stosunkowo niewielką przewagę dokładności nad najbliższymi mu sieciami (Resnet101, Densenet201, Resnet50) dla których czas uczenia był zdecydowanie krótszy i nie przekraczał 8 godzin w dalszych badaniach zrezygnowano z sieci Nasnetlarge. Interesujące jest zestawienie wartości uzyskanych czułości i specyficzności przez różne sieci. W praktyce pożądane jest dobre zrównoważenie obu tych wielkości. Takie oczekiwania spełniają między innymi najlepsze sieci (Nasnetlarge, Densenet201, Inceptionv3, Resnet101, Inceptionresnetv2). W przypadku sieci słabszych (na przykład Alexnet, Efficientnetb0, Squezenet, Googlenet) te różnice są duże, co dyskwalifikuje je z dalszych rozważań jako kandydatów do zespołu.

Analiza wyników poszczególnych sieci pozwala stwierdzić istotne różnice w efektywności rozpoznania klas. Dotyczy to wszystkich miar jakościowych. W tabeli 6.2 przedstawiono zestawienie wyników statystycznych dotyczących analizowanych miar jakości uzyskanych przez 19 badanych struktur sieciowych CNN. Widoczne są bardzo duże różnice między najlepszymi sieciami i najgorszymi (na przykład ACC najgorszej sieci Alexnet jest równa 70.24% a najlepszej Nasnetlarge 92.01%).

Tabela 6.2 – Wartości statystyczne analizowanych miar jakości klasyfikatorów.

| Miara jakości | Średnia | Odchylenie standardowe | Wartość minimalna | Wartość maksymalna |
|---------------|---------|------------------------|-------------------|--------------------|
| AUC | 0.9275 | 0.0555 | 0.7714 | 0.9782 |
| ACC | 0.8502 | 0.0618 | 0.7024 | 0.9201 |
| TPR | 0.8819 | 0.0437 | 0.7848 | 0.9291 |
| TNR | 0.8105 | 0.0909 | 0.5935 | 0.9291 |
| PPV | 0.8559 | 0.0635 | 0.7091 | 0.9417 |
| NPV | 0.8436 | 0.0625 | 0.6916 | 0.9098 |

Na podstawie wyników indywidualnych przeprowadzone zostały próby zbudowania najbardziej efektywnego zespołu sieci CNN zintegrowanego przy zastosowaniu głosowania większościowego opartego na wartościach prawdopodobieństwa klasowego (wynik softmax przed binaryzacją). W skład zespołu proponowane są sieci najlepsze pod względem trzech najważniejszych praktycznie miar jakości (ACC, AUC, TPR). Eksperymenty przeprowadzono z uwzględnieniem sieci Nasnetlarge i po jej eliminacji ze względu na bardzo długi czas uczenia.

Tabela 6.3 – Analizowane składy kombinacji zespołów tworzonych przez różne struktury sieci CNN z uwzględnieniem struktury Nasnetlarge. Poszczególne sieci oznaczone są w postaci numerycznej.

| Nr zespołu | Wybór członków na podstawie | Liczba członków zespołu | Sieci wybrane do poszczególnych zespołów |
|------------|-----------------------------|-------------------------|------------------------------------------|
| 1 | AUC | 3 | [3,4,13] |
| 2 | ACC | 3 | [4,8,13] |
| 3 | TPR | 3 | [4,7,8] |
| 4 | AUC | 5 | [3,4,8,10,13] |
| 5 | ACC | 5 | [3,4,7,8,13] |
| 6 | TPR | 5 | [4,7,8,14,19] |
| 7 | AUC | 7 | [3,4,7,8,10,13,15] |
| | ACC | | |
| 8 | TPR | 7 | [3,4,7,8,13,14,19] |
| 9 | AUC | 11 | [3,4,5,6,7,8,10,13,14,15,19] |
| | ACC | 11 | [3,4,5,6,7,8,10,13,14,15,19] |
| | TPR | 11 | [3,4,5,6,7,8,10,13,14,15,19] |
| 10 | AUC | 15 | [2,3,4,5,6,7,8,9,10,11,13,14,15,18,19] |
| | ACC | | |
| 11 | TPR | 15 | [3,4,5,6,7,8,9,10,11,13,14,15,16,18,19] |
| 12 | - | 19 | wszystkie |

W tabeli 6.3 przedstawiono składy badanych zespołów uwzględniających sieć Nasnetlarge. Zespoły te w wyniku wielu eksperymentów uznano za najbardziej obiecujące. Poszczególne struktury sieci CNN zostały oznaczone numerami (identycznie jak w rozdziale piątym). W kolumnie drugiej przedstawiono miary jakości (ACC, AUC, TPR) względem których dobierano w zespole najlepsze składy sieci CNN. W przypadku zespołu numer 7 i 10 wybór składów zespołu sieci był identyczny niezależnie od użytej miary AUC lub ACC.

Cechą charakterystyczną wyników działania zespołu (tabela 6.4) jest zbliżony poziom wartości poszczególnych miar jakości dla różnych rozwiązań zespołu (na przykład wszystkie wartości dokładności ACC powyżej 92%). Tym nie mniej najlepsze wyniki odpowiadają zespołowi numer 7 zawierającemu siedem następujących sieci: Inceptionv3, Densenet201, Resnet50, Resnet101, Inceptionresnetv2, Nasnetlarge oraz Darknet53, a także zespołowi numer 8 zbudowanemu z siedmiu sieci: Inceptionv3, Densenet201, Resnet50, Resnet101, Nasnetlarge, Darknet19 oraz Vgg19. We wszystkich kategoriach miar jakości zespoły numer 7 i 8 przewyższyły najlepsze indywidualne rozwiązanie odpowiadające najlepszej sieci Nasnetlarge. Na przykład dokładność zespołu siódmego ACC = 93.81% przewyższała wartość ACC = 92.01% najlepszej sieci indywidualnej Nasnetlarge.

Zespół nr 7 przewyższał nieco osiągnięcia zespołu 8 pod względem wartości AUC, ACC, TNR oraz PPV. Z kolei zespół numer 8 charakteryzował się najlepszymi wynikami czułości TPR oraz precyzji NPV w wykrywaniu raka złośliwego. Szczególnie ważne jest podwyższenie czułości w wykrywaniu raka złośliwego (wzrost z wartości 91.3% dla Nasnetlarge do 94.53% w zespole 8).

Tabela 6.4 – Wyniki klasyfikacji różnych składów zespołu przy rozpoznaniu raka złośliwego od postaci łagodnej z uwzględnieniem sieci Nasnetlarge.

| Zespół | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|--------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|-----|
| 1 | 0.9836 | 0.9301 | 0.9300 | 0.9302 | 0.9436 | 0.9137 | 1037 | 78 |
| | | | | | | | 62 | 826 |
| 2 | 0.9838 | 0.9356 | 0.9390 | 0.9313 | 0.9449 | 0.9240 | 1047 | 68 |
| | | | | | | | 61 | 827 |
| 3 | 0.9801 | 0.9331 | 0.9417 | 0.9223 | 0.9383 | 0.9265 | 1050 | 65 |
| | | | | | | | 69 | 819 |
| 4 | 0.9846 | 0.9336 | 0.9363 | 0.9302 | 0.9439 | 0.9208 | 1044 | 71 |
| | | | | | | | 62 | 826 |
| 5 | 0.9846 | 0.9356 | 0.9390 | 0.9313 | 0.9449 | 0.9240 | 1047 | 68 |
| | | | | | | | 61 | 827 |
| 6 | 0.9798 | 0.9311 | 0.9462 | 0.9122 | 0.9312 | 0.9310 | 1055 | 60 |
| | | | | | | | 78 | 810 |
| 7 | 0.9843 | 0.9381 | 0.9435 | 0.9313 | 0.9452 | 0.9292 | 1052 | 63 |
| | | | | | | | 61 | 827 |
| 8 | 0.9833 | 0.9351 | 0.9453 | 0.9223 | 0.9386 | 0.9307 | 1054 | 61 |
| | | | | | | | 69 | 819 |
| 9 | 0.9824 | 0.9356 | 0.9444 | 0.9245 | 0.9402 | 0.9298 | 1053 | 62 |
| | | | | | | | 67 | 821 |
| 10 | 0.9803 | 0.9316 | 0.9390 | 0.9223 | 0.9382 | 0.9233 | 1047 | 68 |
| | | | | | | | 69 | 819 |
| 11 | 0.9807 | 0.9316 | 0.9417 | 0.9189 | 0.9358 | 0.9262 | 1050 | 65 |
| | | | | | | | 72 | 816 |
| 12 | 0.9771 | 0.9281 | 0.9399 | 0.9133 | 0.9316 | 0.9237 | 1048 | 67 |
| | | | | | | | 77 | 811 |

Ze względu na zbyt długi czas uczenia, w następnym eksperymencie pominięto najlepszą sieć Nasnetlarge dla sprawdzenia jak jej eliminacja wpłynie na najlepszy wynik zespołu. Tym razem badaniu poddano składy zespołów przedstawione w tabeli 6.5, wyłonione jako najlepsze we wstępnej fazie eksperymentów.

Tabela 6.5 – Składy kombinacji zespołów tworzonych przez struktury sieci CNN z pominięciem sieci Nasnetlarge.

| Nr zespołu | Wybór członków na podstawie miary jakości | Liczba członków zespołu | Sieci wybrane do poszczególnych zespołów |
|------------|-------------------------------------------|-------------------------|------------------------------------------|
| 1 | AUC | 6 | [3,4,7,8,10,15] |
| | ACC | | |
| 2 | TPR | 6 | [3,4,7,8,14,19] |
| 3 | AUC | 7 | [3,4,7,8,10,14,15] |
| 4 | ACC | 7 | [3,4,6,7,8,10,15] |
| 5 | TPR | 7 | [3,4,7,8,14,15,19] |
| 6 | AUC | 9 | [3,4,6,7,8,10,14,15,19] |
| | ACC | | |
| | TPR | | |
| 7 | AUC | 10 | [3,4,5,6,7,8,10,14,15,19] |
| | ACC | | |
| | TPR | | |
| 8 | AUC | 11 | [3,4,5,6,7,8,9,10,14,15,19] |
| 9 | ACC | 11 | [3,4,5,6,7,8,10,11,14,15,19] |
| | TPR | | |

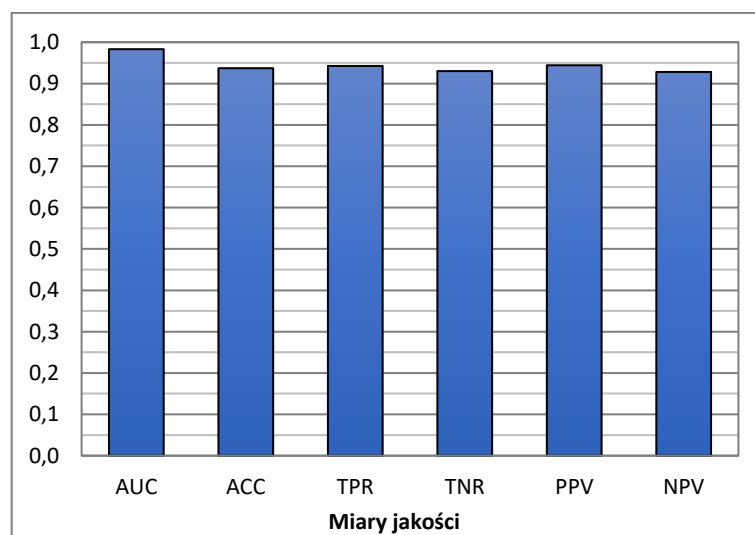
Podobnie jak w przypadku poprzednim wyniki różnych składów zespołów okazały się również bardzo zbliżone do siebie pod względem wartości poszczególnych miar jakości (Tabela 6.6). Najlepsze wyniki uzyskano dla zespołów numer 1 (sześciu członków: Inceptionv3, Densenet201, Resnet50, Resnet101, Inceptionresnetv2, oraz Darknet53) oraz numer 5 (siedmiu członków: Inceptionv3, Densenet201, Resnet50, Resnet101, Darknet19, Darknet53 oraz VGG19). Wyniki uzyskane po odrzuceniu sieci Nasnetlarge tylko minimalnie są gorsze od zespołów zbudowanych z użyciem Nasnetlarge (różnice dopiero na trzeciej pozycji liczbowej). Na przykład średnia dokładność zespołu sześciu sieci to 93.71% w porównaniu z wartością 93.81% w zespole z udziałem Nasnetlarge. Jeszcze mniejsze różnice występują w przypadku miary AUC (0.9836 zespołu siedmiu sieci CNN w porównaniu z wartością 0.9831 zespołu sześciu sieci (bez Nasnetlarge)).

Tabela 6.6 – Wyniki klasyfikacji zespołów sieci CNN zbudowanych bez najlepszej sieci Nasnetlarge przy rozpoznaniu raka złośliwego od postaci łagodnej.

| Zespół | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|--------|--------|--------|--------|--------|--------|--------|--------------|-----|
| 1 | 0.9831 | 0.9371 | 0.9426 | 0.9302 | 0.9443 | 0.9281 | 1051 | 64 |
| | | | | | | | 62 | 826 |
| 2 | 0.9814 | 0.9331 | 0.9462 | 0.9167 | 0.9345 | 0.9314 | 1055 | 60 |
| | | | | | | | 74 | 814 |
| 3 | 0.9824 | 0.9366 | 0.9453 | 0.9257 | 0.9411 | 0.9309 | 1054 | 61 |
| | | | | | | | 66 | 822 |
| 4 | 0.9822 | 0.9331 | 0.9408 | 0.9234 | 0.9391 | 0.9255 | 1049 | 66 |
| | | | | | | | 68 | 820 |

| Zespół | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|--------|--------|--------|--------|--------|--------|--------|--------------|-----|
| 5 | 0.9817 | 0.9341 | 0.9498 | 0.9144 | 0.9330 | 0.9355 | 1059 | 56 |
| | | | | | | | 76 | 812 |
| 6 | 0.9815 | 0.9341 | 0.9471 | 0.9178 | 0.9353 | 0.9325 | 1056 | 59 |
| | | | | | | | 73 | 815 |
| 7 | 0.9813 | 0.9331 | 0.9453 | 0.9178 | 0.9352 | 0.9304 | 1054 | 61 |
| | | | | | | | 73 | 815 |
| 8 | 0.9811 | 0.9341 | 0.9453 | 0.9200 | 0.9369 | 0.9305 | 1054 | 61 |
| | | | | | | | 71 | 817 |
| 9 | 0.9805 | 0.9336 | 0.9480 | 0.9155 | 0.9337 | 0.9334 | 1057 | 58 |
| | | | | | | | 75 | 813 |

Najlepsze wyniki liczbowe dotyczące wszystkich analizowanych miar jakości są wysokie (powyżej 92%) i bardzo bliskie sobie. Jest to widoczne na wykresie z rys. 5.2 dla zespołu numer 1.



Rys. 6.2 – Wykresy miar jakości w rozpoznaniu raka złośliwego od postaci łagodnej dla zespołu numer 1.

Na uwagę zasługuje również zespół numer 5. Jego wyniki są bardzo bliskie najlepszemu (zespół 1), przy czym charakteryzuje się on największą ujemną wartością prognostyczną (NPV = 0.9355), co oznacza najmniejszą liczbę stanów złośliwych zaklasyfikowanych jako łagodne (56 takich błędów w zespole 5 w stosunku do 64 w zespole 1). Ważną zaletą zespołu jest istotne zmniejszenie liczby błędnych rozpoznań raka złośliwego jako zmian łagodnych. Wyniki najlepszych rozwiązań indywidualnych to: 87 takich przypadków w sieci Densenet201, 79 w sieci Resnet101 i 97 w sieci Nasnetlarge, podczas gdy najlepszy wynik zespołu to 56.

6.3. Rozpoznanie zmian nowotworowych od stanu normalnego

W tym badaniu oba rodzaje zmian nowotworowych tworzą jedną klasę, a przypadki normalne klasę przeciwną. We wstępnej fazie badań przeprowadzono dwa rodzaje eksperymentów: bez równoważenia zbiorów oraz ze zrównoważeniem. W przypadku równoważenia liczebności zbiorów zastosowano podział zbioru normalnego na 5 podzbiorów przeciwstawiając każdemu z nich w procesie uczenia ten sam zbiór reprezentujący zmiany nowotworowe. Wynik działania systemu był ustalony przez głosowanie większościowe. Wstępne wyniki symulacji pokazały, że w obu przypadkach wskaźniki jakościowe rozpoznania klas są na podobnym poziomie statystycznym. Stąd dalsze (właściwe badania) zostały przeprowadzone na pełnych zbiorach (bez procedury równoważenia klas).

W efekcie klasę pierwszą w tym badaniach stanowią złączone przypadki zmian nowotworowych (rak złośliwy i rak łagodny jako jedna klasa), natomiast klasę drugą stan normalny. W wyniku przeprowadzonych badań wyodrębniono sześć najlepszych architektur sieci CNN: Inceptionv3, Densenet201, Resnet50, Resnet101, Inceptionresnetv2 oraz Darknet53 jako przyszłych członków zespołu. W tabeli 6.7 przedstawiono wartości miar jakości na danych testujących (tryb 5-krotnej walidacji krzyżowej) uzyskanych indywidualnie przez poszczególne sieci tworzące w przyszłości zespół. Wyniki jednostek są zbliżone do siebie (na przykład wartości ACC zmieniające się od 98.53% uzyskane przez sieć Inceptionv3 do 95.32% dla sieci Resnet50, wartości AUC zmieniające się w zakresie od 0.9868 dla sieci Densenet201 do 0.9972 dla Inceptionresnetv2).

Tabela 6.7 – Wyniki statystyczne poszczególnych sieci CNN przy rozpoznaniu zmian nowotworowych (rak złośliwy i łagodny tworzące wspólna klasę) od stanu normalnego.

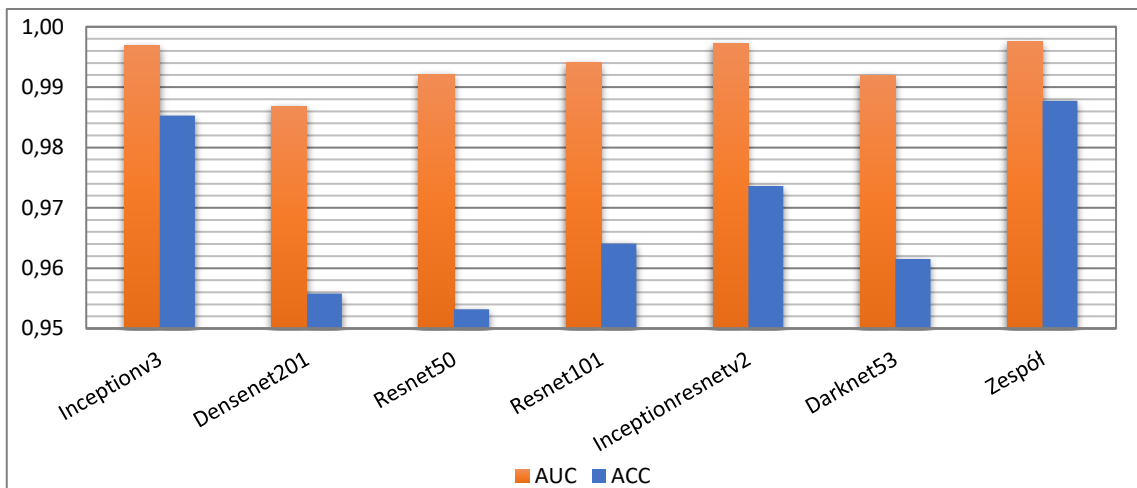
| Sieć CNN | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|-------------------|--------|--------|--------|--------|--------|--------|--------------|------|
| Inceptionv3 | 0.9969 | 0.9853 | 0.9641 | 0.9899 | 0.9541 | 0.9922 | 1931 | 72 |
| | | | | | | | 93 | 9122 |
| Densenet201 | 0.9868 | 0.9558 | 0.9396 | 0.9593 | 0.8339 | 0.9865 | 1882 | 121 |
| | | | | | | | 375 | 8840 |
| Resnet50 | 0.9921 | 0.9532 | 0.9596 | 0.9518 | 0.8123 | 0.9908 | 1922 | 81 |
| | | | | | | | 444 | 8771 |
| Resnet101 | 0.9940 | 0.9641 | 0.9661 | 0.9636 | 0.8524 | 0.9924 | 1935 | 68 |
| | | | | | | | 335 | 8880 |
| Inceptionresnetv2 | 0.9972 | 0.9736 | 0.9785 | 0.9725 | 0.8857 | 0.9952 | 1960 | 43 |
| | | | | | | | 253 | 8962 |
| Darknet53 | 0.9920 | 0.9615 | 0.9451 | 0.9651 | 0.8546 | 0.9878 | 1893 | 110 |
| | | | | | | | 322 | 8893 |

Wyniki działania poszczególnych jednostek zespołu zostały poddane głosowaniu większościowemu na podstawie wartości prawdopodobieństwa klasowego. Uzyskane wartości wskaźników jakościowych przedstawione są w tabeli 6.8. Wszystkie wskaźniki jakościowe zespołu przewyższają najlepsze osiągnięcia indywidualnych członków.

Tabela 6.8 – Wyniki statystyczne zespołu sześciu sieci CNN przy rozpoznaniu zmian nowotworowych (rak złośliwy i łagodny tworzące wspólna klasę) od stanu normalnego.

| AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | |
|--------|--------|--------|--------|--------|--------|--------------|------|
| 0.9975 | 0.9877 | 0.9780 | 0.9907 | 0.9578 | 0.9943 | 1951 | 52 |
| | | | | | | 86 | 9129 |

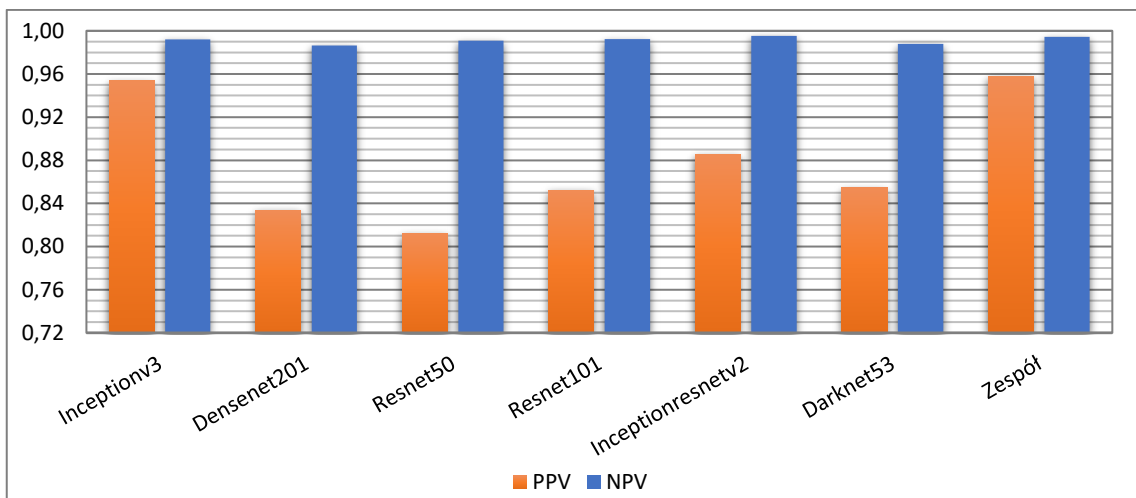
Graficzne porównanie wartości ACC i AUC indywidualnych członków z wynikiem zintegrowanym zespołu zostało zilustrowane na rys 6.3. W obu przypadkach wynik zespołu przewyższa osiągnięcia jego członków, przy czym przewaga nad najlepszym jego członkiem (Inceptionv3) jest stosunkowo niewielka.



Rys. 6.3 – Ilustracja rozkładu wartości ACC i AUC poszczególnych sieci CNN w porównaniu z wynikami zintegrowanego zespołu tych sieci.

Na rysunku 6.4 przedstawiono porównanie wartości precyzji PPV oraz NPV rozpoznania przypadków chorobowych od normalnych uzyskanych przez pojedynczo działające sieci CNN i wyniki zespołu tych samych sieci po ich integracji. Przewaga zespołu w tym przypadku uwidacznia się przede wszystkim dla wskaźników PPV i NPV, które są szczególnie istotne dla niezrównoważonych zbiorów. Widoczna jest stosunkowo duża rozpiętość w przypadku wartości PPV indywidualnych rezultatów sieci (najlepszy wynik to 95.41% dla sieci Inceptionv3, najgorszy 81.23% dla Resnet50). Wskaźnik precyzji rozpoznania klasy negatywnej osiągnął bardzo wysoki poziom dla wszystkich badanych przypadków. Zastosowanie zespołu spowodowało wzrost wartości wskaźnika PPV do

95.78% oraz obniżenie NPV o 0,09% w stosunku do najlepszego rezultatu (sieć Inceptionresnetv2). Analizując wartości liczbowe pomyłek indywidualnych rezultatów sieci należy wyróżnić sieć Inceptionv3, która popełniła łącznie 165 błędów ($93_I + 72_{II}$) co stanowiło znaczną przewagę nad pozostałymi klasyfikatorami. Na wyróżnienie zasługuje również sieć Inceptionresnetv2, która osiągnęła najniższy poziom błędów drugiego rodzaju (43), niestety kosztem niskiej wartości precyzji PPV (253 błędy pierwszego rodzaju). Zastosowanie zespołu zbudowanego z mniej efektywnych jednostek zmniejszyło liczbę błędów do 138 ($86_I + 52_{II}$).



Rys. 6.4 – Porównanie wartości precyzji klasowej PPV i NPV dla poszczególnych sieci CNN z wynikami zintegrowanego zespołu tych sieci.

6.4. Klasyfikacja trójklasowa

W tym eksperymencie rozpoznaniu podlegają trzy klasy:

- klasa 1 – rak złośliwy
- klasa 2 – rak łagodny
- klasa 3 – przypadki normalne.

Na złożoność tego zadania wpływa, podobnie jak w zadaniu poprzednim, fakt znacznego niezrównoważenia danych reprezentujących poszczególne klasy. Wstępnie przeprowadzone badania wskazały, że to niezrównoważenie nie ma jednak istotnego wpływu na wyniki statystyczne. Stąd główne badania przeprowadzono na pełnych zbiorach danych poszczególnych klas, stosując 5-krotną walidację krzyżową.

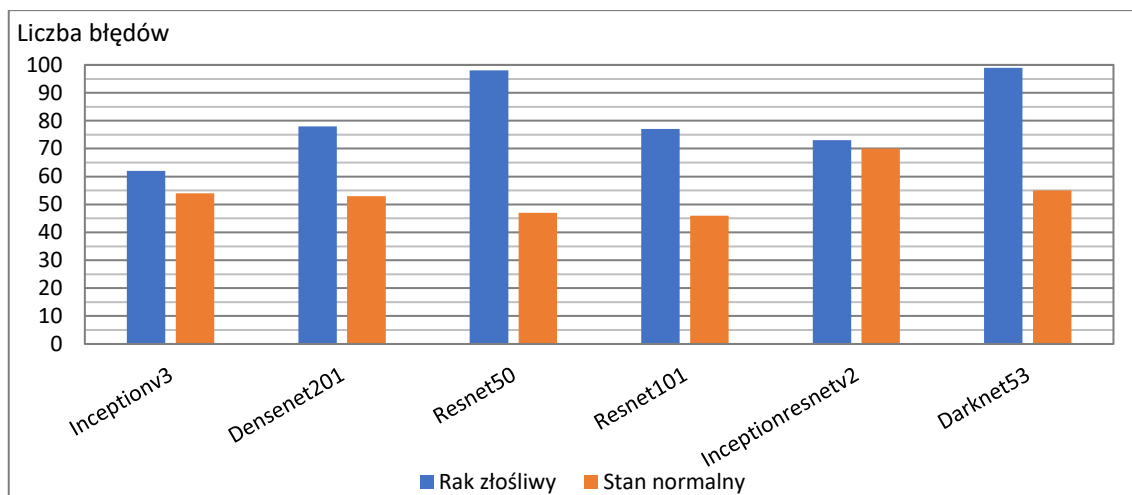
Wobec wystąpienia trzech klas wyniki dotyczące miar jakościowych: czułość klasowa rozpoznawanej klasy (TPR), specyficzność jako parametr odpowiadający pozostałym złączonym 2 klasom (TNR), precyzja rozpoznania danej klasy (PPV) oraz precyzja

klasy większościowej złożonej z dwu pozostałych klas (NPV) będą zaprezentowane dla każdej klasy oddzielnie. W tabeli 6.9 przedstawiono statystyczne wyniki rozpoznania trzech klas na podstawie mammogramów uzyskane przez wybrane struktury klasyfikatorów CNN. Sześć struktur sieciowych poddanych badaniu zostało wyselekcjonowanych we wstępnej fazie eksperymentów jako najlepsze w rozwiązaniu powyższego zadania 3-klasowego. Wyniki uzyskane przez poszczególne sieci reprezentują kolejne analizowane klasy: 1 – rak złośliwy, 2 – rak o postaci łagodnej, 3 – przypadki normalne. Wartości w kolumnie ACC dotyczą średniej dokładności rozpoznania wszystkich klas łącznie.

Tabela 6.9 – Wartości wskaźników jakości przy rozpoznaniu trzech klas mamogramów.

| Sieć CNN | Kl. | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | | | Liczba pomyłek |
|-------------------|-----|---------------|---------------|---------------|---------------|---------------|---------------|--------------|-----|------|----------------|
| Inceptionv3 | 1 | 0.9938 | 0.9717 | 0.9022 | 0.9897 | 0.9063 | 0.9892 | 1006 | 61 | 48 | 109 |
| | 2 | 0.9899 | | 0.8694 | 0.9892 | 0.8733 | 0.9888 | 62 | 772 | 54 | 116 |
| | 3 | 0.9958 | | 0.9899 | 0.9491 | 0.9889 | 0.9534 | 42 | 51 | 9122 | 93 |
| Densenet201 | 1 | 0.9909 | 0.9671 | 0.9184 | 0.9838 | 0.8620 | 0.9909 | 1024 | 51 | 40 | 91 |
| | 2 | 0.9852 | | 0.8525 | 0.9892 | 0.8711 | 0.9873 | 78 | 757 | 53 | 131 |
| | 3 | 0.9935 | | 0.9840 | 0.9536 | 0.9898 | 0.9285 | 86 | 61 | 9068 | 147 |
| Resnet50 | 1 | 0.9880 | 0.9500 | 0.9130 | 0.9751 | 0.8016 | 0.9902 | 1018 | 77 | 20 | 97 |
| | 2 | 0.9810 | | 0.8367 | 0.9766 | 0.7543 | 0.9858 | 98 | 743 | 47 | 145 |
| | 3 | 0.9922 | | 0.9654 | 0.9666 | 0.9925 | 0.8585 | 154 | 165 | 8896 | 319 |
| Resnet101 | 1 | 0.9893 | 0.9531 | 0.9157 | 0.9758 | 0.8071 | 0.9906 | 1021 | 64 | 30 | 94 |
| | 2 | 0.9829 | | 0.8615 | 0.9801 | 0.7878 | 0.9880 | 77 | 765 | 46 | 123 |
| | 3 | 0.9925 | | 0.9665 | 0.9621 | 0.9915 | 0.8618 | 167 | 142 | 8906 | 309 |
| Inceptionresnetv2 | 1 | 0.9933 | 0.9685 | 0.8619 | 0.9892 | 0.8981 | 0.9848 | 961 | 86 | 68 | 154 |
| | 2 | 0.9905 | | 0.8390 | 0.9897 | 0.8754 | 0.9862 | 73 | 745 | 70 | 143 |
| | 3 | 0.9967 | | 0.9939 | 0.9311 | 0.9852 | 0.9708 | 36 | 20 | 9159 | 56 |
| Darknet53 | 1 | 0.9889 | 0.9561 | 0.9103 | 0.9756 | 0.8043 | 0.9900 | 1015 | 64 | 36 | 100 |
| | 2 | 0.9842 | | 0.8266 | 0.9851 | 0.8266 | 0.9851 | 99 | 734 | 55 | 154 |
| | 3 | 0.9926 | | 0.9742 | 0.9546 | 0.9900 | 0.8893 | 148 | 90 | 8977 | 238 |

Uzyskane wyniki numeryczne wskazują, że najtrudniejszym zadaniem jest rozpoznanie przypadków zmian łagodnych od pozostałych (rak złośliwy i przypadki normalne), przy czym częściej dochodziło do błędu jako rozpoznanie ich jako raka złośliwego. Jest to widoczne na rys. 6.5, na którym zobrazowano liczbę błędnych rozpoznań zmian łagodnych jako raka złośliwego bądź stanu normalnego. Najsłabsze wyniki w tym względzie osiągnięto przy użyciu sieci Darknet53 (99 błędnych wskazań na raka złośliwego i 55 przypisań do stanu normalnego).



Rys. 6.5 – Rozkład dwóch rodzajów błędów rozpoznania zmian łagodnych uzyskany przez różne struktury sieci CNN.

Z punktu widzenia medycznego najbardziej krytyczny rodzaj błędów to rozpoznanie raka złośliwego jako stan normalny. Rozrzut tego typu wyników zawiera się w granicach od 20 (Resnet50) do 68 (Inceptionresnetv2). Z analizy wyników przedstawionych w tabeli 6.9 widoczna jest duża różnorodność rozkładów poszczególnych błędów. Najlepsza pod względem średniej dokładności rozpoznania klas (Inceptionv3 o ACC = 97.17%) ustępuje innym rozwiązaniom sieciowym względem innych miar jakości. Najlepsze wyniki w poszczególnych kategoriach jakości zostały zaznaczone w postaci pogrubionej. Cztery spośród sześciu rozwiązań sieciowych może pochwalić się dominacją względem określonej miary jakościowej. Stąd naturalnym rozwiązaniem wydaje się sprzęgnięcie poszczególnych rozwiązań w zespole i agregacja ich wyników. Biorąc pod uwagę zbliżone wartości miar jakościowych poszczególnych rozwiązań zdecydowano się na głosowanie większościowe oparte na prawdopodobieństwie klasowym (podobnie jak w poprzednich rozdziałach). Wyniki zespołu w rozpoznaniu 3 klas zamieszczono w tabeli 6.10.

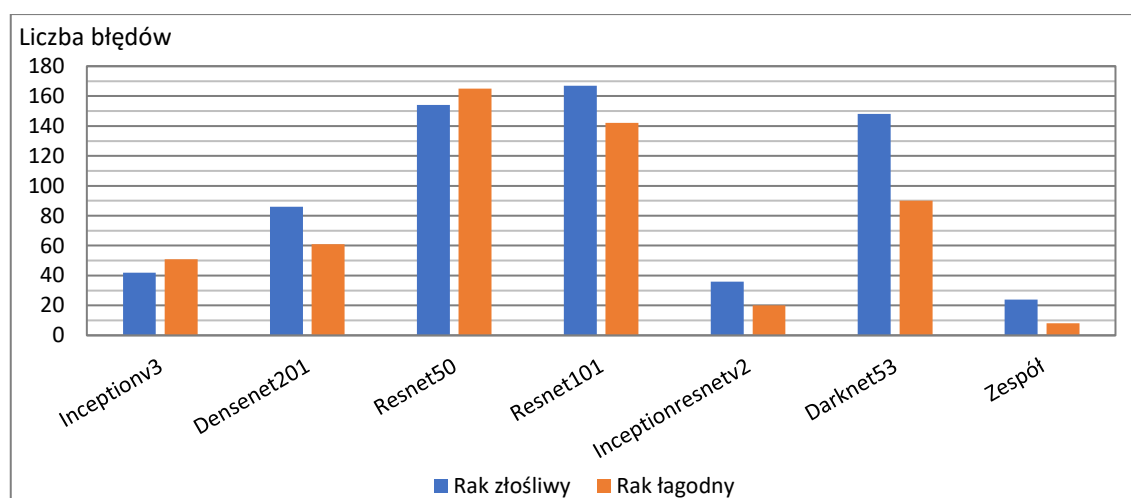
Tabela 6.10 – Wyniki zespołu złożonego z 6 sieci CNN przy rozpoznaniu 3 klas.

| Klasa | AUC | ACC | TPR | TNR | PPV | NPV | Conf. Matrix | | | Liczba pomyłek |
|-------|--------|--------|--------|--------|--------|--------|--------------|-----|------|----------------|
| 1 | 0.9959 | 0.9815 | 0.9327 | 0.9928 | 0.9344 | 0.9926 | 1040 | 47 | 28 | 75 |
| 2 | 0.9943 | | 0.8863 | 0.9947 | 0.9347 | 0.9903 | 49 | 787 | 52 | 101 |
| 3 | 0.9978 | | 0.9965 | 0.9601 | 0.9914 | 0.9836 | 24 | 8 | 9183 | 32 |

Zauważalna jest znaczna poprawa wszystkich wskaźników jakości, a przy tym większe zrównoważenie poziomu liczbowego tych wskaźników. Widoczne jest to zwłaszcza przy rozpoznaniu raka w postaci łagodnej (klasa 2) i złośliwej (klasa 1).

Na przykład wskaźnik PPV dla klasy drugiej zmieniający się od wartości 0.7543 (Resnet50) do 0.8754 (Inceptionresnetv2) został podniesiony w zespole do wartości 0.9344. Podobną poprawę tego wskaźnika przy rozpoznaniu postaci złośliwej raka (od 0.8016 dla sieci Resnet50 do 0.9063 dla Inceptionv3) obserwuje się dzięki zespołowi (PPV = 0.9344).

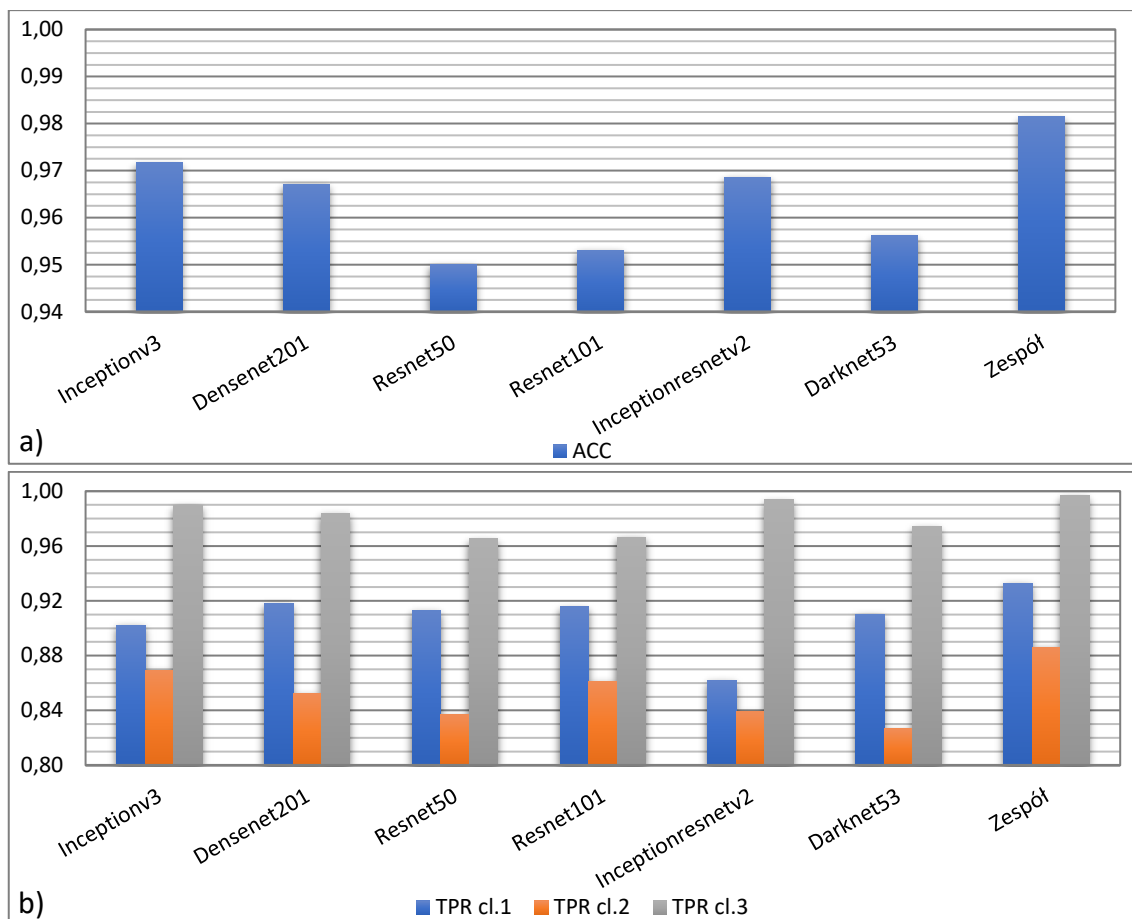
Najbardziej spektakularna poprawa dzięki zastosowaniu zespołu widoczna jest w liczbie pomyłek rozpoznania klasy normalnej jako nowotworu (łącznie złośliwego i łagodnego). Obrazuje to rysunek 6.6, w którym oś pionowa reprezentuje liczbę błędnych rozpoznań przypadków klasy normalnej dla różnych rozwiązań klasyfikatorów.



Rys. 6.6 – Rozkład liczby błędnego rozpoznania klasy normalnej przez poszczególnych członków zespołu w porównaniu z wynikiem zespołu agregowanym metodą głosowania większościowego.

Najgorszy wynik odpowiadający sieci Resnet50 to 319 takich błędów, podczas gdy zespół zdołał ograniczyć ich liczbę do wartości zaledwie 32 (24 pomyłki jako rak złośliwy i 8 jako zmiany łagodne). W przypadku rozpoznania pozostałych klas obserwuje się również znaczącą poprawę w ujęciu ilościowym. Błąd zaliczenia raka złośliwego do pozostałych dwóch klas zmalał z wartości maksymalnej 154 (Inceptioinresnetv2) do 75 w zespole. Liczba błędnych rozpoznań przypadków raka postaci łagodnej zmalała z wartości maksymalnej 154 (Darknet53) do 101 w zespole.

Analiza porównawcza wskaźników jakościowych rozpoznania 3 klas uzyskanych przez poszczególne architektury CNN i zespołu wskazuje na zdecydowaną przewagę zespołu dla każdego rodzaju wskaźnika. Dla lepszego przedstawienia tego problemu na rys. 6.7 zilustrowano graficznie rozkład dokładności średniej ACC (rys. 6.6 a) oraz czułości klasowej TPR (rys. 6.6 b) uzyskany przez rozwiązania indywidualne i wynik ich agregacji większościowej w zespole.



Rys. 6.7 – Rozkład wartości: a) dokładności średniej (ACC) oraz b) czułości klasowej (TPR) uzyskany przez rozwiązania indywidualne i wynik ich agregacji większościowej w zespole.

W każdym przypadku nastąpiła znacząca poprawa w stosunku do najlepszego indywidualnego rozwiązania. W tabeli 6.11 przedstawione jest porównanie miar jakości zintegrowanego zespołu z najlepszym, najgorszym i wynikiem średnim wszystkich 6 członków zespołu.

Tabela 6.11 – Porównanie miar jakości zintegrowanego zespołu z wynikami indywidualnymi.

| Wskaźnik jakości | Klasa | Wynik najlepszy | Wynik najgorszy | Średnia członków | Odchylenie standardowe | Wynik zespołu |
|------------------|-------|-----------------|-----------------|------------------|------------------------|---------------|
| ACC | - | 0.9717 | 0.9500 | 0.9611 | 0.0083 | 0.9815 |
| TPR | 1 | 0.9184 | 0.8619 | 0.9036 | 0.0193 | 0.9327 |
| | 2 | 0.8694 | 0.8266 | 0.8476 | 0.0149 | 0.8863 |
| | 3 | 0.9939 | 0.9654 | 0.9790 | 0.0111 | 0.9965 |
| TNR | 1 | 0.9897 | 0.9751 | 0.9815 | 0.0063 | 0.9928 |
| | 2 | 0.9897 | 0.9766 | 0.9850 | 0.0050 | 0.9947 |
| | 3 | 0.9666 | 0.9311 | 0.9528 | 0.0113 | 0.9601 |
| PPV | 1 | 0.9063 | 0.8016 | 0.8466 | 0.0444 | 0.9344 |
| | 2 | 0.8754 | 0.7543 | 0.8314 | 0.0468 | 0.9347 |
| | 3 | 0.9925 | 0.9852 | 0.9897 | 0.0023 | 0.9914 |
| NPV | 1 | 0.9909 | 0.9848 | 0.9893 | 0.0021 | 0.9926 |
| | 2 | 0.9888 | 0.9851 | 0.9869 | 0.0013 | 0.9903 |
| | 3 | 0.9708 | 0.8585 | 0.9104 | 0.0434 | 0.9836 |

7. Podsumowanie i wnioski końcowe

Celem niniejszej rozprawy doktorskiej było opracowanie systemu zespołowego klasyfikatorów oraz zbadanie jego skuteczności w procesie rozpoznawania obrazów medycznych czerniaka oraz raka piersi. Badania opierały się na wykorzystaniu głębokich sieci neuronowych typu CNN, które przetwarzając obraz na postać wektorową o ograniczonym wymiarze tworzyły wektor cech, stanowiący dane wejściowe dla klasyfikatorów tworzących zespół. Wynik rozpoznania jest tworzony jako fuzja wielu uzyskanych niezależnie rozwiązań poszczególnych jednostek zespołu. Proponowane rozwiązania budowy zespołu klasyfikatorów opierały się na zastosowaniu wielu głębokich sieci neuronowych CNN, których wyniki indywidualne poddano agregacji dla wyłonienia ostatecznej decyzji zespołu. Rozważano kilka możliwych rozwiązań.

Jednym z nich jest współpraca rozwiązań klasycznych (sieć SVM i las losowy drzew decyzyjnych) z siecią głęboką CNN. Wyniki przeprowadzonych eksperymentów pokazały stosunkowo niewielką poprawę wyników takiego rozwiązania w stosunku do najlepszych indywidualnych rozwiązań sieci CNN.

Znacznie lepsze rezultaty osiągnięto budując zespół złożony z wielu współpracujących ze sobą różnych sieci głębokich, których wyniki podlegają fuzji poprzez głosowanie większościowe, oparte na prawdopodobieństwie przynależności klasowej. Niezależnie od zadania klasyfikacji (rozpoznanie czerniaka skóry czy raka piersi) zespół taki wykazał się zwiększoną efektywnością rozpoznania klasy, mierzoną takimi wskaźnikami jakości jak średnia dokładność, czułość, specyficzność, precyzja klasowa czy pole AUC pod krzywą ROC.

Wyniki wielu przeprowadzonych eksperymentów numerycznych potwierdziły słuszność hipotezy, że opracowany system zespołu klasyfikatorów pozwala na istotną poprawę wyników rozpoznania klasy w stosunku do rezultatów osiągniętych przez indywidualne metody klasyfikacji.

Badania eksperymentalne zostały przeprowadzone dla dwóch rodzajów obrazowania. W przypadku czerniaka jest to obraz dermoskopowy skóry, a w przypadku raka piersi – obraz mammograficzny. Niezależnie od rodzaju obrazowania, opracowany system klasyfikacji okazał się bardzo wydajny, generując wyniki na poziomie najlepszych rozwiązań światowych.

W przypadku zbioru ISIC2017 wyniki uzyskane w rozprawie przewyższają najlepsze rezultaty dla tego zbioru prezentowane w literaturze światowej: dokładność

ACC = 96.545, czułość TPR = 94.71%, specyficzność TNR = 97.67%, precyzja rozpoznania obu klas powyżej 96%. Najlepsze wyniki światowe dla tego zbioru [1] omówione w rozdziale pierwszym to ACC = 90.4%, czułość TPR = 82% i specyficzność TNR = 92.5%.

W przypadku raka piersi badania przeprowadzone w ramach trzech zadań rozpoznania również należą do najlepszych na świecie, przy czym należy zaznaczyć, że wyniki porównawcze w literaturze światowej dotyczą w większości różnych baz danych, w tym DDSM, MIAS, ADMIANI itd. Najlepsze prezentowane wyniki rozpoznania nowotworu od przypadków zdrowych dotyczą pracy [63] operującej na bazie danych mammograficznych stworzonych w instytucji „Eastern Health in Newfoundland and Labrador of Canada”. Wyniki badań podano wyłącznie danych uczących: ACC = 96.06%, czułość TPR = 96.45%, specyficzność TNR = 97.45%, stąd trudno wyrokować o ich poziomie dla danych testujących nie uczestniczących w uczeniu.

Wyniki autora rozprawy dotyczą zbioru DDSM i mogą być obiektywnie porównane z innymi wynikami prezentowanymi dla tej samej bazy w literaturze światowej. W rozpoznaniu zmian nowotworowych od przypadków normalnych najlepsze wyniki zaprezentowano w pracy [54], w tym ACC = 89.71%, TPR = 93.54% TNR = 80.58 i AUC = 0.9410. Wyniki autora rozprawy to: ACC = 98.77%, TPR = 97.89%, TNR = 99.07%, AUC = 0.9975.

Przeprowadzone eksperymenty numeryczne wykorzystujące technikę transfer learning pokazały, niezależnie od rodzaju analizowanych danych, że można osiągnąć bardzo dobre rezultaty bez konieczności tworzenia od podstaw nowych struktur sieci neuronowych.

Zastosowanie bardziej zaawansowanych sieci CNN zorganizowane w formie zespołu pozwala zrekompensować jakość zbioru danych uczących zarówno pod kątem unormowania jak i zrównoważenia klas. Kosztem takiej rekompensaty jest wydłużenie czasu badań oraz wzrost mocy jednostki obliczeniowej wymaganej do przeprowadzenia procesu uczenia sieci.

Wnioski płynące z przeprowadzonych badań potwierdzają, że zespołowe uczenie maszynowe, oparte na kombinacji różnych klasyfikatorów, może przynieść znaczące korzyści w procesie analizy obrazów medycznych. Jednocześnie, wyniki te podkreślają wagę odpowiedniego doboru członków zespołu oraz ich optymalnej konfiguracji, aby osiągnąć najlepsze rezultaty.

Wartością dodaną tej pracy jest nie tylko wykazanie skuteczności zespołowego uczenia maszynowego w kontekście rozpoznawania obrazów medycznych, ale również wskazanie konkretnych strategii i podejść, które mogą być wykorzystane w praktyce klinicznej. Dalsze badania w tym obszarze mogą prowadzić do jeszcze bardziej precyzyjnych i skutecznych narzędzi wspomagających diagnozowanie chorób na podstawie obrazów medycznych.

7.1. Główne osiągnięcia rozprawy

Za główne osiągnięcia rozprawy autor uważa.

1. Opracowanie i przebadanie skuteczności działania zespołu klasyfikatorów opartych na głębokich sieciach CNN. W ramach eksperymentów przebadano współpracę rozwiązań klasycznych (sieci SVM i las losowy) z systemami głębokimi.
2. Opracowanie sposobu optymalnego doboru członków zespołu bazującego na analizie wartości trzech głównych miar jakości: dokładności, czułości i AUC. Wyniki badań eksperymentalnych przeprowadzonych w rozpoznaniu zarówno zmian skórnych (czerniak) jak i analizie obrazów mammograficznych raka piersi pokazały przewagę metod zespołowych nad najlepszymi rozwiązaniami indywidualnymi.
3. Przeprowadzenie ogromnej liczby eksperymentów numerycznych przy testowaniu zaproponowanych rozwiązań zespołowych. Wyniki tych eksperymentów potwierdziły skuteczność opracowanych metod tworzenia zespołów. Szczególną zaletą zespołu uwidocznioną w wynikach statystycznych badań jest znacząca poprawa precyzji rozpoznawania zmian nowotworowych w stosunku do najlepszych rozwiązań indywidualnych.

7.2. Kierunki dalszych badań

Głębokie sieci neuronowe typu CNN stanowią uniwersalną technikę pozwalającą na prowadzenie badań i tworzenie rozwiązań problemów z wielu różnych dziedzin nauki i techniki, w tym w inżynierii biomedycznej ukierunkowanej na wspomaganie diagnostyki.

Ciekawym kierunkiem badawczym jest opracowanie nowych metod obrazowania zmian skórnych. Przedstawione w rozprawie badania zostały przeprowadzone w oparciu o zdjęcia wykonane dermatoskopem. Rozszerzenie spektrum elektromagnetycznego

ponad światło widzialne wniesie do problemu nowe informacje, które w połączeniu z opisaną w pracy techniką uczenia maszynowego mogą przyczynić się do zwiększenia skuteczności diagnozowania nowotworów skóry. Rozwój technologii obrazowania, takich jak obrazy o wysokiej rozdzielczości, obrazy wielospektralne czy obrazy trójwymiarowe, może otworzyć nowe możliwości wykorzystania uczenia maszynowego w medycynie, umożliwiając bardziej zaawansowane metody automatycznej analizy i diagnostyki.

Istotny jest dalszy rozwój algorytmów uczenia maszynowego oraz opracowanie nowych struktur konwolucyjnych sieci neuronowych, które mogą w przyszłości doprowadzić do automatyzacji pewnych procedur diagnostycznych, takich jak skuteczne wykrywanie zmian patologicznych na podstawie obrazów medycznych czy identyfikacja obszarów ryzyka chorób. Rozwój w tym kierunku może skrócić czas diagnozy i poprawić dostęp do opieki zdrowotnej, szczególnie w obszarach o ograniczonym dostępie do specjalistów medycznych.

Wykaz literatury

- [1] Acosta M.F.J., Tovar L.Y.C., Garcia-Zapirain M.B., Percybrooks W., *Melanoma diagnosis using deep learning techniques on dermoscopic images*, BMC Med Imaging, vol. 21, 2021.
- [2] Almalki Y.E., Soomro T.A., Irfan M., Alduraibi S.K., Ali A., *Computerized analysis of mammogram images for early detection of breast cancer*, Healthcare (Basel). 2022, 10(5): 801, doi: 10.3390/healthcare10050801 (MIAS).
- [3] Barret Z., Vasudevan V., Shlens J., Quoc V.Le, *Learning Transferable Architectures for Scalable Image Recognition*, arXiv preprint arXiv:1707.07012 2, no. 6, 2017.
- [4] Bengio Y., LeCun Y., Hinton G., *Deep learning*, Nature, 2015, vol. 521, pp. 436–444.
- [5] Breiman L., *Random forests*, Machine Learning, 2001, vol. 45, No 11, pp. 5–32.
- [6] Cassidy B., Kendrick C., Brodzicki A., Jaworek-Korjakowska J., Yap M. H., *Analysis of the ISIC image datasets: usage, benchmarks and recommendations*, Medical Image Analysis, vol. 75, pp. 1-15, 2022.
<https://doi.org/10.1016/j.media.2021.102305>
- [7] Chollet F., *Xception: Deep Learning with Depthwise Separable Convolutions*, arXiv preprint, pp.1610-02357, 2017.
- [8] Codella N., Gutman D., Celebi M.E., Helba B., Marchetti M.A., Dusza S., Kalloo A., Liopyris K., Mishra N., Kittler H., Halpern A., *Skin lesion analysis toward melanoma detection: a challenge at the 2017 International Symposium on Biomedical Imaging (ISBI)*, Hosted by the International Skin Imaging Collaboration (ISIC), arXiv: 1710.05006 [cs.CV], 2017. <https://arxiv.org/abs/1710.05006>
- [9] Codella N., Nguyen Q.B., Pankanti S., Gutman D., Helba B., Halpern A., Smith J.R., *Deep learning ensembles for melanoma recognition in dermoscopy images*. IBM J. of Research and Development, vol. 61, no 4/5, 2017.
- [10] Dhahbi S., Barhoumi W., Zagrouba E., *Breast cancer diagnosis in digitized mammograms using curvelet moments*, Computers in Biology and Medicine 64(1), 79-90, 2015.
- [11] Donahue J., Krahenbuhl P., Darrell T., *Adversarial feature learning*, arXiv:1605.09782v7 [cs.LG], 2017.
- [12] Duda R.O., Hart P.E., Stork P., *Pattern classification and scene analysis*, Wiley, New York, 2003.
- [13] Esteva A., Kuprel B., Novoa R., Ko J., Swetter S., Blu H., Thrun S., *Dermatologist-level classification of skin cancer with deep neural networks*, Nature, vol. 542, pp. 115–118, 2017.

- [14] Frazer H.M.L., Tang J.S.N., Elliott M.S., Kunicki K.M., Hill B., Karthik R., Kwok C.F., Peña-Solorzano C.A., Chen Y., Wang C., et al. *ADMANI: Annotated Digital Mammograms and Associated Non-Image Datasets*. *Radiol. Artif. Intell.*, 5, e220072, 2023
- [15] Gil F., Osowski S., *Feature Selection Methods in Gene Recognition Problem*, *Computational Problems of Electrical Engineering (CPEE)*, pp. 1-4, doi: 10.1109/CPEE50798.2020.9238726), 2020
- [16] Gil F., Osowski S., *Fusion of feature selection methods in gene recognition*, *Bulletin of the Polish Academy of Sciences, Technical Sciences*, Vol. 69, No. 3, pp. 1-8, DOI: 10.24425/bpasts.2021.136748, 2021.
- [17] Gil F., Osowski S., Slowinska M., *Melanoma recognition using deep learning and ensemble of classifiers*, *Computational Problems of Electrical Engineering” (CPEE)*, pp. 1-4, DOI: 10.1109/CPEE56060.2022.9919681, 2022.
- [18] Gil F., Osowski S., Swiderski B., Slowinska M., *Ensemble of classifiers based on deep learning for medical image recognition*, *Metrology and Measurement Systems*, Vol. 30, No. 1, DOI: 10.24425/mms.2023.144400, 2023.
- [19] Goldsmith S. M., *A unifying approach to the clinical diagnosis of melanoma including “D” for “Dark” in the ABCDE criteria*, *Dermatol. Pract. Concepts*, vol. 4, pp. 75–78, 2014.
- [20] Goodfellow I., Bengio Y., Courville A., *Deep learning 2016*, MIT Press, Massachusetts (tłumaczenie polskie: *Deep Learning. Współczesne systemy uczące się*, Helion, Gliwice, 2018).
- [21] Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., *Generative adversarial nets*, *Proceedings of the International Conference on Neural Information Processing Systems (NIPS)*, pp. 2672–2680, 2014.
- [22] Guyon I., Elisseeff A., *An introduction to variable and feature selection*, *J. Mach. Learn. Res.*, vol. 3, pp. 1157-1182, 2003.
- [23] Guyon I., Weston J., Barnhill S., Vapnik V., *Gene selection for cancer classification using Support Vector Machines*, *Machine Learning*, vol. 46, pp. 389-422, 2002.
- [24] Ha Q., Liu B., Liu F., *Identifying melanoma images using EfficientNet ensemble: winning solution to the SIIM-ISIC Melanoma Classification Challenge*, arXiv:2010.05351v1 [cs.CV], 2020.
- [25] He K., Zhang X., Ren S., Sun J., *Deep residual learning for image recognition*, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.

- [26] Heath M., Bowyer K., Kopans D., Moore R., Kegelmeyer P., *The digital database for screening mammography*, in *Digital Mammography*, Springer, Netherlands, 457-460, 1998.
- [27] Howard A., Zhu M., Chen B., Kalenichenko D., *MobileNets: Efficient convolutional neural networks for mobile vision applications*, arXiv: 1704.04861v1 [cs.CV], 2017.
- [28] Huang G., Liu Z., van der Maaten L., Weinberger K., *Densely connected convolutional networks*, arXiv: 1606.06993v5 [cs.CV], 2018.
- [29] Iandola F., Han S., Moskevycz M., Ashraf K., Dally W., Keutzer K., *Squeezenet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size*, Conference ICLR, pp. 1-13, 2017.
- [30] Jafarzadeh H., Mahdianpari M., Gill E., Mohammadimanesh F., Homayouni S., *Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation*, *Remote Sensing*, 13(21), 4405, <https://doi.org/10.3390/rs13214405>, 2021.
- [31] Joson N., Nair M.S., *On the performance of CNN and GAN models for melanoma classification*, International Conference on Artificial Intelligence and Signal Processing (AISP), IEEE | DOI: 10.1109/AISP53593.2022.9760626, 2022.
- [32] Kaur R., Gholamhosseini H., Sinha R., Linden M., *Melanoma classification using a novel deep convolutional neural network with dermoscopic images*, *Sensors*, <https://doi.org/10.3390/s22031134>, 2022.
- [33] Kingma P., Ba J.L., *Adam: A method for stochastic optimization*, arXiv:1412.6980v9, 2014.
- [34] Krizhevsky A., Sutskever I., Hinton G., *ImageNet Classification with Deep Convolutional Neural Networks*, *Communications of the ACM* 60, no. 6, 84–90, <https://doi.org/10.1145/3065386>, 2017.
- [35] Kruk M., Świdorski B., Osowski S., Kurek J., Słowińska M., Walecka I., *Melanoma recognition using extended set of descriptors and classifiers*, *Eurasip Journal on Image and Video Processing*, vol. 43, pp. 1-10, DOI 10.1186/s13640-015-0099-9, 2015.
- [36] Kuncheva L., *Combining pattern classifiers: methods and algorithms*, Wiley, New York, 2015.
- [37] Kurek J., Swiderski B., Osowski S., Kruk M., Barhoumi W., *Deep learning versus classical neural approach to mammogram recognition*, *Bulletin of the Polish Academy of Sciences, Technical Sciences*, Vol. 66, No. 6, pp. 831-840, 2018.
- [38] Lehmann E. L., Romano J. P., *Testing statistical hypotheses*, Springer, 2022.
- [39] *Matlab user manual*, MathWorks, Natick, USA, 2023.

- [40] Melanoma Skin Cancer, American Cancer Society website.
<https://www.cancer.org/cancer/types/melanoma-skin-cancer/about.html>.
- [41] Nicosia L., Gnocchi G., Gorini I., Venturini M., et al., *History of mammography: analysis of breast imaging diagnostic achievements over the last century*, Healthcare, 11(11), doi: 10.3390/healthcare10050801, 2023.
- [42] Osowski S., Leś T., *Deep Learning Ensemble for Melanoma Recognition*, IEEE International Joint Conference on Neural Networks, Glasgow, 2020.
- [43] Osowski S., *Sieci neuronowe do przetwarzania informacji*, Oficyna Wydawnicza PW, 2020.
- [44] Osowski S., Szmurło R., *Matematyczne modele uczenia maszynowego w językach Matlab i Python*, Oficyna Wydawnicza PW, 2023.
- [45] Peng H., Long F., Ding C., *Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 1226-1238, <https://doi.org/10.1109/TPAMI.2005.159>, 2005
- [46] Pesapane F., Trentin C., Ferrari F., Signorelli G., et al., *Deep learning performance for detection on microcalcification on mammography*, European Radiology Experimental, 7(69), 2023.
- [47] Prodan M., Paraschiv E., Stanciu A., *Applying deep learning methods for mammography analysis and breast cancer detection*, Applied Sciences, 13(7), DOI: 10.3390/app13074272, 2023
- [48] Redmon J., *Darknet: Open Source Neural Networks in C*, <https://pjreddie.com/darknet>.
- [49] Robnik-Sikonja M., Kononenko I., *Theoretical and empirical analysis of ReliefF and RReliefF*, Machine Learning, vol. 53, pp. 23–69, 2003.
- [50] Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.C., *MobileNetV2: Inverted Residuals and Linear Bottlenecks*, IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4510-4520), IEEE, 2018.
- [51] Schölkopf B., Smola A., *Learning with kernels*, MIT Press, Cambridge MA, 2002.
- [52] Shahid A., *EfficientNet: Scaling of convolutional neural networks done right*, <https://towardsdatascience.com/efficientnet-scaling-of-convolutional-neural-networks-done-right-3fde32aef8ff>.
- [53] Simonyan K., Zisserman A., *Very deep convolutional networks for large-scale image recognition*, arXiv preprint arXiv:1409.1556, 2014.
- [54] Swiderski B., Gielata Ł., Olszewski P., Osowski S., Kołodziej M., *Deep neural system for supporting tumor recognition of mammograms using modified GAN*, Expert Systems with Applications, vol. 164 (113968), pp. 1-10, <https://doi.org/10.1016/j.eswa.2020.113968>, 2021.

- [55] Swiderski B., Osowski S., Gwardys G., Kurek J., Slowinska M., Lugowska I., Random CNN structure – tool to increase generalization ability in deep learning, *Eurasip Journal on Image and Video Processing*, s.1-18. DOI:10.21203/rs.3.rs-277475/v1, 2022.
- [56] Swiderski B., Osowski S., Kruk M., Barhoumi W., *Aggregation of classifiers ensemble using local discriminatory power and quantiles*, *Expert Systems with Applications*, vol. 46, pp. 316-323, 2016.
- [57] Swiderski B., Osowski S., Kurek J., Kruk M., Lugowska I., Rutkowski P., Barhoumi W., *Novel methods of image description and ensemble of classifiers in application to mammogram analysis*, *Expert Systems with Applications*, vol. 81, pp. 67–78, 2017.
- [58] Szegedy C., Ioffe S., Vanhoucke V., *Inception-v4, Inception-ResNet and the impact of residual connections on learning*, arXiv:1602.07261v2, 2016.
- [59] Szegedy C., Liu W., Yangqing J., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A., *Going deeper with convolutions*, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015.
- [60] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z., *Rethinking the inception architecture for computer vision*, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818-2826, 2016.
- [61] Tan M., Le Q.V., *Efficientnet: rethinking model scaling for convolutional neural networks*, arXiv: 1905.11946v5 [cs.LG], 2020.
- [62] Tan P.N., Steinbach M., Kumar V., *Introduction to data mining*, Pearson Education Inc., Boston, 2014.
- [63] Tang X., Zhang L., Zhang W., Huang X., Iosifidis V., Liu Z., Zhang M., Messina E., Zhang J., *Using machine learning to automate mammogram images analysis*, 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), DOI: 10.1109/BIBM49941.2020.9313247, 2020.
- [64] The International Skin Imaging Collaboration, <https://isic-archive.com/>
- [65] Trajdos P., Kurzynski M., *Building an ensemble of classifiers via randomized models of ensemble members*, arXiv:2109.07861v1. <https://doi.org/10.48550/arXiv.2109.07861>
- [66] Van der Maaten L., Hinton G., *Visualising data using t-SNE*, *Journal of Machine Learning Research*, vol. 9, pp. 2579-2602, 2008.
- [67] Vapnik V., *Statistical learning theory*, Wiley, New York, 1998.
- [68] Yagerman S.E., Marghoob A., *The ABCDs and beyond*, *The Skin Cancer Foundation Journal*, vol. 94, pp. 31-61, 2013.
- [69] Yang X., Zen Z., Yeo S.Y., Tan C., Tey H.L., Su I., *A novel multitask deep learning model for skin lesion segmentation and classification*, 2017.

- [70] Yuexiang L., Linlin S., *Skin lesion analysis towards melanoma detection using deep learning network*, Sensors (Basel), vol. 18, pp. 1-8, 2018.
- [71] Zhang X., Zhou X., Lin M., Sun J., *ShuffleNet: an extremely efficient convolutional neural network for mobile devices*, arXiv: 1707.01083v2 [cs.CV], 2017.
- [72] Zou J., Fu X., Guo L., Ju C., Chen J., *Creating ensemble classifiers with information entropy diversity measure*, Security and Communication Networks, 2021.

Spis rysunków

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| RYS. 2.1 – PRZYKŁADOWE OBRAZY ZBIORU DANYCH NIO. OBRAZY W GÓRNYM WIERSZU PRZEDSTAWIAJĄ CZERNIAKA, A W DOLNYM WIERSZU PRZYPADKI NIEBĘDĄCE CZERNIAKIEM..... | 15 |
| RYS. 2.2 – PRZYKŁADOWE OBRAZY ZBIORU DANYCH ISIC. OBRAZY W GÓRNYM SZEREGU PRZEDSTAWIAJĄ CZERNIAKA, A W DOLNYM SZEREGU PRZYPADKI NIEBĘDĄCE CZERNIAKIEM..... | 18 |
| RYS. 2.3 – PRZYKŁADOWE OBRAZY MAMMOGRAFICZNE. OBRAZY W GÓRNYM SZEREGU PRZEDSTAWIAJĄ RAKA W STADIUM ZŁOŚLIWYM, W ŚRODKOWYM RAKA W STADIUM ŁAGODNYM, A W DOLNYM PRZYPADKI TKANEK NORMALNYCH..... | 20 |
| RYS. 3.1 – SCHEMAT IDEOWY KONWOLUCYJNEJ SIECI NEURONOWEJ..... | 24 |
| RYS. 3.2 – ILUSTRACJA OPERACJI MAX POOLING I AVERAGE POOLING O ROZMIARZE 2x2. | 25 |
| RYS. 3.3 – STRUKTURA BLOKU REZYDUALNEGO W SIECI RESNET. POŁĄCZENIE SKROŚNE (REZYDUALNE) WYSTĘPUJE MIĘDZY DWOMA WARSTWAMI KONWOLUCYJNYMI [52]. | 27 |
| RYS. 3.4 – PRZYKŁADOWA STRUKTURA SIECI RESNET34. POŁĄCZENIE KROPKOWANE WYSTĘPUJE, GDY ROZMIARY WYMIARY WARSTWY STARTOWEJ I DOCELOWEJ RÓŻNIĄ SIĘ. | 28 |
| RYS. 3.5 – STRUKTURA POŁĄCZEŃ SOFTNETU. SYGNAŁY x_i REPREZENTUJĄ ELEMENTY WYPŁASZCZONEGO WEKTORA CECH, LICZBA NEURONÓW WYJŚCIOWYCH JEST RÓWNA LICZBIE KLAS, SYGNAŁY p_i REPREZENTUJĄ WARTOŚCI PRAWDOPODOBIENSTWA PRZYNALEŻNOŚCI DANYCH WEJŚCIOWYCH DO ODPOWIEDNIEJ KLASY. ZWYCIĘŻA KLASA O NAJWIĘKSZEJ WARTOŚCI PRAWDOPODOBIENSTWA..... | 30 |
| RYS. 3.6 – ALTERNATYWNE ROZWIĄZANIE KLASYFIKATORA KOŃCOWEGO WYKORZYSTUJĄCE DETERMINISTYCZNY ALGORYTM SELEKCJI CECH I KLASYCZNE ROZWIĄZANIE KLASYFIKATORÓW. | 33 |
| RYS. 3.7 – WYNIKI DZIAŁANIA MECHANIZMÓW SELEKCJI CECH METODĄ TESTU STATYSTYCZNEGO KRUSKAL-WALLISA, FISHERA I REFIEFF DLA PIERWSZYCH 100 DESKRYPTORÓW NUMERYCZNYCH DLA OBRAZÓW CZERNIAKA..... | 39 |
| RYS. 3.8 – PRZYKŁAD USZEREGOWANIA WAŻNOŚCI 10 NAJLEPSZYCH CECH DIAGNOSTYCZNYCH DLA SELEKCJI METODĄ FISHERA (A) I KRUSKAL-WALLISA (B)..... | 41 |
| RYS. 4.1 – SCHEMAT TYPOWEJ STRUKTURY ZESPOŁU KLASYFIKATORÓW [43]. | 44 |
| RYS. 4.2 – SCHEMAT INTEGRATORA NEURONOWEGO DO AGREGACJI WAŻONEJ WYNIKÓW M CZŁONKÓW ZESPOŁU. ZMIENNE Y_1, Y_2, \dots, Y_M OZNACZAJĄ WEKTORY WYNIKOWE M KLASYFIKATORÓW W ZESPOLE..... | 46 |
| RYS. 4.3 – SCHEMAT PIERWSZEGO ROZWIĄZANIA ZESPOŁU UTWORZONEGO PRZEZ 16 KLASYFIKATORÓW KLASYCZNYCH ZBUDOWANYCH NA BAZIE SVM I RF I 8 METODACH DETERMINISTYCZNYCH SELEKCJI ORAZ KLASYFIKATORA TYPU SOFTMAX | |

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| WYKORZYSTUJĄCEGO LOSOWY DOBÓR CECH DOKONYWANY W KAŻDEJ ITERACJI UCZĄCEJ (ZMIANY)..... | 46 |
| RYS. 4.4 – SCHEMAT POŁĄCZEŃ DRUGIEGO ROZWIĄZANIA ZESPOŁU SKŁADAJĄCEGO SIĘ WYŁĄCZNIE ZE STRUKTUR GŁĘBOKICH CNN..... | 47 |
| RYS. 5.1 – WYKRES AUC I DOKŁADNOŚCI ROZPOZNANIA CZERNIAKA W ZBIORZE NIO PRZY ZASTOSOWANIU 5 WYBRANYCH STRUKTUR SIECI CNN PRZY ZAMROŻENIU WSZYSTKICH WARSTW LOKALNIE POŁĄCZONYCH..... | 55 |
| RYS. 5.2 – ŚREDNIE WARTOŚCI POLA AUC WYBRANYCH KLASYFIKATORÓW W ZALEŻNOŚCI OD ILOŚCI ZAMROŻONYCH WARSTW LOKALNIE POŁĄCZONYCH DLA SIECI GOOGLNET, INCEPTIONV3, DENSENET201, DARKNET19 ORAZ ALEXNET..... | 56 |
| RYS. 5.3 – ŚREDNIE WARTOŚCI DOKŁADNOŚCI ACC WYBRANYCH KLASYFIKATORÓW W ZALEŻNOŚCI OD ILOŚCI ZAMROŻONYCH WARSTW LOKALNIE POŁĄCZONYCH DLA SIECI GOOGLNET, INCEPTIONV3, DENSENET201, DARKNET19 ORAZ ALEXNET..... | 56 |
| RYS. 5.4 – ŚREDNIE WARTOŚCI CZASU UCZENIA WYBRANYCH KLASYFIKATORÓW W ZALEŻNOŚCI OD ILOŚCI ZAMROŻONYCH WARSTW LOKALNIE POŁĄCZONYCH..... | 57 |
| RYS. 5.5 – WYKRES DOKŁADNOŚCI KLASYFIKACJI PRZY ZASTOSOWANIU RÓŻNYCH MODELI SIECI CNN NA ZBIORZE NIO..... | 59 |
| RYS. 5.6 – WYKRES POWIERZCHNI AUC POD KRZYWĄ ROC W KLASYFIKACJI PRZY ZASTOSOWANIU RÓŻNYCH MODELI SIECI CNN NA ZBIORZE NIO..... | 59 |
| RYS. 5.7 – WYKRES CZASU TRENOWANIA I KLASYFIKACJI W TRYBIE 5-KROTNEJ WALIDACJI KRZYŻOWEJ PRZY UŻYCIU RÓŻNYCH MODELI SIECI CNN NA ZBIORZE NIO. GÓRNY WYKRES PRZEDSTAWIA WSZYSTKIE BADANE SIECI CNN. NA DOLNYM WYKRESIE WYKLUCZONO ZE ZBIORU SIEĆ NASNETLARGE DLA UWYPUKLENIA RÓŻNIC CZASOWYCH DLA POZOSTAŁYCH SIECI..... | 60 |
| RYS. 5.8 – WYKRES CZUŁOŚCI (TPR) I SPECYFICZNOŚCI (TNR) BADANYCH MODELI KLASYFIKATORÓW..... | 61 |
| RYS. 5.9 – PRZYKŁADY OBRAZÓW CZERNIAKA SKLASYFIKOWANYCH Z BŁĘDEM II RODZAJU..... | 62 |
| RYS. 5.10 – WYKRES PORÓWNUJĄCY PRECYZJĘ ROZPOZNAWANIA OBU KLAS WSZYSTKICH BADANYCH MODELI KLASYFIKATORÓW. WSKAŹNIK PPV DOTYCZY CZERNIAKA, A NPV BRODAWKI ŁOJOTOKOWEJ..... | 63 |
| RYS. 5.11 – WYKRESY WSKAŹNIKÓW JAKOŚCI ODPOWIADAJĄCE RÓŻNYM ARCHITEKTUROM SIECI CNN: A) AUC, ACC, B) TPR, TNR, C) PPV, NPV..... | 75 |
| RYS. 6.1 – WYKRESY PRZEDSTAWIAJĄCE WARTOŚCI MIAR JAKOŚCI POSZCZEGÓLNYCH ROZWIĄZAŃ SIECI CNN: A) DOKŁADNOŚĆ ACC I MIARA AUC, B) CZUŁOŚĆ TPR I SPECYFICZNOŚĆ TNR..... | 82 |
| RYS. 6.2 – WYKRESY MIAR JAKOŚCI W ROZPOZNANIU RAKA ZŁOŚLIWEGO OD POSTACI ŁAGODNEJ DLA ZESPOŁU NUMER 1..... | 87 |

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| RYS. 6.3 – ILUSTRACJA ROZKŁADU WARTOŚCI ACC I AUC POSZCZEGÓLNYCH SIECI CNN W PORÓWNANIU Z WYNIKAMI ZINTEGROWANEGO ZESPOŁU TYCH SIECI. | 89 |
| RYS. 6.4 – PORÓWNANIE WARTOŚCI PRECYZJI KLASOWEJ PPV I NPV DLA POSZCZEGÓLNYCH SIECI CNN Z WYNIKAMI ZINTEGROWANEGO ZESPOŁU TYCH SIECI. | 90 |
| RYS. 6.5 – ROZKŁAD DWÓCH RODZAJÓW BŁĘDÓW ROZPOZNANIA ZMIAN ŁAGODNYCH UZYSKANY PRZEZ RÓŻNE STRUKTURY SIECI CNN. | 92 |
| RYS. 6.6 – ROZKŁAD LICZBY BŁĘDNEGO ROZPOZNANIA KLASY NORMALNEJ PRZEZ POSZCZEGÓLNYCH CZŁONKÓW ZESPOŁU W PORÓWNANIU Z WYNIKIEM ZESPOŁU AGREGOWANYM METODĄ GŁOSOWANIA WIĘKSZOŚCIOWEGO. | 93 |
| RYS. 6.7 – ROZKŁAD WARTOŚCI: A) DOKŁADNOŚCI ŚREDNIEJ (ACC) ORAZ B) CZUŁOŚCI KLASOWEJ (TPR) UZYSKANY PRZEZ ROZWIĄZANIA INDYWIDUALNE I WYNIK ICH AGREGACJI WIĘKSZOŚCIOWEJ W ZESPOLE. | 94 |

Spis tabel

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| TABELA 2.1 - PARAMETRY STATYSTYCZNE OBRAZÓW ZE ZMIANAMI CZERNAKOWYMI I NIECZERNAKOWYMI ZBIORU NIO DLA WSZYSTKICH OBRAZÓW ZBIORU. | 16 |
| TABELA 2.2 WARTOŚCI INDEKSU PODOBIENSTWA STRUKTURALNEGO WEWNĄTRZKLASOWEGO I MIĘDZYKLASOWEGO OBRAZÓW DLA BAZY NIO. | 17 |
| TABELA 2.3 - PARAMETRY STATYSTYCZNE OBRAZÓW ZE ZMIANAMI CZERNAKOWYMI I NIECZERNAKOWYMI ZBIORU ISIC (ŚREDNIA I ODCHYLENIE STANDARDOWE ZE WSZYSTKICH OBRAZÓW ZBIORÓW). | 19 |
| TABELA 2.4 - PARAMETRY STATYSTYCZNE OBRAZÓW MAMMOGRAFICZNYCH RAKA W STADIUM ZŁOŚLIWYM I ŁAGODNYM ORAZ TKANEK NORMALNYCH DLA WSZYSTKICH OBRAZÓW ZBIORU. | 20 |
| TABELA 2.5 - WARTOŚCI INDEKSU PODOBIENSTWA STRUKTURALNEGO WEWNĄTRZKLASOWEGO I MIĘDZYKLASOWEGO OBRAZÓW MAMMOGRAFICZNYCH DLA BAZY DDSM. | 21 |
| TABELA 3.1 WIELKOŚCI STATYSTYCZNE CHARAKTERYZUJĄCE 300 NAJWAŻNIEJSZYCH CECH DIAGNOSTYCZNYCH OBRAZÓW MELANOMY WYSELEKCYJONOWANYCH PRZE POSZCZEGÓLNE METODY SELEKCJI. | 40 |
| TABELA 4.1 – WYKAZ WSTĘPNIE PRZETRENOWANYCH ARCHITEKTUR SIECI CNN BIORĄCYCH UDZIAŁ W BADANIACH SYMULACYJNYCH. * SIECI NEURONOWE NASNETMOBILE I NASNETLARGE NIE SKŁADAJĄ SIĘ Z LINIOWEJ SEKWENCJI MODUŁÓW.. | 48 |
| TABELA 4.2 – OZNACZENIA MACIERZY POMYŁEK REPREZENTUJĄCEJ ROZKŁADY 2 KLAS (+ OZNACZA KLASĘ POSZUKIWANĄ, NATOMIAST – KLASĘ REPREZENTUJĄCĄ POZOSTAŁE PRZYPADKI (ZWANĄ DALEJ KLASĄ WIĘKSZOŚCIOWĄ) | 50 |
| TABELA 5.1 – PORÓWNANIE WYNIKÓW KLASYFIKACJI PRZY UŻYCIU RÓŻNYCH SIECI CNN W ZALEŻNOŚCI OD ILOŚCI ZAMROŻONYCH WARSTW | 58 |
| TABELA 5.2 – WYNIKI ZBIORCZE DOTYCZĄCE WARTOŚCI WSKAŹNIKÓW JAKOŚCI ROZPOZNANIA CZERNAKA OD BRODAWKI DLA RÓŻNYCH INDYWIDUALNYCH ROZWIĄZAŃ SIECIOWYCH CNN NA ZBIORZE NIO. | 64 |
| TABELA 5.3 – POPULACJA DESKRYPTORÓW GENEROWANYCH PRZEZ RÓŻNE STRUKTURY SIECI CNN. | 65 |
| TABELA 5.4 – WYBRANA LICZBA CECH DLA KAŻDEJ Z BADANYCH METOD SELEKCJI DLA SIECI SVM I RF (INCEPTIONV3). | 66 |
| TABELA 5.5 – WYNIKI ZESPOŁÓW STOSUJĄCYCH WSPÓŁPRACĘ SIECI SVM I RF Z RÓŻNYMI STRUKTURAMI GŁĘBOKIMI CNN (ZBIÓR NIO). | 68 |
| TABELA 5.6 – WYNIKI ZESPOŁÓW STOSUJĄCYCH WSPÓŁPRACĘ SIECI SVM Z RÓŻNYMI STRUKTURAMI GŁĘBOKIMI CNN BEZ UWZGLĘDNIENIA SIECI RF (ZBIÓR NIO). | 69 |

| | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| TABELA 5.7 – SKŁADY ZESPOŁÓW SIECI CNN WYŁONIONE W WYNIKU WSTĘPNYCH EKSPERYMENTÓW (ZBIÓR NIO). | 70 |
| TABELA 5.8 – WYNIKI ZESPOŁÓW ZBUDOWANYCH WYŁĄCZNIE Z RÓŻNYCH STRUKTUR SIECI GŁĘBOKICH (ZBIÓR NIO)..... | 71 |
| TABELA 5.9 – WYNIKI ZBIORCZE DOTYCZĄCE WARTOŚCI WSKAŹNIKÓW JAKOŚCI ROZPOZNANIA CZERNIAKA DLA RÓŻNYCH INDYWIDUALNYCH ROZWIĄZAŃ SIECIOWYCH CNN NA ZBIORZE ISIC. | 74 |
| TABELA 5.10 – ZESTAWIENIE KOMBINACJI SIECI CNN TWORZĄCYCH RÓŻNE SKŁADY ZESPOŁÓW..... | 77 |
| TABELA 5.11 – WARTOŚCI WSKAŹNIKÓW JAKOŚCI DLA ZBIORU ISIC UZYSKANE PRZEZ ZESPOŁY ZŁOŻONE Z WYBRANYCH SIECI CNN (11 KOMBINACJI SIECIOWYCH WSKAZANYCH W TABELI 5.10). | 78 |
| TABELA 6.1 – WYNIKI STATYSTYCZNE ROZPOZNANIA RAKA ZŁOŚLIWEGO OD POSTACI ŁAGODNEJ DLA POSZCZEGÓLNYCH ARCHITEKTUR SIECI CNN..... | 81 |
| TABELA 6.2 – WARTOŚCI STATYSTYCZNE ANALIZOWANYCH MIAR JAKOŚCI KLASYFIKATORÓW. | 83 |
| TABELA 6.3 – ANALIZOWANE SKŁADY KOMBINACJI ZESPOŁÓW TWORZONYCH PRZEZ RÓŻNE STRUKTURY SIECI CNN Z UWZGLĘDNIENIEM STRUKTURY NASNETLARGE. POSZCZEGÓLNE SIECI OZNACZONE SĄ W POSTACI NUMERYCZNEJ. | 84 |
| TABELA 6.4 – WYNIKI KLASYFIKACJI RÓŻNYCH SKŁADÓW ZESPOŁU PRZY ROZPOZNANIU RAKA ZŁOŚLIWEGO OD POSTACI ŁAGODNEJ Z UWZGLĘDNIENIEM SIECI NASNETLARGE.... | 85 |
| TABELA 6.5 – SKŁADY KOMBINACJI ZESPOŁÓW TWORZONYCH PRZEZ STRUKTURY SIECI CNN Z POMIĘCIEM SIECI NASNETLARGE. | 86 |
| TABELA 6.6 – WYNIKI KLASYFIKACJI ZESPOŁÓW SIECI CNN ZBUDOWANYCH BEZ NAJLEPSZEJ SIECI NASNETLARGE PRZY ROZPOZNANIU RAKA ZŁOŚLIWEGO OD POSTACI ŁAGODNEJ. | 86 |
| TABELA 6.7 – WYNIKI STATYSTYCZNE POSZCZEGÓLNYCH SIECI CNN PRZY ROZPOZNANIU ZMIAN NOWOTWOROWYCH (RAK ZŁOŚLIWY I ŁAGODNY TWORZĄCE WSPÓLNA KLASĘ) OD STANU NORMALNEGO..... | 88 |
| TABELA 6.8 – WYNIKI STATYSTYCZNE ZESPOŁU SZĘŚCIU SIECI CNN PRZY ROZPOZNANIU ZMIAN NOWOTWOROWYCH (RAK ZŁOŚLIWY I ŁAGODNY TWORZĄCE WSPÓLNA KLASĘ) OD STANU NORMALNEGO..... | 89 |
| TABELA 6.9 – WARTOŚCI WSKAŹNIKÓW JAKOŚCI PRZY ROZPOZNANIU TRZECH KLAS MAMOGRAMÓW. | 91 |
| TABELA 6.10 – WYNIKI ZESPOŁU ZŁOŻONEGO Z 6 SIECI CNN PRZY ROZPOZNANIU 3 KLAS. | 92 |
| TABELA 6.11 – PORÓWNANIE MIAR JAKOŚCI ZINTEGROWANEGO ZESPOŁU Z WYNIKAMI INDYWIDUALNYMI. | 94 |