

Kierunek: **ISA**
Specjalność: **IPS**

PRACA DYPLOMOWA
MAGISTERSKA

**Analiza skuteczności algorytmów lokalizacji komórek
nowotworowych w oparciu o dane IoT**

Łukasz Kołtun

Opiekun pracy
dr inż. Łukasz Jeleń

Słowa kluczowe: cytologia, uczenie głębokie, konwolucyjne sieci neuronowe, vision transformer

Field of study: **ISA**
Speciality: **IPS**

MASTER THESIS

The Effectiveness Analysis of Cancer Cell Localization Algorithms based on IoT Data

Łukasz Koltun

Supervisor
PhD. Eng. Łukasz Jeleń

Keywords: cytology, deep learning, convolutional neural networks, vision transformer

Streszczenie

Diagnostyka cytologiczna bazuje obecnie na wiedzy patomorfologów przy wykorzystaniu mikroskopów oraz prostych systemów wspomagających. Ocena obrazów zajmuje dużo czasu, co ogranicza możliwości wydajnej pracy. W ostatnich latach cytologiczne analiza obrazów przy wykorzystaniu technik sieci głębokich takich jak np. sieci konwolucyjne poczyniła znaczne postępy osiągając znacznie lepsze wyniki w porównaniu z tradycyjnymi metodami uczenia maszynowego. Głównym celem tej pracy było przeprowadzenie badań porównawczych algorytmów uczenia głębokiego wykorzystywanych w diagnostyce cytologicznej w celu klasyfikacji zmian nowotworowych oraz opracowanie systemu rejestrującego obrazy (przy pomocy urządzeń wykorzystujących IoT) montowanego na dowolnym mikroskopie optycznym. Zaproponowano 4 metody wykorzystujące uczenie głębokie: własną sieć CNN, VGG-16, ResNet-50 oraz Vision Transformer, które zostały przetrenowane oraz zweryfikowane na zbiorze różnorodnych danych, których autorem jest prof. dr hab. n. med. Maria Chosia z Katedry i Zakładu Patomorfologii Pomorskiego Uniwersytetu Medycznego w Szczecinie. Metody te zostały dostosowane do zadania poprzez augmentację danych, dodanie warstw oraz wykorzystanie transferu wag (dla ResNet-50 oraz VGG-16). Wyniki eksperymentów wykazały, że najlepiej działającym modelem okazał się VGG-16 osiągając 86,3% średniej dokładności na zbiorze testowym. Również dla innych metryk: średniego wyniku F1, średniej czułości oraz średniej precyzji sieć VGG-16 zanotowała najlepsze wyniki, wynosząc odpowiednio 84%, 83,7% i 85,7%. Ponadto modelem o najszybszym czasie treningu okazał się model Vision Transformer, a model oparty na własnej sieci CNN posiadał najmniejszą liczbę parametrów. Rezultaty, biorąc pod uwagę, niewielką ilość obrazów w zbiorze są zadowalające. Na potrzeby przyszłych prac metody powinny zostać przetestowane na większym zbiorze danych oraz przy użyciu bardziej zaawansowanych metod augmentacji obrazów. Warto również sprawdzić inne architektury modeli. Algorytmy zostały również sprawdzone na obrazach wykonanych przez opracowany system rejestracji obrazów na mikroskopie optycznym. System ten umożliwia własnoręczne zbieranie danych do badań na potrzeby przyszłych prac.

Abstract

Diagnostic cytology is currently based on the knowledge of pathomorphologists using microscopes and simple imaging systems. Image analysis is time-consuming, which limits efficient work possibilities. In recent years, cytological imaging using deep learning techniques, such as convolutional neural networks, has made significant progress, achieving much better results compared to traditional machine learning methods. The main objective of this study was to conduct comparative research on deep learning algorithms used in cytological diagnostics for the classification of tumor changes and to develop an image registration system (using IoT devices) that can be mounted on any optical microscope. Four deep learning methods were proposed: a custom CNN, VGG-16, ResNet-50, and Vision Transformer, which were trained and verified on a diverse dataset authored by Prof. Md. Maria Chosia from the Department of Pathomorphology at Pomeranian Medical University in Szczecin. These methods were adapted for the task through data augmentation, the addition of layers, and the use of transfer learning (for ResNet-50 and VGG-16). The results of the experiments showed that the VGG-16 model performed best, achieving an average accuracy of 86.3% on the test set. The VGG-16 network also achieved the best results for other metrics: average F1 score, average sensitivity, and average precision, with values of 84%, 83.7%, and 85.7% respectively. Additionally, the Vision Transformer model had the fastest training time, while the custom CNN-based model had the fewest parameters. Considering the small number of images in the dataset, the results are satisfactory. For future work, the methods should be tested on a larger dataset and using more advanced image augmentation techniques. It is also worth exploring other model architectures. The algorithms were also tested on images captured by the developed image registration system on an optical microscope. This system enables manual data collection for future research purposes.

Podziękowania

Pragnę wyrazić szczerze podziękowania dla promotora tej pracy Pana dr Łukasza Jelenia za poświęcony czas, wsparcie, motywację oraz wskazówki i materiały, bez których napisanie niniejszej pracy byłoby niemożliwe.

Jednocześnie składam serdeczne podziękowania Panu prof. dr hab. n. med. Michałowi Jeleniowi za wszystkie cenne wskazówki dotyczące diagnostyki cytologicznej.

Na koniec, chciałbym serdecznie podziękować mojej rodzinie i przyjaciołom, a w szczególności moim Rodzicom, Lucynie oraz Zbigniewowi Kołtun, dzięki którym miałem możliwość kształcić się i zdobywać wiedzę. Wasze wsparcie, cenne rady i niekończąca się motywacja były dla mnie nieocenione przez cały okres trwania studiów na Politechnice Wrocławskiej.

Spis treści

1	Cel i zakres pracy	1
1.1	Sformułowanie problemu	1
1.2	Zakres pracy dyplomowej	1
1.3	Struktura pracy dyplomowej	2
2	Wstęp do diagnostyki cytologicznej	3
2.1	Czym jest diagnostyka cytologiczna?	3
2.2	Historia	4
2.3	Metody pobierania materiału cytologicznego	5
2.4	Diagnostyka raka szyjki macicy	7
2.4.1	Wstęp	7
2.4.2	Metody badań	8
2.4.3	Systemy klasyfikacji	11
3	Wstęp do uczenia maszynowego	15
3.1	Sieci neuronowe	16
3.1.1	Uczenie głębokie	17
3.1.2	Konwolucyjne sieci neuronowe	17
4	Prace o podobnej tematyce	19
4.1	Działanie algorytmów klasyfikacji	19
4.2	Przegląd metod wykorzystujących klasyczne uczenie maszynowe	20
4.3	Przegląd metod wykorzystujących uczenie głębokie	20
5	Materiały	23
5.1	Dane	23
5.2	System rejestracji obrazów w oparciu o dane IoT	26
6	Metody	31
6.1	Architektury sieci neuronowych wykorzystane w pracy	31
6.1.1	VGG16	31
6.1.2	ResNet-50	32
6.1.3	Vision Transformer	32
6.2	Wstępna obróbka oraz podział danych	34
6.3	Schematy modeli	35
6.3.1	Własna sieć CNN	35

6.3.2	VGG-16	36
6.3.3	ResNet-50	37
6.3.4	ViT	39
6.4	Testowanie	41
7	Wyniki	45
7.1	Założenia eksperymentu oraz trening	45
7.2	Rezultaty i analiza eksperymentu	45
7.2.1	Wydajność klasyfikacji modeli na zestawach treningowych i walidacyjnych podczas uczenia	45
7.2.2	Wydajność klasyfikacji modeli na zestawie testowym	50
7.3	Dyskusja oraz wnioski	51
7.4	Działanie systemu rejestracji obrazów w oparciu o dane IoT	52
8	Podsumowanie	55

1. Cel i zakres pracy

Cytologia jest badaniem, którego celem jest wykrywanie raka szyjki macicy - jednego z najczęstszych rodzajów nowotworów, które występują u kobiet. W tym celu pobierane są wymazy komórkowe, które następnie oceniane są przez patomorfologów przy wykorzystaniu mikroskopów oraz prostych systemów skринingowania medycznego. Jest to jednak pracochłonny proces, który wymaga dużego doświadczenia. Dla zapewnienia odpowiedniej jakości musi być wykonywana jedynie przez odpowiednio wykwalifikowanych specjalistów. Dlatego bardzo cenne byłoby zautomatyzowanie tego procesu, by zredukować wysoki nakład pracy (w przypadku zautomatyzowanego systemu lekarze mogliby tylko potwierdzać diagnozę) oraz zwiększyć dokładność opisywania komórek. Działania mające na celu usprawnienie tych procesów to np. detekcja, lokalizacja oraz klasyfikacja komórek na obrazach wymazów komórkowych. Są to przydatne techniki, które pomagają np. w zakresie wczesnej diagnostyki nowotworów. W ostatnich latach obrazowanie przy wykorzystaniu technik sieci głębokich takich jak np. sieci konwolucyjne poczyniło znaczne postępy osiągając znacznie lepsze wyniki w porównaniu z tradycyjnymi metodami uczenia maszynowego. Dużą zaletą wykorzystania sieci neuronowych jest również brak potrzeby wiedzy specjalistycznej dotyczących wyodrębniania cech komórek, sieć samodzielnie ocenia przygotowane obrazy [6].

1.1. Sformułowanie problemu

Głównym celem pracy było przeprowadzenie badań porównawczych algorytmów uczenia głębokiego wykorzystywanych w diagnostyce nowotworowej w celu klasyfikacji obrazów mikroskopowych. Działanie algorytmów zostało sprawdzone na niedużym zbiorze danych (918 zdjęć) zawierającym obrazy preparatów cytologicznych, co pozwoliło zbadać wydajność algorytmu na małym zbiorze danych. W pracy zostały przeanalizowane różne rodzaje sieci, które można wykorzystać do rozpoznawania nowotworów na podstawie informacji obrazowej. Ważnym założeniem było również opracowanie systemu rejestrującego obrazy (przy pomocy urządzeń wykorzystujących IoT) montowanego na dowolnym mikroskopie optycznym, co pozwoliło przeanalizować skuteczność opracowanych algorytmów na własnych danych.

1.2. Zakres pracy dyplomowej

W celu realizacji założeń pracy podjęto następujące działania:

- Opracowania literaturowe na temat diagnostyki cytologicznej oraz uczenia głębokiego
- Przeprowadzenie badań literaturowych na temat wykorzystania uczenia głębokiego w diagnostyce cytologicznej
- Wybór metod uczenia głębokiego do implementacji

- Implementacja 3 algorytmów sieci neuronowych: VGG-16, ResNet-50 oraz Vision Transformer (ViT)
- Opracowanie procedury procesowania danych, testów oraz wizualizacji wyników
- Przeprowadzenie badań zaimplementowanych metod na udostępnionym zbiorze danych
- Skonstruowanie toru wizyjnego do pozyskania obrazów z mikroskopu optycznego
- Analiza skuteczności opracowanych algorytmów na danych pozyskanych z mikroskopu
- Sformułowanie wniosków oraz podsumowania na podstawie szczegółowych wyników badań oraz doświadczeń

1.3. Struktura pracy dyplomowej

Praca dzieli się na 8 rozdziałów: Rozdział 1 dotyczy wprowadzenia do pracy oraz ogólnego przedstawienia celu, zakresu oraz struktury pracy. Rozdział 2 obejmuje opis zagadnień diagnostyki cytologicznej, skupiając się przede wszystkim na opisie diagnostyki raka szyjki macicy. Kolejny rozdział 3 zawiera teoretyczne omówienie uczenia maszynowego, w szczególności skupiając się na uczeniu głębokim oraz sieciach konwolucyjnych. Rozdział 4 przedstawia przegląd prac o podobnej tematyce oraz wyjaśnia zasadę działania algorytmów klasyfikacji. Rozdział 5 prezentuje dane wykorzystywane w pracy oraz zawiera opis systemu rejestrującego obrazy na potrzeby pracy. Rozdział 6 objaśnia architektury sieci wykorzystane w klasyfikacji obrazów, wyjaśnia w jaki sposób sieci te zostały dostosowane na potrzeby pracy oraz przedstawia metryki wykorzystywane do oceny wydajności tych algorytmów. Rozdział 7 dotyczy przeprowadzonych eksperymentów, analizy wyników, prezentacji działania systemu IoT oraz zawiera wnioski z dyskusji nad rezultatami. Ostatni rozdział 8 podsumowuje wnioski z wszystkich etapów pracy oraz wskazuje na działania, które mogą zostać poczynione w celu dalszego rozwoju pracy.

2. Wstęp do diagnostyki cytologicznej

2.1. Czym jest diagnostyka cytologiczna?

Diagnostyka cytologiczna (cytodiagnostyka) jest metodą badania komórek (pochodzących z tkanek lub z płynów ustrojowych organizmu) przy użyciu mikroskopu. Celem badania jest identyfikacja zmian patologicznych w komórkach, które mogą wskazywać na choroby nowotworowe lub stan zapalny. W przypadku wykrycia zmiany nowotworowej, lekarz może zalecić dalsze badania w celu postawienia dokładnej diagnozy i zaplanowania odpowiedniego leczenia.

Zamiarem tego typu badań jest wykrycie zmian przednowotworowych nowotworów we wczesnej fazie (przed wystąpieniem objawów) oraz kontrola stanu zdrowia populacji. W trakcie leczenia może być stosowana również jako metoda monitorowania postępów leczenia lub wczesnego wykrywania nawrotów choroby. Warto podkreślić, że wynik badania cytologicznego nie stanowi rozpoznania. Jednak jako podstawa dla dalszej diagnostyki wykorzystywana jest biopsja aspiracyjna cienkoigłowa (która została opisane w podrozdziale: Metody pobierania materiału cytologicznego) [25].

Cytopatologia jest działem patomorfologii, wobec czego pobrany do badań materiał jest oceniany przez odpowiednio wykwalifikowanych patomorfologów. Cytologia zakłada, że wygląd, charakterystyka komórek odzwierciedla biologiczną specyfikę (kondycję) badanego narządu lub niekiedy całego ciała (zależnie od zakresu badania). Ocena mikroskopowa zawiera analizę elementów takich jak: [25]

- skład komórkowy próbek
- znaki szczególne pojedynczych komórek
- sposób zgrupowania komórek (wzór)
- obecność innych elementów np. ciał obcych, bakterii

Najważniejsze zalety wykorzystania diagnostyki cytologicznej to: [1]

- Bezpieczeństwo - procedury, które są stosowane w celu uzyskania próbek cytologicznych są niezwykle bezpieczne. Powikłania zdarzają się bardzo rzadko, a jeśli już wystąpią, to są stosunkowo łagodne.
- Prostota - uzyskanie większości próbek cytologicznych jest proste w porównaniu z innymi metodami. Obecnie większość instytucji przeprowadzających badania bardzo dobrze znają technologię pracy diagnostycznej.
- Szybkość - wyniki mogą być udzielone niemal natychmiast (w czasie procedury), lub w ciągu kilku godzin (przy zastosowaniu metody Diff-Quik).

- Niski koszt - w porównaniu z biopsją chirurgiczną koszt wykonania badania cytologicznego jest znacznie niższy.
- Wydajność - wysokie wyniki wykrywania zmian nowotworowych (około 80% skuteczności metody).

Wady: [48]

- Niska czułość - niektóre badania cytologiczne mają niską czułość, co oznacza, że mogą pomijać niektóre przypadki choroby.
- Fałszywie ujemne wyniki - w niektórych przypadkach badanie cytologiczne może dać wynik fałszywie ujemny, co oznacza, że wynik jest negatywny, ale choroba występuje.
- Konieczność powtórzenia badania - w przypadku niejednoznacznych lub negatywnych wyników konieczne może być powtórzenie badania, co zwiększa koszty i opóźnia diagnozę.
- Skomplikowane wyniki - wyniki mogą być trudne do interpretacji, co może prowadzić do błędów i opóźnień w leczeniu.

2.2. Historia

Historia diagnostyki cytologicznej rozpoczęła się w XVIII i XIX wieku, kiedy podejmowane były pierwsze próby obserwacji i identyfikacji komórek pod mikroskopem. Ważnym wydarzeniem w rozwoju cytologii było wydanie w 1928 roku pierwszej pracy na temat cytologicznej diagnostyki raka szyjki macicy przez greckiego patologa George Papanicolaou. Jednak pomysł spotkał się ze sceptycyzmem i oporem ze strony innych naukowców oraz środowiska lekarskiego. Kolejną pracą była wydana w 1943 roku przez Papanicolaou wraz z Herbertem Trautem monografia „*Diagnosis of Uterine Cancer by the Vaginal Smear*”, która stała się podstawą nowoczesnej cytopatologii. W książce omówione zostało m.in. przygotowanie wymazów do badania (szyjki macicy i pochwy), opisane zostały zmiany obserwowane w przypadku obecności raka, czy ogólne zmiany cytologiczne w trakcie cyklu miesięczkowego [1, 39, 45].

Co więcej bardzo szczegółowo opisali sposób barwienia i utrwalania preparatów cytologicznych i podali szereg kryteriów rozpoznawczych dla wykrywania komórek nowotworowych. Cała procedura postępowania i identyfikacji opisana w publikacji została nazwana testem Papanicolaou (krócej testem Pap), który obecnie jest powszechnie znany jako badanie cytologiczne, które jest stosowane do wykrywania i zapobiegania rakowi szyjki macicy i innym chorobom żeńskiego układu rozrodczego. Bardzo istotną uwagę przekazaną przez autorów było zastrzeżenie, że cytodiagnostyka nie stanowi metody ostatecznie ustalającej rozpoznanie raka szyjki macicy. Jednak metoda powinna być stosowana jako badanie wstępne i sortujące, które może zasygnalizować występowanie niepożądanych zmian, którą muszą być następnie potwierdzone badaniem mikroskopowym tkanki. Jest to bardzo ważna idea, która jest wciąż aktualna [45]. Wykorzystanie testu Pap w diagnostyce cytologicznej w USA pozwoliło w latach 1950 - 1970 obniżyć zachorowalność nawet o około 50%, a wskaźnik zgonów spowodowanych tego rodzaju nowotworem o około 70%. Pomimo wielu ulepszeń na przestrzeni lat i niejasności dotyczących faktycznego odkrywcy metody procedura wciąż nazywana jest jako „Pap test” [45].

W latach 50. XX wieku cytologia zaczęła być szeroko stosowana, a w kolejnych latach metodykę zaczęto stosować, także w diagnostyce raka piersi i tarczycy. W 1958 roku wprowadzono do diagnostyki cytologicznej badanie płynów ustrojowych, takich jak mocz czy płyn mózgowo-rdzeniowy, a w latach 60. i 70. cytologię narządów wewnętrznych, takich jak płuca, trzustka czy wątroba. W kolejnych dziesięcioleciach rozwój diagnostyki cytologicznej był znaczny. Wprowadzono nowe techniki badawcze, takie jak cytometria przepływowa czy analiza obrazu cyfrowego. Dzięki temu możliwe stało się szybkie i dokładne rozpoznawanie chorób onkologicznych, a także innych chorób, takich jak choroby zakaźne czy autoimmunologiczne.

Dopiero w 1989 roku cytopatologia przeszła proces standaryzacji i w tym roku powstała również oficjalna instytucja przyznająca uprawnienia w dziedzinie cytologii (American Board of Examination Cytopathology). Obecnie diagnostyka cytologiczna jest nieodłączną częścią diagnostyki laboratoryjnej, a jej znaczenie w medycynie jest nieocenione [1].

2.3. Metody pobierania materiału cytologicznego

Materiałem cytologicznym nazywa się pobrany od pacjenta płyn z jam ciała lub guza. Materiał przygotowywany jest przez lekarza klinicystę, radiologa lub patomorfologa. Istnieje wiele różnych metod pobierania materiałów cytologicznych, w zależności od rodzaju wykonywanego badania oraz narządu, który poddawany jest badaniu. Poniżej przedstawiono najczęściej stosowane metody: [1, 10]

- Rozmaz (lub wymaz) - metoda polegająca na pobraniu komórek szczoteczką cytologiczną, a następnie rozmazaniu na szkiełku mikroskopowym. Rozmaz jest najczęściej stosowany w badaniach ginekologicznych, takich jak cytologia szyjki macicy.
- Aspiracja cienkoigłowa - metoda polegająca na pobraniu próbki tkanek lub płynu za pomocą cienkiej igły i strzykawki z guzów w każdej lokalizacji. Aspiracja cienkoigłowa jest stosowana w badaniach takich jak biopsja tarczycy, biopsja piersi, czy pobranie płynu z opłucnej.
- Pobranie wydzieliny - metoda polegająca na pobraniu wydzieliny z narządu, takiej jak np. płwocina, czy wydzielina z szyjki macicy. Pobranie wydzieliny umożliwia analizę komórek w płynach i wydzielinach występujących w narządach.

Wybór odpowiedniej metody pobrania materiału cytologicznego zależy od rodzaju badania oraz narządu, który jest badany. Ważne jest, aby pobranie materiału było bezpieczne i minimalnie inwazyjne dla pacjenta.

Preparat cytologiczny jest to materiał cytologiczny (próbka komórek lub tkanek pobrana w celu diagnostyki cytologicznej) umieszczony na szkiełku podstawowym, utrwalony i zabarwiony. W celu ułatwienia identyfikacji pacjentów na początku badania preparat oznacza się trwale numerem, dzięki któremu łatwo zidentyfikować osobę, od której próbka pochodzi [10].

Do wykonania badania potrzebna jest nieduża próbka, która występuje w dwóch postaciach: [28]

- rozmazu cytologicznego - materiał cytologiczny rozprowadzony i utrwalony na szkiełku

- cytobloku - materiał cytologiczny utrwalony, a następnie zatopiony w bloczku parafinowym (używane np. w diagnostyce nowotworów płuc)

Aby przeprowadzić analizę mikroskopową pobranych komórek lub tkanek konieczne jest odpowiednie przygotowanie preparatu cytologicznego, które zazwyczaj obejmuje następujące kroki: [10, 28]

- Pobranie próbki - metoda pobrania zależy od rodzaju badania oraz narządu, który jest badany.
- Przygotowanie rozmazu (lub cytobloku)
- Utrwalenie preparatu - w celu zabezpieczenia materiału przed zniszczeniem należy utrwalić próbkę z wykorzystaniem środków takich jak np. alkohol.
- Barwienie preparatu - aby zwiększyć widoczność oraz kontrast komórek korzysta się z środków barwiących. Najczęściej stosowanym barwieniem jest wykorzystanie hematoksyliny i eozyny (HE), ale korzysta się również z barwienia Papanicolaou, Diff-quick lub May-Grünwald-Giemsa (MGG).
- Analiza mikroskopowa - przygotowany preparat jest badany pod mikroskopem przez patologa w celu zidentyfikowania i zdiagnozowania wszelkich nieprawidłowości.

Warto odnotować, że konkretne kroki i techniki stosowane podczas przygotowywania preparatu cytologicznego mogą się różnić w zależności od rodzaju próbki i celu badania.

W zależności od sposobu pozyskiwania materiału do badania cytologię dzieli się na cytologię złuszczeniową i biopsję aspiracyjną cienkoigłową.

Cytologia złuszczeniowa, polega na pobraniu i ocenie mikroskopowej próbek komórek z powierzchni narządów (lub tkanek), które uległy samoistnemu złuszczeniu (nazywane złuszczeniem spontanicznym). Do tej kategorii należą cytologiczne badania przesiewowe (screeningowych), czyli profilaktyczne badania dla osób zdrowych, jednak znajdujących się w podwyższonej grupie ryzyka konkretnej choroby nowotworowej.

Obejmuje materiał cytologiczny pobrany z: [1, 28]

- wydzielin i wydalin,
- płynów z jam ciała,
- popłuczyn i wymazów szczoteczkowych

Jako przykład spontanicznego złuszczenia można podać pobieranie komórek jamy opłucnej, które są usuwane przez organizm do płynu opłucnowego. Płyn ten uzyskuje się w czasie punkcji opłucnej (czyli wprowadzenia do jamy opłucnej igły) [38].

Uzyskanie materiału jest możliwe także w sposób ręczny - poprzez zeszkrobanie (lub zeszcotkowanie) komórek z powierzchni tkanek (nazywane złuszczeniem mechanicznym). Przykładem tego rodzaju badania jest opisywane wcześniej badanie Papanicolaou, które stosuje się do badania komórek szyjki macicy u kobiet. W tym badaniu, lekarz zeszkrobuje próbkę komórek z szyjki macicy za pomocą szczoteczki. Cytologia złuszczeniowa jest stosowana

głównie do wykonywania cytologii: szyjki macicy, układu moczowego, płynów z jam ciała czy układu oddechowego [28].

Biopsja aspiracyjna cienkoigłowa (z ang. FNAC - fine-needle aspiration cytology) polega na pobraniu próbki komórek lub tkanek za pomocą cienkiej igły. Do wykonania nakłucia wybiera się miejsca podejrzanego o zmiany chorobowe. W tym rodzaju badania patolog ingeruje w ciało w celu pobrania próbki materiału komórkowego (w odróżnieniu od cytologii złuszczeniowej) [28].

Wyróżnia się następujące rodzaje biopsji cienkoigłowej: [28]

- z aspiracją - do nakłucia wykorzystuje się igłę iniekcyjną (jałową) z podłączoną strzykawką, którą wykorzystuje się do pobrania materiału komórkowego w celu zwiększenia wydajności.
- bez aspiracji - wykonuje się kilkukrotne nakłucie zmiany igłą iniekcyjną (jałową).

Uzyskany z biopsji materiał rozprawdza się na szkiełku podstawowym, a następnie utrwalany w postaci rozmazu cytologicznego lub cytobloku. Wykorzystanie materiału z biopsji (zwłaszcza w postaci cytobloku), pozwala na wykonanie także dodatkowych badań (np. immunohistochemicznych, czy z zakresu biologii molekularnej). Ważną zaletą opisywanego podejścia jest możliwość ostatecznego rozpoznania bez konieczności stosowania bardziej inwazyjnych metod diagnostycznych [28].

FNAC może być wykonywany pod kontrolą palpacyjną (lekarz może wyczuć miejsce niebezpiecznej zmiany poprzez dotyk) w obszarach, takich jak szyja, tarczycza lub piersi. FNAC często wspomagany jest także poprzez USG lub tomografię komputerową w celu pobrania próbek ze zmian występujących głęboko w ciele, które nie mogą być zlokalizowane przez palpację. Metoda jest szeroko stosowana, ale powodzenie badania zależy w dużej mierze od umiejętności lekarza [38].

Przykładem zastosowania biopsji cienkoigłowej jest badanie tarczycy, w którym lekarz pobiera próbkę komórek tarczycy za pomocą cienkiej igły wprowadzonej bezpośrednio do gruczołu.

2.4. Diagnostyka raka szyjki macicy

2.4.1. Wstęp

Rak szyjki macicy jest jednym z najczęstszych rodzajów nowotworów, które występują u kobiet. Szacuje się, że każdego roku rak szyjki macicy rozpoznawany jest u ponad 500 tysięcy kobiet, natomiast nawet 300 tysięcy umiera z powodu efektów działania nowotworu w organizmie. Główną przyczyną występowania jest zakażenie wirusem brodawczaka ludzkiego (HPV), które przyczynia się do rozwoju nieprawidłowych komórek w szyjce macicy. Bardzo ważnym elementem zapobiegania oraz leczenia tego rodzaju nowotworu jest wykonywanie przesiewowego (skriningowego) badania - cytologii szyjki macicy (nazywane również testem Pap - od nazwiska twórcy George'a Papanicolaou). Badanie to polega na pobraniu komórek z powierzchni szyjki macicy, a następnie zbadaniu ich pod mikroskopem w celu wykrycia wszelkich nieprawidłowości [23]. Badanie cytologiczne szyjki macicy opiera się na założeniu, że komórki nowotworowe tracą właściwości kohezyjne (w porównaniu z normalnymi komórkami),

dzięki czemu łatwiej ulegają przemieszczeniu podczas pobierania wymazu z szyjki macicy. Efektem tej wiedzy jest możliwość wykrycia nawet najmniejszych zmian nowotworowych w łatwy sposób [34].

Co ważne wynik badania cytologicznego nie jest ostatecznym rozpoznaniem, lecz jest podstawą dla dalszej diagnostyki przy wykorzystaniu technik takich jak kolposkopia, biopsja czy analiza histopatologiczna (lub molekularna). Średnia czułość cytologii (czyli zdolność do prawidłowego wykrywania choroby) wynosi około 60–70%, a swoistość (wskaźnik identyfikacji osób zdrowych) nawet do 94%. Fałszywie ujemne wyniki zazwyczaj wynikają z niewielkiej liczby nieprawidłowych komórek, obecności krwi lub komórek zapalnych [25]. Według najnowszych standardów Amerykańskiego Stowarzyszenia do Walki z Rakiem (The American Cancer Society) wykonywanie badań cytologicznych zaleca się z różną częstotliwością w podziale na kilka grup wiekowych: [15, 27]

- Kobiety do 21 roku życia - nie wykonuje się badań, również w przypadku, gdy pacjentka jest aktywna seksualnie przed 21 rokiem życia.
- Kobiety w wieku od 21 do 29 - pierwszy test Pap powinien być w wieku 21 lat, a następnie kontrolnie co 3 lata.
- Kobiety w wieku od 30 do 65 - powinny wykonywać zarówno test Pap, jak i test HPV co pięć lat (lub tylko test HPV również co 5 lat) albo tylko test Pap co 3 lata.
- Kobiety w wieku powyżej 65 lat - jeśli przechodziły regularne badania przesiewowe z prawidłowymi wynikami, nie powinny być badane pod kątem raka szyjki macicy.

Kobiety, które w przeszłości chorowały na raka szyjki macicy lub doświadczyły pewnych problemów zdrowotnych, powinny kontynuować badania zgodnie z zaleceniami lekarza. Istotny dla skuteczności badania jest odpowiedni moment wykonania badania w ciągu cyklu miesięczkowego. Badanie najlepiej przeprowadzić w środku cyklu, jednak nie wcześniej niż 4 dni po ostatnim dniu miesiączki i nie później niż 4 dni przed rozpoczęciem miesiączki (badania nie wykonuje się w trakcie krwawienia miesięczkowego) [25].

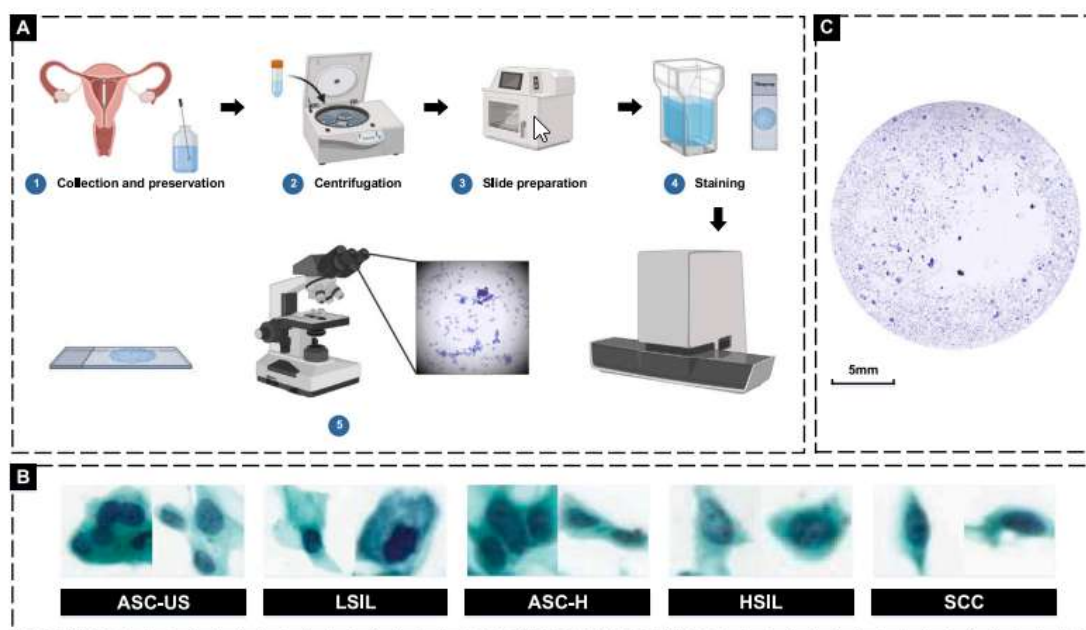
Uzyskiwaniem materiału do cytologii szyjki macicy zajmują się lekarz ginekolog. Pobranie rozmazu przeprowadza się na fotelu ginekologicznym z użyciem wziernika (co zapewnia łatwy dostęp do szyjki macicy). Poprzez wziernik wprowadzana jest specjalna szczoteczka cytologiczna, przy użyciu której pobiera się nabłonek z powierzchni szyjki macicy (pobieranie należy wykonywać delikatnie, ruchem obrotowym). Uzyskiwanie materiału jest praktycznie bezbolesne i jednym skutkiem ubocznym może być krwawienie (plamienie) z szyjki macicy, które wynika ze złuszczenia nabłonka [15, 27].

2.4.2. Metody badań

Cytologia szyjki macicy może być wykonywana przy użyciu różnych technik. Zazwyczaj wyróżnia się dwie metody: [11, 25, 42]

- cytologię konwencjonalną - materiał cytologiczny zebrany przy pomocy szczoteczki lub szpatułki przenoszony jest na szkiełko mikroskopowe tworząc cienką warstwę.

- cytologię na podłożu płynnym (z ang. Liquid-based cytology w skrócie LBC) - inaczej określana również jako cytologia płynna lub cienkowarstwowa. Rysunek 2.1 przedstawia przykładowy proces uzyskiwania preparatu cytologicznego dla metody LBC. W tym przypadku pobranie również odbywa się przy wykorzystaniu szczoteczki lub szpatułki, jednak materiał nie jest od razu przenoszony na szkiełko, tylko najpierw umieszczany jest w pojemniku ze specjalnym płynem, który zapobiega uszkodzeniu komórek. W kolejnym kroku preparat cytologiczny poddawany jest procesowi cytowirowania, filtracji lub sedimentacji, ponieważ komórki są zawieszone w płynie. Efektem jest pozyskanie preparatu zawierającego jedną warstwę komórek pochodzących z osadu rozdzielonego od fazy płynnej, a także oczyszczonego z niepożądanych elementów np. ze śluzu, krwinek czerwonych czy bakterii. Po tym procesie materiał jest gotowy do przeniesienia na szkiełko, procesu barwienia i wreszcie wykonania obserwacji pod mikroskopem. Historia LBC sięga lat 90. XX wieku i została opracowana w celu poprawy efektywności metody konwencjonalnej.



Rysunek 2.1: Przebieg typowego badania - diagnostyki cytologicznej [16]

W ramach LBC rozróżnia się odmienne podejścia do zbierania i obróbki próbek cytologicznych, różniące się m.in. preparatem utrwalającym, narzędziem pobierania czy wielkością próbki na szkiełku. Dwoma najbardziej popularnymi metodami są SurePath i ThinPrep, które zostały zatwierdzone do użytku w USA przez Food and Drug Administration (FDA) i są obecnie najczęściej stosowanymi metodami na świecie. Obydwa systemy są w pełni zautomatyzowane. Porównanie obu metod zostało przedstawione w tabeli 2.1 [42].

Tabela 2.1: Porównanie metod ThinPrep i SurePath [21]

Cechy	ThinPrep	SurePath
Mechanizm działania	Filtr	Cytowirowanie
Utrwalacz	Metanol	Etanol
Obszar próbki	okrąg o średnicy 20 mm	okrąg o średnicy 13 mm
Metoda osadzania	kontrolowane ciśnienie	sedymencja
Cytowirowanie	potrzebne	konieczne
Trwałość preparatu	2 lata	3 lata
Narzędzie pobierania	szczoteczka, szpatułka	szczoteczka
Technika dyspersji	wirowanie	wirowanie oraz przepłukiwanie

Tabela 2.2: Porównanie metod LBC i metody konwencjonalnej [21]

Cechy	LBC	metoda konwencjonalna
Tło próbki: ilość niepożądanych elementów kolor	bardzo niska czyste tło	znaczna ilość nieprzejryste tło
Rozmieszczenie komórek	równomierne	nierównomierne (gęste obszary)
Wielkość komórek	mniejsza	większa
Materiał pozakomórkowy	zredukowany	zachowany
Ilość warstw	pojedyncza	wiele nachodzących na siebie warstw
Transport preparatu	bezpieczny, stabilny	duże ryzyko uszkodzenia materiału
test HPV	możliwe	brak możliwości

Obie metody (metoda LBC i konwencjonalna) są nadal szeroko stosowane w medycynie, porównanie cech obu metod przedstawiono w tabeli 2.2.

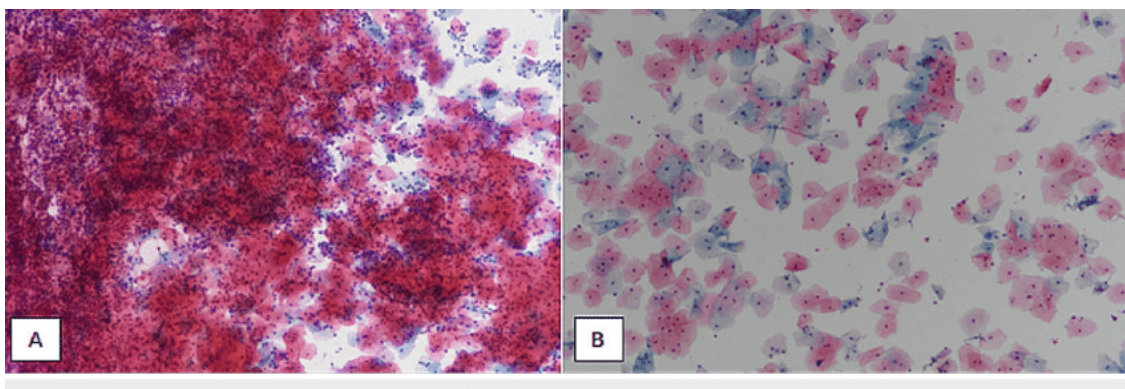
Z tabeli wynika, że metoda LBC zapewnia większą przejrzystość, czytelność próbki (czystsze tło - łatwiej zinterpretować zmiany na próbce). Co więcej próbkę pokrywa pojedyncza warstwa rozproszonych komórek skupionych na jednej powierzchni, czego efektem jest znacznie szybsze rozpoznanie niż w przypadku tradycyjnego rozmazu. Równocześnie metoda LBC jest o wiele bardziej stabilna w transporcie i trwała oraz pozwala na wykonywanie na preparacie także innych badań np. testu HPV. Poza własnościami wyszczególnionymi w tabeli 2.2 LBC zapewnia także inne zalety: [42]

- Preparaty są utrwalane natychmiastowo, dzięki czemu szczegóły budowy jąder komórkowych i cytoplazmy są lepiej uwydatnione
- Z pobranego materiału można sporządzić wiele preparatów
- Znaczna większość preparatów nadaje się do oceny cytologicznej, a każdy do oceny mikroskopowej
- Łatwo zastosować automatyczny skrining

Jednak metoda LBC posiada również wady takie jak: [42]

- Rodzaje rozmazów są odmienne z powodu losowego występowania komórek
- Komórki nieprawidłowe są rozproszone na całej powierzchni próbki
- Rozmazy LBC z małą ilością komórek są trudne do wykonania skriningu i interpretacji
- Elementy niepożądane takie jak: krew, śluz, komórki zapalne, brudne tło (specyficzne dla nowotworu złośliwego) są dalej obecne, jednak mają inny charakter
- Komórki nabłonkowe występują mniej licznie (lub pojedynczo) oraz są nieco mniejsze niż w zwykłych preparatach cytologicznych

W porównaniu z LBC cytologia konwencjonalna jest metodą tanią i znacznie łatwiejszą do wykonania. Dlatego może być stosowana także przez laboratoria o ograniczonych zasobach i zdolnościach. Jednak metoda może być trudniejsza do interpretacji przez patomorfologów ze względu na mniej czytelne obrazy komórek (porównanie wyglądu próbek przedstawiono na rys. 2.2). Cytologia konwencjonalna ma nieco niższą czułość i specyficzność w porównaniu do LBC. Obecnie metoda LBC staje się coraz bardziej popularna, ponieważ jest dużo bardziej wydajna od cytologii konwencjonalnej, pozwala na wykonanie innych testów (przede wszystkim testy na obecność HPV lub testy molekularne) oraz zapewnia lepszą jakość próbek [21, 42].



Rysunek 2.2: Porównanie próbek pozyskanych metodą konwencjonalną (A) oraz LBC (B) [12]

2.4.3. Systemy klasyfikacji

Pierwsza próba klasyfikacji komórek występujących w rozmazie została stworzona przez George'a Papanicolaou (jako elementu całej procedury - testu Papanicolaou). Podział polegał na wyszczególnieniu 5 grup, od obrazów komórek prawidłowych (jako grupy I) do komórek nowotworowych raka szyjki macicy (grupa V). Podczas badania cytologicznego ocenie podlegała grupa, w której znajdują się komórki obecne na rozmazie. Ogromną wadą tej klasyfikacji była informacja jedynie o wyglądzie komórek, nie zawierając żadnych innych przydatnych informacji. Klasyfikacje wg Papanicolaou przedstawiono w tabeli 2.3 [45].

Tabela 2.3: Klasyfikacja mikroskopowa obrazów cytologicznych wg. propozycji Papanicolaou [25, 45]

Grupa	Opis
I	zawiera tylko prawidłowe komórki
II	zawiera prawidłowe komórki oraz komórki zapalne
III	nieprawidłowy rozmaz: podtyp A - nasilony stan zapalny, podtyp B - komórki nabłonka o budowie dysplastycznej wskazujące na CIN (śródnabłonkową neoplazję)
IV	nieprawidłowy rozmaz z nieprawidłowymi komórkami (atypowe), które mogą wskazywać na wczesne stadium nowotworu (bez nacieków na sąsiadujące tkanki)
V	nieprawidłowy rozmaz - liczne atypowe komórki charakterystyczne dla zaawansowanego stadium raka

Obecnie powszechnie stosowaną klasyfikacją, która wyparła procedurę Papanicolaou, jest stworzony w 1988 roku (ulegał również późniejszym modyfikacjom) opisowy system Bethesda (The Bethesda System) opracowany przez amerykański Narodowy Instytut Zdrowia (National Institutes of Health). Motywacją wprowadzenia tego systemu była chęć ulepszenia systemu klasyfikacji Papanicolaou, który okazał się niewystarczająco dokładny wobec postępu badań w zakresie występowania zmian nienowotworowych szyjki macicy. Podstawową niedoskonałość poprzedniego systemu był brak zawarcia zmian nabłonka szyjki macicy powstających pod wpływem infekcji wirusowych (takich jak HPV) oraz zmian w obrębie nabłonka gruczołowego obecnego w kanale szyjki. System Bethesda jest rekomendowany przez Polskie Towarzystwo Ginekologów i Położników dla wykonywania cytologii szyjki macicy już od 2001 roku i jest obecnie najczęściej stosowany w Polsce. Pomimo stosowania klasyfikacji Bethesda badania cytologiczne z szyjki macicy nadal nazywa się testem Pap [25, 37, 40, 45].

Klasyfikacja według systemu Bethesda nie jest oparta na grupach tak jak w oryginalnym badaniu Pap. Zamiast grup wprowadzono opis dotyczący ilości pobranego materiału, jego jakości, oceny zawartości mikroorganizmów w zadanym obszarze, a także sugestii oceniającego dotyczącej np. powtórzenia badania. Każdy uzyskany podczas badania wynik oceniający rozmaz jest sprawdzany przez lekarza według opracowanego schematu, który następnie opisuje komórki i proponuje dalsze zalecenia dla pacjenta np. "powtórne badanie za 12 miesięcy". Ważnym założeniem systemu Bethesda jest także brak wyników pomiędzy np. grupa I/II dla klasyfikacji Papanicolaou [45].

System Bethesda zawiera informacje o typie próbki (cytologia cienkowarstwowa lub konwencjonalna), jakości (liczba i rodzaj komórek, rodzaj środków wykorzystanych do utrwalenia i barwienia preparatu) oraz ogólnej charakterystyce. Jeśli materiał nie spełnia kryteriów diagnostycznych, podawany jest powód niewykonania badania [25].

Nowe ustalenia z zakresu miejsc występowania nowotworu doprowadziły także, do pobierania rozmazu cytologicznego z dwóch miejsc w obrębie szyjki macicy: [45]

- z powierzchni tarczy szyjki (ujścia zewnętrznego szyjki macicy), gdzie znajdują się nabłonek wielowarstwowy płaski

- z kanału szyjki, gdzie znajdują się nabłonek gruczołowy

Zmienione komórki mogą pojawić się w każdym z opisanych rodzajów nabłonka lub obu jednocześnie, co uwzględnia wynik zgodnie ze skalą Bethesda. Najczęstszym nowotworem szyjki macicy jest rak płaskonabłonkowy z nabłonka wielowarstwowego płaskiego, znacznie rzadszy nowotwór nabłonka gruczołowego nazywany jest gruczolakorakiem [29].

Pierwszym aspektem podlegającym ocenie jest jakość pobranego materiału. Diagnosta sprawdza, czy zebrano odpowiednią ilość komórek z każdego nabłonka oraz czy preparat został poprawnie utrwalony i wybarwiony. Następnie przeprowadza się oceny komórek. Jeśli nie stwierdza się żadnych zmian przednowotworowych ani nowotworowych, rozmaz opisuje się skrótem NILM (z ang. negative for intraepithelial lesion or malignancy - negatywny wynik dla zmian śródnabłonkowych lub raka) lub opisowo jako "brak zmian cytoonkologicznych" [29]. W przeciwnym przypadku korzysta się z zapisu: "nieprawidłowy obraz cytoonkologiczny", w połączeniu z którym stosuje się następujące skróty w opisie wyniku: [25, 29]

Nieprawidłowe komórki nabłonka wielowarstwowego płaskiego:

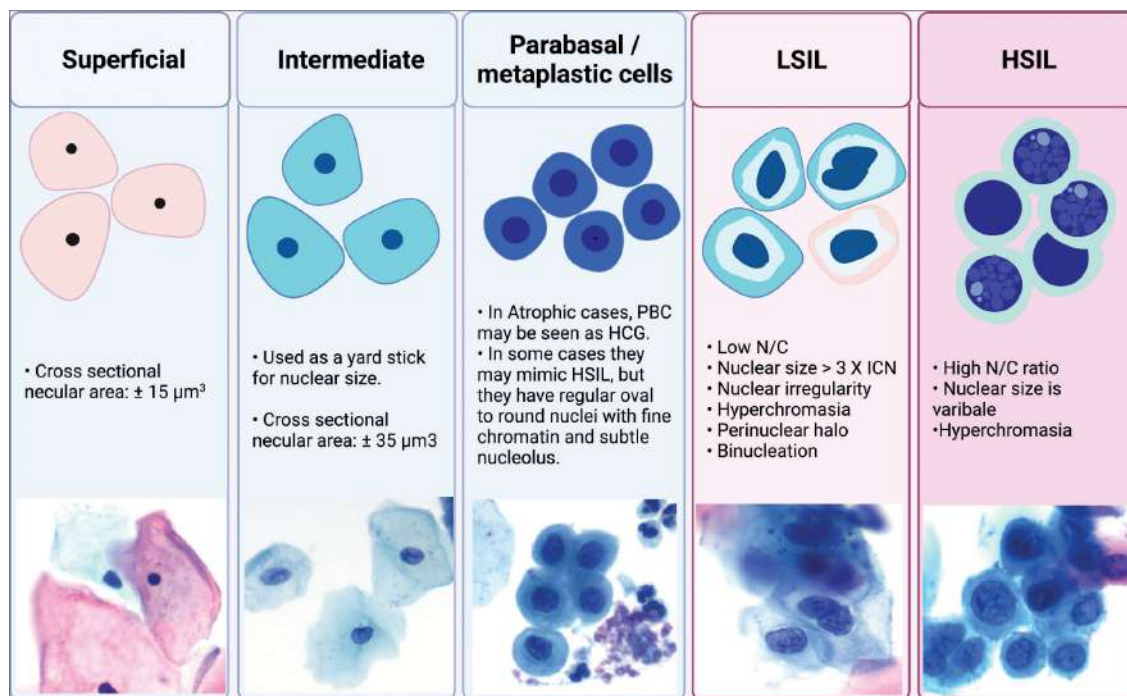
- ASC-US (z ang. atypical squamous cells of undetermined significance) - nieprawidłowe komórki nabłonka wielowarstwowego płaskiego o nieokreślonym znaczeniu, co często świadczy o przewlekłym stanie zapalnym
- ASC-H (z ang. atypical squamous cells, cannot exclude HSIL) - nieprawidłowe komórki nabłonka wielowarstwowego płaskiego, nie można wykluczyć stanu przedrakowego, czyli HSIL
- LSIL (z ang. low-grade squamous intraepithelial lesion) – zmiana śródnabłonkowa niskiego stopnia zaawansowania, może świadczyć o zakażeniu HPV
- HSIL (z ang. high-grade squamous intraepithelial lesion) – zmiana śródnabłonkowa wysokiego stopnia zaawansowania
- CIS (z łac. carcinoma in situ) – rak płaskonabłonkowy in situ czyli „w miejscu”, odpowiada rakowi przedinwazyjnemu

Nieprawidłowe komórki nabłonka gruczołowego:

- AGC (z ang. atypical glandular cells) – nieprawidłowe komórki nabłonka gruczołowego
- AIS (z ang. adenocarcinoma in situ) – rak gruczołowy in situ czyli „w miejscu”, odpowiada rakowi przedinwazyjnemu
- Gruczolakorak - komórki raka gruczołowego szyjki lub trzonu macicy lub nowotworu pozamacicznego

Zgodnie z zaleceniami The College of American Pathologists and American Society for Colposcopy and Cervical Pathology (ASCCP), zaleca się stosowanie dwufazowej klasyfikacji zmian SIL (z ang. Squamous Intraepithelial Lesion). Wyróżnia się dwa stopnie zależne od nasilenia zmian komórkowych i ryzyka progresji nowotworu: LSIL oraz HSIL. Kategorie

LSIL od HSIL różni się na podstawie oceny ilości komórek patologicznych oraz atypii jądra komórkowego. W przypadku gdy ilość nieprawidłowych komórek lub cechy patologiczne są niewystarczające, stosuje się również opisane wcześniej stopnie pośrednie, tj. ASC-US lub ASC-H [26]. Wygląd oraz opis cech LSIL i HSIL przedstawiono na rysunku 2.3. Do odróżnienia komórek LSIL od HSIL wykorzystuje się relację N:C (z ang. nucleus:cytoplasm - jądro:cytoplazma). W LSIL jądra mogą być również znacznie powiększone i pleomorficzne (o bardziej zróżnicowanym kształcie), lecz charakteryzują się niskim stosunkiem N:C, podczas gdy jądra w HSIL są bardziej jednolite i posiadają wysoki stosunek N:C. LSIL ma także tendencję do zajmowania powierzchniowych warstw błony śluzowej, podczas gdy komórki o komórki HSIL rozciągają się na co najmniej jedną trzecią grubości błony śluzowej. W niektórych przypadkach nie można określić czy komórki są LSIL lub HSIL. W takim przypadku zmiany charakteryzują się jako SIL o nieokreślonym stopniu zaawansowania [5].



Rysunek 2.3: Wygląd oraz opis cech komórek szyjki macicy [3]

3. Wstęp do uczenia maszynowego

Uczenie maszynowe (z ang. Machine Learning (ML)) jest działem sztucznej inteligencji (z ang. Artificial Intelligence (AI)), który wykorzystuje algorytmy obliczeniowe do stworzenia modelu matematycznego zdolnego do wykonywania użytecznych zadań (np. przewidywanie wartości) bez dokładnych zewnętrznych wskazówek.

Algorytmy te wykorzystują istniejące dane (treningowe) do opracowania modelu podczas procesu nazywanego trenowaniem, następnie są dostosowywane przez algorytmy optymalizacji w celu uzyskania dokładnych predykcji. Celem modeli jest uzyskanie generalizacji, która pozwoli na dokładną predykcję również dla nowych, niedostępnych wcześniej dla modelu danych. Zdolność do generalizacji modelu jest oceniana przy użyciu nowych danych nazywanych zbiorem walidacyjnym, które są wykorzystywane jako informacja zwrotna dla dalszego dostosowywania modelu. Operacje trenowania i dostosowywania algorytmu są wykonywane w powtarzalnych iteracjach, które kończą się ewaluacją - model oceniany jest na nowym zbiorze danych (testowym), który symuluje zachowanie modelu w przypadku predykcji nowych danych. Jeśli skuteczność algorytmu podczas testowania zestawu danych wzrasta w czasie trenowania wtedy taki algorytm "uczy się" zyskując informację o zbiorze danych. Ważnym wymogiem dla uzyskania skutecznego algorytmu ML jest duża ilość danych, tak by algorytm miał wystarczająco dużo informacji do nauczenia się z generalnej reprezentacji danych.

Algorytmy uczenia maszynowego są przydatne w wielu zadaniach, np. w klasyfikacji, grupowaniu danych czy regresji. Klasyfikacja służy do oznaczania różnych kategorii danych, np. do oddzielania obrazów różnych obiektów (kół od prostokątów). Grupowanie służy do zebrania danych w pewnych podzbiorach danych o podobnych cechach, przykładem mogą być systemy rekomendacji np. filmów, książek na podstawie ocen użytkowników. Natomiast regresja dotyczy przewidywania dokładnej wartości liczbowej (ciągłej, nie dyskretnej - przeciwnie do klasyfikacji) na podstawie danych wejściowych (z których przynajmniej część jest ciągła) [46]. Problemy uczenia maszynowego można podzielić na dwie najważniejsze grupy [46]:

- uczenie nadzorowane - algorytm ma dostęp zarówno do danych wejściowych jak i do oczekiwanych odpowiedzi. Oznacza to, że dane zostały wcześniej skategoryzowane (najczęściej przez człowieka). Przykładem takiego zbioru mogą być zdjęcia ubrań, z których każde jest podpisane według przyjętych nazw etykiet np. spodnie, bluza, koszula. Celem uczenia jest w tym przypadku znalezienie skutecznego modelu przewidującego odpowiedzi na podstawie dostępnych danych.
- uczenie nienadzorowane - w przeciwieństwie do uczenia nadzorowanego algorytm otrzymuje tylko dane wejściowe, nie zapewniając poprawnych danych wyjściowych. W tym przypadku celem uczenia jest znalezienie zależności na podstawie jedynie danych wejściowych. Dla przykładu może być to znów zbiór zdjęć ubrań, który nie jest w żaden sposób opisany tj. nie zawiera żadnych dodatkowych informacji

W ostatnich latach nastąpił znaczny postęp w dziedzinie uczenia maszynowego, znajdując

zastosowanie w wielu dziedzinach np. medycynie, przemyśle. Postęp spowodowany jest przede wszystkim przełomem w jednej z wielu poddziedzin uczenia maszynowego, czyli w sztucznych sieciach neuronowych (z ang. ANN - artificial neural networks) znanych jako głębokie uczenie (z ang. deep learning), które są zbiorem algorytmów pozwalającym na identyfikowanie skomplikowanych zależności w nawet w ogromnych zbiorach danych. Uczenie maszynowe bardzo zyskało na popularności również ze względu na wzrost dostępności danych oraz wydajności elektroniki, co doprowadziło to do powstania nowych i znacznie ulepszonych algorytmów, które znalazły zastosowanie w nowych zadaniach.

3.1. Sieci neuronowe

Sztuczne sieci neuronowe (z ang. ANN - Artificial Neural Networks) stały się obecnie jednym z najczęściej wykorzystywanych modeli uczenia maszynowego. Charakteryzują się wysoką elastycznością oraz zdolnością do modelowania złożonych problemów. Jednak ich wykorzystanie (zostały opracowane w latach 50. XX. wieku) było historycznie ograniczone ze względu na znaczne zasoby obliczeniowe wymagane do procesu trenowania. Jednak na początku XXI. wieku sytuacja uległa zmianie i sztuczne sieci neuronowe stały się jedną z dominujących metod w dziedzinie uczenia maszynowego.

Sieci te, składają się z połączonych ze sobą "neuronów", czyli podstawowych jednostek obliczeniowych, które ułożone są w warstwy. Strukturę sieci neuronowej można przedstawić jako szereg warstw. Zaczynając od warstwy wejściowej, w której dane wchodzi do sieci, kolejno umieszczone są warstwy ukryte (co najmniej jedna), które odpowiedzialne są za transformację danych przechodzących przez sieć. Przetworzone informacje docierają do warstwy wyjściowej, która dokonuje predykcji na podstawie obliczeń całej sieci neuronowej. Trening sztucznej sieci neuronowej polega na identyfikacji wzorców, które występują w danych treningowych. Podczas tego procesu parametry sieci są dostrajane (dokładniej każdy neuron), aby zapewnić dokładne przewidywania dla danych treningowych, co wykonywane jest przez funkcję celu, która porównuje dane treningowe z rzeczywistymi etykietami danych. Gdy sieć z powodzeniem nauczy się podstawowych wzorców, staje się zdolna do przewidywania nowych, wcześniej nieznanymi danych [16, 22].

Podstawowa struktura sztucznych sieci neuronowych opiera się na sparametryzowanych funkcjach matematycznych,

$$y = f(x; \theta)$$

które mapują dane wejściowe x na wyjściowe y poprzez szereg nieliniowych przekształceń typu:

$$f(x) = (f_n \circ \dots \circ f_1)(x)$$

Każdy komponent f_k , nazywany warstwą sieci składa się z liniowej transformacji wyjścia poprzedniego komponentu, po której korzysta się z funkcji nieliniowej:

$$f_k = \sigma_k(\theta_k^T f_{k-1})$$

Typowe wybory dla tych nieliniowych funkcji aktywacji σ_k to funkcje sigmoidalne oraz jednostki typu ReLU. Parametry θ_k natomiast reprezentują macierze liczbowe, znane jako wagi modelu [22].

Podczas fazy uczenia sieć neuronowa korzysta z etykietowanych danych treningowych, a jej celem jest generowanie przewidywań w ostatniej warstwie (wyjściowej), które są zgodne ze znanymi na początku etykietami. Uczenie sieci polega na optymalizacji wag sieci w celu osiągnięcia jak najbardziej dokładnych predykcji. Proces ten jest ułatwiony dzięki metodom optymalizacji np. gradientu prostego (z ang. gradient descent), który działa na podstawie funkcji kosztu lub funkcji straty stając się miarą poprawności wyników. Podczas trenowania sieć przetwarza dane treningowe oraz wyliczany jest gradient funkcji straty w odniesieniu do każdej wagi przy użyciu wzoru na pochodną funkcji złożonej. Wagi są następnie aktualizowane za pomocą metody gradientu prostego w sposób iteracyjny, mając na celu zminimalizowanie strat i poprawę zdolności predykcyjnych sieci. Jednak w przypadku skomplikowanych sieci zawierających tysiące, a nawet miliony parametrów, wraz z wykładniczą liczbą ścieżek między węzłami sieci, złożoność obliczeniowa mocno rośnie [16, 22].

3.1.1. Uczenie głębokie

Głębokie uczenie w odróżnieniu do tradycyjnego podejścia do uczenia maszynowego znacznie ułatwia proces trenowania modeli, automatycznie ucząc się reprezentacji i cech bezpośrednio z surowych danych. Modele głębokiego uczenia wykorzystują sztuczne sieci neuronowe wraz z warstwami pośrednimi. Wykorzystanie sieci pozwala ominąć czasochłonny i pracochłonny proces ręcznego projektowania uczonych cech powodując, że ekstrakcja cech i uczenie modelu stają się zintegrowanym procesem, co prowadzi do ogólnej poprawy wydajności. Przykładem modelu głębokiego uczenia są np. konwolucyjne sieci neuronowe (z ang. CNN - convolutional neural networks), które znacznie uprościły działania na obrazach np. klasyfikację lub detekcję obiektów. CNN okazały się niezwykle skuteczne w uczeniu się reprezentacji obrazów i zastąpiły ręcznie opracowywane modele uczenia maszynowego.

3.1.2. Konwolucyjne sieci neuronowe

Sieć neuronowa składająca się (w części) z warstw splotowych (konwolucji) nazywana jest konwolucyjną siecią neuronową (CNN). W tego typu warstwach iteracyjnie wykonywane są operacje splotu między danymi wejściowymi (obrazami), a filtrami (kernelami), czyli kwadratowymi oknami o znacznie mniejszych wymiarach od obrazu np. 3×3 , które przechodzą pojedynczo przez piksele obrazu. W każdym takim kroku obliczana jest suma wektorów pomnożonych przez wartości występujących w jądrze oraz wartości pikseli w danym fragmencie obrazu, które poddawane są następnie działaniu funkcji aktywacji. Połączenie tych danych skutkuje utworzeniem wspólnego wyjścia, czyli nowego obrazu zawierającego informacje na temat wyodrębnienia cech wejściowych. Jądra są opisane poprzez wagi, które są aktualizowane stale podczas procesu uczenia CNN, co znacznie zmniejsza zapotrzebowanie na pamięć i czas wymagany do obsługi CNN. Do analizy obrazu można użyć prostych sieci neuronowych jednokierunkowych (z ang. feedforward), jednak jest to wysoce nieefektywne, ponieważ sieć tego typu tworzy połączenia z każdego węzła w jednej warstwie do każdego węzła w następnej, czego skutkiem jest wysoka ilość parametrów do obliczeń. Natomiast CNN bardzo dobrze odwzorowują reprezentację danych przy znacznie mniejszej liczbie połączeń między warstwami. Dane wejściowe w CNN przechodzą przez warstwy sieci, zatrzymując część informacji z warstw poprzednich. CNN korzystają z wielu warstw splotu (z ang. convolution layers) i aktywacji (z

ang. activation layers), czasem połączonych z warstwami łączącymi (z ang. pooling layers). Szkolenie tych sieci odbywa się poprzez wsteczną propagację oraz metodę gradientu prostego, identycznie jak dla zwykłych sieci neuronowych [16, 22].

Poniżej opisane zostały najważniejsze warstwy budujące CNN:

- Konwolucyjna (z ang. convolution) - wykonuje operację splotu na warstwie poprzedniej filtrami o różnych sparametryzowanych charakterystykach, jednak o tych samych, niedużych rozmiarach np. 4x4. Podczas działania warstwy filtry współdzielą wagi, co prowadzi do polepszenia efektów ekstrakcji cech [22].
- Aktywacyjna (z ang. activation) - mapy cech wygenerowane przez warstwę konwolucyjną poddawane są działaniu funkcji aktywacji, co pozwala sieci na aproksymację funkcji nieliniowych. Najczęściej stosowanymi funkcjami aktywacji są jednostki liniowe typu ReLU, ale stosują się również funkcje sigmoidalne lub tanh [44].
- Łącząca (z ang. pooling) - powodują zmniejszenie map funkcji poprzez łączenie obszarów (np. poprzez wykorzystanie funkcji maksymalizujących (z ang. max-pooling) lub usredniających (z ang. average-pooling)), czego efektem jest pojedyncza liczba na wyjściu dla pewnego obszaru obrazu [22].
- Warstwa regularyzacji dropout - polega na losowym usuwaniu neuronów podczas uczenia (uśrednienie sieci), co powoduje zastosowaniem różniących się sieci dla każdej partii danych treningowych. Skutkiem jest optymalizacja struktury oraz wartości wag sieci [22].
- Normalizacji próbek uczących (z ang. batch normalization) - zwykle umieszczane są po warstwach aktywacyjnych, wykonują proces normalizacji mapy aktywacji, dla każdej partii danych treningowych, odejmując wartość średnią, a następnie dzieląc przez odchylenie standardowe. Zastosowanie tej warstwy w znacznym stopniu reguluje sieć oraz przyspiesza proces uczenia sieci [22].
- Spłaszczająca (z ang. flatten) - służy do przekształcania wielowymiarowych danych wejściowych w jednowymiarowy wektor. Warstwa najczęściej wykorzystywana jest do przekształcania macierzy z warstwy łączącej w wektor [4].

Projektowanie architektur CNN, polega na łączeniu warstw mając na celu optymalizację architektury sieci dla konkretnego zadania. Podczas projektowania należy wziąć pod uwagę wiele czynników, w tym zrozumienie zadania, wymagań, ograniczeń czasowych czy optymalnego wykorzystania zasobów sprzętowych.

4. Prace o podobnej tematyce

W ostatnich latach opublikowano wiele prac wykorzystujących uczenie maszynowe (w tym uczenie głębokie) w diagnostyce cytologicznej. Rozwój uczenia maszynowego znacznie przyspieszył powstanie systemów samodzielnie rozpoznających lub wspomagających wykrywanie zmian na poziomie komórkowym (dział ten nazywany jest cytologią obliczeniową). Według [16] w latach 2014-2021 liczba publikacji dotyczących wykorzystania uczenia głębokiego w cytologii obliczeniowej dla najbardziej popularnych serwisów publikacyjnych (Google Scholar, PubMed i arXiv) wzrosła dziesięciokrotnie. Co ważne, największe wzrosty publikacji zostały osiągnięte w dwóch ostatnich latach przeprowadzania badań. Przeanalizowane przez autorów artykuły przedstawiają różne podejścia dotyczące wykorzystania uczenia głębokiego do badań cytologicznych, takich jak np. klasyfikacja między normalnymi, a nieprawidłowymi komórkami [49], lokalizacja oraz identyfikacja komórek w rozmazach cytologicznych [30] oraz segmentacja składników komórkowych [50].

4.1. Działanie algorytmów klasyfikacji

Standardowy algorytm klasyfikacji obrazów rozpoczyna działanie od załadowania obrazu, a następnie wstępnego przetwarzania danych (z ang. pre-processing) poprzez np. usuwanie szumów, skalowanie, rotację czy zmianę kontrastu. Na tym etapie często dodaje się proces augmentacji danych, czyli tworzenia syntetycznych obrazów poprzez zmianę wielkości, odwracanie, obrót obrazów oryginalnych, co skutkuje znacznym zwiększeniem zbioru dostępnych danych treningowych. Kolejny etap obejmuje ekstrakcję cech, która znacząco wpływa na dokładność klasyfikacji. Większość badań wykorzystuje nadzorowane metody uczenia się i powszechnie wykorzystuje sztuczne sieci neuronowe takiej jak: VGG, wstępnie wytrenowane CNN, np. AlexNet i ResNet, oraz wielowarstwowe perceptronowe sieci neuronowe (MLPNN) jako ekstraktory cech. Część autorów stosuje także hybrydowe techniki uczenia się cech, łącząc różne algorytmy uczenia maszynowego z głębokim uczeniem się. Na koniec wyodrębnione wektory cech wprowadzane są do klasyfikatorów. Najbardziej popularne klasyfikatory na tym etapie obejmują support vector machine (SVM), drzewa decyzyjne, głębokie sztuczne sieci neuronowe (ANN), warstwy w pełni połączone (z ang. fully-connected), przekazywanie wartości wag oraz regresję softmax [32]. Rozdział ten zawiera przegląd prac korzystających z uczenia maszynowego do klasyfikacji raka szyjki macicy, podkreślając różne podejścia i techniki stosowane w celu uzyskania dokładnych i wiarygodnych wyników klasyfikacji.

4.2. Przegląd metod wykorzystujących klasyczne uczenie maszynowe

Początkowe badania nad klasyfikacją komórek w cytologii koncentrowały się na tradycyjnych technikach przetwarzania obrazów (nie korzystających z uczenia głębokiego), takich algorytmach jak support vector machine (SVM), k-means, fuzzy clustering. Metody te odniosły pewien sukces, jednak opierały się w dużej mierze na ręcznie tworzonych cechach, co ograniczało ich zdolność do uchwycenia złożonych wzorców i zmian w obrazach komórek. Na przykład Chen i inni [6] zaproponowali półautomatyczny system analizy wymazów z komórek szyjki macicy wykonujący segmentację konturów jąder oraz cytoplazmy, którego celem było sklasyfikowanie czterech różnych typów komórek przy wykorzystaniu metody SVM, która okazała się skuteczna osiągając nawet do 98% dokładności.

4.3. Przegląd metod wykorzystujących uczenie głębokie

Głębokie uczenie przyniosło rewolucyjną zmianę w dziedzinie analizy obrazów medycznych, w tym w klasyfikacji komórek raka szyjki macicy, osiągając wysoką skuteczność. Metody głębokiego uczenia, a w szczególności konwolucyjne sieci neuronowe (CNN), wykazały wysoką wydajność w różnych zadaniach rozpoznawania obrazów. Przeprowadzono liczne badania w celu zbadania i zweryfikowania skuteczności różnych technik zwiększających możliwości uczenia głębokiego, niektóre z rozwiązań zostały przedstawione poniżej:

- Różne architektury sieci CNN

Badacze stosują różnorodne architektury CNN do klasyfikacji komórek raka szyjki macicy. Przeprowadzono szereg eksperymentów w celu porównania wydajności architektur CNN w klasyfikacji obrazów cytologicznych [2, 14, 35]. Na podstawie wyników porównania popularnych architektur np. Alexnet, ResNet, VGG, GoogleNet można stwierdzić, że modele te uzyskały wysoką wydajność w zadaniach klasyfikacji komórek. Jednak wyniki między architekturami mogą znacznie różnić się w zależności od wielkości, jakości danych itp. Dlatego ważne jest, aby podczas eksperymentów korzystać z różnego rodzaju architektur i na podstawie wyników wybrać najbardziej uniwersalną dla zadania metodę.

- Transfer początkowych wartości wag (z ang. transfer learning)

Uczenie sieci neuronowej od podstaw zazwyczaj wymaga dużej ilości danych i mocy obliczeniowej, co skutkuje długim czasem uczenia. Rozwiązaniem tego problemu jest wykorzystanie wstępnie przetrenowanych modeli, w którym wstępnie wytrenowany model na pewnym zbiorze danych np. ImageNet jest wykorzystywany do rozwiązania innego zadania. Transfer learning jest szczególnie przydatny w obrazowaniu medycznym ze względu na trudności w pozyskiwaniu dużych zbiorów danych o wysokiej jakości. Na przykład Hussain i inni [14] porównali 6 różnych pre-trenowanych architektur głębokiego uczenia (Alexnet, VGG-16, VGG-19), ResNet-50, ResNet-101, GoogleNet) do klasyfikacji 4 klas na podstawie wymazów Papanicolaou na różnych zbiorach obrazów. Osiągnęli różne wyniki zależnie od wykorzystanego zbioru danych, choć niektóre architektury (np. Resnet) osiągały znacznie większą średnią dokładność od innych metod.

- Augmentacja danych i inne techniki wstępnego przetwarzania danych

Techniki augmentacji danych takiej jak: obracanie, skalowanie, odwracanie obrazów, stosuje się w celu zwiększenia różnorodności danych treningowych ze względu na ograniczoną dostępność zetykietowanych wymazów. Poprzez generację nowych obrazów, poprawiają się możliwości do generalizacji modeli, osiągając w ten sposób wyższą dokładność przewidywań klas. W tym celu stosuje się także inne techniki np. wydzielanie z oryginalnego obrazu mniejszych pól (z ang. patches). Na przykład Zhang i inni [49] przeprowadzili badania, które wykazały skuteczność zastosowania m.in augmentacji danych, wydzielania mniejszych pól obrazu w zwiększaniu wydajności klasyfikacji modeli opartych na CNN

Podsumowując, wcześniejsze badania wykazały skuteczność metod uczenia maszynowego, w szczególności CNN, w diagnostyce cytologicznej (a w szczególności w klasyfikacji komórek nowotworowych). Zastosowanie różnych architektur CNN, opisanych technik procesowania danych przyczyniło się do poprawy dokładności klasyfikacji. Nadal istnieje jednak pole do dalszych badań i ulepszeń w tej dziedzinie. Jednym z nowych rozwiązań w tej kategorii zadań jest wykorzystanie transformat wizyjnych (z ang. vision transformers - ViT), które zostały zastosowane i opisane po raz pierwszy w 2020 roku przewyższając pod pewnymi warunkami wynikami CNN [9] (szerzej opisane w rozdziale 6.). Porównanie ViT z różnymi architekturami CNN jest jednym z celów tej pracy.

5. Materiały

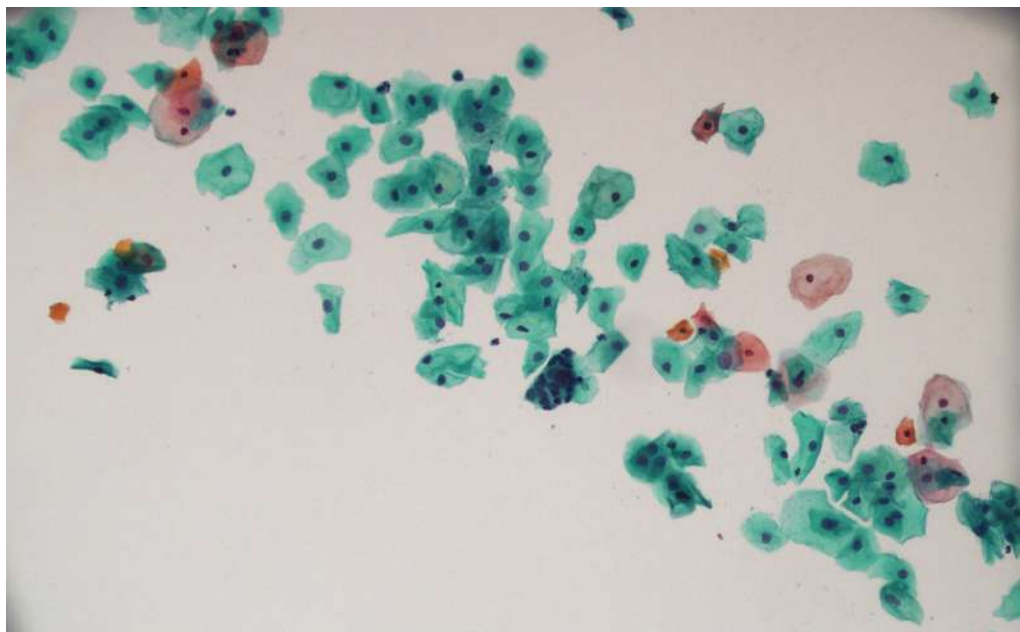
5.1. Dane

Zbiór danych wykorzystanych podczas pracy składał się z 918 zdjęć mikroskopowych preparatów cytologicznych (uzyskanych metodą LBC z tarczy szyjki pochwowej macicy oraz wybarwione metodą Papanicolaou). Obrazy zostały wykonane pod zmiennym powiększeniem (100x dla pełnego obrazu płytek (z ang. WSI - whole slide image) oraz 400x dla pojedynczych komórek) tworząc różnorodny zestaw danych. Wszystkie obrazy zostały zapisane 2 rozdzielczości 2510x1540 pikseli w formacie BMP. Autorem zbioru jest prof. dr hab. n. med. Maria Chosia z Katedry i Zakładu Patomorfologii Pomorskiego Uniwersytetu Medycznego w Szczecinie. Każde zdjęcie zostało poddane ocenie i zakwalifikowane do jednej z trzech kategorii wg zalecanej rekomendacji SIL (co zostało dokładniej opisane w rozdziale 2): LSIL, HSIL oraz komórki prawidłowe opisane jako kategoria NSIL. W ten sposób dla każdej kategorii powstał osobny folder plików z ponumerowanymi według kolejności występowania próbkami, co znacznie ułatwiło pracę z procesowaniem danych. Ilość obrazów dla każdej kategorii została przedstawiona w tabeli poniżej:

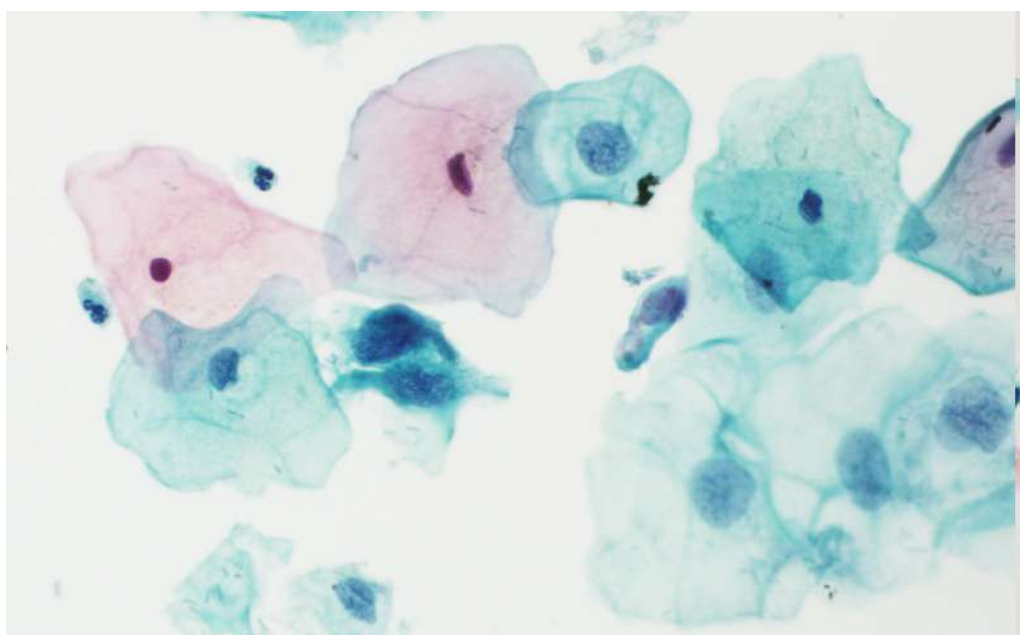
Tabela 5.1: Struktura wykorzystanego zbioru danych

Kategoria	Liczba obrazów
LSIL	210
HSIL	242
NSIL	466

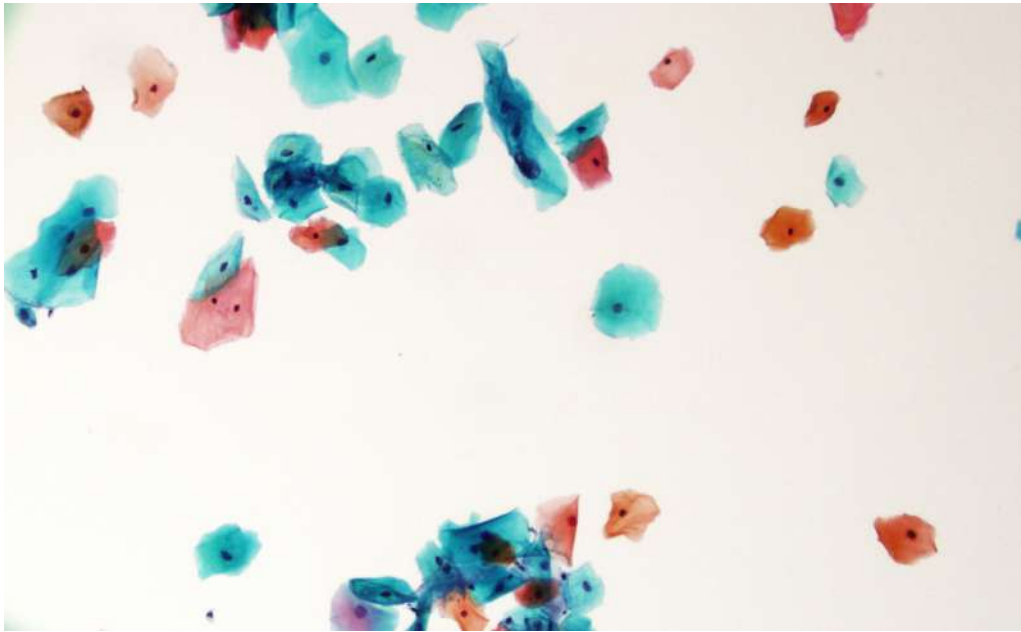
Na poniższych ilustracjach przedstawione zostały przykładowe obrazy komórek z każdej z wymienionych grup w 2 powiększeniach (100x oraz 400x). Rysunki 5.1 oraz 5.2 należą do kategorii HSIL, 5.3 i 5.4 do LSIL, a 5.5 oraz 5.6 to komórki prawidłowe (NSIL).



Rysunek 5.1: Komórki typu HSIL (powiększenie 100x)



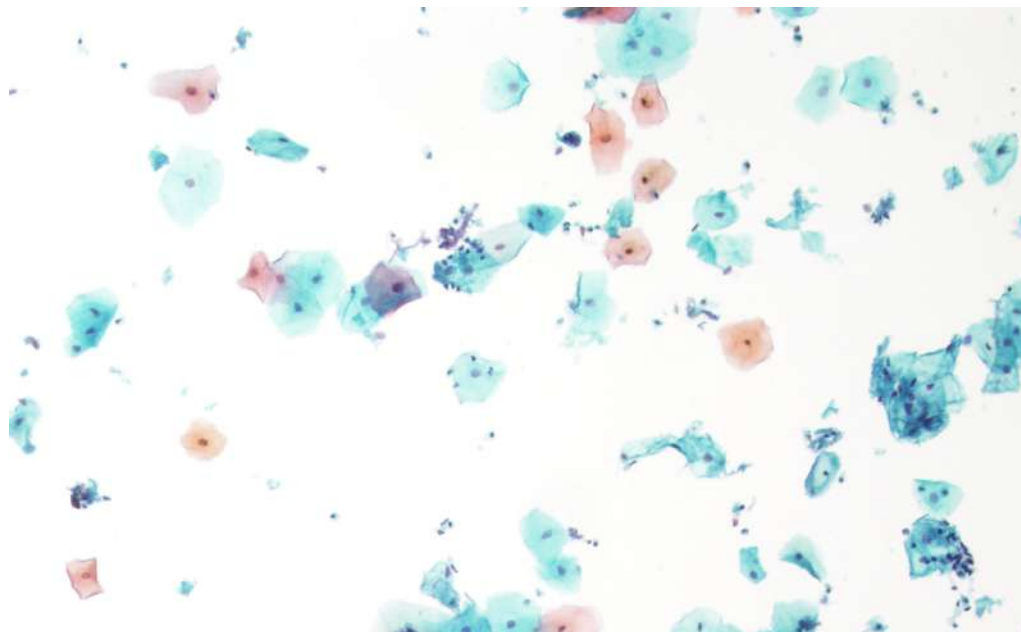
Rysunek 5.2: Komórki typu HSIL (powiększenie 400x)



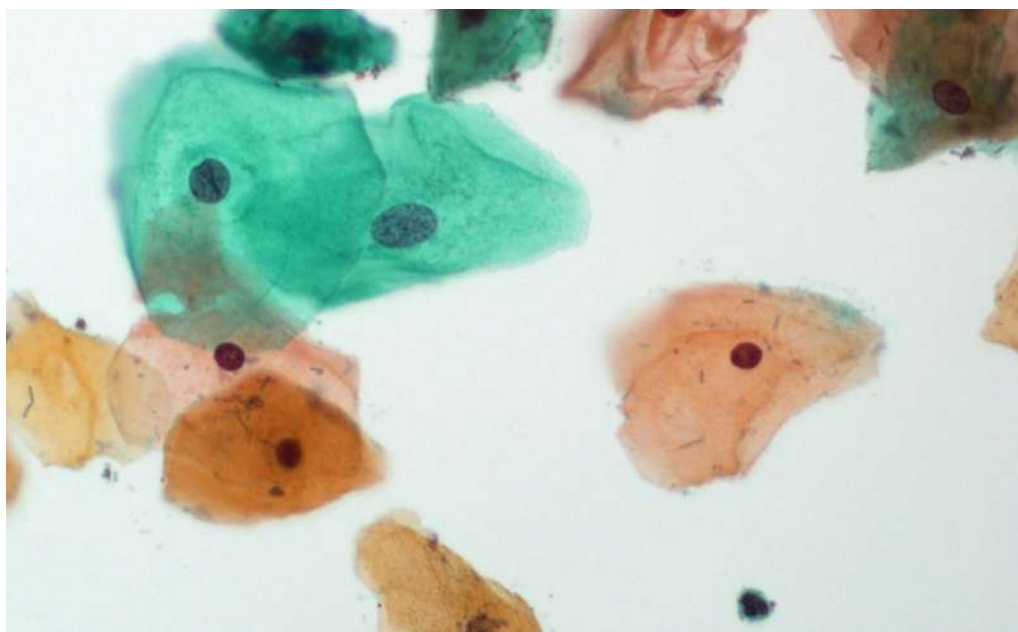
Rysunek 5.3: Komórki typu LSIL (powiększenie 100x)



Rysunek 5.4: Komórki typu LSIL (powiększenie 400x)



Rysunek 5.5: Komórki typu NSIL (powiększenie 100x)



Rysunek 5.6: Komórki typu NSIL (powiększenie 400x)

5.2. System rejestracji obrazów w oparciu o dane IoT

Sprzętem wykorzystanym do budowy systemu rejestrującego obrazy był:

- mikroskop optyczny PZO (przedstawiony na rys.5.7)

- przenośny mini-komputer Raspberry Pi 3 B (przedstawiono na rys.5.8)
- kamera Raspberry Camera Module 3 (przedstawiono na rys.5.8)
- własnoręcznie skonstruowany oraz wydrukowany przez drukarkę 3D adapter na mikroskop (przedstawiona na rys.5.9)
- dioda świetlna wraz z soczewką

Kluczową częścią systemu rejestracji obrazów był mikroskop optyczny, który pozwalał na powiększenie preparatów z wartościami powiększenia 500x oraz 125x. W celu zamontowania kamery na mikroskopie zaprojektowany został adapter dedykowany dla okularu. Założeniem wykonanej części jest uniwersalność oraz łatwy demontaż, dlatego adapter jest jedynie nakładany z góry na okular. Rejestracją obrazu w systemie zajmuje się kamera podłączona do komputera Raspberry, która pozwala na rejestrację z rozdzielczością do 12 Mpx, co pozwoliło na wykonywania zdjęć z rozdzielczością 4608 x 2592 pikseli. Źródłem światła była zwykła dioda LED wraz z soczewką skupiającą. Założeniem działania systemu jest niezależność od innych urządzeń, dlatego jako jednostkę centralną wybrano mini-komputer Raspberry Pi 3 B, który pozwala na instalację bibliotek koniecznych dla modeli uczenia maszynowego w języku Python takich jak np. Tensorflow czy Keras, dzięki czemu łatwo można zaimportować wyuczone modele sieci neuronowych. Inną zaletą Raspberry Pi są małe rozmiary oraz niska waga, dlatego sprzęt ten cechuje się niezwykłą mobilnością. Jest również oparty na Raspberry Pi OS, które jest dystrybucją systemu operacyjnego Linux. W rezultacie pozwoliło to na stworzenie programu (w języku Python), który wykonuje zdjęcie (po wciśnięciu przycisku na klawiaturze), a następnie dokonuje predykcji klas na podstawie zaimportowanego modelu sieci neuronowej. Efekty działania systemu zostały przedstawione w podrozdziale 7.4.

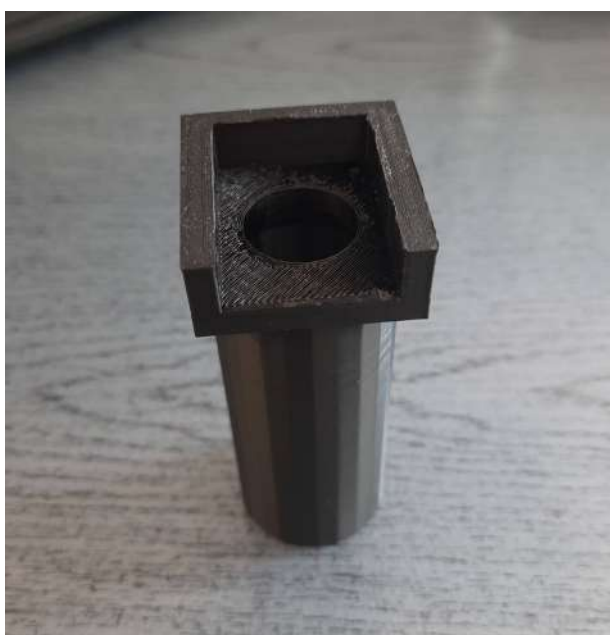
Zdjęcie zmontowanego systemu przedstawiono na rys. 5.10 oraz 5.11.



Rysunek 5.7: Mikroskop optyczny PZO



Rysunek 5.8: Mini-komputer Raspberry Pi 3 B wraz z podłączoną kamerą



Rysunek 5.9: Adapter na mikroskop wydrukowany w 3D



Rysunek 5.10: Zmontowany system rejestracji (widok z góry)



Rysunek 5.11: Zmontowany system rejestracji (widok z boku)

6. Metody

Pierwszym opracowywanym modelem była prosta architektura sieci CNN, która została użyta w celu sprawdzenia cech wykorzystanego zbioru danych, rozstrzygnięcia czy istnieje zagrożenie zjawiskiem nadmiernego dopasowania danych (z ang. overfitting) oraz stanowiła podstawę dla badań porównawczych.

Do opracowywania kolejnych metod postanowiono wykorzystać transfer learning dla 2 wstępnie wytrenowanych (na bazie ImageNet) modeli CNN: VGG16 oraz ResNet-50. Powodem tej decyzji była niewysoka ilość dostępnych danych oraz ograniczone zasoby obliczeniowe. Uczenie w takim modelu polega na dostrojeniu parametrów, czyli dostosowania wartości wag, czy także na stosowaniu niższej szybkości uczenia dla wstępnie wytrenowanych warstw lub dla nowych warstw dodanych do modelu [14, 36]. Natomiast model transformaty wizyjnej (z ang. Vision Transformer) został opracowany od początku na podstawie [9, 17, 18, 19], gdzie kluczowe elementy zostały dostosowane do realizowanego zadania.

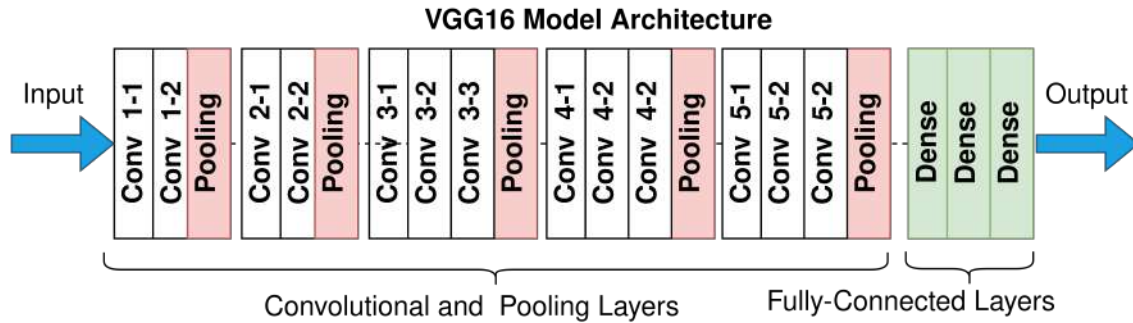
Opracowane modele zostały przeznaczone do klasyfikacji 3 klas: HSIL, LSIL oraz NSIL, dlatego ostatnia warstwa dla wszystkich modeli zawiera 3 wyjścia, które odpowiadają tym klasom. Wszystkie modele zostały zaprogramowane w języku Python 3.10 w środowisku Google Colab.

6.1. Architektury sieci neuronowych wykorzystane w pracy

Na podstawie badań literaturowych oraz doświadczeń (podczas testowania różnych metod) wytypowane zostały 3 perspektywiczne algorytmy uczenia głębokiego: ResNet-50, VGG-16 oraz Vision Transformer, które zostały opisane poniżej.

6.1.1. VGG16

Struktura sieci VGG16 została przedstawiona na rys. 6.1. VGG16 [33] wykorzystuje od 16 do 19 warstw konwolucyjnych, w których używane są najmniejsze filtry konwolucyjne 3x3 z krokiem o wartości jednego piksela, ponadto stosowany jest tutaj filtr konwolucyjny 1x1, który liniowo transformuje dane wejściowe, po czym wywoływana jest funkcja aktywacji ReLU. Po wystąpieniu warstwy splotu dane przechodzą przez pięć warstw łączących typu max-pooling o rozmiarze okna 2x2 i kroku o wartości 2 pikseli. Max-pooling pomaga zmniejszyć wymiarowość i wyodrębnić ważne cechy z obrazu. Na końcu sieci znajdują się trzy warstwy w pełni połączone (fully-connected), a na samym końcu stosowana jest warstwa softmax w celu wygenerowania predykcji prawdopodobieństw dla każdej klasy [32].



Rysunek 6.1: Struktura sieci VGG16 [8]

6.1.2. ResNet-50

Struktura sieci ResNet-50 została przedstawiona na rys. 6.2. ResNet-50 [13] należy do grupy sieci architektury ResNet, która charakteryzują się pomijaniem pewnych połączeń w sieci, co umożliwiło trenowanie głębszych sieci neuronowych i zwiększyło elastyczność sieci. Funkcja ta umożliwia sieci kopiowanie wartości aktywacji z jednego bloku ResNet do drugiego, zachowując oryginalne informacje [32].

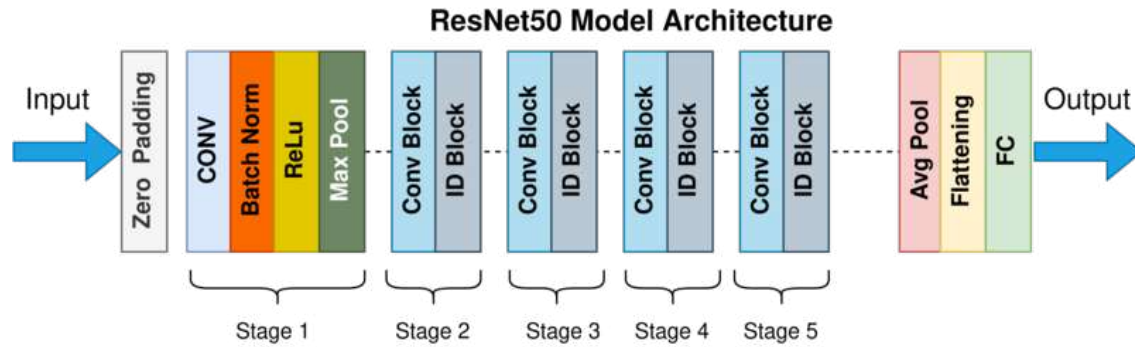
ResNet-50 składa się z 50 warstw, w tym konwolucyjnych, łączących, typu fully-connected oraz z bloków warstw typu residual (która zapewnia wyjątkową funkcjonalność sieci ResNet - pomijanie połączeń).

Architektura zbudowana jest w następujący sposób: [13, 32]

- Warstwa wejściowa: wejściem jest obraz o stałym rozmiarze.
- Warstwy konwolucyjna i łącząca (z wykorzystaniem normalizacji oraz funkcją aktywacji ReLU): Wyodrębniają cechy z obrazu wejściowego. (Stage 1 na rys. 6.2)
- Bloki residual: Składają się z 4 etapów (Stage 2-5 na rys. 6.2) z wieloma blokami residual. Każdy blok łączy dane wejściowe i wyjściowe za pomocą warstw konwolucyjnych, aby nauczyć się mapowania cech z wykorzystaniem warstw residual oraz zoptymalizować działanie sieci.
- Warstwa łącząca typu average pooling: Przekształca dane w jeden wektor.
- Warstwy w pełni połączone: Dwie tego typu warstwy przeprowadzają ostateczną klasyfikację.

6.1.3. Vision Transformer

Vision Transformer (ViT) [9] to model sieci neuronowej stosowany do przetwarzania obrazów, który oparty jest na architekturze transformaty (z ang. transformer) pierwotnie stosowanej do zadań przetwarzania języka naturalnego (z ang. natural language processing - NLP). Obecnie dominującą metodą przetwarzania obrazów są wciąż konwolucyjne sieci neuronowe (CNN), jednak wykorzystanie ViT może doprowadzić do osiągnięcia podobnej, a nawet lepszej wydajności w szybszym czasie poprzez zastosowanie nowej architektury



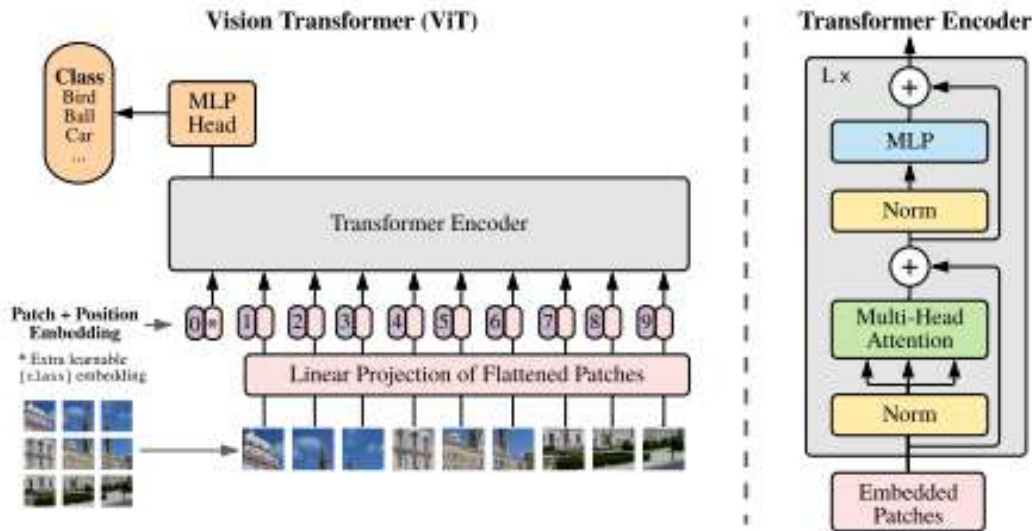
Rysunek 6.2: Struktura sieci ResNet-50 [7]

transformaty wizyjnej. Podstawową koncepcją Vision Transformer jest użycie architektury transformaty do bezpośredniego przetwarzania obrazów wejściowych, bez stosowania warstw konwolucyjnych. W ViT obrazy wejściowe są dzielone na sekwencję pól, które są następnie spłaszczane (transformaty obsługują dane tylko w formie sekwencji) i wprowadzane do modelu. Podstawową jednostką ViT jest blok transformaty, który składa się z warstwy normalizacji, warstwy stosującej wielogłowy mechanizm uwagi (z ang. multi-head self-attention) i warstwy perceptronowej [18, 24, 31, 41].

Natomiast algorytm działa w następujący sposób składając się z kilku etapów [18, 24, 31, 41]:

- Wejście: Obraz jest dzielony na mniejsze fragmenty o jednakowym rozmiarze, zwane polami (z ang. patches). Każde pole jest przekształcane na wektor za pomocą operacji spłaszczenia. Następnie wektory pól są wprowadzane jako sekwencja do modelu.
- Kodowanie pozycji: Do każdego wektora pól dodawany jest wektor kodujący jego pozycję w sekwencji. Kodowanie pomaga modelowi w lokalizacji pól.
- Bloki transformat: Sekwencja pól wraz z kodowaniem pozycji jest poddawana działaniu blokom transformat. Blok transformaty składa się z dwóch głównych warstw:
 - Warstwa wielogłowego mechanizmu uwagi: oblicza wagę dla każdego elementu sekwencji z uwzględnieniem zależności pomiędzy wszystkimi elementami. Wartość "uwagi" obliczana jest na podstawie iloczynu skalarnego między wektorami cech pól. Mechanizm uwagi pozwala modelowi uwzględniać wszystkie pola obrazu podczas przewidywań dla każdego pola.
 - Warstwa perceptronowa: Po warstwie wielogłowego mechanizmu uwagi wyniki są przekazywane przez wielowarstwowy perceptron, który stosuje funkcję aktywacji ReLU, w celu transformacji wyników.
- Warstwa klasyfikacji: Na koniec sekwencja wyników jest przekazywana warstwie klasyfikacji, która przypisuje etykiety obrazom.

Warte podkreślenia jest to, że ViT wymaga dużego zbioru danych treningowych z oznaczonymi obrazami do skutecznego uczenia. Proces uczenia obejmuje optymalizację parametrów modelu przy użyciu metod optymalizacji opartych na gradiencie, takich jak metoda gradientu



Rysunek 6.3: Struktura sieci ViT [9]

prostego lub Adam, oraz minimalizację funkcji straty, która mierzy rozbieżność między przewidywanymi etykietami a rzeczywistymi etykietami. Często stosuje się pre-trenowanie na dużych zbiorach danych, takich jak ImageNet, a następnie dostrojenie (z ang. fine-tuning) na mniejszym zbiorze danych dostosowanym do konkretnego zadania [9].

Jedną z głównych zalet ViT jest wysoka wydajność w różnych zadaniach widzenia komputerowego, w tym w klasyfikacji, wykrywaniu obiektów na obrazach i segmentacji bez użycia warstw splotu. Jednak mogą przy tym wymagać większych zasobów obliczeniowych w porównaniu z tradycyjnymi CNN, zwłaszcza w przypadku obrazów o wysokiej rozdzielczości. Aby rozwiązać ten problem, zaproponowano różne techniki w celu poprawy wydajności i zmniejszenia wymaganych zasobów obliczeniowych [18, 41].

6.2. Wstępna obróbka oraz podział danych

W celu poprawy dokładności modeli oraz zmniejszenia zjawiska overfittingu wszystkie obrazy zostały poddane procesowi augmentacji danych poprzez dodanie następujących warstw:

- normalizacji
- skalowania obrazu do rozmiaru 256 x 256 pikseli (w przypadku ViT do 72 x 72 pikseli)
- losowego horyzontalnego odwracania
- losowej rotacji
- losowego powiększenia lub pomniejszenia

Inne podejścia do augmentacji danych, takie jak np. transformacje kolorów, nie zostały wykorzystane, ponieważ istnieje możliwość zniekształcenia komórek, np. komórka z jądrem

o ciemnym zabarwieniu posiada wysoką tendencję do bycia komórką nieprawidłową. Zbiór danych został podzielony na 3 losowe wybrane podzbiory w następujących proporcjach:

- 70% jako dane treningowe
- 20% jako zbiór testowy
- 10% jako zbiór walidacyjny

6.3. Schematy modeli

6.3.1. Własna sieć CNN

Architektura własnego modelu sieci wykorzystującego CNN została przedstawiona na rysunku 1 oraz 2. Poza początkowymi warstwami augmentacji danych składa się ona z trzech warstw splotowych, trzech warstw łączących, warstwy spłaszczającej oraz z dwóch końcowych warstw perceptronowych. Parametry dla każdej warstwy takie jak rozmiar warstwy, rozmiar oraz liczba użytych filtrów, liczba parametrów warstwy czy użyta funkcja aktywacji przedstawione zostały na rysunku 6.4 oraz 6.5.

Na przykład warstwa spłotu conv2d_3 wykorzystuje 32 filtry z jądrem $3 \times 3 \times 1$, co warstwę $254 \times 254 \times 32$ pikseli. Po każdej warstwie spłotu następuje jednostka liniowa typu ReLU. Zaraz po wykonaniu spłotu warstwa max_pooling2d_3 przeprowadza proces łączenia, co skutkuje uzyskaniem maksymalnej wartości jądrem 3×3 (dla każdego dwóch pikseli), zmniejszając rozmiar warstwy do $127 \times 127 \times 32$ pikseli.

Po trzech warstwach spłotu i trzech warstwach łączenia pojawia się warstwa spłaszczająca redukująca dane do jednego wymiaru. Dwie ostatnie warstwy składające się z wielowarstwowego perceptronu. W ostatniej warstwie wyniki są uzyskiwane przy użyciu funkcji softmax, co pozwala na łatwy odczyt wartości wystąpienia danej klasy.

```
model = Sequential()
model.add(Normalization())
model.add(Resizing(256, 256))
model.add(RandomFlip("horizontal"))
model.add(RandomRotation(factor=0.02))
model.add(RandomZoom(height_factor=0.2, width_factor=0.2))
model.add(Conv2D(32, (3,3), 1, activation='relu', input_shape=img_shape))
model.add(MaxPooling2D())
model.add(Conv2D(16, (3,3), 1, activation='relu'))
model.add(MaxPooling2D())
model.add(Conv2D(16, (3,3), 1, activation='relu'))
model.add(MaxPooling2D())
model.add(Flatten())
model.add(Dense(256, activation='relu'))
model.add(Dense(n_classes, activation='softmax'))
```

Rysunek 6.4: Model własnej sieci CNN

Layer (type)	Output Shape	Param #
normalization_2 (Normalization)	(None, 256, 256, 3)	7
resizing_2 (Resizing)	(None, 256, 256, 3)	0
random_flip_2 (RandomFlip)	(None, 256, 256, 3)	0
random_rotation_2 (RandomRotation)	(None, 256, 256, 3)	0
random_zoom_2 (RandomZoom)	(None, 256, 256, 3)	0
conv2d_3 (Conv2D)	(None, 254, 254, 32)	896
max_pooling2d_3 (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_4 (Conv2D)	(None, 125, 125, 16)	4624
max_pooling2d_4 (MaxPooling2D)	(None, 62, 62, 16)	0
conv2d_5 (Conv2D)	(None, 60, 60, 16)	2320
max_pooling2d_5 (MaxPooling2D)	(None, 30, 30, 16)	0
flatten_1 (Flatten)	(None, 14400)	0
dense_2 (Dense)	(None, 256)	3686656
dense_3 (Dense)	(None, 3)	771
Total params: 3,695,274		
Trainable params: 3,695,267		
Non-trainable params: 7		

Rysunek 6.5: Podsumowanie kompilacji modelu własnej sieci CNN

6.3.2. VGG-16

Struktura modelu oraz podsumowanie z kompilacji dostosowanego modelu VGG-16 zostały przedstawione na rysunku 6.6 oraz 6.7. Pre-trenowany model sieci VGG-16 został przystosowany do zadania klasyfikacji poprzez dodanie warstw augmentacji danych (przed vgg16 na rys. 6.7) oraz warstwy spłaszczającej i 2 warstw typu w pełni połączonych (pierwszej z aktywacją typu ReLu oraz drugą typu softmax) na końcu.

```

dnn_model = Sequential()

imported_model= tf.keras.applications.VGG16(include_top=False,
      input_shape=img_shape,
      pooling='avg',classes=n_classes,)
for layer in imported_model.layers:
    layer.trainable=False

dnn_model.add(Normalization())
dnn_model.add(Resizing(256, 256))
dnn_model.add(RandomFlip("horizontal"))
dnn_model.add(RandomRotation(factor=0.02))
dnn_model.add(RandomZoom(height_factor=0.2, width_factor=0.2))
dnn_model.add(imported_model)
dnn_model.add(Flatten())
dnn_model.add(Dense(512, activation='relu'))
dnn_model.add(Dense(n_classes, activation='softmax'))

```

Rysunek 6.6: Dostosowany model sieci VGG-16

Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 256, 256, 3)	7
resizing (Resizing)	(None, 256, 256, 3)	0
random_flip (RandomFlip)	(None, 256, 256, 3)	0
random_rotation (RandomRotation)	(None, 256, 256, 3)	0
random_zoom (RandomZoom)	(None, 256, 256, 3)	0
vgg16 (Functional)	(None, 512)	14714688
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 512)	262656
dense_1 (Dense)	(None, 3)	1539
Total params: 14,978,890		
Trainable params: 264,195		
Non-trainable params: 14,714,695		

Rysunek 6.7: Podsumowanie kompilacji dostosowanego modelu sieci VGG-16

6.3.3. ResNet-50

Struktura modelu oraz podsumowanie z kompilacji dostosowanego modelu ResNet-50 zostały przedstawione na rysunku 6.8 oraz 6.9. Pre-trenowany model sieci ResNet-50 podobnie jak VGG-16 został przystosowany do zadania poprzez dodanie warstw augmentacji danych

oraz warstwy spłaszczającej i 2 warstw typu fully-connected (pierwszej z aktywacją typ ReLu oraz drugą typu softmax) na końcu.

```
dnn_model = Sequential()

imported_model= tf.keras.applications.ResNet50(include_top=False,
        input_shape=(256,256,3),
        pooling='avg',classes=3,)
for layer in imported_model.layers:
    layer.trainable=False

dnn_model.add(Normalization())
dnn_model.add(Resizing(256, 256))
dnn_model.add(RandomFlip("horizontal"))
dnn_model.add(RandomRotation(factor=0.02))
dnn_model.add(RandomZoom(height_factor=0.2, width_factor=0.2))
dnn_model.add(imported_model)
dnn_model.add(Flatten())
dnn_model.add(Dense(512, activation='relu'))
dnn_model.add(Dense(n_classes, activation='softmax'))
```

Rysunek 6.8: Dostosowany model sieci ResNet-50

Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 256, 256, 3)	7
resizing (Resizing)	(None, 256, 256, 3)	0
random_flip (RandomFlip)	(None, 256, 256, 3)	0
random_rotation (RandomRotation)	(None, 256, 256, 3)	0
random_zoom (RandomZoom)	(None, 256, 256, 3)	0
resnet50 (Functional)	(None, 2048)	23587712
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 512)	1049088
dense_1 (Dense)	(None, 3)	1539
Total params: 24,638,346		
Trainable params: 1,050,627		
Non-trainable params: 23,587,719		

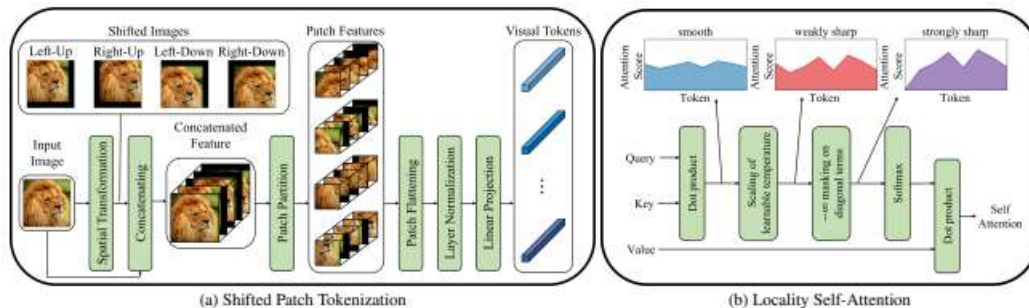
Rysunek 6.9: Podsumowanie kompilacji dostosowanego model sieci ResNet-50

6.3.4. ViT

W artykule [9] autorzy zauważyli, że transformaty wizyjne wymagają znacznie większych ilości danych, aby dorównać skutecznością sieciom konwolucyjnym. W artykule [20] autorzy stwierdzili, że powodem dlaczego ViT słabiej radzi sobie w obliczu małych zbiorów danych jest warstwa samo-uwagi ViT, która nie uwzględnia lokalności błędu indukcyjnego (z ang. inductive bias). Rozwiązaniem tego problemu zaproponowanym przez autorów jest wprowadzenie do oryginalnej architektury ViT dwóch koncepcji:

- Tokenizacja z przesunięciem pola (z ang. Shifted Patch Tokenization - STP) - oryginalnie w ViT obrazy wejściowe dzielone są na pola, które są następnie przekształcane na tokeny. Implementacja STP pozwoliła na zwiększenie recepcji przez algorytm. STP składa się z następujących etapów (na rys. 6.10): 1. Załadowanie obrazu. 2. Przesunięcie obrazu po przekątnych. 3. Połączenie przemieszczonych wcześniej obrazów z oryginałem. 4. Wyodrębnienie fragmentów pól z połączonych obrazów. 5. Spłaszczenie wszystkich pól obrazu. 6. Normalizacja i rzutowanie pól jako wektor.
- Lokalna samo-uwaga (z ang. Locality Self Attention) - autorzy zaproponowali zmianę maskowania iloczynu skalarnego zwiększając poziom uwagi na relacje występujące między tokenami.

Oba pomysły zostały wykorzystane w implementacji ViT opracowanej na podstawie [17, 18, 19]. Na rys 6.10 można zaobserwować wizualizację powyższych pomysłów.



Rysunek 6.10: Wizualne przedstawienie koncepcji Shifted Patch Tokenization oraz Locality Self Attention [20]

Fragmenty powtarzającej się struktury modelu (który zawiera znaczną ilość warstw w związku z czym nie możliwe było umieszczenie schematu całego modelu) pochodzącej z podsumowania z kompilacji modelu ViT zostały przedstawione na rysunku oraz 6.11 oraz 6.12.

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	[(None, 72, 72, 3)]	0	[]
shifted_patch_tokenization (ShiftedPatchTokenization)	((None, 144, 64), (None, 12, 12, 540))	35704	['input_layer[0][0]']
patch_encoder (PatchEncoder)	(None, 144, 64)	9216	['shifted_patch_tokenization[0][0]']
layer_normalization_1 (LayerNormalization)	(None, 144, 64)	128	['patch_encoder[0][0]']
multi_head_attention_lsa (MultiHeadAttentionLSA)	(None, 144, 64)	66369	['layer_normalization_1[0][0]', 'layer_normalization_1[0][0]']
add (Add)	(None, 144, 64)	0	['multi_head_attention_lsa[0][0]', 'patch_encoder[0][0]']
layer_normalization_2 (LayerNormalization)	(None, 144, 64)	128	['add[0][0]']
dense_1 (Dense)	(None, 144, 128)	8320	['layer_normalization_2[0][0]']
dropout (Dropout)	(None, 144, 128)	0	['dense_1[0][0]']
dense_2 (Dense)	(None, 144, 64)	8256	['dropout[0][0]']
dropout_1 (Dropout)	(None, 144, 64)	0	['dense_2[0][0]']
add_1 (Add)	(None, 144, 64)	0	['dropout_1[0][0]', 'add[0][0]']

Rysunek 6.11: Początkowy fragment podsumowania kompilacji modelu sieci ViT

layer_normalization_16 (LayerNormalization)	(None, 144, 64)	128	['add_14[0][0]']
dense_15 (Dense)	(None, 144, 128)	8320	['layer_normalization_16[0][0]']
dropout_14 (Dropout)	(None, 144, 128)	0	['dense_15[0][0]']
dense_16 (Dense)	(None, 144, 64)	8256	['dropout_14[0][0]']
dropout_15 (Dropout)	(None, 144, 64)	0	['dense_16[0][0]']
add_15 (Add)	(None, 144, 64)	0	['dropout_15[0][0]', 'add_14[0][0]']
layer_normalization_17 (LayerNormalization)	(None, 144, 64)	128	['add_15[0][0]']
flatten (Flatten)	(None, 9216)	0	['layer_normalization_17[0][0]']
dropout_16 (Dropout)	(None, 9216)	0	['flatten[0][0]']
dense_17 (Dense)	(None, 2048)	18876416	['dropout_16[0][0]']
dropout_17 (Dropout)	(None, 2048)	0	['dense_17[0][0]']
dense_18 (Dense)	(None, 1024)	2098176	['dropout_17[0][0]']
dropout_18 (Dropout)	(None, 1024)	0	['dense_18[0][0]']
output_dense (Dense)	(None, 3)	3075	['dropout_18[0][0]']
Total params: 21,688,323 Trainable params: 21,688,323 Non-trainable params: 0			

Rysunek 6.12: Końcowy fragment podsumowania modelu sieci ViT

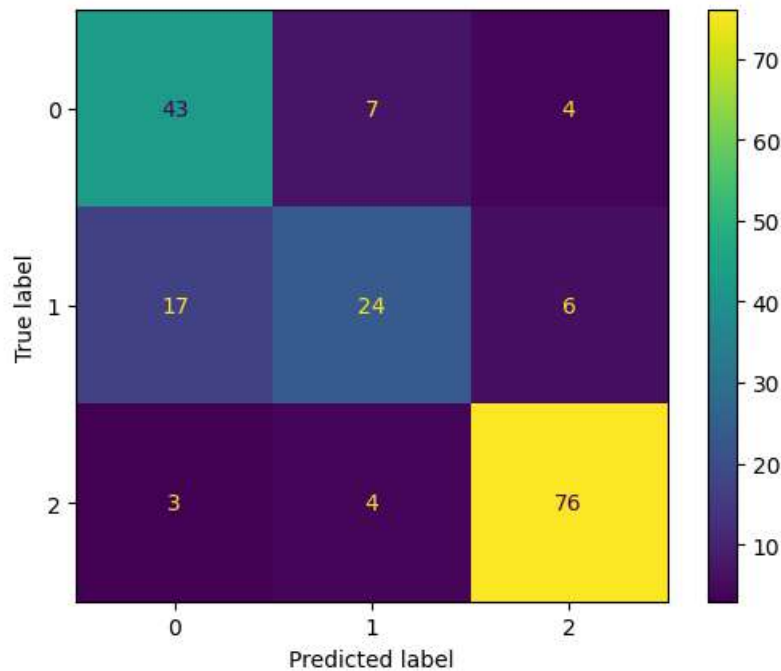
6.4. Testowanie

Sieci neuronowe wymagają określenia wartości wielu parametrów przed rozpoczęciem uczenia m.in. tempa uczenia, liczby epok, wielkości przekazywanej partii danych (z ang. batchy), rodzaju i liczby parametrów warstw. Często problemem jest określenie optymalnych parametrów, dlatego najczęściej spotykanym podejściem (na początku) jest trenowanie sieci z predefiniowanym zestawem parametrów. W następnym etapie (ale w ramach procesu trenowania) parametry sieci są dostosowywane, dzięki znajomości wyników wydajności sieci, które są testowane na walidacyjnym zestawie danych. Po zakończeniu procesu trenowania ostatnim krokiem jest finalne testowanie na zbiorze testowym [43].

Do testowania sieci stosuje się różne metryki, te wykorzystane w pracy zostały opisane poniżej:

- Macierz błędów

Pozwala na porównanie przewidywań wobec wyników rzeczywistych dla każdej z testowanych klas, dlatego jest podstawą do oceny skuteczności algorytmów klasyfikacji [47]. Na rys. 6.13 przedstawiono macierz błędów, która uzyskana została podczas testowania skuteczności algorytmów w tej pracy.



Rysunek 6.13: Macierz błędów uzyskana podczas eksperymentów

Każdy wiersz macierzy przedstawia liczbę przypadków rozpoznanych jako jedna z 3 klas dla prawdziwych wyników danej klasy. Przedstawiona macierz różni się od tradycyjnej macierzy błędów pokazanej w tabeli 6.1, która została przewidziana dla klasyfikacji binarnej składając się z następujących przypadków:

		Klasa rzeczywista	
		Pozytywna	Negatywna
Predykcja	Pozytywna	TP	FP
	Negatywna	FN	TN

Tabela 6.1: Macierz błędu

- prawdziwie pozytywna (TP – True-Positive): przewidywanie pozytywne, rzeczywista klasa pozytywna
- prawdziwie negatywna (TN – True-Negative): przewidywanie negatywne, rzeczywista klasa negatywna
- fałszywie pozytywna (FP – False-Positive): przewidywanie pozytywne, rzeczywista klasa negatywna
- fałszywie negatywna (FN – False-Negative): przewidywanie negatywne, rzeczywista klasa pozytywna
- warunek pozytywny (P – Positive): liczba rzeczywistych pozytywnych przypadków
- warunek negatywny (N – Negative): liczba rzeczywistych negatywnych przypadków

Większość metryk wykorzystuje te 4 kategorie do kalkulacji własnych wskaźników. Jednak macierz z rys. 6.13 nie spełnia tych warunków, dlatego należy zredukować ją do postaci z tab. 6.1. W tym celu wykonuje się macierz (jak na tab. 6.1) dla każdej z 3 klas obecnych w zbiorze (zakładając że wynikiem pozytywnym jest predykcja danej klasy, a negatywnym predykcja jednej z pozostałych klas), a następnie obliczany jest średni wynik dla tych 3 klas.

- Dokładność - określa proporcję wszystkich poprawnych przewidywań względem wszystkich przewidywań (wzór 6.1) Oczekuje się aby wynik był jak najbliższy liczbie 1 (można też przedstawić wynik jako wartość procentową), co oznacza że algorytm przewiduje poprawne etykiety dla wszystkich przypadków. Dokładność pokazuje, jak sieć neuronowa radzi sobie na zrównoważonym zestawie danych.

$$ACC = \frac{TP + TN}{P + N} \quad (6.1)$$

- Precyzja - jest to liczba prawdziwie pozytywnych przypadków względem wszystkich pozytywnych przypadków (również tych pochodzących z błędnej predykcji, czyli razem $TP + FP$) (wzór 6.2). Pożądane jest również, aby wartość to była jak najbliższa liczbie 1. Wskaźnik ten jest przydatny do zauważenia, ile przypadków, dla których przewiduje się spełnienie warunku osiąga ten wynik (dlatego sprawdza się tylko przypadki pozytywne).

$$PPV = \frac{TP}{TP + FP} \quad (6.2)$$

- Czułość - jest to liczba prawdziwie pozytywnych przypadków pośród wszystkich pozytywnych przypadków (ale tym razem uwzględniając przypadki błędnie zaklasyfikowane

do negatywnych, czyli łącznie $TP + FP$) (wzór 6.3). Przy czym warto pamiętać, że jeśli algorytm nie zaklasyfikuje żadnego pozytywnego przypadku błędnie (czyli nie trafi do kategorii FN), to czułość będzie wynosić 1. Nawet jeśli będzie błędnie klasyfikował FP, czyli negatywne przypadki będą trafiać do kategorii pozytywnej. Pożądane jest, aby wartość to była jak najbliższa liczbie 1. Metryka ta jest przydatna do obserwacji ile przypadków, które spełniają warunek, przewidywanych jest, że spełnią ten warunek.

$$TPR = \frac{TP}{TP + FN} \quad (6.3)$$

- Wynik F1 - metryka ta jest średnią harmoniczną precyzji oraz czułości, co pokazano w równaniu 6.4. Również pożądane jest, aby wartość to była jak najbliższa liczbie 1. Metryka ta może być dobrym wynikiem oceniającym skuteczność algorytmu.

$$F1 = \frac{2TP}{2TP + FP + FN} = 2 * \left(\frac{PPV * TPR}{PPV + TPR} \right) \quad (6.4)$$

Częścią wspólną opisanych metryk jest nieuwzględnienie wartości TN. Jednak nie należy się tym przejmować, ponieważ założeniem testowania w tym przypadku jest osiągnięcie rozpoznania choroby, dlatego wystarczy uwzględnienie jedynie TP.

7. Wyniki

7.1. Założenia eksperymentu oraz trening

Eksperymenty zostały przeprowadzone w środowisku Google Colab na wirtualnym GPU NVIDIA Tesla T4, co pozwoliło na znaczne przyspieszenie procesu uczenia. Dane wykorzystywane w eksperymencie zostały umieszczone w środowisku chmurowym Google Drive w celu zwiększenia szybkości przesyłania danych w środowisku. Zbiór danych został podzielony na trzy zestawy, w tym dane treningowe, walidacyjne i testowe. W tym celu 70% wszystkich obrazów zostało przydzielonych do zestawu treningowego, a pozostałe obrazy zostały przeznaczone do zestawów walidacyjnych (10%) oraz testowych (20%). Do treningu wykorzystywane były 4 modele opisane w poprzednim rozdziale. W celu zmierzenia skuteczności treningu oraz walidacji oraz oceny zjawiska overfittingu skorzystano z metryk dokładności (SparseCategoricalAccuracy) oraz funkcji straty (SparseCategoricalCrossentropy). Do oceny wydajności modeli wykorzystano natomiast: dokładność, czułość, precyzję oraz wynik F1.

Kluczowe parametry modeli wykorzystane do eksperymentów zostały przedstawione w tabeli 7.1. Wartości te zostały dobrane podczas prób na zaimplementowanych algorytmach, tak by zapewnić najbardziej równe warunki do wykonania rzetelnego porównania.

Tabela 7.1: Parametry dla algorytmów

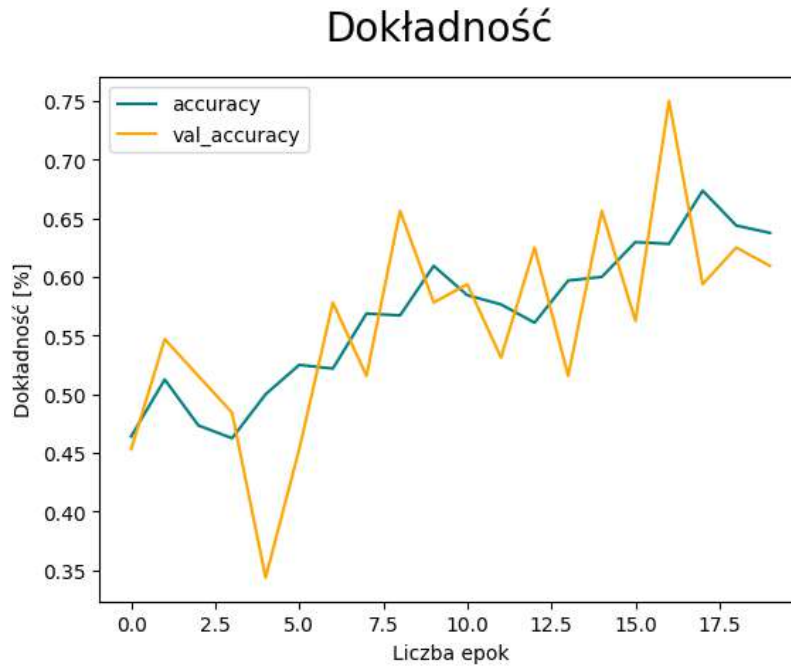
Model	Batch Size	Epoch	Learning rate	Optimizer
CNN	32	20	0,001	adam
VGG16	32	20	0,001	adam
ResNet-50	32	20	0,001	adam
ViT	256	100	0,001	adam

Ustawienia dla 3 modeli konwolucyjnych są identyczne, natomiast dla modelu ViT zdecydowano na zwiększenie liczby epok (do 100) z powodu znacznie niższego czasu wykonywania jednej epoki, co przełożyło się na znacznie niższy czas kompilacji całego modelu. Kolejną różniącą się wartością jest liczba batchy (256 zamiast 32) co zostało spowodowane architekturą sieci ViT.

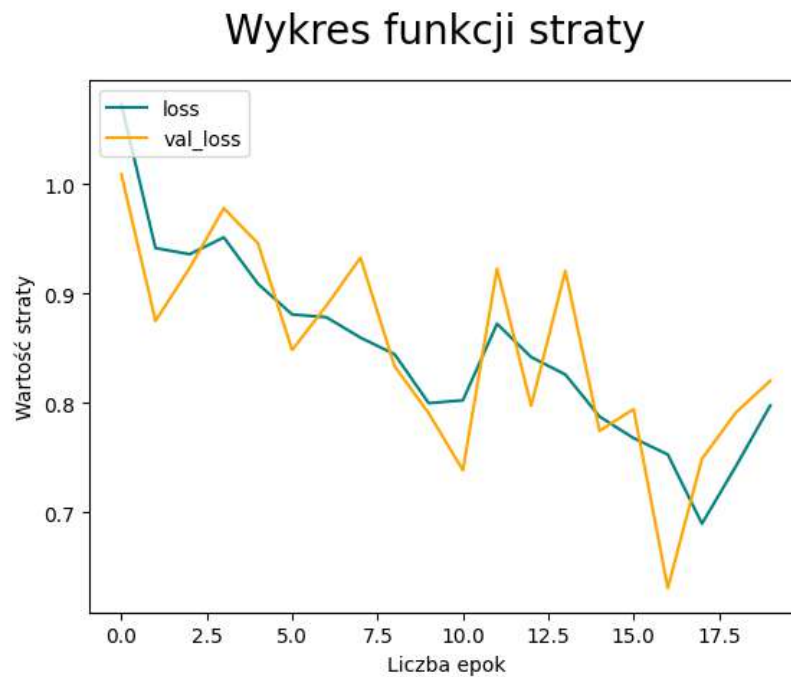
7.2. Rezultaty i analiza eksperymentu

7.2.1. Wydajność klasyfikacji modeli na zestawach treningowych i walidacyjnych podczas uczenia

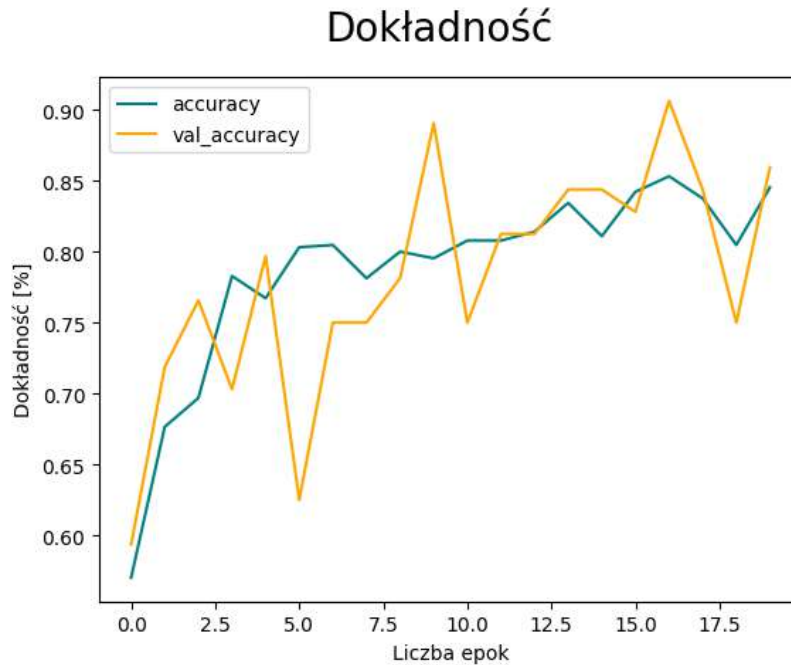
Rysunki 7.1 do 7.8 przedstawiają przykładowe krzywe dokładności i strat modeli na zestawach treningowych i walidacyjnych podczas uczenia.



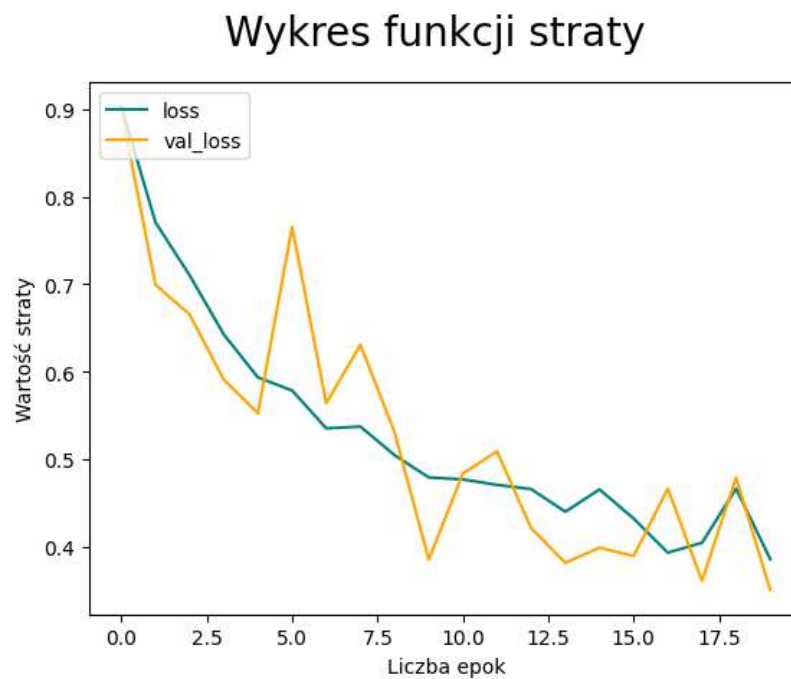
Rysunek 7.1: Wykres dokładności własnego modelu CNN na zestawach treningowych i walidacyjnych podczas uczenia



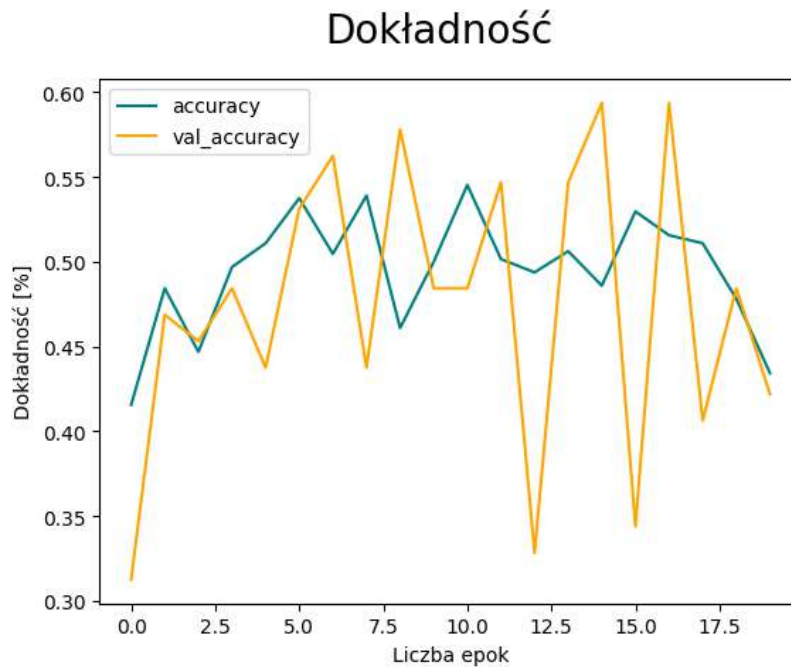
Rysunek 7.2: Wykres funkcji straty własnego modelu CNN na zestawach treningowych i walidacyjnych podczas uczenia



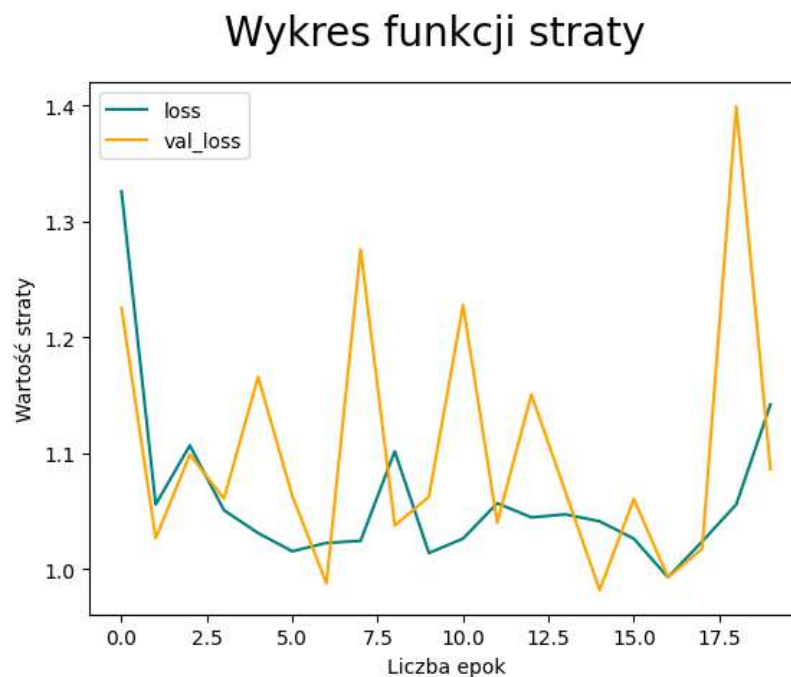
Rysunek 7.3: Wykres dokładności modelu VGG-16 na zestawach treningowych i walidacyjnych podczas uczenia



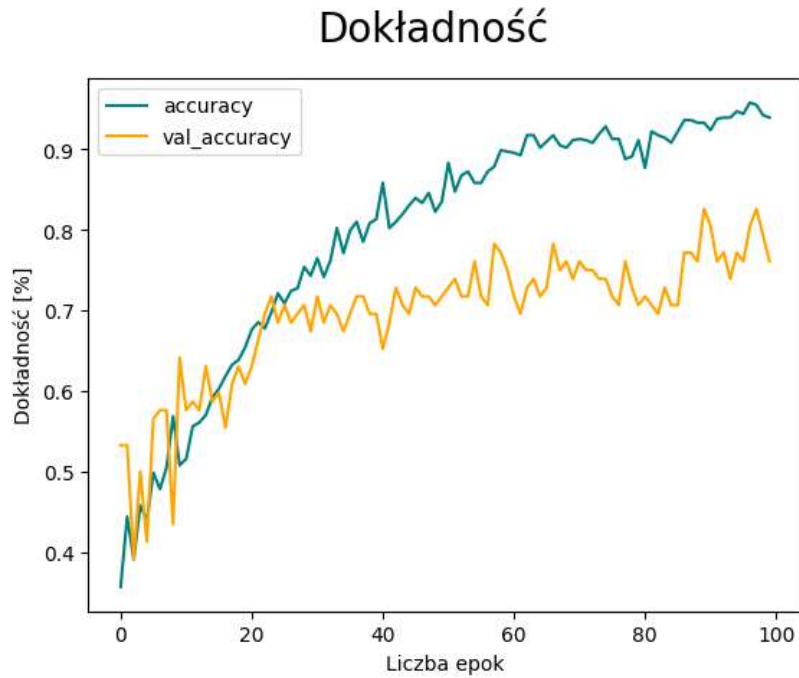
Rysunek 7.4: Wykres funkcji straty modelu VGG-16 na zestawach treningowych i walidacyjnych podczas uczenia



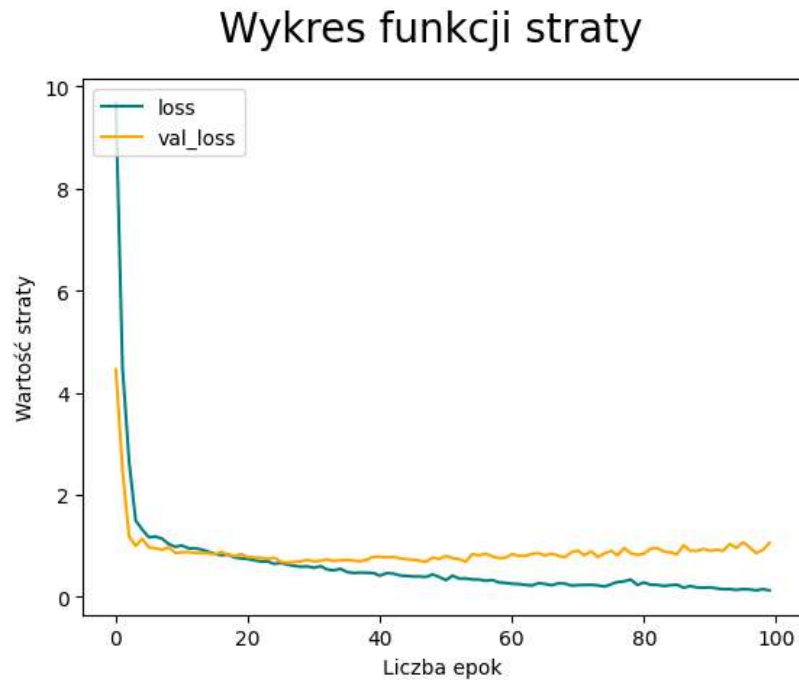
Rysunek 7.5: Wykres dokładności modelu ResNet-50 na zestawach treningowych i walidacyjnych podczas uczenia



Rysunek 7.6: Wykres funkcji straty modelu ResNet-50 na zestawach treningowych i walidacyjnych podczas uczenia



Rysunek 7.7: Wykres dokładności modelu ViT na zestawach treningowych i walidacyjnych podczas uczenia



Rysunek 7.8: Wykres funkcji straty modelu ViT na zestawach treningowych i walidacyjnych podczas uczenia

Jak pokazano na przedstawionych rysunkach, wskaźnik dokładności zestawu treningowego był porównywalny lub nieznacznie mniejszy z zestawem walidacyjnym dla 3 modeli: CNN oraz VGG-16 co oznacza, że dane nie były w znacznym stopniu nadmiernie dopasowane. Dla ResNet-50 dokładność oraz funkcja straty charakteryzują się dużymi wahaniami oraz dokładność na zbiorze walidacyjnym bywa mniejsza od dokładności na zbiorze testowym co wskazuje na zbyt duże dopasowanie modelu do danych (z ang. overfitting). Natomiast dla ViT dokładność dla zestawu treningowego była wyższa niż dla zestawu walidacyjnego, a krzywa funkcji straty dla zestawu walidacyjnego znajdowała się powyżej krzywej funkcji straty dla zestawu treningowego co sugeruje, że model jest zbalansowany.

7.2.2. Wydajność klasyfikacji modeli na zestawie testowym

Tabela 7.2 przedstawia wskaźniki wydajności każdego modelu na zestawie testowym, w tym metryki takie jak: liczba parametrów (Par.), średni czas treningu (Avg. Tre.), średnia precyzja (Avg. Prec.), średnia dokładność (Avg. Acc.), średnia czułość (Avg. Rec.) oraz średni wynik F1 (Avg. F1).

Tabela 7.2: Porównanie wyników klasyfikacji różnych modeli głębokiego uczenia na zbiorze testowym. Wyniki zostały uśrednione dla 3 prób. Najlepsze wyniki w kolumnach zostały **pogrubione**.

Model	#Par.	Avg. Tre. [min]	Avg. Acc. [%]	Avg. F1 [%]	Avg. Rec. [%]	Avg. Prec. [%]
CNN	3.7M	30:37	70.3	60.3	60.7	66.7
VGG-16	15M	36:02	86.3	84.0	83.7	85.7
ResNet-50	24.7M	28:37	53.3	31.3	37.3	36.3
ViT	21.7M	5:41	75.0	70.7	70.7	73.3

Na podstawie wyników zestawu testowego (przedstawione w tab. 7.2) można zaobserwować, że najbardziej dokładnym modelem okazał się VGG-16 osiągając 86,3% średniej dokładności, będąc aż o 11,3% wyższy od drugiego w kolejności modelu ViT. Zaskakująco wysoką dokładność, bliską ViT, (pomimo bardzo prostej struktury) osiągnął model CNN, natomiast niespodziewanie najgorszą dokładnością charakteryzował się model ResNet-50 będąc wyraźnie gorszym od konkurencji. Również dla innych metryk: średniego wyniku F1, średniej czułości oraz średniej precyzji sieć VGG-16 zanotowała najlepsze wyniki, wynosząc odpowiednio 84%, 83,7% i 85,7%. Ciekawą statystyką jest średnia długość trenowania, w której wyraźnie wyróżnił się model ViT poświęcając średnio najmniej czasu na wykonanie zadania - jedynie 5:51 min, co jest 6,34 krotnie mniej niż najlepszy, ale wymagający najwięcej czasu model VGG-16. Czas treningu pozostałych modeli nie różnił się już w tak dużym stopniu. Kolejnym interesującym wskaźnikiem była ilość parametrów sieci. Największą ilością parametrów charakteryzował się model ResNet-50, który osiągnął przy tym najgorszą dokładność. Z tej statystyki można

wywnioskować, że liczba parametrów sieci nie wpływa na czas treningu czy na wydajność sieci.

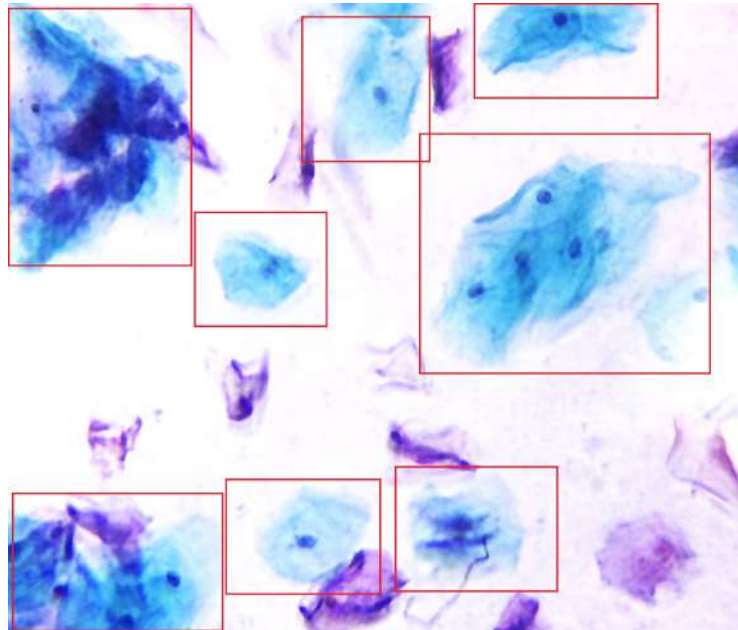
7.3. Dyskusja oraz wnioski

Eksperyment dotyczył wydajności klasyfikacji 4 modeli głębokiego uczenia na małym zbiorze obrazów preparatów cytologicznych (918 obrazów). W celu zwiększenia dokładności wszystkie modele zostały poddane augmentacji danych, a modele VGG-16, ResNet-50 były dodatkowo pretrenowane na dużym zbiorze obrazów ImageNet. Aby poprawić wiarygodność wniosków, eksperyment został wykonany trzykrotnie, a wyniki zostały uśrednione. Rezultaty porównania (zgodnie z przyjętymi metrykami) zostały przedstawione w tabeli 7.2. Podsumowując badania VGG-16 było siecią o najlepszej ogólnej wydajności, model ResNet-50 osiągnął najgorszą ogólną wydajność, a ViT okazał się modelem o najkrótszym czasie działania. Z wyników eksperymentu można odczytać, że sieci charakteryzowały się niewysoką tendencją do nadmiernego dopasowania danych, poza modelem ResNet-50, dla którego ta tendencja jest duża (w tym przypadku zachodzi overfitting). Można przypuścić, że model ResNet-50 mógłby uniknąć tego zjawiska w przypadku zwiększenia liczby danych treningowych. Jednak przyczyną może być również architektura sieci ResNet-50, która składa się z wielu warstw konwolucyjnych, co potwierdza najwyższa liczba parametrów ze wszystkich modeli (tab. 7.2). można przypuszczać, że jest to zbyt dużo dla wykorzystanego zbioru danych. Kolejnym zagadnieniem jest wykorzystanie transferu wag modelu (transfer learning). Modele VGG-16 oraz ResNet-50 były pretrenowane na zbiorze obrazów ImageNet, co powinno znacznie zwiększyć wydajność obu modeli. Z wyników eksperymentu można odczytać, że model ResNet-50 osiągnął najgorszą wydajność, a model VGG-16 najlepszą. Dlatego można wywnioskować, że wykorzystanie transferu wag jest korzystne w przypadku doboru odpowiedniej architektury oraz parametrów sieci. W celu poprawy wyników ViT można zwiększyć ilość danych treningowych oraz zastosować wstępne trenowanie sieci na dużym zbiorze danych (jak sugerują autorzy w artykule [19]). W takim przypadku ViT mógłby także znacznie zwiększać wskaźniki wydajnościowe konkurując z najlepszym w przypadku tego eksperymentu modelem VGG-16. W przypadku modeli opartych na warstwach konwolucyjnych wydajność również mogłaby zostać poprawiona w przypadku skorzystania z większego zbioru danych, jednak powinna być mniejsza niż dla ViT [9].

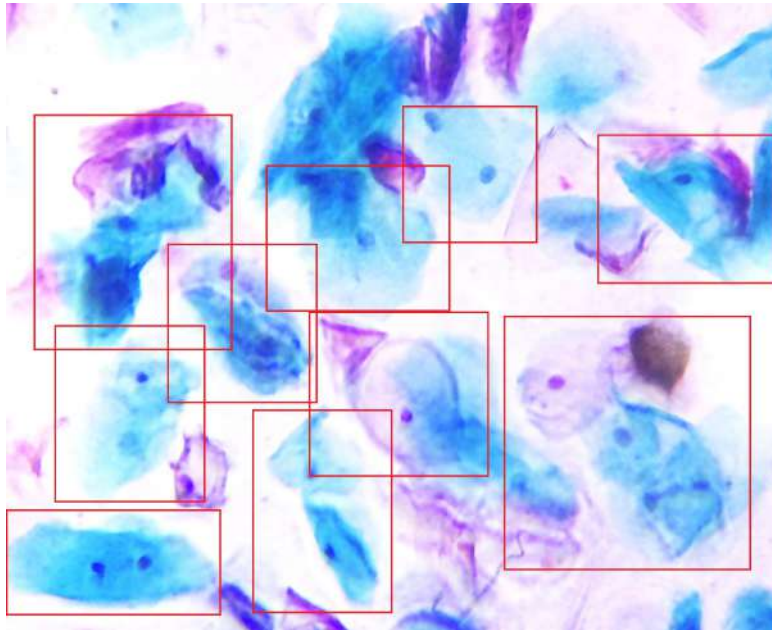
Na podstawie wyników eksperymentu można wywnioskować, że modele głębokiego uczenia nadają się do zastosowania na niedużym zbiorze obrazów cytologicznych bez nadmiernego dopasowania w przypadku gdy wybrana jest odpowiednia architektura sieci (modelu) oraz jeśli poddana jest właściwym modyfikacjom. Kluczowe jest wybranie właściwego, perspektywicznego algorytmu, którego wydajność na niedużym zbiorze danych będzie wysoka oraz warto zastosować techniki takie jak augmentacja danych czy dostrajanie parametrów pretrenowanej sieci (z ang. fine-tuning).

7.4. Działanie systemu rejestracji obrazów w oparciu o dane IoT

Przy wykorzystaniu systemu rejestrującego obrazy (który został opisany w podrozdziale 5.2) wykonano 20 zdjęć w formacie jpeg na podstawie próbek pozyskanych metodą LBC oraz wybarwione metodą Papanicolau. Większej ilości zdjęć nie udało się wykonać ze względu na problemy z uzyskaniem preparatów cytologicznych. Na przygotowanej bazie danych sprawdzono działanie algorytmów, co potwierdziło częściową skuteczność predykcji. Przykładowe obrazy wraz z zaznaczonymi komórkami zostały przedstawione na rys. 7.9 oraz 7.10.



Rysunek 7.9: Przykładowy obraz uzyskany przy pomocy systemu rejestracji



Rysunek 7.10: Przykładowy obraz uzyskany przy pomocy systemu rejestracji

8. Podsumowanie

W niniejszej pracy zostały przeprowadzone badania porównawcze algorytmów uczenia głębokiego wykorzystywanych w diagnostyce nowotworowej w celu klasyfikacji obrazów mikroskopowych. W ramach pracy wykonano szeroki przegląd literatury dotyczącej zastosowań tego typu algorytmów oraz zidentyfikowano istniejące metody i techniki stosowane w tym obszarze. Na podstawie zebranych informacji wytypowane zostały 4 metody wykorzystujące uczenie głębokie: własna sieć CNN, VGG-16, ResNet-50 oraz Vision Transformer, które zostały przetrenowane oraz zweryfikowane na udostępnionym zbiorze danych. Metody te zostały dostosowane do wykonania zadania poprzez zastosowanie kilku technik przetwarzania danych: augmentację danych, dodanie warstw oraz wykorzystanie transferu wag (dla ResNet-50 oraz VGG-16). Istotnym i w pewnym stopniu nowatorskim elementem pracy było sprawdzenie modelu Vision Transformer, który na chwilę obecną jest rzadko stosowany w kategorii klasyfikacji obrazów medycznych. Wyniki eksperymentów wykazały, że najlepiej działającym modelem okazał się VGG-16 osiągając 86,3% średniej dokładności na zbiorze testowym. Również dla innych metryk: średniego wyniku F1, średniej czułości oraz średniej precyzji sieć VGG-16 zanotowała najlepsze wyniki, wynosząc odpowiednio 84%, 83,7% i 85,7%. Ponadto modelem o najszybszym czasie treningu okazał się model Vision Transformer, a model oparty na własnej sieci CNN posiadał najmniejszą liczbę parametrów. Wyniki przeprowadzonych badań wykazały, że zaproponowane metody oparte na uczeniu głębokim są w dużym stopniu skuteczne w rozpoznawaniu i klasyfikacji komórek nowotworowych pochodzących z diagnostyki cytologicznej.

Podczas badań zidentyfikowano również problemy takie jak: niedobór danych treningowych, zróżnicowanie morfologiczne komórek nowotworowych oraz trudności związane z interpretacją wyników. Dlatego na potrzeby przyszłych prac metody powinny zostać przetestowane na większym zbiorze danych oraz przy użyciu bardziej zaawansowanych metod augmentacji obrazów. Warto również rozważyć wykorzystanie innych architektur modeli oraz modyfikację parametrów uczenia.

Ważnym założeniem, które zostało spełnione, było również opracowanie systemu rejestrującego obrazy, który można zamontować na dowolnym mikroskopie optycznym. Przy udziale systemu wykonano zdjęcia, które następnie zostały przetestowane na opracowanych algorytmach. Przyszła praca mogłaby wykorzystać ten system do utworzenia własnego zbioru danych.

Bibliografia

- [1] M. A. Al-Abbadi. Basics of cytology. *Avicenna journal of medicine*, 1(01):18–28, 2011.
- [2] T. Albuquerque, R. Cruz, and J. S. Cardoso. Ordinal losses for classification of cervical cancer risk. *PeerJ Computer Science*, 7:e457, 2021.
- [3] A. Alrajjal, V. Pansare, M. S. R. Choudhury, M. Y. A. Khan, and V. B. Shidham. Squamous intraepithelial lesions (sil: Lsil, hsil, ascus, asc-h, lsil-h) of uterine cervix and bethesda system. *Cytojournal*, 18, 2021.
- [4] C. Andreassen. Melanoma prognosis prediction using image processing and machine learning. Master’s thesis, uis, 2022.
- [5] basicmedicalkey.com. Cervix, Marzec 2023. <https://basicmedicalkey.com/cervix-2/>.
- [6] Y.-F. Chen, P.-C. Huang, K.-C. Lin, H.-H. Lin, L.-E. Wang, C.-C. Cheng, T.-P. Chen, Y.-K. Chan, and J. Y. Chiang. Semi-automatic segmentation and classification of pap smear cells. *IEEE Journal of Biomedical and Health Informatics*, 18(1):94–108, 2014.
- [7] commons.wikimedia.org. Resnet50.png, Czerwiec 2023. <https://commons.wikimedia.org/wiki/File:ResNet50.png>.
- [8] commons.wikimedia.org. Vgg16.png, Czerwiec 2023. <https://commons.wikimedia.org/wiki/File:VGG16.png>.
- [9] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [10] dr. hab. n. med. Magdalena Bodnar prof. UMK. Zabezpieczenie. przygotowanie oraz transport materiału cytologicznego do zakładu patomorfologii, placentologii i hematopatologii klinicznej, Marzec 2023. https://bizei.umk.pl/assets/files/cenniki/2022/PR13_PJ10_w3.pdf.
- [11] R. Dunleavy. *Cervical cancer: a guide for nurses*. John Wiley & Sons, 2008.
- [12] A. Hashmi, S. Naz, O. Ahmed, S. Yaqeen, M. Asif, A. kamal, N. Faridi, and M. Irfan. Comparison of liquid-based cytology and conventional papanicolaou smear for cervical cancer screening: An experience from pakistan. *Cureus*, 12, 12 2020.
- [13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

- [14] E. Hussain, L. B. Mahanta, C. R. Das, and R. K. Talukdar. A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network. *Tissue and Cell*, 65:101347, 2020.
- [15] N. C. Institute. Cervical cancer screening, Marzec 2023. <https://www.cancer.gov/types/cervical/screening>.
- [16] H. Jiang, Y. Zhou, Y. Lin, R. C. Chan, J. Liu, and H. Chen. Deep learning for computational cytology: A survey. *Medical Image Analysis*, page 102691, 2022.
- [17] kaggle.com. Vision transformers for brain tumor classification, Czerwiec 2023. <https://www.kaggle.com/code/raagulbharatwajk/vision-transformers-for-brain-tumor-classification>.
- [18] keras.io. Image classification with vision transformer, Czerwiec 2023. https://keras.io/examples/vision/image_classification_with_vision_transformer/.
- [19] keras.io. Train a vision transformer on small datasets, Czerwiec 2023. https://keras.io/examples/vision/vit_small_ds/.
- [20] S. H. Lee, S. Lee, and B. C. Song. Vision transformer for small-size datasets, 2021.
- [21] Y. H. Lee. Liquid-based cytology in gynecologic cytology. *Korean Journal of Pathology*, 43:291–300, 2009.
- [22] A. S. Lundervold and A. Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2):102–127, 2019.
- [23] A. Manna, R. Kundu, D. Kaplun, A. Sinitca, and R. Sarkar. A fuzzy rank-based ensemble of cnn models for classification of cervical cytology. *Scientific Reports*, 11(1):14538, 2021.
- [24] mirosławmamczur.pl. Vision transformer, Maj 2023. <https://mirosławmamczur.pl/czym-jest-i-jak-działa-transformer-siec-neuronowa/>.
- [25] mp.pl. Cytologia - kiedy robić badanie, co oznaczają wyniki, Marzec 2023. <https://www.mp.pl/pacjent/ginekologia/badania-i-zabiegi/292320, cytologia-badanie-cytologiczne-wszystko-co-musisz-wiedziec>.
- [26] A. Nasierowska-Guttmejer, W. Kędzia, W. Rokita, S. Wojtylak, D. Lange, R. Jach, and M. Wielgoś. Rekomendacje dotyczące diagnostyki i leczenia płaskonabłonkowych zmian śródnabłonkowych szyjki macicy na podstawie wytycznych cap/asccp. *Ginekologia i Perinatologia Praktyczna*, 1(3):130–138, 2016.
- [27] T. A. C. of Obstetricians and Gynecologists. *Updated Cervical Cancer Screening Guidelines*, Kwiecień 2021. <https://www.acog.org/clinical/clinical-guidance/practice-advisory/articles/2021/04/updated-cervical-cancer-screening-guidelines>.

- [28] J. R. Opracowanie zbiorowe pod przewodnictwem Renaty Langfort, Andrzeja Marszałka. *Patomorfologia: standardy i przykłady dobrej praktyki oraz elementy diagnostyki różnicowej. Wytyczne dla zakładów/pracowni patomorfologii*. zdrowie.gov.p, Marzec 2023. https://zdrowie.gov.pl/uploads/pub/pages/page_1053/text_images/wytyczne_patomorfologia_1.pdf.
- [29] K. Pawlak. Cytologia wyniki – system bethesda (tbs) i skala papanicolaou (pap), Marzec 2023. <https://wylecz.to/badania-diagnostyczne/cytologia-wyniki-system-bethesda-tbs-i-skala-papanicolaou-pap/>.
- [30] A. Pirovano, L. G. Almeida, S. Ladjal, I. Bloch, and S. Berlemont. Computer-aided diagnosis tool for cervical cancer screening with weakly supervised localization and detection of abnormalities using adaptable and explainable classifier. *Medical image analysis*, 73:102167, 2021.
- [31] radekbialowas.pl. Vision transformer, Maj 2023. <https://radekbialowas.pl/architektura-transformer-nlp-pod-mikroskopem-transformer-architecture-explaine/>.
- [32] M. M. Rahaman, C. Li, X. Wu, Y. Yao, Z. Hu, T. Jiang, X. Li, and S. Qi. A survey for cervical cytopathology image analysis using deep learning. *IEEE Access*, 8:61687–61710, 2020.
- [33] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
- [34] J. H. Smith. Cytology, liquid-based cytology and automation. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 25(5):585–596, 2011.
- [35] S. Sornapudi, G. T. Brown, Z. Xue, R. Long, L. Allen, and S. Antani. Comparing deep learning models for multi-cell classification in liquid-based cervical cytology image. In *AMIA Annual Symposium Proceedings*, volume 2019, page 820. American Medical Informatics Association, 2019.
- [36] towardsdatascience.com. Guide to using pretrained models in keras, Czerwiec 2023. <https://towardsdatascience.com/step-by-step-guide-to-using-pretrained-models-in-keras-c9097b647b29>.
- [37] Wikipedia. Bethesda system, Marzec 2023. https://en.wikipedia.org/wiki/Bethesda_system.
- [38] Wikipedia. Cytopathology, Marzec 2023. <https://en.wikipedia.org/wiki/Cytopathology>.
- [39] Wikipedia. Georgios papanikolaou, Marzec 2023. https://en.wikipedia.org/wiki/Georgios_Papanikolaou.
- [40] Wikipedia. System papanicolaou, Marzec 2023. https://pl.wikipedia.org/wiki/System_Papanicolaou.
- [41] wikipedia.org. Vision transformer, Maj 2023. https://en.wikipedia.org/wiki/Vision_transformer.

- [42] win.eurocytology.eu. Automatyzacja skryningu cytologicznego i cytologia fazy płynnej, Marzec 2023. <https://win.eurocytology.eu/static/eurocytology/POL/cervical/LP1ContentKcontE.html>.
- [43] www.analyticsvidhya.com. Important model evaluation error metrics, Czerwiec 2023. <https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/>.
- [44] www.geeksforgeeks.org. Activation functions neural networks, Maj 2023. <https://www.geeksforgeeks.org/activation-functions-neural-networks/>.
- [45] www.histeroskopia.info. Cytologia, Marzec 2023. <https://www.histeroskopia.info/cytologia-warszawa.html>.
- [46] www.statystyczny.pl. Co to jest machine learning, Maj 2023. <https://www.statystyczny.pl/co-to-jest-machine-learning/>.
- [47] www.statystyczny.pl/. Macierz błędów raport dokładność czułość precyzja, Czerwiec 2023. <https://www.statystyczny.pl/macierz-bledow-raport-dokladnosc-czulosc-precyzja/>.
- [48] www.webedcafe.com. Limitations of cytology, Marzec 2023. https://www.webedcafe.com/extern/program_media/amjmed.com/2012/CCS-HPV/printable.php?speaker=monk&figure=14.
- [49] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu, and J. Yao. Deeppap: deep convolutional networks for cervical cell classification. *IEEE journal of biomedical and health informatics*, 21(6):1633–1643, 2017.
- [50] Y. Zhou, H. Chen, H. Lin, and P.-A. Heng. Deep semi-supervised knowledge distillation for overlapping cervical cell instance segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*, pages 521–531. Springer, 2020.

Spis rysunków

2.1	Przebieg typowego badania - diagnostyki cytologicznej [16]	9
2.2	Porównanie próbek pozyskanych metodą konwencjonalną (A) oraz LBC (B) [12]	11
2.3	Wygląd oraz opis cech komórek szyjki macicy [3]	14
5.1	Komórki typu HSIL (powiększenie 100x)	24
5.2	Komórki typu HSIL (powiększenie 400x)	24
5.3	Komórki typu LSIL (powiększenie 100x)	25
5.4	Komórki typu LSIL (powiększenie 400x)	25
5.5	Komórki typu NSIL (powiększenie 100x)	26
5.6	Komórki typu NSIL (powiększenie 400x)	26
5.7	Mikroskop optyczny PZO	27
5.8	Mini-komputer Raspberry Pi 3 B wraz z podłączoną kamerą	28
5.9	Adapter na mikroskop wydrukowany w 3D	28
5.10	Zmontowany system rejestracji (widok z góry)	29
5.11	Zmontowany system rejestracji (widok z boku)	29
6.1	Struktura sieci VGG16 [8]	32
6.2	Struktura sieci ResNet-50 [7]	33
6.3	Struktura sieci ViT [9]	34
6.4	Model własnej sieci CNN	35
6.5	Podsumowanie kompilacji modelu własnej sieci CNN	36
6.6	Dostosowany model sieci VGG-16	37
6.7	Podsumowanie kompilacji dostosowanego modelu sieci VGG-16	37
6.8	Dostosowany model sieci ResNet-50	38
6.9	Podsumowanie kompilacji dostosowanego modelu sieci ResNet-50	38
6.10	Wizualne przedstawienie koncepcji Shifted Patch Tokenization oraz Locality Self Attention [20]	39
6.11	Początkowy fragment podsumowania kompilacji modelu sieci ViT	40
6.12	Końcowy fragment podsumowania modelu sieci ViT	40
6.13	Macierz błędów uzyskana podczas eksperymentów	41
7.1	Wykres dokładności własnego modelu CNN na zestawach treningowych i walidacyjnych podczas uczenia	46
7.2	Wykres funkcji straty własnego modelu CNN na zestawach treningowych i walidacyjnych podczas uczenia	46
7.3	Wykres dokładności modelu VGG-16 na zestawach treningowych i walidacyjnych podczas uczenia	47

7.4	Wykres funkcji straty modelu VGG-16 na zestawach treningowych i walidacyjnych podczas uczenia	47
7.5	Wykres dokładności modelu ResNet-50 na zestawach treningowych i walidacyjnych podczas uczenia	48
7.6	Wykres funkcji straty modelu ResNet-50 na zestawach treningowych i walidacyjnych podczas uczenia	48
7.7	Wykres dokładności modelu ViT na zestawach treningowych i walidacyjnych podczas uczenia	49
7.8	Wykres funkcji straty modelu ViT na zestawach treningowych i walidacyjnych podczas uczenia	49
7.9	Przykładowy obraz uzyskany przy pomocy systemu rejestracji	52
7.10	Przykładowy obraz uzyskany przy pomocy systemu rejestracji	53

Spis tabel

2.1	Porównanie metod ThinPrep i SurePath [21]	10
2.2	Porównanie metod LBC i metody konwencjonalnej [21]	10
2.3	Klasyfikacja mikroskopowa obrazów cytologicznych wg. propozycji Papanicolaou [25, 45]	12
5.1	Struktura wykorzystanego zbioru danych	23
6.1	Macierz błędu	42
7.1	Parametry dla algorytmów	45
7.2	Porównanie wyników klasyfikacji różnych modeli głębokiego uczenia na zbiorze testowym. Wyniki zostały uśrednione dla 3 prób. Najlepsze wyniki w kolumnach zostały pogrubione .	50