

**WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI,**

**INFORMATYKI I INŻYNIERII BIOMEDYCZNEJ**

Projekt dyplomowy

*Automatyczna diagnoza zmian skórnych z wykorzystaniem głębokich sieci neuronowych*

*Automatic diagnosis of skin lesions with the use of deep neural networks*

Autor: *Gabriela Hanna Piwar*

Kierunek studiów: Inżynieria Biomedyczna

Opiekun pracy: *prof.* *dr hab. Małgorzata Jakubowska*

Kraków, 2022

**SPIS TREŚCI**

[Wstęp 3](#_Toc121569609)

[Cel i założenia projektu 4](#_Toc121569610)

[Rozdział 1. Zagadnienia teoretyczne związane z projektem 5](#_Toc121569611)

[1. Dermatologiczne spojrzenie na zmiany skórne 5](#_Toc121569612)

[1.1. Budowa i funkcje skóry 5](#_Toc121569613)

[1.2. Podstawowe rodzaje zmian skórnych 6](#_Toc121569614)

[1.3. Przyczyny powstawania zmian nowotworowych skóry 7](#_Toc121569615)

[1.4. Zachorowalność i metody leczenia nowotworów złośliwych skóry 9](#_Toc121569616)

[2. Uczenie maszynowe – zautomatyzowane rozwiązywanie problemów 11](#_Toc121569617)

[2.1. Konwolucyjne sieci neuronowe 12](#_Toc121569618)

[2.2. Sposoby ewaluacji rezultatów klasyfikacji 14](#_Toc121569619)

[2.3. Modele sieci konwolucyjnych w medycynie 16](#_Toc121569620)

[3. Środowisko i język programowania użyte w pracy 19](#_Toc121569621)

[Rozdział 2. Implementacja algorytmów klasyfikacyjnych 20](#_Toc121569622)

[1. Opis wybranego zbioru zdjęć dermatoskopowych 20](#_Toc121569623)

[2. Analiza danych dołączonych do zbioru HAM10000 23](#_Toc121569624)

[3. Przygotowanie i równoważenie zbioru 29](#_Toc121569625)

[4. Przebieg klasyfikacji zmian skórnych 33](#_Toc121569626)

[4.1. Optymalizacja wstępnego algorytmu 33](#_Toc121569627)

[4.2. Klasyfikacja przy użyciu sieci pretrenowanych 36](#_Toc121569628)

[5. Ewaluacja rezultatów uczenia 39](#_Toc121569629)

[Podsumowanie i możliwości rozwoju 44](#_Toc121569630)

[Wnioski 45](#_Toc121569631)

[Bibliografia 46](#_Toc121569632)

# Wstęp

Czerniak jest jednym z groźniejszych nowotworów ze względu na jego pozostawanie przez przedłużony okres w stanie utajonym oraz wysoką podatność na wczesne powstawanie przerzutów. Rozwój medycyny pozwolił na stosowanie w leczeniu raka takich technik jak chemioterapia, radioterapia i różnego rodzaju terapie farmaceutykami. Są to metody, które dają zadowalające rezultaty, ale ze swoimi skutkami ubocznymi. Najważniejszy pozostaje jednak czas diagnozy. Im szybciej zmiana zostanie zdiagnozowana, tym większa szansa na pełne wyleczenie. Wpływ na szybką diagnozę ma przede wszystkim zwiększona świadomość ludzi i wczesne podjęcie odpowiednich kroków. W ostatnich latach zaczęto eksplorować możliwości uczenia maszynowego jako narzędzia do detekcji zmian nowotworowych. Nie jest to zupełną nowością, gdyż medycyna – a w dużej mierze radiologia – praktycznie opiera się na systemach komputerowych. Powstanie sztucznej inteligencji, zdolnej do rozpoznawania chorób na podstawie zdjęć czy wywiadu nie jest już nierealnym pomysłem, a rzeczywistością. Technologia ta jest ciągle rozwijana i już znajduje swoje pierwsze zastosowania, ułatwiając tym samym złożone zadania pracowników ochrony zdrowia.

Motywacją do napisania niniejszej pracy była chęć zoptymalizowania algorytmów klasyfikacyjnych, wykorzystywanych do przyśpieszenia diagnozy analizowanych zmian skórnych. Przebieg klasyfikacji jest ściśle związany z budową wykorzystywanej sieci neuronowej, jak również doborze sterujących nią hiperparametrów. Ze względu na podjęte kroki możliwe jest zmniejszenie czasu uczenia, jak również zwiększenie wartości metryk, opisujących skuteczność modelu.

W treści projektu wyodrębniono dwa rozdziały. Pierwszy z nich skupia się na poruszeniu zagadnień teoretycznych związanych z pracą. Omówione w nim zostały aspekty takie jak zachorowalność na nowotwory skóry oraz podstawy związane z przyczynami ich powstawania. Opisane zostały również modele uczenia głębokiego, które bazują na architekturze sztucznych sieci neuronowych i jakie możliwości daje ich wykorzystanie w medycynie. Drugi rozdział natomiast, prezentuje dokumentację całego procesu związanego z implementacją algorytmu do automatycznej diagnozy. W pierwszym etapie dokonano analizy wybranego zbioru danych oraz przeprowadzono szereg modyfikacji dopasowujących go do klasyfikacji binarnej. W rozdziale tym przedstawiono też architektury proponowanych sieci, jak również skuteczność ich działania. W końcowej części poddano je ewaluacji stosując metryki wyznaczające wartości liczbowe w ocenie przypisania etykiet obrazom w porównaniu do rzeczywistych klas.

# Cel i założenia projektu

Celem pracy była implementacja algorytmu uczenia maszynowego, który posłuży do automatycznej klasyfikacji binarnej zmian skórnych na podstawie zdjęć dermatoskopowych. Kluczowe było odpowiednie dobranie parametrów modelu, jak i warstw, z których był zbudowanych, aby możliwe było osiągnięcie jak najwyższej trafności, która pozwoliłaby na sprawniejszy przebieg diagnozy patologii skóry.

Założenia projektu obejmowały w pierwszym etapie analizę i dobór danych z podziałem na dwie klasy - zmiany łagodne i złośliwe. Następnie implementację modelu sieci neuronowej zaprojektowanego do klasyfikacji zmian skórnych, a na koniec wytypowanie optymalnego narzędzia ewaluacji uzyskanych wyników.

# Rozdział 1. Zagadnienia teoretyczne związane z projektem



## Dermatologiczne spojrzenie na zmiany skórne

Skóra to największy narząd organizmu ludzkiego, który pełni szereg istotnych funkcji. Stanowi ona przede wszystkim barierę oddzielającą środowisko wewnętrzne organizmu od zewnętrznego, chroniąc nas przed groźnymi infekcjami i drobnoustrojami. Odpowiada również za termoregulację i gospodarkę wodną ciała. Jej kondycja ma istotny wpływ na nasze samopoczucie jak i zdrowie. Wraz z rozwojem medycyny zwiększyła się świadomość na temat dbania o skórę i ochrony przed niekorzystnymi zaburzeniami. Dziedziną poświęconą badaniom nad budową, funkcjami oraz chorobami skóry jak i jej przydatków jest dermatologia. Stanowi ona stale rozwijający się dział medycyny, ale szczególną popularnością cieszy się rozwój działalności dermokosmetyków dostępnych na rynku. Najczęściej spotykane przeznaczone są do codziennej pielęgnacji skóry twarzy i ciała. Skupiają się one na wyglądzie skóry, mają na celu jej ujędrnienie, odmłodzenie przez zniwelowanie zmarszczek czy powierzchowne oczyszczenie. Jednak często poważniejsze zaburzenia pozostają niewykryte i zaniedbane. Może to doprowadzić do rozwoju chorób zagrażających nawet życiu, w szczególności są nimi nowotwory. Z tego względu kładzie się coraz większy nacisk na regularne badania skóry.

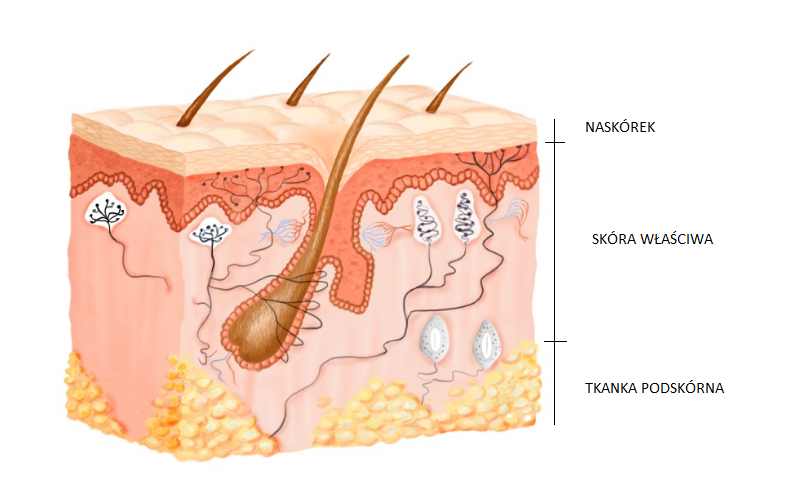
### Budowa i funkcje skóry

Przed przystąpieniem do szczegółowego omówienia zagadnienia zmian skórnych, konieczne jest przedstawienie podstawowych informacji na temat samej budowy skóry. Zaczynając od najważniejszego podziału, w jej makrostrukturze wyróżniamy trzy warstwy: naskórek (łać. *epidermis*), skórę właściwą (łać. *dermis*) oraz tkankę podskórną (łać. *hypodermis*). Omawiane struktury zilustrowano na Rysunku 1.

Naskórek jest najbardziej zewnętrzną częścią skóry, stanowiącą rolę mechanicznej i wodoodpornej bariery oddzielającej środowisko zewnętrzne. Zbudowany jest z w dużej mierze z keranocytów, ale obecne są też w nim komórki melanocytowe. Tlen dostarczany jest do komórek w procesie dyfuzji, ponieważ naskórek nie posiada sieci naczyń krwionośnych [1]. W jego budowie możemy wyróżnić pięć lub cztery warstwy – w zależności od lokalizacji na ciele. Są to warstwa podstawna, kolczysta, ziarnista, jasna i rogowa (Rysunek 2).

Skóra właściwa natomiast składa się z tkanki łącznej właściwej zbitej. Jej głównym celem jest nadanie wysokiej wytrzymałości i sprężystości skórze, dzięki zawartym w macierzy międzykomórkowej włóknom kolagenowym i elastynowym. To w niej znajdują się liczne naczynia krwionośne jak również zakończenia nerwowe tworzące struktury takie jak mechanoreceptory czy termoreceptory.

Ostatnim elementem skóry jest tkanka podskórna, czyli wewnętrzna warstwa skóry. Jej głównym elementem budulcowym jest tkanka łączna właściwa luźna. Tworzy ona połączenie skóry z mięśniami szkieletowymi oraz kośćmi, ale dzięki luźnej strukturze komórek umożliwia częściowy ruch nad strukturami do których jest przymocowana. W jej skład wchodzi również tkanka tłuszczowa, pełniąca rolę izolacyjną, ale jest również ona źródłem energii i miejscem gromadzenia się wody. Ilość tkanki tłuszczowej w tkance podskórnej jest silnie zależna od czynników takich jak wiek, płeć, gospodarka hormonalna czy dieta i tryb życia.



Rysunek 1. Schemat budowy skóry z podziałem na warstwy [2].

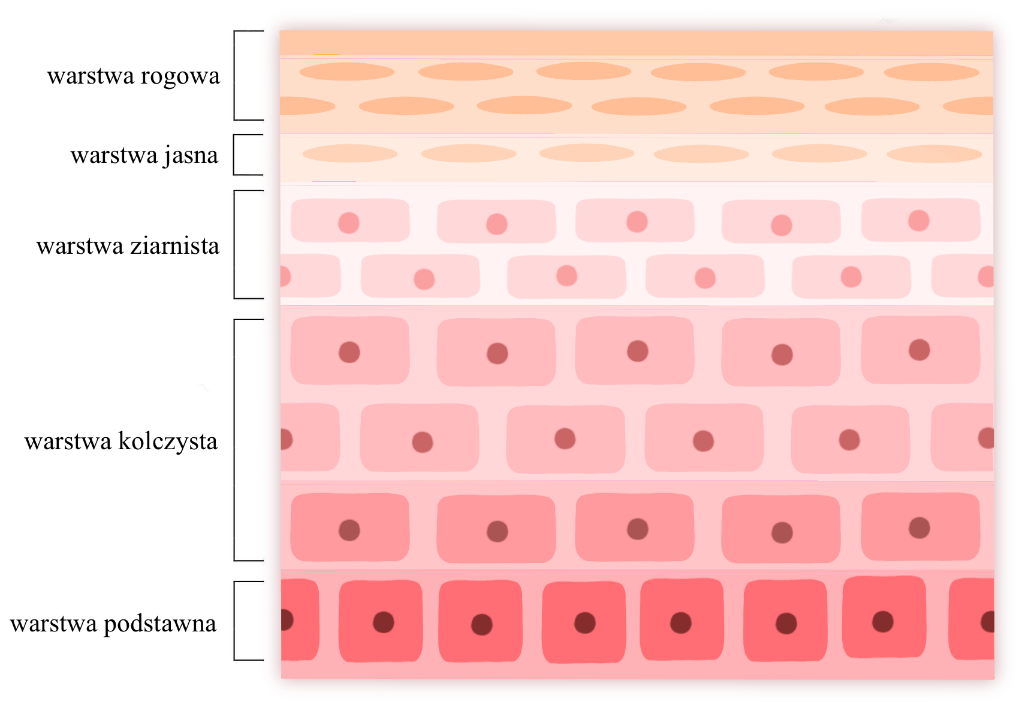
### Podstawowe rodzaje zmian skórnych

Zmiany skórne możemy określić jako obszary na skórze, o cechach odróżniających je od otaczającej skóry. Mogą one mieć nieregularny kształt, powiększony rozmiar, inny kolor czy też nietypową powierzchnię. Powodów powstawania zmian skórnych jest wiele, mogą one być wywołane miejscowym uszkodzeniem skóry lub też przejawem ogólnoustrojowych zaburzeń. Z tego powodu można wyróżnić dwie grupy zmian skórnych: pierwotne i wtórne [3]. Wykwity pierwotne powstają na skutek działania znanego czynnika na zdrowy fragment skóry. Są to przede wszystkim pieprzyki, piegi, strupy, ale również pęcherze po oparzeniach czy krostki. Nie leczone lub poddane większym urazom mogą przemienić się w zmiany wtórne, wywołujące większe powikłania w dalszej perspektywie.

W diagnostyce zmian skórnych ważniejszą rolę odgrywa podział ze względu na ich charakter. Zmiany łagodne nie mają podłoża nowotworowego i mogą występować na każdej części ciała. Mają one bardzo wiele postaci, najczęściej spotykane to zmiany barwnikowe na tle melanocytowym lub naczyniowym. Nie są one groźne dla zdrowia, nie wymagają leczenia, chyba że pojawiają się objawy takie jak świąd czy dyskomfort. Problem stanowią natomiast zmiany złośliwe, często wizualnie zbliżone do łagodnych, ale o znacznie groźniejszym działaniu. Są one uznawane za nowotwory, najczęściej spotykanymi są czerniak i rak podstawnokomórkowy. Cechami charakteryzującymi zmiany złośliwe jest gwałtowny rozrost, zmiana koloru oraz trwałe, niegojące się strupy [3]. Wyróżniamy również tak zwane stany przedrakowe, do których zalicza się rogowacenie słoneczne, nazywane też starczym. Są to zmiany skórne o statystycznie zwiększonym ryzyku przemiany w nowotwór.

### Przyczyny powstawania zmian nowotworowych skóry

Naskórek to warstwa skóry, w której zachodzi ciągła wymiana komórek. Przyjmuje się, że cykl pełnej wymiany u dorosłej osoby trwa w zakresie od 28 do 50 dni. Cały proces jest możliwy dzięki stałej proliferacji komórek w warstwie podstawnej naskórka (Rysunek 2), które z czasem przesuwają się w stronę coraz bardziej zewnętrznych warstw, a na końcu dochodzi do ich złuszczania w warstwie rogowej. Nic dziwnego więc, że najczęstszym miejscem powstawania nowotworów w skórze jest właśnie warstwa podstawna naskórka. Z jej komórek powstaje rak podstawnokomórkowy stanowiący 75-80% wszystkich rozpoznawanych raków skóry [4]. Drugim nowotworem na czele diagnozowanych zmian jest rak płaskonabłonkowy (kolczystokomórkowy) złożony z komórek warstwy kolczystej. Szansa na zachorowanie w ciągu życia na jeden z przedstawionych typów wynosi aż 20% [4]. W przypadku czerniaka zmiany nowotworowe zachodzą w melanocytach – komórkach barwnikowych naskórka, kiedy pod wpływem danego czynnika zaburzone są ich cykle metaboliczne oraz procesy apoptozy, zdolności do podziału i wzrostu komórek. Uznaje się, że ponad połowa wszystkich diagnozowanych przypadków czerniaka to nowe zmiany, a blisko 40% powstaje z przeobrażenia już istniejących zmian barwnikowych łagodnych [5].



Rysunek 2. Rozmieszczenie warstw naskórka.

Czynnikiem mającym największy wpływ na ryzyko powstania nowotworu złośliwego skóry jest ekspozycja na światło słoneczne. Zwiększona liczba poparzeń słonecznych w ciągu życia ma silny wpływ na podwyższone ryzyko zachorowania. Najbardziej narażone są osoby o jasnym fenotypie skóry. W przypadku długotrwałego wystawienia na szkodliwe promieniowanie UVA i UVB zalecane jest pełne zabezpieczenie w postaci filtrów o wysokiej wartości wskaźnika SPF (ang. Sun Protection Factor). Do środowiskowych czynników zaliczamy również te związane z zawodem, jak na przykład przedłużony kontakt ze związkami arsenu, czy też wystawienie na działanie promieniowania rentgenowskiego. Duży wpływ na ryzyko mają również predyspozycje genetyczne, w szczególności duża liczba pieprzyków na skórze. Podstawowym jednak czynnikiem jest wiek – u osób starszych odnotowuje się więcej przypadków nowotworów skóry. Równie istotna jest historia zachorowań w bliskiej rodzinie oraz osoby u których już raz wytworzył się rak skóry, ponieważ mają zwiększoną szansę na powtórne zachorowanie. Poza wymienionymi przyczynami odnotowuje się zwiększone ryzyko wytworzenia nowotworu w wyniku współistnienia innej choroby, tak jak w przypadku zakażenia brodawczakiem ludzkim (HPV) czy immunosupresji spowodowanej nabytym niedoborem odporności czy u pacjentów przyjmujących leki o takim działaniu.

### Zachorowalność i metody leczenia nowotworów złośliwych skóry

Współczynnik zachorowalności na nowotwór skóry w ostatnich latach XXI wieku ma znaczącą tendencję wzrostową, a w przypadku czerniaka rośnie on szybciej niż dla większości innych nowotworów (wyjątek stanowi rak płuc i piersi u kobiet) [6]. Szczególnie widoczne jest to w krajach z przewagą jasnego fenotypu skóry w populacji. Polska nie jest wyjątkiem, w roku 2010 rak skóry stanowił 6,8% u mężczyzn i 7,8% u kobiet wszystkich zachorowań na nowotwory [7]. Tabela 1 zawiera dane na temat zwiększającej się liczby zachorowań wśród mężczyzn jak i kobiet. Wśród nich znajdują się również wartości surowego współczynnika zachorowalności, czyli liczby zachorowań przypadającej na 100 000 osób w populacji oraz współczynnika standaryzowanego, który również opisuje ta zależność ale dodatkowo standaryzowaną metodą bezpośrednią względem struktury wieku w populacji.

Tabela 1. Zachorowalność na raka skóry w Polsce w latach 1990-2010 [7].



W roku 1990 w sumie liczba przypadków nowotworu skóry oscylowała poniżej 4 tyś, natomiast w roku 2010 wyniosła już ponad 10 tyś. przypadków. Obserwujemy, więc prawie 3-krotny wzrost liczby jak i współczynnika zachorowań wśród obu płci w przeciągu 20 lat. Największą grupę ryzyka stanowią osoby w przedziale wiekowym 60+, a najbardziej zagrożone są osoby po 80 roku życia [7].

Niepodważalne jest zagrożenie jakie stanowi rak skóry, dlatego też w Polsce wprowadzony został system ochrony zdrowia mający na celu wczesne go wykrywanie. Ogólnopolski program profilaktyki nowotworów skóry obejmuje zakres badań dermatoskopowych dla wszystkich osób powyżej 50 roku życia i dla osób w przedziale 19-49 lat w przypadku stwierdzenia niepokojących objawów przez lekarza. Badania te są nieinwazyjne i polegają na oglądaniu zmiany za pomocą urządzenia naświetlającego i powiększającego. Głównym celem dermatoskopii jest rozróżnienie zmian łagodnych od złośliwych, ale służy również do przeprowadzania fotograficznej dokumentacji ewolucji danej zmiany. Podstawą diagnozy zmian skórnych jest natomiast badanie mikroskopowe, zwane histopatologicznym. Daje ono największą ilość informacji na temat charakteru i rodzaju zmiany. Aby je wykonać, wymagane jest pobranie wycinka skóry obejmującego badaną zmianę wraz z obszarem zdrowym wokół niej. Tak pobrany wycinek jest odpowiednio preparowany, a następnie poddawany ocenie przez lekarza specjalistę. W celu diagnozy wykonuje się również badania radiologiczne jak RTG i USG, aby wykluczyć możliwość wystąpienia przerzutów. Równie pomocne są wyniki morfologii krwi pacjenta, podwyższone stężenie dehydrogenazy mleczanowej (LDH) może świadczyć o uszkodzeniu tkanek w organizmie pacjenta.

Leczenie nowotworów skóry takich jak rak podstawnokomórkowy czy płaskonabłonkowy w głównej mierze może być ograniczone do wycięcia zmiany z marginesem zdrowej tkanki. W 95% pozwala to na całkowite usunięcie choroby, bez ryzyka nawrotu [4]. Zabieg odbywa się przy znieczuleniu miejscowym i wymaga założenia pojedynczych szwów. Inaczej sytuacja wygląda w przypadku czerniaka, który bardzo często daje przerzuty. W przypadku pojawienia się ich w węzłach chłonnych często konieczna jest limfadenektomia, czyli wycięcie wszystkich dotkniętych chorobą fragmentów układu limfatycznego. Natomiast kiedy przerzuty są bardziej rozsiane po organizmie zalecane są metody takie jak radioterapia, chemioterapia, immunoterapia, ale też leczenie molekularne. Wymienione techniki należą do bardziej inwazyjnych i związane są z nimi długotrwałe skutki zdrowotne.

## Uczenie maszynowe – zautomatyzowane rozwiązywanie problemów

Od momentu wynalezienia pierwszych komputerów powstała idea zautomatyzowanej i samodzielnej nauki sztucznej inteligencji. Programowanie systemów będących w stanie podejmować logiczne decyzję z taką samą dokładnością jak ludzie jest uznawane za krok, który pozwoli wejść nam w nową erę komputeryzacji. Dlatego też wywierana jest duża presja na rozwój technik związanych ze sztuczną inteligencją.

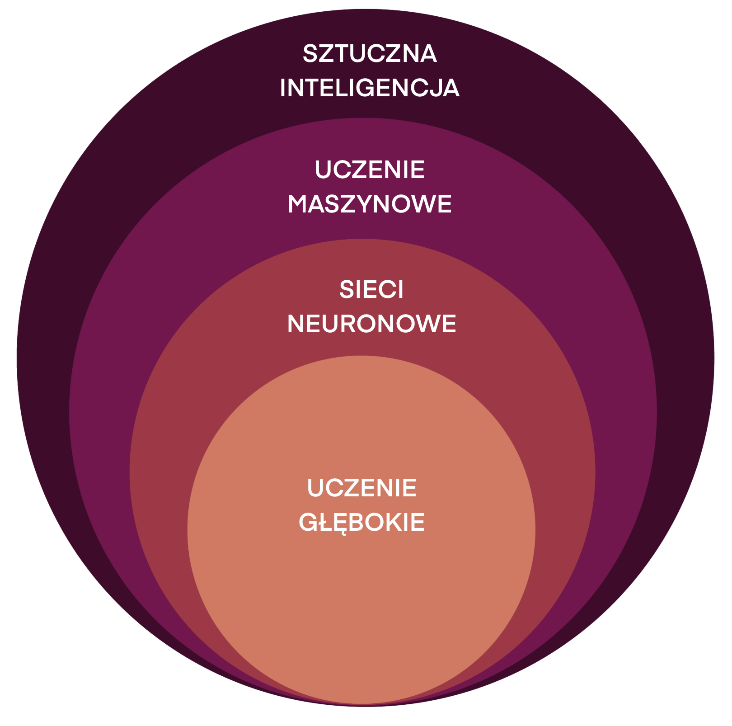
Algorytmy uczenia maszynowego skupiają się na analizie wielkich zbiorów danych z różnych dziedzin nauki. Podejście to łączy ze sobą dyscypliny takie jak informatyka, statystyka, złożoność komputerowa, ale też biologia, filozofia i kognitywistka [8]. Dzięki dostępnym narzędziom możliwa jest szeroka eksploracja danych na podstawie, których algorytmy dokonują klasyfikacji i przewidywań, wykorzystywanych później w podejmowaniu decyzji. Do przykładów zastosowań uczenia maszynowego możemy zaliczyć personalizację reklam na portalach internetowych, rozpoznawanie twarzy przez kamerę aparatu oraz proponowanie diagnozy na podstawie powtarzających się symptomów w wywiadach medycznych.

Z powodu szerokich możliwości wykorzystania uczenia maszynowego, możemy wyróżnić wiele kategorii metod uczenia algorytmów. Podstawowy podział obejmuje cztery różne techniki [9]. Pierwsza z nich to uczenie nadzorowane wykorzystujące zbiór danych z przypisanymi im już etykietami. Umożliwia to przebieg szczegółowej klasyfikacji, gdyż model uczy się na zbiorze treningowym z podanymi odpowiedziami, a następnie uzyskaną wiedzę weryfikuje na zbiorze testowym, gdzie otrzymuje nowe dane bez etykiet. Ten typ uczenia stosowany jest głównie do klasyfikacji i regresji. Kolejna metoda to uczenie bez nadzoru, będąca pewnego rodzaju uzupełnieniem pierwszego. W tym przypadku algorytm pracuje na danych bez etykiet, a tym samym konkretnego polecenia co z nimi zrobić. Jego celem jest znalezienie ukrytych podobieństw i różnic wśród badanego zbioru i pogrupowanie elementów na ich podstawie. Ponieważ, nie istnieje tak zwany „ground truth” do opracowywanych danych, nie jest możliwe zdefiniowanie poprawność działania algorytmu. Mimo tego ta technika znajduje szerokie zastosowanie, ze względu na trudność w nabyciu zbiorów z etykietami, koniecznych podczas uczenia z nadzorem Metodę stanowiącą most między uczeniem z nadzorem i bez nazywamy uczeniem częściowo nadzorowanym. Jest to połączenie zalet obu rozwiązań, stosując mniejszy zbiór danych z etykietami do przypisania ich do obiektów z większego zbioru. Omija to problem związany z dużym nakładem pracy wynikającym z ręcznego przypisania etykiet przez człowieka do elementów całej bazy. Szczególnie przydatna jest w inżynierii biomedycznej w przypadku diagnozowania obrazów medycznych np. zdjęć RTG.

Ostatnią metodą jest uczenie wzmacniane, polegające na systemie nagród i kar. Algorytm ma za zadanie samodzielne znalezienie rozwiązania problemu, dzięki otrzymywaniu nagród za podejmowanie kroków zbliżających go do celu. Dzięki temu brany pod uwagę jest większy wachlarz możliwości, co może poskutkować lepiej zoptymalizowanym systemem.

### Konwolucyjne sieci neuronowe

Ponieważ coraz częściej w życiu codziennym stykamy się z technologiami wykorzystującymi możliwości sztucznej inteligencji, terminy takie jak, uczenie maszynowe, sieci neuronowe czy uczenie głębokie zaczęto w potocznej mowie używać zamiennie, do opisania tego samego zjawiska. Jest to jednak błędne podejście, pomimo że pojęcia te są ze sobą powiązane, istnieją znaczące różnice w ich definicji. W najprostszym ujęciu, każde z nich zawiera się w drugim. Graficzną reprezentacją tych zależności jest diagram z Rysunku 3.



Rysunek 3. Schemat zależności pomiędzy przedstawionymi terminami.

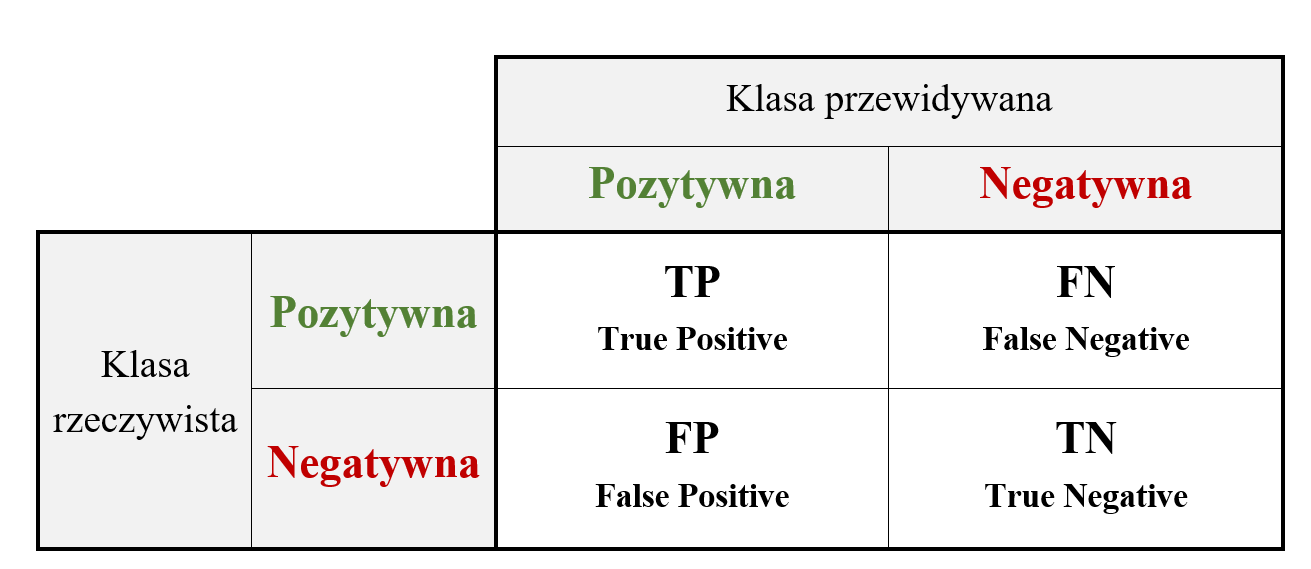
Najbardziej obszernym pojęciem spośród nich jest sztuczna inteligencja (ang. AI, Artificial Inteligence), obejmująca wszystkie zagadnienia związane z tworzeniem modeli programistycznych i matematycznych naśladujących zachowanie inteligentne jakie spotykamy w naturze. Jej częścią jest uczenie maszynowe, które skupia się na zdolności algorytmów do podejmowania decyzji na bazie własnych doświadczeń. Sieci neuronowe, a konkretniej sztuczne sieci neuronowe (ang. ANN, Artificial Neural Network) mają na celu naśladowanie biologicznej architektury mózgu, odtwarzając funkcjonowanie systemu nerwowego [10]. Dzięki temu algorytmy posiadają zwiększoną zdolność uczenia się na podstawie przykładów i potrafią rozwiązywać problemy bez konieczności tworzenia odpowiedniego modelu matematycznego. W momencie, gdy sieć neuronowa zbudowana jest z więcej niż trzech warstw, wkraczamy w pojęcie uczenia głębokiego.

Konwolucyjne sieci neuronowe (ang. CNN, Convolutional Neural Network) to specjalny typ uczenia głębokiego, przeznaczony do pracy na danych w postaci macierzy. Najczęstsze zastosowanie praktyczne znajdują w analizie obrazów, które można przedstawić jako macierze 2D wartości poszczególnych pikseli [11]. W podstawowej architekturze sieci przeznaczonej do klasyfikacji możemy wyróżnić poszczególne warstwy, które na różny sposób przekształcają dane wejściowe na wyjściowe. Wśród nich najważniejsze to:

* **Warstwa wejściowa** (ang. Input Layer) to nic innego jak obraz wejściowy, przedstawiony w postaci macierzy. W zależności od jego rodzaju, możemy wyróżnić jeden kanał wejściowy dla obrazów w skali szarości lub trzy kanały dla obrazów w skali RGB. Skala ta opisuje wartości koloru czerwonego, zielonego i niebieskiego przypadające dla każdego piksela.
* **Warstwa konwolucyjna** (ang. Convolutional Layer) jak sama nazwa wskazuje przeprowadza operację konwolucji, czyli w pojęciu matematycznym splotu. Jest to działanie na dwóch funkcjach, w wyniku czego powstaje nowa, zmodyfikowana wersja pierwotnych funkcji. Pod względem operacji na zdjęciach, jest to przekształcenie macierzowe obrazu, w celu wyodrębnienia poszukiwanych cech. W tej warstwie istotne jest dobranie wartości hiperparametrów takich jak rozmiar jądra (ang. kernel size), wypełnienie (ang. padding) czy liczba pikseli, o które jądro ma być przesuwane (ang. strides). Dzięki nim możliwe jest sterowanie procesem uczenia algorytmu.
* **Warstwa łącząca** (ang. Pooling Layer) ma na celu zmniejszenie liczby parametrów, które podlegają treningowi i złożoności obliczeniowej sieci, poprzez stopniową redukcję rozmiaru danych np. obrazu. Pozwala to uniknąć zjawiska przeuczenia modelu. Najczęściej używanym typem warstwy łączącej jest MaxPooling, która przesuwając się po macierzy wybiera z okna np. 2x2 największą wartość [12].
* **Warstwa porzucająca** (ang. Dropout Layer) podobnie jak w przypadku poprzedniej warstwy, stosowana jest, aby zapobiec przeuczeniu sieci. Działa poprzez przypisanie prawdopodobieństwa, z którym dojdzie do losowego usunięcia neuronów z sieci. Sieć dzięki temu zmuszona jest do nauki w sposób mniej ukierunkowany.
* **Warstwa spłaszczająca** (ang. Flatten Layer) przekształca wielowymiarowe dane sieci w jednowymiarowy wektor. Stosowana jest przed użyciem warstwy gęstej, ponieważ do poprawnego przebiegu klasyfikacji wprowadzane dane musza być spłaszczone.
* **Warstwa połączeń każdy-z-każdym/gęsta** (ang. Fully Connected/Dense Layer) warstwa przeprowadzająca ostateczną klasyfikację obiektu wejściowego, poprzez zwrócenie prawdopodobieństwa jego przynależności do każdej z klas. Jej nazwa wynika z faktu, że łączy ona każdy neuron, z neuronami z poprzedniej warstwy.

### Sposoby ewaluacji rezultatów klasyfikacji

Ze względu na złożoność procesu implementacji modeli klasyfikacyjnych bazujących na sieciach neuronowych, konieczna jest stała kontrola jakości modelu. Aby możliwa była miarodajna i szybka ewaluacja wydajności algorytmu stosuje się metryki oceny klasyfikacji. Zwracają one wartości liczbowe odpowiadające konkretnym parametrom, a tym samym pozwalają na szybką ocenę jakości modelu. Podstawę ewaluacji dla klasyfikacji binarnej stanowi macierz pomyłek zilustrowana na Rysunku 4.

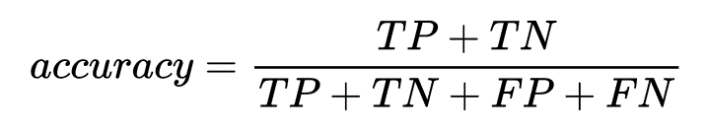


Rysunek 4. Macierz pomyłek.

Tablica ma dwa wiersze i dwie kolumny, gdzie wiersze przedstawiają klasy rzeczywiste, kolumny zaś klasy przewidywane. Oznaczenia w komórkach macierzy to odpowiednio: TP – przypadki prawdziwie pozytywne, FN – przypadki fałszywie negatywne, FP – przypadki fałszywie pozytywne, TN – przypadki prawdziwie negatywne.

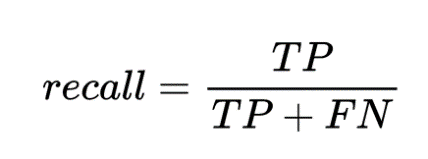
Dzięki graficznej wizualizacji już na pierwszy rzut oka można wyciągnąć wnioski na temat przebiegu klasyfikacji. Dostęp do bardziej precyzyjnych informacji umożliwia obliczenie szczegółowych parametrów macierzy pomyłek, do których zaliczamy:

* **Dokładność** (ang. accuracy) – parametr opisujący, ile przypadków w modelu zostało prawidłowo zaklasyfikowanych, zarówno jako pozytywne, jak i negatywne. Określa ją wzór (1):



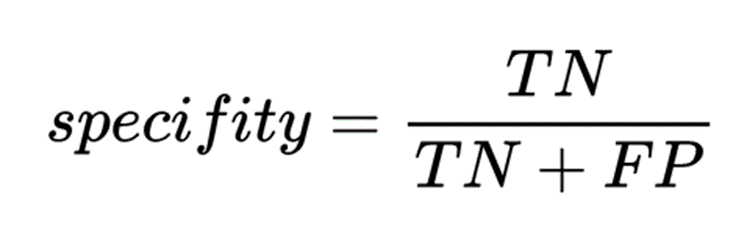
(1)

* **Czułość** (ang. recall) – parametr opisujący, ile pozytywnych przypadków zostało wykrytych ze wszystkich prawdziwie pozytywnych. Określa on jak dobrze dany model klasyfikuje jedną z klas. Obliczana jest ze wzoru (2):



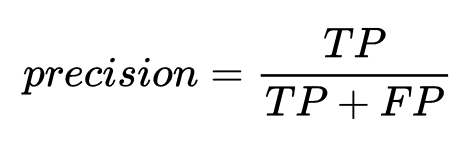
(2)

* **Swoistość** (ang. specifity) – parametr opisujący, ile negatywnych przypadków zostało wykrytych ze wszystkich prawdziwie negatywnych. Często przedstawiana jest w parze z czułością, gdyż daje informację o poprawności klasyfikacji drugiej z klas. Opisywana jest przez wzór (3):



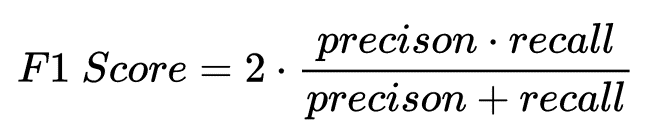
(3)

* **Precyzja** (ang. precision) – parametr opisujący, ile przypadków z uznanych przez klasyfikator za negatywne jest prawdziwie pozytywnych. Do jej wyznaczenia wykorzystuje się wzór (4):



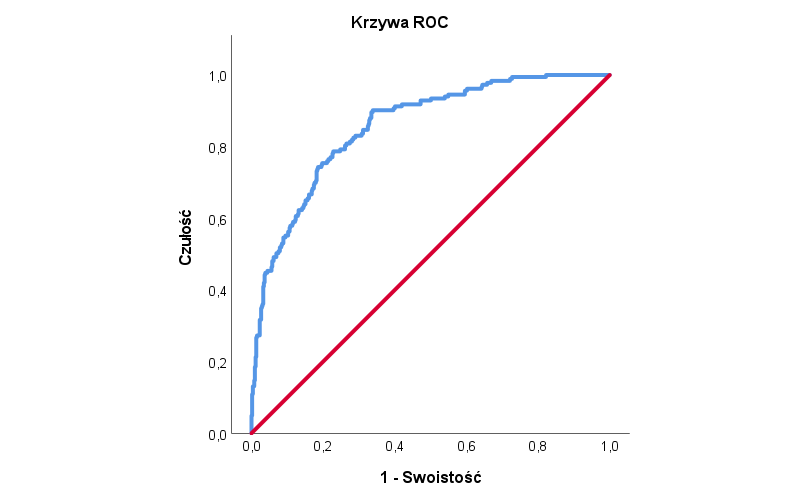
(4)

* **Miara F1** (ang. F1 Score) – średnia harmoniczna parametrów precyzji i czułości. Służy do szybkiego porównania metryk różnych modeli. Im bliższa wartości 100%, tym klasyfikator lepiej się sprawuje. Parametr opisywany jest przez wzór (5):



(5)

Dodatkowym sposobem prezentacji graficznej otrzymanych wartości metryk jest krzywa ROC (Rysunek 5). Wykres ilustruje związek pomiędzy czułością, a swoistością klasyfikatora, obliczając dla różnych punktów odcięcia ich wartości. Generuje on również nowy rodzaj metryki nazywany AUC (ang. Area Under Curve), który obliczamy jako pole pod krzywą ROC. Przyjmuje ona wartości w przedziale [0,1] – im większa, tym model jest bardziej trafny. Jest jedną z popularniejszych technik oceny modelu klasyfikacyjnego, jednak nie daje pełnej informacji o jej przebiegu. Z tego powodu analiza powinna odbywać się łącznie z porównaniem kształtu samego wykresu ROC.



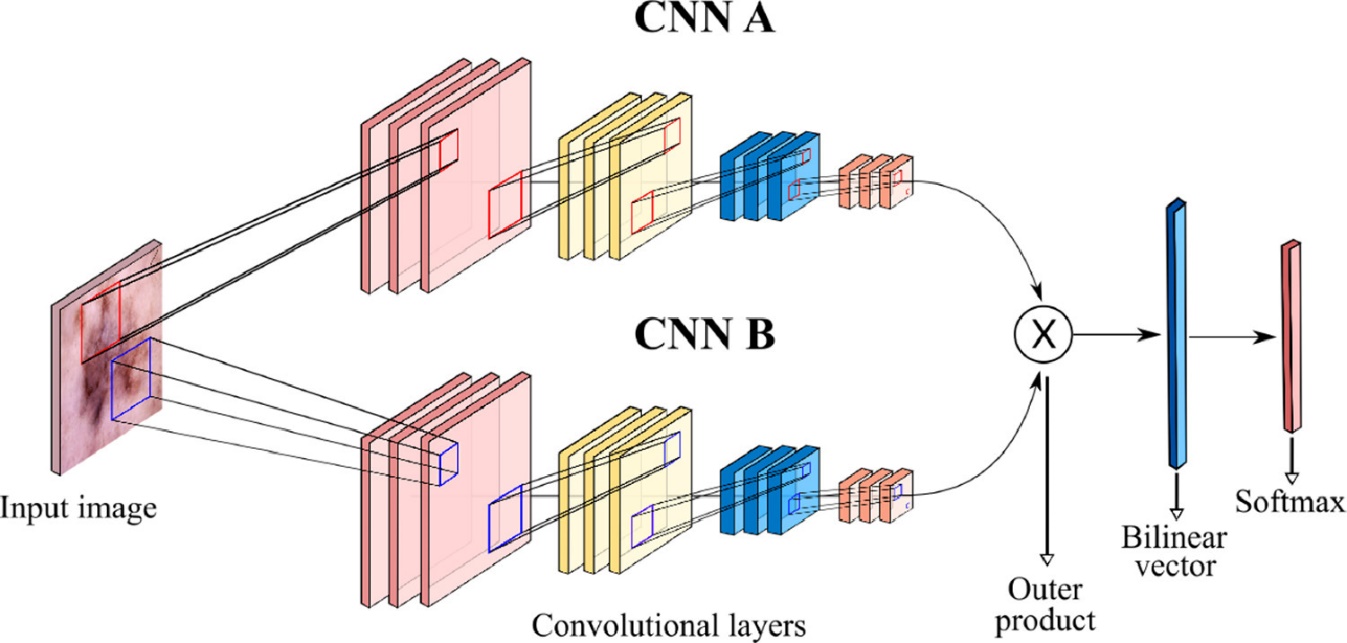
Rysunek 5. Przykładowy wykres krzywej ROC [13].

W tym celu na wykresie umieszcza się również przekątną, która opisuje przypadek, gdzie model dokonuje klasyfikacji poprawnej i błędnej z taka samą częstotliwością. Im wyżej wyliczona krzywa ROC oddalona jest od przekątnej, tym model osiągną znacznie lepsze wyniki.

### Modele sieci konwolucyjnych w medycynie

Uczenie maszynowe staje się coraz częściej wykorzystywanym narzędziem w medycynie. Możliwości zastosowania algorytmów są bardzo obszerne. Do przykładów obejmujących ten szeroki zakres aplikacji możemy zaliczyć modele pozwalające na klasyfikację typów choroby śródmiąższowej płuc na podstawie zdjęć z tomografii komputerowej. Podobne zastosowanie wykorzystano do detekcji zmian zapalnych w płucach wywołanych przez chorobę COVID-19. Sieci konwolucyjne okazały się przydatne również w detekcji raka piersi. Stanowi on najczęstszy typ nowotworu występujący u kobiet, dlatego tak istotne jest jego wczesne wykrycie. Wykorzystano do tego automatyczną segmentacją zdjęć mammograficznych, przeprowadzoną na podstawie zmian gęstości piersi. Innym podejściem natomiast była detekcja zmian na zdjęciach histopatologicznych węzłów wartowniczych. Imponujący jest również rozwój sieci takich jak AlexNet i GoogleNet, które osiągają dotychczas niespotykane wyniki pod względem wysokiej poprawności i niskiego wskaźnika błędu. Dowiedziono, że detekcja nowotworów płuc za pomocą sieci GoogleNet przeprowadzana jest z dokładnością na poziomie 89%, w momencie, gdy patolodzy osiągali wyniki jedynie na poziomie 70% [14].

Wymienione przykłady są tylko małym fragmentem możliwości uczenia głębokiego w medycynie. Można powiedzieć, że istnieją zastosowania dla każdego jej działu. Dermatologia nie jest pod tym względem wyjątkiem, a szczególnie jeżeli chodzi o klasyfikację zmian skórnych. Testowane są różne podejścia do tego zagadnienia, aby możliwe było uzyskanie jak najlepszych wyników. Do nich możemy zaliczyć bilinearną architekturę sieci zaproponowaną w artykule naukowym z 2021 roku [15]. Wykorzystano do tego dwa różne, pretrenowane modele połączone ze sobą na wejściu i wyjściu (Rysunek 6).



Rysunek 6. Schemat architektury sieci bilinearnej [15].

Posłużono się następującymi parami sieci: ResNet50 z VGG16 oraz ResNet50 z Xception. Uzyskane wyniki poddano ewaluacji wykorzystując podstawowe metryki, z których następnie przeliczono średnią ważoną dla wszystkich klas zmian skórnych. Najwyższe wartości otrzymane dla dokładności, precyzji oraz miary F1 były na poziomie 93%, różniąc się od siebie w granicach 0,1 punktu procentowego. Porównując z poprzednio stosowanymi metodami do klasyfikacji wieloklasowej, są to najwyższe wyniki dotychczas uzyskane.

Ciekawym podejściem do detekcji patologii w zmianach skórnych był algorytm zaproponowany w 2022 roku, który wykorzystuje klasyfikację zmian skórnych ze względu na ich symetrię [16]. Motywacją autorów był fakt, że jednym z dominujących czynników podczas diagnozy czerniaka jest właśnie symetria. Problem stanowią natomiast liczne techniki stosowane do ewaluacji tej cechy, często dające sprzeczne informacje. W artykule zaproponowano ujednorodnioną metodę klasyfikacji, wyróżniając trzy klasy zmian: kompletnie asymetryczne, symetryczne względem jednej osi oraz symetryczne względem dwóch osi. Wykorzystano w tym celu sieć przetrenowaną VGG16, ale z modyfikacjami zmniejszającymi jej stopień złożoności. Otrzymano następujące wartości metryk: precyzja 69%, czułość 62,7% oraz miara F1 64,5%. Natomiast jeżeli dokonano by klasyfikacji binarnej, łącząc klasy zmian symetrycznych w jedną, wyniki metryk wynosiłyby: precyzja 73,5%, czułość 72,2% i miara F1 71,8%. W ogólnym spojrzeniu nie są to w pełni zadowalające osiągnięcia, należy jednak wziąć pod uwagę, że we wcześniejszych metodach nie zastosowano sieci konwolucyjnych. Porównując otrzymane wyniki do klasycznych metod, znacząco przewyższają one dotychczas uzyskane wartości.

Ostatnim przykładem wartym wspomnienia jest metoda opublikowana w artykule z 2019 roku, w której najistotniejszym elementem było nowe podejście do regularyzacji procesu uczenia sieci konwolucyjnej. Dzięki temu osiągnięto wartości metryk przewyższające wyniki najnowocześniejszych algorytmów. Proponowany regularyzator bazował na odchyleniu standardowym macierzy wag klasyfikatora. Umożliwiło to sterowanie złożonością modelu, stosując kary, kiedy dochodziło do zbyt dużego rozrzutu wartości wag. Przeprowadzana klasyfikacja była binarna, przypisując etykiety zmian łagodnych i złośliwych dla różnych kombinacji zmian skórnych. Do przykładowych par możemy zaliczyć zmiany melanocytowe i czerniak, jak również rogowacenie łojotokowe i rak podstawnokomórkowy. Podczas ewaluacji metryk najwyższe wartości jakie otrzymano to dokładność na poziomie 97,5% oraz pole AUC równe 98%. Są to bardzo zadowalające wyniki, przedstawiające skuteczność stosowania regularyzacji nad jej brakiem.

## Środowisko i język programowania użyte w pracy

Implementację kodu projektu wykonano w Google Colaboratory będącym darmowym środowiskiem notatników Jupyter, obsługiwanym przez serwery Google. Jego podstawową zaletą jest zapisywanie danych w chmurze. Pozwala on również na udostępnianie użytkownikom klasycznych procesorów graficznych GPU jak i opracowanych specjalnie do uczenia maszynowego procesorów TPU. Dzięki niemu każdy ma szansę wykorzystania wszystkich możliwości notatników Jupyter bez konieczności ich instalacji na komputerze lokalnym.

Programy wykorzystane w projekcie napisano w języku Python 3. Jest on wysokopoziomowym językiem o dużej popularności, którą zawdzięcza dzięki wygodzie i spójności jego obsługi. Posiada on pełną bibliotekę standardową jak również bardzo szeroki wachlarz bibliotek dodatkowych. Podstawę obliczeń matematycznych na macierzach stanowiła biblioteka NumPy, zawierająca narzędzia do szybkiej i prostej obróbki danych. Jej rozwinięciem jest biblioteka Pandas, która skupia się na danych relacyjnych. Szczególnie przydatna była podczas wstępnej analizy danych w trakcie projektowania sieci neuronowych. Natomiast do wizualizacji omawianych danych posłużył moduł Matplotlib, umożliwiający tworzenie wykresów słupkowych, jak i moduł Seaborn, poszerzający możliwości pierwszego o bardziej zaawansowane techniki wizualizacji – na przykład mapy cieplne.

Klasyfikacja przeprowadzona w projekcie opiera się na możliwościach biblioteki Keras. Zbudowana jest ona na bazie biblioteki TensorFlow i daje szerokie możliwości tworzenia modeli uczenia maszynowego, dzięki ułatwionemu dostępowi do podstawowych elementów budujących sieć. Realizuje wiele rodzajów warstw, funkcji aktywacyjnych i optymalizatorów potrzebnych do zbudowania funkcjonalnej sieci neuronowej. Posiada również wsparcie dla sieci konwolucyjnych i rekurencyjnych.

# Rozdział 2. Implementacja algorytmów klasyfikacyjnych



## Opis wybranego zbioru zdjęć dermatoskopowych

Do zrealizowania założeń projektu wykorzystano zbiór danych HAM10000 (Human Against Machine with 10000 training images). Opublikowany został w 2018 roku na stronie Harvard Dataverse Repository, będącej internetowym repozytorium dla danych badawczych. Autorem jest profesor Philipp Tschandl działający na rzecz Vienna Dermatologic Imaging Research będącego częścią Wydziału Dermatologii Uniwersytety Medycznego w Wiedniu. Jako pierwsze zastosowanie zbioru HAM10000 możemy określić jego udział w konkursie ISIC 2018 Challenge (ang. Intrenational Skin Imaging Collaboration), gdzie posłużył za dane treningowe dla zadania klasyfikacji zmian chorobowych skóry.

Powodem do powstania omawianego zbioru był wcześniejszy brak zróżnicowanych i licznych danych w zakresie zmian skórnych. Małe zbiory są jednym z najczęstszych przyczyn niepowodzeń podczas trenowania sieci neuronowych, przeznaczonych do automatycznej diagnozy. Dlatego też HAM10000 składa się z aż 10015 zdjęć dermatoskopowych, które zostały pozyskane z różnych populacji jak i źródeł. Spośród nich wyróżniono 7 klas odzwierciedlających najistotniejsze w dermatologii rodzaje zmian na skórze. Zaliczyć do nich możemy:

* **Actinic Keratoses and Intraepithelial Carcinoma (akiec)** – rogowacenie słoneczne, nazywane również starczym oraz Choroba Bowena;
* **Basal Cell Carcinoma (bcc)** – rak podstawnokomórkowy skóry;
* **Benign Keratosis-like Lesion** **(bkl)** – zmiany podobne do rogowacenia o charakterze niezłośliwym takie jak: plama soczewicowata słoneczna, brodawka łojotokowa i rogowacenie liszajowate;
* **Dermatofibroma (df)** – włókniak twardy;
* **Melanoma (mel)** – czerniak;
* **Melanocytic Nevus (nv)** – znamiona melanocytowe o charakterze niezłośliwym;
* **Vascular Lesion** **(vasc)** – zmiany naczyniowe takie jak: naczyniak, rogowiec krwawy, ziarniak ropotwórczy i krwiak podskórny.

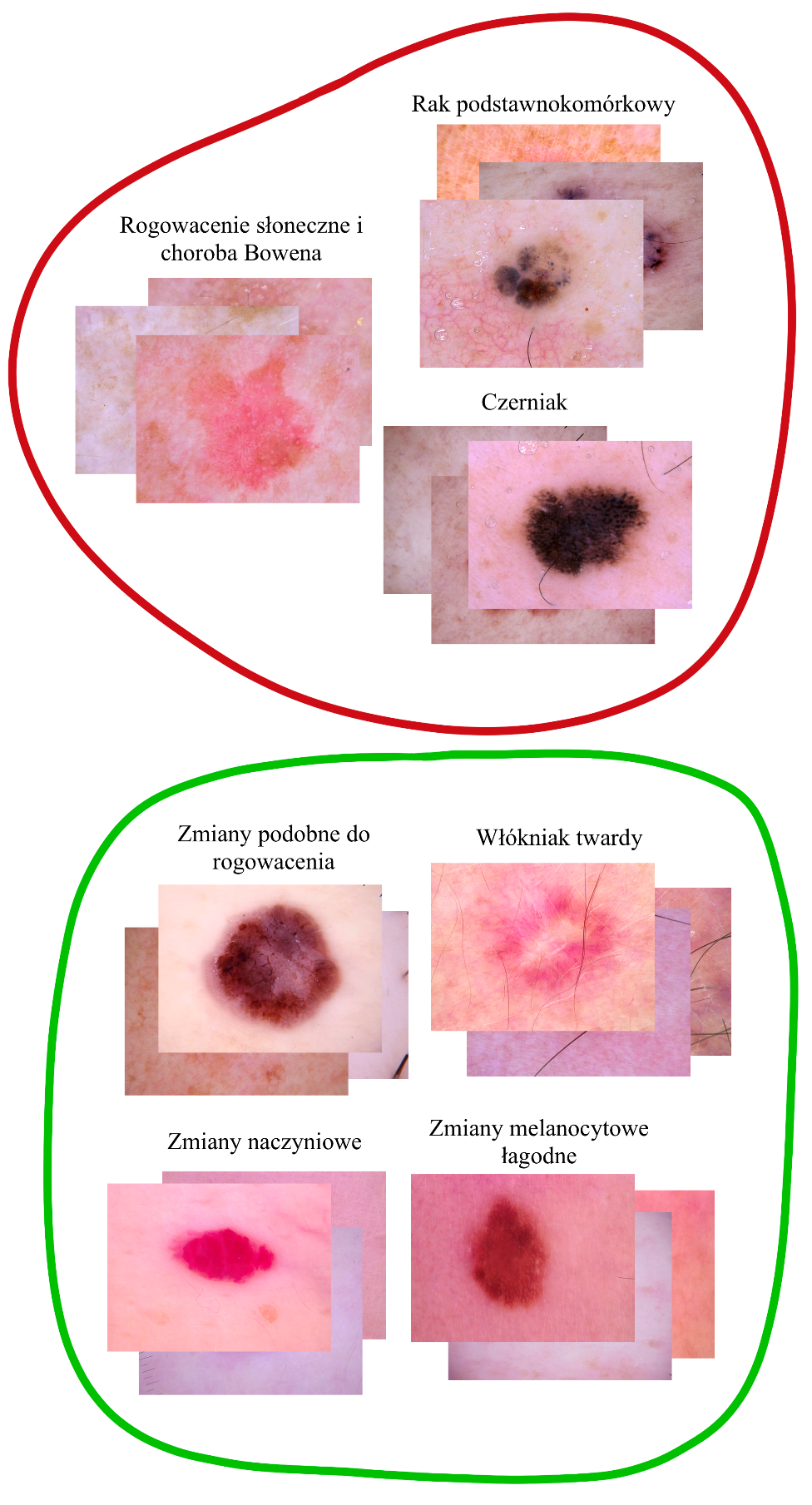
Aby możliwa była pełna analiza zbioru, dołączony do niego został również plik HAM10000\_metadata.csv zawierający istotne informacje na temat każdego zdjęcia. W jego skład wchodzi 7 kolumn opisujących dane takie jak: identyfikator zmiany **(***lesion\_id*), identyfikator zdjęcia **(***image\_id*), rodzaj zmiany skórnej (*dx*), jakim sposobem zdiagnozowano rodzaj zmiany **(***dx\_type*), wiek (*age*) i płeć (*sex*) pacjenta oraz lokalizację zmiany na ciele (*localization*). Ponieważ w zbiorze pojawiają się różne zdjęcia tej samej zmiany, dokonano rozdzielenia na dwa identyfikatory. Przykładowe wartości zamieszczono w Tabeli 2.

Tabela 2. Pięć pierwszych wartości zawartych w pliku HAM10000\_metadata.csv.

Obraz zawierający stół

Opis wygenerowany automatycznie

Ze względu na to, że w zbiorze pojawiają się różne klasy zmian skórnych, przeprowadzono dodatkowe zgrupowanie omawianych danych na dwa skupienia. Wytypowano dwie grupy: zmiany złośliwe oraz zmiany łagodne, co nawiązuje do charakteru wybranych zmian skórnych. Na Rysunku 7 zilustrowano podział obiektów. W przypadku klasy *akiec* zawierającej zdjęcia zmiany przedrakowego rogowacenia słonecznego i choroby Bowena zadecydowano o przyłączeniu jej do grupy zmian złośliwych. Podział ten pozwala uprościć zagadnienie klasyfikacji, gdyż podczas diagnozy najważniejszy jest fakt, czy dana zmiana stanowi zagrożenie dla organizmu pacjenta.



Zmiany złośliwe

Zmiany łagodne

Rysunek 7. Podział obiektów w zbiorze na zmiany złośliwe i łagodne.

## Analiza danych dołączonych do zbioru HAM10000

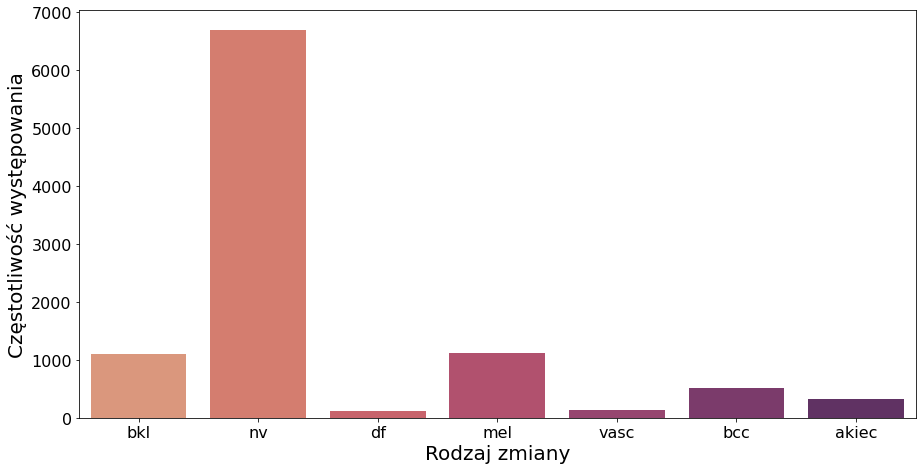
Przed podjęciem zagadnienia klasyfikacji przeprowadzono wstępną analizę wykorzystywanego zbioru danych. W tym celu posłużono się możliwościami pakietu Pandas z zakresu analizy danych jak i bibliotekami Matplotlib i Seaborn do ich wizualizacji.

W pierwszym kroku rozpoczęto od odczytania informacji dla kolumn zawartych w pliku CSV. Otrzymane wyniki zamieszczono w Tabeli 3. Na jej podstawie dowiadujemy się, że w zbiorze zawarte jest 7470 unikatowych przypadków co oznacza, że 26,4% wszystkich zdjęć przedstawia tą samą zmianę. Ponadto okazało się, że w kolumnie*age* istnieje 57 brakujących danych. Jest to istotna informacją, na którą trzeba zwrócić uwagę podczas dalszych kroków.

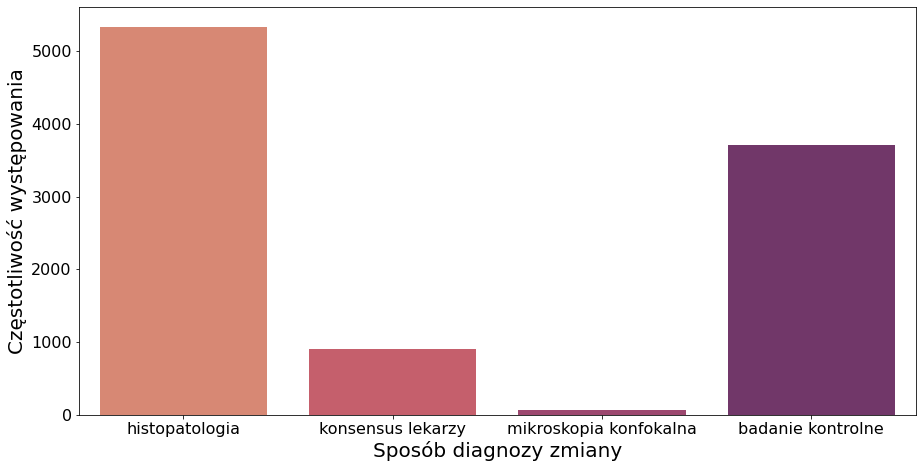
Tabela 3. Zestawienie wybranych informacji o zbiorze danych.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | lesion\_id | image\_id | dx | dx\_type | age | sex | localization |
| Liczebność | 10015 | 10015 | 10015 | 10015 | 9958 | 10015 | 10015 |
| Wartości unikatowe | 7470 | 10015 | 7 | 4 | - | 3 | 15 |
| Typ danych | object | object | object | object | float64 | object | object |
| Liczba brakujących | 0 | 0 | 0 | 0 | 57 | 0 | 0 |

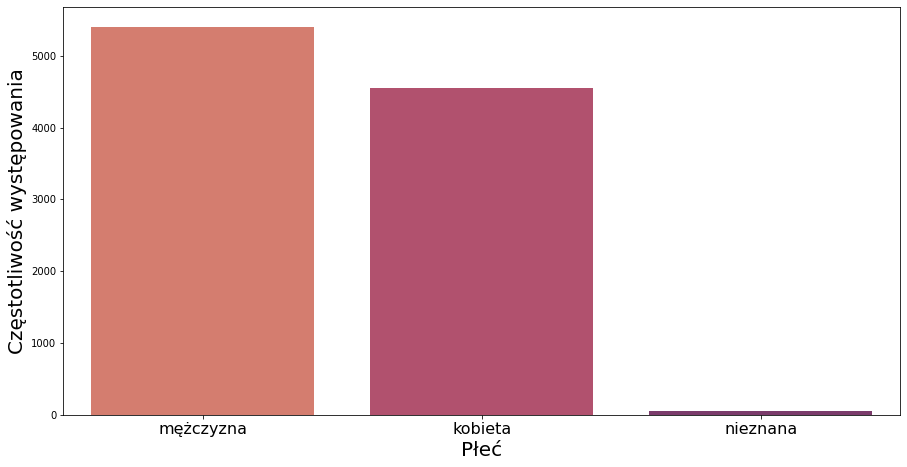
W następnym etapie przeprowadzono analizę liczebności danych. Jest to zagadnienie stanowiące podstawę przeprowadzania klasyfikacji na danych medycznych, ponieważ ze względu na ich naturę, bardzo często występuje niezrównoważenie zbioru. W tym celu zilustrowano liczebności poszczególnych wariantów w omawianych cechach zbioru. Za pomocą przedstawionych danych można rozwiązywać różne zadania klasyfikacyjne, gdyż każdy z parametrów macierzy (z wyjątkiem kolumn *lesion\_id* i *image\_id*) stanowi zmienna odpowiedź, którą można interpretować niezależnie. Ze względu na fakt, że w przypadku danych medycznych niektóre z cech są ze sobą silnie powiązane, zilustrowano również liczebność klas zmian skórnych w zależności od wybranych cech zbioru. Wszystkie wykresy zamieszczono na Rysunkach 8 - 15.



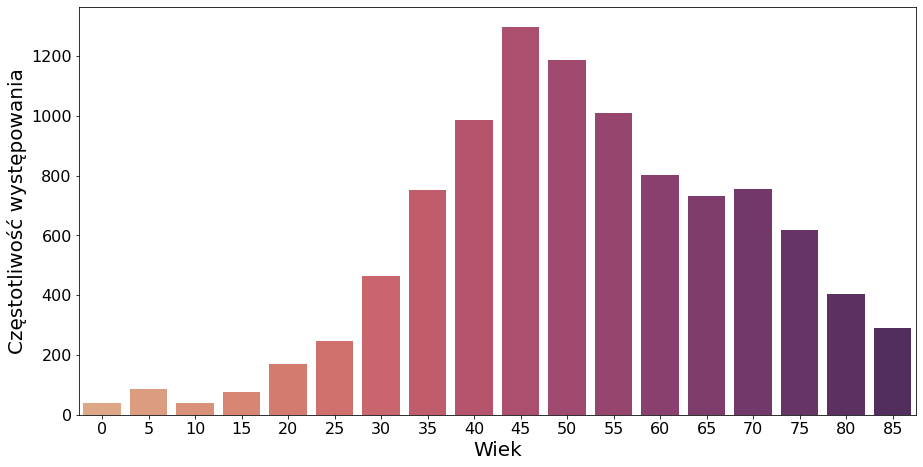
Rysunek 8. Wykres słupkowy rozkładu liczebności klas zmian skórnych.



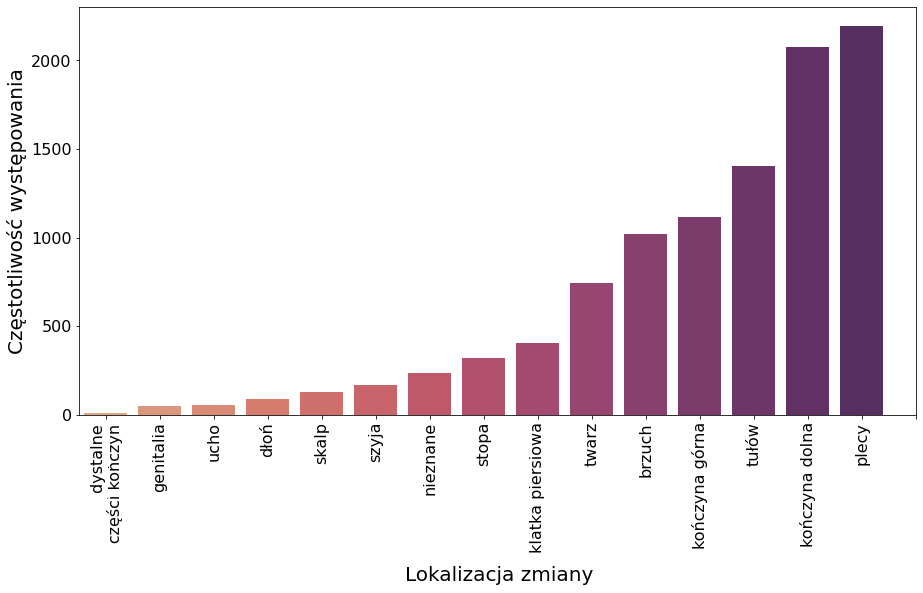
Rysunek 9. Wykres słupkowy rozkładu liczebności klas technik diagnozy.



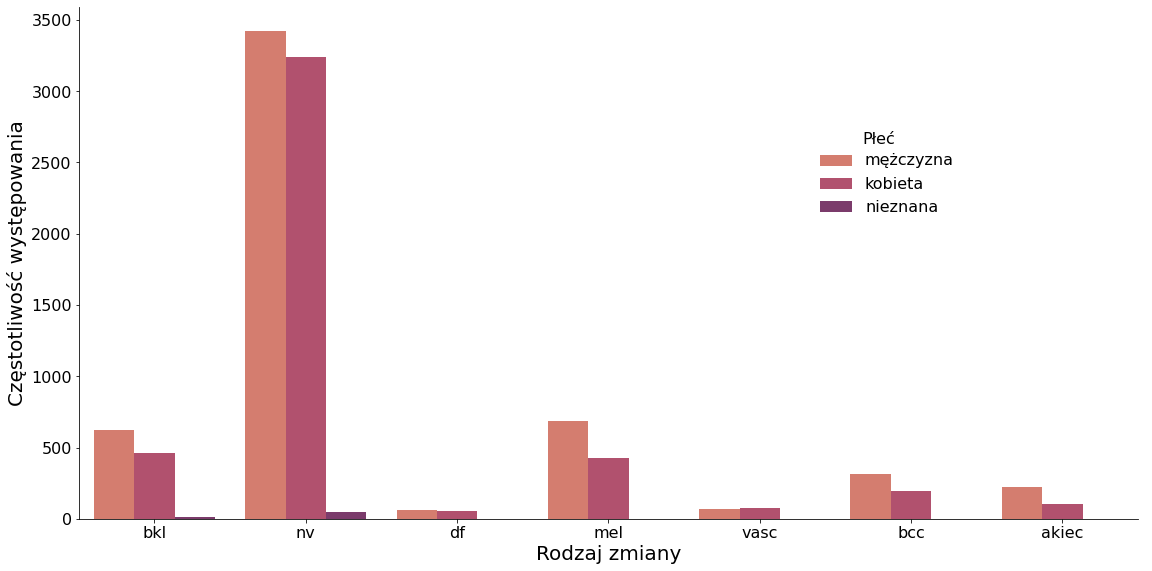
Rysunek 10. Wykres słupkowy rozkładu liczebności klas płci wśród pacjentów.



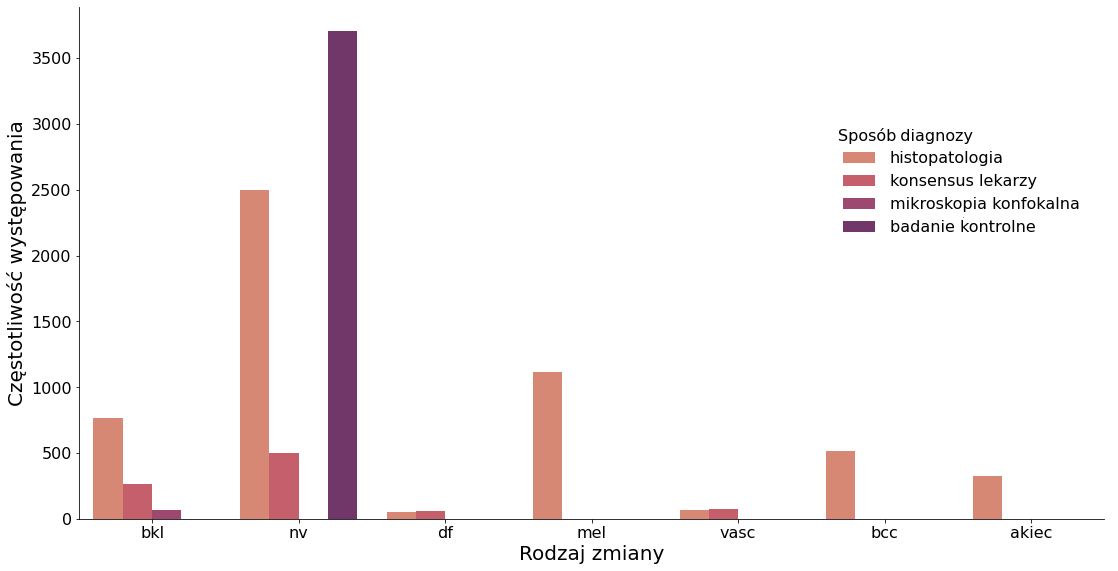
Rysunek 11. Wykres słupkowy rozkładu liczebności klas wieku wśród pacjentów.



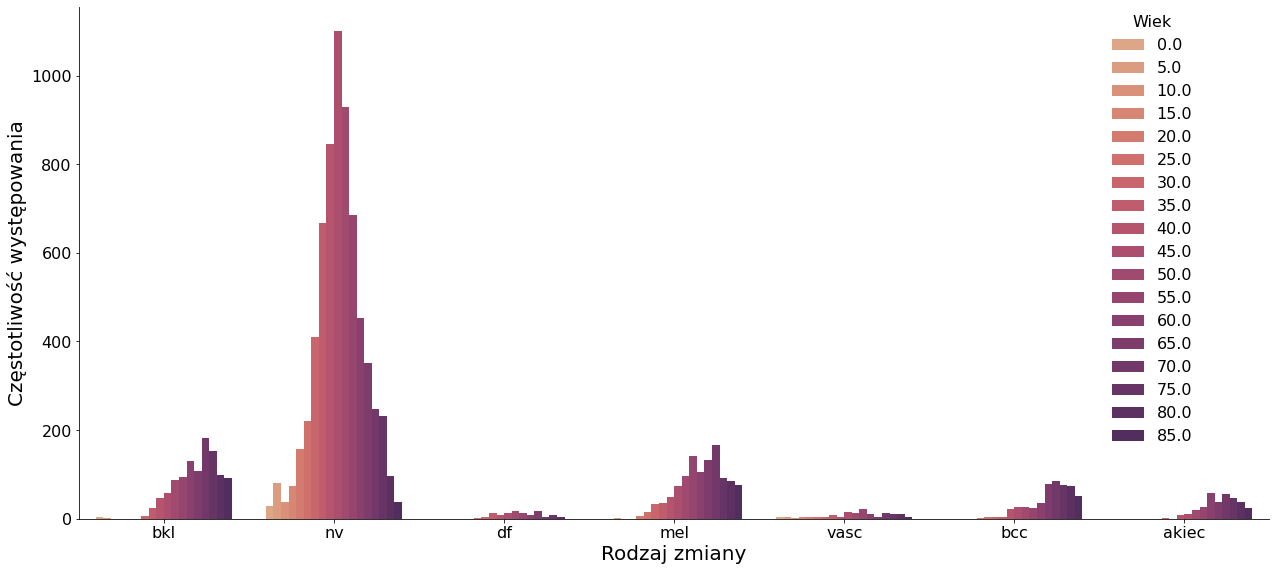
Rysunek 12. Wykres słupkowy rozkładu liczebności klas lokalizacji zmian skórnych.



Rysunek 13. Wykres słupkowy rozkładu liczebności klas płci w zależności od klas zmian skórnych.



Rysunek 14. Wykres słupkowy rozkładu liczebności klas technik diagnozy w zależności od klas zmian skórnych.



Rysunek 15. Wykres słupkowy rozkładu liczebności klas wieku pacjentów w zależności od klas zmian skórnych.

Na podstawie przedstawionych rysunków jednoznacznie stwierdzono o dużym niezrównoważeniu zbioru. Przede wszystkim największe znaczenie miała przewaga klasy *Melanocytic Nevus*(nv), która stanowiła, aż 70% omawianych danych.  Natomiast klasy *Vascular Lesion* (vasc) i *Dermatofibroma* (df) były najmniej liczne, zawierając około 1% wszystkich zdjęć. Spośród zmian złośliwych najliczniejsza była klasa *Melanoma* (mel), która stanowiła 67% zdjęć w tej grupie.

Zebrane dane pochodziły w 55% od pacjentów płci męskiej, nie stwierdzono znaczącej przewagi żadnej z płci, z wyjątkiem obrazów czerniaka i rogowacenia słonecznego. Dla klas *Melanoma* (mel) oraz *Actinic kerstoses* (akiec) zauważono przewagę mężczyzn o prawie 50%. Informacja ta jest szczególnie istotna ze względu na złośliwy charakter tych zmian, a konkretniej w odniesieniu do czerniaka. Oznacza to, że w grupie badawczej mężczyźni mieli większe ryzyko zachorowania na omawiany nowotwór skóry.

Duże różnice w liczebności widoczne były również w przypadku technik diagnozy zmian, z przewagą histopatologii i badań kontrolnych. Co ciekawe analizując wykres z Rysunku 14 odznacza się bardzo wysoka liczba badań kontrolnych dla zmian *Melanocytic Nevus*(nv), niezauważalna dla innych klas. Zaobserwowano również fakt, że trzy klasy należące do zmian złośliwych, czyli: *Melanoma* (mel), *Basal Cell Carcinoma* (bcc) oraz *Actinic kerstoses* (akiec) były wyłącznie diagnozowane na podstawie histopatologii.

Wartym uwagi był także rozkład wieku w zależności od typu zmiany, w którym zauważono przewagę badanych w średnim i starczym wieku. Zilustrowany na Rysunku 15, wizualizuje pewien trend obecny dla wszystkich klas. Mianowicie dla każdej z nich liczba próbek zaczęła podwajać się po bramce 25 lat, osiągając maksimum dla około 50 roku życia. Po tym liczba przypadków utrzymywała się na stałym poziomie lub malała z mniejszym tempem. Świadczy to o zwiększonej zachorowalności wraz ze starszym wiekiem wśród badanej populacji.

W przypadku lokalizacji zmian na ciele, największą przewagę miały miejsca takie jak plecy, tułów, kończyny górne i dolne. Prawdopodobnie związane to było ze stosunkowo większą powierzchnią skóry w tych obszarach niż w przypadku innych kategorii. Są to też fragmenty skóry najczęściej wystawione na działanie promieniowania słonecznego, co potwierdza jego wpływ na zwiększoną zachorowalność.

## Przygotowanie i równoważenie zbioru

Analiza danych pozwoliła na wytypowanie najistotniejszych wad zbioru HAM10000, które mogą być przeszkodami podczas jego wykorzystania do treningu sieci. W celu uproszczenia implementowanych modeli klasyfikacyjnych, podjęto decyzję o przeprowadzeniu treningu sieci na podstawie samych zdjęć. Powodem tego jest szeroki zakres informacji dołączony do zbioru, które wykazują duże niezrównoważenie w obrębie każdej cechy. Dzięki temu, ominięto problem brakujących danych z pliku CSV i również pozwoliło to na ograniczenie niezrównoważenia zbioru tylko do klas zmian skórnych.

Aby lepiej wpasowywać się w ramy założeń projektowych, wybrany zbiór danych został poddany przetwarzaniu, przystosowującym go do zagadnienia klasyfikacji binarnej. Ponieważ, oryginalnie zaprojektowano go z myślą o klasyfikacji wieloklasowej, konieczne było wydzielenie dwóch klas. Dokonano następującego podziału: **MELANOMA** – grupa, którą stanowiły zmiany określane mianem czerniaka, czyli klasa *Melanoma* (*mel*) o liczebności 1113 zdjęć oraz **NOT MELANOMA** – składająca się z pozostałych rodzajów zmian skórnych, licząca 8902 zdjęć. Zdecydowano o takim podziale, ze względu na fakt, że klasa *Melanoma* (*mel*) stanowi największą część zmian złośliwych w zbiorze. Również reprezentuje ona najpoważniejszy typ nowotworu skóry, dlatego też jego wczesne wykrycie stanowi priorytet podczas diagnozy.

W zaproponowanym podziale nadal widnieje duża różnica w liczebności klas – grupa MELANOMA jest prawie 8-krotnie mniejsza niż NOTMELANOMA. Jest to powód dla którego w dalszej części projektu posłużono się zbiorem dostępnym na stronie Kaggle, nazwanym jedynie „*melanoma*” [17]. Dokonany w nim jest identyczny podział zdjęć zbioru HAM10000 na dwie klasy, jaki zaproponowano w projekcie. Jest on rozwiązaniem na niezrównoważenie pomiędzy wybranymi klasami, gdyż wyrównuje on ich liczebność. Zwiększona zostaje liczba zdjęć w grupie MELANOMA, dzięki procesowi augmentacji. Polega on na pobraniu oryginalnego zdjęcia zmiany skórnej, a następnie przeprowadzeniu szeregu transformacji pozwalających na wytworzenie syntetycznych zdjęć. Są one ustalane przypadkowo, z przypisanym prawdopodobieństwem. Możemy do nich zaliczyć modyfikacje takie jak: zbliżenie obrazu, przechylenie o zadany kąt, jak i obrót w pionie lub poziomie. Przykład wygenerowanych w ten sposób zdjęć zamieszczono na Rysunku 16.

.

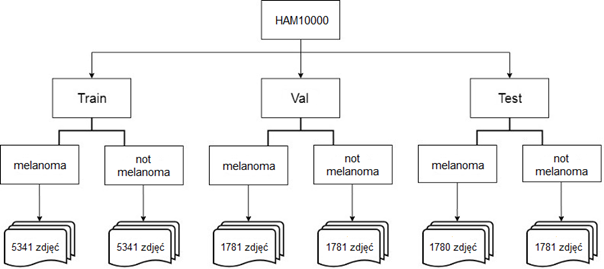
Obraz zawierający kwadrat

Opis wygenerowany automatycznie

Rysunek 16. Możliwości syntetycznego tworzenia zdjęć przez augmentację [18].

Po zastosowaniu tej techniki, liczbę obrazów w zbiorze zwiększono do 17 805 i otrzymano następujące liczebności omawianych grup: MELANOMA – 8 903 zdjęć oraz NOT MELANOMA – 8 902 zdjęć. Poza uzyskaniem zrównoważonych klas, przeprowadzone modyfikacje wpływają na generalizację uczącego się modelu. Jest to zdolność do uogólniania zdobytej wiedzy, dla przykładu, kiedy obraz przedstawia czerniaka nachylonego w prawo, to w przypadku, gdy będzie on nachylony w lewo, model nadal będzie go klasyfikował jako czerniak.

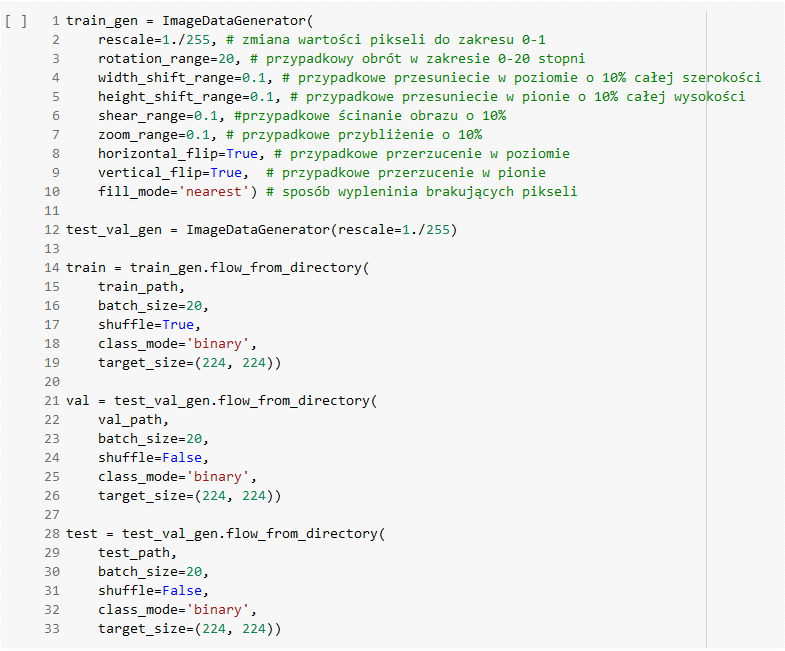
W kolejnym etapie wyodrębniono trzy zbiory, w których zamieszczono zdjęcia z obu wytypowanych klas zmian skórnych. Podziału dokonano w stosunku 60%-20%-20% wszystkich dostępnych zdjęć. Pierwszym był zbiór uczący „Train”, który został wykorzystany do treningu sieci neuronowej i zawierał w sobie 10 682 zdjęć. Następnym był zbiór walidacyjny „Val”, o liczebności 3 562 zdjęć, służący do ewaluacji przebiegu uczenia modelu, pozwalając na dopracowanie hiperparametrów w czasie rzeczywistym. Natomiast końcowym był zbiór testowy „Test” zawierający 3 561 zdjęć, stanowiący fragment danych nie wykorzystywany podczas treningu sieci, dlatego też ewaluacja modelu za jego pomocą służyła obiektywnej ocenie jakości algorytmu. Ostateczna architektura zbioru wykorzystanego w dalszych etapach projektu została przedstawiona na Rysunku 17.



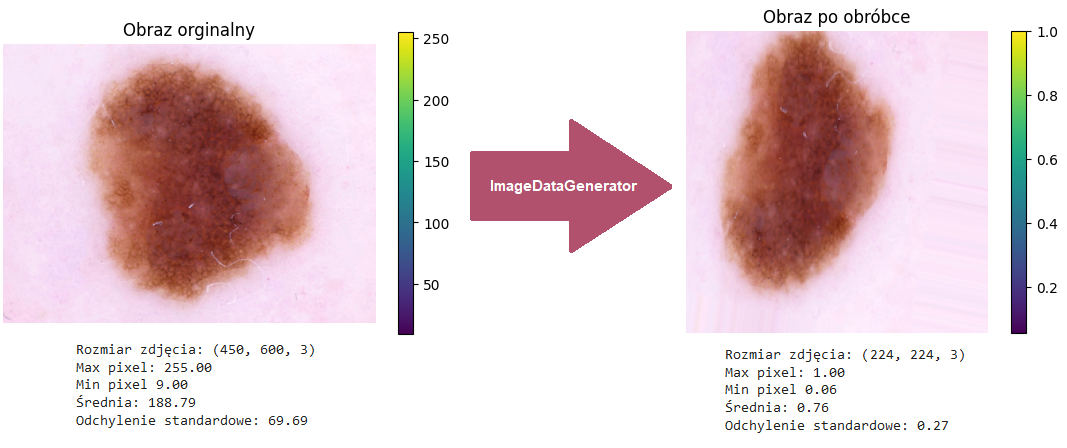
Rysunek 17. Architektura zbioru danych wykorzystanego do klasyfikacji.

Zilustrowany podział na podzbiory nie został wykonany bez powodu, ponieważ był on konieczny do wygenerowania danych w postaci tensorów. Umożliwiła to klasa ImageDataGenerator z biblioteki Keras, która tworzy obiekt generatora, o zadanych parametrach. Generator ten odpowiada za pobranie określonej liczby obrazów (ang. batch), następnie przeprowadzenie przypadkowych modyfikacji i wygenerowanie zdjęć, które posłużą w treningu sieci. Do przykładowych transformacji zastosowanych w projekcie możemy zaliczyć: obrót o zadany kąt, przesunięcie lub przerzut w pionie i poziomie, normalizację wartości pikseli czy przybliżenie lub oddalenie obrazu. Ten typ augmentacji danych różni się od zastosowanego do zwiększenia zbioru MELANOMA tym, że nowo powstałe zdjęcia nie są zapisywane, a jedynie istnieją tymczasowo na potrzeby uczenia modelu - otrzymane modyfikacje są różne dla każdej ich iteracji. Na rysunku 18 zademonstrowano fragment kodu, w którym widnieje implementacja omawianego generatora, wraz z zastosowaniem metody flow\_from\_directory pozwalającej na szybkie pobranie zdjęć z uprzednio przygotowanych folderów.

Po przeprowadzonej augmentacji zdjęć, końcowy efekt jaki uzyskano to przekształcone obrazy o rozmiarze 224 × 224 pikseli i trzech kanałach barwnych. Aby czas potrzebny na trening implementowanej sieci nie był nadmiernie wydłużony, przeprowadzono normalizację wartości pikseli do zakresu 0.0 – 1.0.



Rysunek 18. Fragment kodu zawierający implementację generatora danych.

****

Rysunek 19. Wizualizacja zmian w zdjęciach po modyfikacjach klasy ImageDataGenerator.

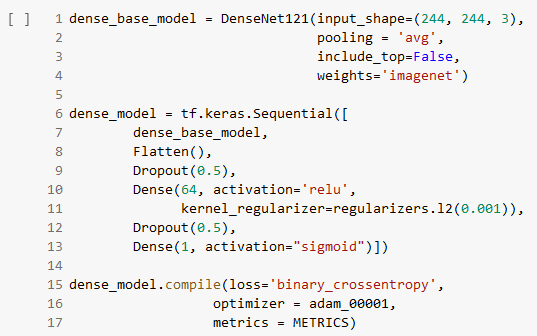
## Przebieg klasyfikacji zmian skórnych

W tej części projektu opisany został krok po kroku, proces tworzenia modeli klasyfikacyjnych do detekcji klas MELANOMA i NOT MELANOMA, które utworzono w trakcie przygotowywania danych. Zdecydowano, o implementacji sieci w wykorzystaniem techniki transferlearningu, czyli w dosłownym tłumaczeniu transferu wiedzy. Jednak najczęściej używa się określenia: uczenie transferowe. Polega ona na użyciu już wytrenowanych sieci neuronowych jako bazy do własnego modelu. W przypadku klasyfikacji obrazów, najpopularniejsze pretrenowane sieci są uczone na zbiorze ImageNet, ponieważ zawiera on 1000 różnych klas zdjęć. Dzięki temu możliwe jest łatwe dopasowanie wiedzy modeli pretrenowanych do klasyfikacji wybranego zagadnienia. W literaturze bardzo często spotyka się modele wykorzystujące transfer wiedzy, a przykładem na to są omówione w Rozdziale 1 zastosowania sieci konwolucyjnych w medycynie.

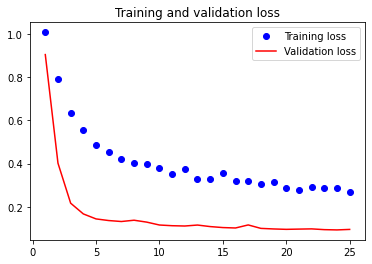
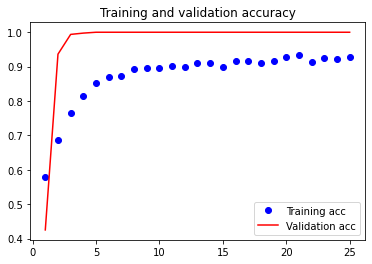
W niniejszej pracy postanowiono dokonać implementacji pięciu modeli, posługujących się różnymi sieciami pretrenowanymi. Zaliczyć do nich możemy sieci: DenseNet121 (428 warstw), ResNet50 (176 warstw), EfficientNetB7 (815 warstw), InceptionV3 (312 warstw) oraz VGG16 (20 warstw). Proponowane modele cechują się różnym stopniem złożoności, dlatego też po przeprowadzonej klasyfikacji, oczekiwane były różnorodne wyniki pozwalające wyłonić najlepszą architekturę do detekcji zmian skórnych. Na tym etapie wytypowano również podstawowe założenia, które implementowane sieci powinny spełniać. Aby uznać otrzymane modele za adekwatne, ich dokładność na zbiorze testowym powinna wynosić co najmniej 90%. Równocześnie nie powinno wystąpić zjawisko overfittingu, czyli przeuczenia na zbiorze treningowym. Dlatego też bardzo istotna była wizualizacja zmian wartości metryk takich jak dokładność (ang. accuracy) i starta (ang. loss) otrzymanych na zbiorze treningowym i walidacyjnym w czasie uczenia, które pozwoliły na szczegółową analizę przebiegu szkolenia modelu.

### Optymalizacja wstępnego algorytmu

Pierwsze próby przeprowadzono z użyciem modelu DenseNet121, aby przetestować wstępną architekturę sieci. Zaproponowano model klasyfikacyjny, w którym bazę stanowiła wspomniana sieć. Na jej wierzch dołożono dodatkowe warstwy, zaczynając od warstwy spłaszczającej Flatten(). Następnie wykorzystano warstwę Dropout(), o hiperparametrze porzucenia równym 0.5, co miało na celu zabezpieczenie sieci przed przeuczeniem. Kolejno dołożono warstwę Dense(), w której zastosowano funkcję aktywacji „ReLu” oraz regularyzację typu L2. Następnie ponowiono warstwę Dropout() i zakończono warstwą Dense(), która zwracała prawdopodobieństwo przynależności danego zdjęcia do jednej z dwóch klas. Przedstawiony model skompilowano dla binarnej funkcji starty, optymalizatora Adam o parametrze learning\_rate równym 0,0001 oraz dla wybranych mierników jakości wśród których znalazły się: dokładność, precyzja, czułość oraz pole AUC. Całość omówionych kroków przedstawiono w formie kodu na Rysunku 20. Po wytrenowaniu modelu na zbiorze treningowym i walidacyjnym, wykreślono przebieg zmian wartości dokładności i straty w zależności od liczby epok. Otrzymane grafy zilustrowano na Rysunku 21.



Rysunek 20. Fragment kodu z definicją architektury modelu w pierwszym podejściu.

****

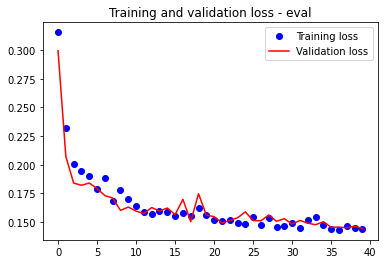
Rysunek 21. Wykresy zmian dokładności i funkcji straty modelu w pierwszym podejściu.

Na podstawie otrzymanych wykresów stwierdzono, że zaszło rzadko spotykane zjawisko underfittingu. Rozpoznawalną cechą dla tego problemu są wyższe wartości metryk dla zbioru walidacyjnego niż treningowego. Oznacza to, że model ma trudność z dopasowaniem danych treningowych oraz generalizacją nowych danych. Aby rozwiązać ten problem zdecydowano się na zamrożenie początkowych warstw modelu pretrenowanego i przeprowadzenie szkolenia tylko na ostatnich warstwach. Wpływ na underfitting ma również zbyt wysoka regularyzacja procesu uczenia, dlatego zrezygnowano z zastosowania L2 w pierwszej warstwie Dense().

Po głębszej analizie problemu, stwierdzono również o błędnym sposobie wykreślania omawianych wykresów oceny jakości modeli podczas treningu. Wynika to z faktu, że wcześniej wykorzystywana funkcja history() w metodzie model.fit() – która zapisuje obliczone dokładności i starty podczas uczenia – inaczej liczy je dla obu omawianych zbiorów. W przypadku treningowego, metryki liczone są na bieżąco w trakcie uczenia, dlatego też aktywne są mechanizmy regularyzacji, takie jak np. Dropout(). Jest to przyczyna zaniżenia otrzymanych wyników. Natomiast podczas fazy walidacyjnej, model przechodzi w tryb, w którym regularyzacja jest wyłączona.

Z tego też powodu zdecydowano się na inne podejście. Mianowicie po każdej epoce uczenia, posłużono się metodą model.evaluate() do wyliczenia dokładności i straty na zbiorze treningowym i walidacyjnym. Dzięki temu uzyskano obiektywne wyniki dla sprawności modelu po każdej iteracji uczenia. Nową architekturę sieci z poprawionymi hiperparametrami poddano uczeniu i na Rysunku 22 zilustrowano nowo otrzymane wykresy dokładności i starty, o znacznie lepszym dopasowaniu.

Obraz zawierający tekst

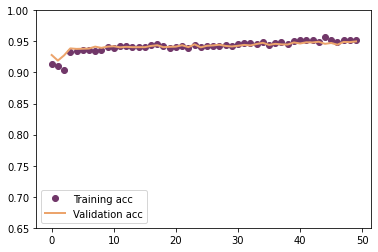
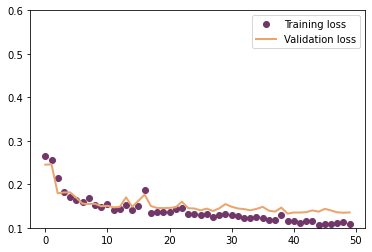
Opis wygenerowany automatycznie

Rysunek 22. Wykresy zmian dokładności i starty modelu po wprowadzonych zmianach.

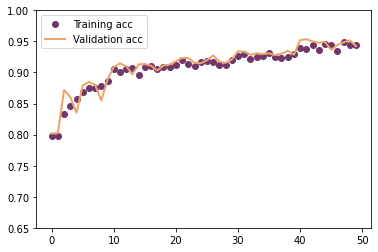
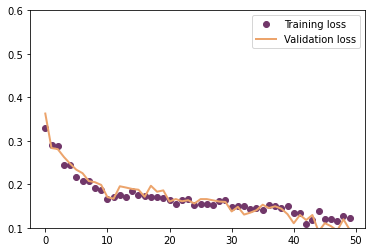
Niestety dużą wadą tego podejścia było znaczące wydłużenie czasu całego procesu, ponieważ metoda model.fit() musiała być ponawiana dla każdej kolejnej epoki. Niemniej zdecydowano się w dalszych etapach przyjąć przedstawiony algorytm.

### Klasyfikacja przy użyciu sieci pretrenowanych

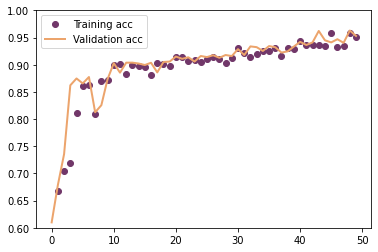
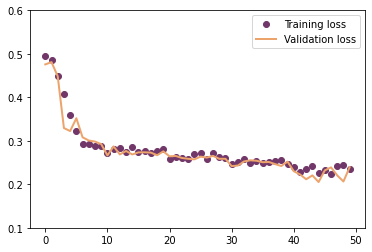
W dalszej części dokonano końcowej optymalizacji proponowanej architektury. Wartość hiperparametru prawdopodobieństwa porzucenia w Dropout() została zmniejszona do 0.3, oraz zwiększono parametr learning\_rate na 0.001 w optymalizatorze Adam. Wprowadzone zmiany wpłynęły korzystnie na zniwelowanie zjawiska underfitingu. Dodatkowo dla każdego modelu pretrenowanego dokonano odmrożenia ostatniego bloku konwolucyjnego. Mianowicie dla modelu DenseNet121 było to 10 ostatnich warstw, ResNet50 11 warstw, EfficientNetB7 19 warstw, InceptionV3 10 warstw, a dla modelu z VGG16 nie dokonano zamrożenia warstw ze względu na jego małą złożoność. Wszystkie sieci zostały przetrenowane przez 50 epok, z liczbą kroków równą 100 dla zbioru treningowego i 20 dla walidacyjnego. Posłużono się również funkcjami zwrotnymi dostępnymi w bibliotece Keras, takimi jak: EarlyStopping – funkcja służąca do zapobiegania przeuczeniu modelu, zatrzymując jego szkolenie po osiągnięciu docelowych parametrów oraz ReduceLROnPlateau – funkcja dokonująca zmniejszenia wartości parametru learning\_rate, kiedy wartości wybranej metryki przestają rosnąć. Następnie za pomocą metody model.evaluate() po każdej epoce obliczono wartości dokładności i starty dla zbioru treningowego i walidacyjnego. Otrzymane wyniki przedstawiono w formie wykresów, które wyskalowano, aby osie zawierały się w tym samym zakresie. Zilustrowano je na Rysunkach 23-27.

**** ****

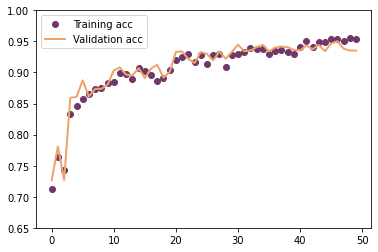
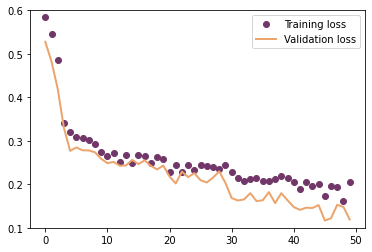
Rysunek 23. Wykresy zmian dokładności i starty modelu z DenseNet121.

**** ****

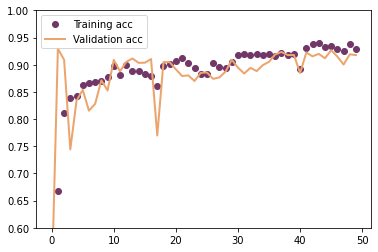
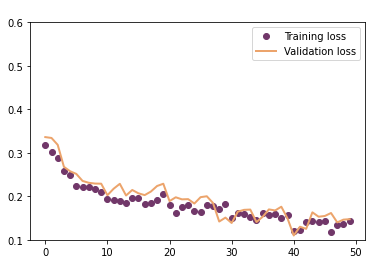
Rysunek 24. Wykresy zmian dokładności i starty modelu z ResNet50.

**** ****

Rysunek 25. Wykresy zmian dokładności i starty modelu z EfficientNetB7.

**** ****

Rysunek 26. Wykresy zmian dokładności i starty modelu z InceptionV3.

**** ****

Rysunek 27. Wykresy zmian dokładności i starty modelu z VGG16.

Po obserwacji otrzymanych wykresów, można jednoznacznie stwierdzić o braku przeuczenia oraz underfittingu przedstawionych modeli. Wszystkie wykresy dla zbioru treningowego i walidacyjnego mają bardzo zbliżony przebieg, a tym samym otrzymane wartości. Świadczy to o wysokim dopasowaniu modeli do danych. W przypadku niektórych modeli zauważalny było niewielkie odchylenie. Działo się tak dla modelu z siecią DenseNet121, gdzie strata podczas ewaluacji na zbiorze treningowym była w końcowym etapie niższa niż dla walidacyjnego. Odwrotną sytuacje zaobserwowano natomiast w przypadku modelu z siecią InceptionV3. Uznano jednak, że otrzymane różnice są wystarczająco małe i można przyjąć je za pomijalne. Najwyższą wartość starty podczas uczenia zauważono dla modelu EfficentNetB7, który pod koniec osiągnął wartość delikatnie poniżej 0,3. Może to wynikać z faktu, że cechował się on najwyższą złożonością, ze względu na wysoką liczbę warstw w modelu bazowym. Spośród wszystkich modeli, wykres dokładności dla modelu VGG16 miał największe wahania wartości, szczególnie na początku szkolenia, co prawdopodobnie wynika z jego prostoty.

## Ewaluacja rezultatów uczenia

Przeprowadzona analiza przebiegu uczenia omawianych modeli dała jedynie podstawowy wgląd na sprawność implementowanych sieci neuronowych. Aby móc obiektywnie ocenić ich jakość, konieczne było przeprowadzenie ewaluacji na surowych danych, z którymi modele nie miały styczności. W tym celu wykorzystany został zbiór testowy, w którym znajduje się 3561 zdjęć, z podziałem na 1780 obrazów należących do klasy MELANOMA i 1781 do klasy NOT MELANOMA. Za pomocą metody model.evaluate() wyliczono na tym zbiorze wybrane metryki oceny klasyfikatora, a otrzymane wartości dla każdego modelu zamieszczono w Tabeli 4.

Tabela 4. Wartości metryk wyrażone w procentach, wyliczone ze zbioru testowego.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Dokładność | Precyzja | Czułość | AUC | Miara F1 |
| DenseNet121 | 94,07 | 90,05 | 99,10 | 98,16 | 94,36 |
| ResNet50 | 93,38 | 94,94 | 91,69 | 97,61 | 93,28 |
| EfficientNetB7 | 94,80 | 92,67 | 97,25 | 99,32 | 94,90 |
| InceptionV3 | 93,11 | 88,78 | 98,65 | 98,08 | 93,45 |
| VGG16 | 91,02 | 92,40 | 89,44 | 97,96 | 90,89 |

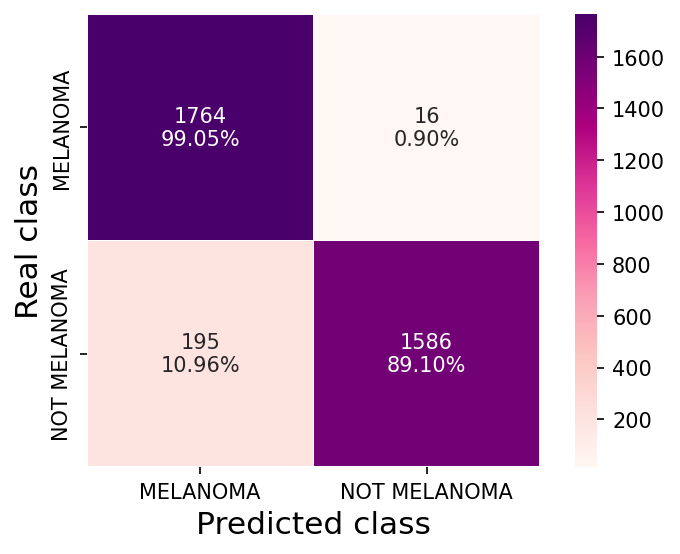
Ewaluację rozpoczęto od parametru dokładności, mówiącego o ilości poprawnie sklasyfikowanych zdjęć w całym zbiorze. Jest on najbardziej podstawowym miernikiem jakości modelu, który służy do szybkiej oceny. Dla wszystkich modeli osiągnięta została wartość powyżej 90%, co spełnia pierwsze założenia wytypowane przed wdrożeniem algorytmów. Spośród nich najwyższy wynik uzyskał model bazujący na sieci EfficientNetB7 na poziomie 94,8%, a najniższy okazał się dla modelu z siecią VGG16 o dokładności równej 91,02%. Różnica pomiędzy uzyskanymi dokładnościami w pięciu modelach była jedynie na poziomie paru punktów procentowych, dlatego również przeprowadzono porównanie innych metryk.

W przypadku omawianego zagadnienia klasyfikacji jednostek chorobowych, najważniejsza była zdolność modelu do poprawnego wykrywania zdjęć czerniaka. Tą właściwość opisywał wyliczony parametr czułości, gdzie jego wartość oznaczała odsetek poprawnie przypisanych etykiet w klasie MELANOMA. Pod tym względem najwyższy wynik osiągnął model bazujący na sieci DenseNet121 o wielkości 99,1%. Oznacza to, że mniej niż 1% obrazów czerniaka nie zostało wykryte. Natomiast najniższy wyniki ponownie przypada dla modelu VGG16, który wyniósł jednie 89,44%.

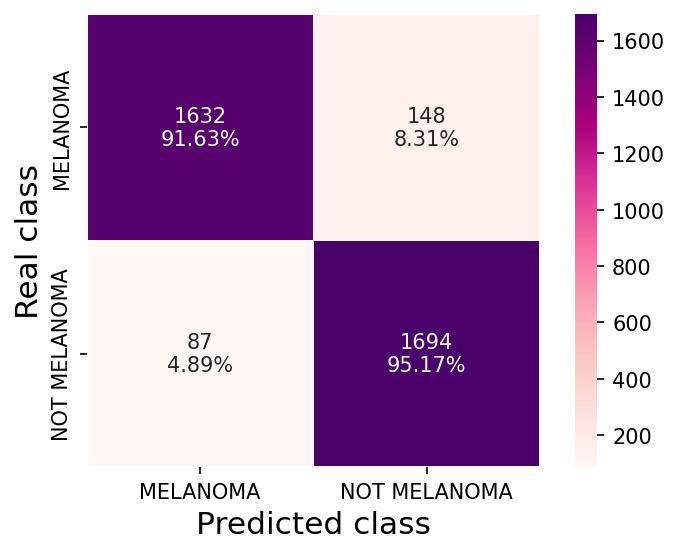
Biorąc pod uwagę sam parametr czułości, należy również mieć świadomość możliwości, w której model przypisuje wszystkie etykiety tylko jednej klasie. W takim przypadku czułość osiągnie bardzo wysoki poziom, ale sieć będzie kompletnie niezdatna do diagnozy patologii. Dlatego też w parze z czułością, omawiana powinna być metryka precyzji, opisująca, ile obrazów przypisanych przez model do klasy MELANOMA rzeczywiście do niej należy. Dzięki temu zauważono znaczne różnice między testowanymi modelami. Najciekawszym przypadkiem spośród nich była sieć z InceptionV3, dla której czułość i precyzja różniły się o prawie 10 punktów procentowych. Oznacza to, że model ten miał tendencję to przypisywania niektórym zdjęciom zmian z grupy NOT MELANOMA klasę czerniaka. Nie była to może cecha przekreślająca jego zastosowanie, natomiast nietrafna diagnoza jest jedną z przyczyn wydłużenia czasu przystąpienia do leczenia.

Metryką łączącą omawiane parametry czułości i precyzji jest miara F1. Stanowi ją średnia harmoniczna obu parametrów, która służy do szybkiego porównania jakości różnych modeli. Uznano ją za najbardziej miarodajną metrykę, opisującą możliwości modeli w bardziej złożonym aspekcie. W kolejności najwyższej do najniższej wartości miary F1 możemy wymienić: model EfficientNetB7 na poziomie 94,9%, zaraz po nim model DenseNet121 - 94,36%, następnie model InceptionV3 – 93,45%, model ResNet50 – 93,28% i na koniec model VGG16 z metryką o wartości 90,89%. Co ciekawe przedstawiona kolejność pokrywa się ze stopniem złożoności każdego z wybranych modeli. Sieć EfficientNetB7 jest największa i posiada łącznie 815 warstw, z kolei sieć VGG16 stanowi tylko 20 warstw. Może to świadczyć o dużym wpływie stopnia złożoności na jakość modelu w zagadnieniu klasyfikacji zmian skórnych.

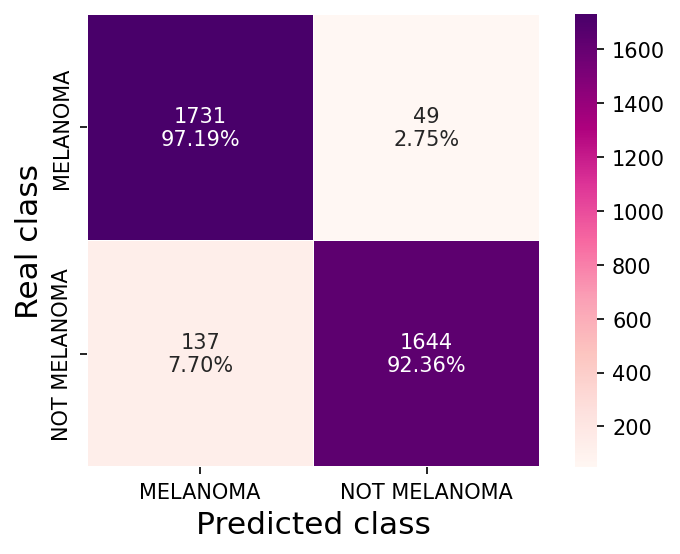
W końcowym etapie dokonano graficznej wizualizacji omówionych parametrów za pomocą macierzy pomyłek. Zostały one wyliczone na zbiorze testowym, gdzie przypadki pozytywne stanowiły obrazy czerniaka, a przypadki negatywne pozostałe zdjęcia. W macierzach oprócz liczby obserwacji klasyfikowanych do poszczególnych klas zamieszczono również ich odsetek w stosunku do wielkości całej klasy, wyrażony w procentach. Wykresy zilustrowano na Rysunkach 28 - 32.



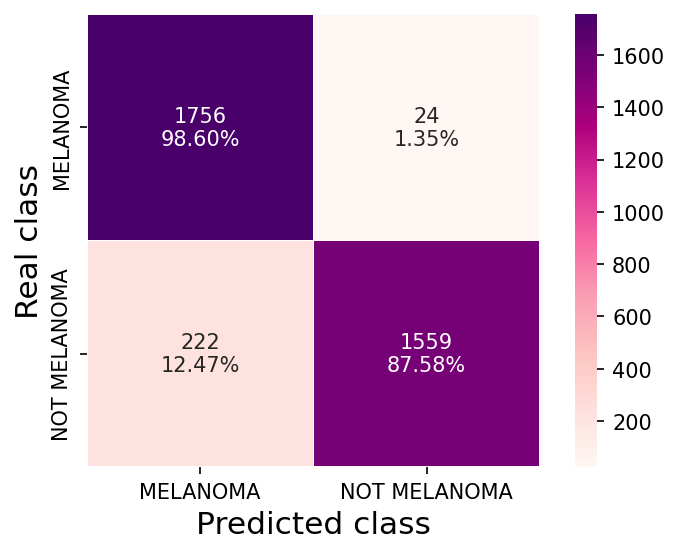
Rysunek 28. Macierz pomyłek dla modelu DenseNet121



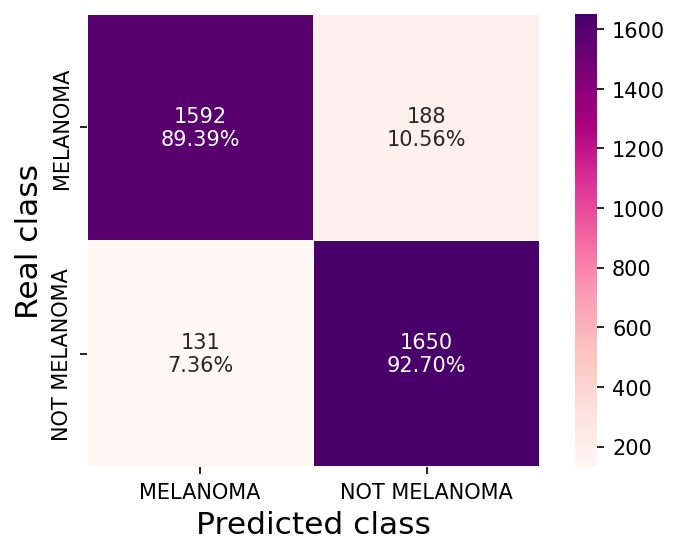
Rysunek 29. Macierz pomyłek dla modelu ResNet50



Rysunek 30. Macierz pomyłek dla modelu EfficientNetB7



Rysunek 31. Macierz pomyłek dla modelu InceptionV3



Rysunek 32. Macierz pomyłek dla modelu VGG16

Analizując uzyskane wykresy, zaobserwowano te same zależności, które wytypowano z analizy wartości liczbowych omawianych metryk. Jedną z nich jest najwyższy odsetek przypadków prawdziwie pozytywnych przypadający dla sieci DenseNet121 - 99,05%, a najniższy dla sieci VGG16 – 89,39%. Natomiast wybrana reprezentacja danych dodatkowo pozwala na uzyskanie informacji o liczbie błędnie przypisanych etykiet. Dzięki temu możliwe jest określenie, z którą klasą model radzi sobie gorzej. Zauważono, że algorytmy bazujące na sieciach EfficientNetB7, DenseNet121 i InceptionV3 popełniły więcej błędów w przypisywaniu etykiet obrazom należącym do grupy NOT MELANOMA. Z kolei modele z sieciami ResNet50 i VGG16 wykazały odwrotną tendencję, lepiej klasyfikując zdjęcia należące do wymienionej klasy. Był to niepożądany efekt ze względu na fakt, że najistotniejsze było wykrycie czerniaka. Oczekiwany algorytm powinien skupiać się na poprawnym wykrywaniu etykiet klasy MELANOMA, z jak najmniejszą liczbą przypadków fałszywie pozytywnych.

# Podsumowanie i możliwości rozwoju

W projekcie zaprezentowano możliwości sieci neuronowych jako narzędzia w klasyfikacji zmian skórnych. W rozdziale 1 przybliżone zostały zagadnienia związane z przyczynami oraz skutkami powstawania raka skóry. Szczególną uwagę zwrócono na zagrożenie, które stanowi stale rosnąca liczba zachorowań na czerniaka, co wywiera presję na coraz szybszą diagnozę badanych patologii skóry. W tej części omówiono również podstawowe założenia uczenia maszynowego oraz sprecyzowano ideę sieci konwolucyjnych. Przytoczone zostały metody ewaluacji implementowanych modeli, a na koniec przedstawiono wybrane przykłady zastosowań CNN w dermatologii zebrane z przeglądu literaturowego. Drugi rozdział pracy skupiał się na wdrożeniu modelu klasyfikacyjnego. Rozpoczęto w nim od przedstawienia wykorzystywanego zbioru danych, który stanowiły zdjęcia dermatoskopowe ze zbioru HAM1000. Dokonano analizy jego liczebności i stwierdzono, że występuje bardzo duże niezrównoważenie wśród wyznaczonych klas. Z tego powodu konieczne było odpowiednie przygotowanie zbioru. Kroki jakie podjęto w tym kierunku opierały się na wytypowaniu dwóch nowych klas zdjęć, z podziałem na obrazy ilustrujące czerniaka oraz pozostałe zmiany skórne. Na drodze augmentacji, syntetycznie zwiększono mniej liczną grupę i doprowadzono do wyrównania liczebności klas. W kolejnym etapie opisany został proces implementacji algorytmów, które bazowały na pięciu różnych, konwolucyjnych sieciach pretrenowanych. Należały do nich modele: DenseNet121, ResNet50, EfficientNetB7, InceptionV3 oraz VGG16. W projekcie dokonano opisu napotkanych problemów i działań mających na celu ich zniwelowanie. Przeprowadzono analizę przebiegu procesu uczenia poszczególnych modeli. Na koniec dzięki wyliczonym metrykom oceny jakości modelu i graficznej ich reprezentacji w postaci macierzy pomyłek, dokonano ewaluacji wykorzystanych sieci. Wszystkie z nich osiągnęły zadowalające wyniki, z dokładnością i miara F1 na poziomie powyżej 90%. Model bazujący na EfficientNetB7 wyróżniono jako najdokładniejszy, natomiast VGG16 cechował się najniższymi wynikami.

Jako że uczenie maszynowe jest stosunkowo nową technologią, a w szczególności uczenie głębokie istnieje wiele ścieżek rozwoju proponowanych algorytmów. Najbardziej oczywistą z nich jest ciągła optymalizacja parametrów modeli, która pozwoliłaby na uzyskanie jeszcze lepszych rezultatów klasyfikacji. Patrząc w dalszej perspektywie, proponowane modele mogłyby być rozbudowane o posługiwanie się informacjami na temat pacjenta takimi jak: wiek, płeć czy lokalizacja badanej zmiany w celu zwiększenia dokładności przypisywanych etykiet.

# Wnioski

Głównym celem zrealizowanego projektu było stworzenie modelu sieci konwolucyjnej do automatycznej klasyfikacji zmian skórnych z podziałem na zmiany łagodne i złośliwe. Proponowane algorytmy jako źródło danych wykorzystywały zbiór zdjęć dermatoskopowych, należących do publicznie dostępnego zbioru HAM10000. W czasie realizacji projektu poruszono zagadnienie wstępnego przetwarzania obrazów oraz równoważenia zbioru danych rzeczywistych.

W niniejszej pracy dowiedziono, że stopień złożoności ma duży wpływ na osiągnięcia modelu stworzonego do detekcji groźnych zmian skórnych. Wśród zaproponowanych algorytmów, model bazujący na sieci EfficientNetB7 okazał się być najdokładniejszy. Świadczy o tym wysoki poziom metryki F1 równy 94,9%, opisujący zdolność klasyfikatora do poprawnej detekcji czerniaka, jak również wysoka wartość wskaźnika F1 równa 94,9%. Natomiast model, który opierał się na sieci VGG16 wykazywał najniższe wartości metryk, przy dokładności równej 91,02% i czułości 89,44%. Niższy parametr czułości świadczy o gorszych możliwościach modelu do wykrywania obrazów czerniaka. Pomimo tego uzyskane wyniki dla wszystkich modeli są obiecujące, gdyż nie doszło do ich przeuczenia. Nie dorównują one jednak najnowszym osiągnięciom z dziedziny klasyfikacji binarnej zmian skórnych. Powodem może być zawiłość dopasowania poszczególnych aspektów budowy modelu do wykorzystywanych danych, a także odpowiednia realizacja ich przetworzenia. Są to bardzo trudne zagadnienia, których dopracowywanie może trwać latami. Jednak rozwój techniki uczenia maszynowego w przyszłości może pozwolić na ułatwienie całego procesu, a tym samym możliwe będzie wdrożenie tej technologii w większej liczbie placówek medycznych.

# Bibliografia

[1] M. Stücker, A. Struk, P. Altmeyer, M. Herde, H. Baumgärtl, i D. W. Lübbers, „The cutaneous uptake of atmospheric oxygen contributes significantly to the oxygen supply of human dermis and epidermis”, *The Journal of Physiology*, t. 538, nr 3, s. 985–994, luty 2002, doi: 10.1113/jphysiol.2001.013067.

[2] Anna, „Dermestetica Tarnów - Budowa i funkcja skóry - jak to działa?”, *dermestetica*, 3 luty 2021. https://dermestetica-tarnow.pl/blog/dermatologia/jak-dziala-skora/ (dostęp 3 listopad 2022).

[3] „Skin Lesions: What Are They, Types, Causes, Diagnosis, Treatment, and More | Osmosis”. https://www.osmosis.org/answers/skin-lesions (dostęp 3 listopad 2022).

[4] „Skóra | KRN”. http://onkologia.org.pl/nowotwory-zlosliwe-skory/ (dostęp 3 listopad 2022).

[5] „Jak szybko rozwija się czerniak? Jak powstaje? - Akademia Czerniaka”. https://www.akademiaczerniaka.pl/o-czerniaku/jak-rozwija-sie-czerniak-skory-czynniki-ryzyka (dostęp 3 listopad 2022).

[6] „ABCDE samokontroli znamion - OGÓLNOPOLSKI PROGRAM PROFILAKTYKI NOWOTWORÓW SKÓRY”, https://www.pacjent.gov.pl/programy-profilaktyczne/profilaktyka-nowotworow-skory (dostęp 3 listopad 2022)

[7] „Nowotwory skóry (C44) | KRN”. http://onkologia.org.pl/nowotwory-skory-c44/ (dostęp 3 listopad 2022).

[8] T. M. Mitchell, "*Machine Learning"*. New York: McGraw-Hill, 1997.

[9] M. Jakubowska, F. Ciepela, „Chemometria wykłady”, Akademia Górniczo-Hutniczna, 2022.

[10] „Artificial Neural Network - an overview | ScienceDirect Topics”. https://www.sciencedirect.com/topics/neuroscience/artificial-neural-network (dostęp 3 listopad 2022).

[11] „Deep Learning”. https://www.deeplearningbook.org/ (dostęp 3 listopad 2022).

[12] „SztucznaInteligencja-UczenieGłębokichSieciNeuronowych.pdf”. https://home.agh.edu.pl/~horzyk/lectures/ai/SztucznaInteligencja-UczenieG%C5%82%C4%99bokichSieciNeuronowych.pdf (dostęp 3 listopad 2022)

[13] „Krzywe ROC i ocena jakości klasyfikacji - Predictive Solutions”. https://predictivesolutions.pl/krzywe-roc-i-ocena-jakosci-klasyfikacji (dostęp 3 listopad 2022).

[14] D. R. Sarvamangala i R. V. Kulkarni, „Convolutional neural networks in medical image understanding: a survey”, *Evol. Intel.*, t. 15, nr 1, s. 1–22, mar. 2022, doi: 10.1007/s12065-020-00540-3.

[15] C. Calderón, K. Sanchez, S. Castillo, i H. Arguello, „BILSK: A bilinear convolutional neural network approach for skin lesion classification”, *Computer Methods and Programs in Biomedicine Update*, t. 1, s. 100036, sty. 2021, doi: 10.1016/j.cmpbup.2021.100036.

[16] L. Talavera-Martínez, P. Bibiloni, A. Giacaman, R. Taberner, L. J. D. P. Hernando, i M. González-Hidalgo, „A novel approach for skin lesion symmetry classification with a deep learning model”, *Computers in Biology and Medicine*, t. 145, s. 105450, cze. 2022, doi: 10.1016/j.compbiomed.2022.105450.

[17] „melanoma”. https://www.kaggle.com/datasets/drscarlat/melanoma (dostęp 24 listopad 2022).

[18] „Machine Learning Primer for Clinicians–Part 12”. https://histalk2.com/2019/01/16/machine-learning-primer-for-clinicians-part-12-2/ (dostęp 24 listopad 2022)