AKADEMIA GÓRNICZO-HUTNICZA

im. Stanisława Staszica w Krakowie

WYDZIAŁ ZARZĄDZANIA

Sieci neuronowe i uczenie głębokie

Porównanie metod klasyfikacji ras psów z wykorzystaniem głębokich sieci neuronowych – analiza modeli CNN, ResNet50 i VGG16

Autor: Daria Twardy, Aleksandra Rosa

Informatyka i Ekonometria

Stopień II, rok II, grupa 2 (nst)

Kraków, styczeń 2025

Spis treści

Wstęp

Rozpoznawanie ras psów na podstawie zdjęć to ciekawe wyzwanie w dziedzinie komputerowego rozpoznawania obrazów. Psy należą do gatunku o ogromnej różnorodności fenotypowej – ich rasy różnią się znacząco pod względem wielkości, kształtu ciała, kolorystyki sierści i rysów pyska. Z tego względu klasyfikacja ras psów na podstawie zdjęć wymaga zastosowania skutecznych metod ekstrakcji cech oraz odpowiednich algorytmów uczenia maszynowego.

Opis problemu i cel projektu

Celem projektu było opracowanie i porównanie różnych modeli głębokiego uczenia, obejmujących zarówno tradycyjne sieci konwolucyjne (CNN), jak i bardziej zaawansowane architektury transfer learningu, takie jak ResNet50 i VGG16, w zadaniu klasyfikacji sześciu ras psów. Projekt miał na celu określenie, które podejście pozwoli uzyskać najwyższą skuteczność klasyfikacji, a także jakie techniki przetwarzania danych i regularyzacji mogą poprawić zdolność modeli do generalizacji na nowe dane. Dane wykorzystane w projekcie pochodziły z publicznie dostępnego zbioru 6 Dog Breeds, dostępnego na platformie RoboFlow.

Przegląd istniejących rozwiązań.

W ostatnich latach sieci neuronowe zdominowały dziedzinę analizy obrazów, zastępując tradycyjne metody oparte na ręcznej ekstrakcji cech, takie jak Histogram of Oriented Gradients (HOG) czy deskryptory SIFT i SURF. Modele głębokiego uczenia potrafią samodzielnie uczyć się istotnych cech wizualnych na podstawie dostarczonych danych, co znacząco zwiększa ich skuteczność.

Przełomowym momentem w tej dziedzinie było wprowadzenie architektury AlexNet w 2012 roku przez Alexa Krizhevsky'ego, Ilyę Sutskevera i Geoffreya Hintona. AlexNet, wykorzystując głęboką sieć konwolucyjną, osiągnął znaczący sukces w konkursie ImageNet Large Scale Visual Recognition Challenge, co zapoczątkowało nową erę w głębokim uczeniu. Kolejne modele, takie jak VGG16 i ResNet, wprowadziły jeszcze głębsze i bardziej zoptymalizowane architektury, umożliwiając skuteczniejsze rozpoznawanie skomplikowanych wzorców w obrazach. VGG16 (Visual Geometry Group) został zaprojektowany przez Simonyana i Zissermana w 2014 roku, wyróżniał się prostotą architektury, bazując na wielu warstwach konwolucyjnych o małych filtrach 3x3 i warstwach poolingowych, co prowadziło do bardzo dobrej skuteczności w zadaniach klasyfikacji. ResNet (Residual Networks) wprowadzony został przez He et al. w 2015 roku zrewolucjonizował dziedzinę, rozwiązując problem zanikającego gradientu i umożliwiając budowę wyjątkowo głębokich sieci neuronowych poprzez zastosowanie residual connections. GoogLeNet (Inception) opracowany przez Szegedy'ego i współpracowników wprowadził moduły Inception, które pozwalały na efektywne przetwarzanie obrazów o różnych poziomach abstrakcji, poprawiając wydajność modelu przy jednoczesnym zmniejszeniu liczby parametrów.

W niniejszym projekcie zdecydowano się na porównanie różnych podejść: od klasycznych sieci konwolucyjnych (CNN), poprzez modele wykorzystujące augmentację danych, aż po transfer learning. Celem było nie tylko określenie, które z podejść jest najskuteczniejsze, ale także zrozumienie, jakie modyfikacje mogą poprawić jakość klasyfikacji. Szczególna uwaga została poświęcona wpływowi regularyzacji, augmentacji danych oraz dostrajania wybranych warstw w gotowych sieciach, aby zminimalizować problem przeuczenia i zwiększyć zdolność modelu do uogólniania wiedzy na nowych danych.

Metodologia

Opis wybranych modeli

W projekcie zastosowano kilka różnych strategii modelowania, aby znaleźć optymalne podejście. Podjęte kroki obejmowały zarówno klasyczne konwolucyjne sieci neuronowe (CNN) budowane od podstaw, jak i wykorzystanie transfer learningu, który umożliwia korzystanie z wcześniej wytrenowanych sieci na dużych zbiorach obrazów. Pierwszym krokiem było stworzenie własnych modeli CNN, które były stopniowo rozwijane i ulepszane, aby zidentyfikować, jak głęboka i złożona architektura sieci może skutecznie klasyfikować obrazy psów. Budowano różne wersje sieci, od prostych, dwuwarstwowych modeli po bardziej zaawansowane, kilkustopniowe architektury z warstwami dropout, batch normalization i większą liczbą filtrów konwolucyjnych. Jednak klasyczne CNN napotkały pewne ograniczenia – przede wszystkim trudności w uzyskaniu wysokiej dokładności ze względu na ograniczoną ilość danych. W związku z tym, drugim etapem było zastosowanie techniki transfer learningu, czyli wykorzystania gotowych modeli wcześniej wytrenowanych na dużych zbiorach obrazów (np. ImageNet). Dzięki temu modele miały już wyuczone reprezentacje cech wizualnych i wymagały jedynie dostosowania do konkretnego zadania klasyfikacji ras psów. Zastosowano dwa popularne modele takie jak ResNet50 oraz VGG16. ResNet50 to głęboka sieć neuronowa o 50 warstwach, która charakteryzuje się zastosowaniem mechanizmu residual connections, pozwalającym na efektywne uczenie bardzo głębokich sieci a VGG16 to prostszy model z 16 warstwami, który, mimo swojej relatywnie płytkiej struktury w porównaniu do ResNet, często osiąga bardzo dobre wyniki w klasyfikacji obrazów. Obydwa modele zostały przetestowane w trybie zamrożonych wag – oznacza to, że wykorzystano wcześniej nauczone cechy wizualne, a jedynie dodano nową warstwę klasyfikacyjną. Takie podejście pozwalało uniknąć problemu zbyt długiego treningu i lepiej dopasować model do dostępnych danych. Ostatnim krokiem było ulepszenie modelu VGG16 z powodu, iż ze wszystkich stworzonych modeli najlepiej sobie radził. Ulepszono go poprzez fine-tuning, czyli odblokowanie i ponowne wytrenowanie kilku ostatnich warstw sieci. Dodatkowo wprowadzono zaawansowaną augmentację danych, aby zwiększyć różnorodność zbioru treningowego, regularyzację dropout, aby zmniejszyć ryzyko przeuczenia oraz dynamiczne dostosowanie learning rate (ReduceLROnPlateau), które automatycznie zmniejszało tempo uczenia, gdy model przestawał się poprawiać. Dzięki temu ulepszony model VGG16 osiągnął najlepsze wyniki, co potwierdziło skuteczność transfer learningu oraz fine-tuningu jako metody optymalizacyjnej.

Projekt zaczęto od budowy modelu CNN, który kolejno testowano pod kątem różnych parametrów. Z powodu niezadowalających wyników, postanowiono wykorzystać modele transfer learning. Zatem podjęte decyzje dotyczące architektury i strategii uczenia miały na celu jak najlepsze dopasowanie modelu do problemu klasyfikacji ras psów. Wybór różnych podejść wynikał z kilku kluczowych czynników takich jak efektywność w analizie obrazów, dostosowanie do dostępnej ilości danych oraz wydajność i optymalizacja

Modele CNN budowane od podstaw pozwalały na pełną kontrolę nad architekturą, ale wymagały dużej ilości danych, aby osiągnąć satysfakcjonujące wyniki. Z tego powodu nie były wystarczająco skuteczne w przypadku naszego zbioru. Transfer learning z ResNet50 i VGG16 pozwolił na wykorzystanie wcześniej nauczonych cech wizualnych, co znacząco przyspieszyło trening i poprawiło wyniki klasyfikacji. Model ResNet50 jest głębszym modelem, który dobrze sprawdza się w dużych zbiorach, ale jego wyniki na tym konkretnym zadaniu nie były zadowalające – prawdopodobnie dlatego, że struktura residual connections nie była w pełni wykorzystana przy ograniczonej ilości danych. VGG16 lepiej pasował do zadania, ponieważ

jego struktura bazuje na prostych warstwach konwolucyjnych, które dobrze radzą sobie z klasyfikacją o stosunkowo niewielkiej liczbie klas.

Dodatkowo własne modele CNN wymagałyby znacznie większej liczby próbek, aby uzyskać dobre wyniki, ponieważ uczą się wszystkiego od podstaw. Natomiast modele transfer learningowe mogły efektywnie działać nawet na ograniczonym zbiorze, ponieważ ich wcześniejsze treningi na ImageNet dostarczyły im ogólnej wiedzy na temat cech wizualnych.

Opis danych

Zbiór danych zawiera obrazy sześciu różnych ras psów, podzielone na trzy podzbiory czyli zbiór treningowy zawierający największą liczbę próbek, wykorzystywany do nauki modeli, zbiór walidacyjny służący do oceny wydajności modeli podczas treningu i zapobiegania przeuczeniu oraz zbiór testowy przeznaczony do końcowej oceny skuteczności wytrenowanych modeli. Obrazy w zbiorze miały różne rozdzielczości, dlatego przed treningiem były przeskalowywane do jednolitego rozmiaru, zależnie od wybranego modelu (np. 224x224 dla VGG16 i ResNet50 oraz 640x640 dla pierwszej i drugiej sieci CNN). Dane były również poddawane normalizacji, a w niektórych przypadkach stosowano augmentację, aby zwiększyć różnorodność próbek i poprawić zdolność modeli do generalizacji.

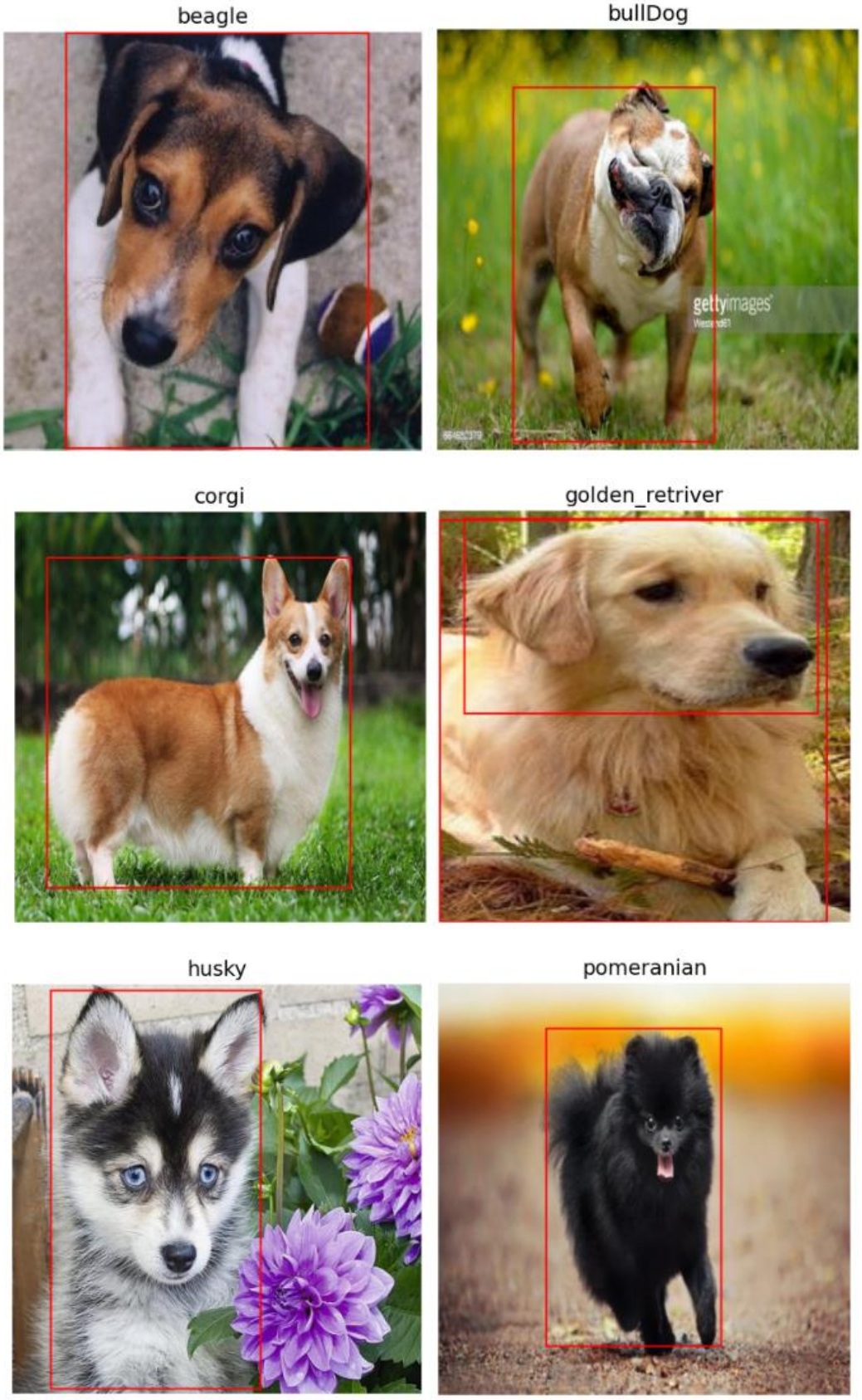
Tabela 1. Ilość obrazów w zestawach

Rasa	Zestaw treningowy	Zestaw walidacyjny	Zestaw testowy
Beagle	171	52	25
Bulldog	126	46	25
Corgi	173	37	23
Golden Retriever	177	54	21
Husky	137	35	21
Pomeranian	96	27	11

Rozkład liczby próbek w poszczególnych klasach nie jest równomierny, co może mieć istotny wpływ na skuteczność modeli klasyfikacyjnych. W szczególności widoczne są różnice w liczbie dostępnych obrazów dla różnych ras psów, zarówno w zbiorze treningowym, walidacyjnym, jak i testowym. Nierównomierność w ilości obrazów może powodować, że modele mogą preferować klasy z większą liczbą próbek, co może doprowadzić do sytuacji, w której mniej reprezentowane klasy będą częściej klasyfikowane błędnie lub przypisywane do bardziej licznych kategorii. Mniej liczne klasy mogą powodować większą wariancję w wynikach walidacyjnych i testowych, ponieważ mniejsza liczba przykładów mogła nie odzwierciedlać pełnego zakresu cech wizualnych danej rasy psa.

Poniżej przedstawiono przykładowe zdjęcia z każdej klasy

Wykres 1. Przykładowe obrazy z każdej klasy



Implementacja

Pierwszym zastosowanym modelem w eksperymencie była klasyczna sieć konwolucyjna (CNN) o stosunkowo prostej architekturze, zaprojektowana w celu oceny, jak podstawowa struktura radzi sobie z klasyfikacją sześciu ras psów. Model składał się z trzech warstw konwolucyjnych, których zadaniem była ekstrakcja kluczowych cech wizualnych, oraz warstw poolingowych, które redukowały wymiar danych wejściowych, zmniejszając liczbę parametrów do przetworzenia. Obrazy były wprowadzane do sieci w pełnej skali 640x640 pikseli, co oznaczało dużą liczbę operacji obliczeniowych, ale jednocześnie dawało możliwość uchwycenia bardziej szczegółowych wzorców charakterystycznych dla poszczególnych ras.

Na początku model przekształcał obrazy za pomocą warstwy konwolucyjnej z 32 filtrami o rozmiarze 3x3, stosując funkcję aktywacji ReLU, co pozwalało na wydobycie podstawowych krawędzi i tekstur. Następnie zastosowano operację max pooling, która zmniejszała rozmiar obrazu, ułatwiając sieci generalizację. W kolejnych etapach dodano drugą i trzecią warstwę konwolucyjną, każdą z większą liczbą filtrów (64 i 128), aby umożliwić modelowi uchwycenie coraz bardziej złożonych cech charakterystycznych dla poszczególnych ras. Po przetworzeniu przez warstwy konwolucyjne i poolingowe dane były spłaszczane do jednowymiarowego wektora, który następnie przechodził przez gęsto połączoną warstwę z 128 neuronami, gdzie model miał możliwość analizy abstrakcyjnych zależności między cechami. Ostateczna warstwa wyjściowa, zawierająca sześć neuronów aktywowanych funkcją softmax, przypisywała wejściowe obrazy do jednej z sześciu klas.

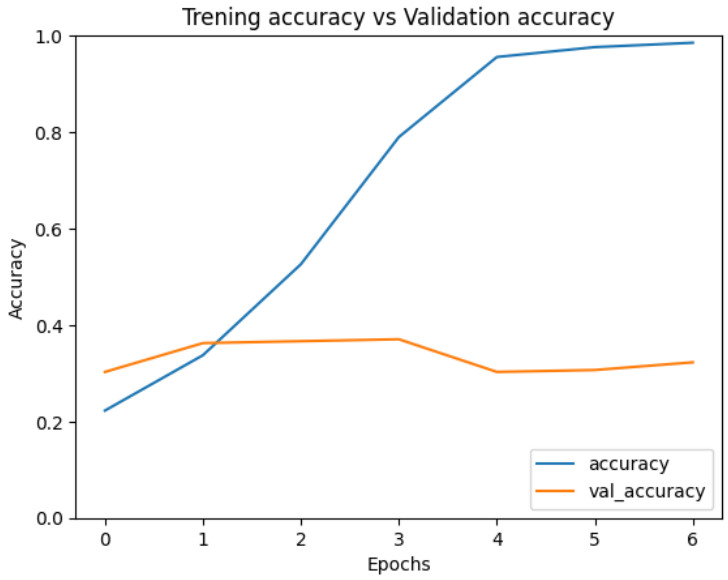
W celu optymalizacji procesu uczenia model skompilowano z wykorzystaniem algorytmu Adam, który dobrze radzi sobie z dynamicznymi zmianami w danych i jest często stosowany w zadaniach klasyfikacji obrazów. Jako funkcję straty zastosowano sparse categorical crossentropy, która jest szczególnie przydatna w przypadku klasyfikacji wieloklasowej, gdy etykiety reprezentowane są jako indeksy klas zamiast wektorów one-hot. Ten podstawowy model stanowił punkt odniesienia dla dalszych, bardziej zaawansowanych architektur. Jego wydajność na zbiorze walidacyjnym oraz testowym pozwoliła ocenić, czy konieczne są dodatkowe modyfikacje, takie jak dodanie głębszej architektury, regularyzacja, czy też wykorzystanie transfer learningu z gotowych, sprawdzonych modeli. Analiza wyników tego pierwszego podejścia pokazała, że klasyczna sieć konwolucyjna nie radziła sobie wystarczająco dobrze z klasyfikacją ras psów, co skłoniło do dalszej optymalizacji i testowania bardziej złożonych rozwiązań.

Trening, walidacja i testowania.

Podczas przetwarzania i przygotowania danych do treningu wykorzystano narzędzie ImageDataGenerator, które pozwoliło na normalizację oraz odpowiednie podział danych na trzy kluczowe zbiory. Zbiór treningowy (train) posłużył do nauki modelu poprzez dostarczanie mu przykładów, na podstawie których mógł dopasowywać swoje wagi. Zbiór walidacyjny (valid) umożliwiał monitorowanie postępów i wykrywanie ewentualnego przeuczenia, natomiast zbiór testowy (test) został wykorzystany do końcowej oceny skuteczności modelu na nowych, niewidzianych wcześniej obrazach. Aby ułatwić proces uczenia sieci, wartości pikseli wejściowych obrazów zostały przeskalowane do zakresu [0,1] poprzez podzielenie ich wartości przez 255. Taka normalizacja pozwoliła na stabilniejszą i szybszą konwergencję modelu podczas treningu. Dodatkowo zastosowano mechanizm EarlyStopping, który monitorował walidacyjną dokładność (val_accuracy). Dzięki temu trening był automatycznie przerywany, jeśli przez trzy kolejne epoki nie obserwowano poprawy wyników. W takim przypadku wagi

modelu były przywracane do tych, które osiągnęły najlepszą skuteczność, co zapobiegało przeuczeniu i poprawiało ogólną zdolność modelu do generalizacji. Złożoność sieci oraz duże wymiary wejściowych obrazów (640x640 pikseli) sprawiły, że czas trenowania jednej epoki wynosił średnio od 145 do 178 sekund. To pokazuje, że im większa głębia modelu i wyższa rozdzielczość wejściowych obrazów, tym większe były wymagania obliczeniowe, co miało wpływ na ogólną efektywność treningu.

Wykres 2. Dokładność modelu 1 na zbiorze treningowym i walidacyjnym



Wyniki i obserwacje

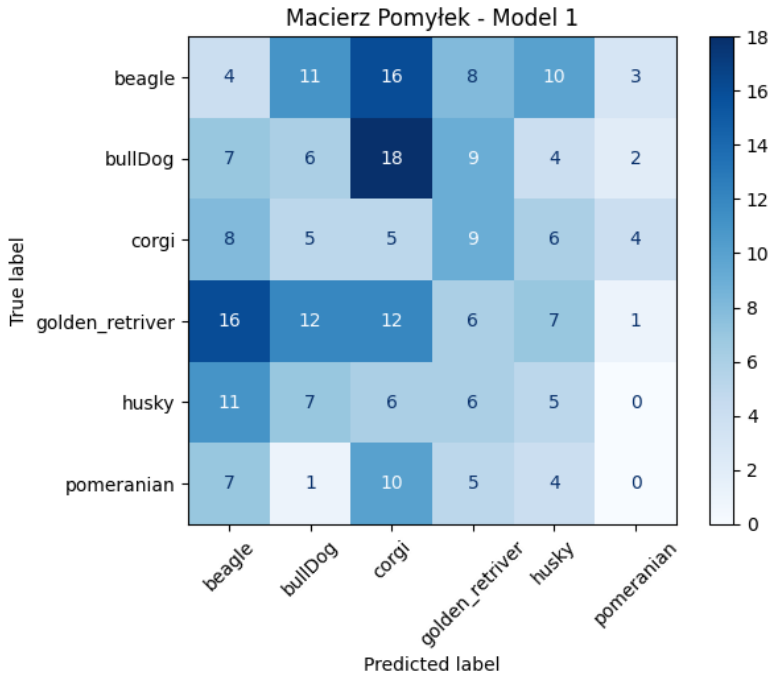
Podczas treningu model początkowo wykazywał stopniową poprawę dokładności, osiągając 77.39% na zbiorze treningowym po czterech epokach. Jednak w kolejnych epokach zaczęło pojawiać się przeuczenie, gdzie dokładność na zbiorze treningowym rosła (osiągając nawet 98%), podczas gdy dokładność walidacyjna pozostawała stosunkowo niska, na poziomie około 30-37%. Wskazywało to na problem ze zdolnością modelu do uogólniania wiedzy na nowych, niewidzianych wcześniej danych. Zastosowanie mechanizmu EarlyStopping pozwoliło na zatrzymanie treningu po siódmej epoce, przywracając najlepsze parametry z epoki czwartej. Wskazuje to, że dalsze trenowanie jedynie pogłębiało problem przeuczenia, zamiast poprawiać zdolność generalizacji modelu. Wydajność na zbiorze testowym również nie była zadowalająca, co sugeruje, że model dobrze zapamiętywał wzorce z danych treningowych, ale nie potrafił skutecznie rozpoznawać psów na nowych obrazach. Dodatkowym wyzwaniem była duża liczba parametrów modelu (99,7 miliona), co mogło wpłynąć na trudność w efektywnej optymalizacji oraz na czas trenowania. Każda epoka trwała średnio 145-178 sekund, co w połączeniu z szybkim przeuczeniem modelu sugerowało, że architektura może być zbyt złożona w stosunku do wielkości zbioru danych i charakterystyki problemu.

Podsumowując, klasyczny model CNN w swojej podstawowej formie nie sprawdził się jako skuteczne rozwiązanie do klasyfikacji ras psów, głównie ze względu na problem przeuczenia oraz trudności w generalizacji wyników. Konieczne stało się zastosowanie bardziej zaawansowanych podejść, takich jak rozbudowane modele konwolucyjne z dodatkowymi

mechanizmami regularyzacyjnymi lub sieci pretrenowane na dużych zbiorach danych, które mogłyby poprawić wyniki klasyfikacji.

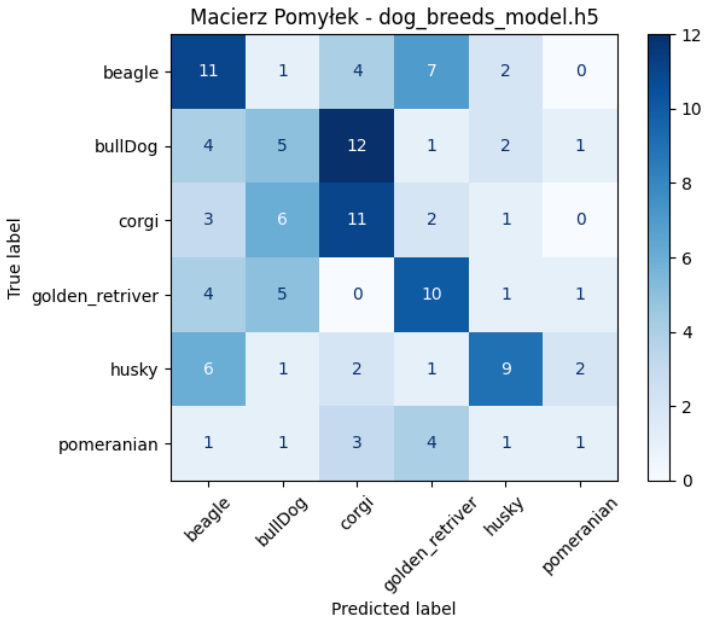
Dla modelu stworzono również macierz pomyłek na zbiorze walidacyjnym. Jak widać na poniższym obrazie model nie osiąga wysokiej skuteczności – widoczna jest znaczna liczba błędów klasyfikacyjnych, a wartości na przekątnej (prawidłowe klasyfikacje) są relatywnie niskie w stosunku do liczby błędnych dopasowań. Zauważalny jest również brak wyraźnej tendencji, która sugerowałaby, że model dobrze radzi sobie przynajmniej z jedną konkretną klasą.

Wykres 3. Macierz pomyłek modelu 1 na zbiorze walidacyjnym



Po zakończeniu procesu trenowania i walidacji model został przetestowany na zbiorze testowym, aby ocenić jego rzeczywistą skuteczność w klasyfikacji ras psów. Dokładność modelu na zbiorze testowym wyniosła 37.30%, co wskazuje na umiarkowaną skuteczność, ale jednocześnie potwierdza istnienie istotnych problemów w rozróżnianiu klas. Najlepsze wyniki są dla Bulldog i Corgi. Model najczęściej poprawnie klasyfikował te rasy (12 prawidłowych dla Bulldog i 11 dla Corgi), co sugeruje, że posiada dobrze dopasowane filtry konwolucyjne do tych konkretnych wzorców. Model napotkał problemy z rozpoznawaniem Golden Retrievera. W tym przypadku pojawiłą się znaczna liczba błędnych klasyfikacji (np. często mylony z Beagle i Bulldog) może sugerować, że model nie nauczył się dobrze cech charakterystycznych tej rasy. Pomyłki występowały również dla pozostałych ras. Niska dokładność modelu na zbiorze testowym (37.30%) sugeruje, że model nie generalizuje dobrze na nowe dane. Może to wynikać z niedostatecznego dopasowania hiperparametrów, niewystarczającej liczby warstw konwolucyjnych lub niedostatecznego zróżnicowania danych treningowych.

Wykres 4. Macierz pomyłek modelu 1 na zbiorze testowym



Eksperymenty:

Model pierwszy wykazywał trudności w klasyfikacji ras psów. Częste błędne klasyfikacje między podobnymi rasami wskazują, że architektura sieci może być niewystarczająca do skutecznego rozpoznawania wzorców wizualnych. Z powodu widocznego na wykresie accuracy przeuczenia w kolejnym modelu dodano dropout w celu ograniczenia przeuczenia.

Model 2 – Model z warstwą Dropout

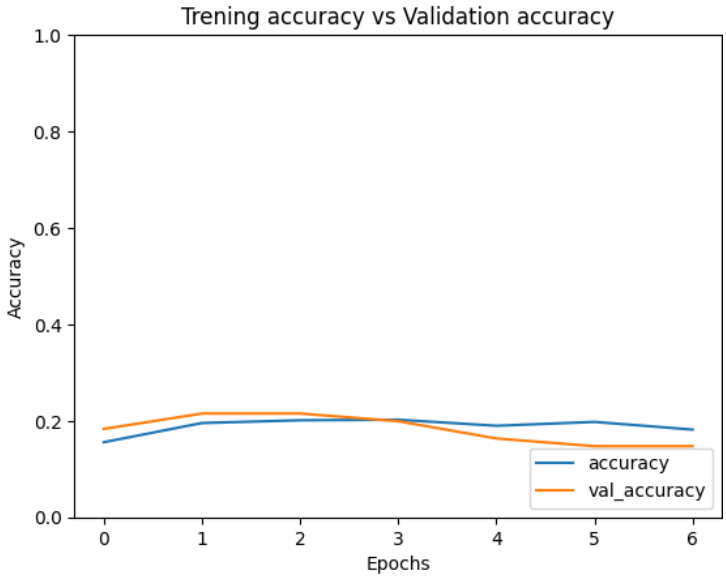
Główne zmiany w architekturze w stosunku do poprzedniego modelu obejmowały dodanie warstw Dropout - Po każdej operacji MaxPooling wprowadzono warstwę Dropout, aby losowo dezaktywować część neuronów, co miało na celu zmniejszenie ryzyka przeuczenia poprzez lepszą generalizację modelu. Liczba warstw konwolucyjnych była podobna do modelu wyjściowego. Model nadal składał się z trzech bloków konwolucyjnych każda z warstwą dropout. Warstwa gęsta o 128 neuronach została uzupełniona o dodatkową warstwę Dropout, co miało poprawić odporność modelu na przeuczenie. Liczba parametrów tego modelu pomimo dodania warstw regularizacyjnych, nie uległa istotnym zmianom względem Modelu 1, pozostając na poziomie 99,774,406.

Model również został wytrenowany z wykorzystaniem EarlyStopping, co oznacza, że trening kończył się po wykryciu braku poprawy dokładności walidacyjnej przez 3 epoki. W porównaniu do Modelu 1, wprowadzenie warstw Dropout pomogło w ograniczeniu przeuczenia, co można zaobserwować na wykresie przebiegu treningu. Czas trenowania wynosił średnio dla jednej epoki około 140-160 sekund, co było podobne do Modelu 1. Model 2 osiągnął wyższą dokładność niż Model 1, choć nadal nie był wystarczająco skuteczny w rozróżnianiu podobnych ras psów.

Na poniższym wykresie widać, że dokładność na zbiorze treningowym była wyraźnie wyższa niż na zbiorze walidacyjnym. To wskazuje, że model dobrze dopasował się do danych treningowych, ale nie był w stanie generalizować dobrze na nowych danych. Po kilku epokach

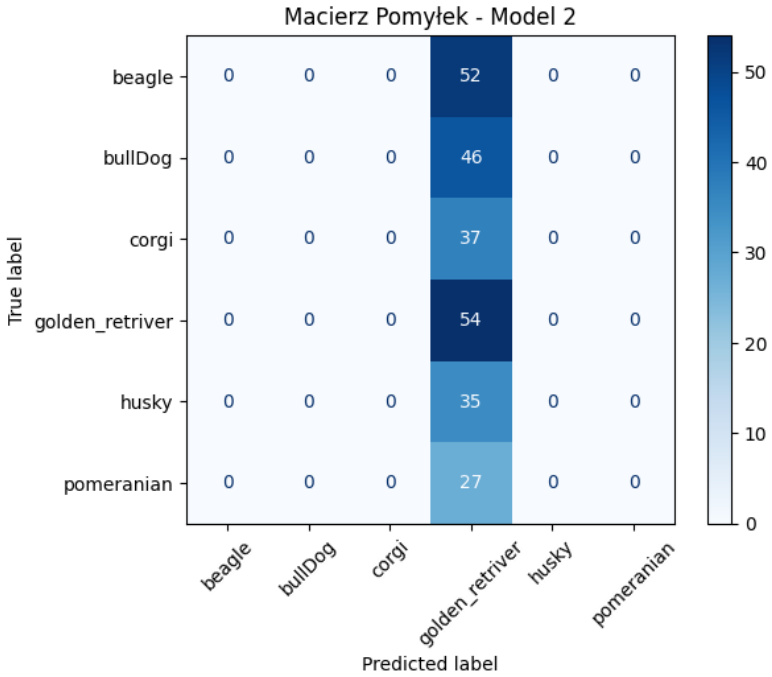
dokładność walidacyjna przestała rosnąć, a czasami nawet lekko spadała, co oznacza, że model przestał uczyć się nowych wzorców i zaczął zapamiętywać cechy treningowe.

Wykres 5. Dokładność modelu 2 na zbiorze treningowym i walidacyjnym



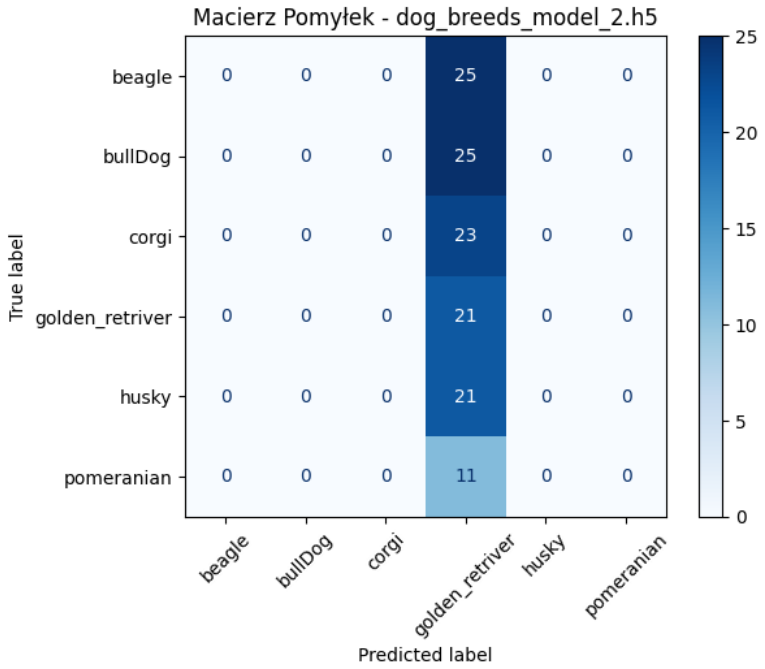
Mechanizm EarlyStopping zatrzymał trening po kilku epokach bez poprawy, co sugeruje, że model osiągnął swój limit i dalszy trening mógłby tylko pogorszyć wyniki walidacyjne. Dodanie warstw Dropout pomogło w redukcji przeuczenia w porównaniu do Modelu 1, ale nie wyeliminowało go całkowicie. Dropout sprawia, że model staje się bardziej odporny na dopasowywanie się do specyficznych danych treningowych, jednak sama regularizacja nie była wystarczająca do pełnej poprawy generalizacji. O jakości modelu świadczy również macierz pomyłek, która wskazuje, że model miał problem z przewidywaniem różnych klas, koncentrując się na jednej dominującej kategorii. Jak widać na ilustracji, model przypisywał większość obrazów do jednej klasy golden retriever, co oznacza, że pomimo dodania regularizacji, sieć nie nauczyła się skutecznie rozróżniać cech między rasami psów.

Wykres 6. Macierz pomyłek modelu 2 na zbiorze walidacyjnym



Model 2 zbadano również na zbiorze testowym, w którym dokładność modelu 2 wyniosła 16.67% Podsumowując, Model 2 miał mniej przeuczenia niż Model 1, ale nadal nie generalizował dobrze na nowe dane. Zarówno na zbiorze walidacyjnym, jak i testowym, można zauważyć bardzo poważne problemy z generalizacją. Mimo że wprowadzenie warstw Dropout miało na celu ograniczenie przeuczenia, rezultaty wskazują na odwrotny efekt – model nie nauczył się rozróżniać ras i przypisuje wszystkie próbki do jednej klasy ("golden retriever").

Wykres 7. Macierz pomyłek modelu 2 na zbiorze testowym



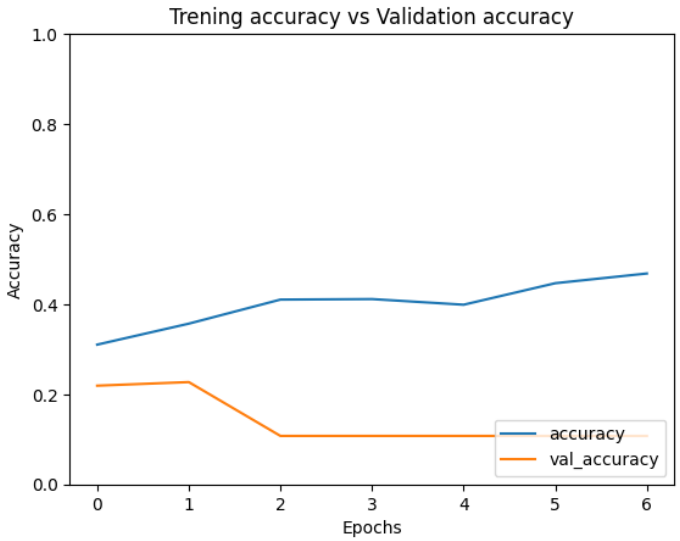
Model 3 - Sieć z Augmentacją Danych i Normalizacją

Model 3 został zaprojektowany jako rozwinięcie poprzednich wersji sieci konwolucyjnych. Kluczową zmianą było dodanie mechanizmu automatycznej augmentacji obrazów podczas treningu, co miało na celu zwiększenie różnorodności danych i poprawę zdolności generalizacji modelu. Dodatkowo, wprowadzono warstwy normalizacji BatchNormalization, które stabilizują proces uczenia i pomagają w efektywniejszym propagowaniu gradientów.

Augmentacja danych w zbiorze treningowym pozowliła na wprowadzenie losowcyh operacji m.in. obrót obrazów (do 20°), przesunięcia w pionie i poziomie (do 20% szerokości/wysokości), transformacja "shear", skalowanie (zoom), odbicie poziome czy regulacja jasności. Te operacje miały na celu stworzenie bardziej zróżnicowanego zestawu treningowego i poprawę zdolności modelu do rozpoznawania psów w różnych warunkach. Dodatkowo dodano warstwy normalizacji BatchNormalization. Umieszczono je po każdej warstwie konwolucyjnej w celu poprawy stabilności uczenia i przyspieszenia konwergencji. Zastosowano także warstwy GlobalAveragePooling2D zamiast pełnej warstwy spłaszczającej (Flatten), co pozwoliło na znaczną redukcję liczby parametrów, a tym samym ograniczenie ryzyka przeuczenia. Została zwiększona także liczba filtrów w kolejnych warstwach konwolucyjnych – wprowadzono 256 filtrów w końcowej warstwie konwolucyjnej, co miało na celu wyłapanie bardziej złożonych wzorców oraz dropout (0.3) w warstwie gęstej – dodany w celu poprawy generalizacji modelu i uniknięcia przeuczenia.

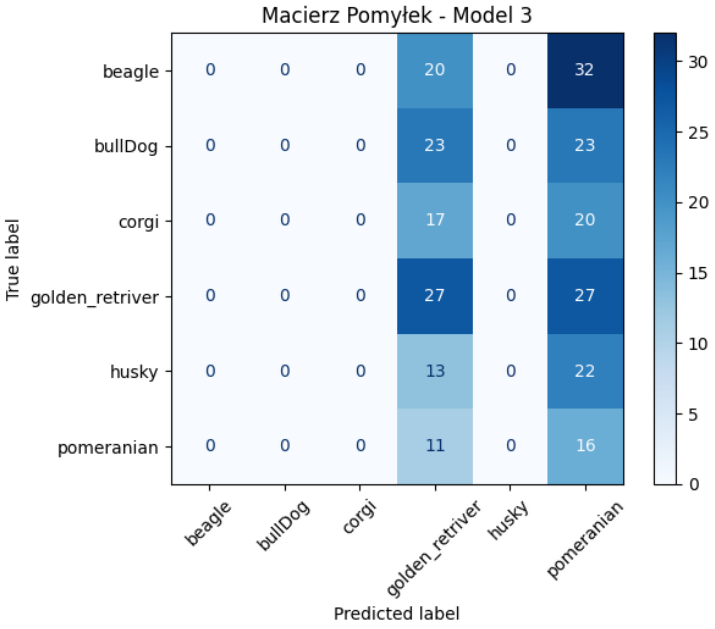
Na poniższym wykresie zaprezentowano krzywe uczenia dla tego modelu. Wyniki treningu modelu 3 pokazały wyraźny problem z generalizacją. Na zbiorze treningowym model osiągnął dokładność ok. 45,7%, co sugeruje, że nauczył się rozpoznawać pewne wzorce.

Wykres 8. Dokładność modelu 3 na zbiorze treningowym i walidacyjnym



Dokładność walidacyjna natomiast wynosiła jedynie 10,76%, co oznacza, że model kompletnie nie radził sobie z nowymi danymi. Co więcej, widać dużą rozbieżność między dokładnością treningową a walidacyjną, co wskazuje na przeuczenie – mimo zastosowania augmentacji i normalizacji model nie nauczył się generalizować. Czas treningu dla jednej epoki wynosił 22–25 sekund, co było zauważalnie dłuższe niż w poprzednich modelach. Wynikało to z dodatkowych obliczeń związanych z augmentacją oraz złożoną architekturą z normalizacją i większą liczbą filtrów konwolucyjnych.

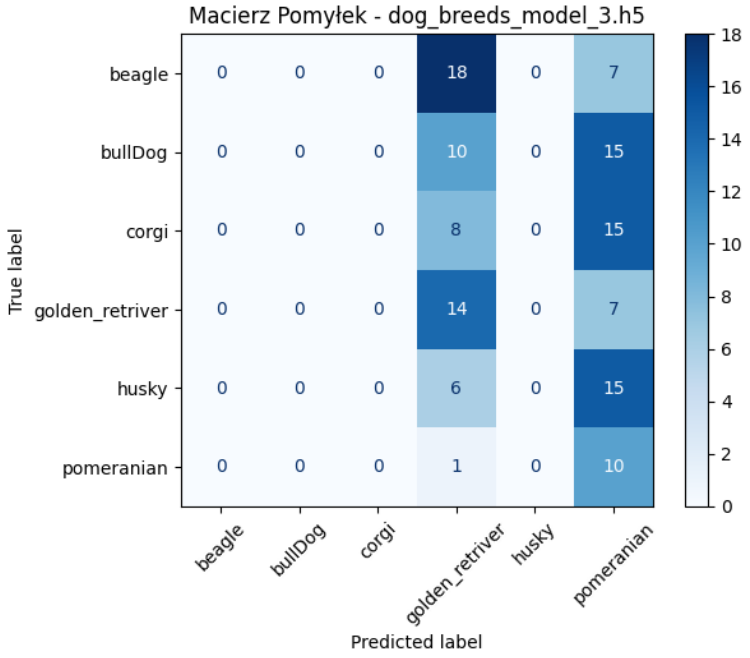
Wykres 9. Macierz pomyłek modelu 3 na zbiorze walidacyjnym



Macierz błędów na zbiorze walidacyjnym ujawniła bardzo istotny problem – model przypisywał niemal wszystkie obrazy do jednej klasy tak jak w modelu 2. Zamiast poprawnie rozróżniać rasy, model dominująco klasyfikował psy do jednej kategorii. To sugeruje, że sieć nie nauczyła się użytecznych reprezentacji cech – zamiast uczyć się różnic między rasami, zaczęła „zgadywać" jedną z kategorii, w której najczęściej pojawiały się obrazy. Regularizacja i augmentacja nie wystarczyły do poprawy wyników, co oznacza, że problem mógł wynikać z niewystarczającej ilości danych treningowych lub źle dobranej architektury.

Macierz pomyłek dla modelu 3 na zbiorze testowym ujawnia te same problemy jak w przypadku zbioru walidacyjnego czyli brak zrównoważonej klasyfikacji, dominację błędnej klasy

Wykres 10. Macierz pomyłek modelu 3 na zbiorze testowym



Podsumowując Pomimo zastosowania augmentacji i normalizacji, model nie poprawił zdolności generalizacji, co skutkowało niską dokładnością testową (19,05%). Wynik ten sugeruje, że architektura sieci i augmentacja danych mogły nie być wystarczające do skutecznej klasyfikacji. Możliwe, że problem wynikał z niedostatecznej liczby danych treningowych, co utrudniło modelowi nauczenie się zróżnicowanych cech każdej rasy. Model wciąż miał problem z generalizacją. Sieć stała się bardziej skomplikowana, ale nie poprawiło to dokładności – wręcz przeciwnie, błąd na zbiorze walidacyjnym sugerował, że model „utknął" w jednym wzorcu klasyfikacji.

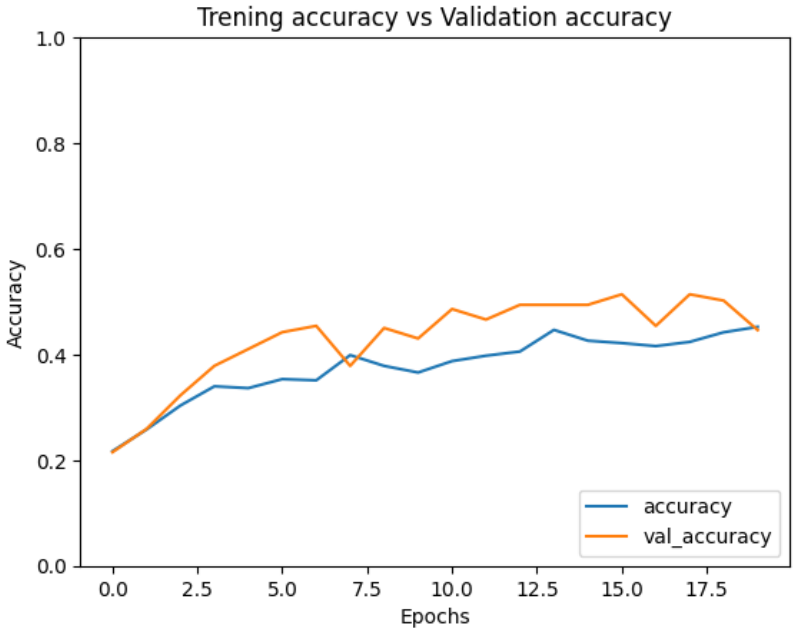
Model 4 - Redukcja złożoności sieci przy zachowaniu augmentacji

Model 4 stanowi uproszczenie w stosunku do Modelu 3, w którym sieć neuronowa była bardziej złożona, zawierała więcej warstw konwolucyjnych i operacji normalizacyjnych. W Modelu 4 usunięto najgłębsze warstwy ekstrakcji cech, pozostawiając jedynie dwie warstwy konwolucyjne. co miało na celu zmniejszenie liczby parametrów i skrócenie czasu trenowania,

co rzeczywiście miało miejsce – czas przetwarzania jednej epoki zmniejszył się do około 17-20 sekund, w porównaniu do 22-25 sekund w Modelu 3. Mimo zmniejszenia liczby warstw, pozostawiono augmentację danych, ponieważ w poprzednich modelach pozwalała ona na zwiększenie różnorodności zbioru treningowego, co w teorii powinno poprawić zdolność modelu do generalizacji.

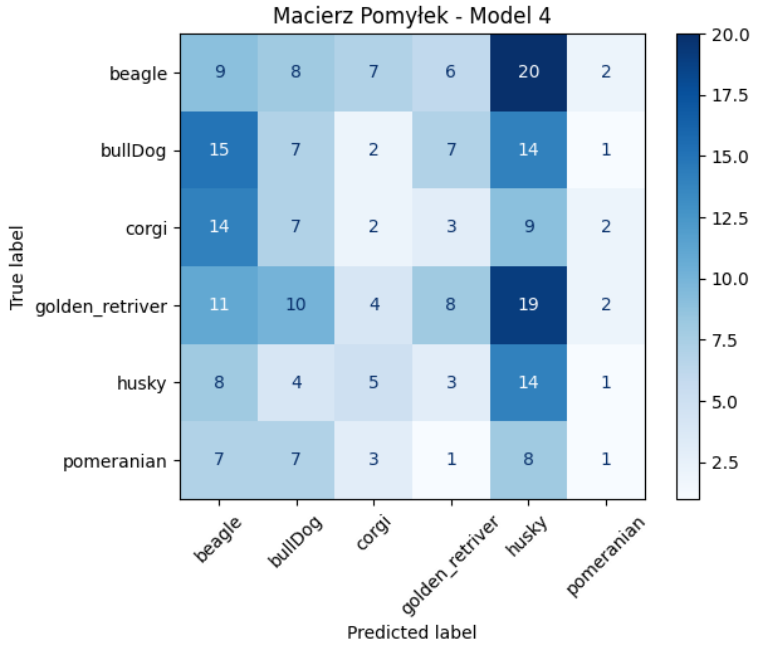
Analizując wykres dokładności treningowej i walidacyjnej, można zauważyć kilka istotnych trendów. W pierwszych kilku epokach model poprawnie uczył się wzorców – dokładność zarówno na zbiorze treningowym, jak i walidacyjnym rosła w sposób zbliżony. Model osiągnął maksymalną dokładność walidacyjną na poziomie około 51%, co jest lepszym wynikiem niż w Modelu 3 (gdzie dokładność zatrzymała się na ~22%). Pomimo wzrostu dokładności na zbiorze treningowym, w późniejszych epokach nastąpiło pogorszenie wyników walidacyjnych, co może sugerować pewne przeuczenie modelu. W porównaniu do Modelu 3 uproszczenie architektury przyczyniło się do bardziej stabilnego uczenia się modelu. Wyższa dokładność na zbiorze walidacyjnym oznacza, że model miał lepszą zdolność do generalizacji.

Wykres 11. Dokładność modelu 4 na zbiorze treningowym i walidacyjnym



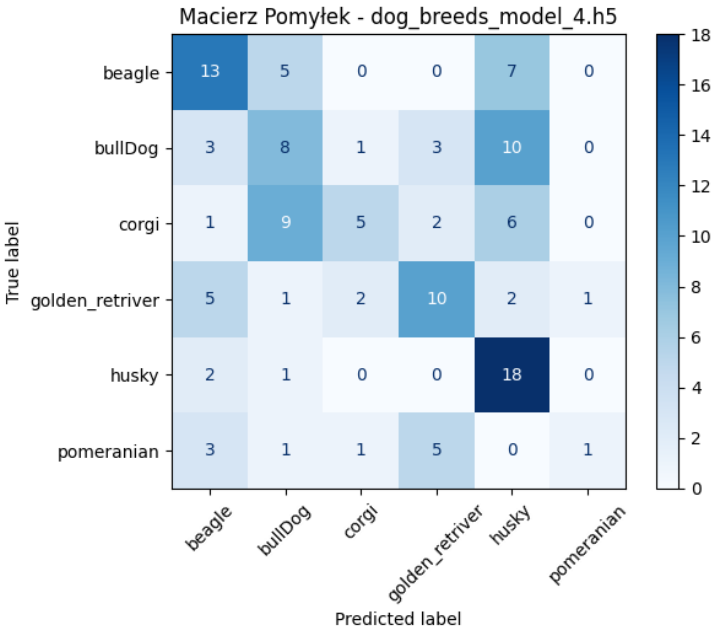
Analiza macierzy pomyłek pokazuje, że model miał trudności w poprawnej klasyfikacji psów każdej rasy, jednak częściej niż w Modelu 3 dokonywał poprawnych predykcji. Szczególnie często błędnie klasyfikował bulldogi i corgi – te rasy były zamiennie przypisywane do innych klas. Część ras, jak golden retriever i beagle, również była często mylona. W porównaniu do Modelu 3 zwiększyła się liczba poprawnie sklasyfikowanych przykładów. Model nadal ma problemy z pewnymi klasami, co sugeruje, że prosta architektura może nie być wystarczająca do skutecznego rozpoznawania ras.

Wykres 12. Macierz pomyłek modelu 4 na zbiorze walidacyjnym



W macierzy pomyłek dla zestawu testowego model wciąż dobrze klasyfikował husky (18 poprawnych), ale za to golden retrievery zaczęły być częściej błędnie klasyfikowane. Więcej błędów pojawiło się dla bulldogów i beagle - model miał trudności z ich jednoznacznym rozróżnieniem. Zwiększyła się również liczba błędów w pomeranianach co sugeruje, że model mógł nie nauczyć się dobrze cech tej rasy.

Wykres 13. Macierz pomyłek modelu 4 na zbiorze testowym



Model 4 wykazał się większą stabilnością niż wcześniejsze modele, co potwierdza mniejsza różnica między dokładnością walidacyjną (51%) a testową (43.65%)co oznacza, że bardziej efektywnie przetwarzał cechy obrazów i lepiej radził sobie z nowymi danymi. Choć nadal

występowały trudności w klasyfikacji niektórych ras, szczególnie bulldogów, beagle i pomeranianów, to model znacznie lepiej generalizował niż poprzednie wersje. W przeciwieństwie do wcześniejszych sieci, gdzie spadek dokładności między walidacją a testem był większy (10-15%), tutaj wyniósł jedynie 7.35%, co sugeruje lepszą zdolność do przewidywania na nowych danych. Redukcja liczby warstw pozwoliła na skrócenie czasu trenowania, a pozostawienie augmentacji pozwoliło poprawić zdolność modelu do generalizacji.

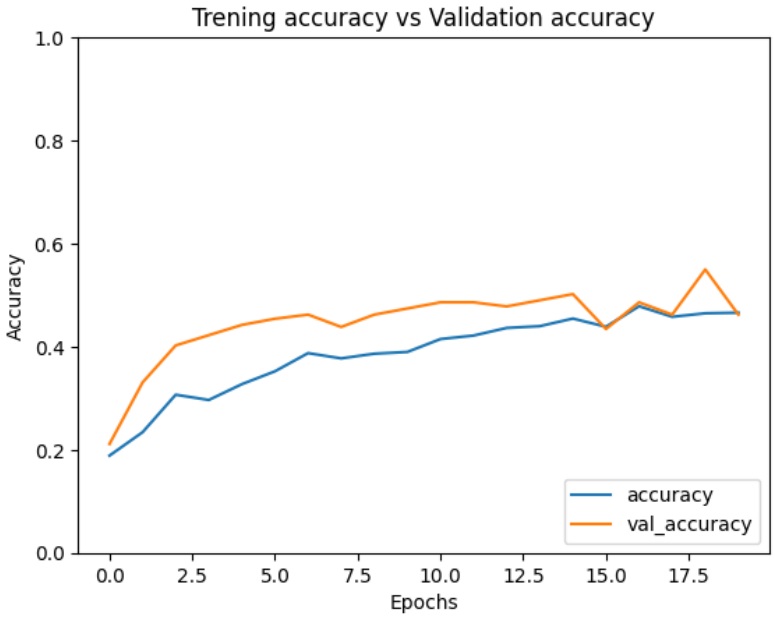
Model 5 - Optymalizacja modelu 4

Model 5 jest zoptymalizowaną wersją Modelu 4, zachowującą jego kluczowe założenia, ale wprowadzającą poprawki w architekturze, optymalizacji i regularyzacji. W porównaniu do wcześniejszych modeli wyróżnia się bardziej zrównoważonym podejściem, które poprawiło stabilność treningu oraz dokładność zarówno na zbiorze walidacyjnym, jak i testowym.

Podobnie jak Model 4, Model 5 składa się tylko z dwóch warstw konwolucyjnych i poolingowych. Jednak różni się optymalizacją parametrów sieci, co wpłynęło na lepszą skuteczność. Liczba filtrów konwolucyjnych pozostała na tym samym poziomie, czyli 32 filtry w pierwszej warstwie i 64 w drugiej, co okazało się wystarczające do ekstrakcji kluczowych cech ras psów. W dalszym ciągu zachowano warstwę Dropout (0.3) aby ograniczyć ryzyko przeuczenia oraz dodano warstwę Dropout przed warstwą w pełni połączoną. Zmniejszono liczbę neuronów w warstwie gęstej. Zastosowanoi 128 neuronów zamiast bardziej rozbudowanych warstw z wcześniejszych modeli. To przyspieszyło uczenie i zmniejszyło liczbę parametrów, co poprawiło generalizację. Dodatkowo w modelu 5 dodano mechanizm wczesnego zatrzymywania treningu, który pozwalał uniknąć przeuczenia. Trening zatrzymywano, gdy dokładność walidacyjna przestała się poprawiać przez 5 kolejnych epok. Zmniejszono także learning rate (0.001), który pozwolił na bardziej stabilne uczenie modelu. Wcześniejsze modele czasami zbyt szybko dopasowywały się do danych treningowych, co prowadziło do przeuczenia. W tym modelu pozostawiono augumentację.

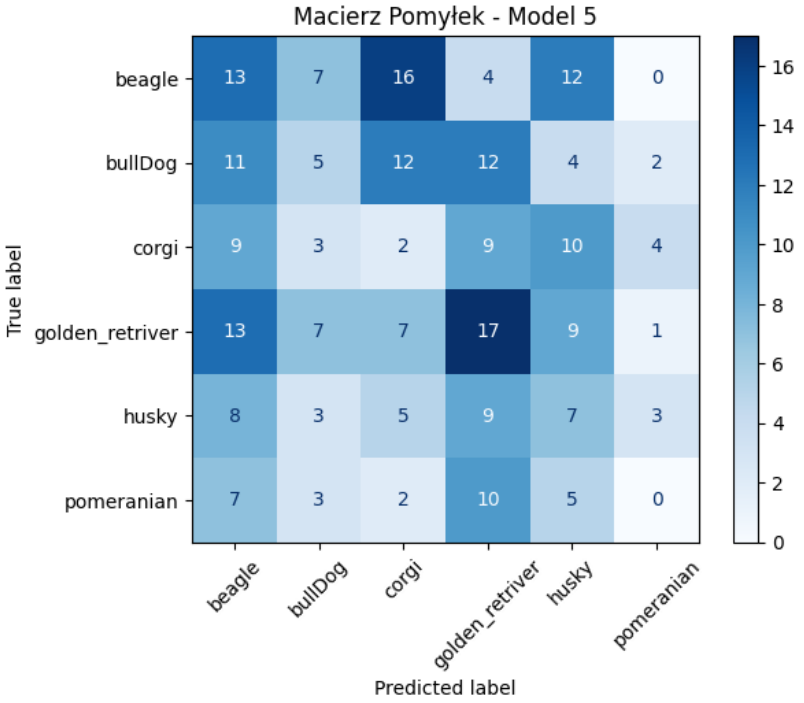
Model 5 wykazał stabilny przebieg uczenia bez gwałtownych skoków dokładności. W pierwszych epokach nastąpił stopniowy wzrost dokładności treningowej i walidacyjnej, co wskazuje na prawidłowe dopasowywanie się modelu do danych. W środkowej fazie uczenia tempo poprawy spowolniło, a pod koniec procesu osiągnęło maksimum na poziomie około 55% dokładności walidacyjnej. Zastosowanie EarlyStopping pozwoliło zatrzymać trening, gdy nie było już dalszej poprawy, zapobiegając przeuczeniu. W porównaniu do wcześniejszych modeli Model 5 wykazał lepszą równowagę między dokładnością treningową a walidacyjną, co oznacza, że dobrze uogólniał nauczone wzorce. Podsumowując, Model 5 był najbardziej skuteczny spośród testowanych wersji, łącząc stabilność, brak przeuczenia i lepszą zdolność generalizacji na nowe próbki.

Wykres 14. Dokładność modelu 5 na zbiorze treningowym i walidacyjnym



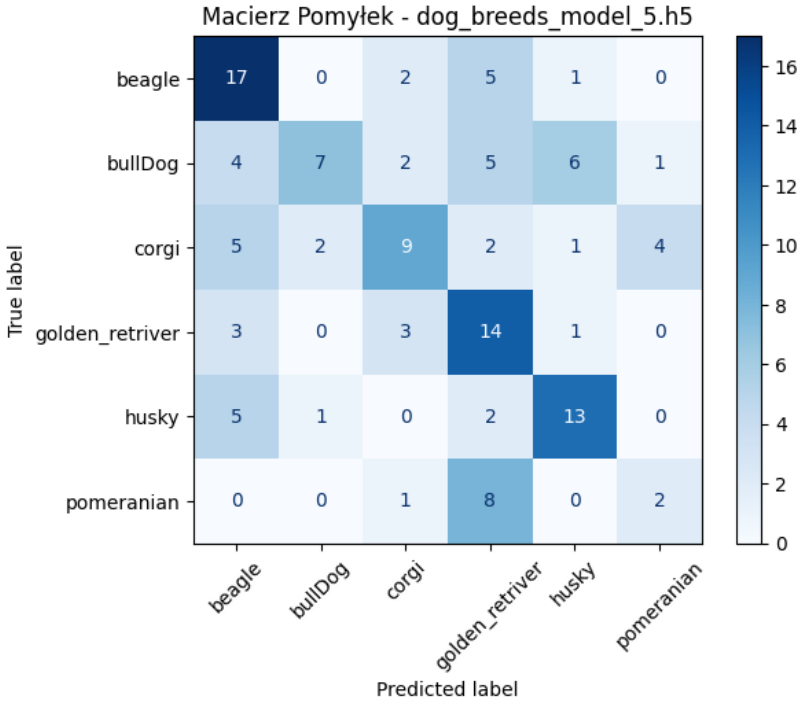
Macierz pomyłek dla Modelu 5 na zbiorze walidacyjnym pokazuje pewne postępy w klasyfikacji w porównaniu do poprzednich modeli, ale nadal występują liczne błędne przypisania. Klasa golden retriever jest rozpoznawana stosunkowo najlepiej, ale wiele innych ras jest często błędnie klasyfikowanych jako corgi lub golden retriever. Przykładowo: Beagle jest często mylony z corgi i golden retrieverem. Bulldog również jest często klasyfikowany jako corgi. Z kolei Husky ma bardziej rozproszoną liczbę błędnych klasyfikacji, co może wskazywać na trudność w odróżnieniu tej rasy od innych. Pomimo dodania augmentacji i ulepszenia modelu, niektóre rasy nadal nie są rozpoznawane poprawnie, co może wskazywać na potrzebę dalszej optymalizacji, np. lepszego doboru cech lub większej ilości danych.

Wykres 15. Macierz pomyłek modelu 5 na zbiorze walidacyjnym



Na zbiorze testowym Model 5 wykazał zauważalną poprawę w stosunku do wcześniejszych modeli, osiągając 49,21% dokładności. To najlepszy wynik spośród dotychczas testowanych architektur. Jednak, analiza macierzy pomyłek wskazuje, że błędy klasyfikacyjne nadal występują, ale są bardziej równomiernie rozłożone. Beagle został poprawnie sklasyfikowany w 17 przypadkach, ale nadal pojawiają się błędy, głównie w kierunku golden retrievera. Golden retriever osiągnął stosunkowo dobry wynik, ale część obrazów była błędnie przypisywana do corgi i bulldoga. Husky za to został poprawnie rozpoznany w większości przypadków, ale nadal zdarzają się pomyłki z beagle i bulldogiem. Poprawa wyników na zbiorze testowym w stosunku do walidacyjnego sugeruje, że model lepiej generalizuje i jest bardziej odporny na dane, których wcześniej nie widział. Ostatecznie, Model 5 wykazał największą stabilność spośród dotychczas analizowanych wariantów.

Wykres 16. Macierz pomyłek modelu 5 na zbiorze testowym



Podsumowanie modeli CNN

W trakcie eksperymentów przeanalizowano pięć różnych architektur CNN, stopniowo optymalizując model w celu poprawy dokładności klasyfikacji sześciu ras psów. Każdy kolejny model wprowadzał ulepszenia, takie jak regularyzacja, augmentacja danych czy modyfikacja liczby warstw, aby poprawić zdolność sieci do rozpoznawania cech istotnych dla klasyfikacji. Model 1 (bazowy CNN) to najprostszy model składający się z trzech warstw konwolucyjnych i warstw w pełni połączonych. Osiągał stosunkowo niską dokładność zarówno na zbiorze walidacyjnym, jak i testowym, a także wykazywał tendencję do przeuczenia. W celu ograniczenia przeuczenia wprowadzono warstwy Dropout w modelu 2, co poprawiło stabilność tego modelu. Niestety, poprawa nie była znacząca, a model nadal miał trudności z rozróżnianiem niektórych ras. Dla modelu 3 wprowadzenie dynamicznych transformacji obrazów poprawiło zdolność modelu do generalizacji. Pomimo lepszej odporności na przeuczenie, dokładność klasyfikacji nie wzrosła znacząco, a model nadal miał problem z jednoznacznym rozpoznawaniem niektórych klas. W modelu 4 natomiast zmniejszono liczbę

warstw i uproszczono architekturę. Miało na celu poprawę efektywności obliczeniowej i uniknięcie nadmiernego dopasowania do zbioru treningowego. Wyniki były bardziej stabilne, a model osiągnął lepszą równowagę między dokładnością treningową a walidacyjną. Model 5 to ulepszony CNN z regularyzacją i optymalizacją parametrów. Najlepszy spośród testowanych wariantów. Wykorzystał dropout oraz augmentację danych przy nadal zmniejszonej liczbie warstw, co pozwoliło osiągnąć najwyższą dokładność na zbiorze testowym (49,21%). Model miał lepszą generalizację i mniej błędnych klasyfikacji niż wcześniejsze wersje.

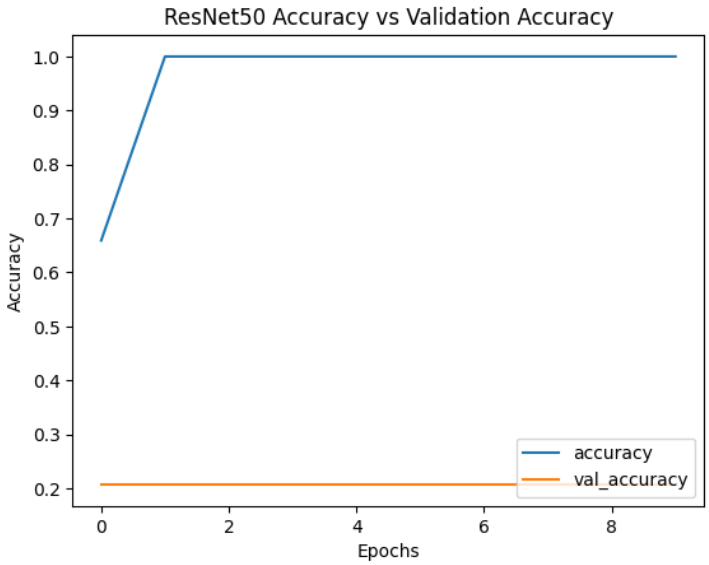
Podsumowując, testowane modele CNN wykazały, że sama głębsza sieć niekoniecznie prowadzi do lepszych wyników. Kluczowe znaczenie miały regularyzacja, augmentacja oraz odpowiednie dobranie architektury do problemu. Model 5 okazał się najlepszy, ale dokładność nadal pozostawia pole do dalszych ulepszeń, co będzie analizowane przy użyciu bardziej zaawansowanych architektur, takich jak ResNet i VGG.

ResNet50

Model ResNet50 to głęboka sieć neuronowa, która składa się z 50 warstw i wykorzystuje mechanizm residual learning. Kluczowym elementem tej architektury są bloki resztkowe, które umożliwiają skuteczniejsze trenowanie bardzo głębokich sieci, minimalizując problem zanikania gradientu. W tym przypadku wykorzystano pretrenowaną wersję modelu na zbiorze ImageNet, co oznacza, że wagi warstw konwolucyjnych zostały zamrożone i nie były aktualizowane podczas treningu. Dzięki temu model mógł wykorzystywać wcześniej wyuczone cechy, jednak nie dostosowywał ich do specyficznych wzorców ras psów. Na końcu modelu dodano warstwę Global Average Pooling (GAP), która zamiast tradycyjnych warstw w pełni połączonych pozwala na redukcję liczby parametrów i lepszą generalizację. Po niej umieszczono gęstą warstwę z 256 neuronami aktywowanymi funkcją ReLU, której zadaniem było przekształcenie wyodrębnionych cech w reprezentację gotową do klasyfikacji. Aby ograniczyć przeuczenie, zastosowano Dropout (0.3), który losowo deaktywował część neuronów w trakcie treningu. Ostatnim elementem była warstwa wyjściowa, składająca się z 6 neuronów aktywowanych funkcją softmax, które odpowiadały za przypisanie obrazu do jednej z sześciu klas psów. Zamrożenie wag miało na celu zachowanie wyuczonych cech ogólnych, jednak ograniczyło możliwość dostosowania modelu do konkretnego zadania. W efekcie model wykazywał bardzo wysoką dokładność na zbiorze treningowym, ale nie potrafił poprawnie klasyfikować obrazów w zbiorach walidacyjnym i testowym, co świadczyło o całkowitym przeuczeniu.

Podczas treningu modelu ResNet50 zaobserwowano znaczące różnice między dokładnością na zbiorze treningowym a walidacyjnym. Już po drugiej epoce model osiągnął 100% dokładność na zbiorze treningowym, co wskazuje na całkowite przeuczenie. Jednocześnie dokładność walidacyjna pozostała na stałym poziomie około 20%, co oznacza, że model nie nauczył się właściwego rozpoznawania ras psów w nowych, nieznanych danych. ResNet50, mimo dużej liczby warstw, był stosunkowo szybki dzięki wykorzystaniu pretrenowanych wag i zamrożeniu większości warstw. Trening każdej epoki trwał około 34-40 sekund, co było znacznie krótsze niż w przypadku wcześniejszych modeli CNN, gdzie czas przetwarzania jednej epoki dochodził nawet do kilku minut.

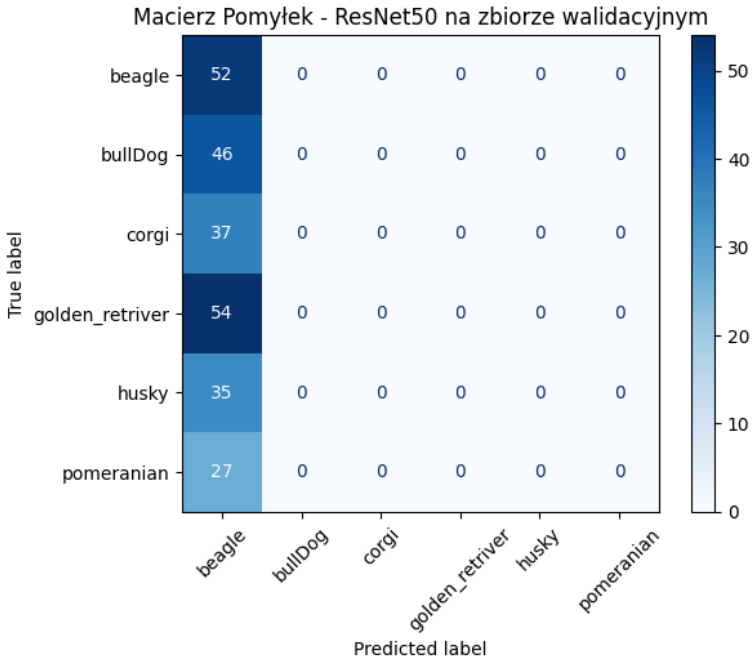
Wykres 17. Dokładność ResNet50 na zbiorze treningowym i walidacyjnym



Na wykresie przedstawiającym dokładność treningową i walidacyjną w kolejnych epokach widoczny jest gwałtowny wzrost dokładności treningowej, która już w drugiej epoce osiągnęła maksymalny możliwy poziom. W przeciwieństwie do tego dokładność walidacyjna utrzymywała się bez żadnej poprawy, co wskazuje na brak uogólnienia modelu. Z kolei funkcja straty walidacyjnej stale rosła, co potwierdza, że model przestał uczyć się nowych wzorców i całkowicie dopasował się do zbioru treningowego. Podsumowując, model ResNet50 nie zdołał uogólnić nauki na zbiór walidacyjny, co skutkowało niską jakością klasyfikacji nowych danych. Powodem było prawdopodobnie zamrożenie wszystkich warstw konwolucyjnych, przez co model nie mógł dostosować się do specyficznych cech psów z użytego zestawu danych.

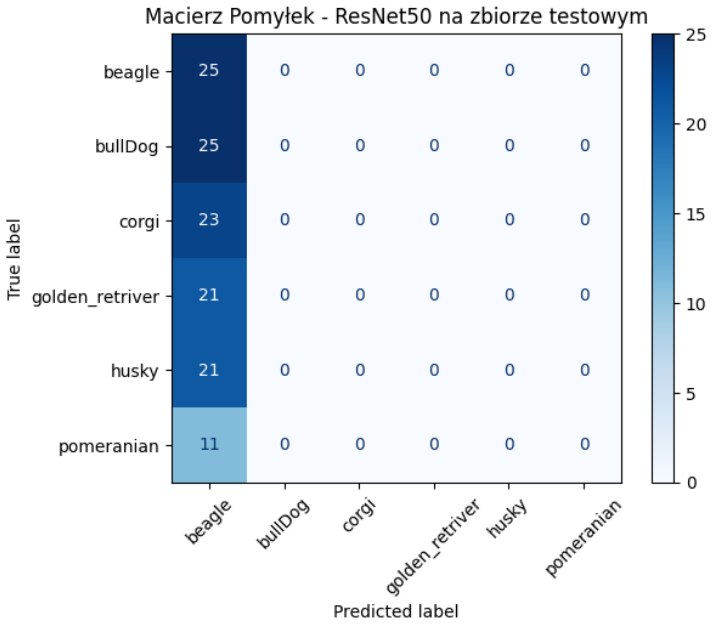
Macierz pomyłek na zbiorze walidacyjnym pokazuje, że model nie nauczył się poprawnie klasyfikować żadnej z klas. Wszystkie próbki zostały przypisane do jednej kategorii (beagle), co oznacza, że model nie rozróżniał cech charakterystycznych innych ras psów. Jest to typowe dla sytuacji, w której model nie został dostatecznie dopasowany do nowego zadania po transferze uczenia.

Wykres 18. Macierz pomyłek ResNet50 na zbiorze walidacyjnym



Analogiczny efekt widoczny jest w macierzy pomyłek na zbiorze testowym, gdzie ponownie wszystkie próbki zostały zaklasyfikowane do jednej kategorii. Oznacza to, że model nie wykorzystywał ekstrakcji cech do klasyfikacji i najprawdopodobniej "nauczył się" przypisywać każdą próbkę do jednej dominującej klasy.

Wykres 19. Macierz pomyłek ResNet50 na zbiorze walidacyjnym



Wcześniejsze modele CNN, mimo że miały niższą dokładność treningową, wykazywały pewien stopień zdolności do rozróżniania ras. Nawet jeśli pojawiały się liczne błędy, klasyfikacja była bardziej zróżnicowana. W przypadku ResNet50 model wykazał brak jakiejkolwiek zdolności do generalizacji, co sprawia, że jego predykcje były bezużyteczne w tym kontekście. Model ResNet50, mimo potężnej architektury, nie dostosował się do specyfiki zadania i całkowicie przeuczył się na zbiorze treningowym, co skutkowało katastrofalnie niską jakością predykcji na zbiorach walidacyjnym i testowym. Aby poprawić wyniki, należałoby odmrozić część warstw i zastosować fine-tuning, tak aby sieć mogła dostosować się do rozpoznawania ras psów na podstawie dostępnych danych.

VGG

Model VGG16 to głęboka sieć konwolucyjna składająca się z 16 warstw (konwolucyjnych i w pełni połączonych), pierwotnie wytrenowana na zbiorze ImageNet. W naszym eksperymencie wykorzystaliśmy transfer learning, gdzie zamroziliśmy wszystkie warstwy ekstrakcji cech z VGG16 i dodaliśmy własne warstwy w pełni połączone.

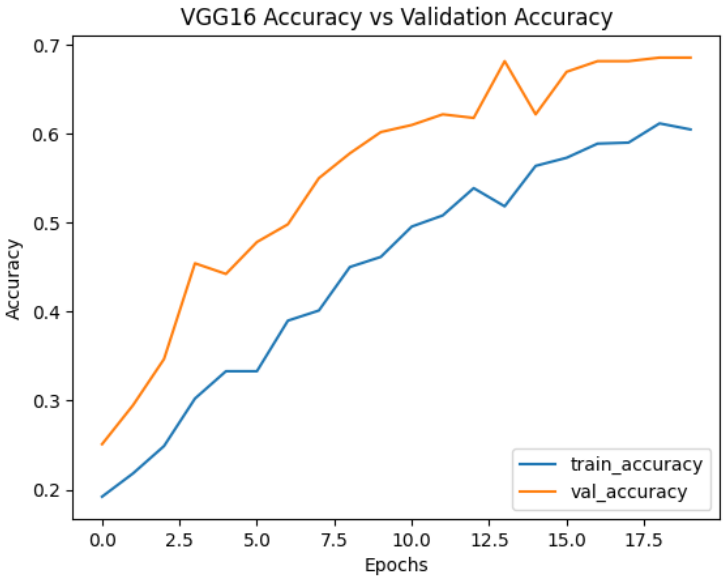
Architektura modelu składa się z:

- Zamrożonego bloku konwolucyjnego VGG16, który przetwarza obrazy o wymiarach 224x224x3.

- Global Average Pooling, który zmniejsza wymiar wyjścia i ułatwia dalsze przetwarzanie.
- Gęstej warstwy z 256 neuronami i aktywacją ReLU, która uczy się nowych reprezentacji istotnych dla klasyfikacji ras psów.
- Dropout (30%), aby zapobiec przeuczeniu.
- Warstwy wyjściowej (Softmax) z 6 neuronami, odpowiadającej liczbie klas psów.

Podczas treningu modelu VGG16 obserwowano stopniowy wzrost dokładności zarówno na zbiorze treningowym, jak i walidacyjnym. Początkowo model osiągał niskie wartości dokładności, jednak z każdą kolejną epoką poprawiał swoje wyniki. Już po kilku epokach dokładność walidacyjna zaczęła przewyższać dokładność treningową, co sugerowało dobrą zdolność do generalizacji.W kolejnych epokach nastąpiło dalsze polepszanie wyników – dokładność na zbiorze walidacyjnym wzrosła do około 68,5%, co jest znaczącą poprawą w porównaniu do wcześniejszych modeli. Wartość funkcji straty sukcesywnie malała, co potwierdzało skuteczność procesu uczenia. Czas trenowania każdej epoki wynosił średnio 70-74 sekundy, co było dłuższym czasem niż w prostszych modelach CNN, ale wciąż znacznie krótszym niż w przypadku ResNet50. Model wykazywał stabilny wzrost dokładności, a brak gwałtownych spadków wskazywał na dobrze dobraną architekturę oraz skuteczne wykorzystanie transfer learningu.

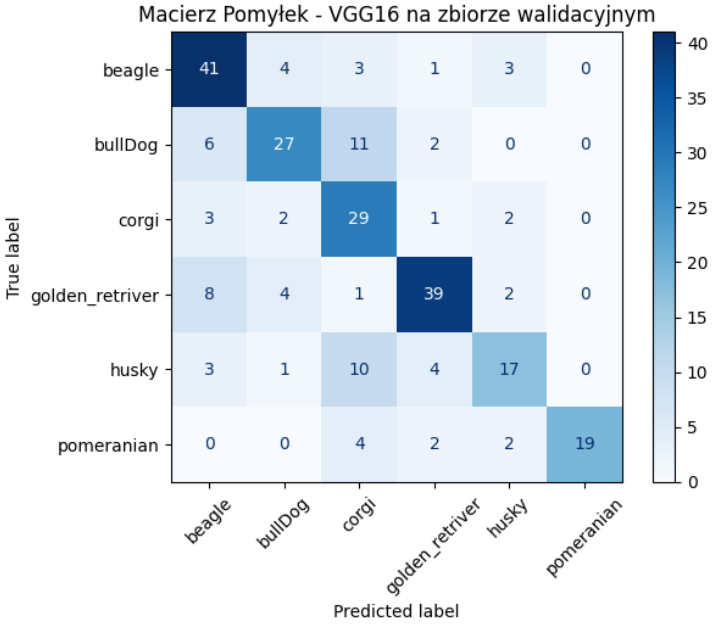
Wykres 20. Dokładność VGG16 na zbiorze treningowym i walidacyjnym



Model VGG16 okazał się najbardziej skuteczny spośród testowanych architektur, osiągając wysoką dokładność zarówno na zbiorze walidacyjnym (68.53%), jak i testowym (61.90%). Jego zdolność do generalizacji jest znacznie lepsza niż w przypadku ResNet50, który miał problem z poprawnym klasyfikowaniem obrazów.

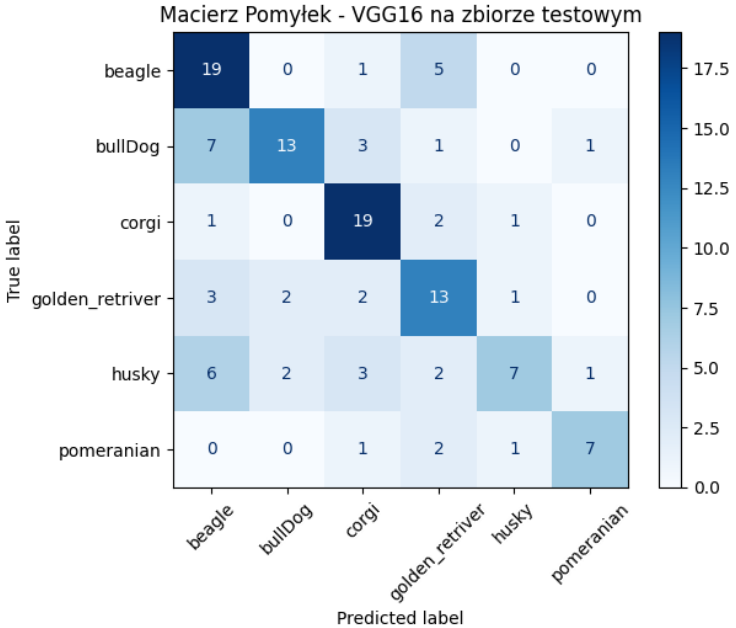
Podczas analizy macierzy pomyłek można zauważyć, że model bardzo dobrze rozpoznaje rasy takie jak Beagle, Golden Retriever oraz Corgi, co świadczy o skutecznym przetwarzaniu charakterystycznych cech tych psów. Nieco większe trudności pojawiają się w przypadku Bulldoga i Husky, które są często mylone z innymi rasami. Wyniki na zbiorze testowym potwierdzają ten trend – model zachowuje wysoką skuteczność, choć nadal występują błędy w rozpoznawaniu podobnych wizualnie ras.

Wykres 21. Macierz pomyłek VGG16 na zbiorze walidacyjnym



W porównaniu do wcześniejszych modeli CNN i ResNet50, VGG16 lepiej radzi sobie z klasyfikacją obrazów, prawdopodobnie dzięki głębszej strukturze i bardziej wyrafinowanemu ekstraktowi cech. Chociaż wciąż istnieją pewne obszary do poprawy, model ten wykazuje największy potencjał w zadaniu rozpoznawania ras psów.

Wykres 22. Macierz pomyłek VGG16 na zbiorze walidacyjnym



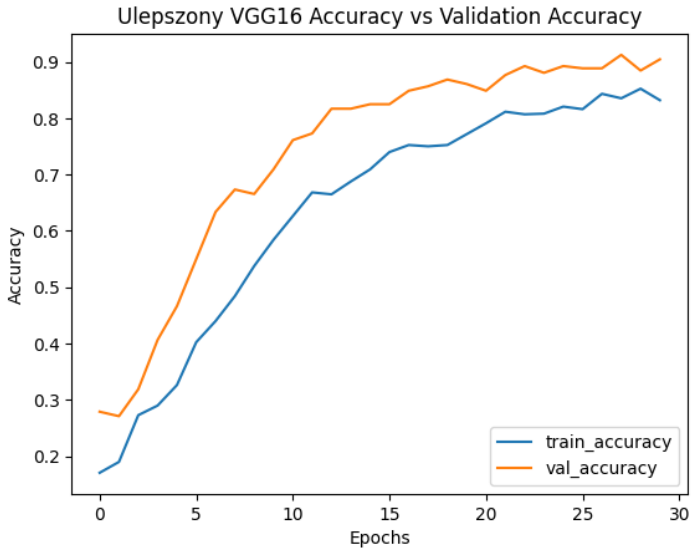
VGG – Ulepszony

W ulepszonym modelu VGG16 wprowadzono kilka kluczowych zmian, które miały na celu poprawę dokładności klasyfikacji ras psów oraz lepszą generalizację modelu. Przede wszystkim zastosowano bardziej zaawansowaną augmentację danych, co pozwoliło na zwiększenie różnorodności obrazów w zbiorze treningowym. Zwiększono zakres

transformacji, takich jak rotacja, przesunięcia poziome i pionowe oraz zoom, co pomogło modelowi lepiej radzić sobie z różnymi wariantami tych samych obrazów. Kolejną istotną zmianą było częściowe odmrożenie ostatnich czterech warstw VGG16. W poprzednim modelu wszystkie warstwy były zamrożone, co ograniczało zdolność modelu do dostosowywania się do konkretnego problemu. Odblokowanie kilku warstw pozwoliło na lepsze wykorzystanie wcześniej wyuczonych cech i dostosowanie ich do klasyfikacji ras psów. Dodatkowo, zwiększono liczbę neuronów w warstwach gęstych. Do modelu dodano większą warstwę z 512 neuronami przed dotychczasową warstwą 256-neuronową, co zwiększyło zdolność modelu do reprezentowania skomplikowanych wzorców. Aby ograniczyć ryzyko przeuczenia, zwiększono dropout do 0.4 w pierwszej warstwie gęstej, co poprawiło stabilność modelu. Wprowadzono również bardzo niski learning rate (1e-5), co spowolniło proces uczenia, ale jednocześnie zapewniło większą precyzję w dostrajaniu wag. Dodatkowo zastosowano mechanizm ReduceLROnPlateau, który automatycznie zmniejsza tempo uczenia, gdy model przestaje poprawiać wyniki.

Wszystkie te zmiany, w połączeniu z dłuższym treningiem (30 epok zamiast 20), przyczyniły się do znacznej poprawy dokładności. Model osiągnął 90,44% dokładności na zbiorze walidacyjnym i 85,71% na zbiorze testowym, co czyni go najlepszym spośród wszystkich testowanych architektur.

Wykres 23. Dokładność ulepszonym VGG16 na zbiorze treningowym i walidacyjnym



W ulepszonym modelu VGG16 wprowadzono kilka kluczowych zmian, które miały na celu poprawę dokładności klasyfikacji ras psów oraz lepszą generalizację modelu. Przede wszystkim zastosowano bardziej zaawansowaną augmentację danych, co pozwoliło na zwiększenie różnorodności obrazów w zbiorze treningowym. Zwiększono zakres transformacji, takich jak rotacja, przesunięcia poziome i pionowe oraz zoom, co pomogło modelowi lepiej radzić sobie z różnymi wariantami tych samych obrazów.

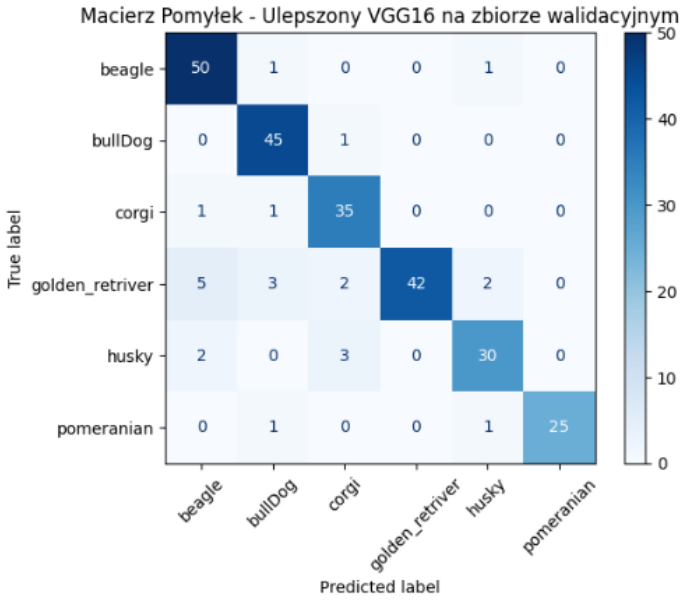
Kolejną istotną zmianą było częściowe odmrożenie ostatnich czterech warstw VGG16. W poprzednim modelu wszystkie warstwy były zamrożone, co ograniczało zdolność modelu do dostosowywania się do konkretnego problemu. Odblokowanie kilku warstw pozwoliło na lepsze wykorzystanie wcześniej wyuczonych cech i dostosowanie ich do klasyfikacji ras psów.

Dodatkowo, zwiększono liczbę neuronów w warstwach gęstych. Do modelu dodano większą warstwę z 512 neuronami przed dotychczasową warstwą 256-neuronową, co zwiększyło zdolność modelu do reprezentowania skomplikowanych wzorców. Aby ograniczyć ryzyko przeuczenia, zwiększono dropout do 0.4 w pierwszej warstwie gęstej, co poprawiło stabilność modelu.

Wprowadzono również bardzo niski learning rate (1e-5), co spowolniło proces uczenia, ale jednocześnie zapewniło większą precyzję w dostrajaniu wag. Dodatkowo zastosowano mechanizm ReduceLROnPlateau, który automatycznie zmniejsza tempo uczenia, gdy model przestaje poprawiać wyniki. Wszystkie te zmiany, w połączeniu z dłuższym treningiem (30 epok zamiast 20), przyczyniły się do znacznej poprawy dokładności. Model osiągnął 90,44% dokładności na zbiorze walidacyjnym i 85,71% na zbiorze testowym, co czyni go najlepszym spośród wszystkich testowanych architektur.

Ulepszony model VGG16 znacząco poprawił swoje wyniki w klasyfikacji ras psów, co można zauważyć na podstawie macierzy pomyłek zarówno dla zbioru walidacyjnego, jak i testowego. W zbiorze walidacyjnym model bardzo dobrze radził sobie z poprawnym rozpoznawaniem większości ras. Beagle został prawidłowo sklasyfikowany w pięćdziesięciu przypadkach, popełniając jedynie dwie pomyłki, które przypisały go do bulldoga i pomeraniana. Bulldog również uzyskał wysoki wynik, z czterdziestoma pięcioma poprawnymi klasyfikacjami i tylko dwoma błędami. Golden Retriever miał czterdzieści dwa prawidłowe przypisania, a corgi trzydzieści pięć, co również wskazuje na wysoką skuteczność modelu. Najmniej błędów wystąpiło w przypadku pomeraniana, który został niemal zawsze poprawnie sklasyfikowany, co świadczy o łatwości rozpoznania tej rasy przez model.

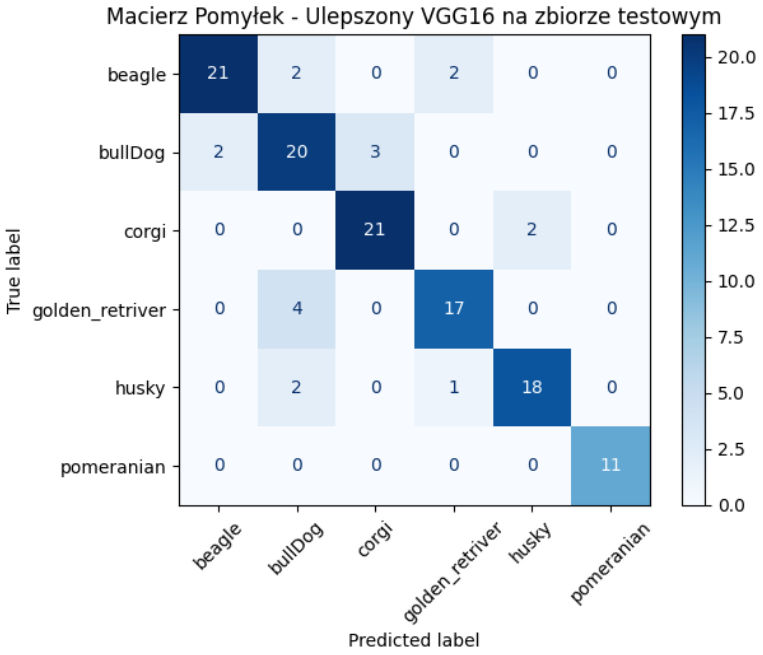
Wykres 24. Macierz pomyłek ulepszonego VGG16 na zbiorze walidacyjnym



Podobna sytuacja występuje w zbiorze testowym, gdzie model wykazał wysoką skuteczność klasyfikacji. Beagle został prawidłowo sklasyfikowany w dwudziestu jeden przypadkach, natomiast bulldog w dwudziestu, co wskazuje na dużą pewność modelu w rozpoznawaniu tych ras. Corgi osiągnął wysoki wynik dwudziestu jeden poprawnych

klasyfikacji, a golden retriever siedemnaście. Husky został poprawnie rozpoznany osiemnaście razy, a pomeranian w jedenastu przypadkach, co oznacza, że model niemal całkowicie wyeliminował błędy w jego klasyfikacji.

Wykres 25. Macierz pomyłek ulepszonego VGG16 na zbiorze testowym



Ogólnie rzecz biorąc, ulepszony VGG16 wykazuje znacznie lepszą skuteczność niż wcześniejsze modele. Liczba błędnych klasyfikacji została mocno ograniczona, a model nie wykazuje tendencji do mylenia konkretnych ras w sposób systematyczny. To wskazuje na skuteczniejszą generalizację i większą zdolność do rozpoznawania unikalnych cech wizualnych każdej klasy. Model doskonale odróżnia poszczególne rasy, co widać szczególnie w niewielkiej liczbie pomyłek na zbiorze testowym.

Poniżej przedstawiono przykładowe obrazy, które model VGG16 w ulepszonej wersji, źle zaklasyfikował.

Wnioski

Przeprowadzone eksperymenty pozwoliły na ocenę skuteczności różnych modeli CNN w zadaniu klasyfikacji ras psów. Testowane były zarówno podstawowe konwolucyjne sieci neuronowe (CNN), jak i zaawansowane modele transfer learning, takie jak ResNet50 oraz VGG16. W toku analizy stopniowo ulepszano architekturę sieci oraz techniki przetwarzania danych, co skutkowało zauważalnym wzrostem dokładności.

Podstawowe konwolucyjne sieci neuronowe (CNN) wykazały znaczące ograniczenia w klasyfikacji ras psów. Model bez zastosowania augmentacji danych charakteryzował się bardzo niską dokładnością zarówno na zbiorze walidacyjnym, jak i testowym, co wskazuje na niedouczenie modelu i jego niską zdolność do generalizacji. Dodanie augmentacji (Model 3) poprawiło ogólną wydajność, jednak dokładność wciąż była niska – model nadal miał trudności z prawidłowym rozpoznawaniem klas i wykazywał oznaki nadmiernego dopasowania do zbioru treningowego. Zmniejszenie architektury sieci (Model 4) przyniosło zauważalne usprawnienie. Dokładność modelu na zbiorze testowym wzrosła do 43,65%, co oznaczało lepszą generalizację w porównaniu do poprzednich wariantów. Niemniej jednak, wciąż pojawiały się liczne błędy klasyfikacyjne, a model nie był w stanie efektywnie rozróżniać wszystkich kategorii. Dalsza optymalizacja (Model 5) doprowadziła do jeszcze lepszego wyniku – 49,21% na zbiorze testowym, co było najlepszym wynikiem spośród wszystkich testowanych modeli CNN. Mimo to, model nadal nie osiągał wystarczającej dokładności do praktycznego zastosowania.

Tabela 3. Porównanie modeli

Model	Dokładność walidacyjna	Dokładność testowa	Przeuczenie	Zastosowane Ulepszenia
CNN 1	37,0	37,3	Wysokie	Brak
CNN 2	21,5	16,7	Średnie	Dropout
CNN 3	22,7	19,1	Średnie	Augmentacja

CNN 4	51,2	43,7	Lepsza Generalizacja	Mniejsza architektura
CNN 5	55,0	49,2	Lepsza Generalizacja	Optymalizacja architektury
ResNet50	20,7	19,8	Wysokie	Transfer Learning
VGG16	68,5	61,9	Niskie	Transfer Learning
Ulepszony VGG16	90,4	85,7	Niskie	Fine-tuning, Zaawansowana Augmentacja

Z kolei model ResNet50, mimo swojej zaawansowanej architektury, nie sprawdził się w tym zadaniu. Już po dwóch epokach osiągnął 100% dokładność na zbiorze treningowym, jednak jego dokładność na zbiorze walidacyjnym i testowym utrzymywała się na poziomie około 20%, co sugeruje silne przeuczenie. ResNet50 kompletnie nie generalizował do nowych danych – analiza macierzy pomyłek wykazała, że model przypisywał wszystkie obrazy do jednej kategorii. Może to wynikać z nieodpowiedniego dostosowania parametrów modelu lub niewystarczającej ilości danych do skutecznego treningu na tej architekturze. Natomiast Model VGG16, w przeciwieństwie do ResNet50, osiągnął znacznie lepsze wyniki. Jego dokładność na zbiorze walidacyjnym wyniosła 68,53%, a na zbiorze testowym 61,90%, co czyniło go najlepszym spośród dotychczas testowanych modeli. Analiza macierzy pomyłek wykazała, że model był w stanie skutecznie rozpoznawać większość klas, choć pojawiały się przypadki błędnej klasyfikacji, zwłaszcza w rasach o podobnym wyglądzie. Ze względu na sukces VGG16, przeprowadzono dodatkowy eksperyment z jego ulepszoną wersją. Wprowadzono zaawansowaną augmentację danych, odblokowano ostatnie warstwy modelu i zmniejszono learning rate, aby dostosować wagę modelu do nowego zadania. Ulepszony VGG16 osiągnął 90,44% dokładności na zbiorze walidacyjnym oraz 85,71% na zbiorze testowym, co czyni go najlepszym modelem spośród wszystkich testowanych. Macierze pomyłek wskazały na minimalne błędy klasyfikacyjne, a model poprawnie rozpoznawał niemal wszystkie klasy.

Tabela 4. Porównanie sieci

Model	Warstwy konwolucyjne	Dropout	Augmentacja	Batch Normalization	Pooling
CNN 1	3	Brak	Brak	Nie	MaxPooling
CNN 2	4	0.3	Brak	Tak	MaxPooling
CNN 3	4	0.3	Tak	Tak	MaxPooling
CNN 4	2	0.3	Tak	Nie	MaxPooling
CNN 5	2	0.3	Tak	Nie	MaxPooling
ResNet50	Transfer Learning	0.3	Tak	Tak	GlobalAveragePooling
VGG16	Transfer Learning	0.3	Tak	Tak	GlobalAveragePooling
VGG16 Ulepszony	Transfer Learning + Fine-Tuning	0.4 + 0.3	Zaawansowana	Tak	GlobalAveragePooling

Możliwe kierunki dalszych usprawnień

Pomimo znacznych ulepszeń wprowadzonych w kolejnych modelach, wyniki wskazują, że istnieją jeszcze obszary, w których można poprawić skuteczność klasyfikacji ras psów. Najlepsze rezultaty osiągnął ulepszony model VGG16, uzyskując dokładność 90.44% na zbiorze walidacyjnym oraz 85.71% na zbiorze testowym. Mimo to istnieje kilka możliwości dalszej optymalizacji, które mogłyby wpłynąć na jeszcze lepsze dopasowanie modelu do zadania klasyfikacji. Jednym z aspektów, który można usprawnić, jest optymalizacja hiperparametrów. W szczególności warto przeanalizować wpływ różnych wartości learning rate, testując zarówno stałe, jak i dynamiczne wartości dostosowane do etapu treningu. Alternatywnie można eksperymentować z innymi funkcjami aktywacji, takimi jak Leaky ReLU, które mogą pozwolić na lepsze uchwycenie szczegółów obrazów. Dodatkowo można sprawdzić wpływ regularizacji L1/L2 na poprawę generalizacji modelu, co mogłoby zapobiec potencjalnemu przeuczeniu.

Kolejnym istotnym kierunkiem rozwoju może być bardziej zaawansowany fine-tuning modelu. W ulepszonej wersji VGG16 odblokowano jedynie cztery ostatnie warstwy do dalszego trenowania. Możliwe jest odblokowanie większej liczby warstw, co pozwoliłoby modelowi na większą adaptację do danych. W szczególności, zamiast zamrażać całą sieć poza końcowymi warstwami, można eksperymentować z odblokowaniem dwóch ostatnich bloków konwolucyjnych, by umożliwić modelowi lepsze dopasowanie do specyfiki klasyfikowanych obrazów. Osiągnięcie lepszych wyników może być również możliwe poprzez rozszerzenie zbioru danych treningowych. Augmentacja w ulepszonym modelu VGG16 została już znacznie rozbudowana, jednak dalsze wzbogacenie zbioru mogłoby przynieść dodatkowe korzyści. Możliwe jest zastosowanie technik syntetycznego generowania obrazów, takich jak Style Transfer czy MixUp, co mogłoby pomóc w zwiększeniu różnorodności danych i poprawieniu zdolności modelu do uogólniania wzorców. Zastosowany w eksperymencie ResNet50 nie sprawdził się dobrze w klasyfikacji, uzyskując stosunkowo niską dokładność, co sugeruje, że warto rozważyć testowanie innych modeli pretrenowanych. Nowsze architektury, takie jak ResNet101, EfficientNet czy MobileNetV3, mogą oferować lepszą generalizację i efektywniejsze wykorzystanie parametrów. Alternatywnie, można również spróbować ensemblowania kilku modeli, co pozwoliłoby na połączenie ich zalet i redukcję błędów w klasyfikacji.

Kolejnym potencjalnym usprawnieniem mogłoby być dalsze eksperymentowanie z technikami augmentacji danych. Możliwe jest zastosowanie jeszcze bardziej agresywnych transformacji, takich jak CutMix, które łączą fragmenty dwóch różnych obrazów w celu poprawy zdolności modelu do rozpoznawania wzorców niezależnie od tła. Testowanie różnych poziomów intensywności augmentacji mogłoby pomóc w określeniu optymalnej konfiguracji dla danych o wysokiej zmienności wizualnej. Warto również zastanowić się nad optymalizacją modelu pod kątem zastosowań praktycznych. Chociaż ulepszony VGG16 osiągnął bardzo wysoką dokładność, jest to stosunkowo ciężki model, co może utrudniać jego wykorzystanie w aplikacjach mobilnych czy systemach działających w czasie rzeczywistym. Możliwe jest zastosowanie technik pruning oraz quantization, które pozwalają na redukcję liczby parametrów i zmniejszenie rozmiaru modelu przy jednoczesnym zachowaniu jego skuteczności. Alternatywnie można rozważyć zastosowanie lżejszych sieci CNN, takich jak MobileNetV3 lub SqueezeNet, które oferują dobrą równowagę między wydajnością a dokładnością klasyfikacji.

Podsumowując, chociaż uzyskane wyniki są obiecujące, istnieje wiele potencjalnych kierunków dalszej optymalizacji. Poprawa hiperparametrów, bardziej zaawansowany fine-tuning, rozszerzenie zbioru danych oraz eksperymentowanie z nowoczesnymi architekturami sieci mogłyby przyczynić się do jeszcze lepszej skuteczności modelu. W zależności od zastosowania, optymalizacja pod kątem wydajności obliczeniowej może być również kluczowym czynnikiem w dalszym rozwoju systemu klasyfikacji ras psów.

Źródła:

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks.*

GeeksforGeeks. (n.d.). *VGG-16 CNN model.*

Mostafid, T. (2021, March 18). *Overview of VGG16, Xception, MobileNet, and ResNet50 neural networks.* Medium.

Stack Overflow. (2021, October 7). *What is the difference between ResNet50, VGG16, etc., and RCNN, Faster RCNN?*

Roboflow. (n.d.). *6 Dog Breeds Dataset.*