

Tytuł projektu:
**Wykrywanie zapalenia płuc na obrazach RTG z wykorzystaniem
ensemble learningu**

Zespół:

Hubert Balcerzak

Jan Chyb

Justyna Hnatiuk

Miłosz Włoch

Jan Trynda

Bartłomiej Plewnia

Cel projektu:

Projekt ma na celu klasyfikację obrazów RTG ze zbioru za pomocą techniki ensemble learningu z wykorzystaniem modeli sieci neuronowych, <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia> na obrazy płuc dotkniętych zapaleniem płuc i płuc zdrowych.

Opis datasetu:

Dataset składa się z 5863 obrazów RTG w formacie .jpg, podzielonych na katalogi: train, test, val, zawierających po dwie kategorie (NORMAL, PNEUMONIA). Obrazy RTG zostały wykonane w procesach diagnostycznych pacjentów w wieku 1-5 lat, usunięto nieczytelne lub niskiej jakości skany, a następnie zostały one zaklasyfikowane przez dwóch lekarzy ekspertów.

Dataset jest niezbalansowany (jest około 3 razy więcej obrazów płuc z zapaleniem niż płuc zdrowych), co stwarza dodatkowe wyzwanie.

Plan eksperymentów:

1. Augmentacja danych.
2. Wyuczenie sieci CNN (5 bloków sieci składających się z warstw: konwolucyjnej, max-pooling, batch-normalization + warstwa spłaszczająca). Funkcją aktywacji będzie Relu, w ostatniej warstwie będzie to funkcja sigmoid (ze względu na binarność klasyfikacji), będzie to nasza sieć referencyjna.
3. W zależności od otrzymanych wyników:
 - **Stworzenie nowej sieci i porównanie wyników**

- **Połączenie stworzonej sieci z istniejącymi pretrenowanymi sieciami (np. GoogLeNet, ResNet), i obserwacja wyników.** Wagi przypisane modelom składowym będą wyliczane na podstawie precision, recall, f1-score i AUC.
- Zmiana datasetu i porównanie wyników działania modelu/modeli na innym podobnym zestawie danych (np. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>)

Wykonanie:

1. Augmentacja danych:

Dokonano augmentacji danych za pomocą ImageDataGenerator dostępnego w bibliotece Keras.

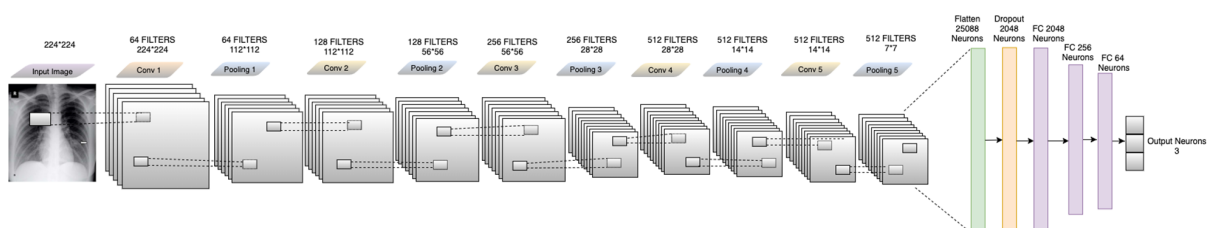
```
train_datagen = ImageDataGenerator(rescale=1./255,
    zoom_range=0.3,rotation_range=10,width_shift_range=0.2,
    height_shift_range=0.2,brightness_range=[0.4,1.5])
```

Dane w zbiorze testowym zostały m.in. przeskalowane, zrotowane o max. 10 stopni, przesunięte oraz zmienione zostały ich jasności, żeby otrzymać pożądany rozmiar danych treningowych.

Porównaliśmy wyniki uczenia za pomocą zaugmentowanych danych i bez augmentacji - w przypadku braku augmentacji trenowany model wykazywał większy overfitting (otrzymywaliśmy znacznie gorsze wyniki na zbiorze testowym niż w modelu trenowanym na augmentowanych danych). Dlatego pozostaliśmy przy augmentacji danych.

2. Wyczenie sieci CNN:

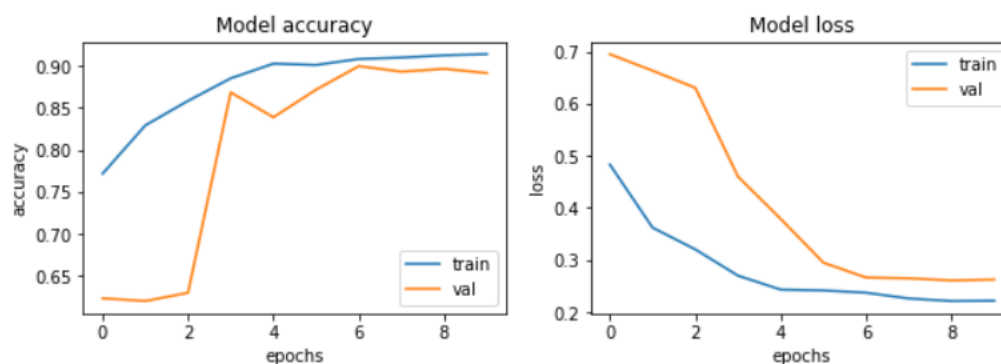
Architektura modelu została zaczerpnięta z artykułu "Automated detection of COVID-19 through convolutional neural network using chest x-ray images"



Rys. 1. Architektura sieci CNN użytej w pierwszym etapie projektu [2]

Zbudowano model CNN, składający się z 5-ciu bloków konwolucyjnych (z funkcją ReLU i max-poolingiem 2x2). Do dwóch ostatnich bloków dodano Dropout z prawdopodobieństwem usunięcia neuronu równym 0.2. Na końcu dodano warstwę spłaszczającą.

Uzyskano następujące wyniki:



CONFUSION MATRIX

[[213 21]

[45 345]]

OTHER METRICS

Accuracy: 89.42307692307693%

Precision: 94.26229508196722%

Recall: 88.46153846153845%

F1-score: 91.26984126984127

TRAIN METRIC -----

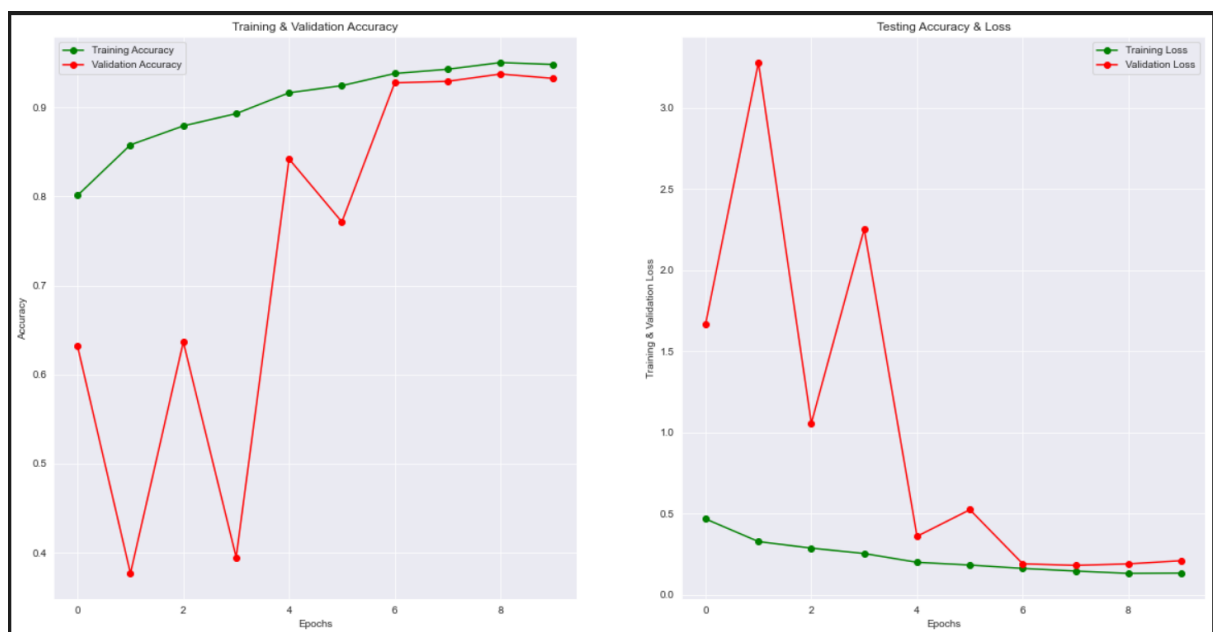
Train acc: 91.39

W przypadku diagnozy chorób, ważnym parametrem jest Recall, który przedstawia stosunek poprawnych pozytywnych predykcji do prawdziwych pozytywnych przypadków. Początkowa sieć CNN ma Recall na poziomie 88,46%.

3. Stworzenie nowej sieci i porównanie wyników

Połączenie stworzonej sieci z istniejącymi pretrenowanymi sieciami (np. GoogLeNet, ResNet), i obserwacja wyników.

Zbudowano sieć ResNet-18, dokonując pomiarów dokładności na zbiorze treningowym i walidacyjnym.

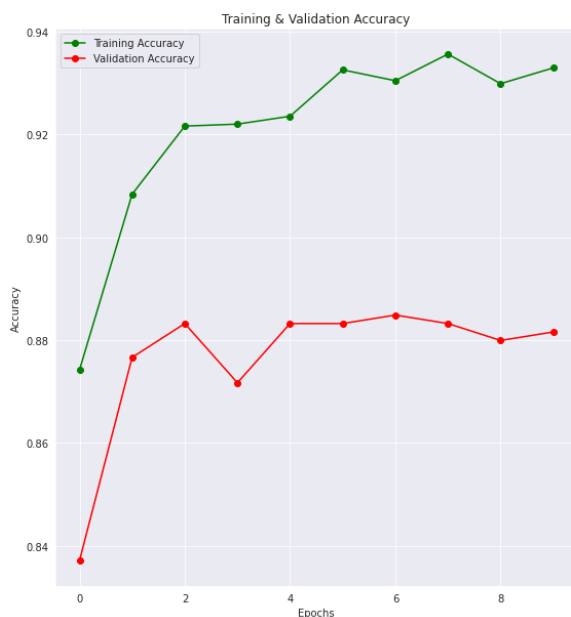


W tym przypadku wyniki są porównywalne z siecią referencyjną.

W celu polepszenia wyników postanowiliśmy wykorzystywać sieć ResNet w połączeniu z DenseNet i siecią referencyjną.

W następnym etapie zaimportowaliśmy z modułu keras model DenseNet201 (pretrenowany na zbiorze ImageNet) i przetrenowaliśmy go na danych treningowych z badanego datasetu.

Uzyskaliśmy następujące wyniki:



Następnie zastosowaliśmy technikę Stacked Ensemble Learning, w której budowany jest meta-learner, który przyjmuje predykcje dokonane przez modele poziomu 0 (CNN, ResNet18, DenseNet201) i uczy się tworzyć jak najlepsze predykcje. Jako meta-learner został wykorzystany model regresji logistycznej.

Dzięki wykorzystaniu Stacked Ensemble Learningu udało się polepszyć ostateczną dokładność modelu złożonego z trzech wspomnianych sieci.

Model	Accuracy
CNN 5-cio blokowy	89,3%
ResNet-18	92,9%
DenseNet201	89,33%
Stacked Ensemble Learning na trzech powyższych sieciach	94,1%

Zastosowanie Stacked Ensemble Learningu pozwoliło polepszyć dokładność modeli o 1,2 punktów procentowych w stosunku do najlepszego modelu będącego częścią ostatecznego modelu. W porównaniu z początkową siecią udało nam się również znacząco poprawić Recall (o 8 punktów procentowych), i dokładność o 5.

Szczegółowe statystyki:

Statystyki	Sieć Stacked Ensemble	Sieć referencyjna - CNN
Dokładność [%]	94.0705	89.42307
Precyzja [%]	94.2355	94.262
Recall [%]	96.41	88.4615
F1-score [%]	95.31	91.2698
ROC-AUC	0.9329	

Wnioski i podsumowanie:

Co prawda nie udało się nam pobić wyniku z referencyjnego artykułu, ale projekt miał stanowić próbę polepszenia wyników klasyfikacji obrazów RTG. Na niższe wyniki sieci mogą mieć wpływ: użycie zbyt płytkiej w porównaniu do reszty sieci konwolucyjnej (składała się z 5 bloków). Innym powodem może być niewielka ilość epok uczenia (10) lub zastosowanie Stacked Ensemble Learningu zamiast metody średnich ważonych. Ponadto, autorzy referencyjnego artykułu[1] oceniali modele na podstawie większej ilości parametrów (np. ROC), co potem uwzględniali w wagach.

W celu rozwinięcia projektu należałoby przede wszystkim powtórzyć eksperymenty na modelach z wykorzystaniem innych metod ensemble learningu. Możliwe, że zastosowanie dwóch sieci ResNet o zróżnicowanej głębokości (np. ResNet-18 i ResNet-54) polepszyłyby wyniki, ze względu na to, że model oparty o architekturę ResNet najczęściej uzyskiwał najwyższe wyniki w klasyfikacji. Należałoby również rozszerzyć testy o testy na innych zbiorach danych z obrazami RTG.

Literatura:

- 1) **Pneumonia detection in chest X-ray images using an ensemble of deep learning models.** Kundu, Rohit & Das, Ritacheta & Geem, Zong Woo & Han, Gi-Tae & Sarkar, Ram (2021)
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0256630>
- 2) **Automated Detection of COVID-19 through Convolutional Neural Network using Chest x-ray images**
Rubina Sarki, Khandakar Ahmed, Hua Wang, Yanchun Zhang, Kate Wang
medRxiv 2021.02.06.21251271; doi: <https://doi.org/10.1101/2021.02.06.21251271>