Identifying pneumonia in chest X-ray images

Maciej Lorens
419763

Introduction to the problem

Annually, pneumonia takes the lives of around 2 million children under the age of 5, persistently standing as the foremost cause of childhood mortality according to the World Health Organization (WHO). WHO highlights that nearly all cases (95%) of new-onset childhood clinical pneumonia surface in developing regions, primarily Southeast Asia and Africa, where swift interpretation of radiographic data is often lacking. This project aims to address this challenge by developing an ensemble deep learning model capable of accurately identifying pneumonia in chest X-ray images.

Dataset description

Source:
https://data.mendeley.com/datasets/rscbjbr9sj/2

The dataset comprises 5,856 JPEG chest X-ray images from children, each accompanied by a binary label (1 for Pneumonia, 0 for Normal). Within the training set, 3,883 images portray pneumonia (2,538 bacterial and 1,345 viral cases), while 1,349 images represent chest X-rays without anomalies. Moving to the test set, it consists of 234 normal images and 390 chest X-ray images indicating pneumonia.
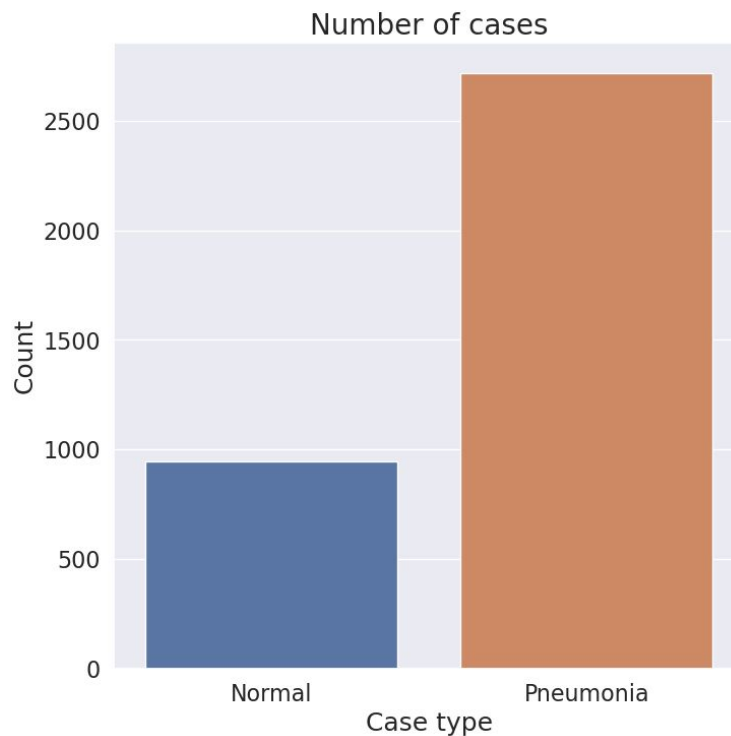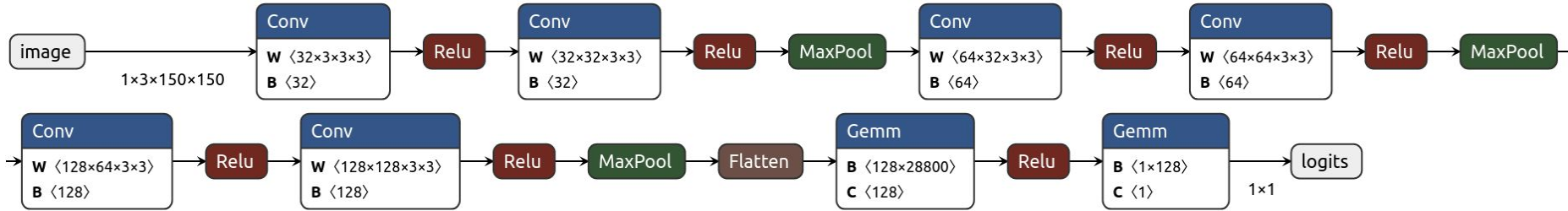
Examples of images

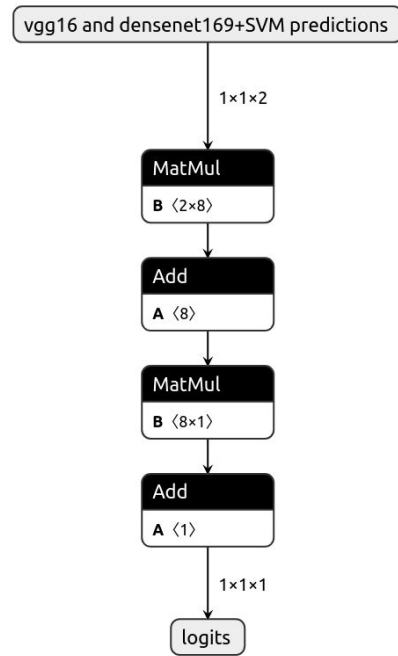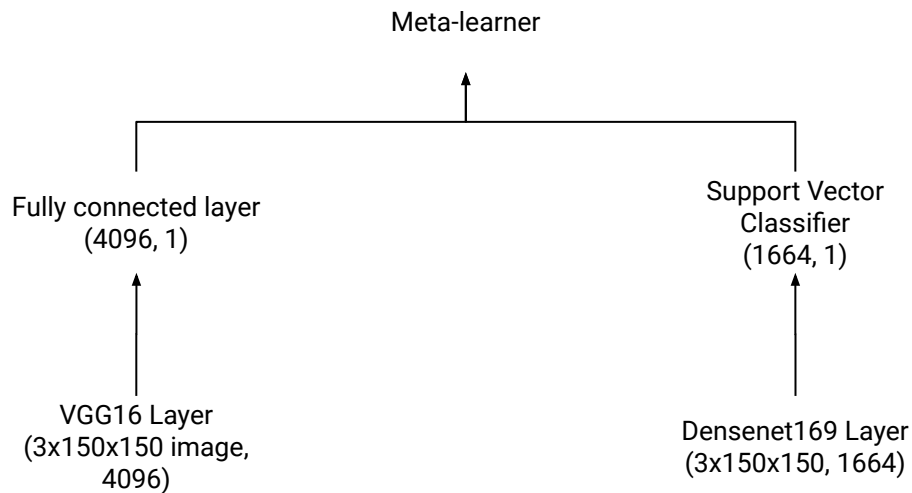# Selected models and chosen hyperparameters for optimization

- Benchmark model:
  a. Simple Convolutional Neural Network (CNN)
- Ensemble of two Transfer Learning Models (outputs of the models are combined into the final prediction via Single Layer Perceptron):
  a. Frozen VGG16 model with the final layer replaced by a trainable Fully Connected Layer
  b. Frozen DenseNet169 model with the final layer replaced by a trainable Support Vector Classification (SVC) model

The ensemble model's architecture was inspired by the paper "A novel ensemble CNN model for COVID-19 classification in computerized tomography scans".

Model architectures - benchmark model

Model architectures - stacking model

Meta-learner

Fully connected layer
(4096, 1)

Support Vector
Classifier
(1664, 1)

VGG16 Layer
(3x150x150 image,
4096)

Densenet169 Layer
(3x150x150, 1664)

vgg16 and densenet169+SVM predictions

1×1×2

MatMul
B ⟨2×8⟩

Add
A ⟨8⟩

MatMul
B ⟨8×1⟩

Add
A ⟨1⟩

1×1×1

logits

Performance metrics

In the context of identifying pneumonia in developing regions, it is crucial to account for limited resources, emphasizing the need to balance Precision and Recall.

- **Area Under the Receiver Operating Characteristic (ROC-AUC)** - Evaluates the model's ability to distinguish between positive and negative classes across different threshold values, providing a comprehensive measure of its discriminatory power. It was also used in the original paper that introduced the dataset.
- **Balanced Accuracy** - It takes into account the imbalance between the classes and provides a more balanced view of the model's performance.
- **F1 Score** - Harmonic mean of precision and recall, a single metric that considers both false positives and false negatives.

VGG16 best hyperparameters

Search space:

- Batch size = [64, 128, 256]
- Dropout = [0.2, 0.4]
- Learning rate = [0.01, 0.001]

Best hyperparameters (after 10 epochs with median pruning)

- **Batch size = 64**
- **Dropout = 0.2**
- **Learning rate = 0.001**
- **Best threshold on validation = 0.5**

Densenet169 + SVC best hyperparameters

Search space:

- C = (0.01, 0.1)
- Kernel = ["linear", "poly", "rbf"]
- Class weight = ["balanced", None]

Best hyperparameters

- **C = 0.01**
- **Kernel = "linear"**
- **Class weight = "balanced"**
- **Best threshold on validation = 0.86**

Stacking model best hyperparameters

Search space:

- Batch size = [64, 128, 256]
- Dropout = [0.2, 0.4]
- Learning rate = [0.001, 0.01]
- Number of neurons = [2, 4, 6, 8]
- Activation function = [Identity, Sigmoid, ReLU]

Best hyperparameters (after 10 epochs with median pruning)

- **Batch size = 64**
- **Dropout = 0.2**
- **Learning rate = 0.01**
- **Number of neurons = 8**
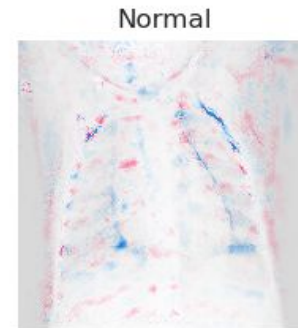- **Activation function = Identity**
- **Best threshold on validation = 0.63**

Comparison of results on test set

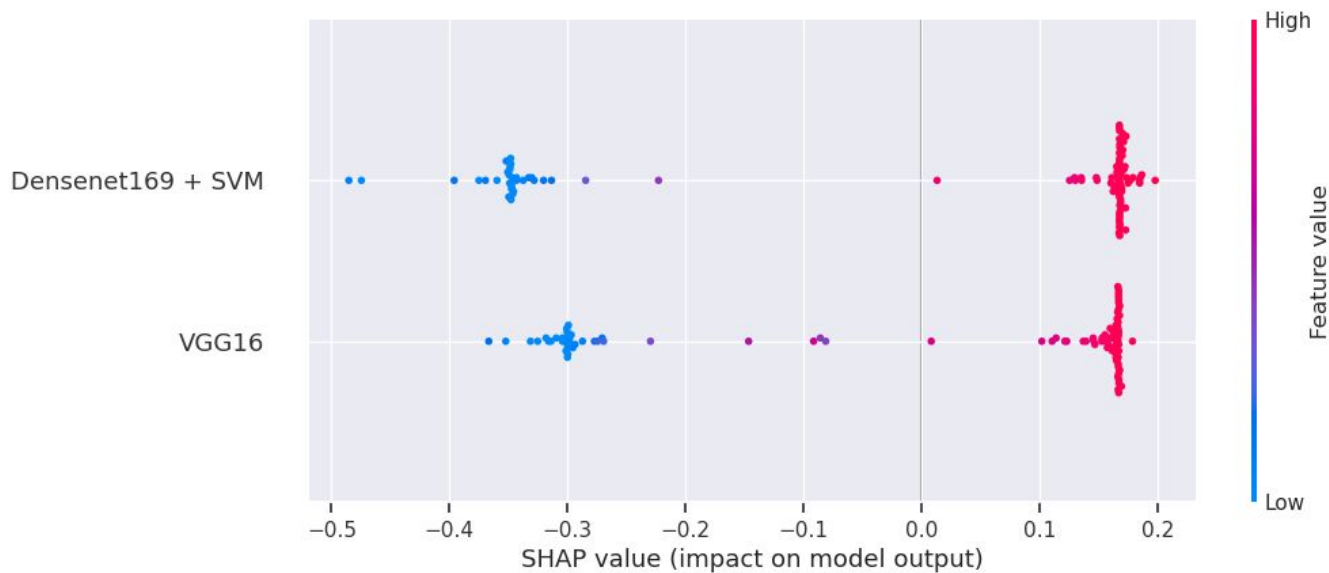|	Benchmark	VGG16	Densenet169+SVC	Stacking model
AUC	0.8515	0.9215	0.8746	0.9402
Balanced accuracy	0.7936	0.7983	0.7543	0.7962
F1-score	0.8362	0.8826	0.8611	0.8816

Stacking model confusion matrix on test set
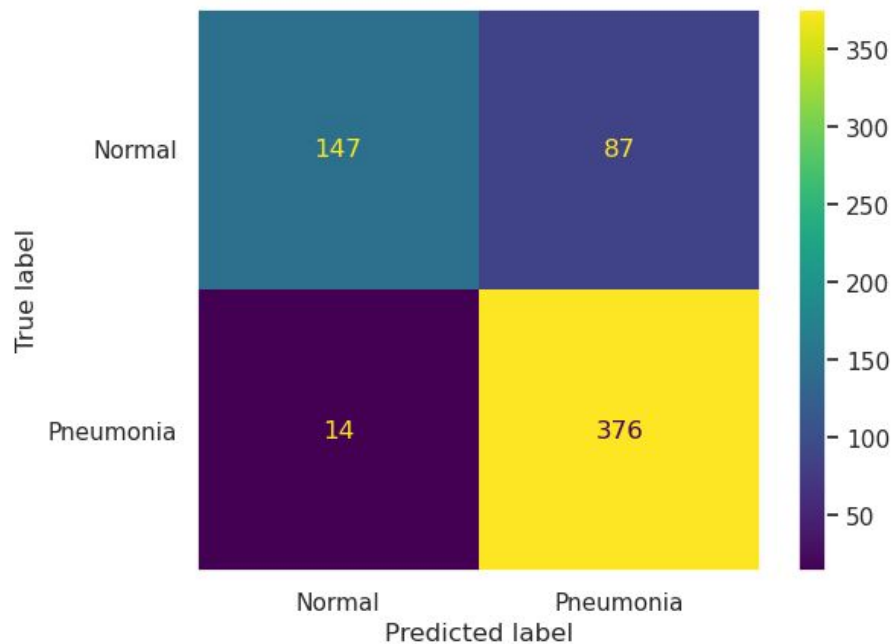
Shapley values of VGG16 model

Shapley values of stacking model

References

- Kermany D, Goldbaum M, Cai W et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018; 172(5):1122-1131. doi:10.1016/j.cell.2018.02.010.
- Cortes O, Diniz J, Silva LDJ (2023) "A novel ensemble CNN model for COVID-19 classification in computerized tomography scans"
- Biloglav Z, Boschi-Pinto C, Campbell H, Mulholland K, Rudan I (2008) "Epidemiology and etiology of childhood pneumonia"