Pneumonia Detection from Chest X-Ray Images

Hemanth Kumar 19BM6JP27 • Karthikeya Racharla 19BM6JP32 • Subramania Bharathi 19BM6JP36

Presented at Final End Term Project Evaluation

for

BM69006 Data Science Laboratory 2020
Indian Institute of Technology Kharagpur
June 14, 2020

INTRODUCTION

- According to the World Health Organization (WHO), **pneumonia kills** about **2 million children** under 5 years old every year and is consistently estimated as the single leading cause of childhood mortality (Rudan et al., 2008), killing more children than HIV/AIDS, Malaria, and Measles combined.

- **Chest X-rays - C**urrently the best available method for diagnosing pneumonia as stated by WHO (2001). However, detecting pneumonia is a challenging task that relies on the availability of expert radiologists.

MOTIVATION

- **Bacterial Pneumonia** requires urgent referral for immediate antibiotic treatment, while **Viral Pneumonia** is treated with supportive care
- Therefore, accurate and timely diagnosis is imperative
- However, rapid radiologic interpretation of images is not always available, particularly in the low-resource settings where childhood pneumonia has the highest incidence and highest rates of mortality

PROBLEM STATEMENT

Build an AI model which can automatically detect Pneumonia from Chest X-ray images

Sub-problems:

1) Pneumonia vs. Normal Cases Classification with Chest X-ray

2) Bacterial vs. Viral Cases Classification with Chest X-ray

IMPORTANT METRICS

- Immediate treatment for someone with pneumonia is utmost important, so everyone with pneumonia has to be found out correctly with the model, i.e., in evaluation terms of Machine Learning:

  Recall has to be high for Pneumonia vs Normal cases

- Once some one is diagnosed with Pneumonia, we have to identify whether it is a viral pneumonia case or bacterial pneumonia, as the patient has to be treated accordingly because mistreatment as a viral case could be fatal. Moreover, the treatments for both are different.

  In this case, Precision is very important

SIGNIFICANCE OF THE PROJECT

- Potential for generalized high-impact application in biomedical imaging

- To quote an example, in Africa’s 57 nations, a gap of 2.3 million doctors and nurses exists. For these populations, accurate and fast diagnosis means everything. It can guarantee timely access to treatment and save much needed time and money for those already experiencing poverty

- The techniques used in the projects could be extended for detecting other lung diseases where X-ray based detection techniques are currently used

- Ideal example for the same is the current ongoing research and diagnosis on detecting COVID-19 with Chest X-ray images

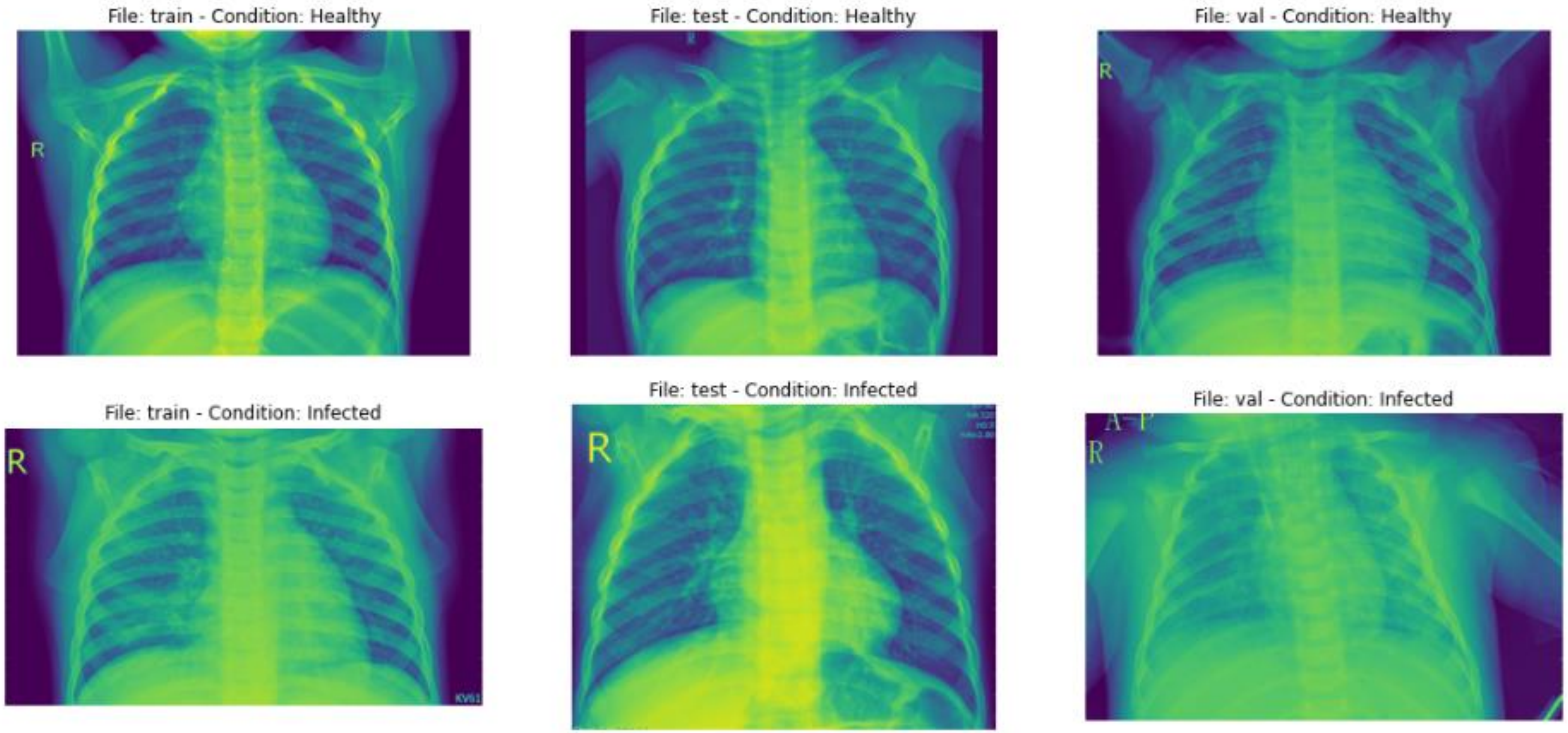
ABOUT THE DATASET

- There are a total of **5,856** X-Ray Grey-Scale images (JPEG) and 2 categories, namely Pneumonial vs. Normal images.
- These **pediatric chest X-rays** of pneumonial patients is further distinguished into viral and bacterial pneumonia to facilitate rapid referrals for children needing urgent intervention.

Class Imbalance

SET	Total Images	Normal	Pneumonia	Bacterial	Viral
Training	5,216	1,341	3,875	2,530	1,345
Validation	16	8	8	8	0
Testing	624	234	390	242	148

PNEUMONIA VS NORMAL X-RAY IMAGES



DATA AUGMENTATION

- Technique to artificially create new training data from existing training data

- *Implementation:* ImageDataGenerator class from Keras

- Transformations used include:
  - **Zoom**
  - **Horizontal shift, Vertical shift**
  - **Horizontal flip, Rotation and shear**

PREVIOUS WORK AND RESULTS

- Kermany et. al (2018) have developed an Image-based deep learning classifies macular degeneration and diabetic retinopathy using retinal optical coherence tomography images, that is generalized to have potential for applications in biomedical image interpretation and medical decision making

[Pneumonia vs. Normal] They achieved an accuracy 92.8%, with sensitivity 93.2%, specificity 90.1% and Precision 98.45%

[Bacterial vs. Viral Pneumonia] resulted in test accuracy 90.7%, with sensitivity 88.6%, specificity 90.9% and Precision 48.112%

Reference: Kermany et. al (Cell Journal 2018)

OUR CONTRIBUTION / MODEL USED

- Training Convolution Neural Network from scratch
- Transfer learning model
  - **InceptionV3**
  - **DenseNet**
  - **ResNet50**
  - **VGG16**

Note: Results of these models will be compared with results obtained in prior benchmark studies

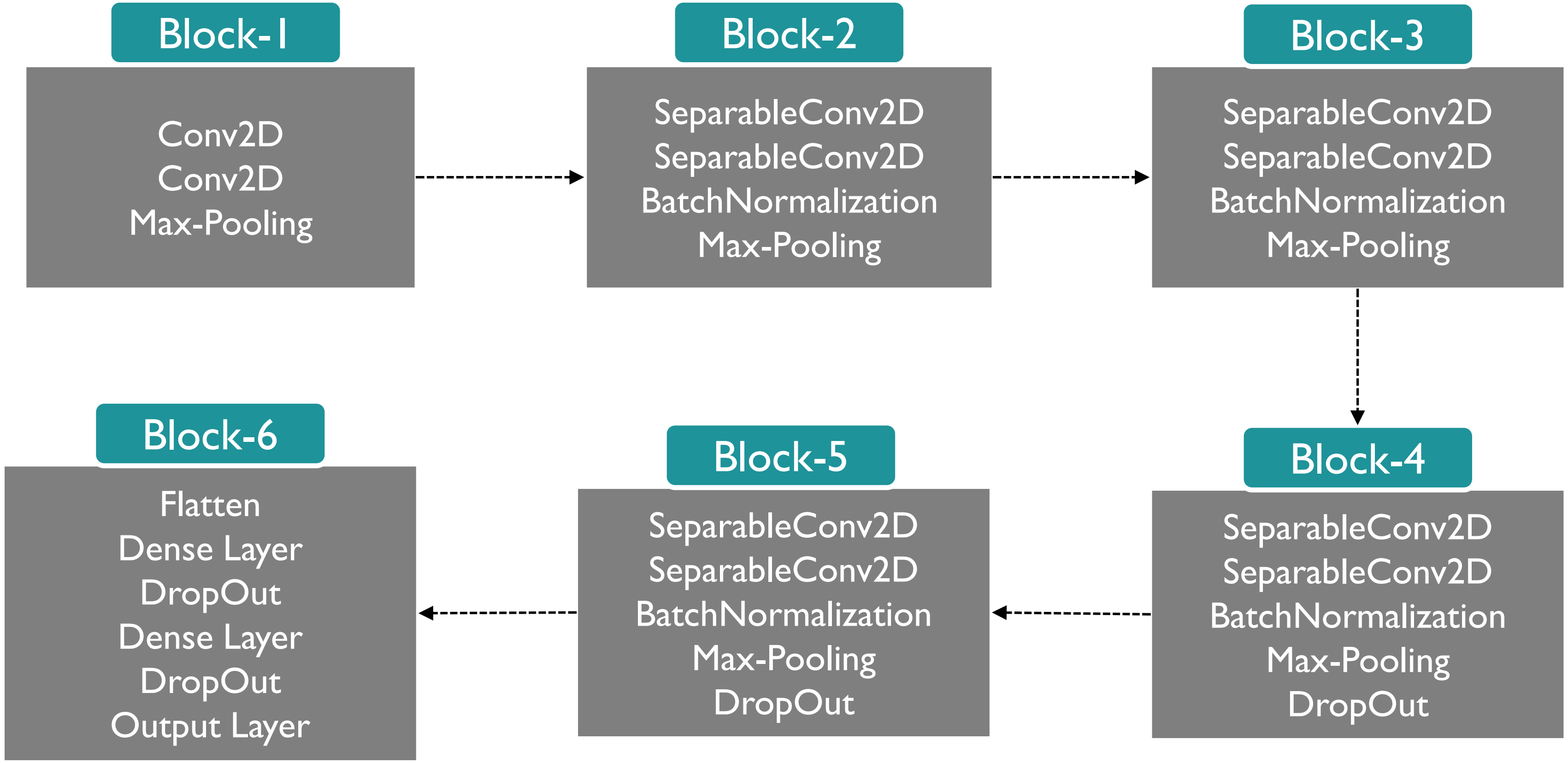
CNN ARCHITECTURE

INPUT → FEATURE EXTRACTION → CLASSIFICATION OUTPUT

Feature extraction - The feature extraction component of a convolutional neural network is what distinguishes CNNs from other multilayered neural networks. It typically comprises of repeating sets of these sequential steps

CONVOLUTION LAYER → POOLING LAYER → NON-LINEAR ACTIVATION UNIT

CNN ARCHITECTURE: DEBRIEF



CONVOLUTION LAYERS

Action

- Apply filters to extract features
- Filters are composed of small kernels, learned
- One bias per filter
- Apply activation function on every value of feature map

Parameters

- Number of kernels
- Size of kernels
- Activation function
- Stride
- Padding
- Regularization type and value

POOLING LAYERS

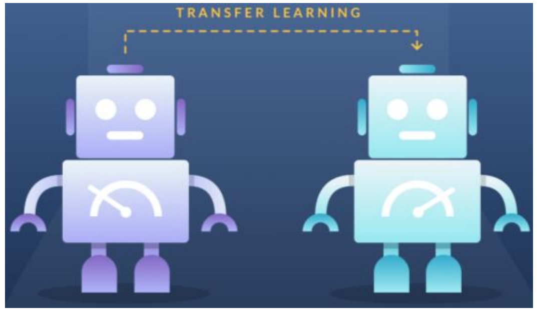
Action

- Reduce dimensionality
- Extra maximum of average of a region
- Sliding window approach

Parameters

- Stride
- Size of window

TRANSFER LEARNING



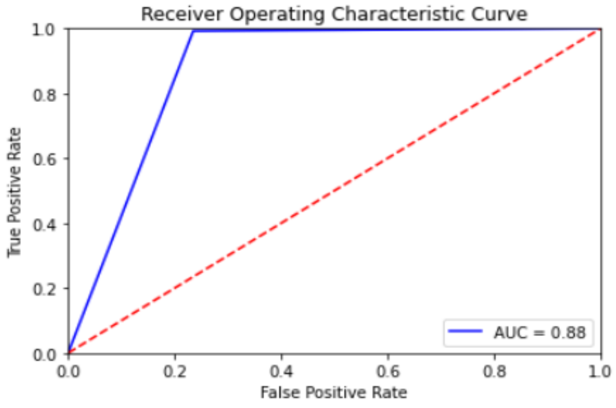
- *Intuition:* What has been learned in one setting is exploited to improve generalization in another setting
- **Why Transfer Learning?**
  - CNNs are data-hungry: require significant amounts of data and Resources to train
  - ImageNet ILSVRC Model was trained on 1.2 million images over the period of 2–3 weeks across multiple GPUs
  - We can make use of the features learned in those pre trained model
- **Advantages of transfer learning**
  - Reduces the training time and data needed to achieve a custom task
  - Very popular in Computer Vision; often produces robust models with improvement in Metrics
- **How to use it?**

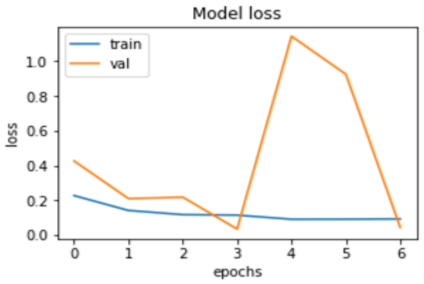
It takes a CNN that has been pre-trained (typically ImageNet), removes the last fully-connected layer and replaces it with our custom fully-connected layer, treating the original CNN as a feature extractor for the new dataset. Once replaced, the last fully-connected layer we train the classifier for the new dataset

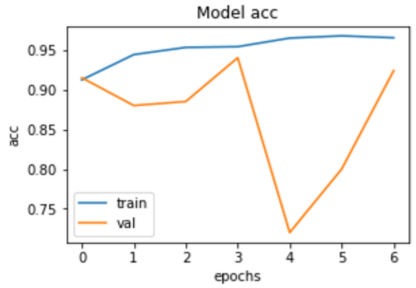
MODEL RESULTS

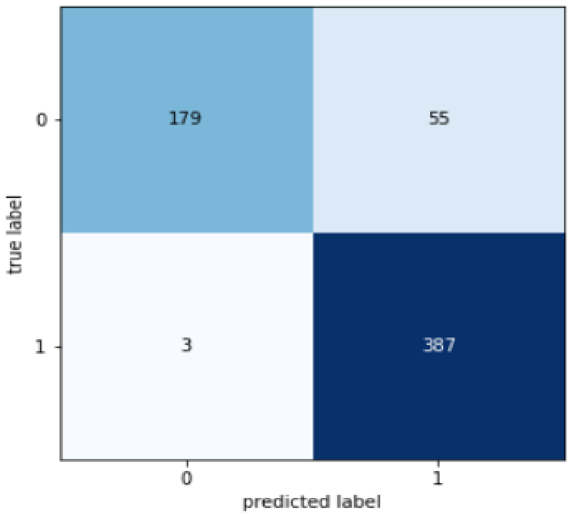
NORMAL VS PNEUMONIA RESULTS

VGG16 MODEL









TEST METRICS

Accuracy: 90.71%

Precision: 87.55%

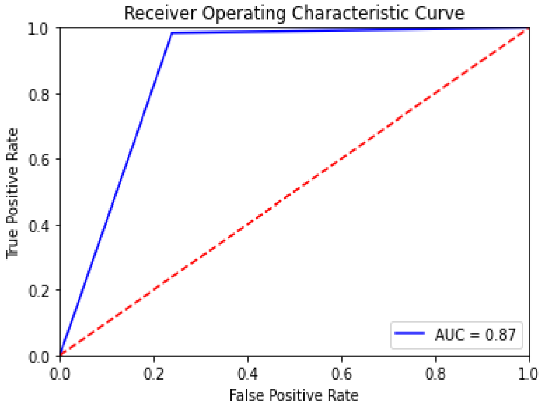
Recall: 99.23%

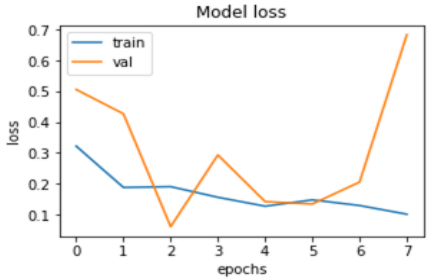
F1-score: 93.02%

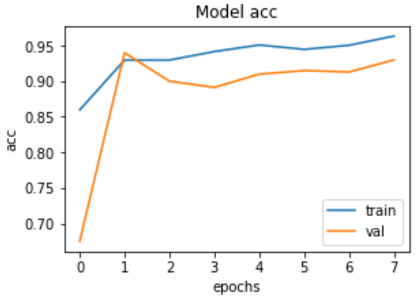
BENCHMARK RECALL VALUE : 93.2%

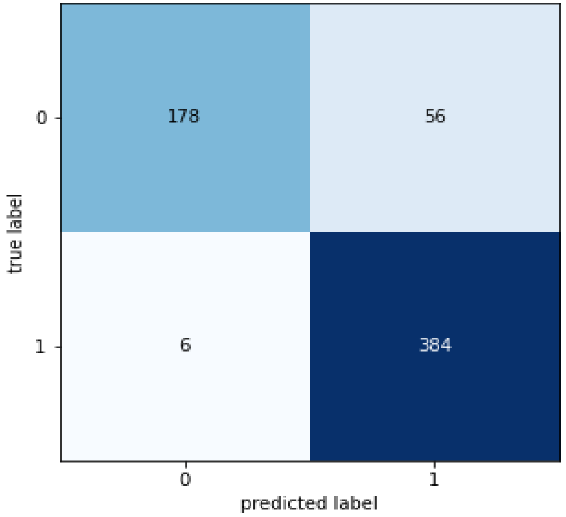
ACHIEVED BETTER THAN BENCHMARK

INCEPTION-V3 MODEL









TEST METRICS

Accuracy: 90.06%

Precision: 87.27%

Recall: 98.46%

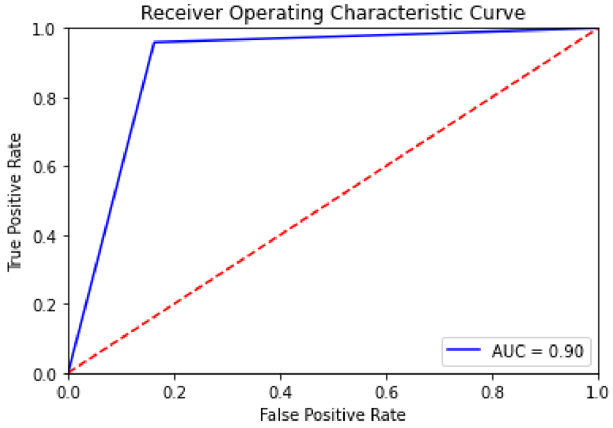
F1-score: 92.53%

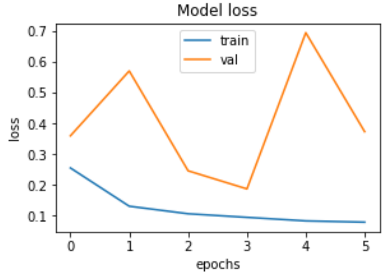
BENCHMARK RECALL VALUE : 93.2%

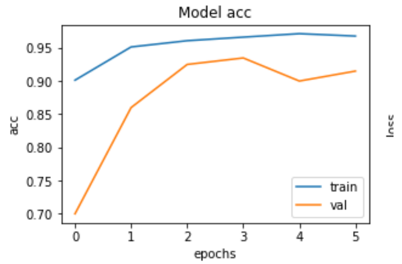
ACHIEVED BETTER THAN BENCHMARK

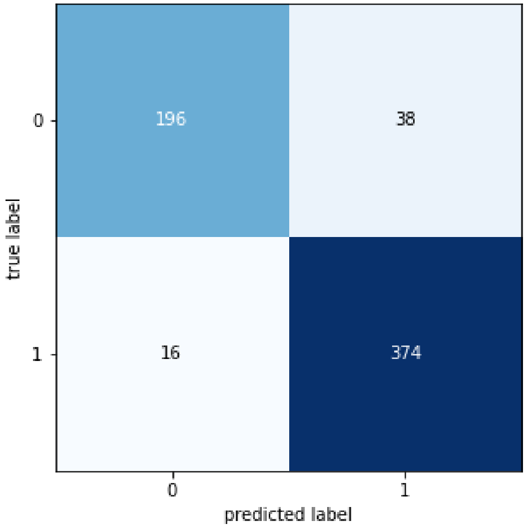
NORMAL VS PNEUMONIA RESULTS

DENSENET MODEL









Test metrics

Accuracy: 91.34%

Precision: 90.77%

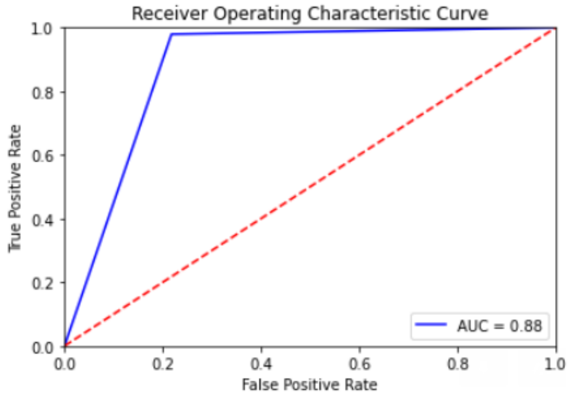
Recall: 95.89%

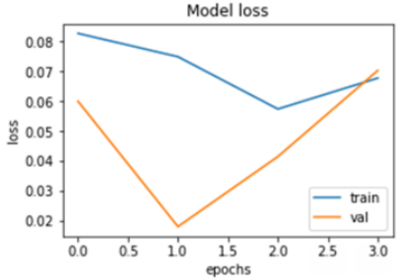
F1-score: 93.26%

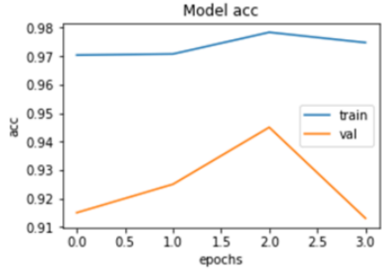
BENCHMARK RECALL VALUE : 93.2%

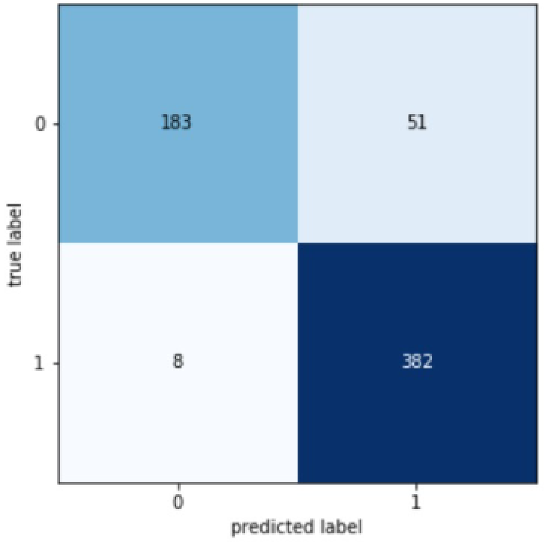
ACHIEVED BETTER THAN BENCHMARK

RESNET50 MODEL









Test metrics:

Accuracy: 92.95%

Precision: 92.95%

Recall: 95.64%

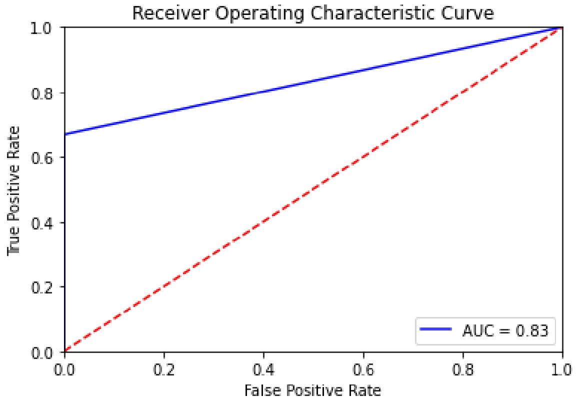
F1-score: 94.43%

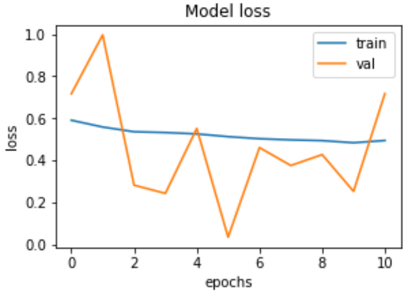
BENCHMARK RECALL VALUE : 93.2%

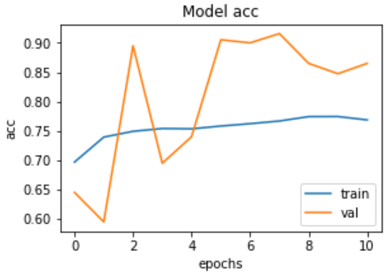
ACHIEVED BETTER THAN BENCHMARK

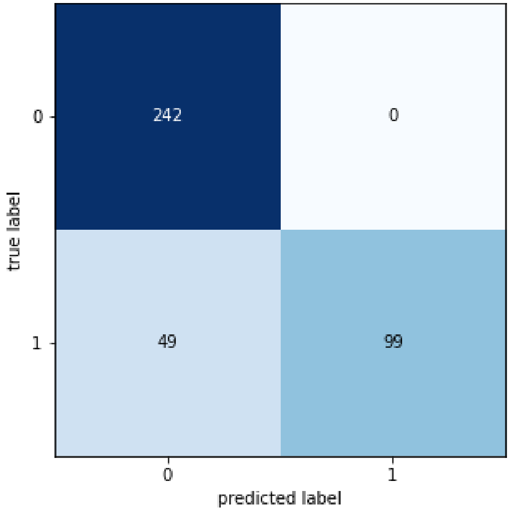
BACTERIAL VS. VIRAL CLASSIFICATION RESULTS

VGG16 MODEL









Test metrics

Accuracy: 87.43%

Precision: 100.0%

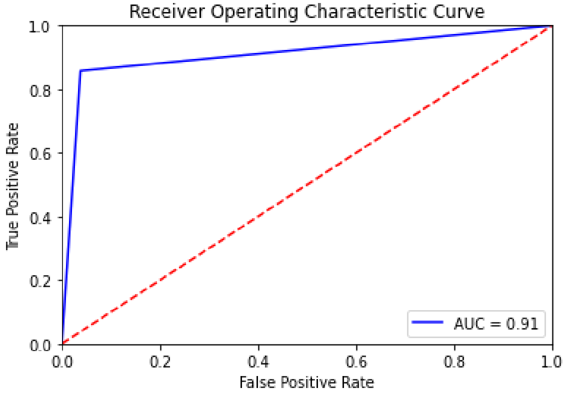
Recall: 66.89%

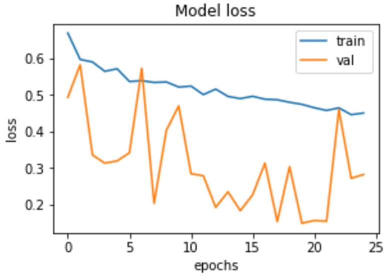
F1-score: 80.16%

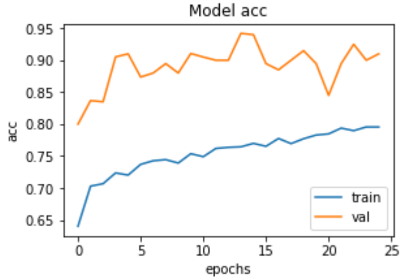
BENCHMARK PRECISION,F1: 48.11%, 62.36%

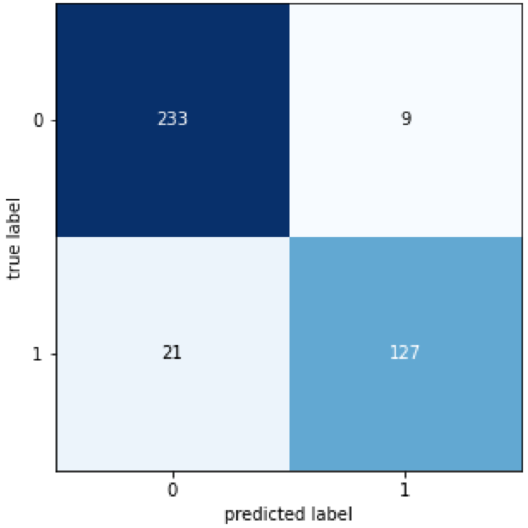
ACHIEVED BETTER THAN BENCHMARK

INCEPTION-V3 MODEL









TEST METRICS

Accuracy: 90.06%

Precision: 87.27%

Recall: 98.46%

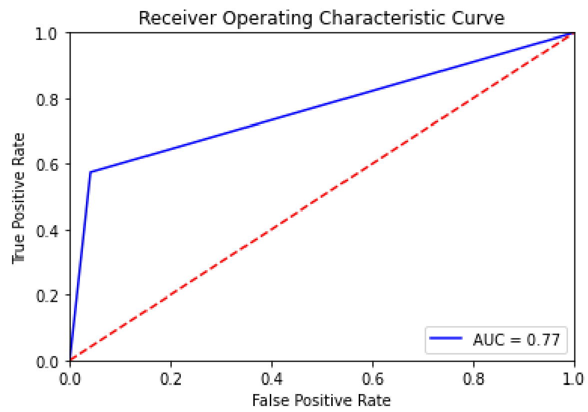
F1-score: 92.53%

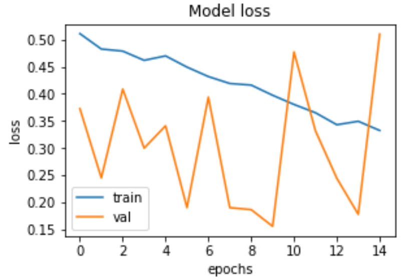
BENCHMARK PRECISION,F1: 48.11%, 62.36%

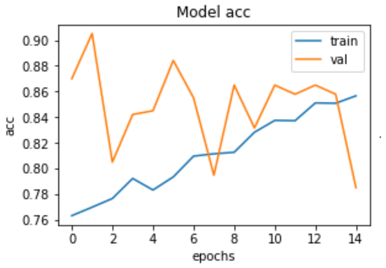
ACHIEVED BETTER THAN BENCHMARK

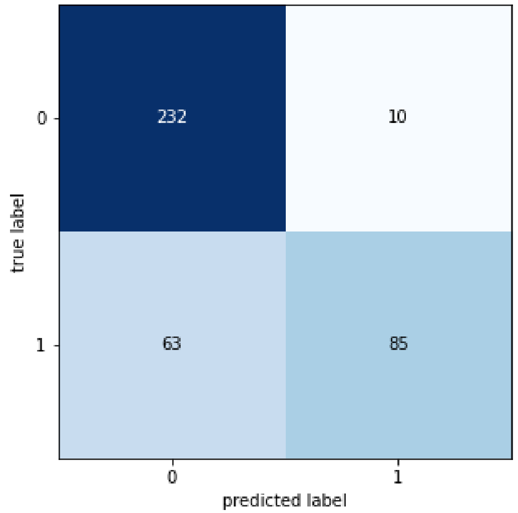
BACTERIAL VS. VIRAL CLASSIFICATION RESULTS

DENSENET MODEL









Test metrics:

Accuracy: 81.28%

Precision: 89.47%

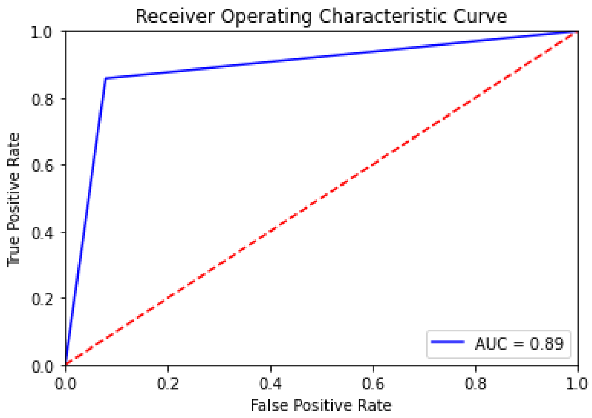
Recall: 57.43%

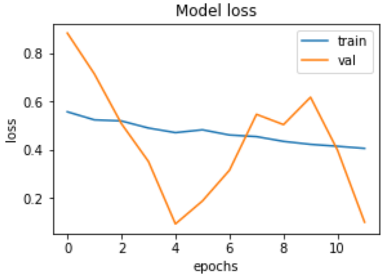
F1-score: 69.96%

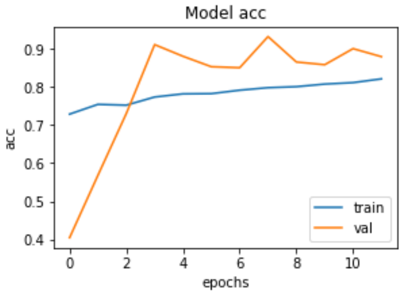
BENCHMARK PRECISION,F1: 48.11%, 62.36%

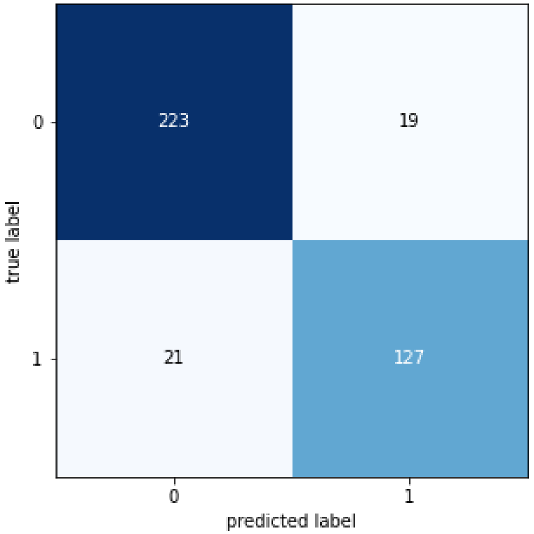
ACHIEVED BETTER THAN BENCHMARK

RESNET50 MODEL









Test metrics:

Accuracy: 89.74%

Precision: 86.98%

Recall: 85.81%

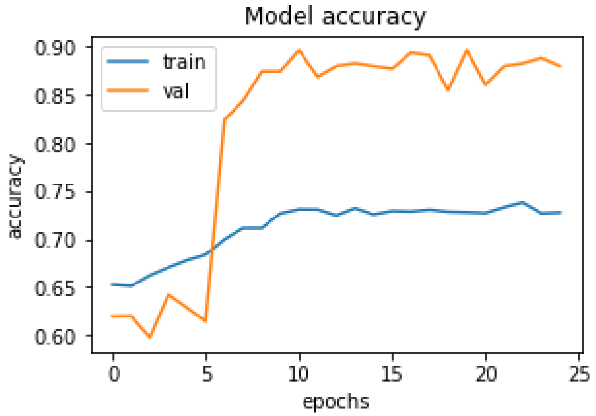
F1-score: 86.39%

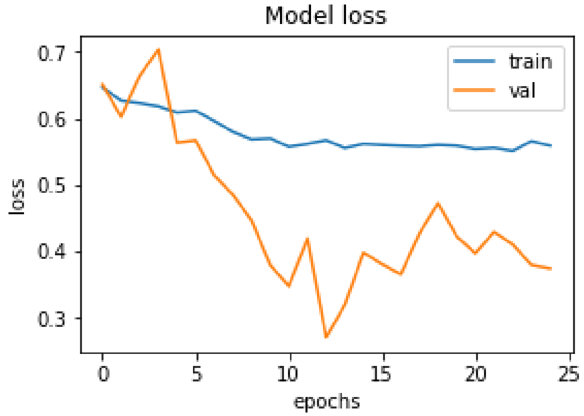
BENCHMARK PRECISION,F1: 48.11%, 62.36%

ACHIEVED BETTER THAN BENCHMARK

CNN RESULTS

BACTERIAL VS. VIRAL CLASSIFICATION RESULTS





Test metrics:

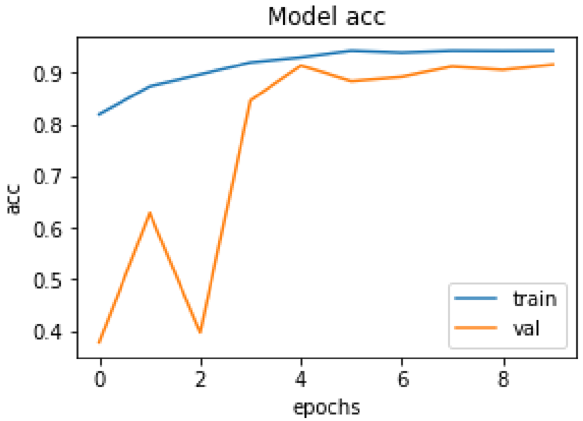
Accuracy: 88.21%

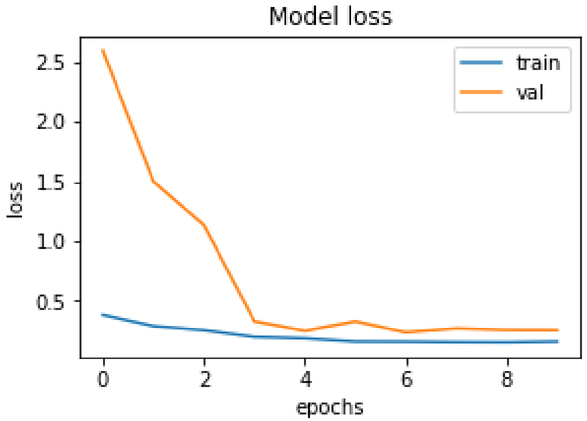
Precision: 94.74%

Recall: 72.97%

F1-score: 82.4%

NORMAL VS PNEUMONIA RESULTS





Test metrics:

Accuracy: 91.03%

Precision: 89.76%

Recall: 96.66%

F1-score: 93.08%

CONSOLIDATED RESULTS

NORMAL VS PNEUMONIA RESULTS

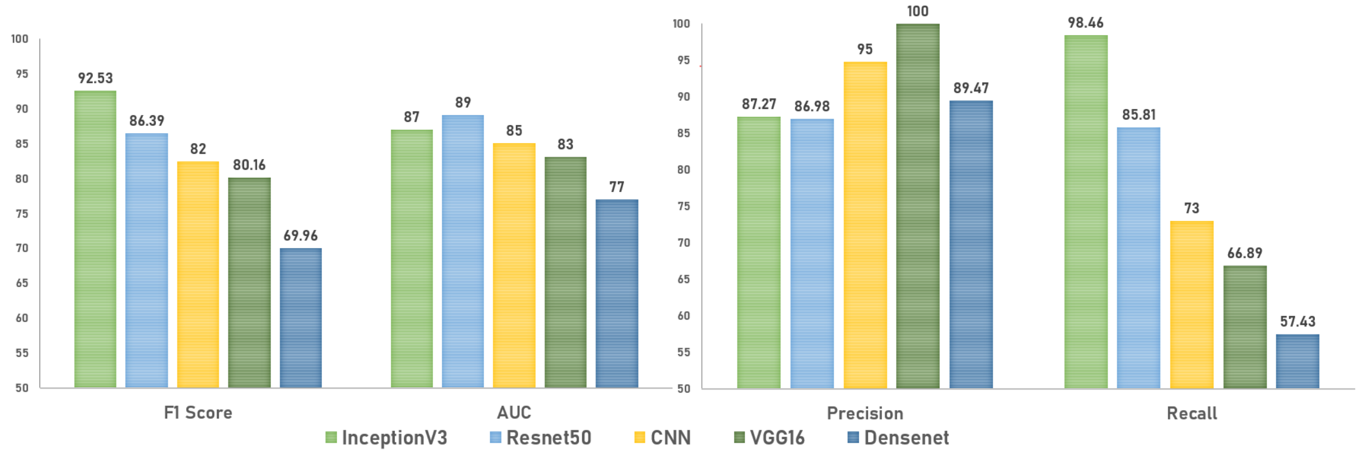
Models	Precision	Recall	Accuracy	F1 Score	AUC
VGG16	87.55	**99.23**	90.71	93.02	88
InceptionV3	87.27	98.46	90.06	92.53	87
DenseNet	90.77	95.89	91.34	93.26	90
ResNet50	**92.95**	95.64	**92.95**	**94.43**	**92**
CNN	89.76	96.66	91.03	93.08	91

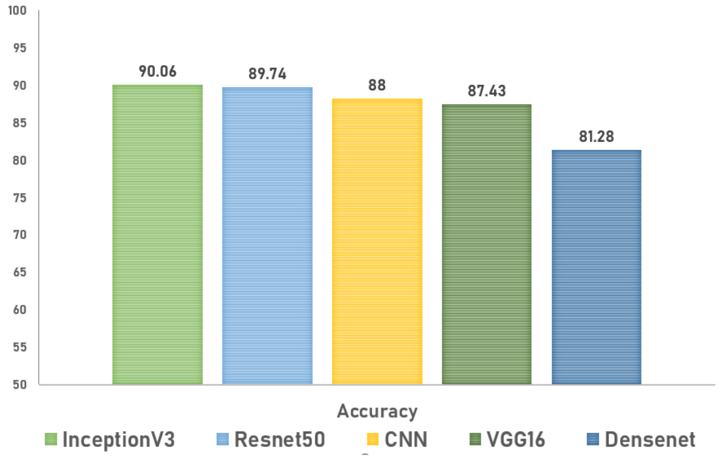
BACTERIAL VS. VIRAL CLASSIFICATION RESULTS

Models	Precision	Recall	Accuracy	F1 Score	AUC
VGG16	**100**	66.89	87.43	80.16	83
InceptionV3	87.27	**98.46**	**90.06**	**92.53**	87
DenseNet	89.47	57.43	81.282	69.958	77
ResNet50	86.98	85.81	89.74	86.39	**89**
CNN	94.736	72.97	88.205	82.44	85

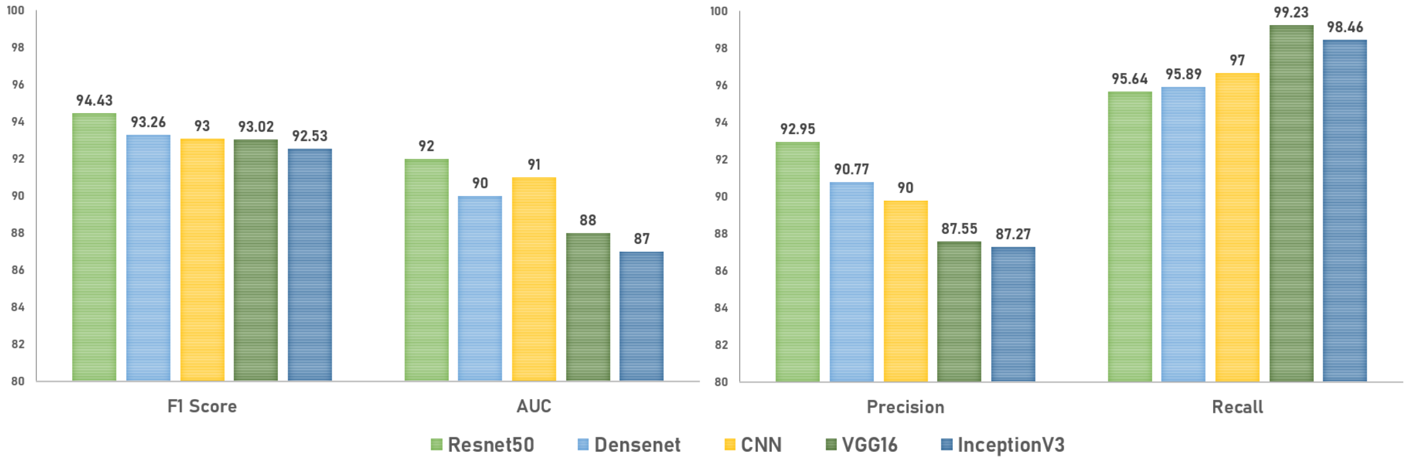
SUMMARY OF RESULTS

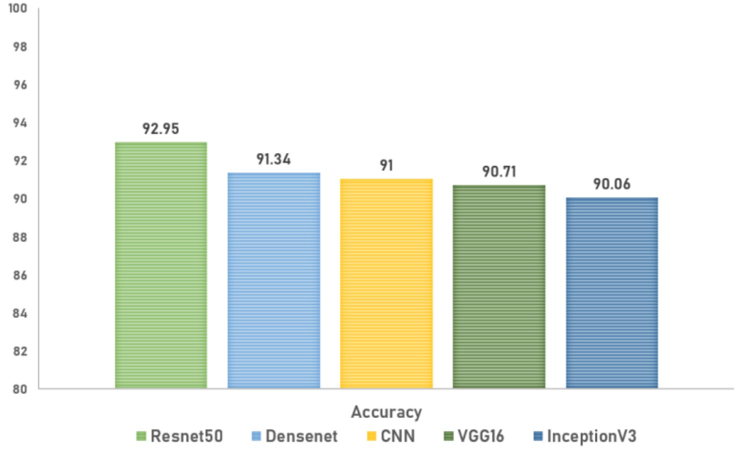
BACTERIAL VS. VIRAL CLASSIFICATION RESULTS





NORMAL VS PNEUMONIA RESULTS





CONCLUSION

- We can clearly see from the results that our **transfer learning models** are **performing better** than our **conventional CNN model**
- For **Normal vs Pneumonia case** - Our transfer learning model has performed remarkably well and has **produced Recall better than the benchmark score (93.2%)**; **VGG16 (Recall: 99.23)** being the champion model
- For **Bacterial vs. Viral case** - Our models have **produced better F1 scores and Precision than the benchmarks of F1 (62.36%) and Precision(48.11%)**; **CNN (F1-Score 93.08%)** being the champion model
- Showcased the **power of the transfer learning** system to make highly effective classifications (better than benchmarks), even with a very limited training data

FUTURE SCOPE

- Long way in ***improving the health of at-risk children*** in energy-poor environments
- Our study was ***limited by depth of Data***. With increased access to data and training of the model with radiological data from patients and nonpatients in different geography, significant improvements can be made
- Future studies could ***entail use of images from varied manufacturers***, so that the system will be universally useful
- **Ophthalmology, CT Scans** —in principle, the techniques we have described here could potentially be extended in a wide range of medical images across multiple disciplines
- Perform ***Occlusion testing*** to identify areas of greatest importance used by model while assigning a diagnosis

REFERENCES & ACKNOWLEDGEMENTS

- *Kermany et. al*, Michael Goldbaum, Michael Goldbaum, Wenjia Cai, Carolina C.S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown. Identifying medical diagnoses and treatable diseases by image-based deep learning. In Cell 172, 1122–1131, 2018
- *Kang Zhang*., MD, PhD, professor of Ophthalmology at Shiley Eye Institute and founding director of the Institute for Genomic Medicine at ***UC San Diego School of Medicine***
- ***Kaggle Dataset Link***
- *Github Code Repository:* https://github.com/KarthikeyaR/pneumonia-detection

We are grateful to **Prof. Sujoy Bhattacharya** for allowing us to pursue such an interesting topic for our course project, guiding us throughout and evaluating the project results

Thank You...