Chest X-Ray Classification

COMP 6721

Group Q

Problem overview

Why early diagnosis is important?

Early diagnosis of diseases like pneumonia and COVID19 leads to decreased mortality rate. Also a powerful way to manage pandemic.

Methods of Diagnosis

Diseases can be diagnosed by using variety of tests like CT scan, PCR, pulse oximetry but X-rays are the most accessible.
Since the X-rays are available in minutes it is one of the fastest ways of diagnosis.
Main bottleneck is expert radiologist needed to evaluate the scan.

Earlier attempts to solve the problem

Many attempts to solve this problem with deep learning but none that can replace radiologists. Other challenges include small data problem and model explainability as the ability to understand the reason for diagnosis is important for patients and doctors.

Main Challenges

Small and highly imbalanced datasets, different radiographic contrast of x-ray images across scanners.

This project

This project is an attempt to compare the three backbone architectures and lung disease datasets to find the model that works best for lung disease classification.

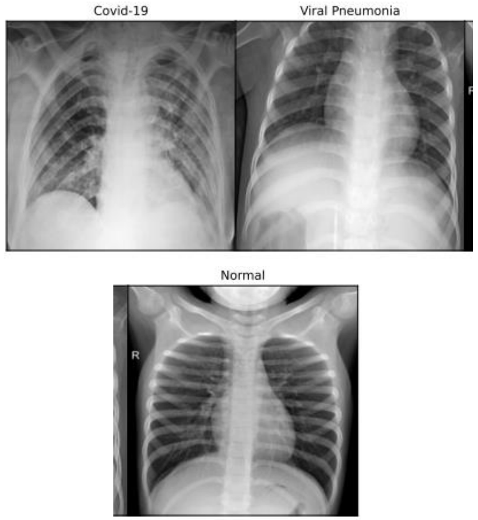
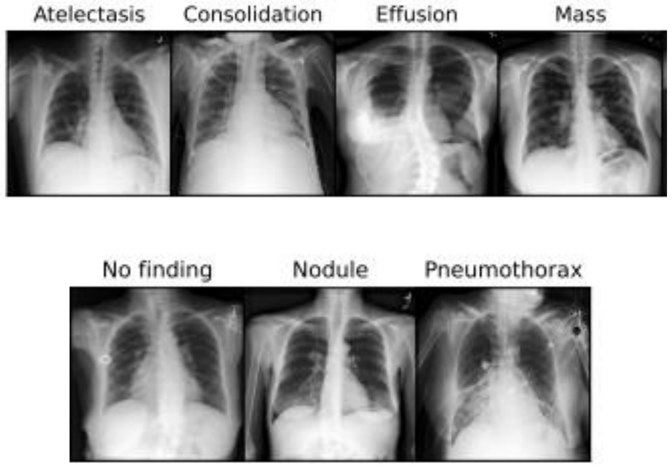
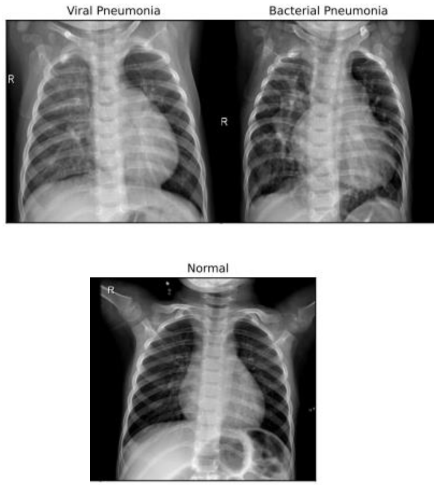
A detailed comparison of results across architectures and datasets along with the best model in terms of efficiency and F1 score is provided.

Goals

Chest X-ray Classification

- Perform data pre-processing on chest X-rays
- Perform Ablation Study
- Comparison across backbone architectures
- Comparison with transfer learning
- Model explainabiltiy using GradCAM
- Model visualization using t-SNE

Type of data

	COVID	Chest X-Ray 8	Pneumonia
# IMAGES	3.6K : 3.6K : 1.3K	7.2K : 7K : 7K : 4.1K : 3.9K : 3.5K : 2.9K	3K : 1.5K : 1.5K
SIZE	299 X 299	1024 X 1024	224 X 224
CLASSES	3	7	3
Sample			

Methodology

MODELS TRAINED

12 models i.e. 4 for each dataset were trained and fourth model was trained using transfer learning. Backbone models used are Resnet34, Mobile Net V3 large and Efficient Net B1.

MODEL VISUALIZATION

GRADCAM and TSNE visualization plots created for model explainability

DATA AUGUMENTATION

During training, images were augmented using RandomHorizontalFlip, RandomAdjustSharpness and RandomAutoContrast in PyTorch.

HYPERPARAMETERS

Hyperparameters will be fixed across models to produce comparable results.

DATA PREPROCESSING

Histogram Equalization and Gaussian Blur with a 5X5 filter.

LEARNING RATE SCHEDULER USED

Cosine Annealing LR

HYPERPARAMETER TUNING

Different values of learning rate (0.1, 0.05, 0.01, 0.001, 0.005) taken for ablative study.

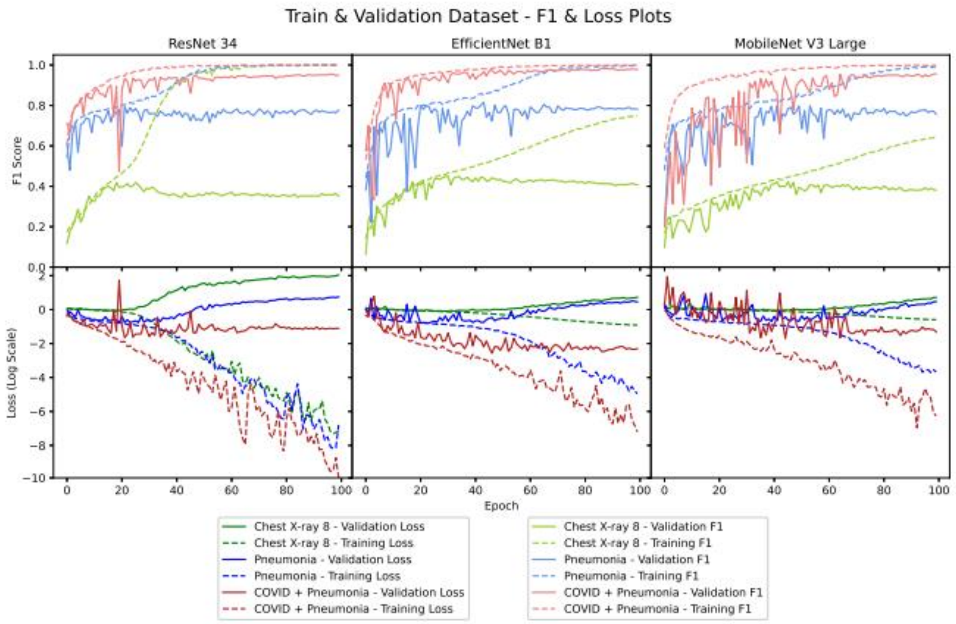
TRAIN VALIDATION TEST SPLIT

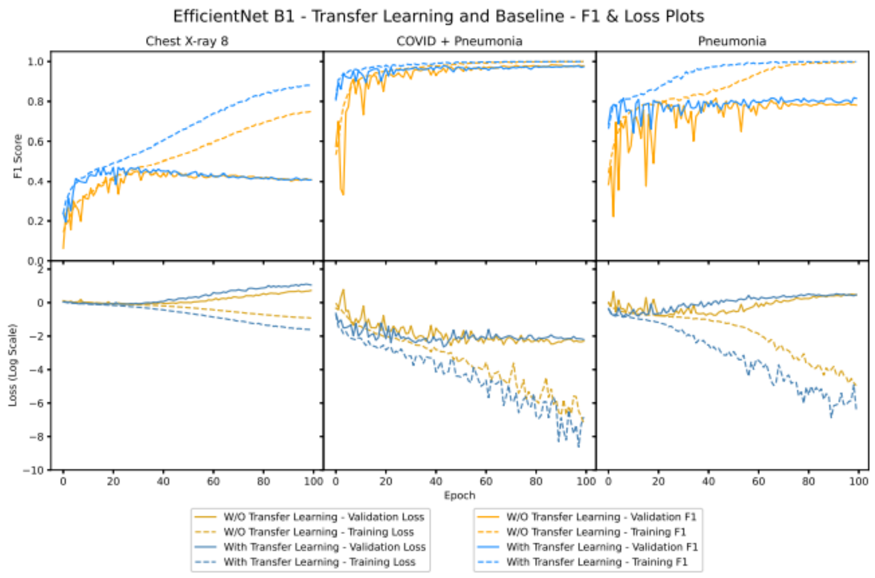
All scans divided into 70:15:15 split for train, validation and test set

OPTIMIZER AND LOSS FUNCTION

ADAM optimizer used along with cross entropy loss

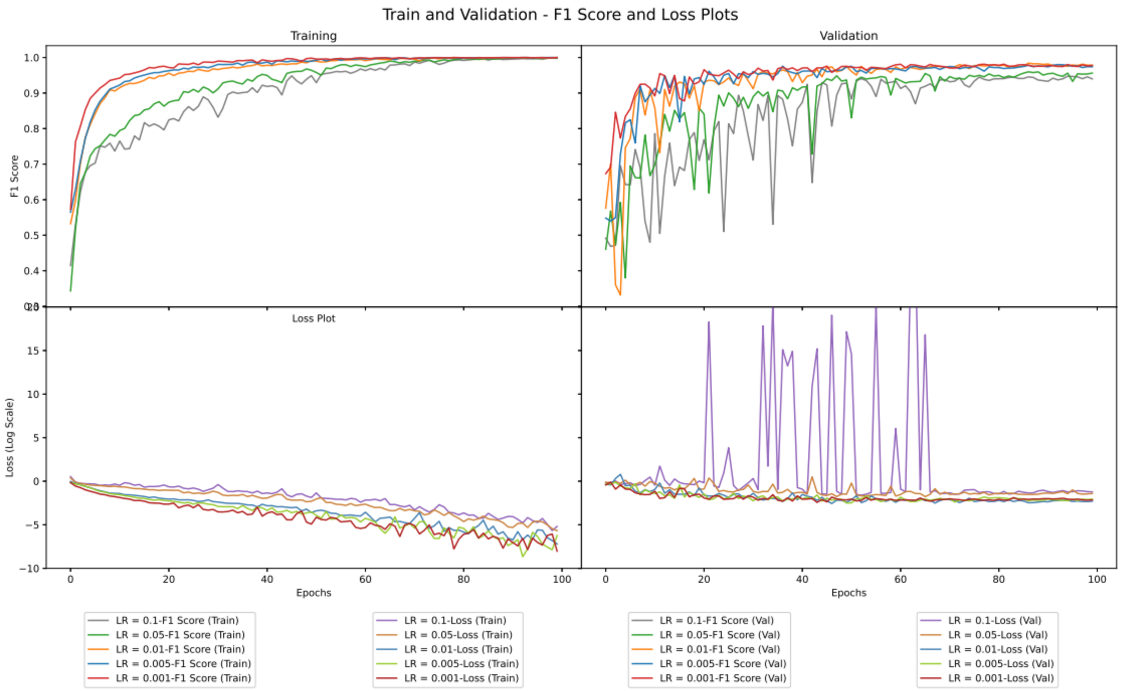
Results

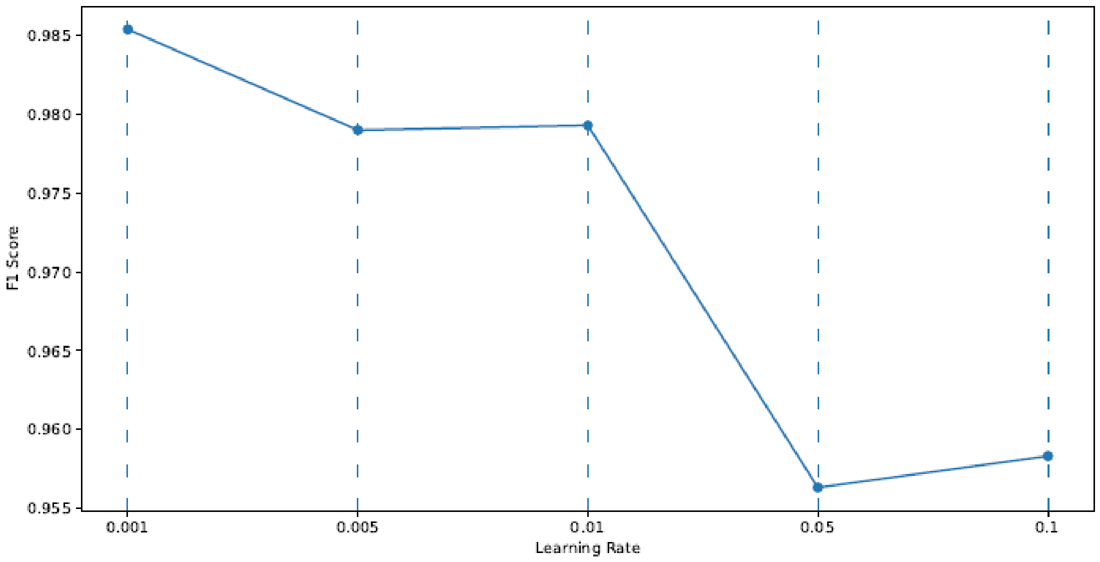




- ✓ From small to bigger architecture makes the model overfit earlier.
- ✓ MobileNet model was most unstable for early epochs among the three datasets.
- ✓ The EfficientNet architecture performs best for the COVID and Chest X-ray 8 dataset.
- ✓ Surprisingly, the pneumonia dataset performed worse than the COVID pneumonia dataset .
- ✓ Transfer learning model converged much quicker than the models trained from scratch.
- ✓ EfficientNet took the longest to train.

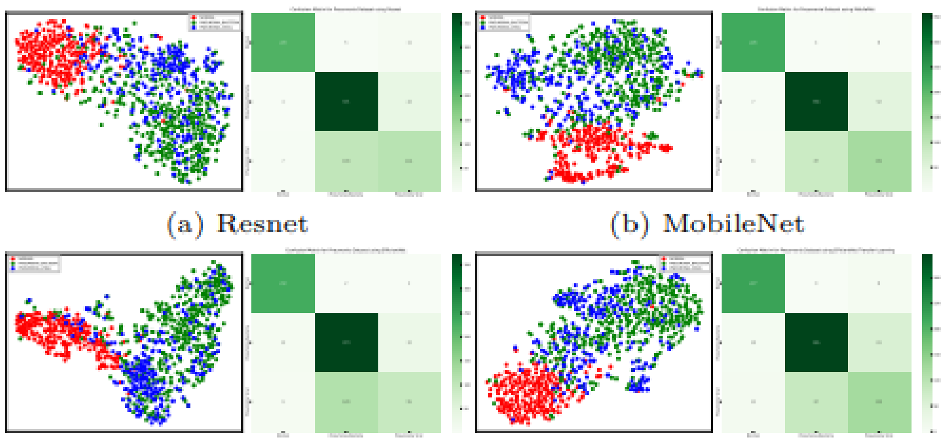
Ablation Study





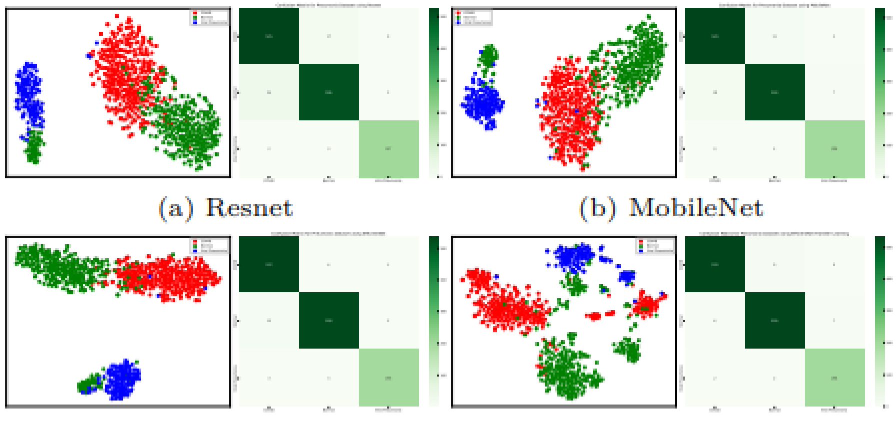
- ✓ A high learning rate of 0.1 makes the model highly unstable and prevents the model from reaching close to the global minima.
- ✓ The learning rate of 0.001 was the most stable and reached the highest F1 score earlier as compared to other learning rates.
- ✓ The learning rate of 0.001 gave the best F1 score on the test set with 0.005, 0.01 being close second and 0.05, 0.1 performing the worst.

Results



(a) Resnet (b) MobileNet
(c) EfficientNet (d) EN - Transfer Learning

Pneumonia Dataset

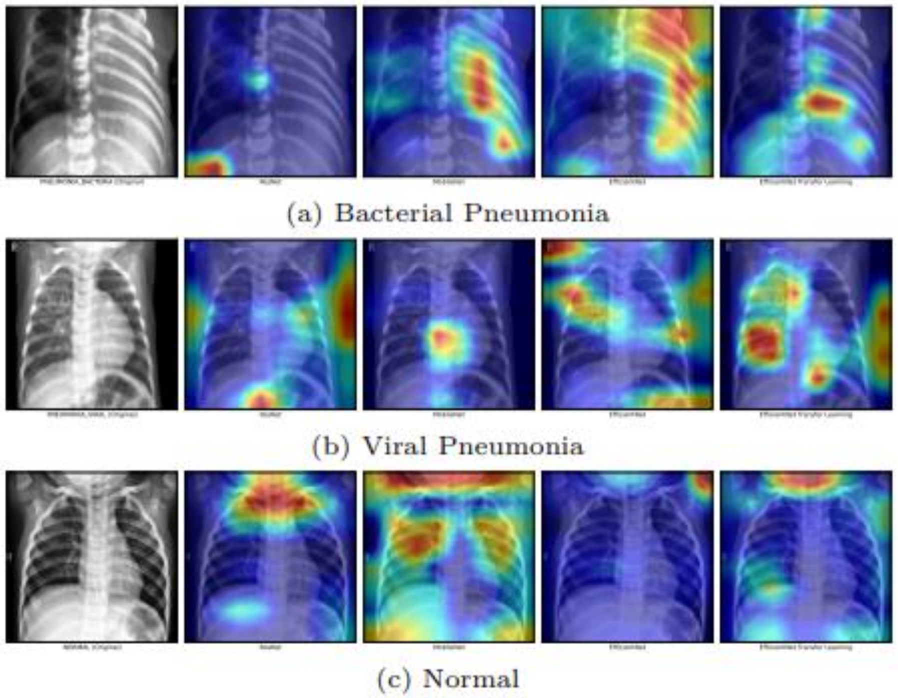


(a) Resnet (b) MobileNet
(c) EfficientNet (d) EN - Transfer Learning

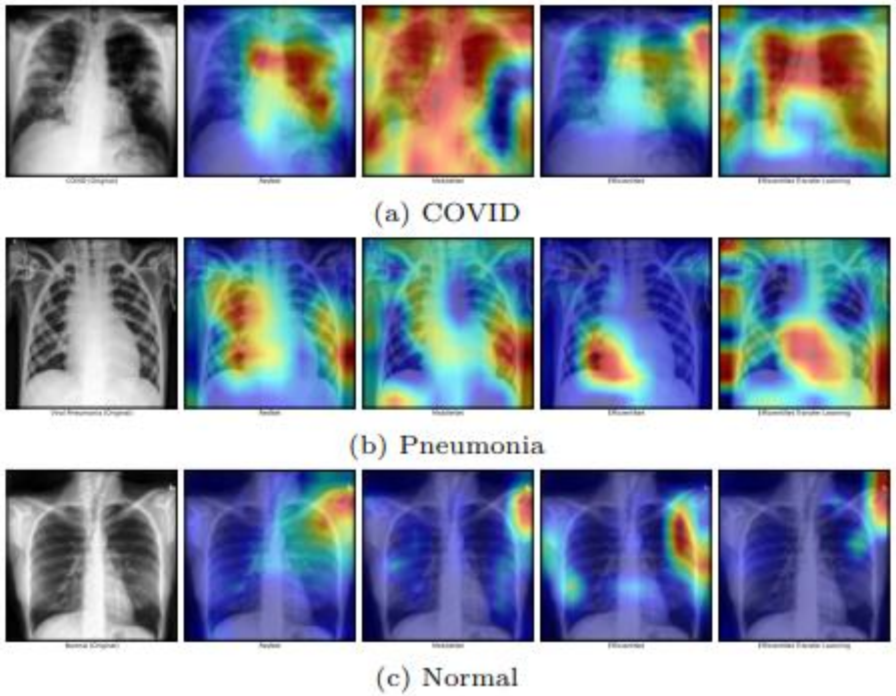
COVID Dataset

- ✓ The models are able to differentiate well between the normal and pneumonia classes but struggle with the viral pneumonia vs bacterial pneumonia classification.
- ✓ The EfficientNet transfer learning model performs the best amongst all models and hence would generalize well on new unseen data.
- ✓ All the models of the COVID dataset do a good job of separating classes to create distinct clusters but, the transfer learning model creates better clusters with separate smaller clusters.
- ✓ The performance of all the models from the confusion matrix and the t-SNE plots correlate.

Results



Pneumonia Dataset



COVID Dataset

- ✓ GradCAM shows that ResNet is learning completely different features as compared to the other models which could be the reason for low performance.
- ✓ EfficientNet transfer learning models activated the correct features and thus provided the highest accuracy.
- ✓ GradCAM activations correlate with the actual damage inflicted by the diseases on the lungs.

References

1) Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. and Summers, R.M., 2017. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097-2106).

2) Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F. and Dong, J., 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell, 172(5), pp.1122-1131.

3) National Institutes of Health, 2022. NIH clinical center provides one of the largest publicly available chest x-ray datasets to scientific community

4) Wang, G., Liu, X., Shen, J., Wang, C., Li, Z., Ye, L., Wu, X., Chen, T., Wang, K., Zhang, X. and Zhou, Z., 2021. A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images. Nature biomedical engineering, 5(6), pp.509-521.

5) Xu, L., Li, D., Ramadan, S., Li, Y. and Klein, N., 2020. Facile biosensors for rapid detection of COVID-19. Biosensors and Bioelectronics, 170, p.112673.

6) Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).