CS 470/670 – Introduction to Artificial Intelligence

Group 6 Term Project Final Report

Training Classification Models for Detecting Pneumonia Using X-Ray Images

Shabnam Azizi [02010937]

Ahmed Fassassi [01919774]

Kaif Khalifa [01956440]

Maxim Moskalenko [02033143]

Putha Sri Hari Reddy [02185878]

Trinadhreddy Seelam [02126243]

Abstract

This report outlines the development and evaluation of various machine learning models' performance on detecting pneumonia through X-ray images. Using the Kaggle "Chest X-Ray Images (Pneumonia)" dataset, we investigated transfer learning with ResNet50 and MobileNetV2, as well as a custom convolutional neural network (CNN). Our experiments demonstrated the effectiveness of data augmentation and the benefits of transfer learning for small datasets. Although the best-performing model achieved 84% accuracy, further improvements are necessary for real-world clinical applications.

Introduction

Pneumonia is a common and serious lung infection, with over three million cases reported annually in the United States (source: Mayo Clinic). Diagnosing pneumonia from chest X-rays can be challenging, even for expert radiologists, particularly in early stages. This project aims to develop a machine learning model capable of accurately detecting pneumonia from chest X-ray images, thereby streamlining the diagnostic process and enabling cost-effective early intervention.

The dataset used for this project is Kaggle's "Chest X-Ray Images (Pneumonia)" dataset, containing 5,863 labeled X-ray images: 32% labeled as "Normal" (healthy patients) and 68% labeled as "Pneumonia." The uneven distribution presented challenges in achieving balanced model performance, which we discuss in a further section.

Problem Statement

Pneumonia is difficult to detect with the naked eye through X-ray observation, necessitating the development of computational tools for accurate diagnosis. Our goal is to create a model that can reliably detect pneumonia from X-ray images, providing scalable and cost-effective preliminary screenings to assist medical professionals.

Project Procedure Breakdown

Brainstorming and Model Selection

We explored three main model types:

- ResNet50: A 50-layer CNN pretrained on ImageNet, evaluated using transfer learning and fine-tuning.
- MobileNetV2: A lightweight CNN optimized for efficient feature extraction and transfer learning.
- A custom CNN: Designed from scratch without pretrained weights to test its feasibility.

Data Preprocessing

In phase 1 of the project (before the first PPU), we partitioned the dataset into training, validation, and testing sets. Images were resized to a uniform resolution and batched for processing.

In phase 2, we applied data augmentation techniques, including random flipping, rescaling, zooming, and rotation, in order to mitigate overfitting.

Evaluation metrics included accuracy, validation loss monitoring, and confusion matrix visualizations.

Implementation

Part I - ResNet50 Models

Over the course of our project, we trained 4 models that built upon the ResNet50 deep learning model. One variation of the model utilized transfer learning, meaning the pretrained ImageNet weights remained frozen during training, and only our few added layers were trained using our x-ray dataset.

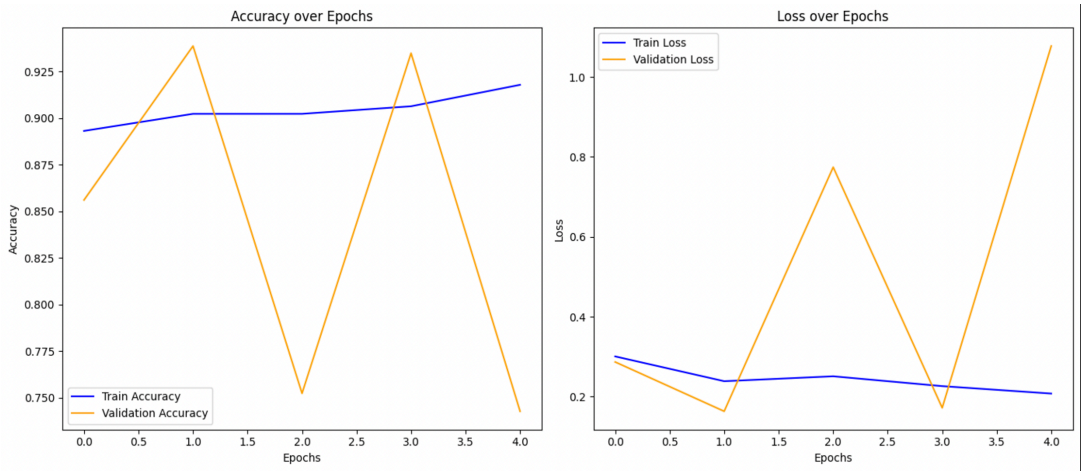
Of each of the two model variations, we trained them on the dataset without augmented training data, and with augmented training data, totaling to 4 final variations of the ResNet50 based models. The following is the resulting performance:

i. **Transfer Learning model without Data Augmentation:** Achieved **76%** accuracy over 10 epochs.

ii. **Transfer Learning with Data Augmentation:** Achieved **84%** accuracy over 10 epochs; note that performance slightly degraded when trained for 15 epochs due to overfitting.

iii. **Fine-Tuned Model without Data Augmentation:** Achieved **64%** accuracy over 10 epochs. Early stopped at 6 epochs due to continuous increase in validation loss.

iv. **Fine-Tuned Model with Data Augmentation:** Achieved **82%** accuracy over 10 epochs. Early stopped at 4 epochs due to continuous increase in validation loss.



- *ResNet50-based model #4 showing continuous loss increase after first epoch, stopping early at epoch 4 (early stopping with a patience level of 3 epochs).*

Part II - MobileNetV2 and Custom CNN

In Part I, it can be observed that data augmentation significantly improved model performance, and that transfer learning was much more efficient than finetuning.

Due to these observations, for or MobileNetV2-based model, we utilized only transfer learning, and for both that model and our custom CNN model, we used the training data with data augmentation applied. In fact, for all remaining subsequent models we used augmented training data due to its efficiency.

Layer (type)	Output Shape	Param #
conv2d_9 (Conv2D)	(None, 254, 254, 32)	896
batch_normalization_9 (BatchNormalization)	(None, 254, 254, 32)	128
max_pooling2d_9 (MaxPooling2D)	(None, 127, 127, 32)	0
dropout_12 (Dropout)	(None, 127, 127, 32)	0
conv2d_10 (Conv2D)	(None, 125, 125, 64)	18,496
batch_normalization_10 (BatchNormalization)	(None, 125, 125, 64)	256
max_pooling2d_10 (MaxPooling2D)	(None, 62, 62, 64)	0
dropout_13 (Dropout)	(None, 62, 62, 64)	0
conv2d_11 (Conv2D)	(None, 60, 60, 128)	73,856
batch_normalization_11 (BatchNormalization)	(None, 60, 60, 128)	512
max_pooling2d_11 (MaxPooling2D)	(None, 30, 30, 128)	0
dropout_14 (Dropout)	(None, 30, 30, 128)	0
flatten_3 (Flatten)	(None, 115200)	0
dense_6 (Dense)	(None, 128)	14,745,728
dropout_15 (Dropout)	(None, 128)	0
dense_7 (Dense)	(None, 1)	129

- *Custom CNN layers*

The results were as follows:

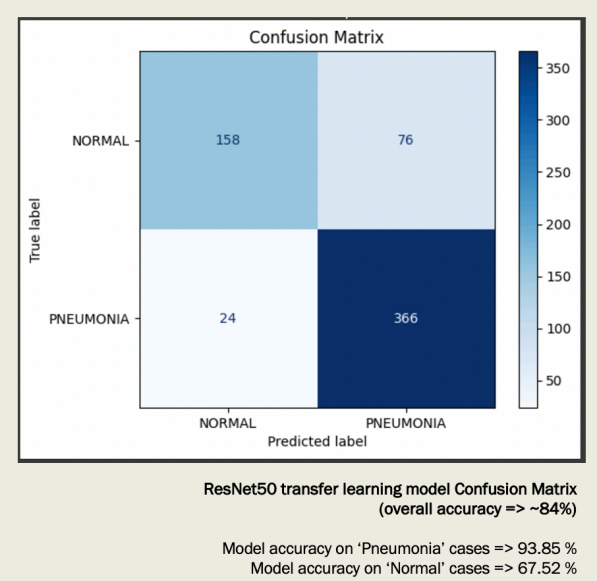
i. **MobileNetV2 transfer learning model:** This model resulted in an accuracy of **80%**, with early stopping applied after 6 epochs to prevent overfitting.

ii. **Custom CNN:** While this model was simpler in architecture, its lack of complexity and pretrained features led to a lower performance, achieving an accuracy of **61%** after 7 epochs of training.

Part III - Attempted Improvements

For this part, we noticed that our accuracy was plateauing at ~80%, and we did some investigating to uncover ways to improve accuracy to the 90% range.

During our investigation, we noticed that our models consistently performed better on pneumonia cases than normal cases, likely due to dataset imbalance (3,800 pneumonia examples vs. 1,300 normal examples).



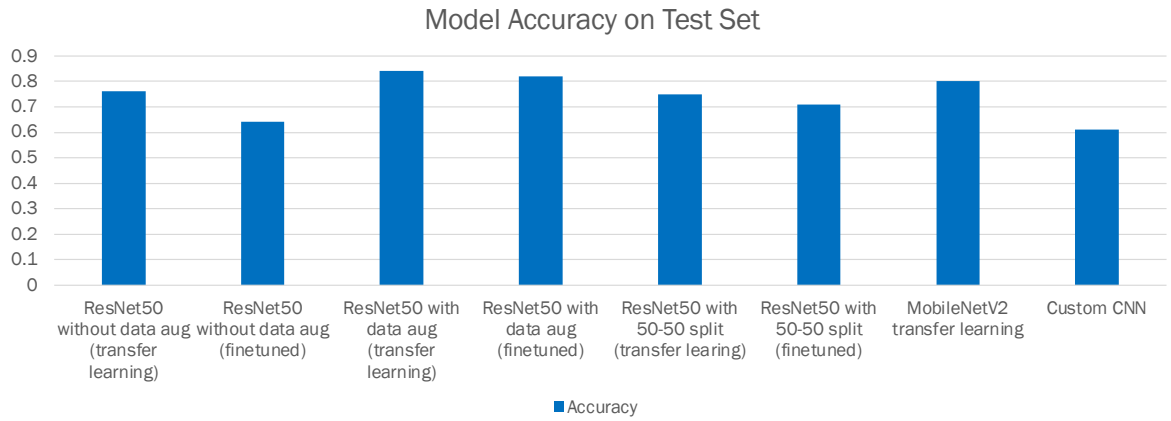
- Our best performing model's accuracy on 'Pneumonia' cases vs 'normal' cases

To address dataset imbalance, we tested equal splits of pneumonia and normal cases during training.

- *New dataset split—the majority examples (Pneumonia) were reduced to match the amount of minority examples (Normal), reducing the training data to ~1,300 x-ray images vs the prior ~5000 total images used for training (~3500 P, ~1500 N).*

However, this approach resulted in decreased accuracy for both transfer learning (75%) and fine-tuned ResNet50 (71%) models. Early stopping and data augmentation were still applied.

Overall Model Comparisons



Conclusion

The ResNet50 transfer learning model with data augmentation emerged as the best-performing approach, achieving an accuracy of 84%. Although promising, this level of accuracy remains inadequate for real-world medical applications without oversight from a qualified human expert. To address these limitations, future work could focus on enhancing the dataset by increasing its size and diversity, as well as implementing more sophisticated techniques to achieve better class balance. This project underscored the effectiveness of transfer learning for small datasets and emphasized the crucial role of data augmentation in improving model performance and reducing overfitting. Together, these insights lay a strong foundation for further advancements in leveraging AI for medical imaging tasks.

References

- Akay, Metin et al. "Deep Learning Classification of Systemic Sclerosis Skin Using the MobileNetV2 Model." https://www.researchgate.net/publication/350152088_Deep_Learning_Classification_of_Systemic_Sclerosis_Skin_Using_the_MobileNetV2_Model.
- Sandler, Mark et al. "MobileNetV2: Inverted Residuals and Linear Bottlenecks." https://arxiv.org/abs/1801.04381.
- "Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images." https://www.sciencedirect.com/science/article/pii/S235291482200065X.
- "Transfer Learning." Wikipedia. https://en.wikipedia.org/wiki/Transfer_learning