Springboard Data Science Capstone 3 -

Optimizing a Convolutional Neural Network for Pneumonia Classification in Radiographs

Table of Contents

1. Executive Summary

This project involved developing a convolutional neural network for pneumonia detection in chest x-rays. A key challenge was balancing precision and recall - minimizing false positives while not missing cases. Threshold adjustment optimized this tradeoff, achieving 80% precision and 97% recall. This improved diagnostic accuracy while limiting unnecessary workups.

2. Introduction

2.1 Background:

Computer vision techniques like convolutional neural networks (CNNs) provide an automated approach to analyzing medical images and extracting diagnostic information. For chest x-rays specifically, prior research has shown promise in using CNNs to identify abnormalities like pneumonia that can support clinical diagnosis and treatment.

2.2 Objective:

The objective of this project was to develop a CNN model to classify chest x-ray images into normal or pneumonia categories. However, misdiagnoses create consequences. False positives lead to unnecessary workups, while false negatives miss treatments. <u>The goal was building a model balancing precision and recall to improve accuracy while managing costs</u>

2.3 Scope:

The model development process included collecting and preprocessing a pneumonia image dataset, designing and training a CNN architecture, and evaluating model performance on a held-out test set. Additional techniques like class balancing, early stopping, and threshold optimization were employed to improve model training and maximize classification precision and recall.

3. Data Collection and Preprocessing

The chest x-ray images were sourced from the [Kaggle Chest X-Ray Images (Pneumonia)] dataset. This dataset contains 5,863 JPEG images across two categories - Pneumonia and Normal.

The images were split into training, validation, and testing sets in a 5216, 16, 624 ratio respectively. Prior to model training, the training data underwent several preprocessing steps:

- The images were resized to 224 x 224 pixels
- Oversampling was done with *SMOTE to balance the two classes
- The images and labels were shuffled and normalized
- The pneumonia/normal labels were encoded

4. Model Building and Compilation

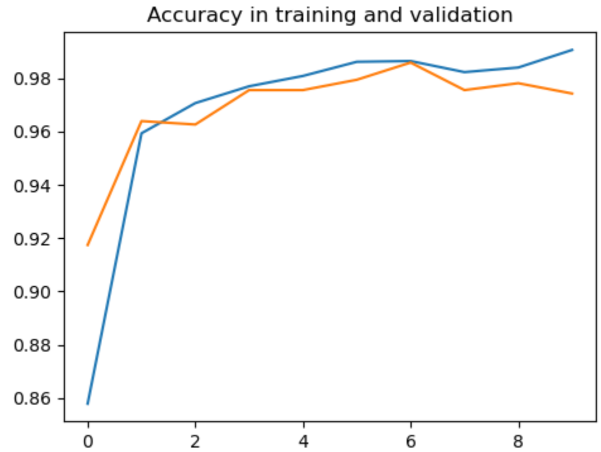
A CNN architecture was developed, consisting primarily of convolutional layers interspersed with max pooling layers for downsampling. The convolutional feature maps were flattened and passed through fully connected dense layers before a binary classification output layer with sigmoid activation. The model was compiled with binary cross entropy loss to optimize for the pneumonia classification task. The Adam optimization algorithm was used along with accuracy metrics. The final CNN model had approximately 2.8 million trainable parameters.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 222, 222, 32)      320

 conv2d_1 (Conv2D)           (None, 220, 220, 64)      18496

 max_pooling2d (MaxPooling2D  (None, 110, 110, 64)     0
 )

 conv2d_2 (Conv2D)           (None, 108, 108, 32)      18464

 max_pooling2d_1 (MaxPooling  (None, 54, 54, 32)       0
 2D)

 conv2d_3 (Conv2D)           (None, 52, 52, 64)        18496

 max_pooling2d_2 (MaxPooling  (None, 26, 26, 64)       0
 2D)

 flatten (Flatten)           (None, 43264)             0

 dense (Dense)               (None, 64)                2768960

 dropout (Dropout)           (None, 64)                0

 dense_1 (Dense)             (None, 1)                 65

=================================================================
Total params: 2,824,801
Trainable params: 2,824,801
Non-trainable params: 0
_________________________________________________________________
```

5. Model Training

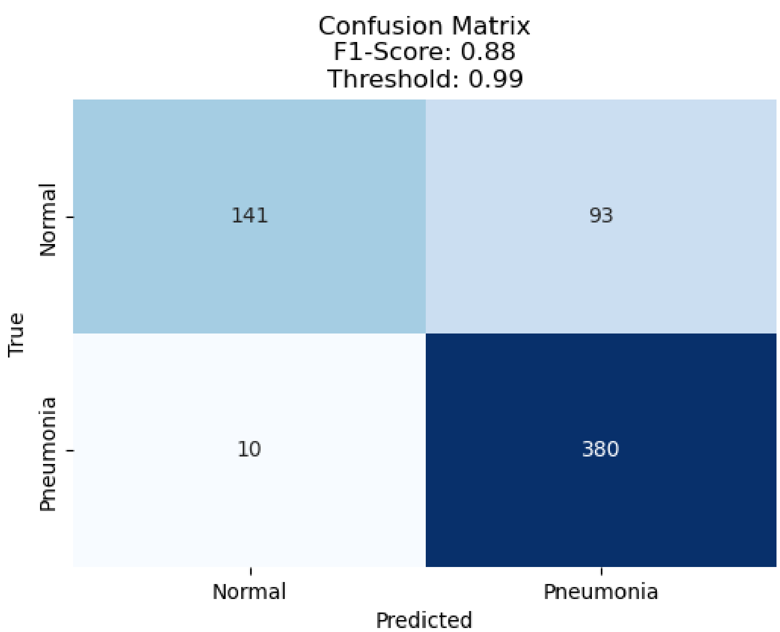
The CNN model was set to train for 15 epochs using the prepared training data. An early stopping patience of 3 epochs was used to prevent overfitting. The model stopped early after the 10th epoch. The Adam optimizer and binary cross-entropy loss function were used to update weights and minimize loss. Validation accuracy was monitored at the end of each epoch to check for overfitting.

The model achieved a training accuracy of 99% and training loss of 0.02. The validation accuracy was 97% with a validation loss of 0.08, indicating acceptable generalization.



6. Model Evaluation

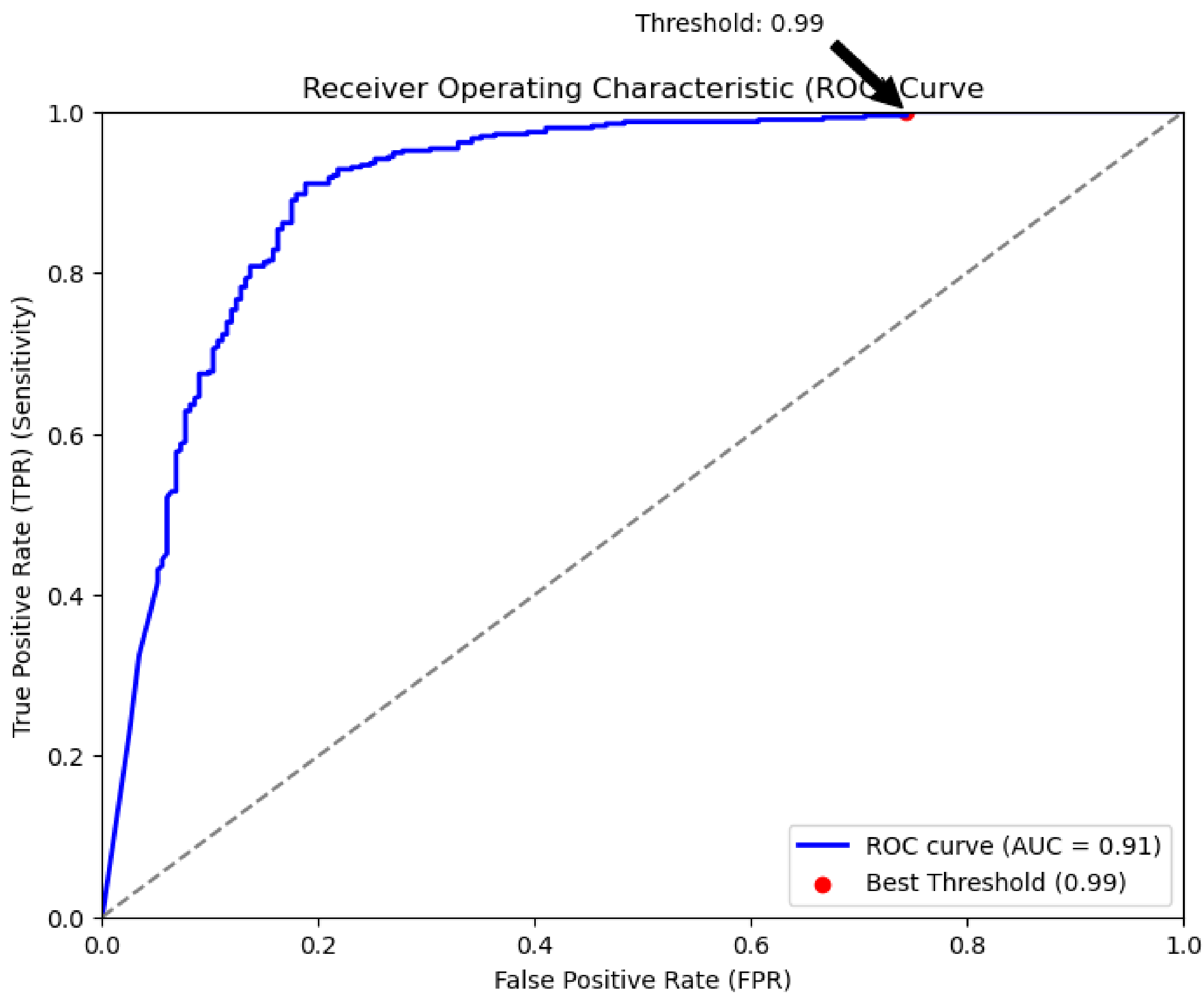
The trained model was evaluated on the test set to determine its real-world pneumonia classification performance. The primary goal was to balance false positives and false negatives, maximizing accuracy, precision and recall. The default model achieved 73% accuracy, 70% precision and 99% recall.



By adjusting the classification probability threshold to 0.989, the precision rose to 80% and recall dropped slightly to 97%.

This is the best threshold to balance the precision and recall of this model.This increased the overall F1 score, reflecting a better balance of precision and recall, optimizing diagnostic performance and managing total workup costs.

We were able to prevent 76 incorrect diagnoses but at the cost of 8 missed ones.



7. Assumptions and Limitations

The model made the assumption that the test data would match the patterns in the training data. With more variability in real-world x-rays, the model performance could potentially degrade. The relatively small dataset also limits the robustness of the model. Additionally, the training classes were balanced artificially with SMOTE oversampling, which may impact generalizability. Finally, biases inherent in the data could affect generalizability.

8. Future Work

Additional work on this model could help improve its viability for clinical use. Collecting more varied training data could make the model more robust. Further tuning the CNN architecture and training hyperparameters could potentially boost accuracy further. Additionally, employing ensemble techniques with multiple models could lead to more stable predictions.

9. Conclusion

In conclusion, this project demonstrated successful development and application of a CNN model for classifying chest x-rays as normal or depicting pneumonia. Using a pneumonia image dataset, the model was trained on over 5,000 images and optimized using techniques like early stopping and threshold adjustment. The final CNN model achieved an accuracy of 83% on the test set. Thresholding the classification probabilities improved the precision from 69% to 80%. Overall, the project demonstrated effective application of a CNN architecture and training techniques for classifying x-ray images relevant to pneumonia detection. With further improvements to the training data and model architecture, this type of AI technology could become a valuable tool for radiologists and clinicians working to detect pneumonia cases.