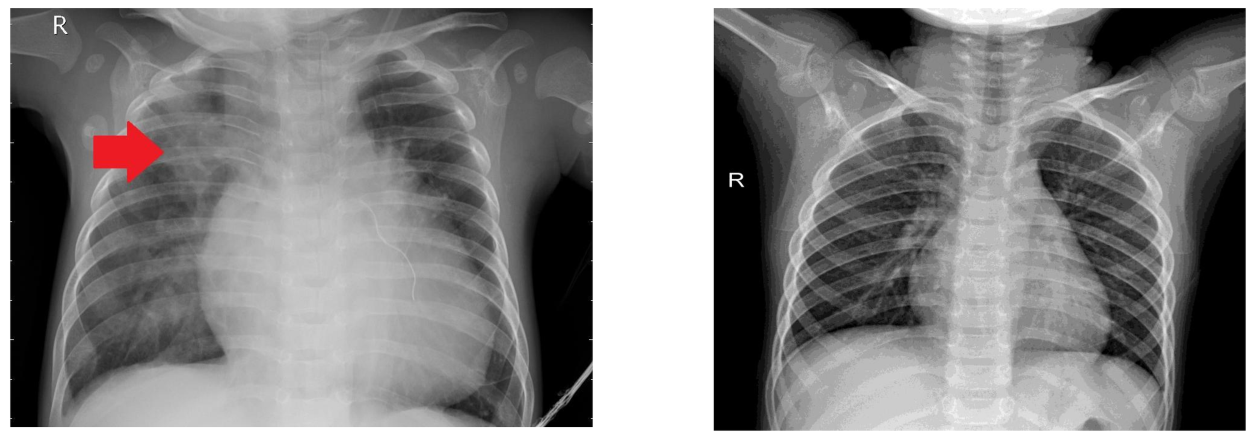
Identification of Pneumonia from X-Ray Images using Machine Learning Techniques

Ruchit Doshi, Lalit Ghule, Akanksh Shetty, Jeffrey Woodford

Carnegie Mellon University, Department of Mechanical Engineering

INTRODUCTION

In the United States, more than 250,000 people go to the hospital because of pneumonia and around 50,000 people die from it each year [1]. Quick and accurate diagnosis of pneumonia is critical, as the symptoms often overlap with more mild conditions such as colds and the flu, meaning it can grow unsuspected until it is too late. According to the National Heart, Lung, and Blood Institute, the best test for diagnosing pneumonia is a chest X-Ray [2]. However, detecting pneumonia can be challenging and requires experienced radiologists, and radiologists are not consistent amongst each other when diagnosing pneumonia based upon X-Rays [3]. To help counteract this variability in performance of trained radiologists, a machine learning model can be created that can flag X-Rays for more in-depth evaluation for potentially having pneumonia. This project seeks to create different models to assist radiologists when diagnosing pneumonia.



(a) (b)

Figure 1. Frontal Chest X-Ray to diagnose Pneumonia. (a) X-Ray with higher density region corresponding to Pneumonia pointed out. (b) normal healthy chest X-Ray.

METHOD – Shallow Algorithms

Dataset: A dataset from Kaggle was used for this project [4][5]. It consists of three sets of images of chest X-Rays labeled as Test, Train and Validation. There are 3742 images in the training set, consisting the chest X-Rays of patients with and without Pneumonia. The Testing folder consists of 624 images with their labels which can be used to verify the accuracy of the models. In order to use the raw pixels of the images as the input to the algorithms, every image was resized to 150x150 pixels. Different sizes of images were also tested, such as 224x224. The images were converted to grayscale and normalized.

Shallow Machine Learning Algorithms such as Logistic Regression, Support Vector Machine, Random Forest, and XG-boost were applied using the normalized pixel values of the images, and later, the convolutional layer's flattened output was used as well. To improve the performance of the models, they were combined and the prediction of an X-Ray image was determined by taking the mode of the model outputs, as shown in Figure 2.

Tuning of hyper parameters were done as follows:

- Random Forest: The number of estimators were tweaked, keeping all other parameters as default. N_estimators was tweaked from 10 to 30. The optimum result was found to be 20.
- Support Vector Machine: Different Kernels were tested such as Linear, Poly, and Radial basis Function (RBF). The best results were obtained with the Linear kernel.
- XG_boost: Different learning rates and N_estimators were tried, and the optimum results were obtained on N_estimators=2000 and learning rate=0.01.

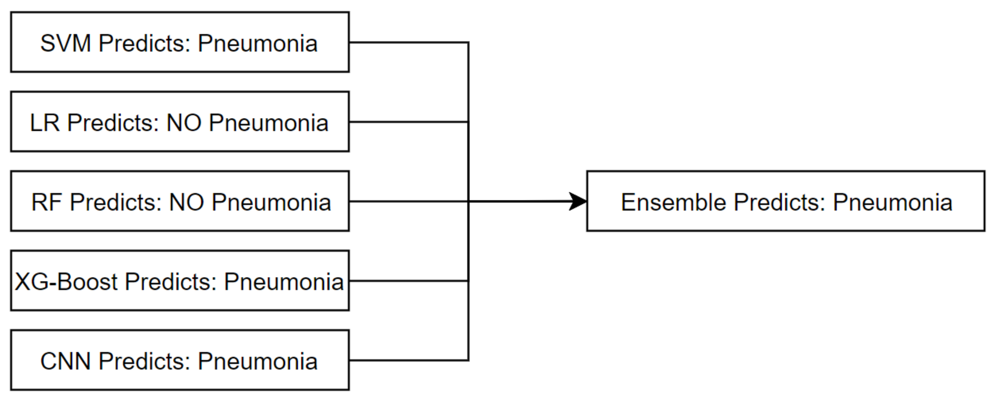


Figure 2. Ensemble Method takes mode of all models and returns it as prediction

METHOD – Deep Algorithm

A Convolutional Neural Network was implemented using Keras with TensorFlow backend. Many different architectures were implemented before arriving at the one shown below in Figure 3. The network below has a total of eight convolution layers . In addition to this, a Dropout of 0.3 was used once after the first two convolution layers. Max pooling was performed every two convolution layers. After every max pooling, we normalized the obtained output. These operations were then followed by four fully connected layers. The final layer of the fully connected network has only one output to which we applied a sigmoid nonlinearity. The network is trained using the Adam optimizer. A Mini-batch of size 32 was used. Binary Cross-entropy loss was used in this model.

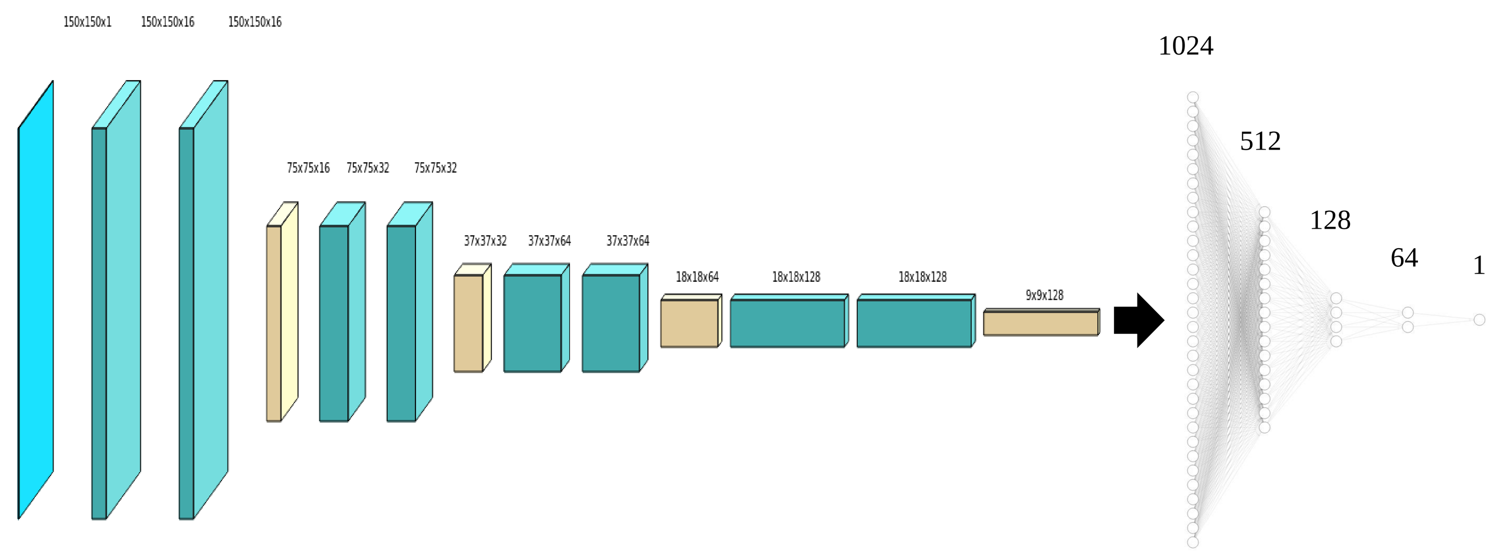


Figure 3. Visualization of Convolutional Neural Network Model

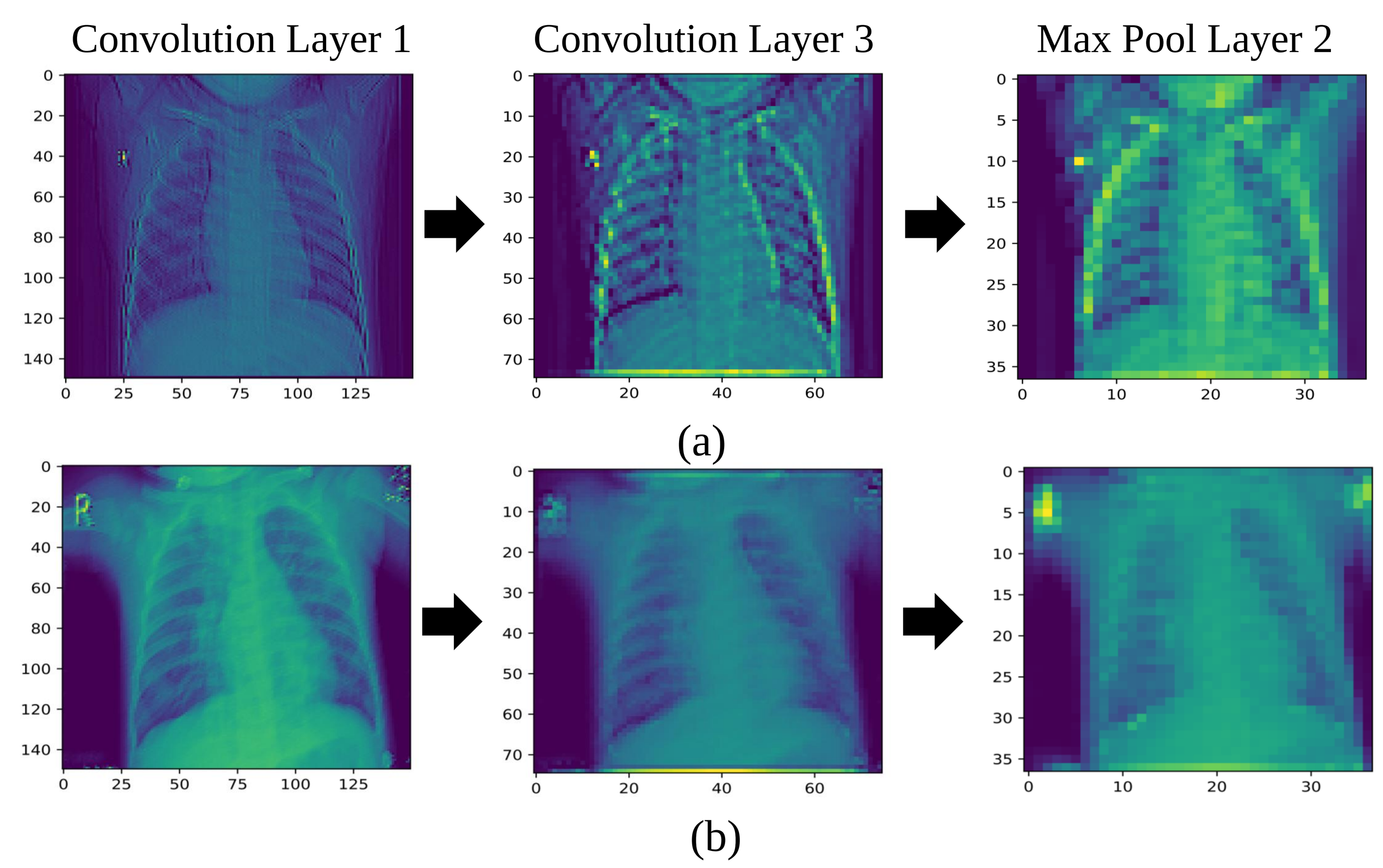
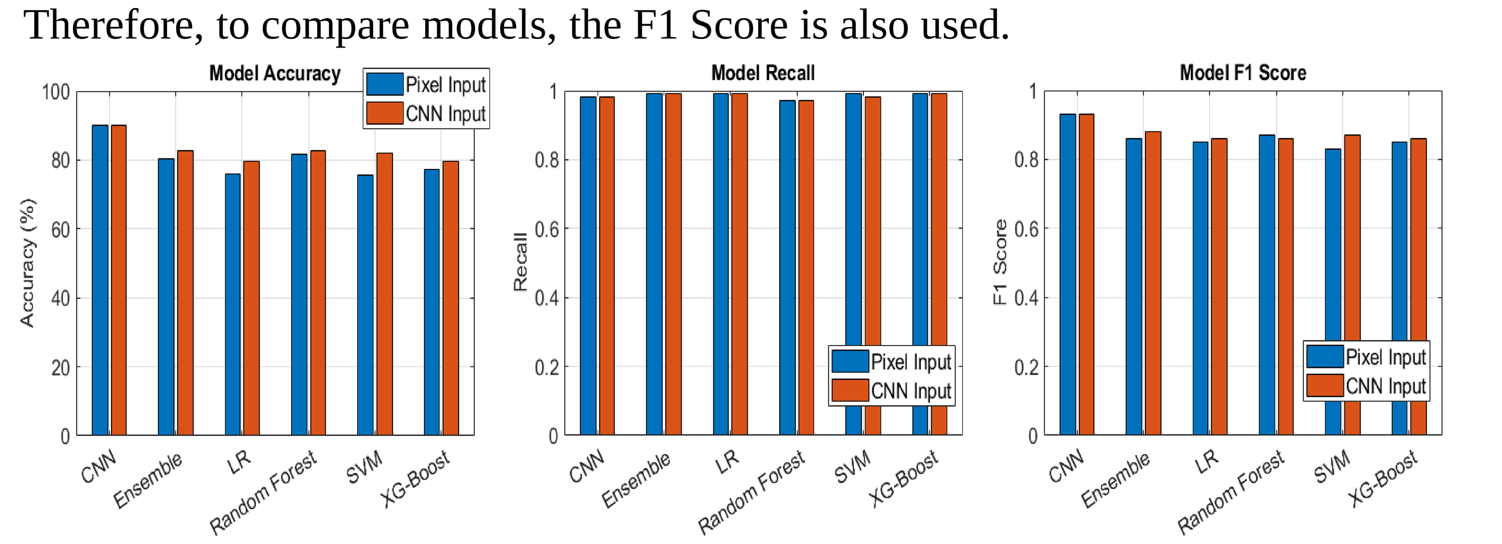


Figure 4. Visualization of select convolutional filters. (a) Normal (b) Pneumonia

RESULTS – Shallow Algorithm

In evaluating the performance of the models, a different method is needed than just comparing accuracy alone. The recall of the models is more important than the accuracy because it is associated with how well the model can catch cases of pneumonia. But recall should be balanced with precision so that the model is useful. Therefore, to compare models, the F1 Score is also used.



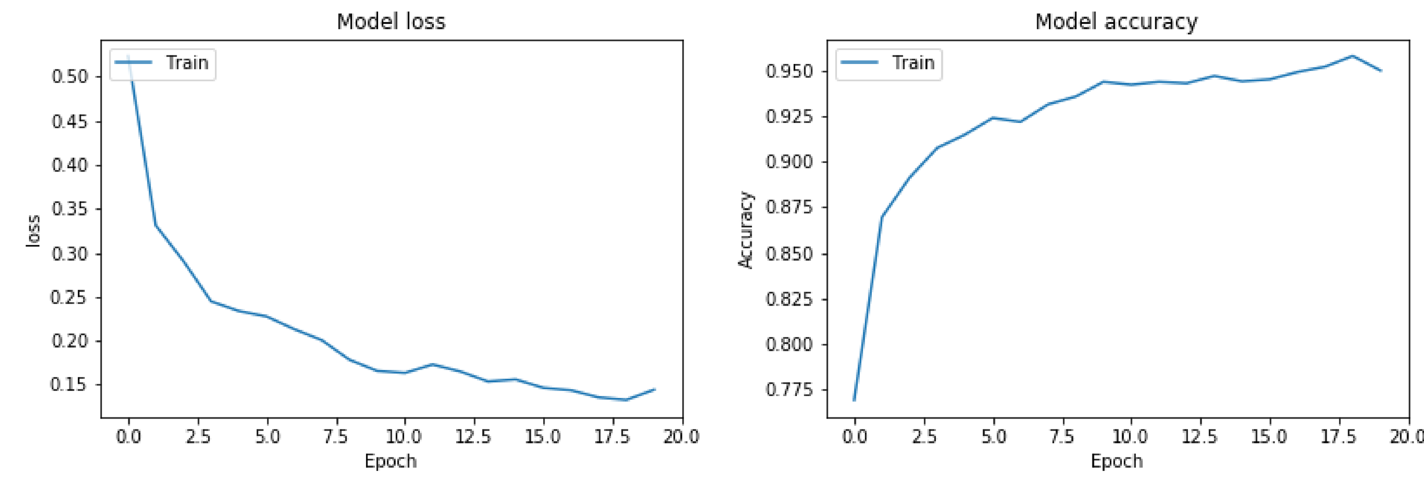
(a) (b) (c)

Figure 5. Comparison of shallow and deep models for different inputs. (a) Accuracy (b) Recall (c) F1 Score

RESULTS – Deep Algorithm

Changing the input to the CNN output increased the individual accuracies and F1 scores without significantly changing the Recall. For the ensemble, the accuracy increased by 2% and the F1 Score increased by 0.02 while maintaining the Recall at 0.99. Therefore, the input of the models should be the output of the convolutional layers.

However, the recommended method of use would be to drop the shallow models and ensemble and exclusively use the CNN. If only the CNN is used, the model will have an accuracy of 90% and a F1 Score of 0.93 while maintaining a Recall of 0.98.



(a) (b)

Figure 6. Convolutional Neural Network (a) Model Loss (b) Model Accuracy.

CONCLUSIONS

The best performing model with an accuracy of 90.06%, a F1 Score of 0.93, and a Recall of 0.98 was the CNN. This is not surprising as CNN's are preferred when handling large inputs and images because of the convolution layers. With an accuracy of 82.69%, a F1 Score of 0.88, and a Recall of 0.99, the combined model did not perform as well, but is still useful. One of the most likely reasons for the poorer performance of the combined model is that the input of the 224x224 pixel image was too large for the models to handle. All the shallow algorithms that were used started to overfit in the first few Epochs. To improve the combined model, different methods of extracting features from images should be implemented, as was found by changing the input to the output of the convolutional layers rather than the pixel values. Overall, both methods could be used by radiologists to assist in the diagnosis process, but care should still be taken when using these models and a radiologist should not rely on the model to make a diagnosis, but rather use it as a tool help identify X-Rays that may indicate the presence of pneumonia.

ACKNOWLEDGEMENTS

Prof. Amir Barati Farimani, Mechanical Engineering, CMU; Rishikesh Magar, CA, 24-787; Mechanical and Artificial Intelligence Lab (MAIL, CMU);

REFERENCES

[1] "Pneumonia Can Be Prevented-Vaccines Can Help," Centers for Disease Control and Prevention, 21- Nov-2019. [Online]. Available: https://www.cdc.gov/pneumonia/prevention.html?CDC_AA_refVal=https://www.cdc.gov/features/pneumonia/index.html.

[2] "Pneumonia," National Heart Lung and Blood Institute. [Online]. Available: https://www.nhlbi.nih.gov/health-topics/pneumonia. [Accessed: 02-Dec-2019].

[3] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, R. L. Ball, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning."

[4] P. Mooney, "Chest X-Ray Images (Pneumonia)," *Kaggle*, 24-Mar-2018. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia.

[5] Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", Mendeley Data, v2 http://dx.doi.org/10.17632/rscbjbr9sj.2