PREDICTING PNEUMONIA IN LUNGS X-RAYS

Dhruv Ragunathan

BUSINESS OBJECTIVES

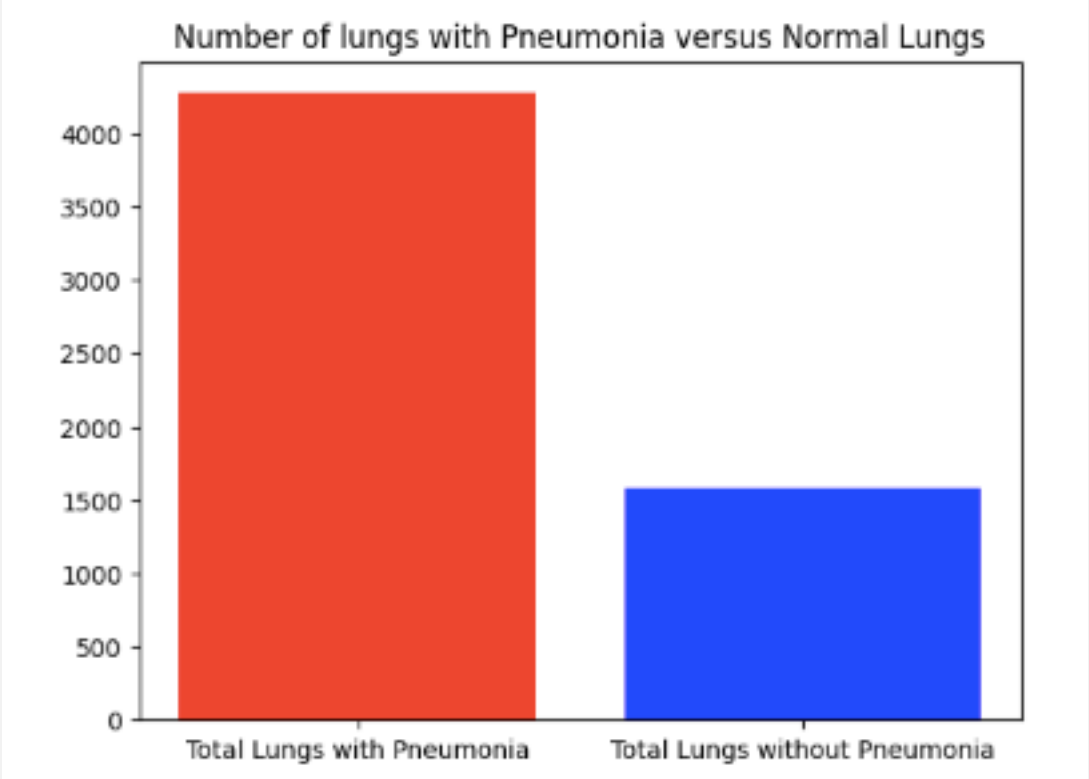
- We have been tasked to improve identification of pneumonia by the CIO.
- Staffing resources are in a crunch and there aren't enough radiologists to review all x-rays.

- We use a machine learning model called a convolutional neural network (CNN) to identify pneumonia in lung x-rays.

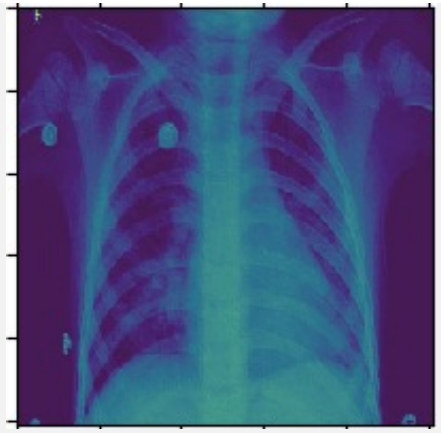
- We believe our findings show promise in using this technique to alleviate hospital staff time.

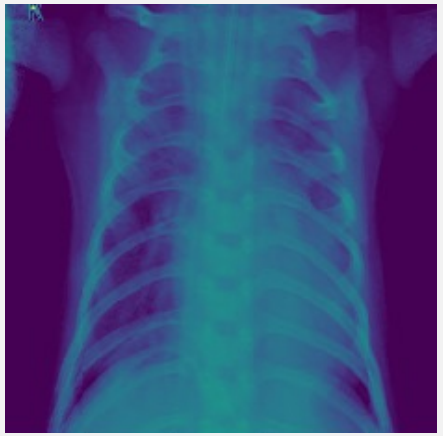
DATA

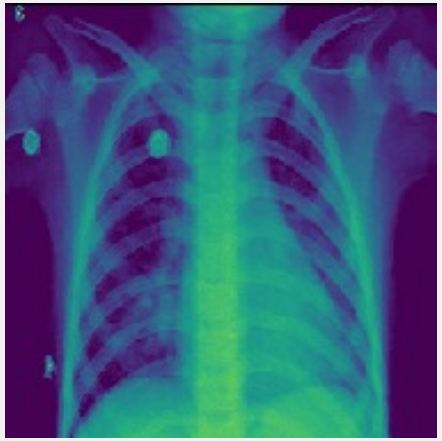
- Data sets off lungs with Pneumonia and healthy lungs were provided by Guangzhou Women and Children's Medical Center, Guangzhou.
- The data contains 3 data sets, train, test, and validation. In each file, there are two folders: Pneumonia and normal.
- The data only contained chest x-rays of patients that are less than 5 years old.



DATA CONTINUED





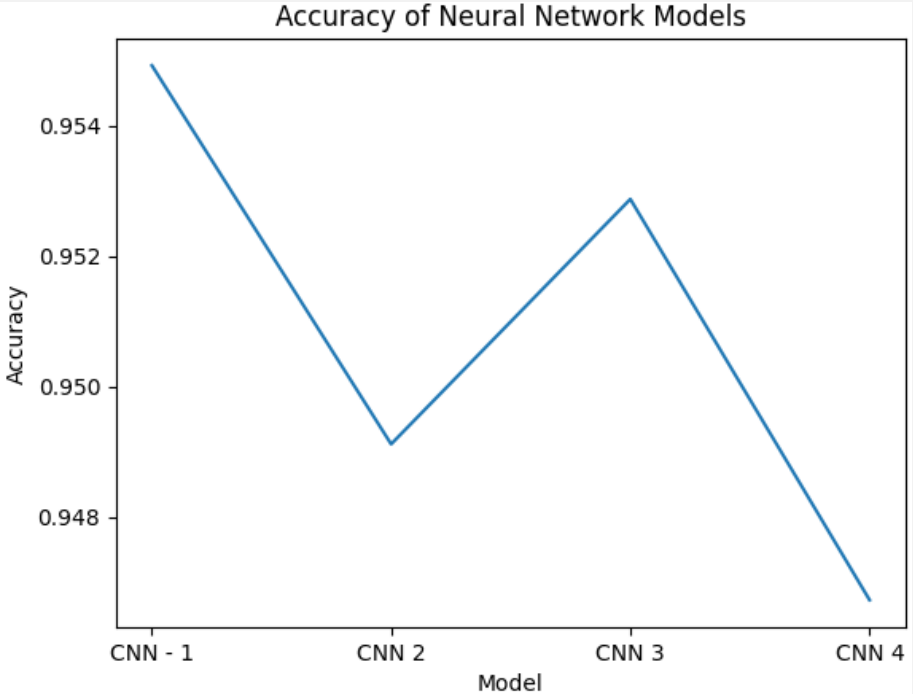


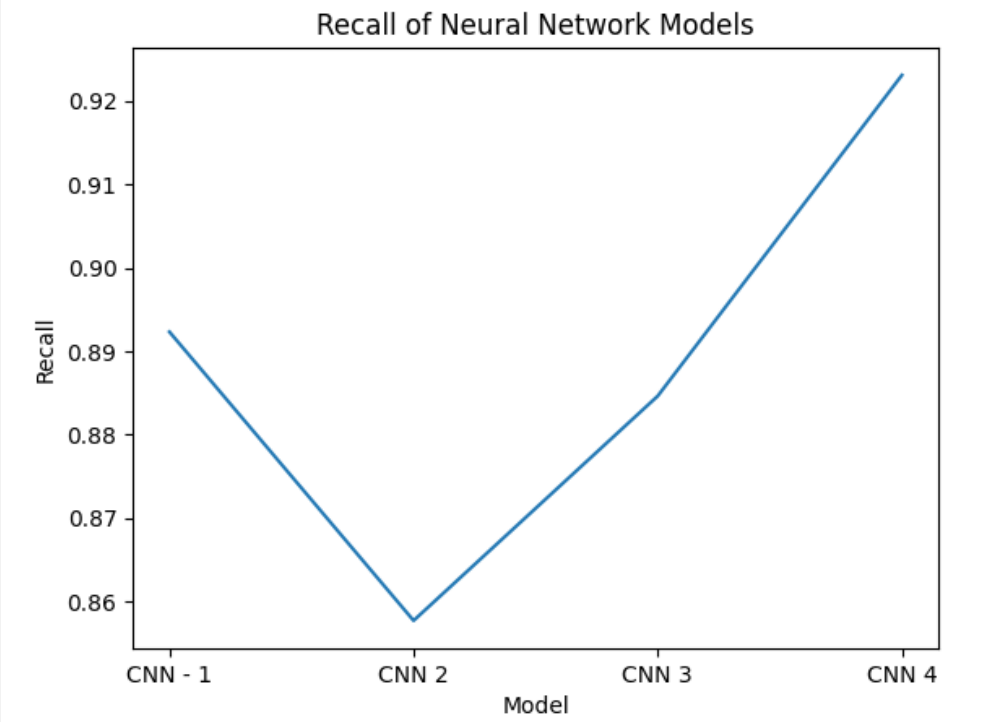
MODELS AND EVALUATION CRITERIA

- Due to resource and the limits off the machines, only 2500 off the 5800 X-rays could be used in training the model.
- Four CNN models were created as part off this project.
- The evaluation criteria were accuracy and recall.
  - Accuracy: Higher accuracy means more model predictions were correct (lower false positive)
  - Recall: Higher recall means more pneumonia x-rays were identified (lower false negatives)
- Recall was prioritized over accuracy because a higher recall implies that more lives were saved.

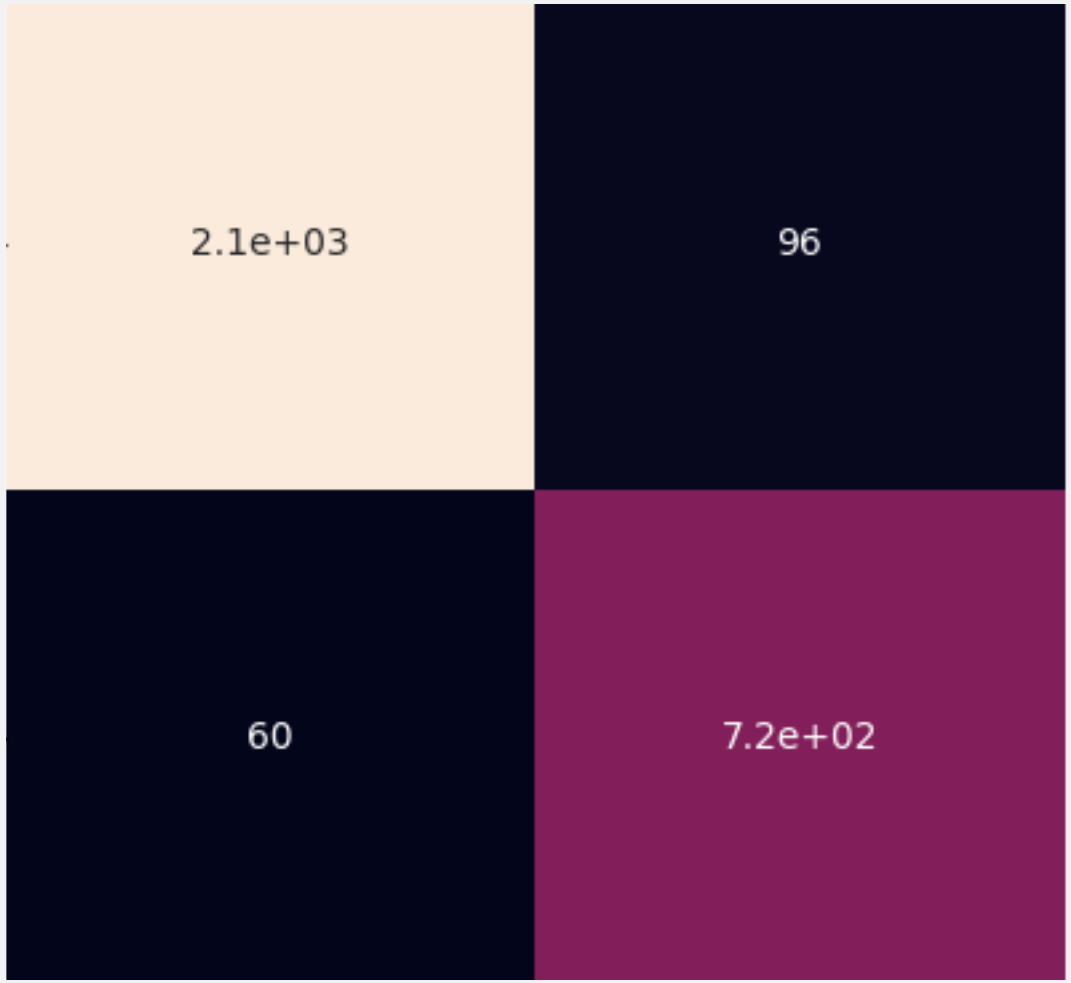
MODEL RESULTS

- The model's accuracies ranged from 94.5% to 95.5%. They were only off by 1%.
- The model's recall ranged from 86 to 92%.
- Since recall is more important than accuracy, we are selecting model 4.





MODEL RESULTS – CONFUSION MATRIX



- Model correctly identified 2,100 lungs with pneumonia
- Model correctly identified 720 lungs without pneumonia.
- Model incorrectly identified 96 normal lungs as having pneumonia.
- Model incorrectly identified 60 pneumatic lungs as if they had pneumonia.

RECOMMENDATIONS

- Recommend that you move forward with using the final model since it had the highest recall by 3% and only was 1% less accurate than the other models.
- Further research on how to operationalize this model in hospital workflows. Examples:
  - How are false positives and negatives handled in a hospital?
  - Is there a method for clinicians to quickly determine that the patient does not have pneumonia?
- Training on larger and more diverse data sets is necessary before it's deployed.

FUTURE PROJECTS

- Aggregate significantly more lung data across groups off different ages.
- Enhance the model by allowing it too identify bacterial and viral pneumonia.
- Deploy the model in a small hospital. Observe the following:
  - Record time saved by radiologists
  - impact of false positives on clinician workflows.

CONTACT INFORMATION

- Email: Dhruvragunathan@gmail.com
- LinkedIn: https://www.linkedin.com/in/dhruv-ragunathan-908993b1/
- Github: https://github.com/dragunat2016/Pneumonia_Image_Detection