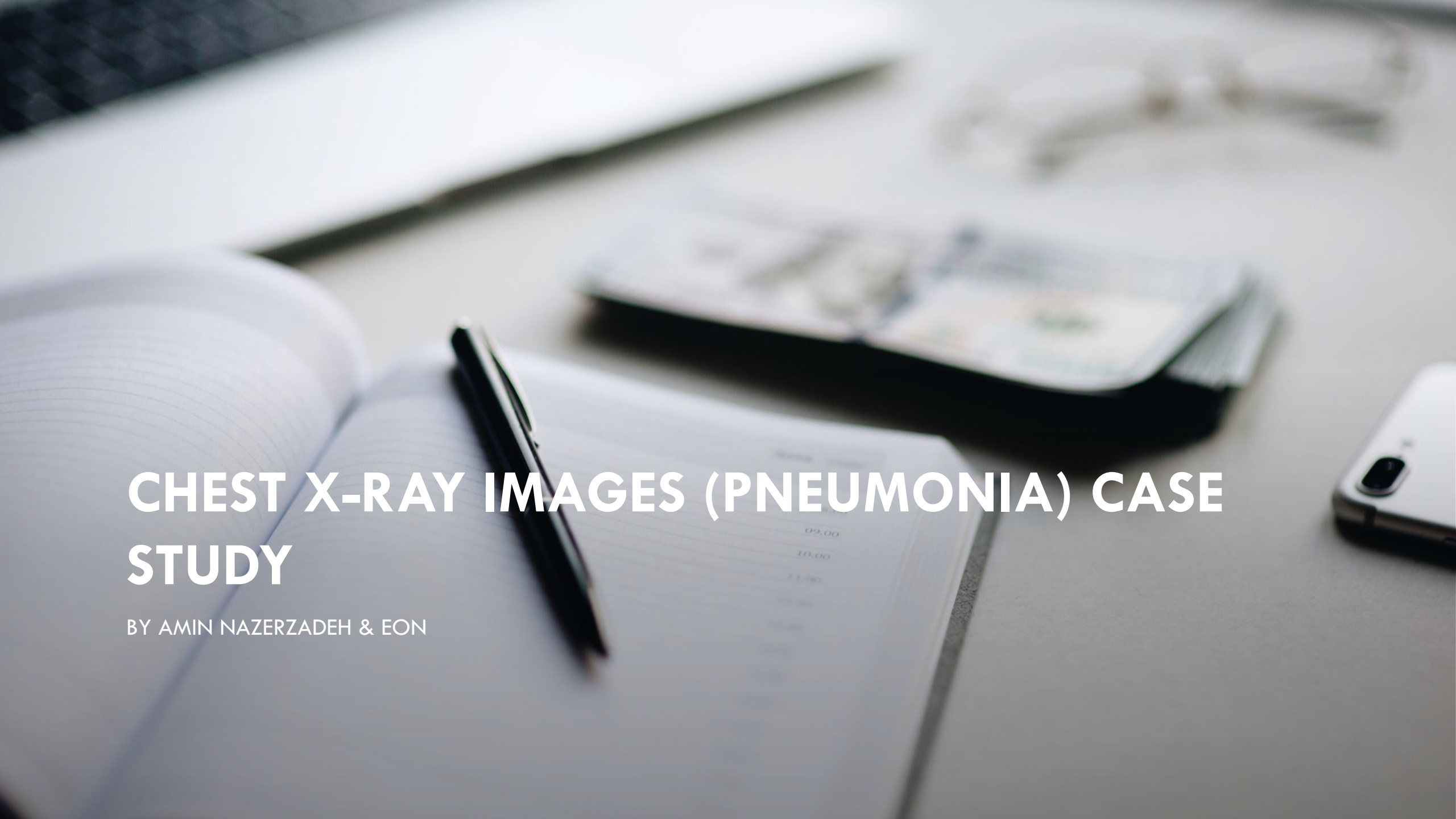


CHEST X-RAY IMAGES (PNEUMONIA) CASE STUDY

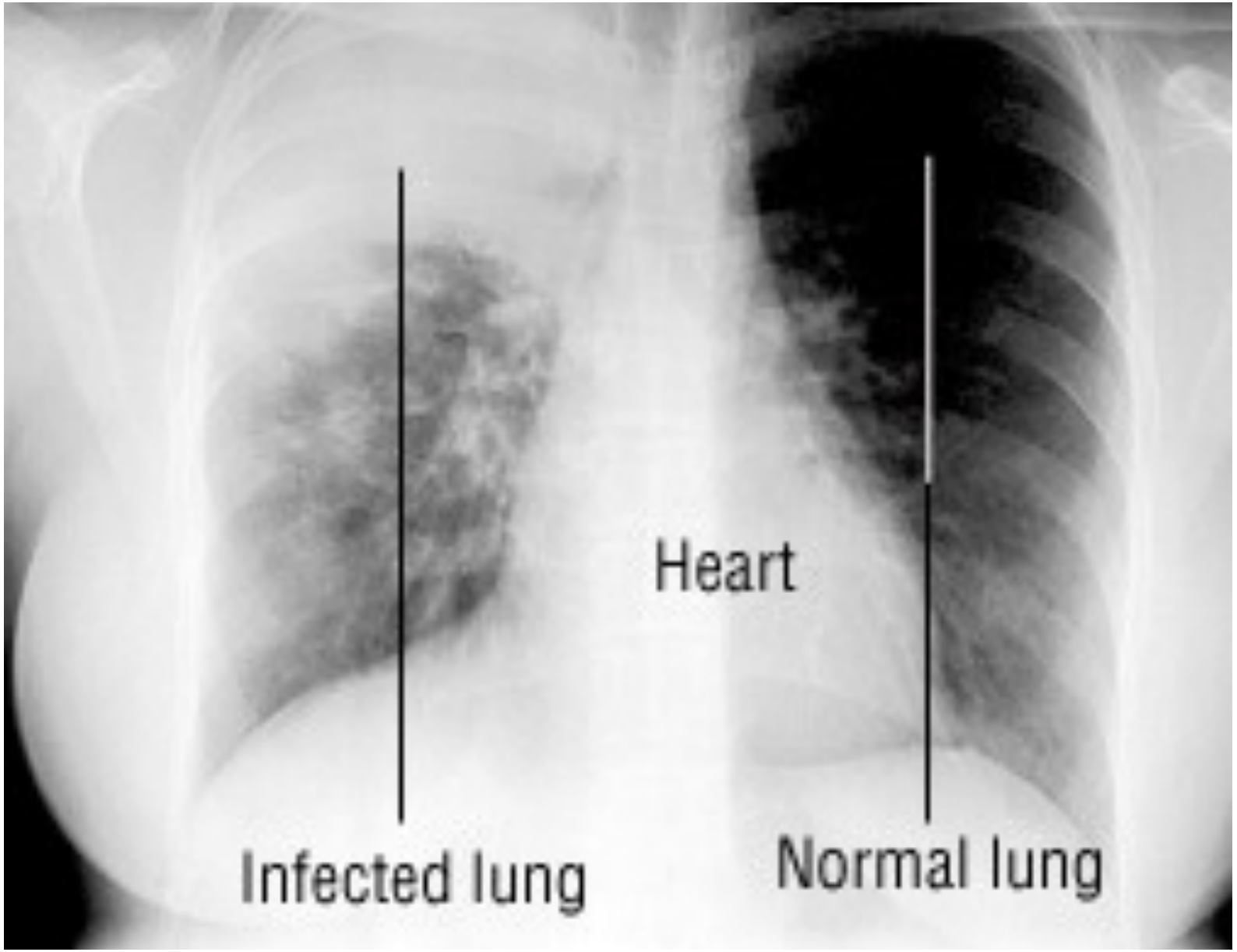
BY AMIN NAZERZADEH & EON

DATA

- The dataset comes from Kermany et al. on Mendeley
- Kagel website (https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia)
- There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal).
- All chest X-ray imaging was performed as part of patients' routine clinical care.
- All chest radiographs were initially screened for quality control by removing all low quality or unreadable scans
- The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system

WHAT IS PNEUMONIA

- Pneumonia is the filling of air vesicles in the lung with an inflamed fluid.
- Viruses, bacteria, and rarely fungal infections cause it.
- Pneumonia can be diagnosed by examining the X-Ray chest radiography by doctors.
- We will build a model that will do this instead of doctors

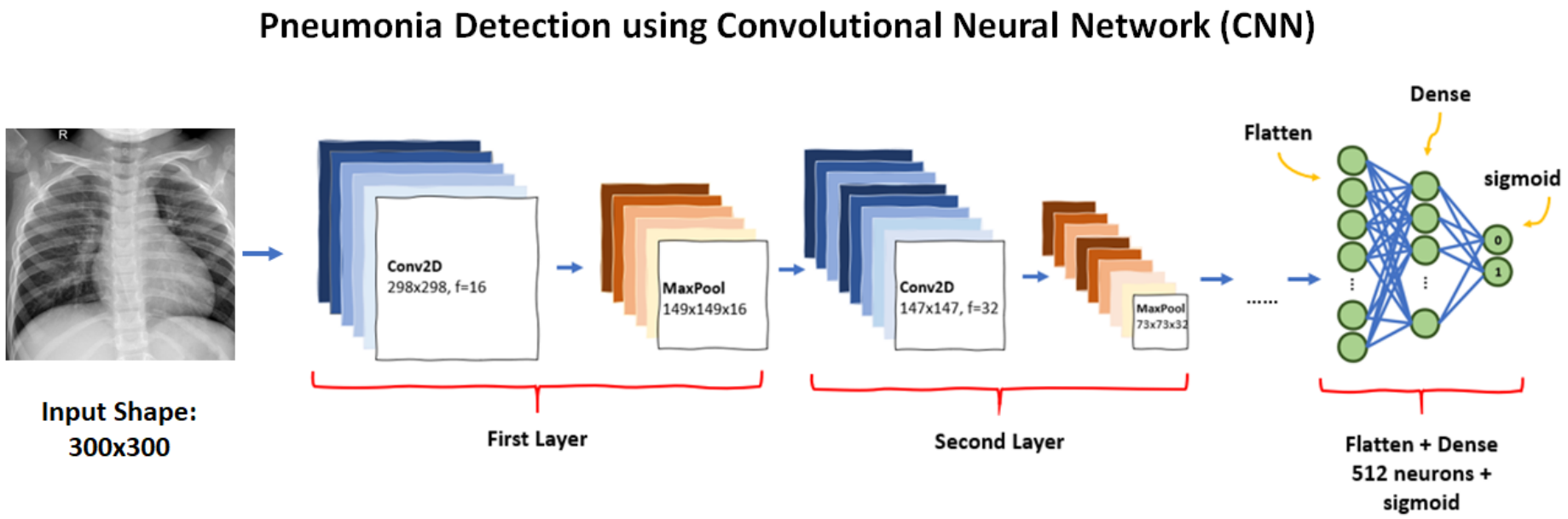


BUSINESS GOAL

Finding	Finding the best model that can predict if someone has Pneumonia or not
Building	Building different models and find the best model that has the best accuracy without overfitting

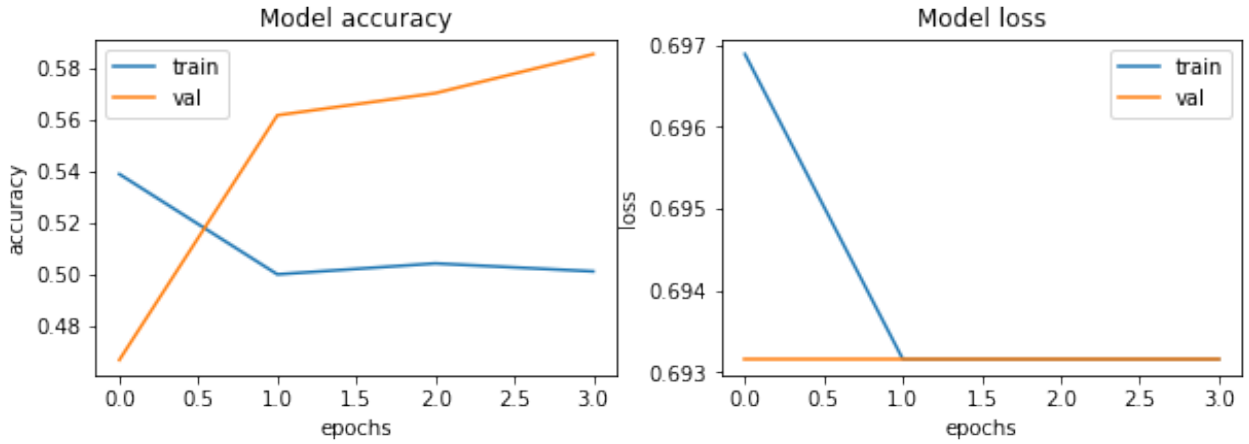
MODEL STRUCTURES

- Structure of CNN model will be like (Conv2D->relu -> MaxPool2D -> Dropout)x2 -> Flatten -> Dense -> Dropout -> Out
- I used five convolutional blocks comprised of convolutional layer and max-pooling.
- I have used dropouts to reduce over-fitting on the third Model.
- Activation function was Relu throughout except for the last layer where it was Sigmoid as this is a binary classification problem.
- I have used Adam as the optimizer
- I have also used EarlyStopping to prevent overfitting



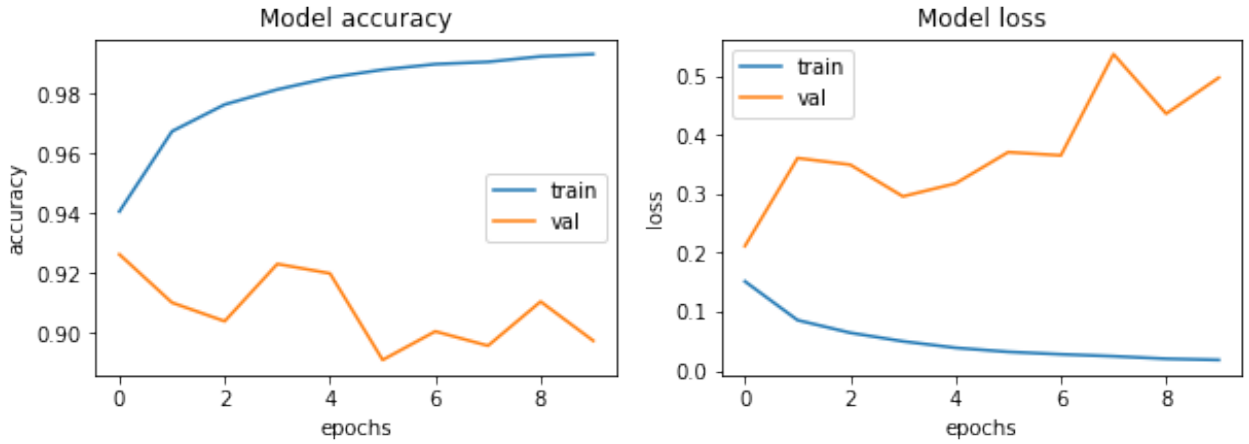
MODEL 1

- After running our first model we got an accuracy score of: 0.5011 and al_accuracy: 0.5855
- Unfortunately, our model has low accuracy scores
- Based on the graph was unable to learn the training dataset



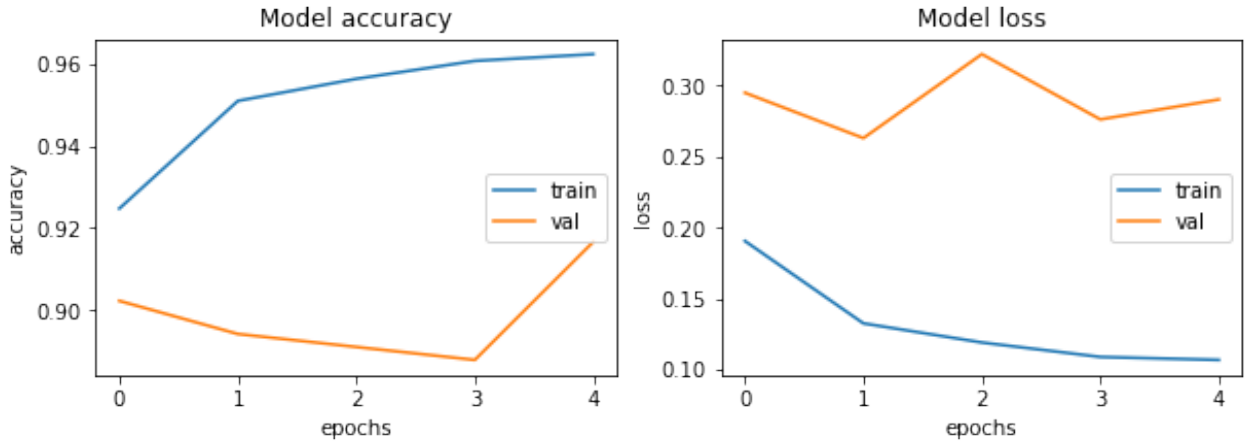
MODEL 2

- After running our second model we got an accuracy score of: 0.9933 and val_accuracy: 0.8974 which is better than the last Model
- You can clearly see from the preceding outputs that we still end up overfitting the model
- After the the 3 epochs



MODEL 3

- After running our third model we got an accuracy score of: 0.9624 and val_accuracy: 0.9166
- The val_accuracy was higher than the second model
- There is a decent gap between the model train and validation accuracy
- As we can see our model is overfitting on the training data



FURTHER ANALYSIS

- Adding more chest x ray images to the dataset is likely to help
- Adding more layers to improve our model and reduce overfitting
- Increasing the dropout might be helpful to reduce the overfitting
- It would be nice to see how our model perform by comparison with doctors
- Can our model do a better classification than the doctors?