# Pneumonia Detection

A DEEP LEARNING PROJECT BY VINAYAK MODGIL

Business Problem

- Pneumonia is an inflammatory disease caused by infection with viruses or bacteria in the lungs.

- Affects about 445 million (about 7% of the total population) and results in about 4.3 million deaths

- Can be lethal if not treated – Can be treated if diagnosed during the early onset

- Can we use Deep Learning to detect Pneumonia?

- Can we use Deep learning to making informed medical decisions on other diseases?
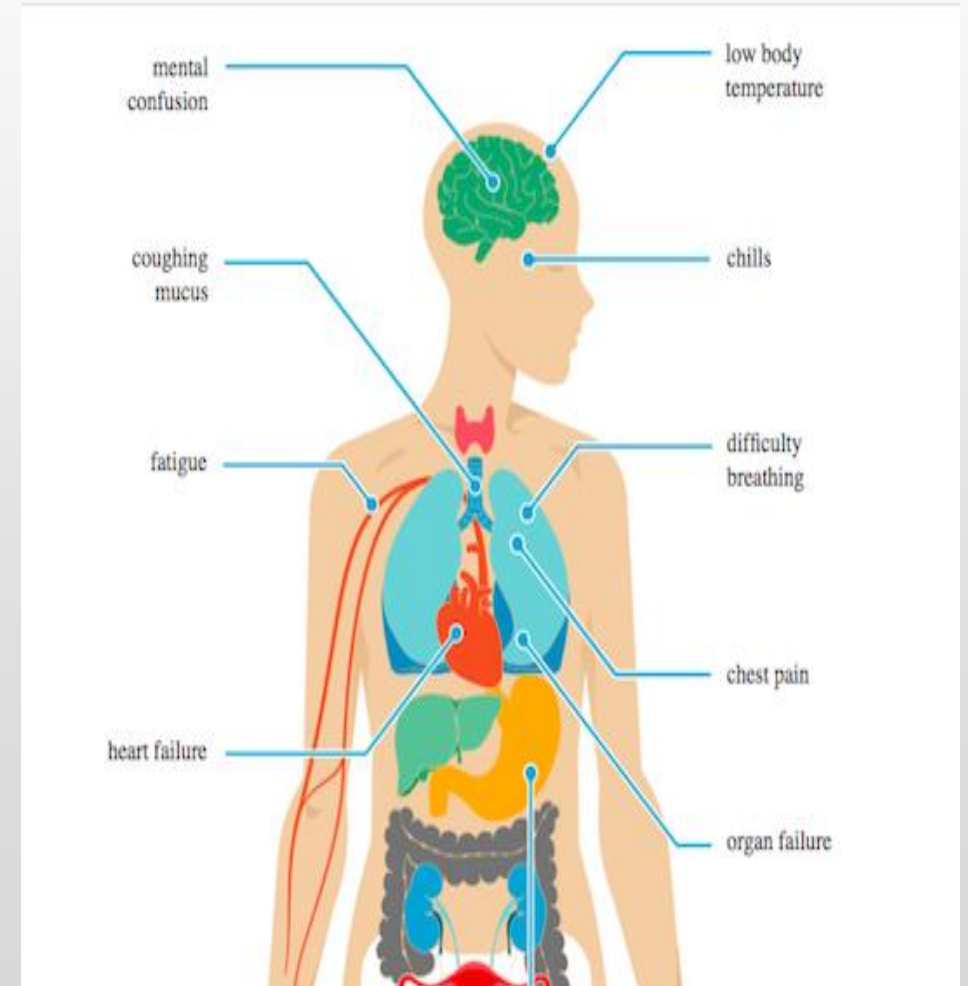
Diagnosing Pneumonia

- Doctor's Expertise based on evaluating the patient and symptoms.

- X-ray imaging (The red portion in the image) – A Radiologist or a Physician.
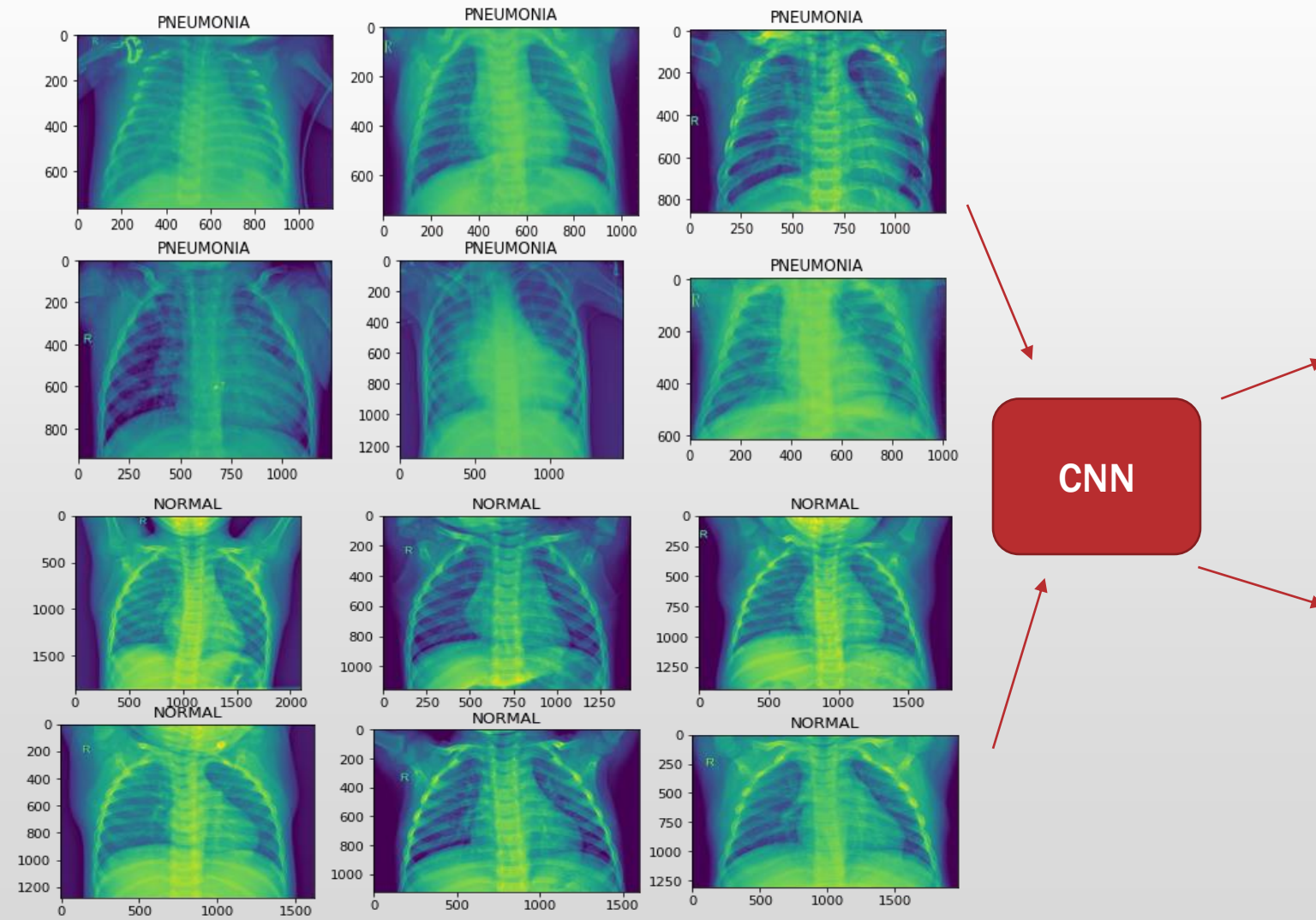
Convolutional Neural Networks –The Solution

"Convolutional neural networks (CNNs) are effective tools for image understanding. They have outperformed human experts in many image understanding tasks" (Sarvamangala & Kulkarni, 2021).
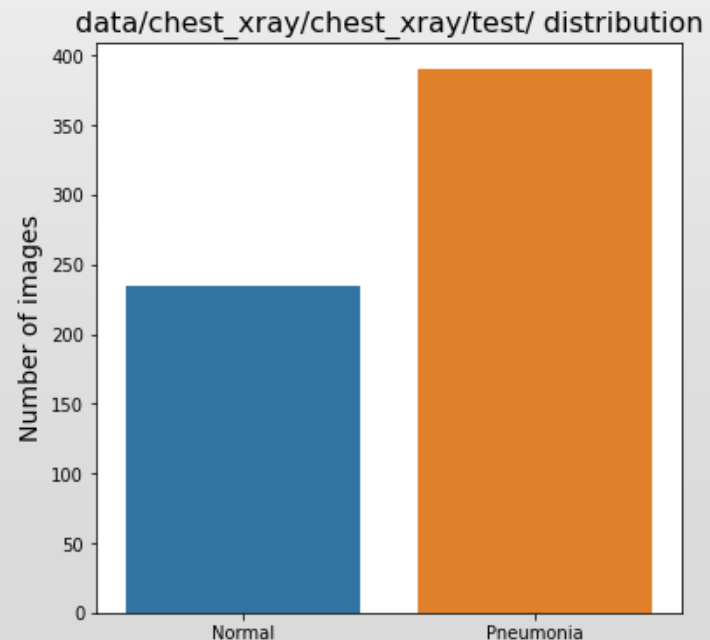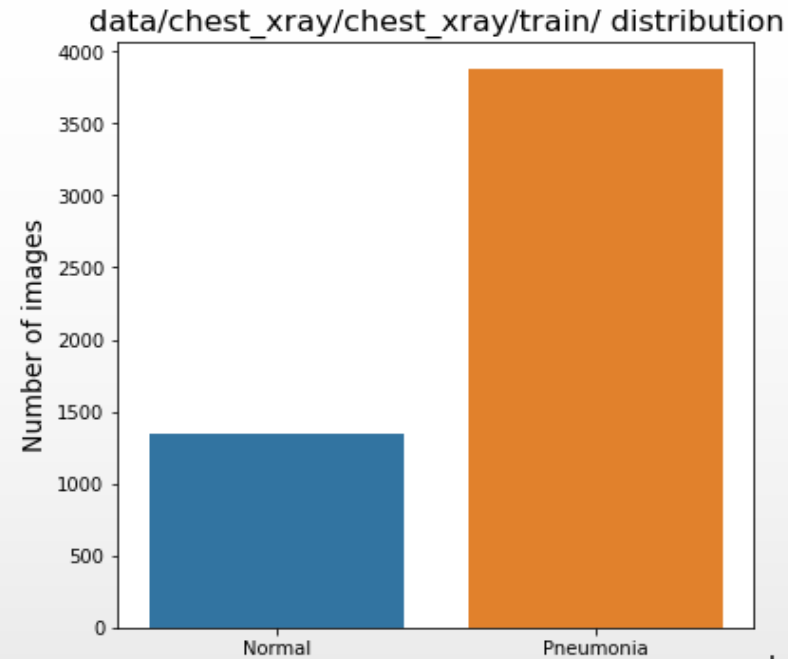

Pneumonia Detection using Convolutional Neural Network (CNN)
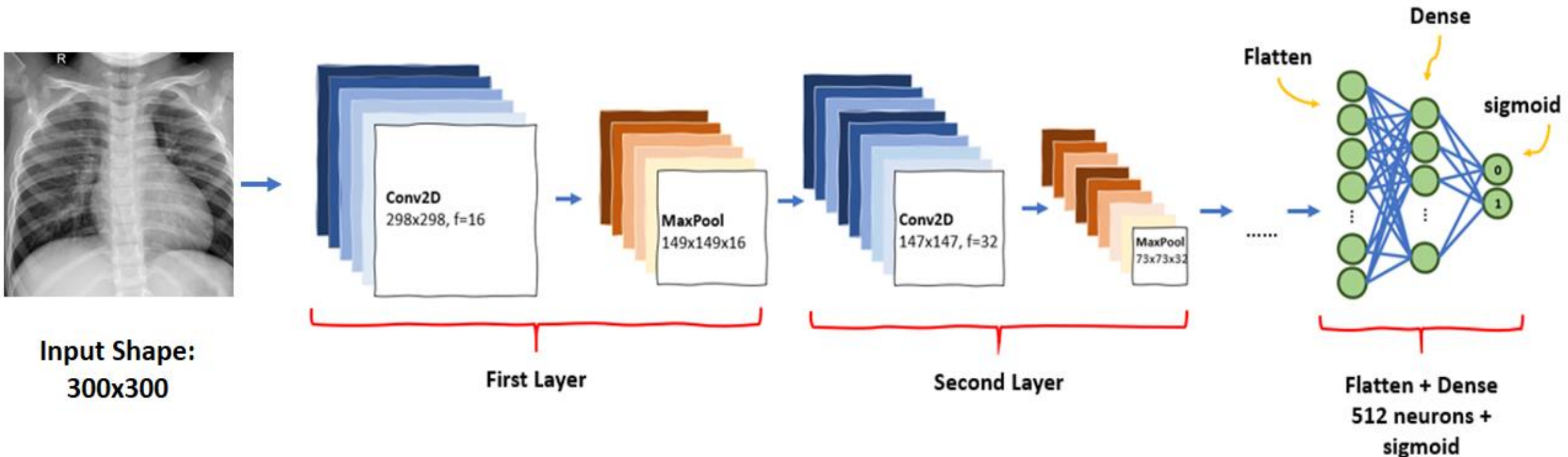
CNN Working

Layer (type)	Output Shape	Param #
conv2d_17 (Conv2D)	(None, 80, 80, 32)	896
max_pooling2d_17 (MaxPooling	(None, 40, 40, 32)	0
conv2d_18 (Conv2D)	(None, 40, 40, 63)	18207
max_pooling2d_18 (MaxPooling	(None, 20, 20, 63)	0
conv2d_19 (Conv2D)	(None, 20, 20, 128)	72704
max_pooling2d_19 (MaxPooling	(None, 10, 10, 128)	0
conv2d_20 (Conv2D)	(None, 10, 10, 128)	147584
max_pooling2d_20 (MaxPooling	(None, 5, 5, 128)	0
flatten_4 (Flatten)	(None, 3200)	0
dense_14 (Dense)	(None, 64)	204864
dense_15 (Dense)	(None, 128)	8320

data/chest_xray/chest_xray/train/ distribution

data/chest_xray/chest_xray/test/ distribution

Exploratory Data Analysis

- The data we trained the NN model on contain normal and pneumonia X-rays as the 2 class

- Test data balance: Normal = 37.4% Pneumonia=62.3%

- Train data balance: Normal = 25.7% Pneumonia=74.2%
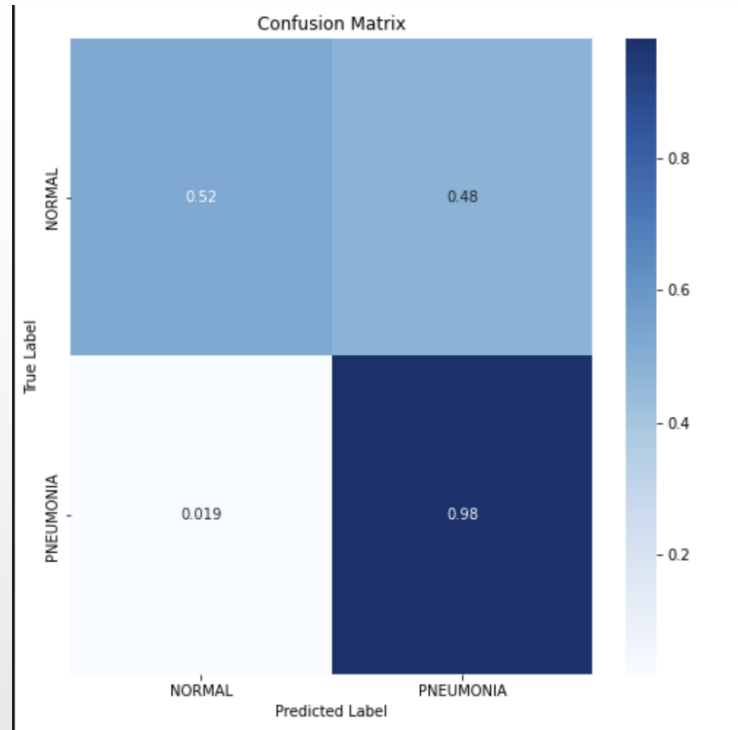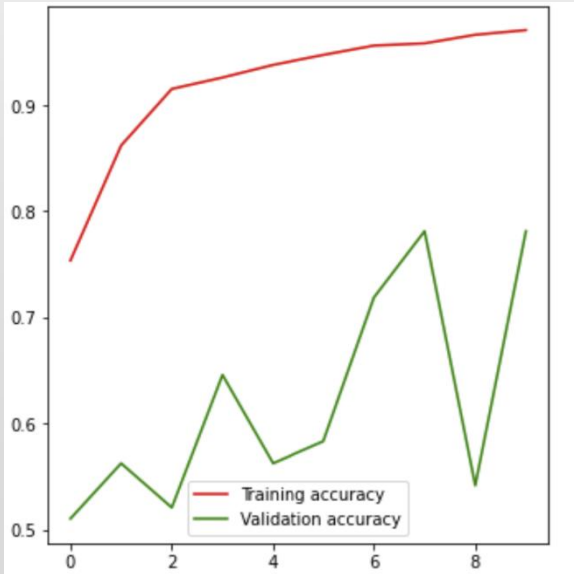
Modelling

- Baseline Model

- Model structure (2 CNN layers, 2 Max Pool layers and 2 Dense layers

- Training Accuracy = 97%, val accuracy = 78.12& (Overfitting)

- Confusion Matrix - The model does extremely well with detecting pneumonia but has limited functionality because of high false positive rate.

Best Model

Training and validation accuracy

- Training accuracy
- Validation accuracy
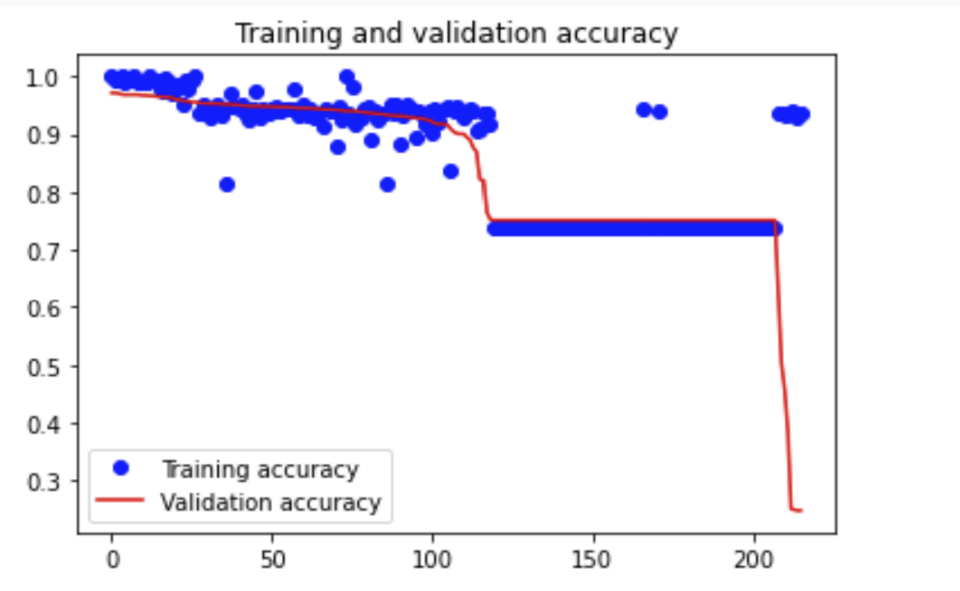
loss	accuracy	val_loss	val_accuracy
0.007022	0.999726	0.097082	0.971246

- Model Structure: 1 CNN layer, 1 MaxPool layer, 2 Dense layers, Dropout layer.

- Talos for the params

- Talos ran 216 models and the best model parameters were:

- Activation-fnc (CNN and dense): relu, Dropout:0.25, Nodes(CNN):128, Nodes(Dense), Optimizer:adam

- "A recent publication estimated that by 2020, there will be 200 times more medical information than what a single individual would be able to read in his/her entire life" (Chumbita et al., 2020)

- The goal of the model was achieved with over 95% validation accuracy

- However, the accuracy could be further increased by adding additional CNN and dense layers to the model structure, and then using the best parameters from Talos

Further Studies

Thank You