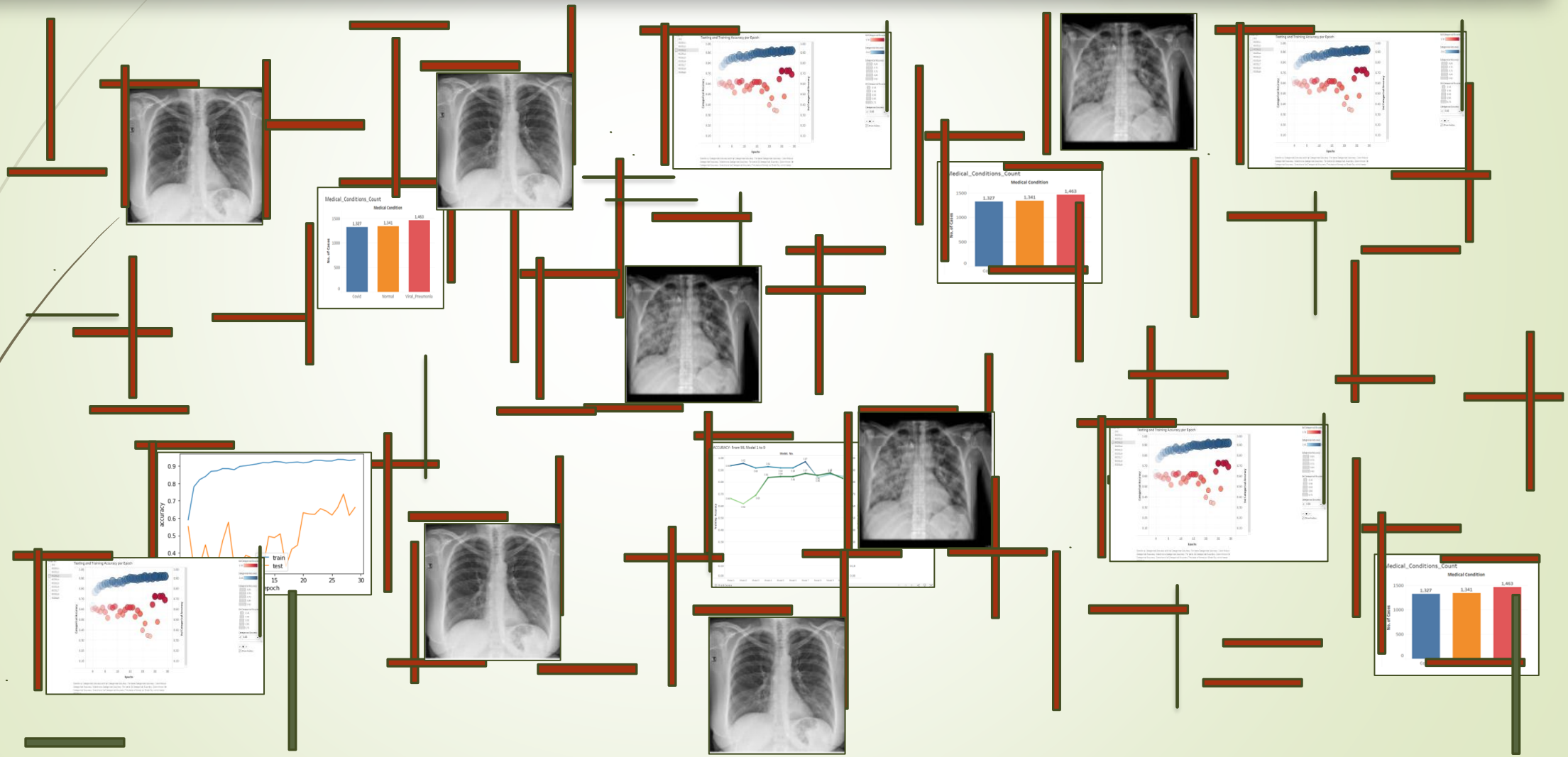
# COVID PREDICTION ML MODEL: VISUALIZATION

- STUTI BUDHWAR

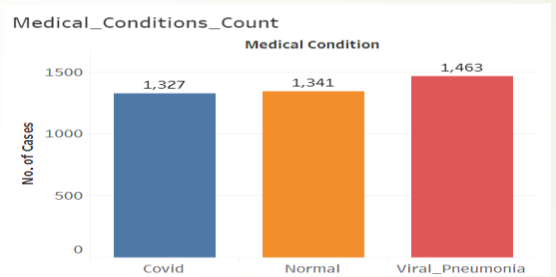


# PROJECT OVERVIEW:

## PURPOSE OF THE PROJECT:

- Arriving at a Machine Learning prediction model for various COVID-19, Viral Pneumonia or Normal condition from frontal chest X-ray images of the patients. This would give physicians an edge and allow them to act with more confidence while they wait for the analysis of a radiologist by having a digital second opinion confirm their assessment of a patient's condition. Also, these tools can provide quantitative scores to consider and use in studies.

## DATASET:




- **Type of Data :** Frontal Chest X-ray Images
- **Size of Data:** 4013 Images
- **Source:**
  - https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
  - Plus 184 COVID-19 image added from https://www.kaggle.com/bachrr/covid-chest-xray
- **ML Models applied :** Basic, VGG_16, RESNET

## MACHINE LEARNING STRATEGY:

- We started with the basic model that was just a little overfitted with excellent training accuracy. The next step was to slightly modify the model each time, so it can generalize better.
- If we got a model with a better testing accuracy than its training accuracy, it would indicate that the model is not learning the most important features . And since it's just that the validation dataset, it is easier to predict. This is feature common with data augmentation.
- Base model was a very simple model to compare against. Then we tried different pretrained models like VGG-16 and Resnet-50 and added layers at the end to suit our needs using trial and error.

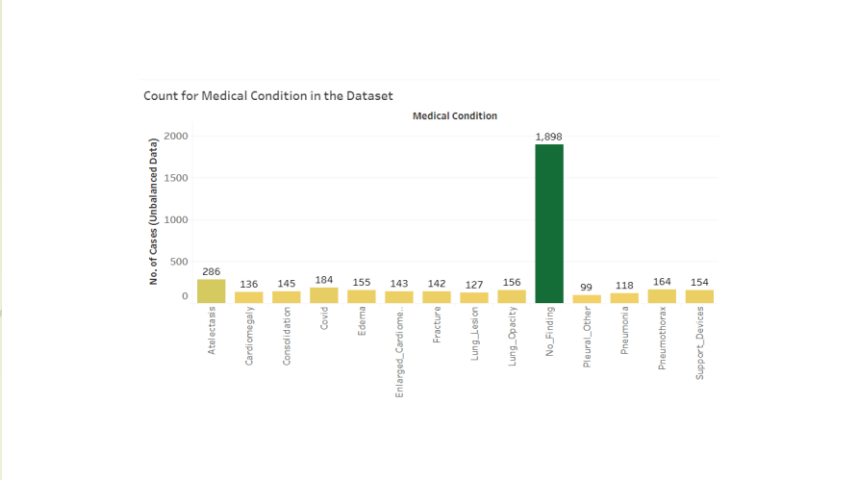
# EVOLVING OF TOOLS AND TECHNOLOGIES:

Our journey so far has been quite interesting and challenging. We, as a team have evolved to following technologies/ tools based on our experience so far.

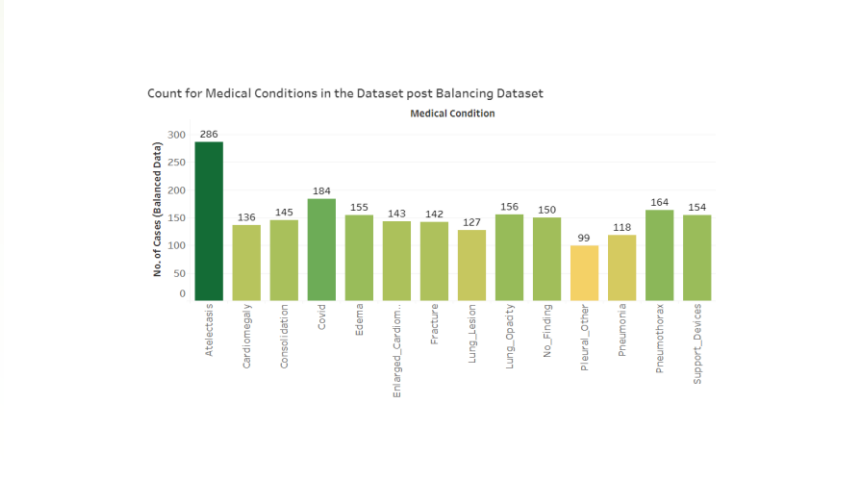
- **DATASET:**
  - **BEFORE:**
    - https://www.dropbox.com/s/09b5nutjxotmftm/data_upload_v2.zip?dl=0
    - Dataset with 5000 chest X-ray images to predict one of the 14 medical conditions.
  - **NOW:**
    - https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
    - Plus 184 COVID-19 image added from https://www.kaggle.com/bachrr/covid-chest-xray
    - Dataset with 3829 + 184 chest X-ray images dataset to predict COVID-19, Viral Pneumonia or No finding/Normal.
  - **REASON:**
    - The number of chest X-ray images after balancing the data turned out too less for 14 medical conditions that they had to be trained for. This created a model that did not perform well at all.
    - We had a limitation of looking for a dataset with more COVID-19 infected chest X-rays images.
    - The imbalanced data however performed better giving an accuracy of about 50%, which was due to almost 50% of the images belonging to “No finding” category. This model would have been a biased ML model.

# DATASET BEFORE:

## COUNT OF 14 MEDICAL CONDITIONS: BEFORE BALANCING DATA:

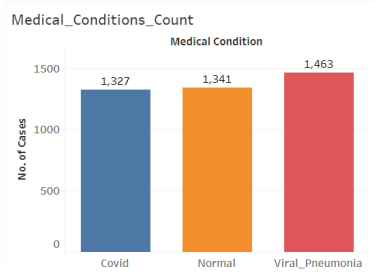



## COUNT OF 14 MEDICAL CONDITIONS: AFTER BALANCING DATA:

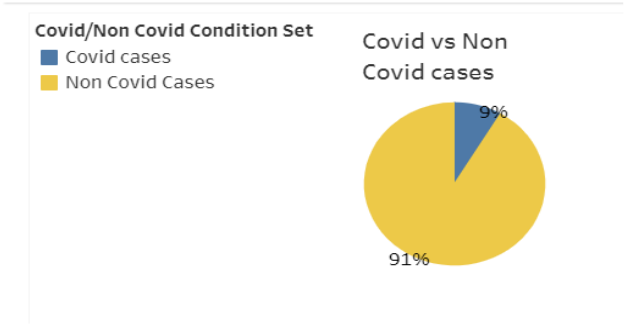



# DATASET NOW:

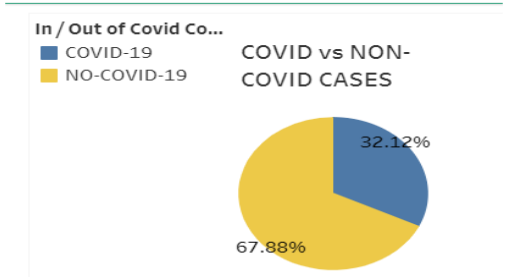
## COUNT OF 2 MEDICAL CONDITIONS: ALREADY BALANCED DATA



## DATASET BEFORE:
### COVID Vs NON COVID COUNT:



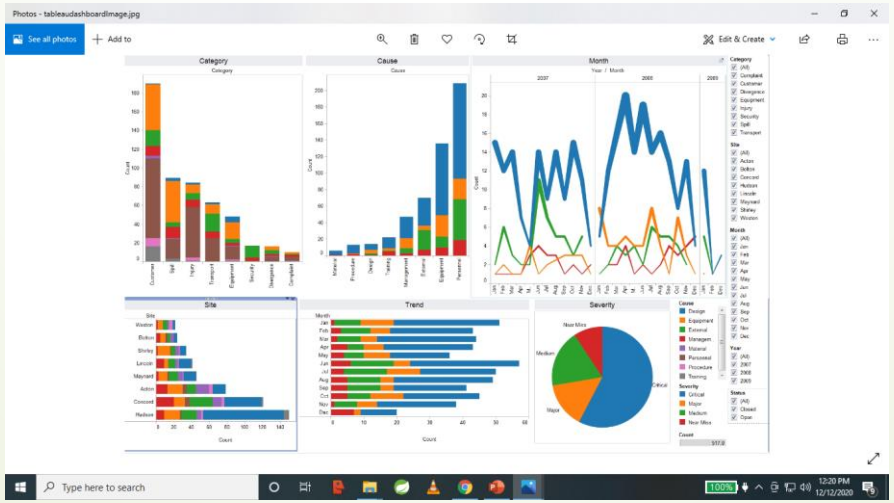

## DATASET NOW:
### COVID Vs NON COVID COUNT:



- **DATABASE:**
  - **OPTION 1:**
    - We had earlier planned that the images would be stored in Amazon Webservices S3 bucket and the labels for the same in AWS RDB with PostgreSQL flavor. We would integrate Flask to display the data.
  - **OPTION 2:**
    - Then we started using Amazon SageMaker to save our image data that integrates with Jupyter Notebook to build our ML models. We would integrate Flask to display the data.
  - **OPTION 3:**
    - Now we are using local Jupiter Notebook and Google Collab as IDE and storing all data on local computers.
  - **OBSERVATIONS:**
    - The free tier of Amazon Webservices S3 bucket was getting exhausted.
    - Amazon SageMaker is a fully-managed service that enables us to quick and easily build, train, and deploy machine learning models at any scale. Amazon SageMaker includes modules that can be used together or independently to build, train, and deploy your machine learning models. But runs too slow.
    - Google Collab and Local Jupiter Notebook are both working good so far with our image dataset.
    - All image data is stored in our local machines.

# VISUALIZATION TOOLS USED TO CREATE FINAL DASHBOARD:

- We predominantly used **Tableau to create the final dashboard along with some other visualizations as required.**
- Tableau is one of the most popular visualization tool in todays time due to being proficient in analyzing large data sets using statistical & analytical knowledge to understand data trends, generate business insights and forecast outcomes.
- We integrated **Tableau** for developing effective storytelling and fully functional, intuitive and interactive dashboards.
- It was published on the **public tableau server**.



# JOURNEY OF VISUALIZATIONS PLANNED:

## FROM BEGINNING TO THE FINAL DASHBOARD:

- **Visualizing original dataset**
  - Total chest X-ray images per medical condition
  - COVID-19 vs Non COVID-19 cases
  - Visualization Tool: **Tableau**
- **Analysis after every ML model is applied:** (*ML strategy coming ahead subsequently)
  - ML Model* vs model accuracy for the chest X-ray images having one of the 2 medical conditions* or images with no finding/Normal. (using confusion matrix)
  - ML Model* vs loss for the chest X-ray images having one of the for the 2 medical conditions* or images with no finding.
  - Animated visualization showing the performance change of the ML model with every little feature change done for the betterment of model in chronological order to show how each model evolved over different models/feature change of each model.
  - Create an animated plot to show performance of the model over epochs.
  - Visualization Tool: **Tableau, matplotlib library**

- **Analysis of accuracy and loss for pretrained models like VGG16, VGG19, ResNet18, ResNet50, SqueezeNet, DenseNet-121 vs a regular model.**
- **Final performance analysis of the model**
  - Plot all models vs accuracy to conclude most optimal model.
  - Plot all models vs loss to conclude most optimal model.
  - Animated visualization showing the performance change in models in chronological order to show how performance evolved over different models.
  - Final 3 shortlisted models for visualizing the performance in terms of loss and accuracy.
  - Best performing model  for visualizing the performance in terms of loss and accuracy.
  - Final prediction based on the best ML model, showing predicted vs actual value (class it belongs to)
  - Visualization Tool: **Tableau**

# VISUALIZATIONS PLANNED FOR THE FINAL DASHBOARD:

- Plot all models vs accuracy to conclude most optimal model.
- Plot all models vs loss to conclude most optimal model.
- I am stretching myself to work on an animated visualization showing the performance change of the ML model with every little feature change done for the betterment of model in chronological order to show how each model evolved over different models/feature change of each model.
- Also, have a final animated visualization showing performance change during all models in chronological order to show how performance evolved over different models. I am not sure if it will work, but it would be extremely useful if I am able to find a pattern using the above 2 visualizations.

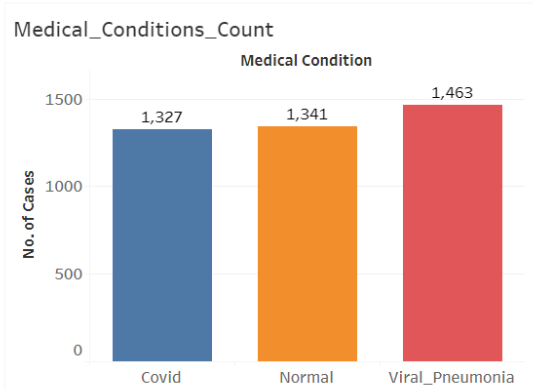
* ML Model: We will try different pertained machine learning models including VGG16, ResNet50 and add layers at the end to adjust for our needs.

* Medical conditions: COVID-19, Viral Pneumonia and no finding/Normal.

# DESCRIPTION OF INTERACTICE ELEMENTS

- **Tooltips and viz in tooltips-** This was used in all visualizations to show main features of the point in the pop up box as required.
- **Highlighting & highlighting actions-** This was used to highlight regular model as against the pretrained models in visualizations as required.
- **Filters** - This would be used in all visualizations to filters different parameters like Model no., loss, accuracy etc.
- **Groups**- This was used in a visualization to group certain number of Epochs to which the machine learning model was being run.
- **Animation**- This would be used to show performance change during feature enhancement of 1 model and eventually evolving of the best model from all models in chronological order. (Used in this segment)

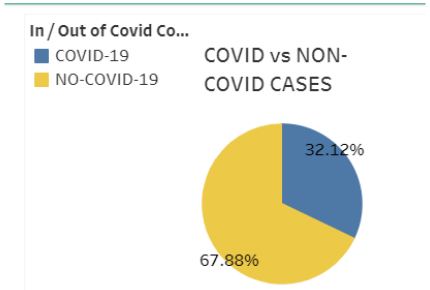
# BASIC DATASET VISUALIZATION-1:
## COUNT FOR MEDICAL CONDITIONS




* Tableau public link:

https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_New_Dataset_Medical_Conditions_Count/Medical_Conditions_Count?publish=yes

# BASIC DATASET VISUALIZATION-2: COUNT OF COVID VS NON COVID CASES




* Tableau public link:

https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_New_Dataset_Covid_vs_NonCovid_Count/CovidVsNonCovid?publish=yes

# ARCHITECTURE/STRATEGY FOR MACHINE LEARNING MODELS AND VISUALIZATION:

- **STEP 1: Image classification model creation**
  - Image classification for multiple classification (COVID-19, Viral Pneumonia, No finding/Normal) dataset.
  - TensorFlow API 'ImageDataGenerator' used to pre process/ rescale the images.
- **STEP 2: CNN model creation**
  - Different pretrained models including VGG16, ResNet50 are being tried.
  - Adding/applying hidden layers, neurons, dense layers, activation for hidden and dense layers etc by trial and error to improve the model.
- **STEP 3: Compilation of the mode**
- **STEP 4: Fitting and training the model**
- **STEP 5: Visualization of the model performance**
  - The accuracy and error plotted to check the model performance.

# MACHINE LEARNING MODEL APPROACH:

We tried the following pertained machine learning models and added layers at the end to adjust for our needs:

- Basic Model
- VGG16
- ResNet50

We started with the basic model that was just a little overfitted with excellent training accuracy. The next step was to slightly keep modifying the model so it could generalize better.
But if we got a model with a better val_accuracy than its training accuracy, it would be a very bad indicator implying that the model was not learning the most important features. It's just that the validation dataset is easier to predict. This is common issue with data augmentation.

Base model was a very simple model to compare against. Then we just try different pretrained models and added layers as per our requirements.

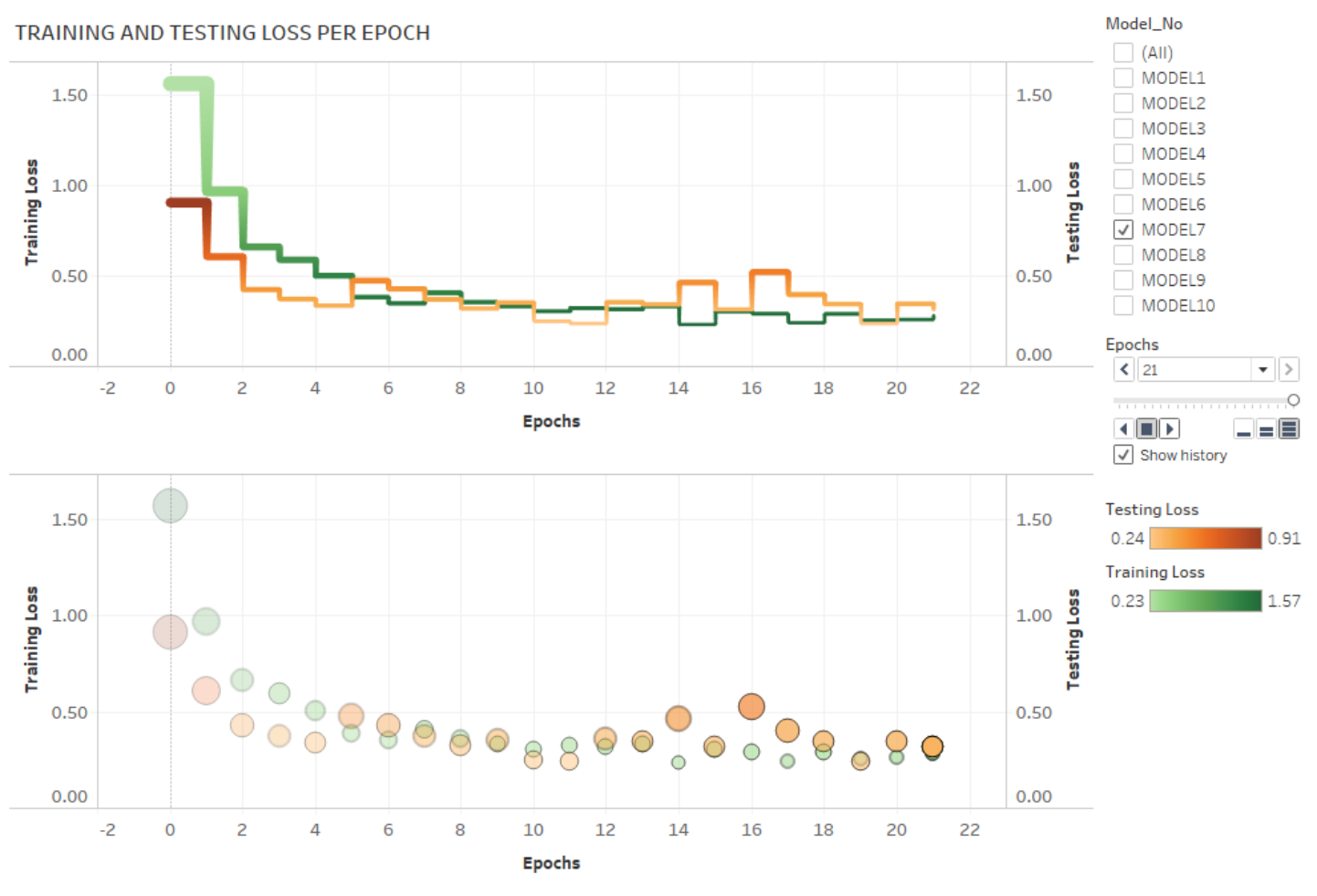
Note:

- ML Model details are provided in Annexure_Model_Details.md
- Some results of the models would be presented differently since 2 people on the team had been working on it. Since they are basic models, it didn't seem worth to run those models again just for the sake of presentation or printing results differently as the models were going to evolve in up-coming cases.

# MODEL SUMMARY :

| Model. No. | Basic Model Name | No of Epochs | File_Name | Training Loss | Training Accuracy | Testing Loss | Testing Accuracy | Reference Git Branch for Code |
|---|---|---|---|---|---|---|---|---|
| Model 1 | Basic Model-A | 30 | Model1_history | 0.1909 | 0.936 | 1.6882 | 0.6615 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Bijan_Samimi_ml/COVID_ML_3classes.ipynb |
| Model 2 | Basic Model-B | 30 | Model2_history | 0.1354 | 0.9548 | 2.6053 | 0.6181 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Bijan_Samimi_ml/COVID_ML_3classes.ipynb |
| Model 3 | Basic Model-C | 30 | Model3_history | 0.2423 | 0.9172 | 0.9551 | 0.6892 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Bijan_Samimi_ml/COVID_ML_3classes.ipynb |
| Model 4 | Basic Model-D | 50 | Model4_history | 0.198 | 0.9278 | 0.385 | 0.8368 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Gabriel_Cuchacovich/GoogleColab_COVID_ML.ipynb |
| Model 5 | VGG-16-A | 100 | Model5_history | 0.194 | 0.9378 | 0.3526 | 0.8595 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Bijan_Samimi_ml/COVID_ML_3classes.ipynb |
| Model 6 | VGG-16-B | 200 | Model6_history | 0.1792 | 0.9382 | 0.3636 | 0.8663 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Bijan_Samimi_ml/COVID_ML_3classes.ipynb |
| Model 7 | VGG-16-C | 50 | Model7_history | 0.2848 | 0.8965 | 0.3187 | 0.8736 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Bijan_Samimi_ml/optimization_gcl_vgg16_final.ipynb |
| Model 8 | Resnet50-A | 50 | Model8_history | 0.3309 | 0.8718 | 0.3371 | 0.8748 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Bijan_Samimi_ml/covid_resnet.ipynb |
| Model 9 | Resnet50-B | 50 | Model9_history | 0.3603 | 0.8653 | 0.3061 | 0.8748 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Gabriel_Cuchacovich/GoogleColab_COVID_ML.ipynb |
| Model 10 | Resnet50 -C | 100 | Model10_history | 0.3564 | 0.8614 | 0.3382 | 0.8796 | https://github.com/Coachnmomof3/UCB_COVID_Prediction_Model/blob/Gabriel_Cuchacovich/GoogleColab_COVID_ML.ipynb |

# TABLEAU REPRESENATION OF LOSS FOR ALL MODELS:

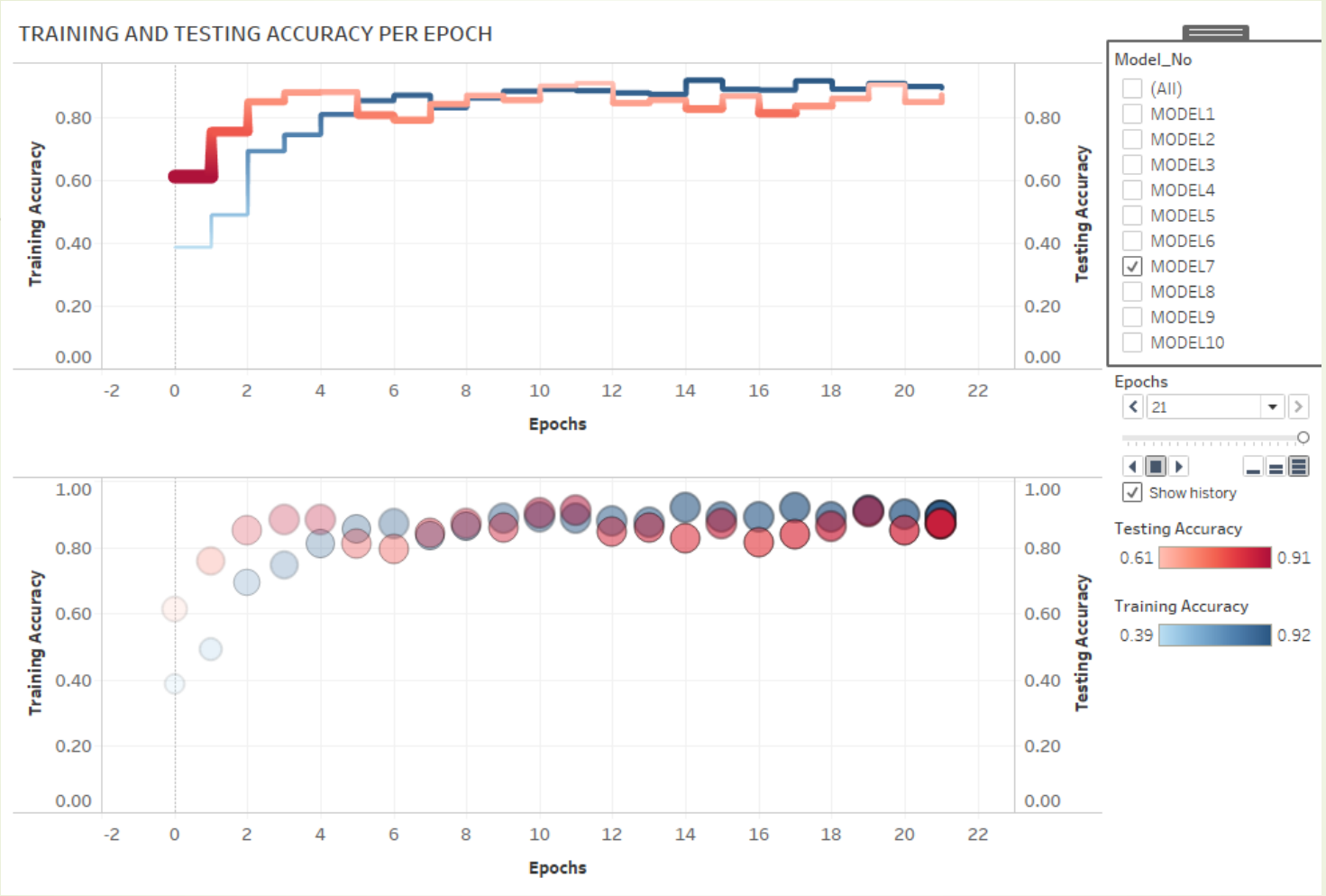



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENATION OF ACCURACY FOR ALL MODELS:




Note:

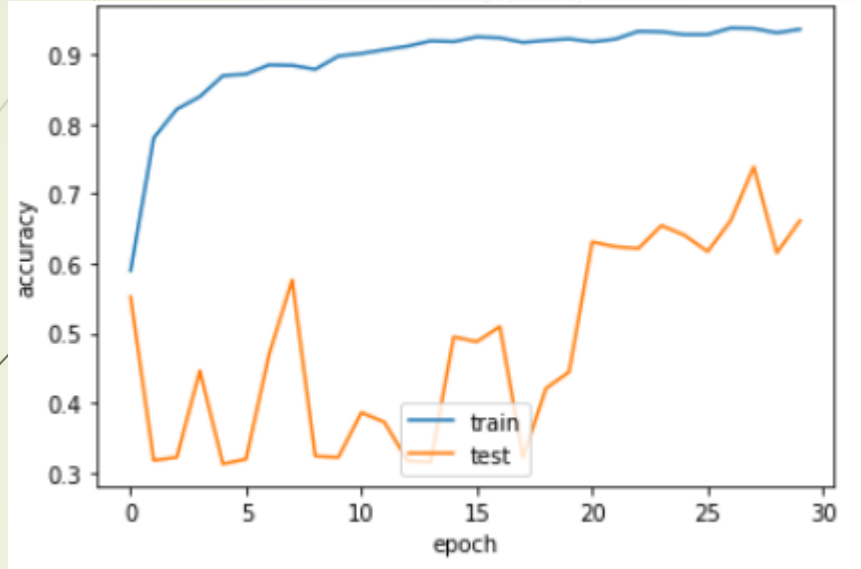
- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

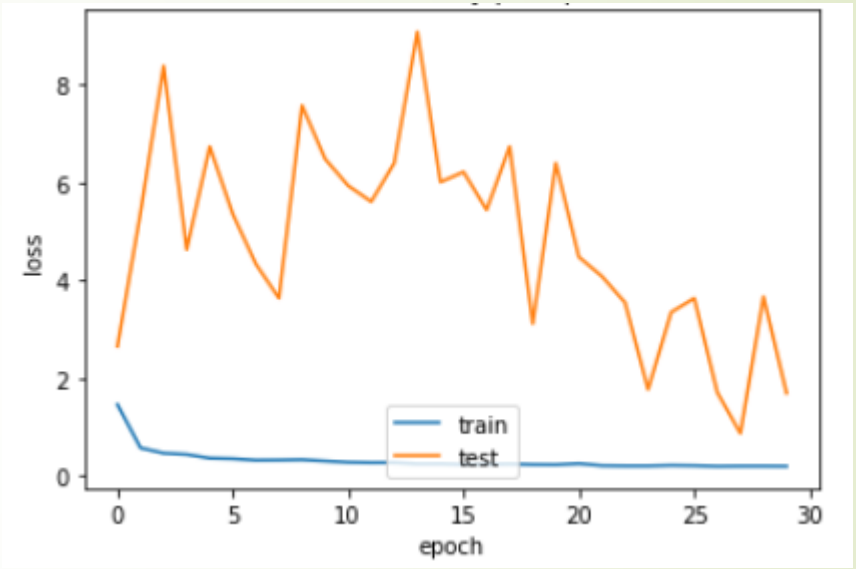
# ANALYSIS AND VISUALIZATION FOR MODEL-1: Accuracy and Loss

## MODEL-1: Basic Model-A

### Model accuracy per Epoch:




### Model loss per Epoch:



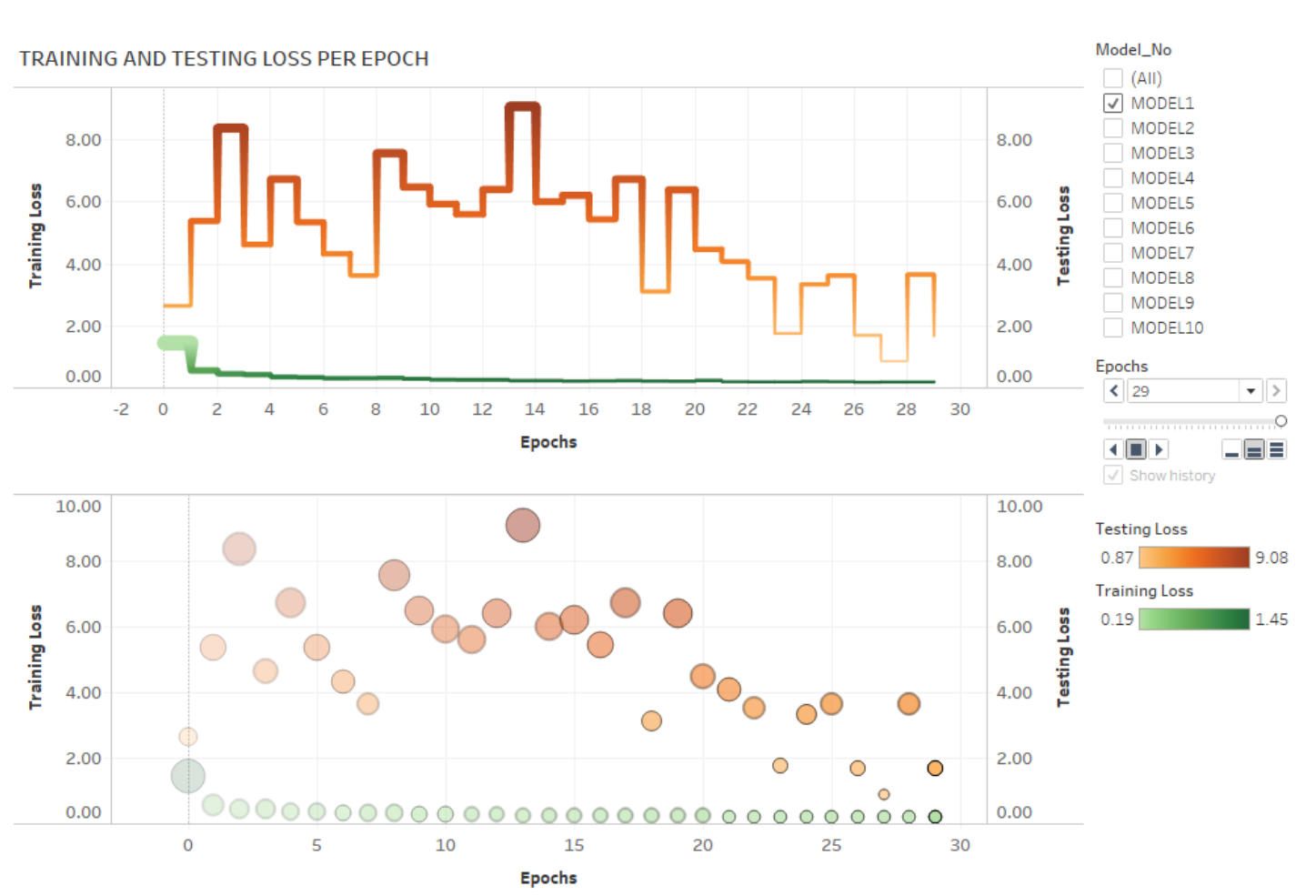

### ANALYSIS:

- The training curve for accuracy was way above the testing curve implying that the model is overfitted with excellent training accuracy just the way we wanted to start our first basic model.
- The next step would be to slightly modify the model so it can generalize better.

Note:

- ML Model details are provided in Annexure_Model_Details.md

# TABLEAU REPRESENTATION- Loss Trail per Epoch

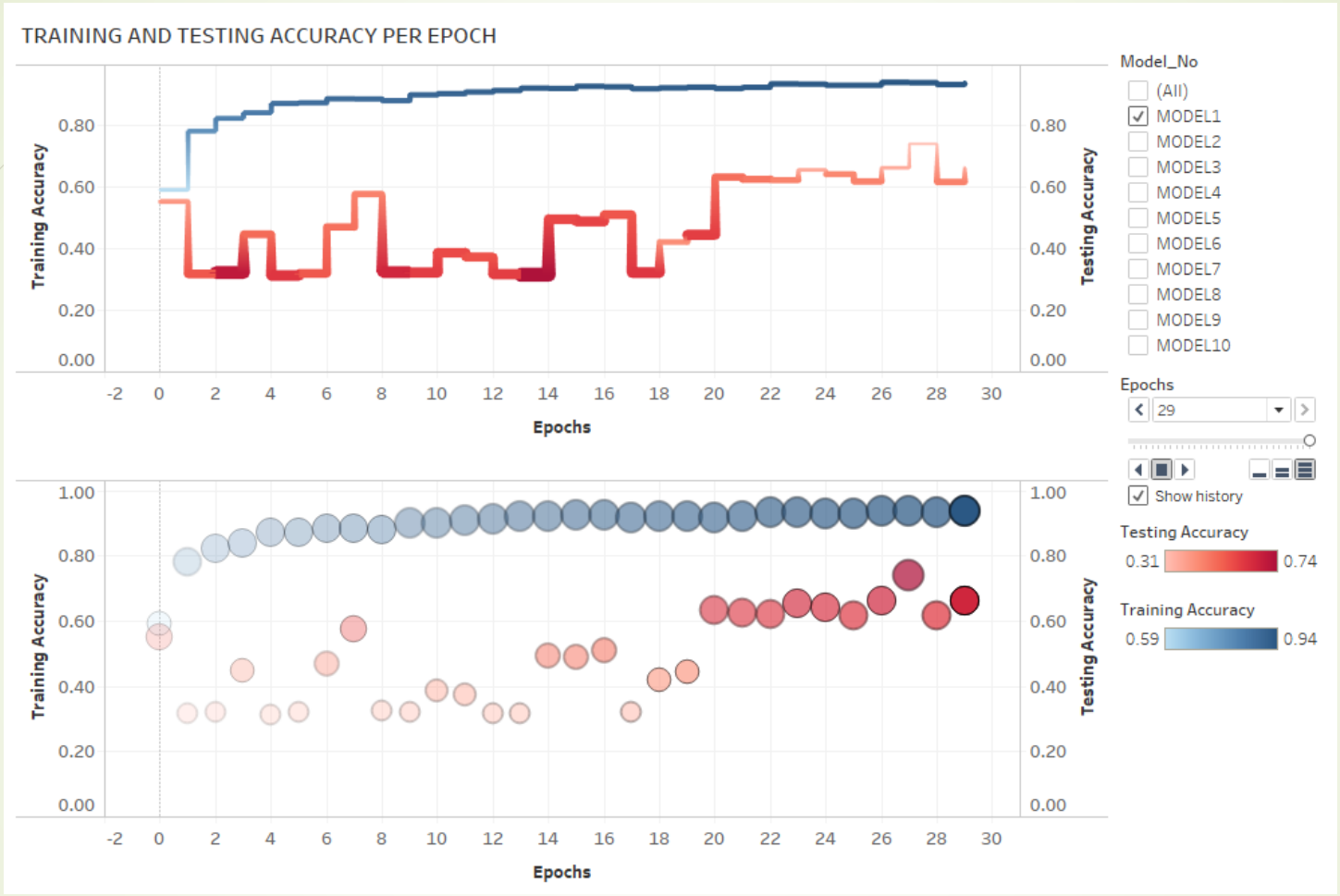



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION- Accuracy Trail per Epoch




Note:

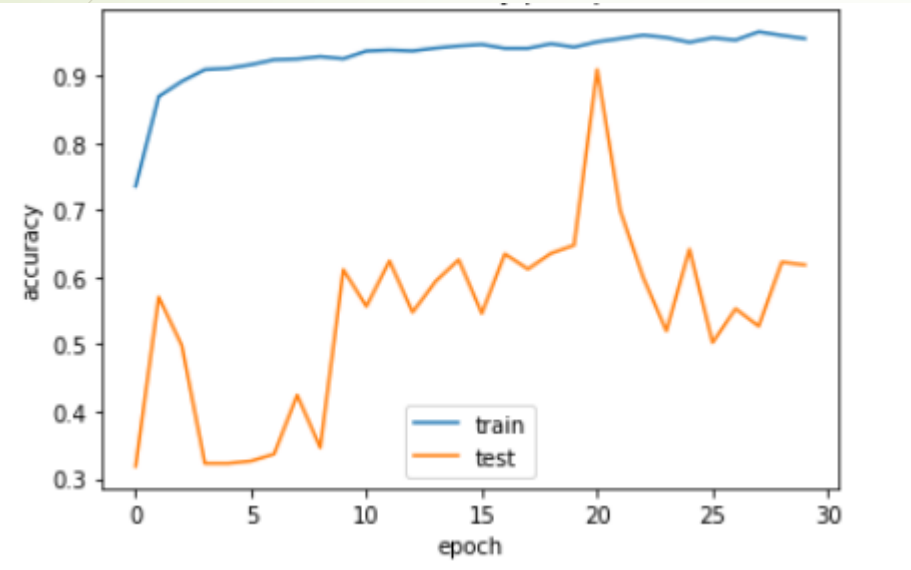
- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

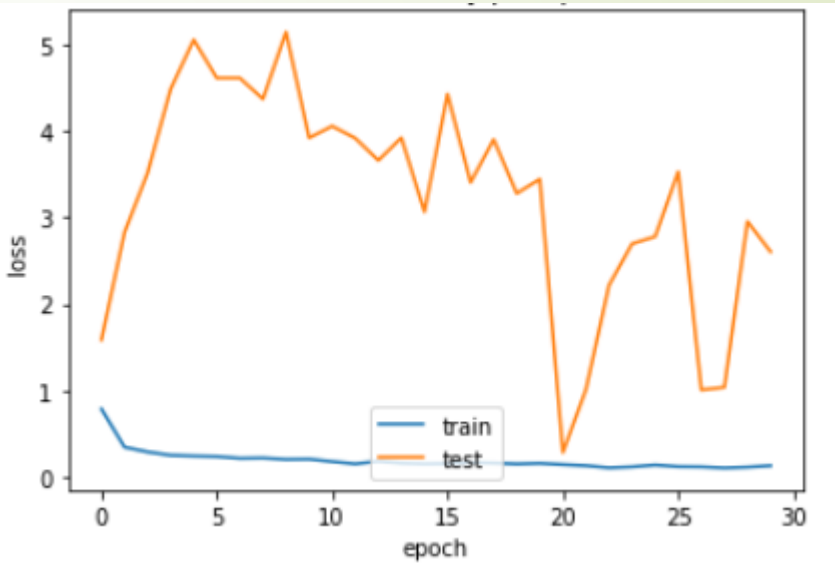
# ANALYSIS AND VISUALIZATION FOR Model-2: Accuracy and Loss

## MODEL-2: Basic Model-B

### Model accuracy per Epoch:



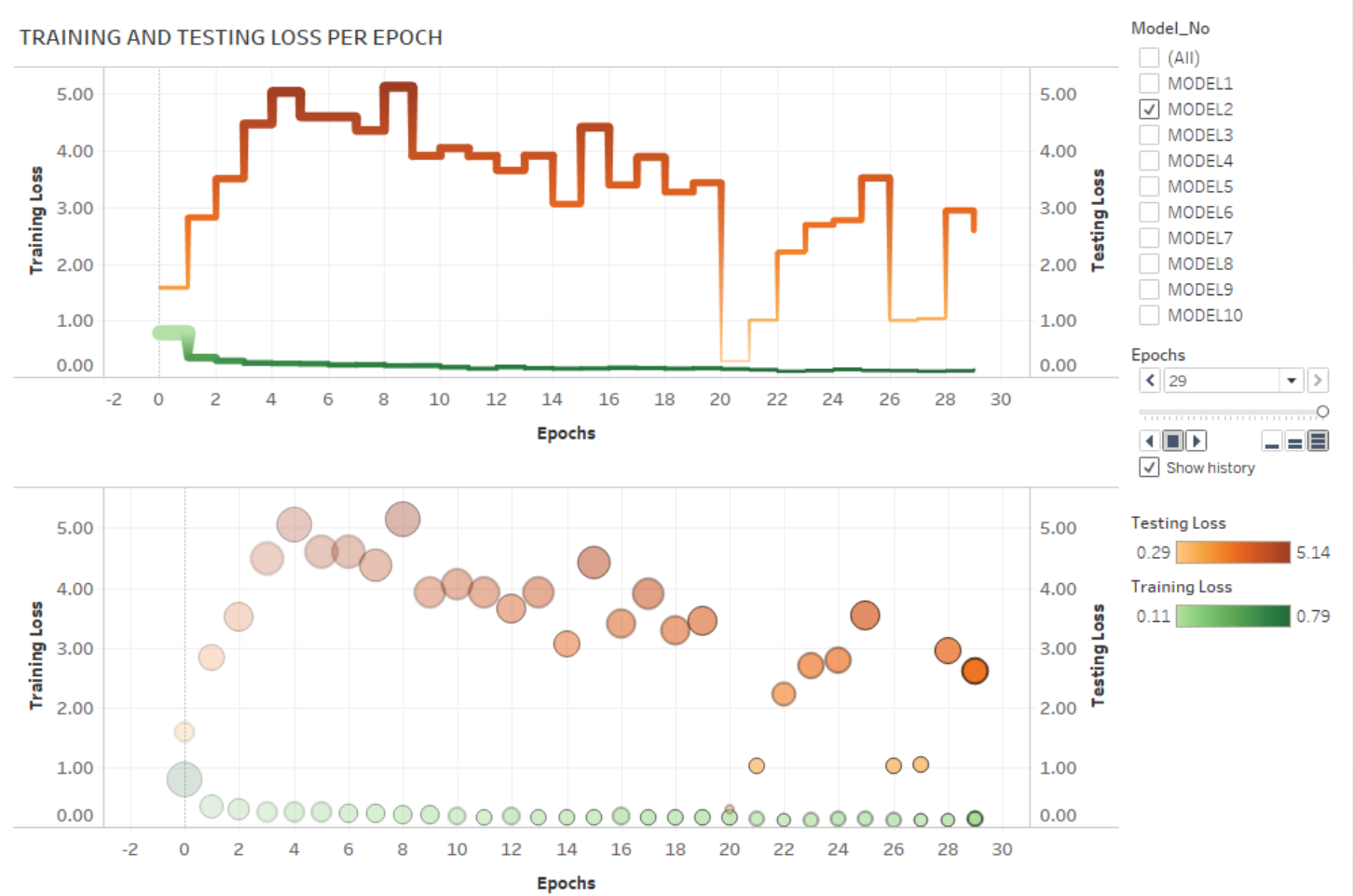

### Model loss per Epoch:




Note:

- ML Model details are provided in Annexure_Model_Details.md

# TABLEAU REPRESENTATION- Loss Trail per Epoch

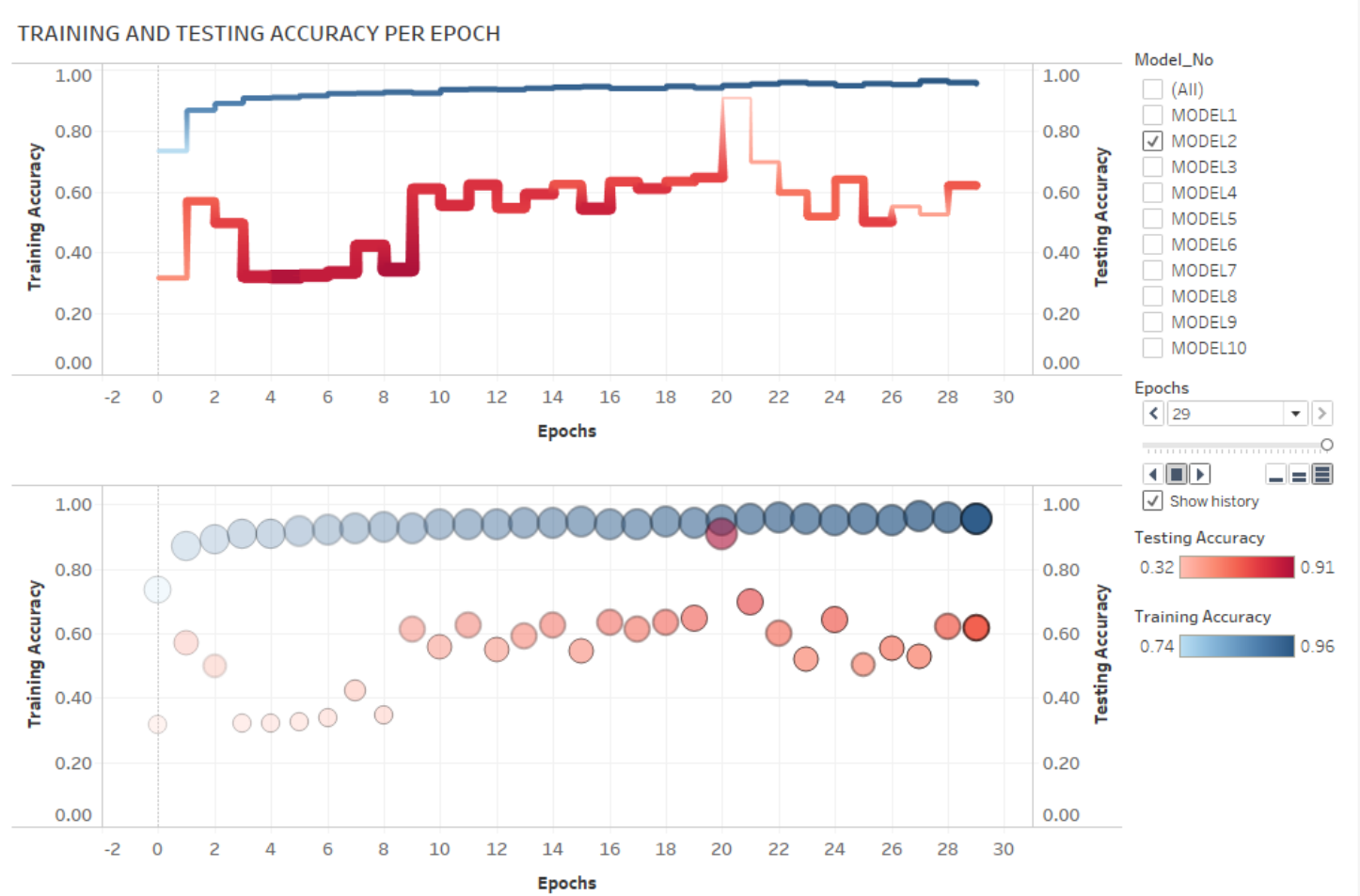



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION- Accuracy Trail per Epoch




Note:

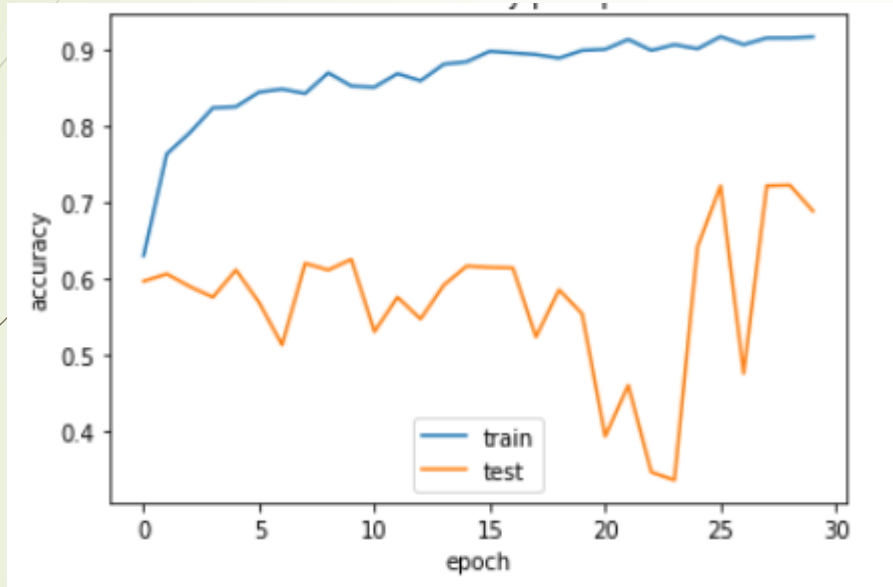
- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

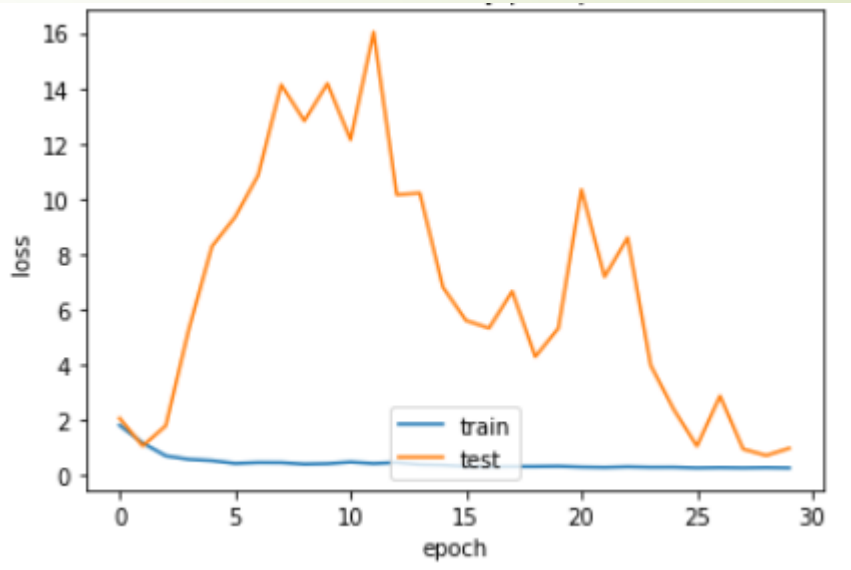
# ANALYSIS AND VISUALIZATION FOR Model-3: Accuracy and Loss

## MODEL-3: Basic Model-C

### Model accuracy per Epoch:



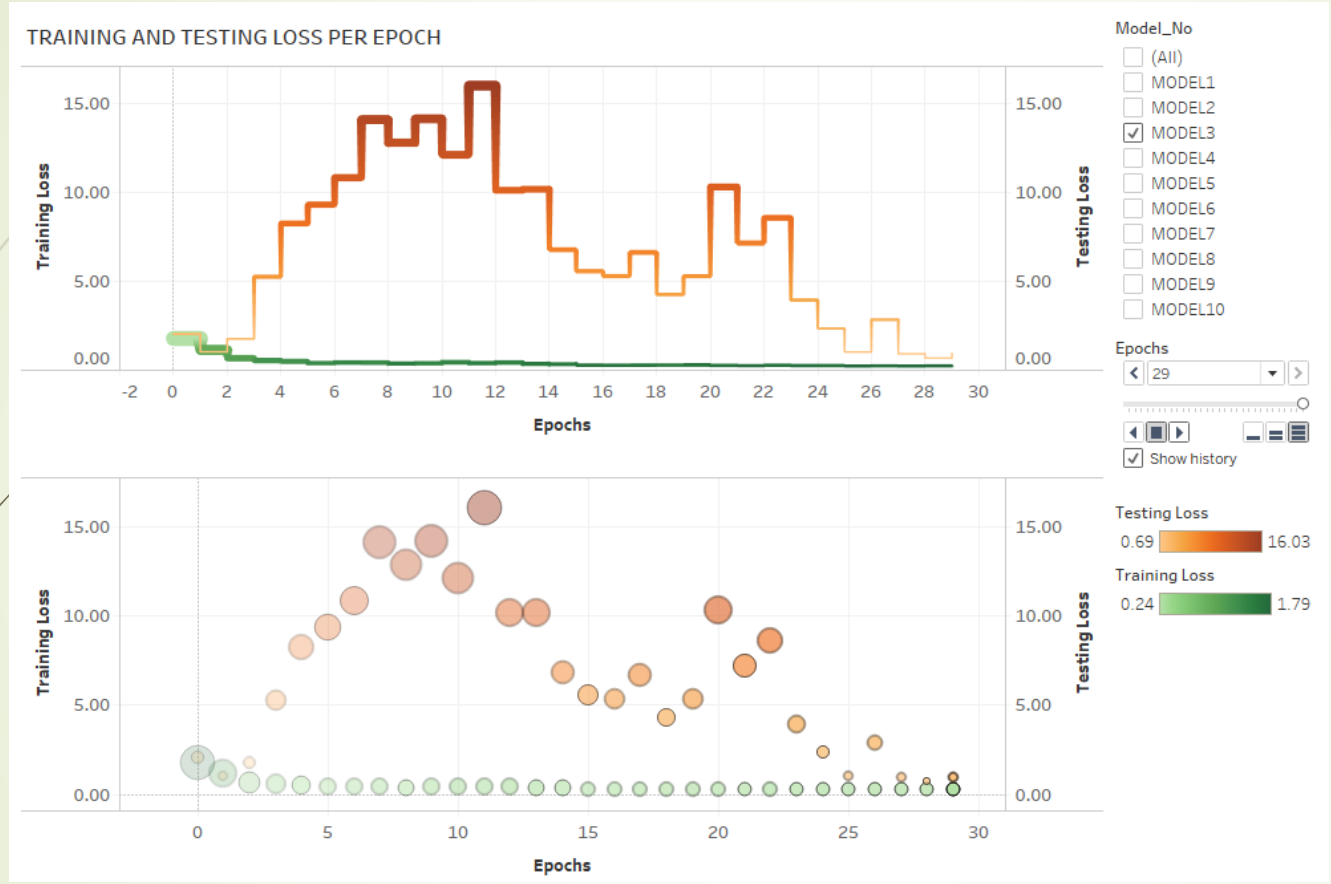

### Model loss per Epoch:




Note:

- ML Model details are provided in Annexure_Model_Details.md

# TABLEAU REPRESENTATION

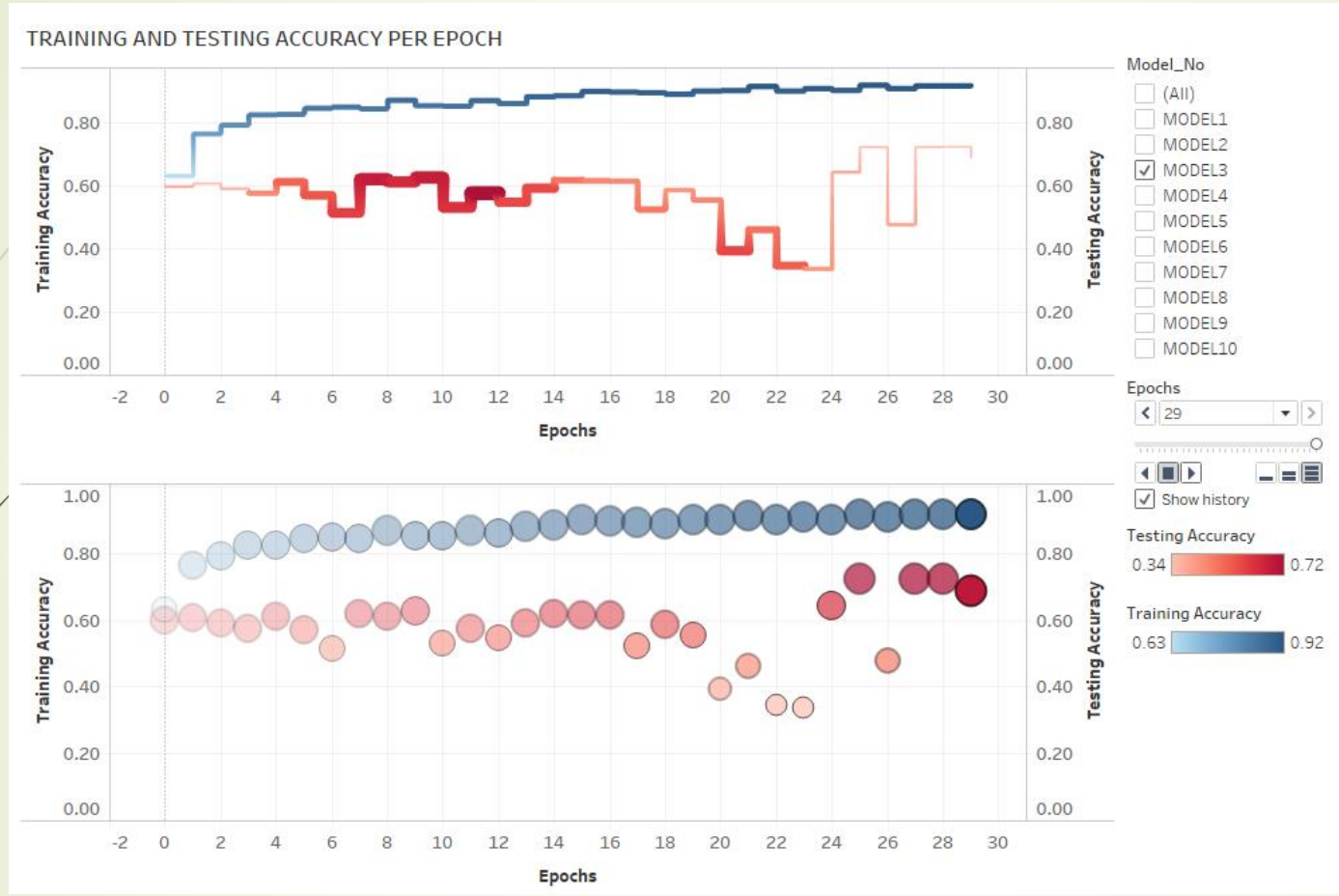



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION



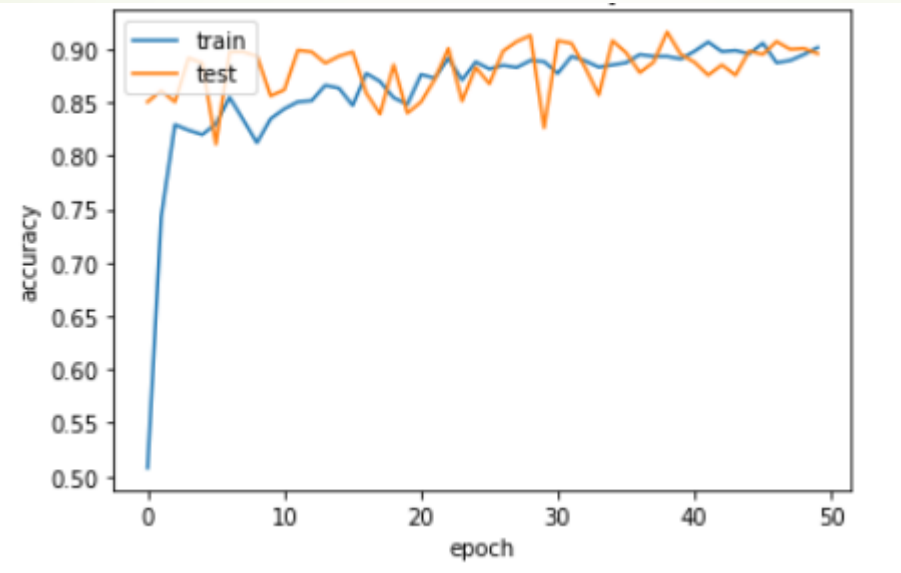

Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

# ANALYSIS AND VISUALIZATION FOR Model-4: Accuracy and Loss
# MODEL 4 – BASIC MODEL-D

## Model accuracy per Epoch:



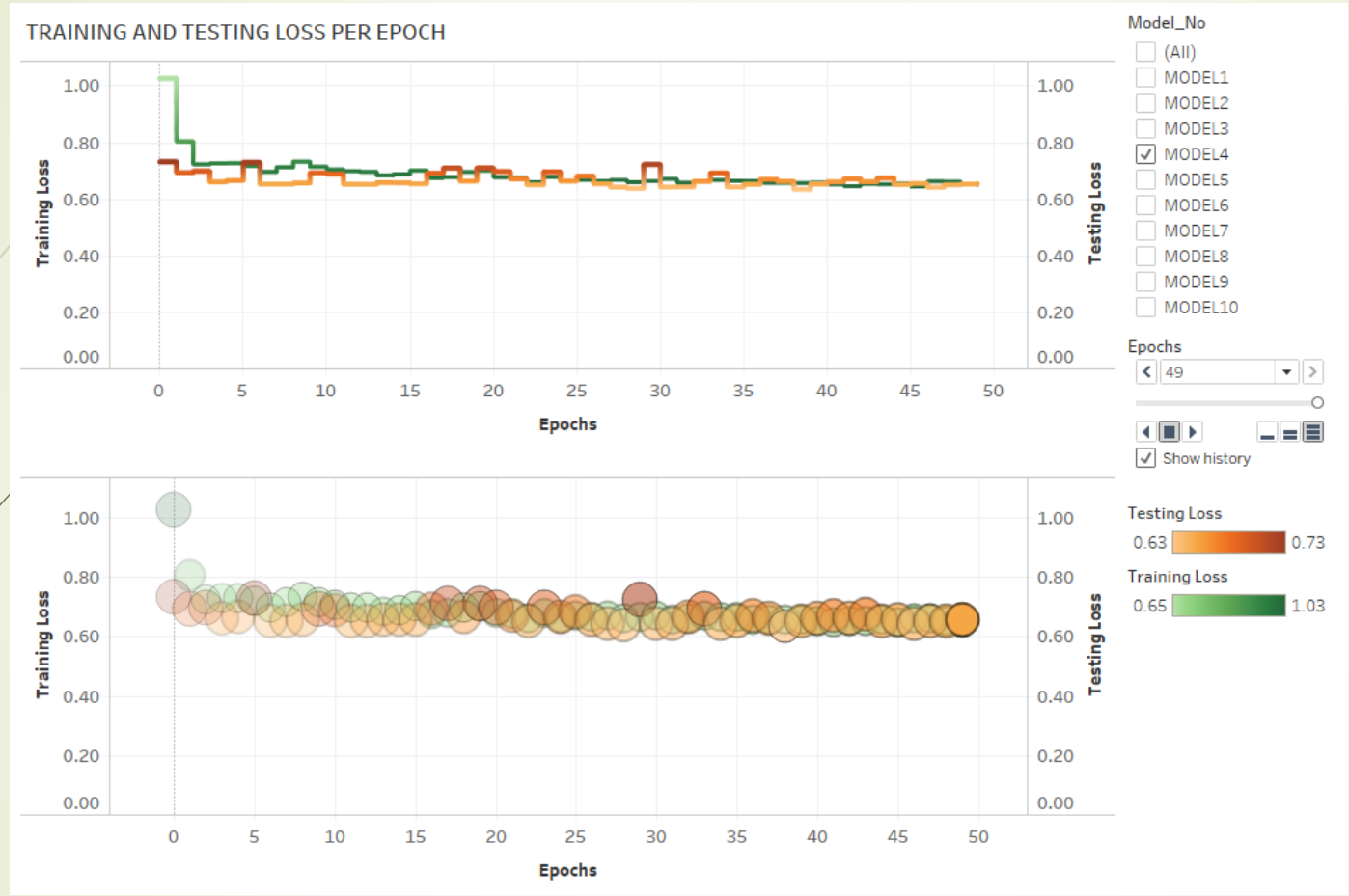

**ANALYSIS:**

- The training and testing curve for accuracy have started to overlap a bit implying out model is learning well now.
- The next step would be to apply a pretrained model VGG16 to train our model.

Note:

- ML Model details are provided in Annexure_Model_Details.md

# TABLEAU REPRESENTATION

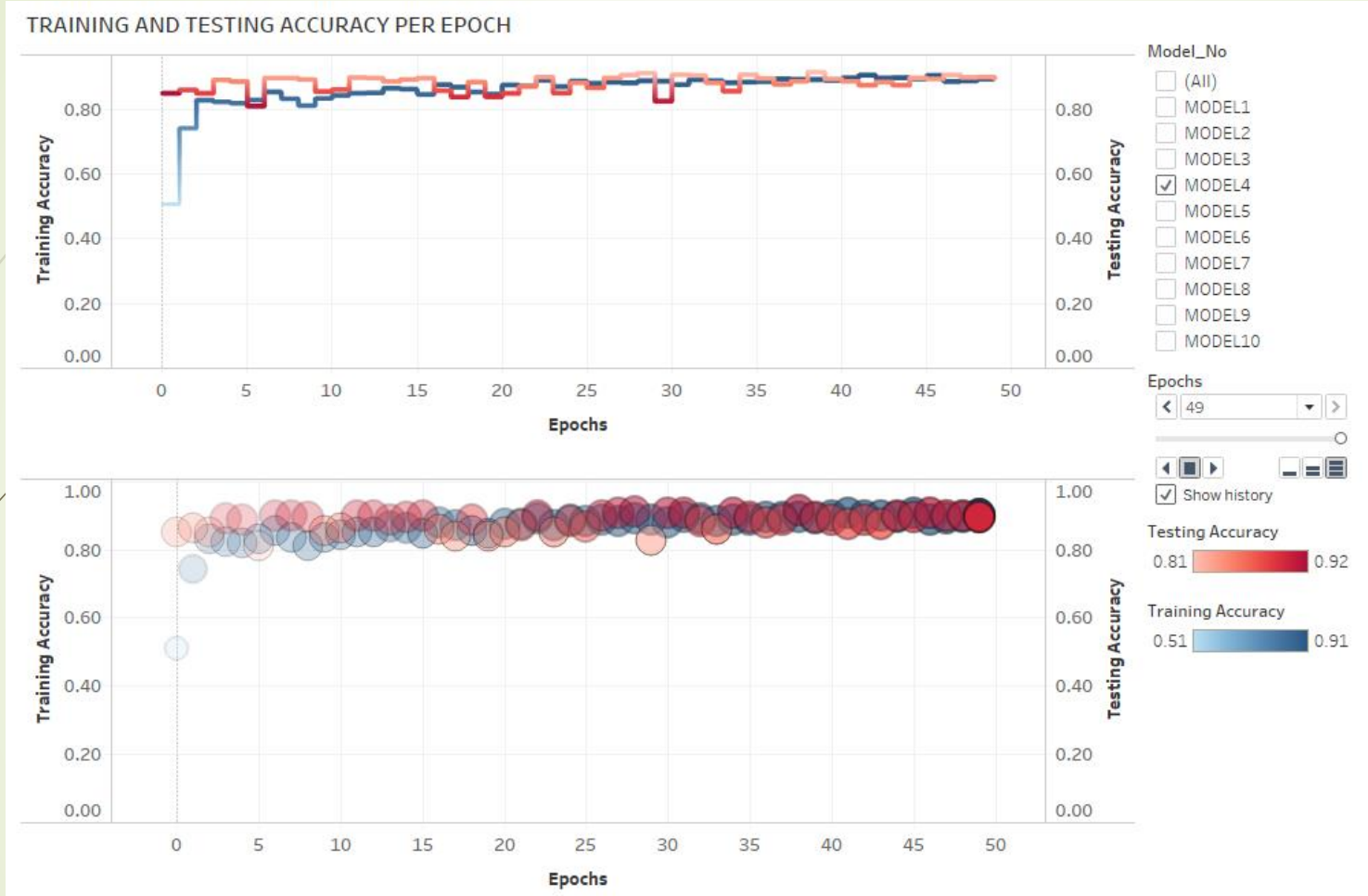



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION




Note:

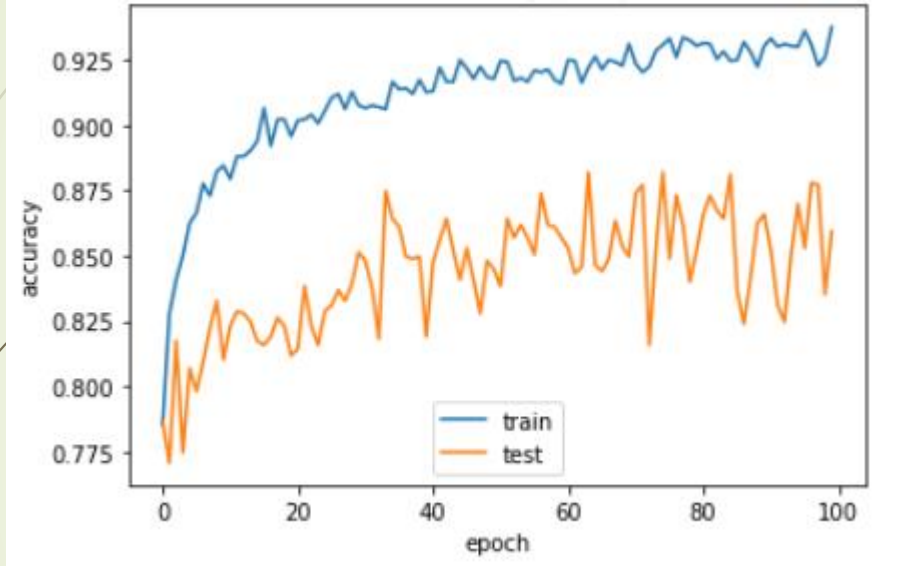
- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

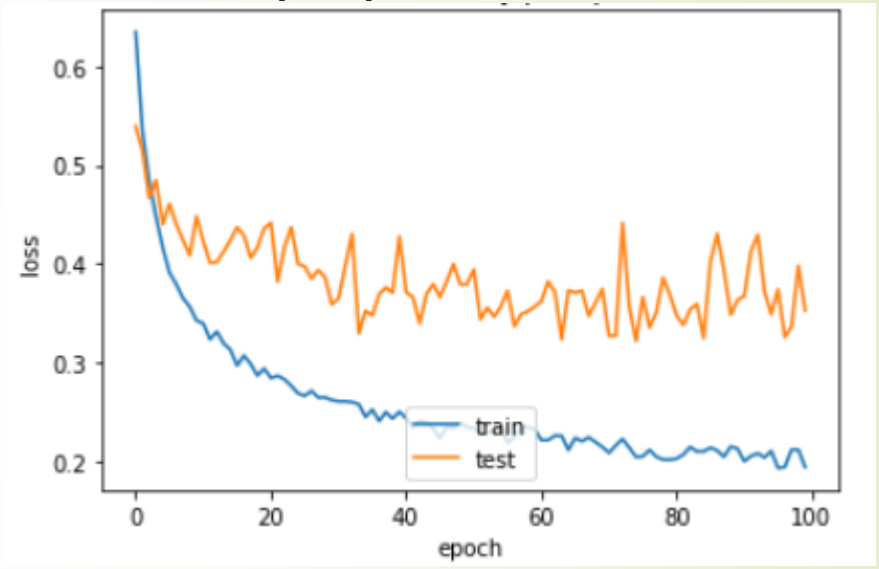
# ANALYSIS AND VISUALIZATION FOR Model-5: Accuracy and Loss

## MODEL-5: VGG-16-A

### Model accuracy per Epoch:



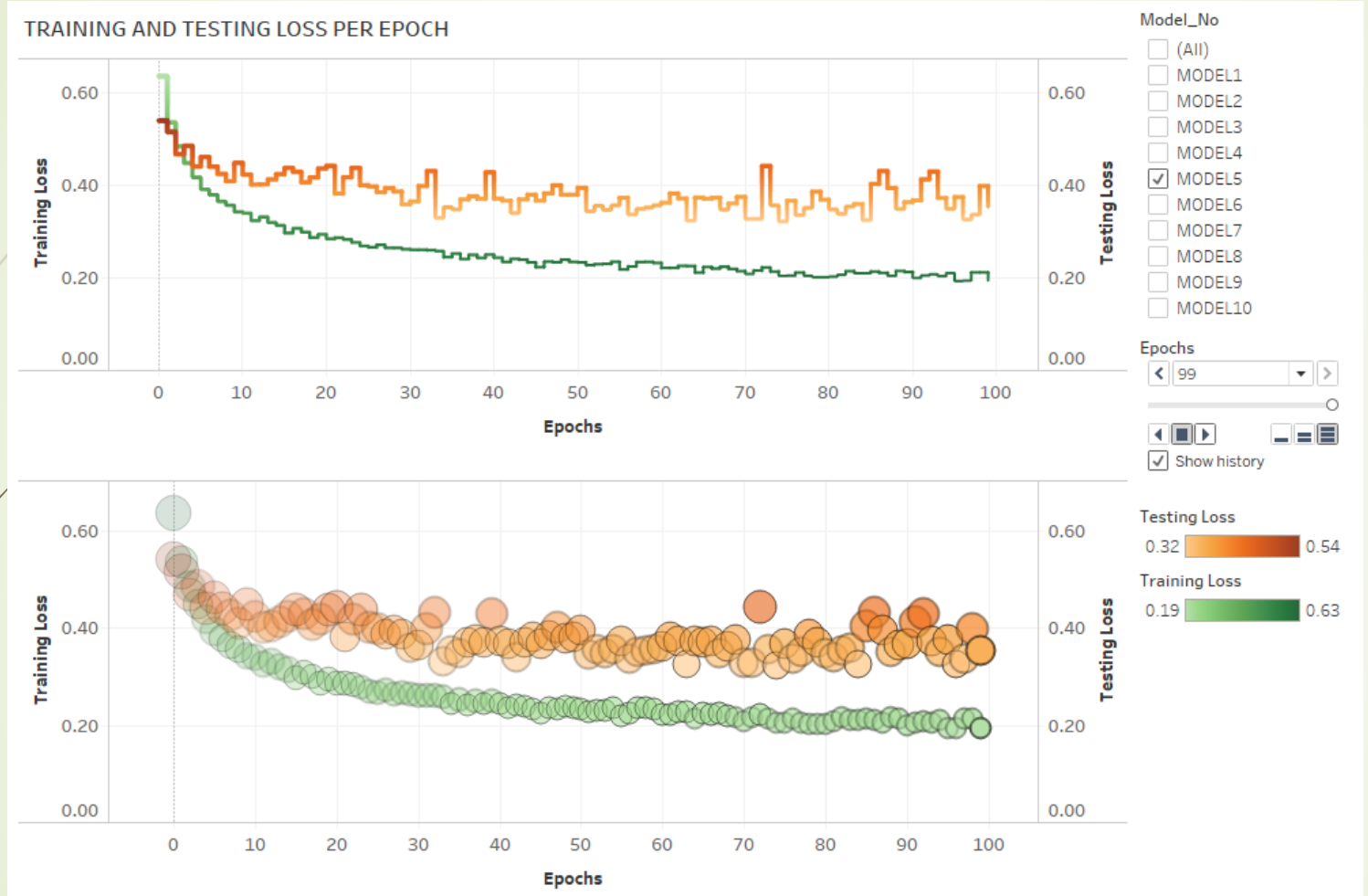

### Model loss per Epoch:




Note:

- ML Model details are provided in Annexure_Model_Details.md

# TABLEAU REPRESENTATION

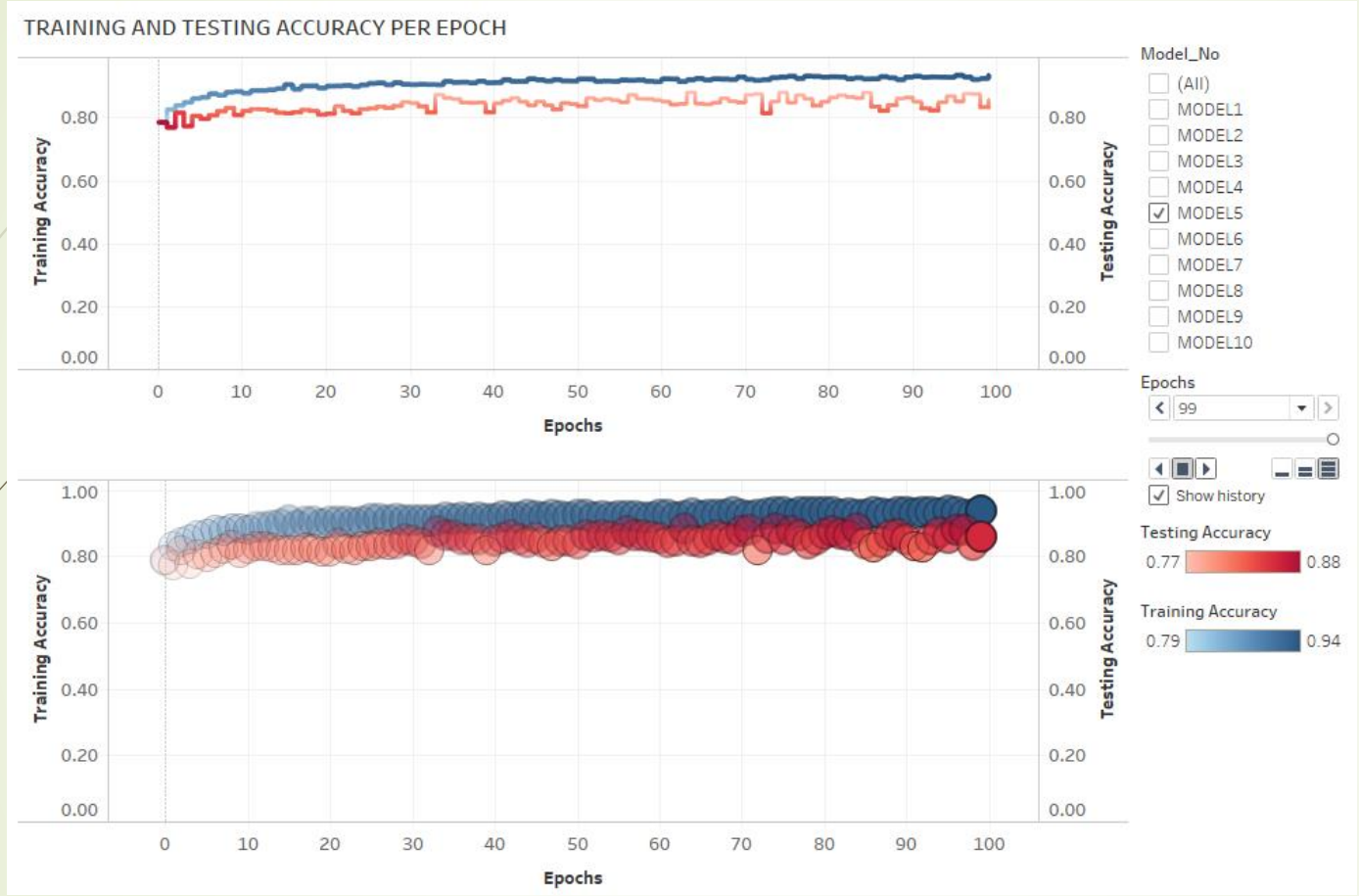



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION




Note:

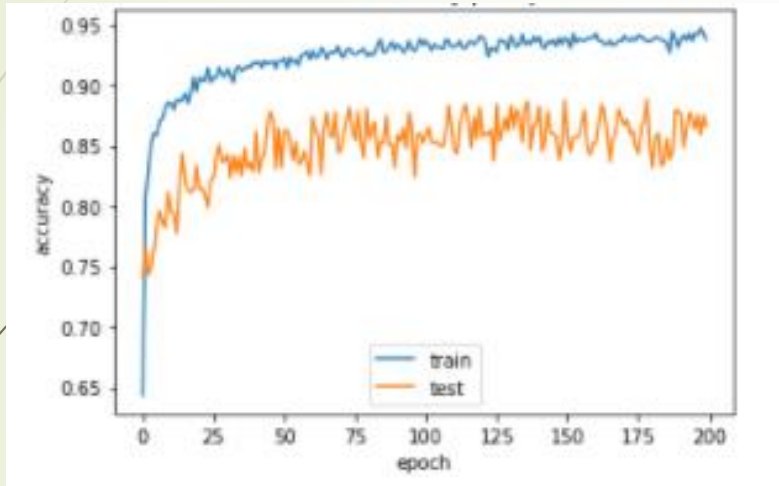
- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

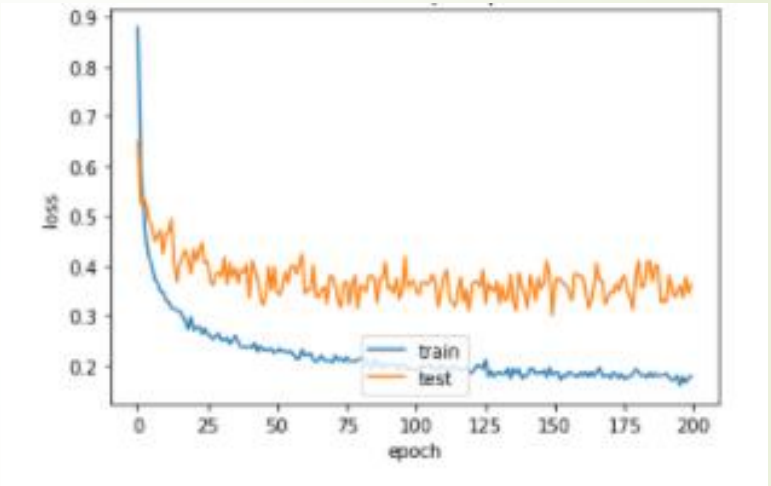
# ANALYSIS AND VISUALIZATION FOR Model-6: Accuracy and Loss

## MODEL6: VGG-16-B

### Model accuracy per Epoch:



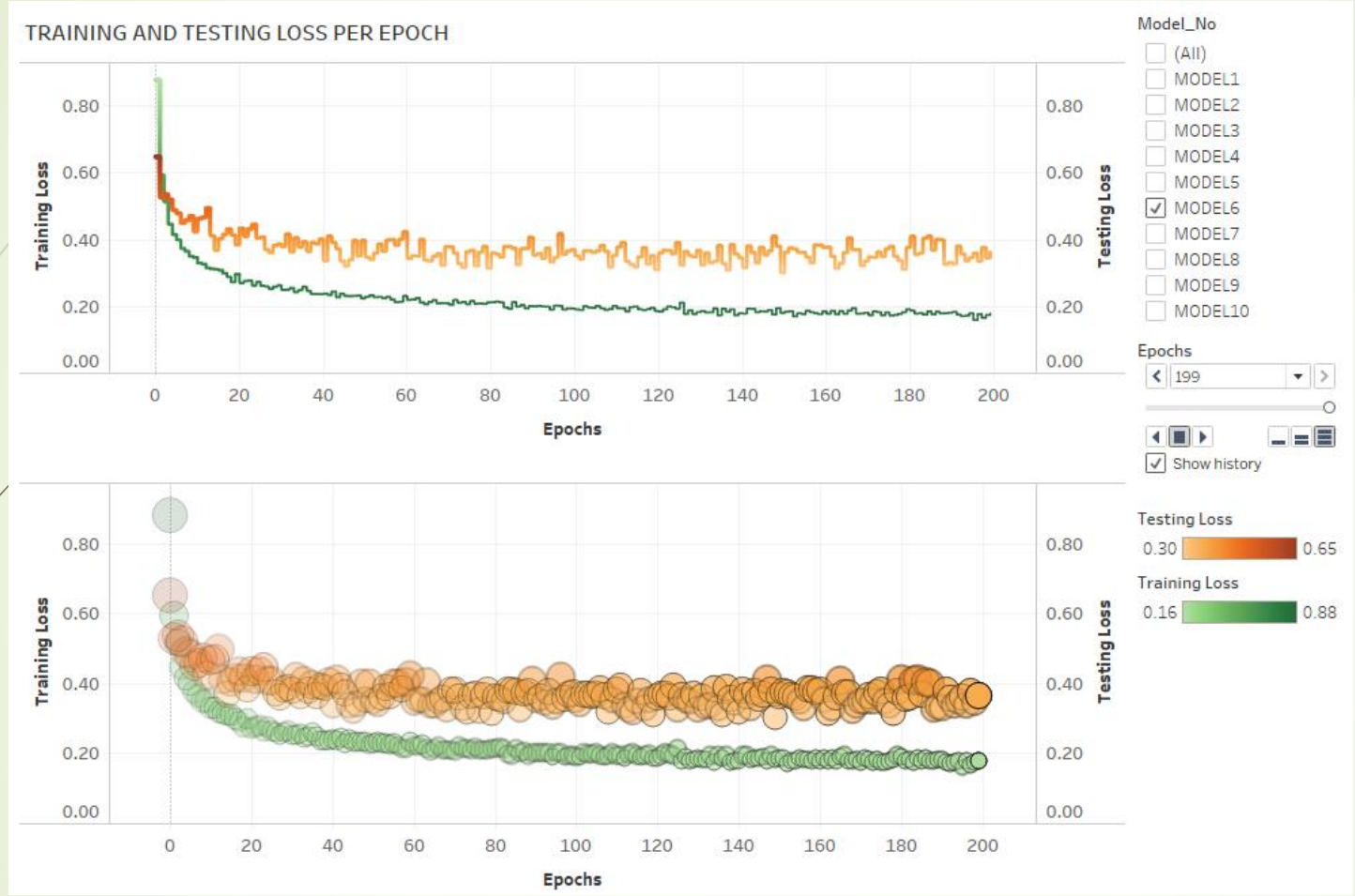

### Model loss per Epoch:




Note:

- ML Model details are provided in Annexure_Model_Details.md

# TABLEAU REPRESENTATION

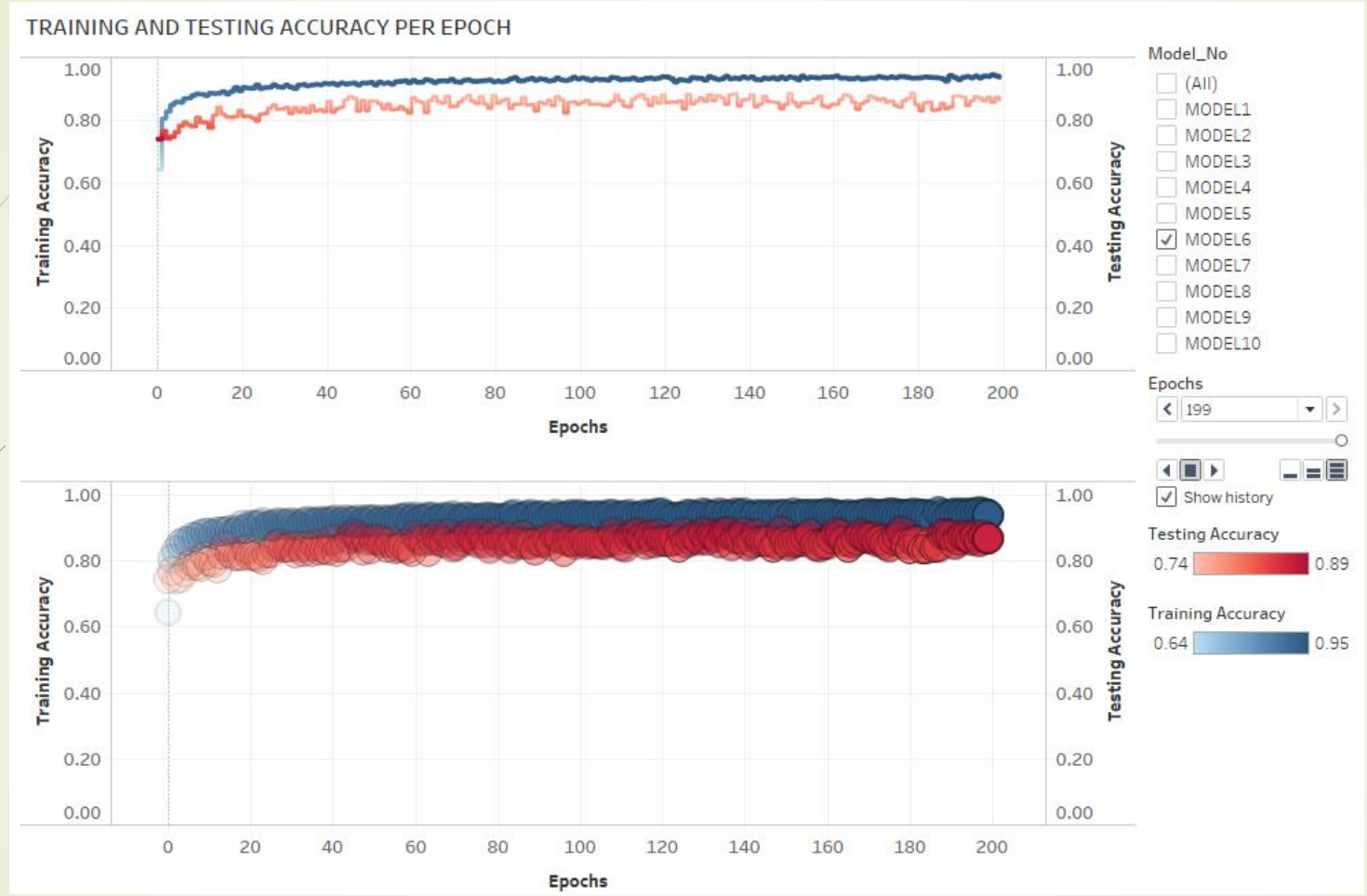



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION




Note:

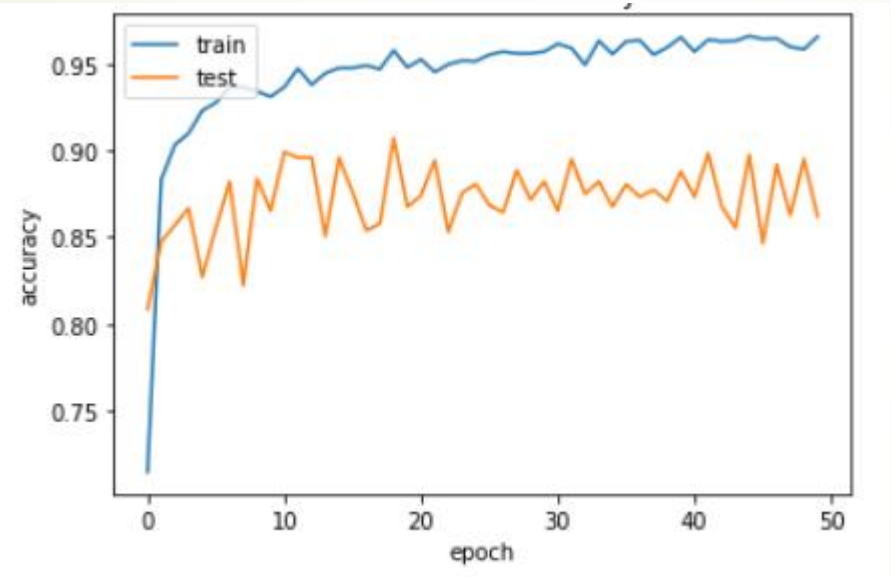
- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

# ANALYSIS AND VISUALIZATION FOR Model-7: Accuracy and Loss

## MODEL 7: VGG-16-C

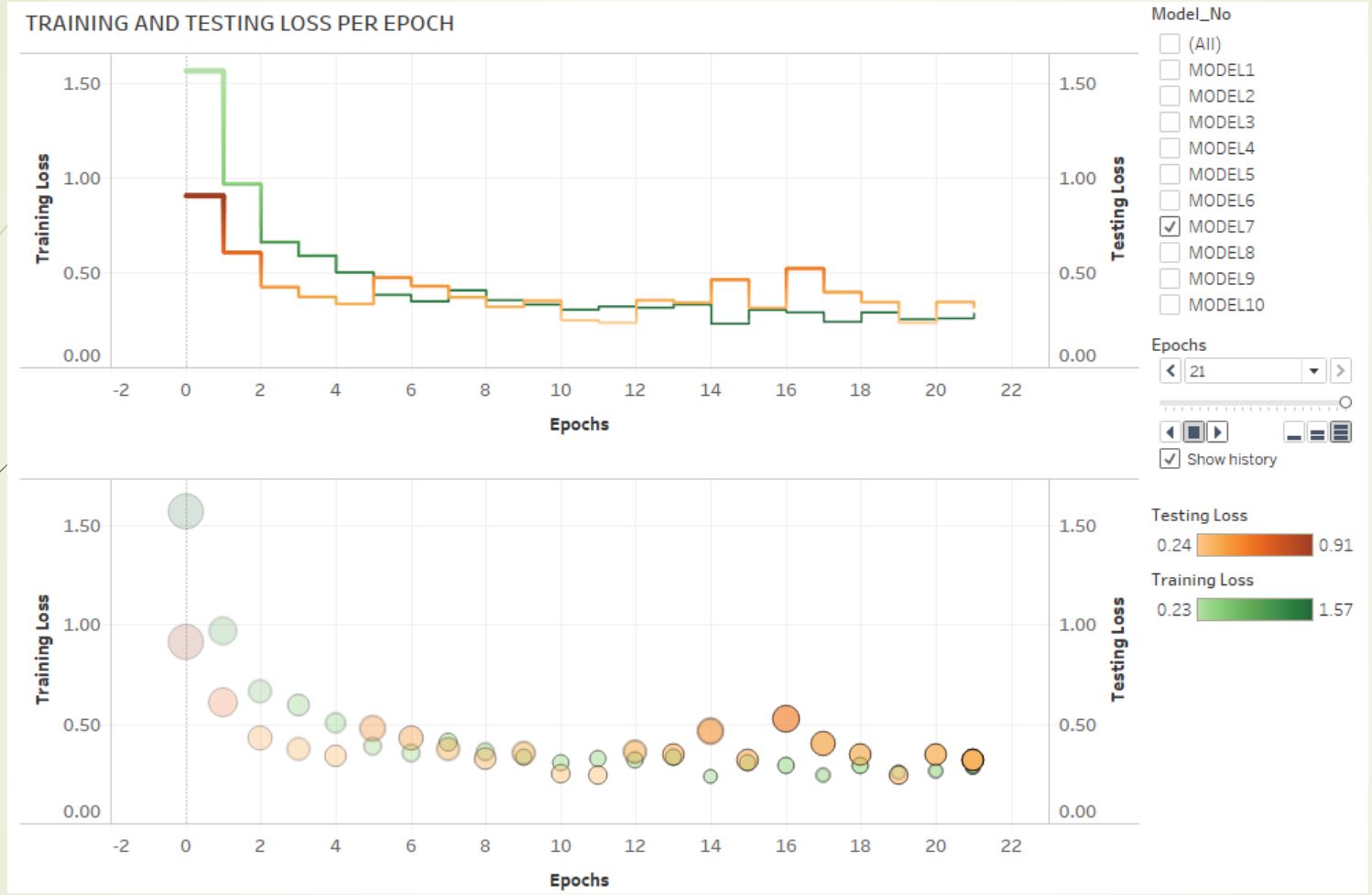
### Model accuracy per Epoch:




Note:

- ML Model details are provided in Annexure_Model_Details.md

# TABLEAU REPRESENTATION

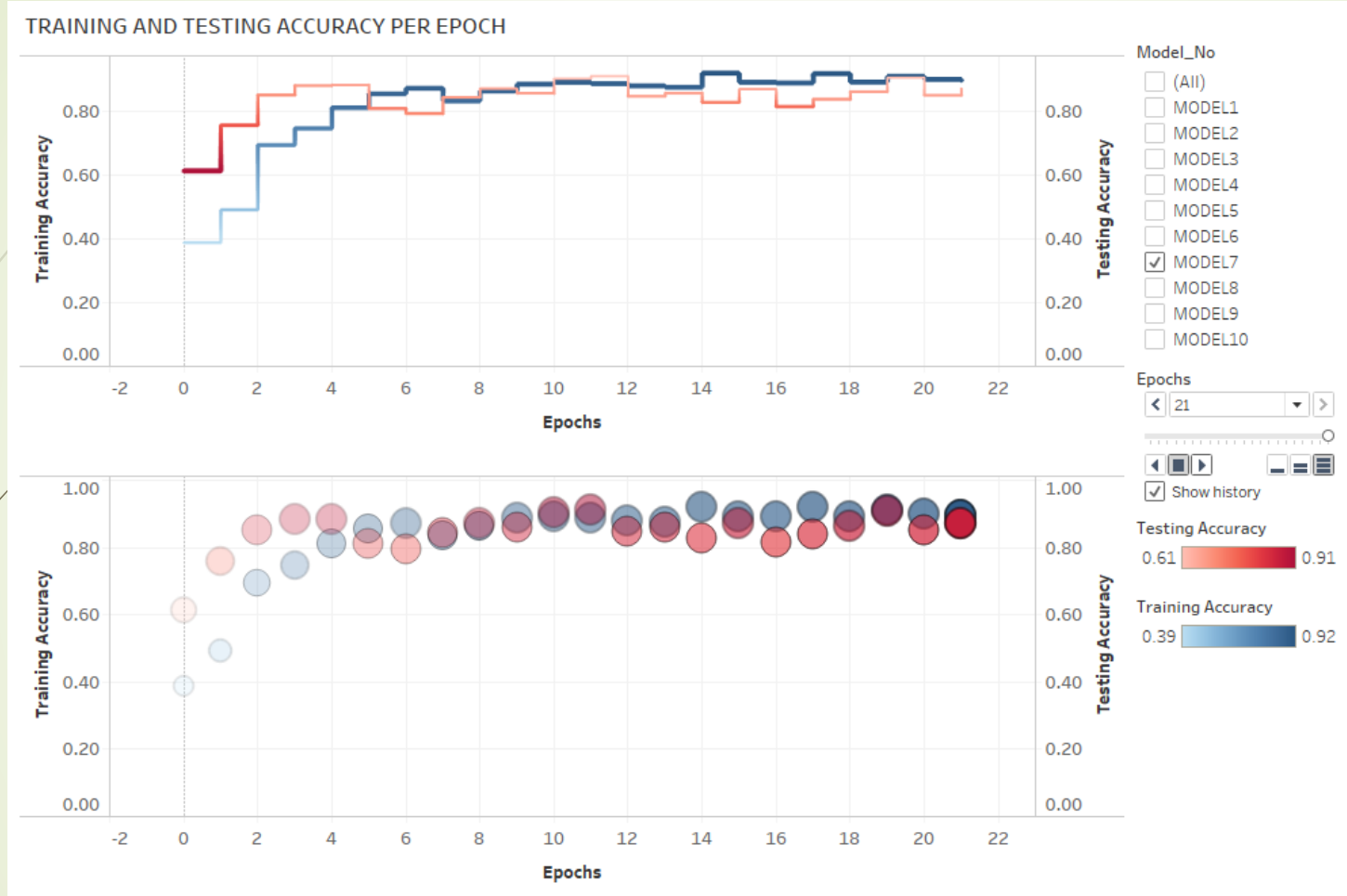



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION




Note:

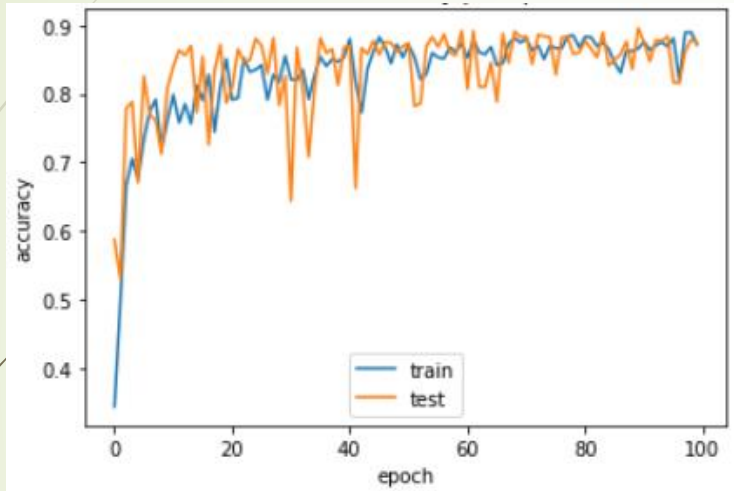
- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

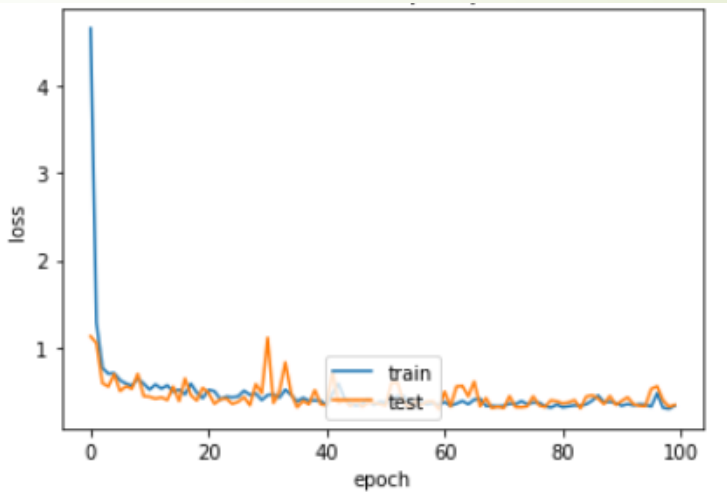
# ANALYSIS AND VISUALIZATION FOR Model-8: Accuracy and Loss

## MODEL 8: Resnet50-A

### Model accuracy per Epoch:



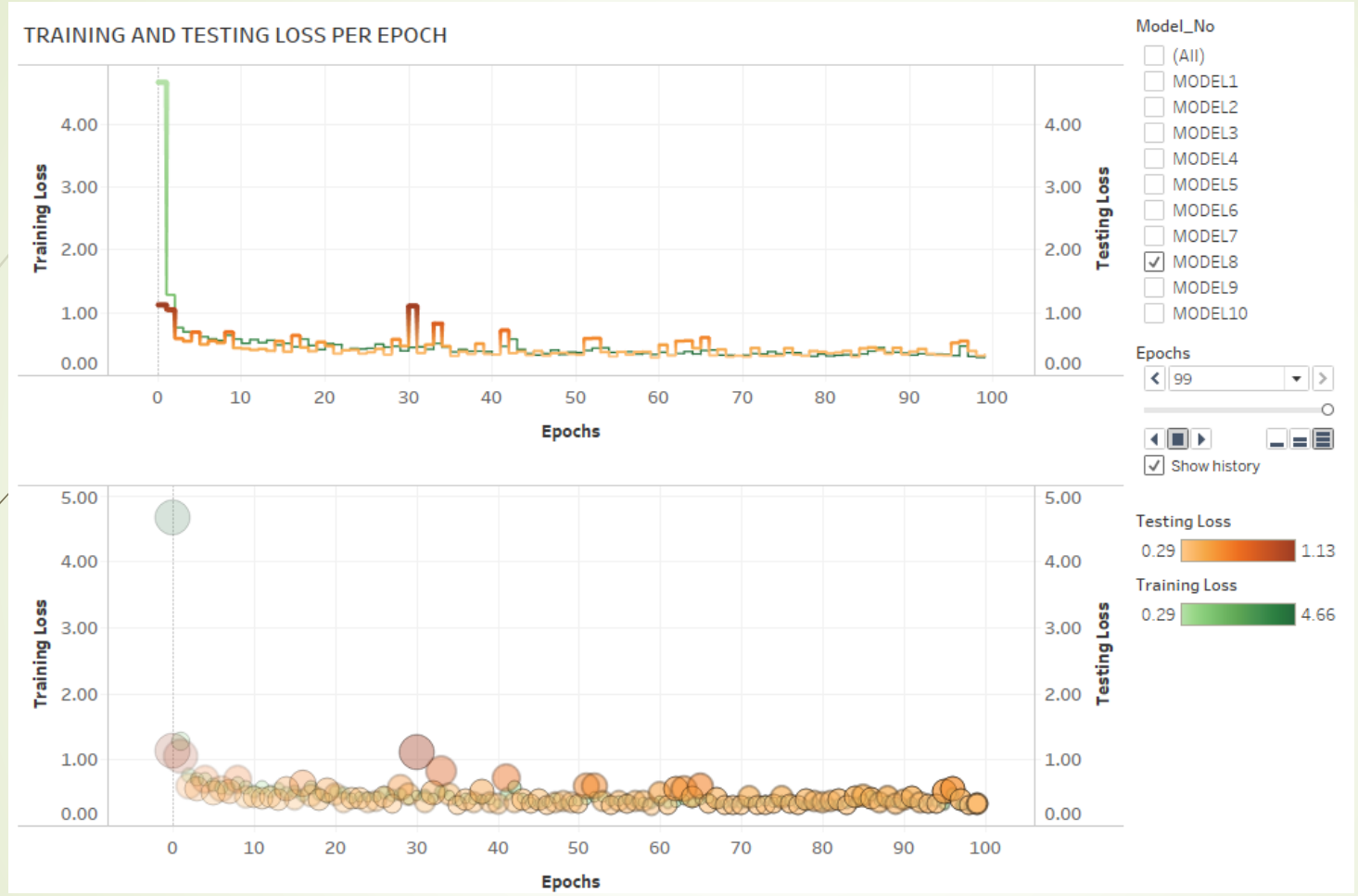

### Model loss per Epoch:




Note:

- ML Model details are provided in Annexure_Model_Details.md

# TABLEAU REPRESENTATION

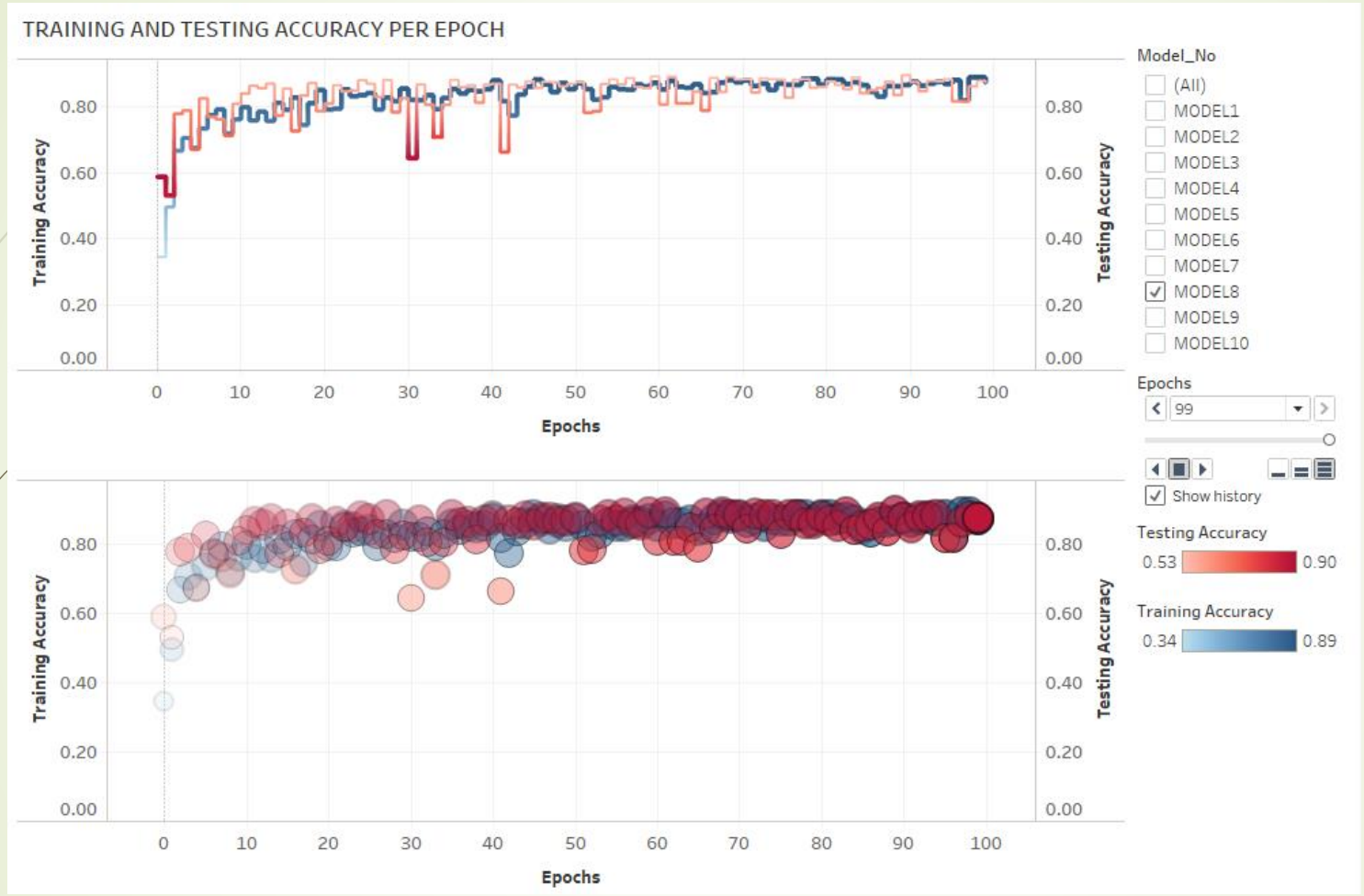



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION




Note:

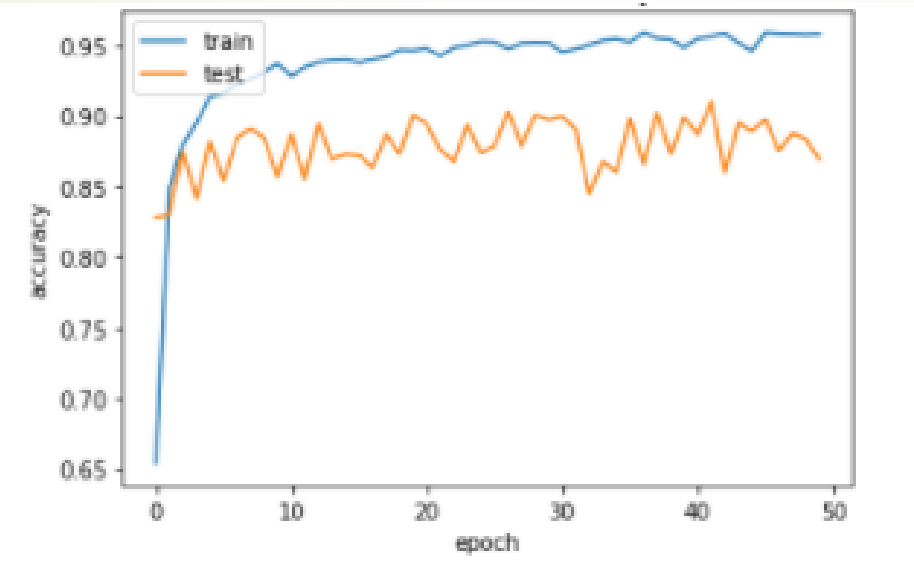
- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

# ANALYSIS AND VISUALIZATION FOR Model-9: Accuracy and Loss

## MODEL-9: Resnet50-B

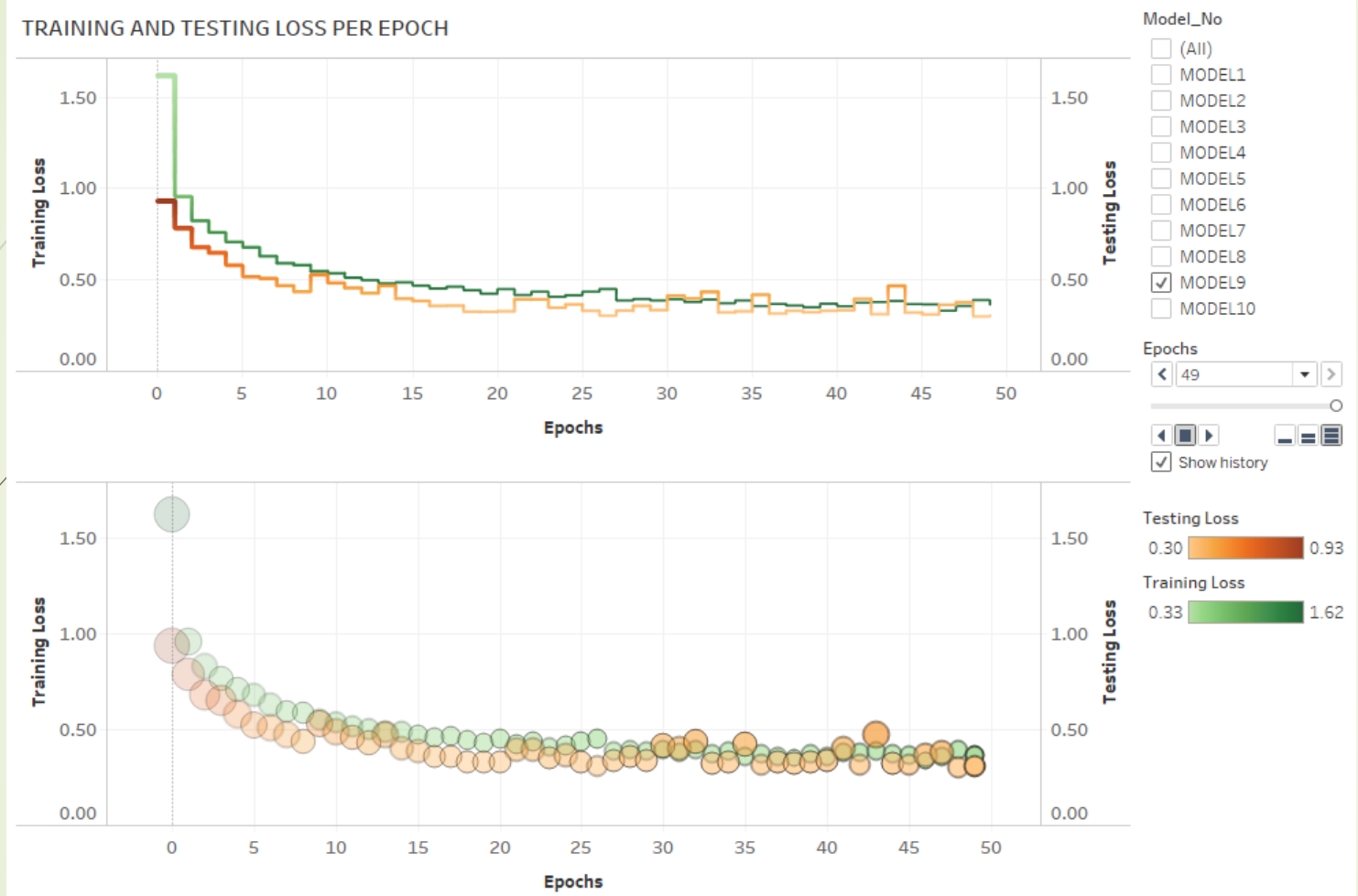
### Model accuracy per Epoch:




Note:

- ML Model details are provided in Annexure_Model_Details.md

# TABLEAU REPRESENTATION

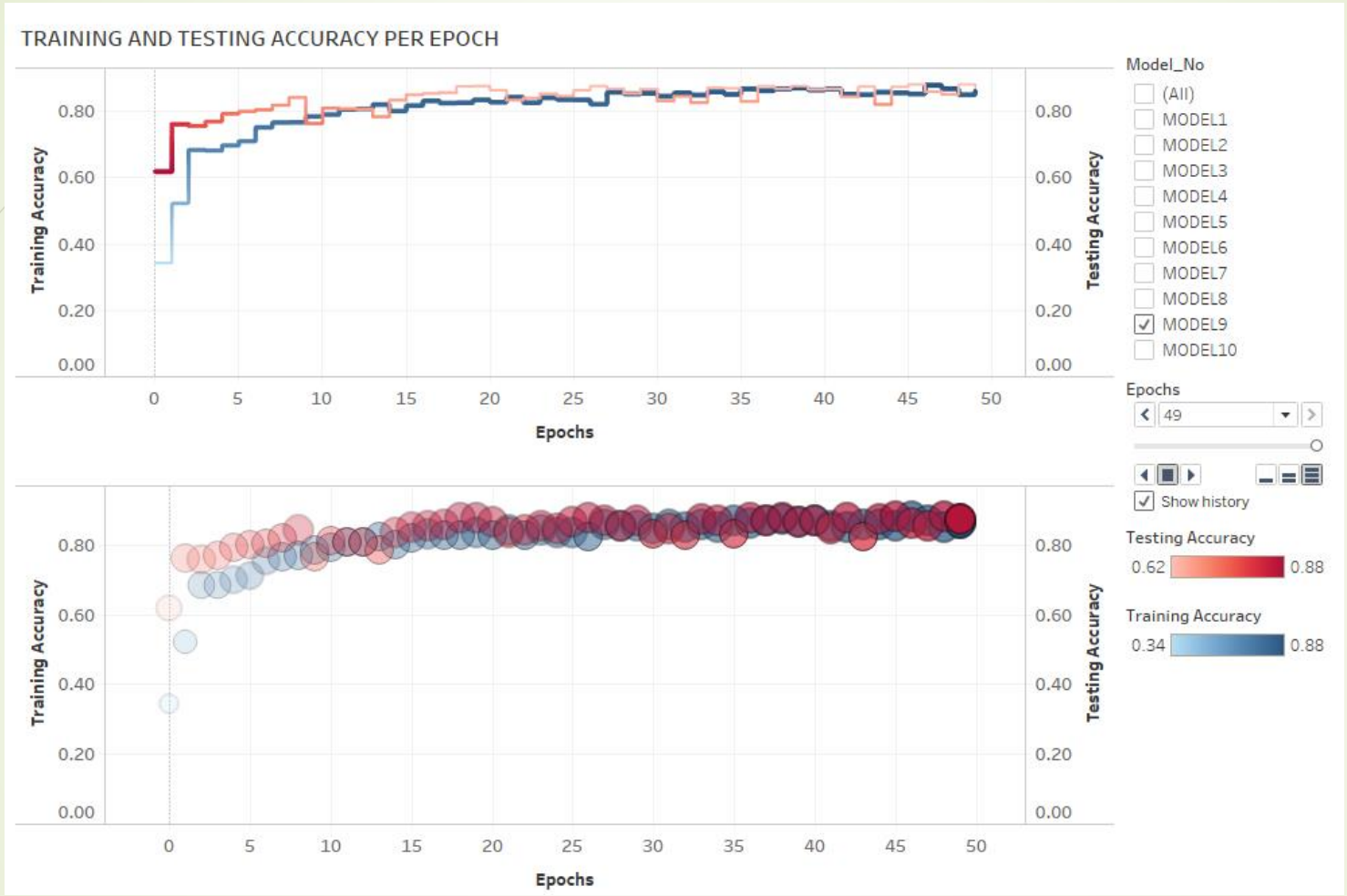



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION




Note:

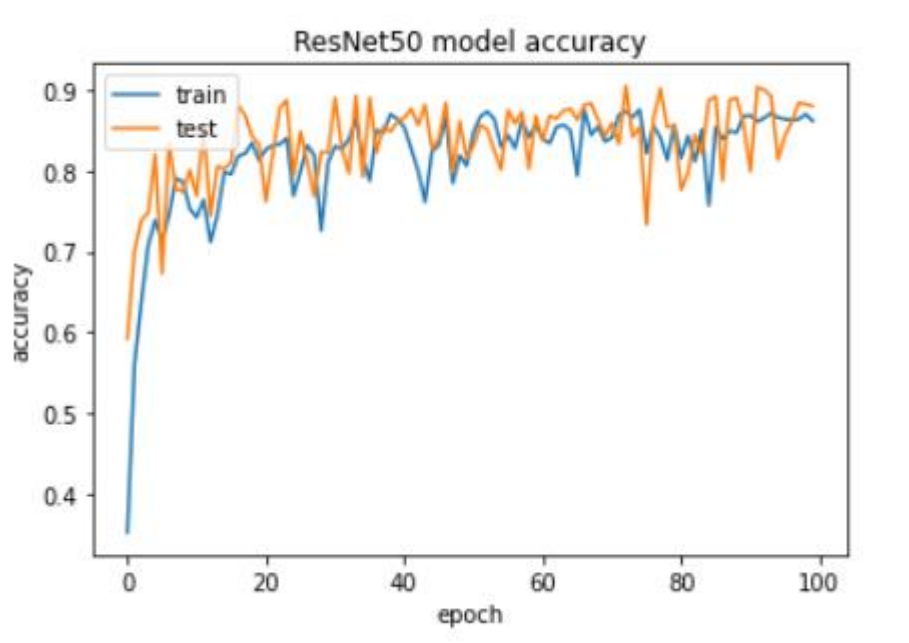
- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

# ANALYSIS AND VISUALIZATION FOR Model-10: Accuracy and Loss

## MODEL-10: Resnet50-C

### Model accuracy per Epoch:

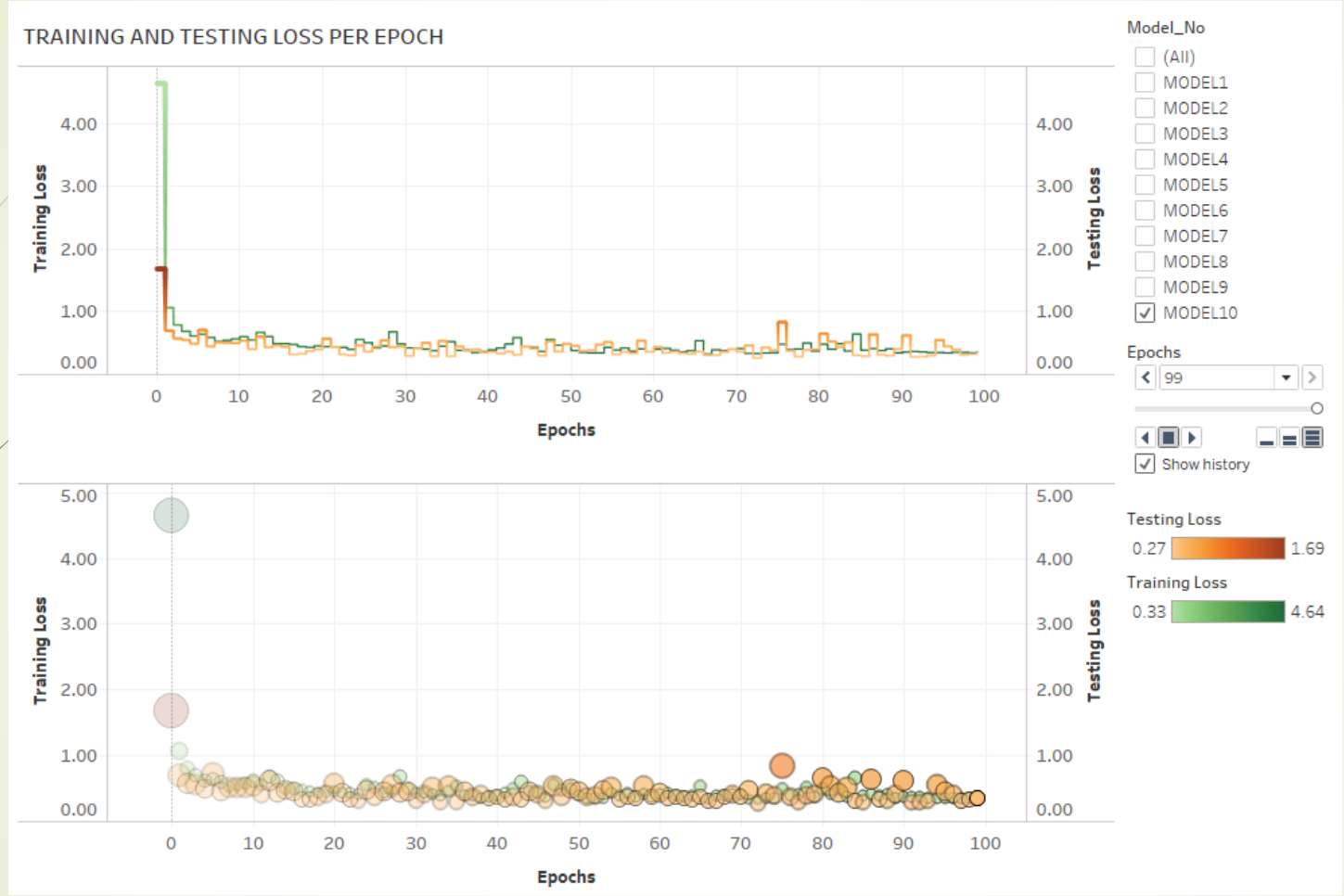



Note:

- ML Model details are provided in Annexure_Model_Details.md

# ANALYSIS AND VISUALIZATION FOR Model-10: Accuracy and Loss: TABLEAU REPRESENTATION

## MODEL-10: Resnet50-C

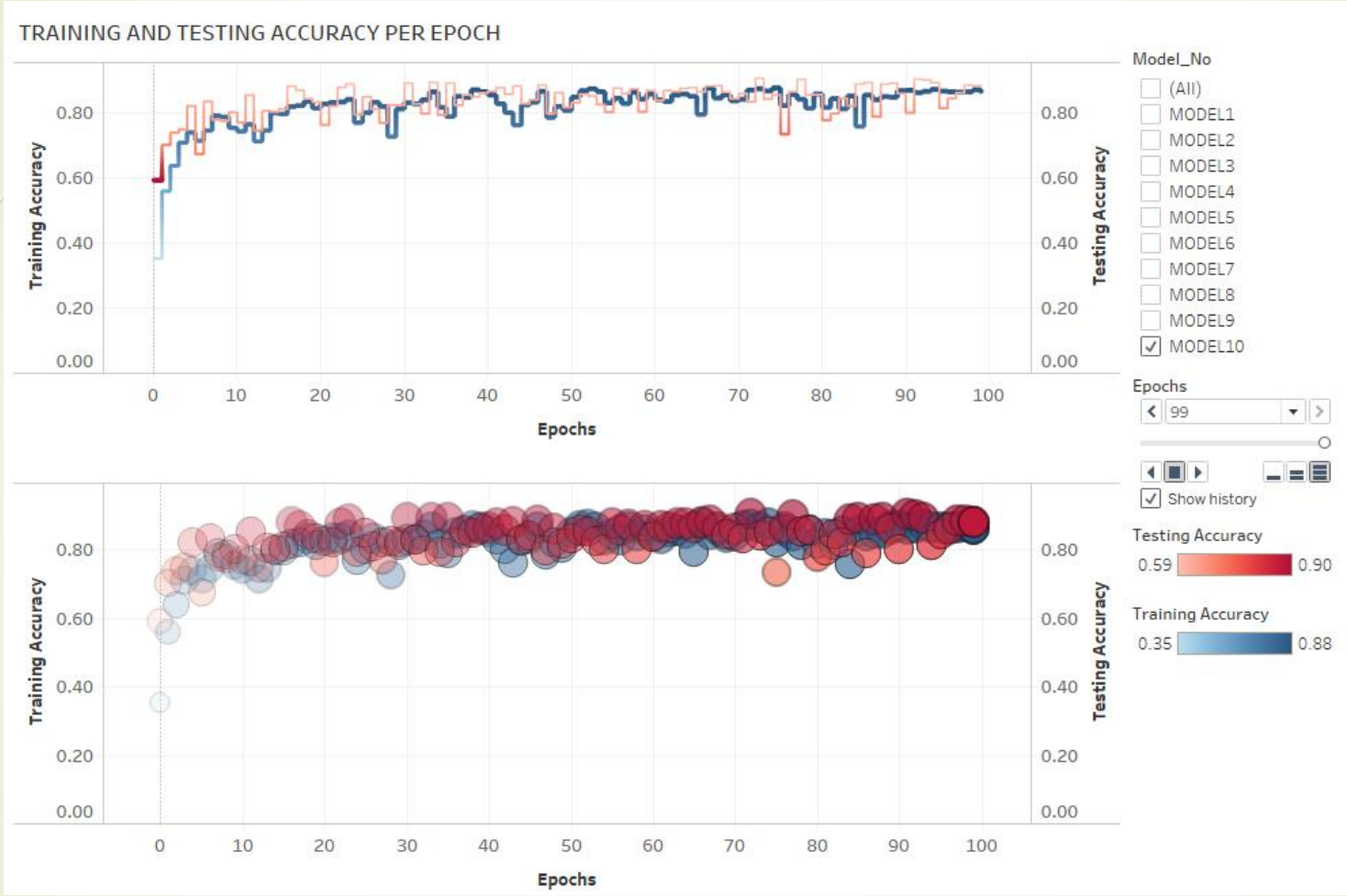



Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing loss.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Loss

# TABLEAU REPRESENTATION



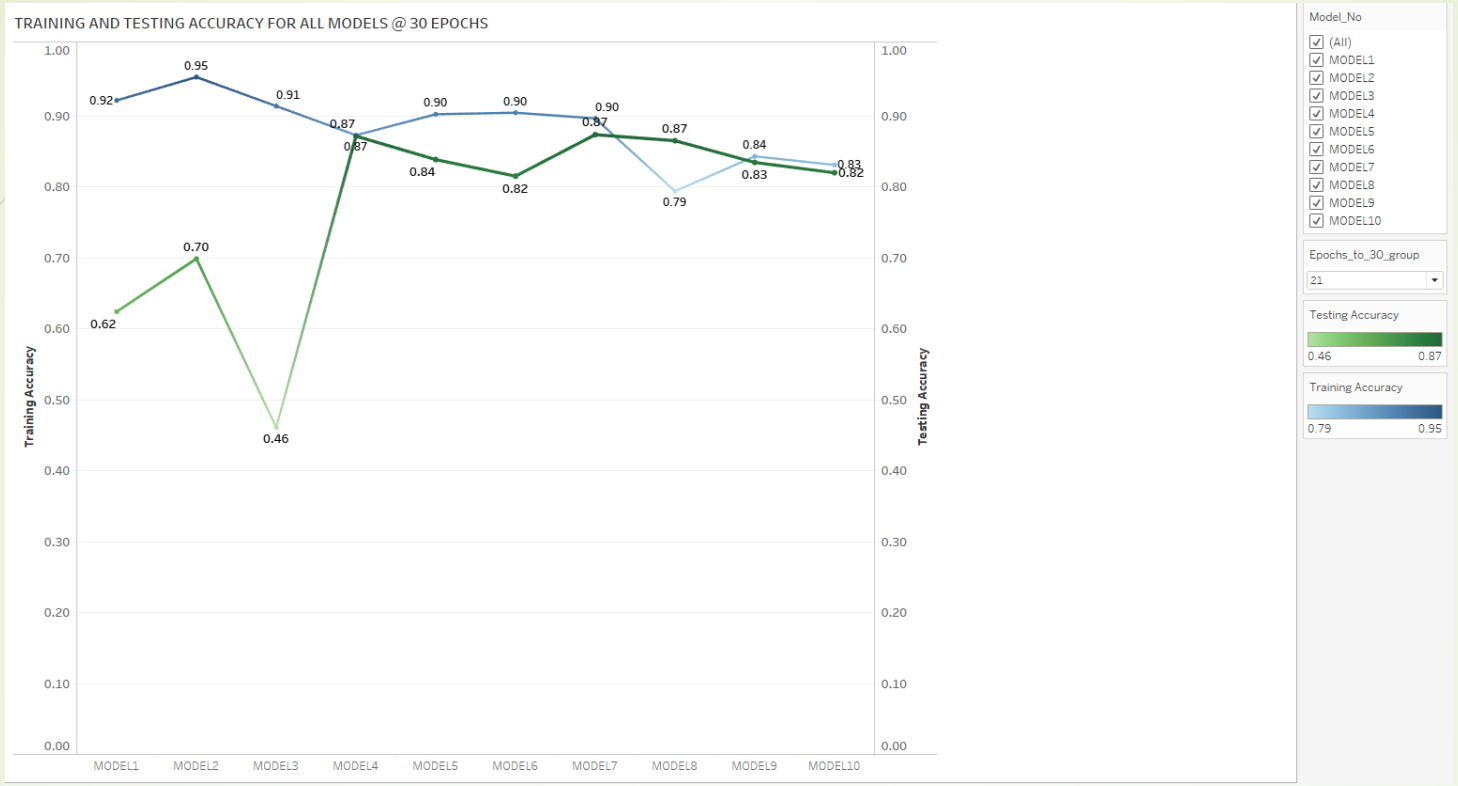

Note:

- Please refer tableau public link below to see the animation showing trail of the training and testing accuracy.
- Please select the Model No on the screen and click on the play button to see the animated trail.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/Model_Accuracy

# ANALYSIS AND VISUALIZATION: OVERALL MODEL PERFORMANCE

## Testing and training Loss from Model 1 to Model 10



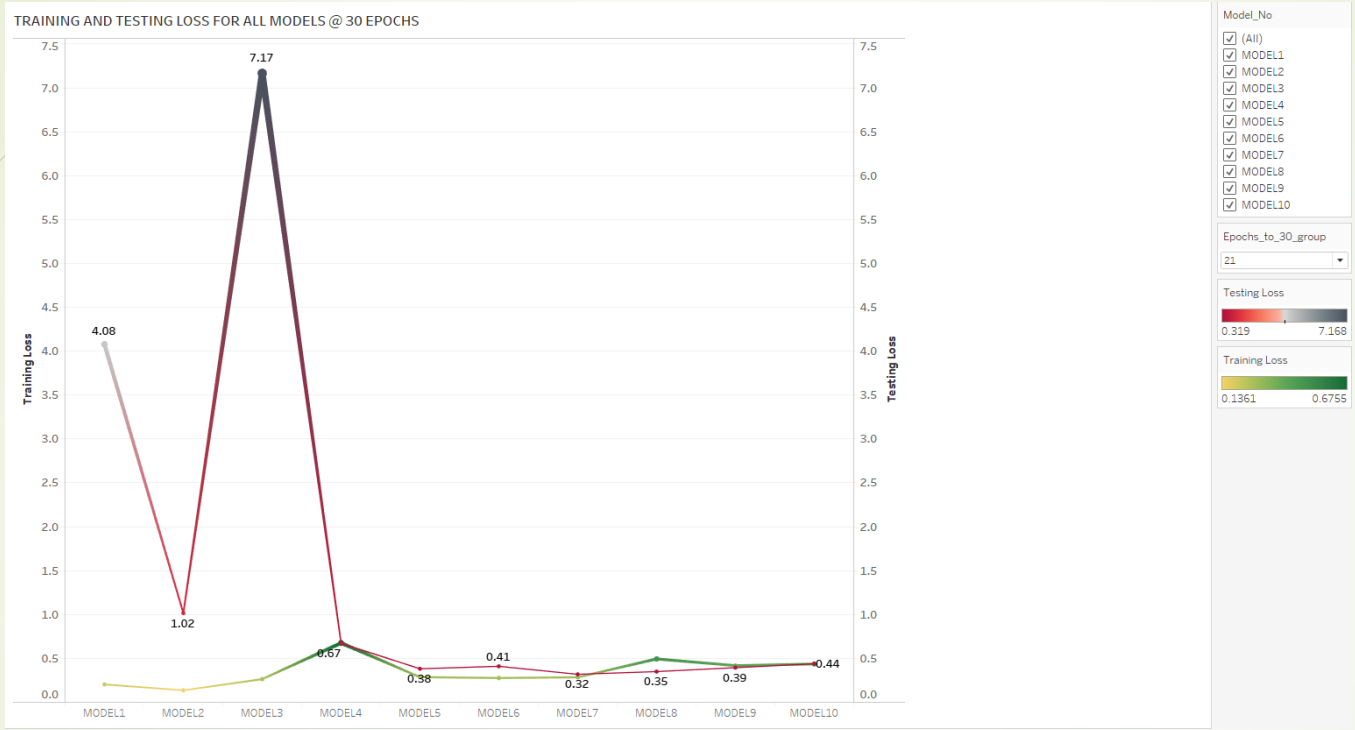

NOTE:

- To summarize all models, although the models have been run to different Epochs as required, I am plotting the loss at 30-50 Epochs to have a fair performance comparison of all models.
- Please refer to the tableau link provided below to see chart interactivity.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/AccuracyLoss-Model1to10

# ANALYSIS AND VISUALIZATION: MODEL PERFORMANCE

- Testing and training Accuracy from Model 1 to Model 10




NOTE:

- To summarize all models, although the models have been run to different Epochs as required, I am plotting the accuracy at 30-50 Epochs to have a fair performance comparison of all models.
- Please refer to the tableau link provided below to see chart interactivity.

Tableau Public link:
https://public.tableau.com/profile/stuti.budhwar#!/vizhome/Covid_Prediction_Model_Dashboard/AccuracyLoss-Model1to10

## OBSERVATIONS AND ANALYSIS OF MODEL PERFORMANCE:

- The old dataset of 5000 images of chest X-rays with 14 classes of medical condition, created a model that did not perform well at all. The number of images after balancing the data were too less for 14 medical conditions that they had to be trained for.
- The imbalanced data however performed better giving an accuracy of about 50%, which was due to almost 50% of the images belonging to "No finding" category. This model would have been a biased ML model.
- The change in the dataset to with just 3 classes showed great improvement in the performance of our models.
- With this new dataset. we started with the basic model that was just a little overfitted with excellent training accuracy. We kept modifying the model slightly so it can generalize better.
- VGG16 performed better than the base model. Despite the base model having a slightly better Val accuracy; because the training accuracy of VGG16 is way better, this implies VGG16 has more potential.
- VGG16 did seem overfitted initially, so we added a dropout layer in subsequent models to see if it generalizes better.
- After running a few basic and pretrained models, we realized that class 1 which is 'Viral Pneumonia' is being grossly misinterpreted. Almost a third of Viral Pneumonia is being predicted as normal.
- This changed our strategy for the upcoming models. We increased the complexity of the models to improve of feature engineering.

## CONCLUSION:

- Among all base models, VGG-16, Resnet-50 models, **Model-7: VGG-16** model did the best giving maximum accuracy, least loss, with model that's neither over-fitted not under-fitted like others and performing consistently over multiple runs.
- It also predicted the correct medical condition of COVID-19, Viral pneumonia or Normal more correctly among all the 10 models that we designed.