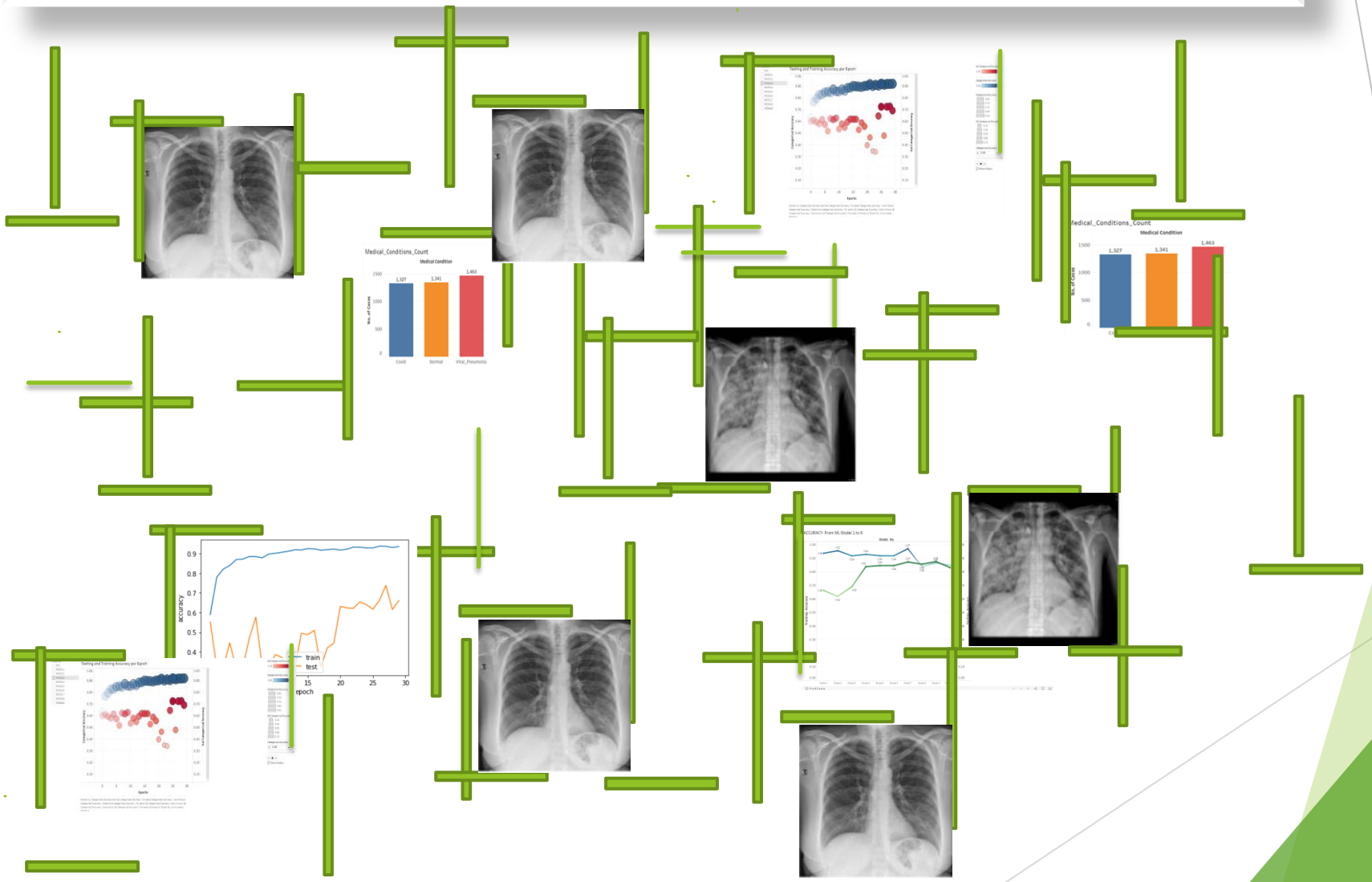
COVID PREDICTION ML MODEL

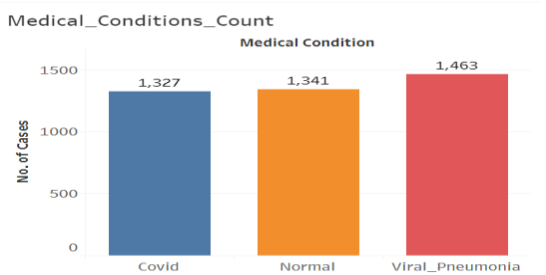


PROJECT OVERVIEW:

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.

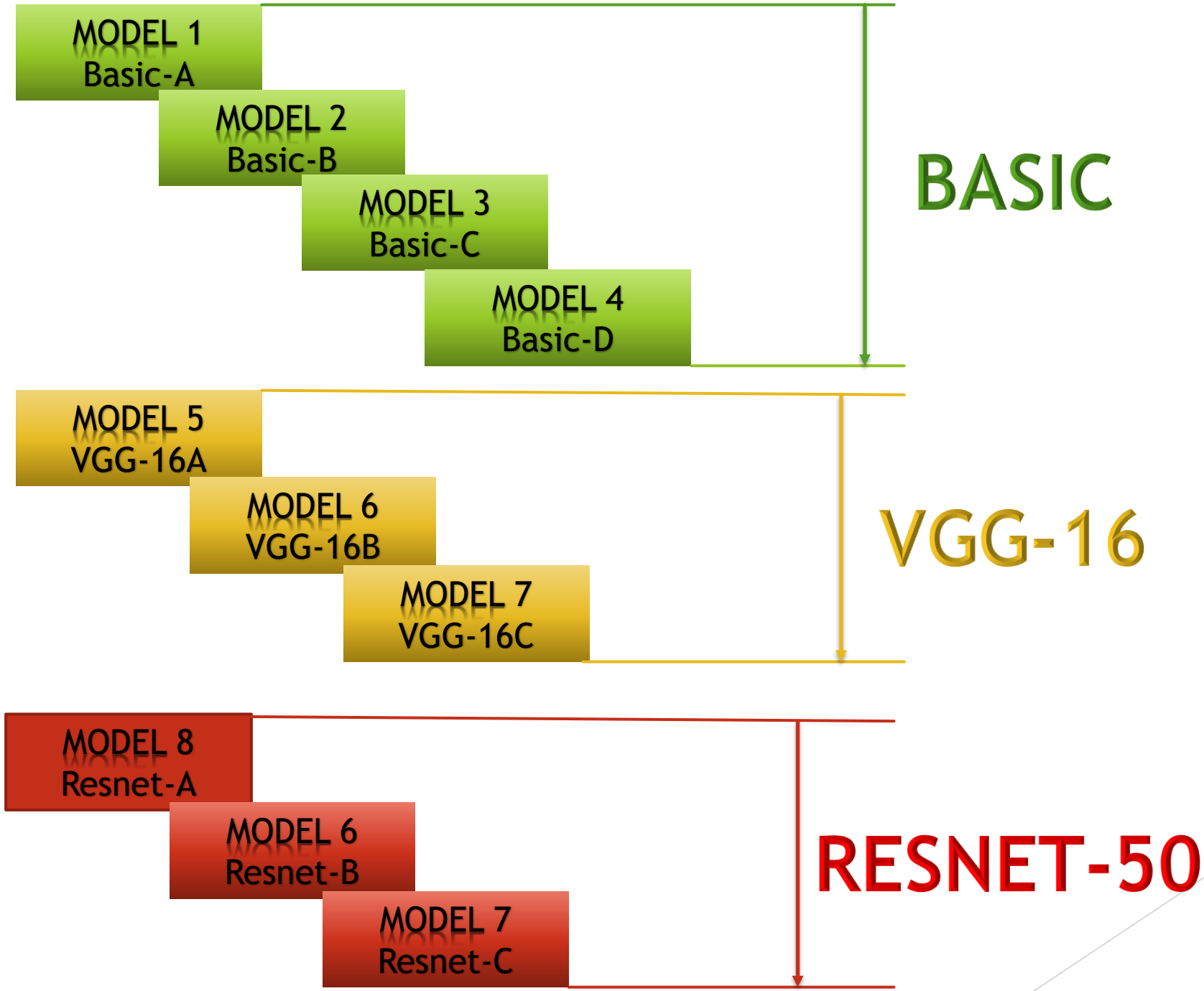
DATASET:



- **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:

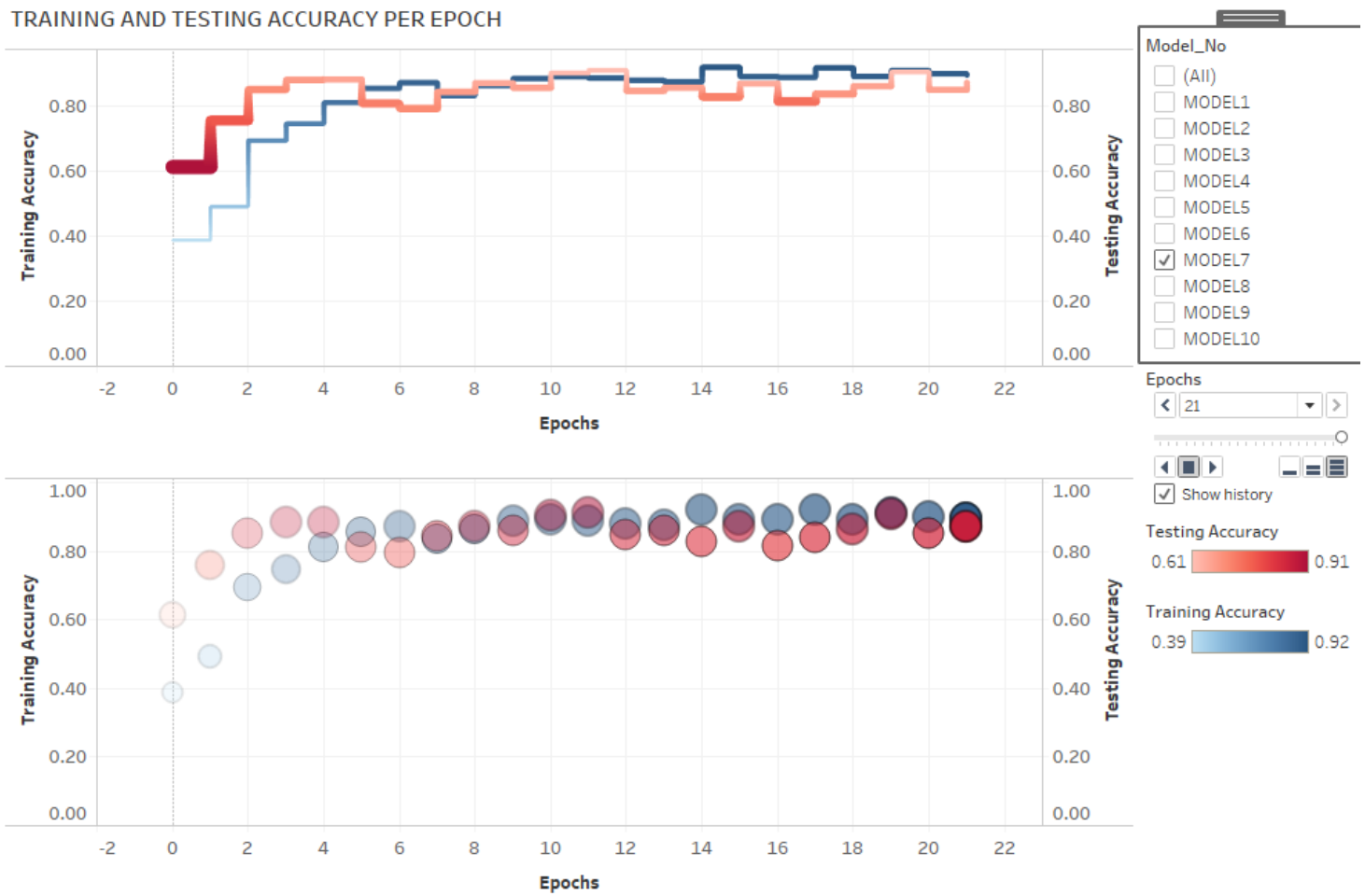
- We started with the basic model that was just a little overfitted with excellent training accuracy. The next step was to slightly modify the model each time, so it can generalize better.
- If we got a model with a better testing accuracy than its training accuracy, it would indicate that the model 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.


MODEL 1
Basic-A
MODEL 2
Basic-B
MODEL 3
Basic-C
MODEL 4
Basic-D
BASIC
MODEL 5
VGG-16A
MODEL 6
VGG-16B
MODEL 7
VGG-16C
VGG-16
MODEL 8
Resnet-A
MODEL 6
Resnet-B
MODEL 7
Resnet-C
RESNET-50

ML MODEL PERFORMANCE SUMMARY:

Model. No.	Basic Model Name	No of Epochs	File_Name	Loss	Categorical_ Accuracy	Val_Loss	Val_ Categorical_ 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 ACCURACY FOR ALL MODELS:

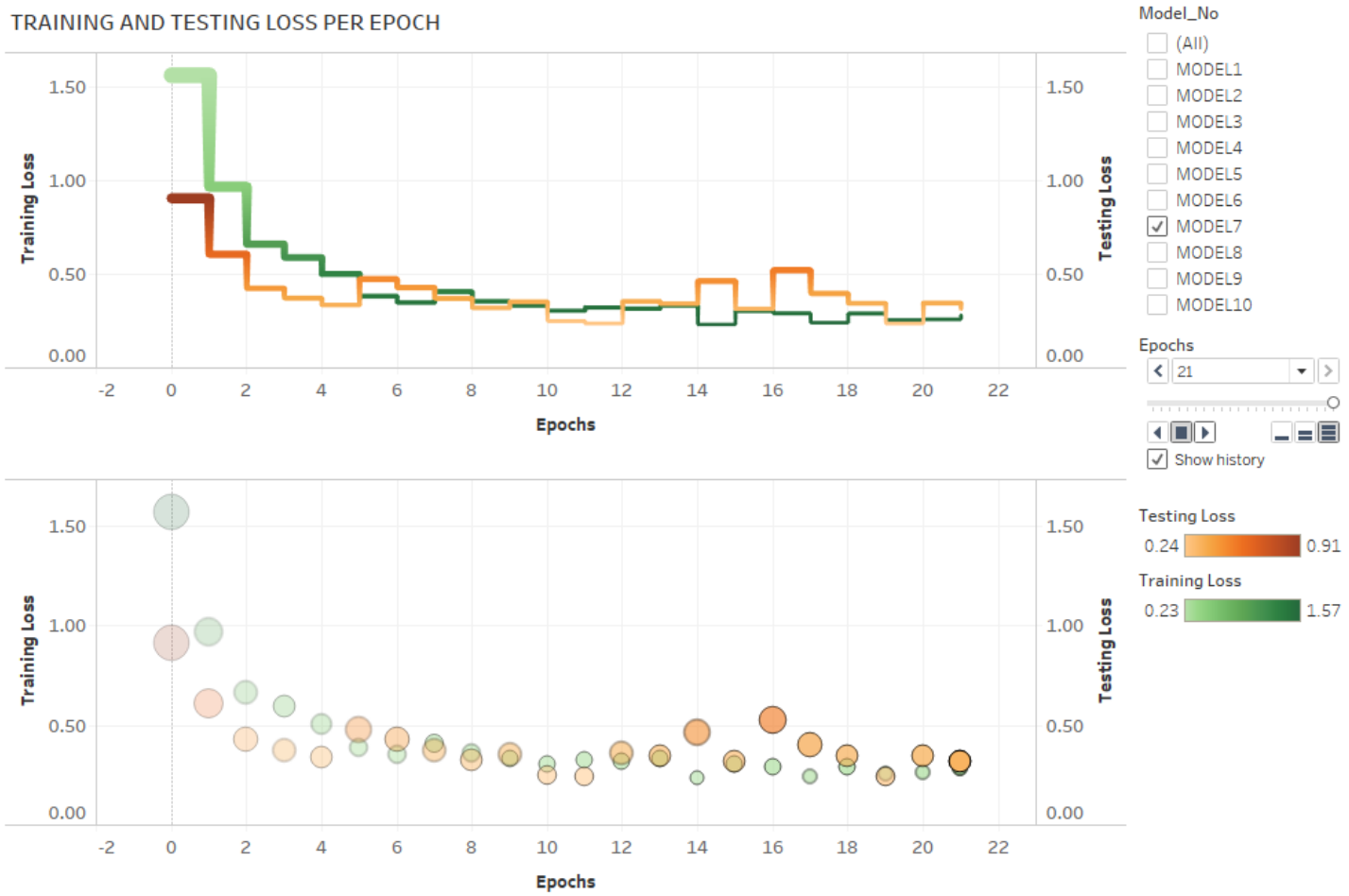


Note:

- 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/ACCURACY?publish=yes

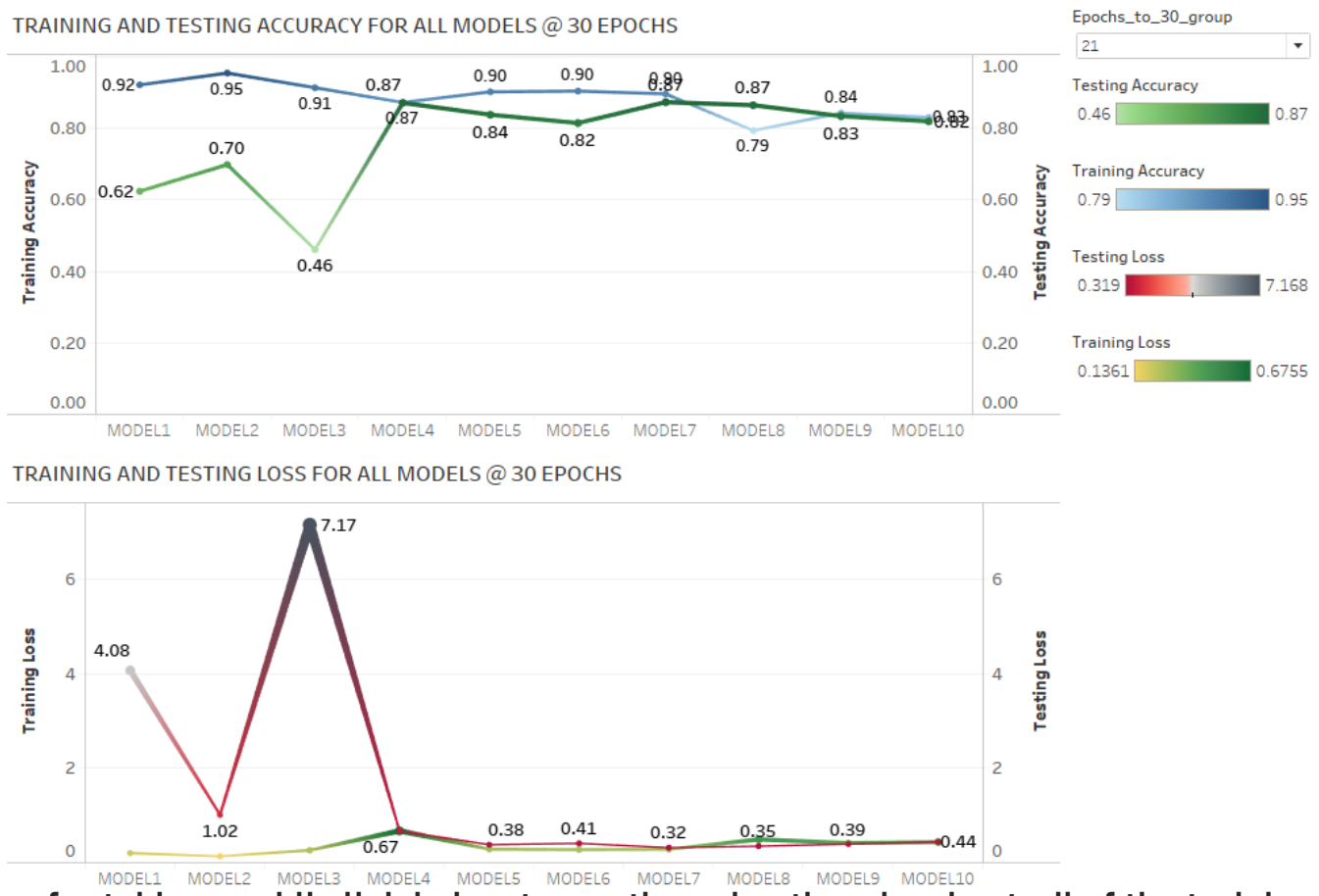
TABLEAU REPRESENATION OF LOSS FOR ALL MODELS:



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



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/ACCURACY?publish=yes

WINNER

VGG-16

RESNET-50

BASIC

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, 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.