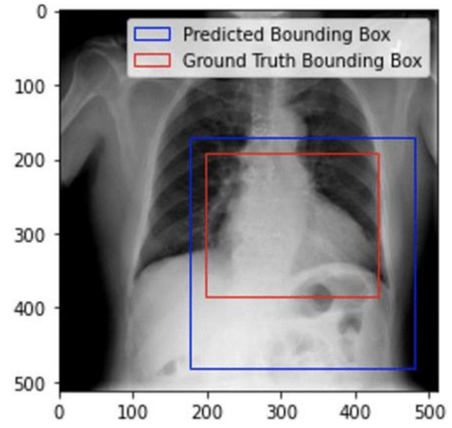
Cyclical Learning Rates (CLR’s) for Improving Training Accuracies and Lowering Computational Cost

Rushikesh Chopade, Aditya Stanam, Anand Narayanan, & Shrikant Pawar

Indian Institute of Technology, University of Iowa, Yale University, & Claflin University

Image classification problems?

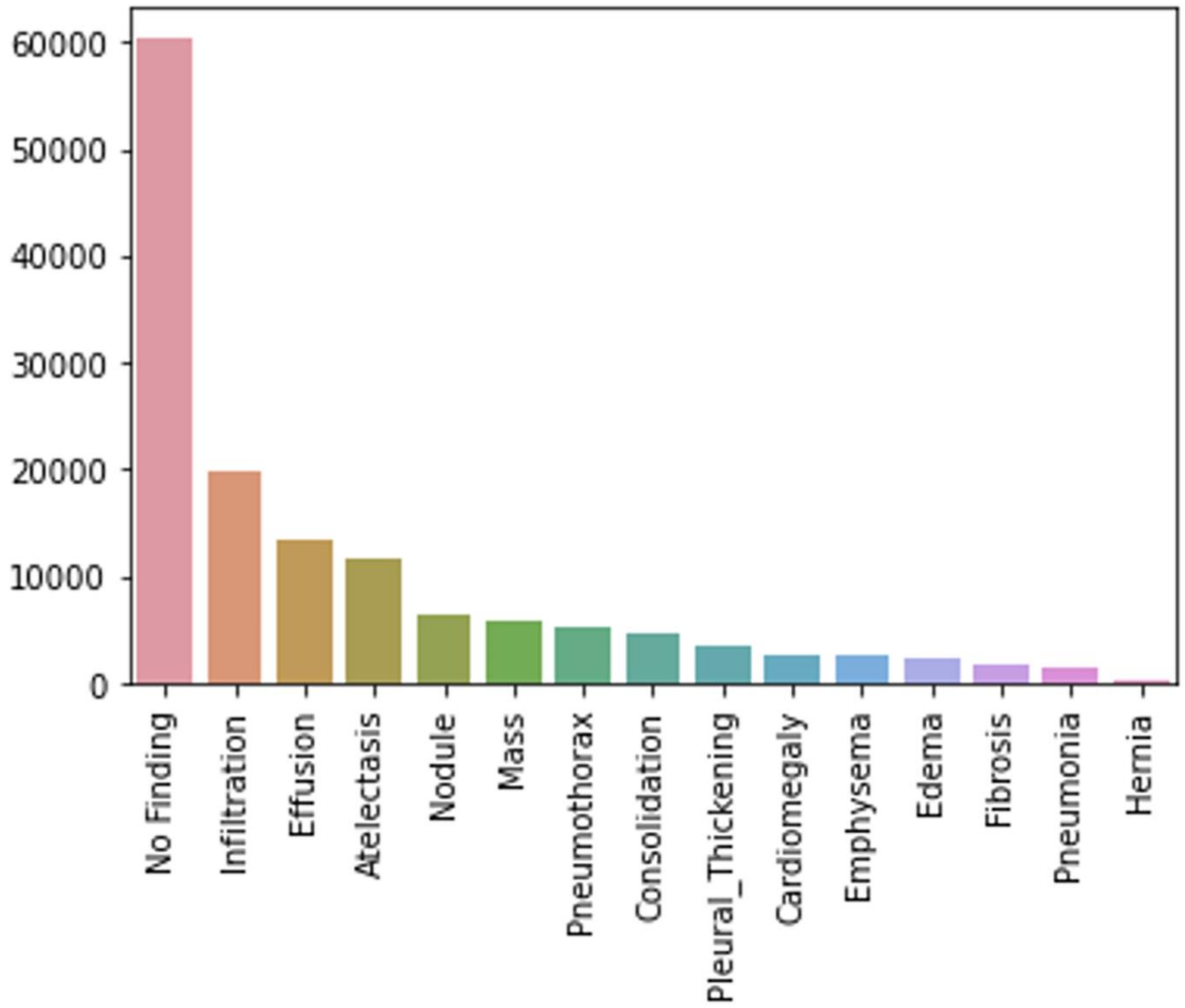
- Prediction of different lung pathologies using chest X-ray images is a challenging task requiring robust training and testing accuracies. In this article, one-class classier (OCC) and binary classification algorithms have been tested to classify 14 different diseases (atelectasis, cardiomegaly, consolidation, effusion, edema, emphysema, fibrosis, hernia, infiltration, mass, nodule, pneumonia, pneumothorax and pleural-thickening).
- We have utilized 3 different neural network architectures (MobileNetV1, Alexnet, and DenseNet-121) with four different optimizers (SGD, Adam, and RMSProp) for comparing best possible accuracies.



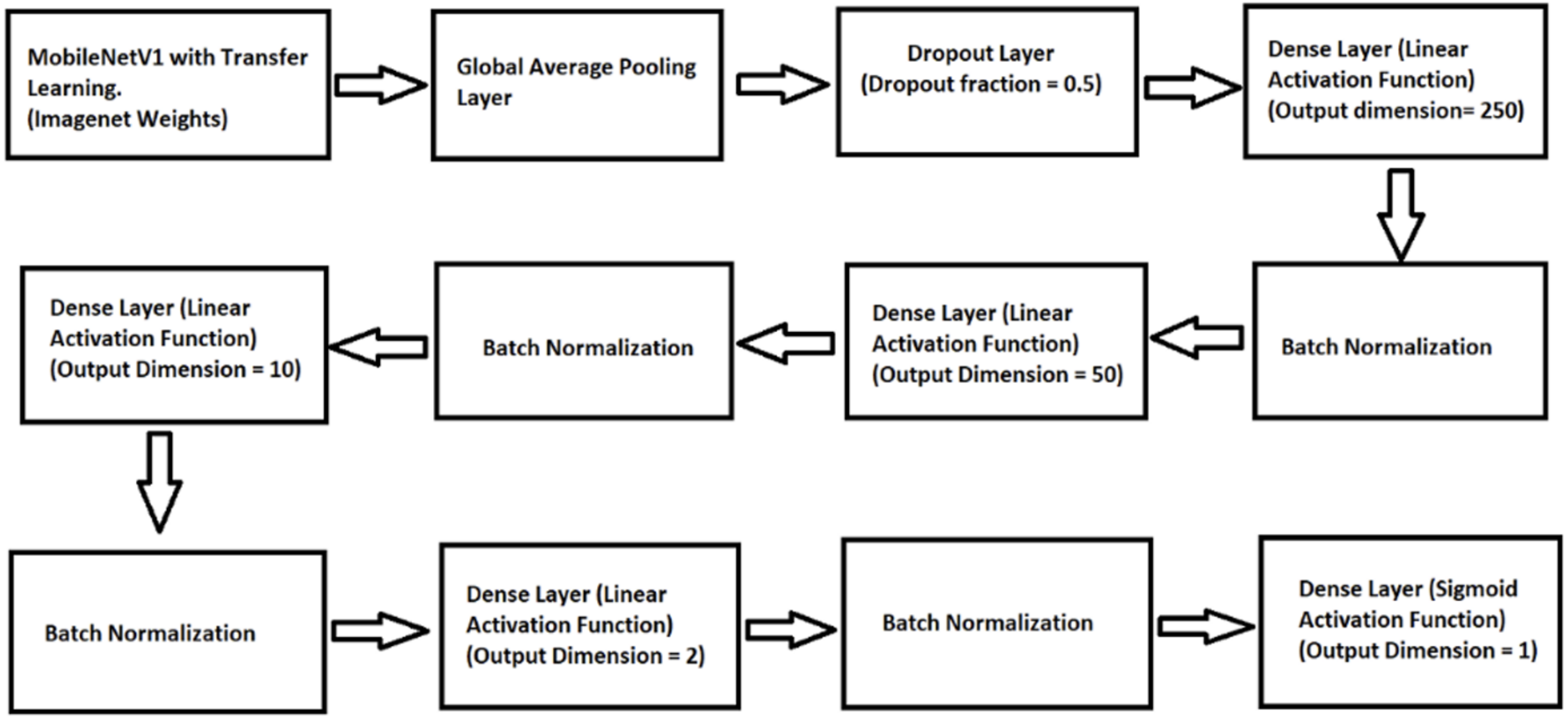
Cyclical learning rate (CLR)’s

- Cyclical learning rate (CLR), a tuning hyperparameters technique was found to have a faster convergence of the cost towards the minima of cost function.
- Here, we present a unique approach of utilizing previously trained binary classication models with a learning rate decay technique for retraining models using CLR’s. Doing so, we found signicant improvement in training accuracies for each of the selected conditions.

Number of top 15 unique labels



Model architecture used for all the binary classifiers

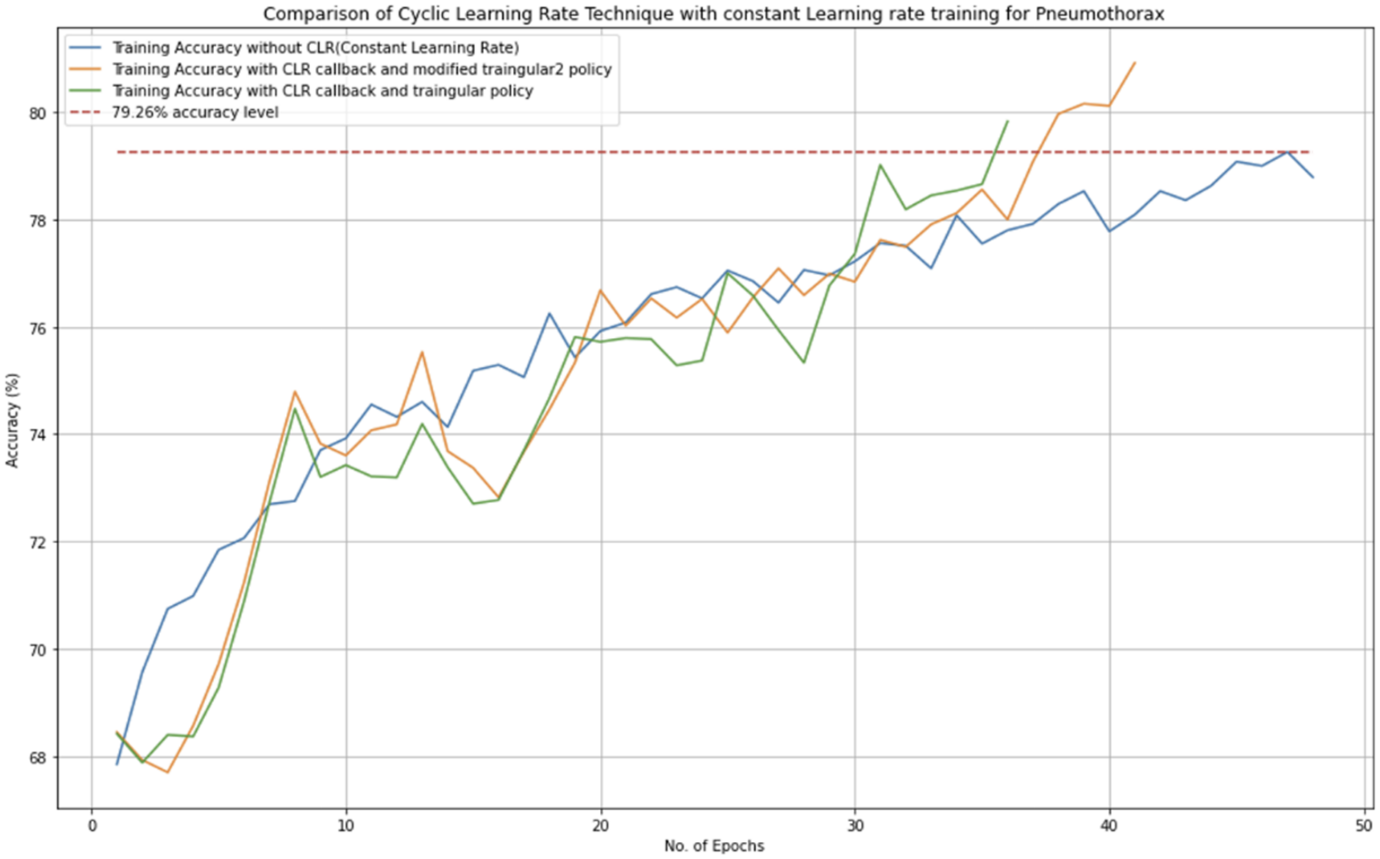


Accuracy comparisons for binary classifiers

Binary Classifier	No. of Epochs	Accuracy (in %)
Atelectasis	10	75.10
Cardiomegaly	12	75.78
Consolidation	10	73.32
Edema	12	93.37
Emphysema	10	85.60
Effusion	10	86.53
Fibrosis	10	66.58
Infiltration	10	64.60
Mass	10	70.11
Nodule	10	68.23
Pneumothorax	10	70.12
Pneumonia (with CLR)	30	88.43
Pleural Thickening	10	71.67
Hernia	30	90.81

Binary Classifier	Accuracy before CLR application (in %)	Epochs taken to achieve the accuracy before CLR application	Accuracy after CLR application (in %)	Epochs taken to achieve the accuracy after CLR application	Policy Used
Atelectasis	75.10	10	79.59	32	Triangular
Infiltration	64.6	10	76.15	10	Modified Triangular2
Fibrosis	66.58	10	88.96	32	Modified Triangular2
Pneumothorax	70.12	10	79.83	36	Triangular
Pneumonia	-	-	88.43	30	Triangular

Accuracy plot for “Pneumothorax” binary classier with constant learning rate, CLR with “triangular” and CLR with “modified triangular2” policies.



Future scope

- Primarily, we found that there are two main advantages of training with CLR’s over constant learning rates, with decay learning rates the model can get stuck into the saddle points or local minima due to low learning rates, and secondly CLR’s reduces the effort of choosing an optimal learning rate by hit and trial method.
- Poor choice of initial learning rate can make the model circle infinitely. In setting a learning rate, there is a trade-off between the rate of convergence and overshooting, a high learning rate will make the learning jump over minima but a too low learning rate will either take too long to converge or get stuck in an undesirable local minimum.
- Tuning the batch size hyper-parameter for adjusting learning rates have also been shown to improve learning accuracy.
- Some hyperparameter tools like Hyperopt, SMAC, and Optuna, using grid search, random search and bayesian optimization have been seen ecient in tuning batch sizes.