Deep learning based diagnosis of COVID-19 using chest CT-scan images

1st Talha Anwar
Department of Computer Science
National University of Computer and Emerging Sciences
Islamabad, Pakistan
chtalhaanwar@gmail.com

2nd Seemab Zakir
Department of Biomedical Engineering
Riphah International University
Islamabad, Pakistan
seemabzakir2@gmail.com

Abstract—**The Coronavirus disease (COVID-19) is an infectious disease that primarily affects lungs. This virus has spread in almost every continent. Countries are racing to slow down the spread by testing and treating patients. To diagnose the infected people, reverse transcription-polymerase chain reaction (RT-PCR) test is used. Because of colossal demand; PCR kits are under shortage, and to overcome this; radiographic techniques such as X-rays and CT-scan can be used for diagnostic purpose. In this paper, deep learning technology is used to diagnose COVID-19 in subjects through chest CT-scan. EfficientNet deep learning architecture is used for timely and accurate detection of coronavirus with an accuracy 0.897, F1 score 0.896, and AUC 0.895. Three different learning rate strategies are used, such as reducing the learning rate when model performance stops increasing (reduce on plateau), cyclic learning rate, and constant learning rate. Reduce on plateau strategy achieved F1-score of 0.9, cyclic learning rate and constant learning rate resulted in F1-score of 0.86 and 0.82, respectively. Implementation is available at github.com/talhaanwarch/Corona_Virus/tree/master/CT_scan**

Index Terms—**COVID-19, deep learning, classification, CT scan images**

I. INTRODUCTION

SARS-CoV-2 pandemic has affected millions of people worldwide. To date, several million people have gone to the valley of death and many more are getting affected by it on daily bases. Hundreds of subjects don't get diagnosed with COVID-19 at their early stage due to the shortage of testing equipment. When these undiagnosed early stage subjects with no physical symptoms come in contact with the healthy individuals, they transfer the virus to healthy people and this chain continues. The current approach to test COVID-19 is by using reverse transcription-polymerase chain reaction (RT-PCR). Patient's specimen is collected through subject's nose or throat and then sent to the laboratories for diagnostic purpose. This process is time consuming and due to its huge demand, there is a shortage of these kits. Round the globe, researchers are trying to find alternative techniques to diagnose coronavirus in affected people. Radiological equipment such as X-ray and CT-scan came up as potential alternatives for COVID-19 diagnosis.

In a report consist of correlation of CT-scan and PCR for diagnosis of COVID-19, it is found that CT-scan can be used to diagnose COVID as it is more sensitive in diagnosis of virus than PCR. The study comprised of 1014 subjects and 59% were PCR positive and 88% had CT-scan positive results [1]. This is why we discussed scope of CT-scan for diagnosis of COVID-19 in this paper. Zhao *et al.* [2] collected the CT-scan images both for COVID and non-COVID subjects and He *et al.* [3] used self-supervised learning along with transfer learning to achieve an F1-score of 0.85. Wang *et al.* [4] trained the deep learning model on CT-scan images and resulted in an F1-score of 0.77. Polsinelli [5] proposed a light convolution neural network for medium specs system to diagnose COVID-19 from chest CT-scans. An accuracy, recall precision and F1-score of 0.83 and 0.85, 0.817, 0.83 respectively was achieved by light CNN. Amyar *et al.* [6] achieved 0.86 accuracy and 0.93 AUC score while classifying COVID CT-scan images. Another study compared 3D convolution neural networks and found that they worked well as compared to 2D. Authors claimed that DenseNet3D121 achieved highest accuracy of 88.63% and F1 score of 88.14% [11]. Zheng *et al.* [12] proposed a weakly supervised 3D deep neural network to predict COVID-19 from chest CT-scan. Data was comprised of 313 COVID patients and 229 non-COVID patients. They achieved accuracy of 90% but positive predicted values (recall) is 84%. Mobiny *et al.* [13] used detail oriented capsule networks to identify corona virus from CT-scan images and an accuracy and F1 score of 0.87 is achieved. A.K.Mishra *et al.* [14] used ensemble of VGG16, inceptionV3, ResNet50, DenseNet121, DenseNet201 and achieved accuracy of 88.34% and F1-score of 86.7%. Mohammed *et al.* [16] proposed weekly supervised approach ResNext+ that require volume level input and resulted in slice level prediction with F1-score of 81.4%. Researcher also tried to diagnose COVID patients from community acquired pneumonia (CAP) cases using CT-scan images. Sun *et al.* [15] proposed Adaptive Feature Selection guided Deep Forest algorithm for this purpose. They used 1495 CT-scan images of COVID patients and 1027 images of CAP cases. They extracted volume, infected lesion number, histogram distribution,surface area, radiomics features, age and gender as features and used adaptive feature selection guided deep forest as classifier resulted in an accuracy of 91.79% and F1-score of 93.07%.

Table I shows accuracy and F1-score of previous work compared with this study.

In this work, we used EfficientNet architecture and per-

Author	Number of Images		Accuracy	F1-score
	Covid	Non Covid		
Xuehai He* [3]	349	397	0.86	0.85
Wang [4]	325	740	0.89	0.77
Polsinelli* [5]	360	397	0.83	0.83
Amyar* [6]	347	397	0.86	
Xin He [11]	1515	965	0.88	0.88
Zheng [12]	313	229	**0.9**	
Mobiny* [13]	349	397	0.87	0.87
Mishra* [14]	360	397	0.88	0.86
Mohammed [16]	3520	19,353	0.77	0.81
Our Work*	351	396	**0.9**	**0.9**

formed 5 fold cross-validation strategy to predict the test data in each fold. The test predictions of each fold are averaged and evaluated against the ground truth. Test images are the same as used by some other researchers [3], [5], [6] and [13]. The paper is organized as follows. Section 2 presents details of the dataset used and the methodology for the classification. Section 3 deals with the results in detail. Section 4 concludes the paper and presents some ideas for future work.

II. METHODOLOGY

A. Dataset

Dataset comprises of 349 CT scan images collected from 216 COVID patients and 396 images from non-COVID patients. [2]. Training data consists of 253 COVID images and 291 non-COVID images, whereas test set comprises of 98 COVID and 105 non-COIVD CT-scan images. As the dataset is limited, no separate validation-dataset is used in this study. Instead, 5 fold cross-validation strategy is applied such that test data is predicted in each fold. After training; test predictions of each fold are averaged and evaluated against the ground truth.

As small dataset for training may lead to over-fitting, so the dataset is augmented to increase the number of images, along with transfer learning. Transfer learning can be domain specific such as training on other CT-scan image dataset or domain generic such as training on very large image dataset containing millions of images and hundreds of classes such as Imagenet dataset [7]. In this approach, domain generic transfer learning is used, because to date; no other COVID CT-scan dataset is available publicly. We don't want to train on healthy CT-scan images to avoid pre-learning on normal CT-scan images which may lead to weight shifting towards the normal class. For data augmentation, different transformation techniques are used to synthesize data such as:

- Flip: Images are flipped horizontally and vertically.
- Rotate: Images are rotated by an angle
- Shift: Images are shifted left, right, upward and downward

After transformation, these images are considered as separate images. Only flipping is performed in this study as it yields much better results. Rotation mostly works better in circular images, and shifting may result in loss of some information.

	COVID	Non-COVID	Flipping	Total
Train	253	291	x2	1088
Test	98	105	x2	406
Total	351	396		1494

Table II shows the image distribution in training and testing set. After horizontal and vertical flipping there are 1088 images for training, among 544 are original images and 544 augmented images. Similarly, flipping results in 203 more test images.
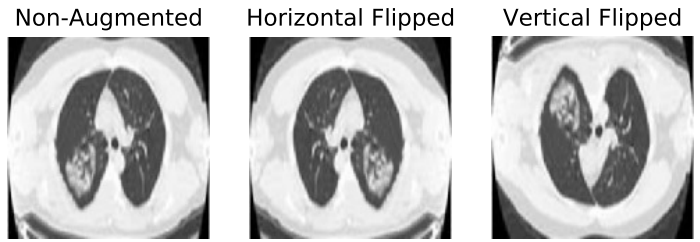
COVID CT scan images

Fig. 1. Non-Augmented, horizontally and vertically flipped augmented COVID CT-scan images
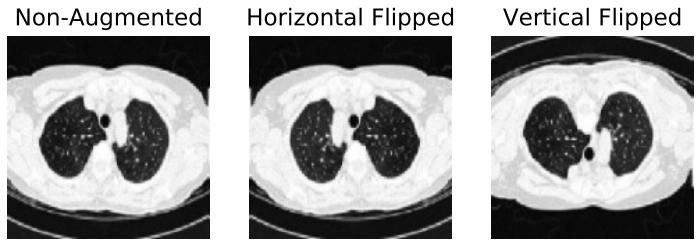
Non-COVID CT scan images

Fig. 2. Non-Augmented, horizontally and vertically flipped augmented COVID CT-scan images

Fig. 1 and 2 shows original non augmented and augmented COVID and non-COVID images of chest CT-scan, respectively.

B. Deep learning architecture

EfficientNet family of deep learning architecture is known for their smaller number of parameters and nearly state of the art performance on ImageNet benchmark dataset [9]. EfficientNet B4 is used as deep learning architecture for this study. This model has only 19 million parameters and it surpassed the other models having 7 times more parameters such as SENet with 146 million parameters, on ImageNet benchmark. The model is trained on google colab. Batch size of 16, image size of 348x348 and adam is used as optimizer.

C. Learning Rate

Learning rate plays an important role in model convergence leading to better performance. We tried a fixed learning rate of 0.0001. Apart from the fixed learning rate, cyclic learning rate [8] and learning rate reduced on plateau strategy is also used. In cyclic learning rate, model learning rate cycles between an upper and lower bound while training. We used triangular2 strategy, in which learning rate becomes half of the previous triangle, as shown in fig. 3. While in reduce on plateau, model learning rate decreases when model performance stops improving. This helps to take small steps when the model is near convergence. Cyclic learning rate is independent of validation data, while reduce on plateau considers model performance on validation data by decreasing the learning rate, as shown in fig. 4.
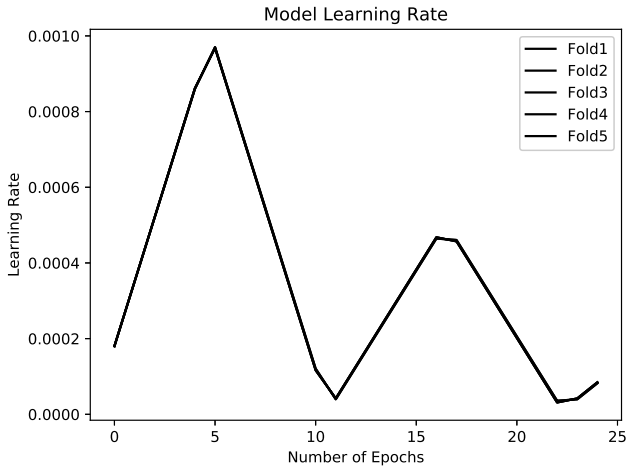

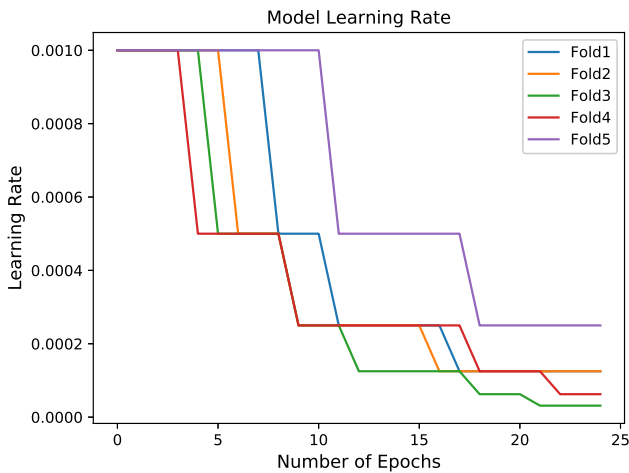Fig. 3. Cyclic learning rate triangular2 policy for 5 fold cross validation

Fig. 4. Reduce on plateau learning rate for 5 fold cross validation

D. Loss Function

To minimize the error; instead of binary cross-entropy, focal loss function [10] is used. Focal loss function puts less weights for the easily classified examples and gives more importance to hard examples thus increasing the probability of correct predictions for difficult examples.

E. Evaluation Criteria

We used accuracy, precision, recall, F1 score and AUC score as the evaluation criteria. Accuracy is the total number of correct predictions out of all predictions. Recall depicts out of total actual coronavirus cases; how many are identified as coronavirus affected ? Precision is out of total subjects predicted as corona virus cases; how many actually are affected by coronavirus ? F1 score is harmonic mean between precision and recall. AUC is the area under the ROC curve and it provides an aggregate measure of performance across all possible classification thresholds. Accuracy, recall, precision and F1-score are shown in 1,2, 3 and 4 respectively.

$$Accuracy = \frac{TP + TN}{FP + TP + FN + TN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1 - score = \frac{2 * precision * recall}{precision + recall} \quad (4)$$
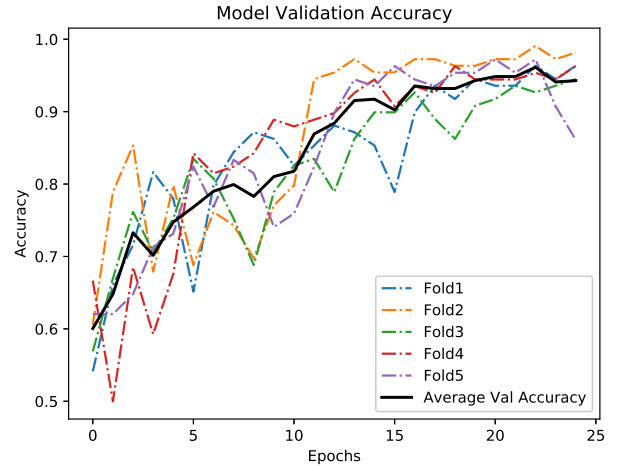
III. RESULTS

Fig. 5. Validation accuracy of model using reduce on plateau learning rate strategy

Fig. 5 shows 5 fold validation accuracy and average validation accuracy while using reduce on plateau strategy of learning rate. The 5 fold cross-validation accuracies are nearly stable because learning rate is reduced based on each fold validation loss; leading closer to global minima for that fold.
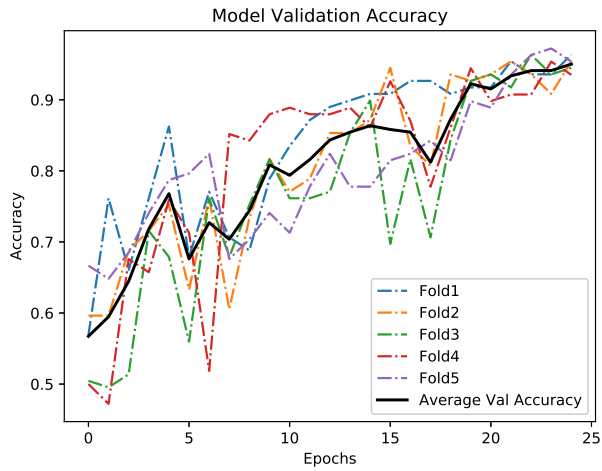
Fig. 6. Validation accuracy of model using cyclic learning rate

Fig. 6 shows that cyclic learning rate 5 fold cross-validation accuracies are more volatile than the plateau reducing learning rate, because cyclic learning rate is independent of validation split. In each fold, the learning rate cycles with out considering the validation data of that fold as shown in fig. 3.
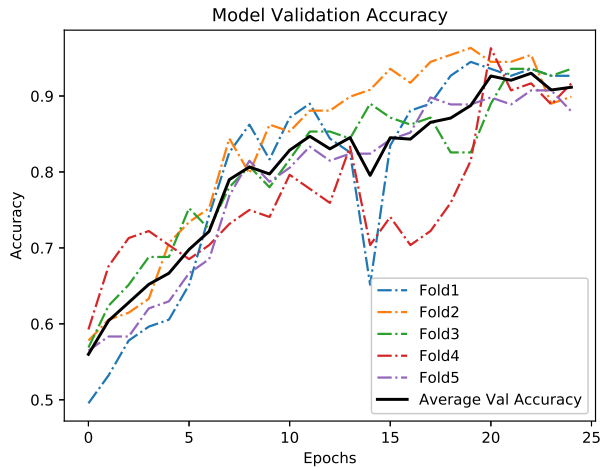

Fig. 7. Validation accuracy of model using constant learning rate

Fig. 7 shows that validation accuracy when model is trained with fixed learning rate. The validation accuracies are stable during initial epochs and then become unstable in later epochs in some folds.

Fig. 8 and fig. 9 shows that plateau based learning rate strategy achieved better results because the learning rate is reduced when validation performance stops improving, making the model to take smaller learning steps. Though the constant learning rate did not have fluctuations in the beginning as compared to plateau based learning rate and cyclic learning rate, but it converged less as compared to plateau based
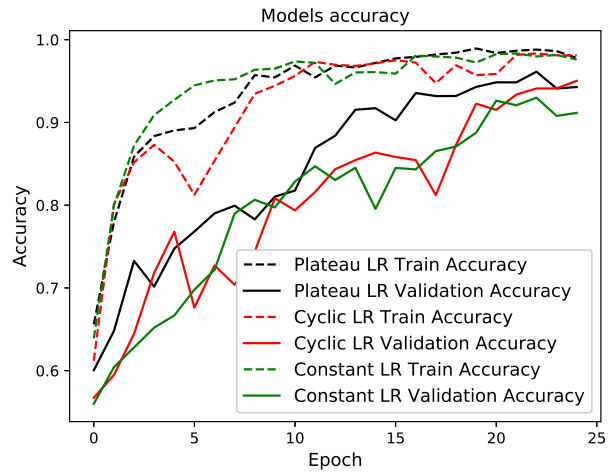

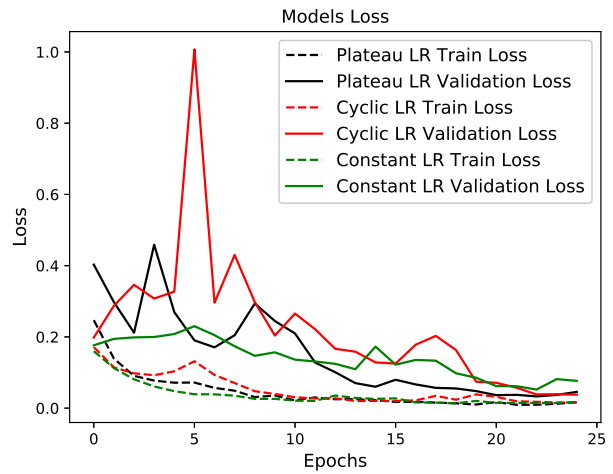Fig. 8. Comparison of average training accuracy and validation accuracy

Fig. 9. Comparison of average training loss and validation loss

learning rate strategy. Cyclic learning rate showed some sharp peaks in validation accuracy.

Table III shows precision, recall and F1-score for COVID and non-COVID classes separately. The highest precision of 0.97 in COVID class is obtained using cyclic learning rate strategy, while in non-COVID class 0.94 precision is obtained using reduce on plateau strategy. Highest recall in COVID

TABLE III
CLASSIFICATION PERFORMANCE COMPARISON ON TEST DATA BASED ON INDIVIDUAL CLASSES

LR Methodology	Classes	Precision	Recall	F1 Score
Plateau LR	COVID	0.94	0.84	0.89
	Non-COVID	0.86	0.95	**0.90**
Cyclic LR	COVID	**0.97**	0.73	0.84
	Non-COVID	0.80	**0.98**	0.88
Constant LR	COVID	0.94	0.68	0.79
	Non-COVID	0.77	0.96	0.85

	Plateau LR	Cyclic LR	Constant LR
Macro Average Precision	0.90	0.89	0.85
Macro Average Recall	0.89	0.86	0.82
Macro Average F1 Score	0.90	0.86	0.82
Accuracy	0.90	0.86	0.86
AUC score	0.90	0.86	0.82

class is 0.84; obtained by reduce on plateau strategy, whereas in non-COVID it is 0.98 by cyclic learning rate.

Table IV shows accuracy, precision, recall, F1 score and AUC score of the model trained on three different learning rate strategies i.e reduce on plateau, cyclic learning rate and constant learning rate. The best performance on test set is achieved by reduce on plateau strategy, resulting in 0.897, 0.896, 0.895 accuracy, F1 score and AUC score respectively. Though the cyclic learning rate, did not look much promising in validation graphs, achieved an accuracy, F1 score and AUC value of 0.86,0.86,0.86 respectively. Constant learning rate strategy performed worst on test dataset and an accuracy, F1 score and AUC value of 0.83,0.82,0.82 is achieved respectively.

IV. CONCLUSION

This paper used EfficientNet b4 to discriminate between COVID CT-scan images and normal CT-scan images. Different learning rate policies are studied, such as reduce on plateau, cyclic learning rate and constant learning rate. Instead of using binary cross-entropy, focal loss is used to tackle hard examples. Previous studies claimed highest accuracy of 0.86 and F1-score of 0.85 using this dataset. Our results with 0.90 F1 score and accuracy on the same dataset are significantly better. Future work includes training the model on the same dataset used in this paper and test on some other collected dataset.

REFERENCES

[1] Tao. Ai et al "Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases" 2020,Radiology, 200642

[2] J. Zhao, Y. Zhang, X. He, and P. Xie, "COVID-CT-Dataset: A CT scan dataset about COVID-19,"2020, arXiv:2003.13865,2020.

[3] X. He, X. Yang,S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P Xie. "Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans," 2020, medRxiv doi:10.1101/2020.04.16.20064709

[4] S. Wang et al "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," 2020, medRxiv, doi: 10.1101/2020.02.14.20023028.

[5] M. Polsinelli, L Cinque and G. Placidi. "A Light CNN for detecting COVID-19 from CT scans of the chest," 2020, Pattern recognition letters vol. 140: 95-100. doi:10.1016/j.patrec.2020.10.001

[6] A Amyar, R Modzelewski, S. Ruan. "Multi-task Deep Learning Based CT Imaging Analysis For COVID-19 Pneumonia: Classification and Segmentation."2020, Computers in Biology and Medicine.

[7] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[8] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, 2017, pp. 464-472, doi: 10.1109/WACV.2017.58.

[9] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in The International Conference on Machine Learning (ICML), 2019.

[10] T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollar. Focal loss for dense object detection. IEEE ICCV, 2017.

[11] X He, et al. "Benchmarking Deep Learning Models and Automated Model Design for COVID-19 Detection with Chest CT Scans," 2020, medRxiv. doi:10.1101/2020.06.08.20125963

[12] C. Zheng et al. "Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label," 2020, medRxiv. doi:10.1101/2020.03.12.20027185

[13] A. Mobiny et al."Radiologist-Level COVID-19 Detection Using CT Scans with Detail-Oriented Capsule Networks," 2020,arXiv:2004.07407.

[14] A.K. Mishra, S.K. Das, P.Roy and S. Bandyopadhyay 2020. Identifying COVID19 from Chest CT Images: A Deep Convolutional Neural Networks Based Approach. Journal of Healthcare Engineering, 2020, pp.1-7.

[15] L. Sun et al., "Adaptive Feature Selection Guided Deep Forest for COVID-19 Classification with Chest CT," in IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/JBHI.2020.3019505.

[16] A. Mohammed et al., "Weakly-Supervised Network for Detection of COVID-19 in Chest CT Scans," in IEEE Access, vol. 8, pp. 155987-156000, 2020, doi: 10.1109/ACCESS.2020.3018498.