Data analysis of COVID-19 pandemic using CT images

Aravind Jadhav[a], Dr Sanjay Pujari [b]

[a] Assistant Professor, Department of Electronics and Communication Engineering, Angadi institute of technology and management Belagavi, Karnataka, India, jadhav.aravind@gmail.com

[b] Professor, Department of Electronics and Communication Engineering, KLE College of Engineering & Technology,Chikodi,Belagavi, , Karnataka, India, pujarisap@gmail.com

Abstract

The COVID-19, infectious disease which is caused by the new virus called corona virus [SARS-COV-2], majorly affects the lungs which can be identified by the CT scan of the corona affected patients. We have collected for about 353 lung CT frames from each COVID-19 patient. it is about 369 non COVID-19 CT frames for the purpose of testing and training which gives the best identification technique in this pandemic situation. The identification technique we have introduced in this paper is data augmentation technique which gave best results that will be discussed here in further. From the collected data, we have utilized 75% of the lungs CT frames for training and another 25% frames for difficult attributes. This research paper encompasses of the results specified on CT images of corona complete, improved and loss problems which helps in comparative analysis. So, this comparative analysis of CT images is the illustration investigative statistics examination.

Keywords: COVID-19 lung CT scan, Data augmentation, performance metrics

DOI Number: 10.14704/NQ.2022.20.12.NQ77292 **NeuroQuantology2022;20(12): 2957-2964**

I. INTRODUCTION

The COVID-19 has majorly affected the health which may be mild to serve attacks to respiratory system of a human body. As of now, for the screening of COVID-19 attack, the best tool used is Compound Tomography. Thus, from the observation of chest CT reports of COVID-19 patients, there will be a massive pathological change in the lungs of a body and severely damages the tissues of the affected organ. The core symptoms of this COVID-19 are fever, cold and cough [3].

A real-time in validate dictation translation sequence response [RT-PCR] is one of the nuclear-derived laboratory technique to identify the severe acute respiratory infection in human body, rapidly by the occurrence of detailed inherent objects within several pathogen which includes virus also [4]. If the infection was detected in early stages, CT of the lungs along with the model of deep learning will be the best solution. Depending upon the technical skills of the lab technician, the accuracy of the ultrasound images can be analyzed. But, the internal density, size and texture at different angles and planes can be analyzed by height resolution images of CT scan reports.

The concept called "Ground-Glass Opacity (GGO)"is adopted to detect the lung infection which relates radiological examination like slender slice of lungs CT images shows the accurate internal structures of tissues without overlapping.

There are many different techniques and perspectives to analyze the huge dataset. From the CT image of the COVID patient, there is a need to calculate the length, min-max and area of the infected lungs. The satisfied data analytics techniques are required to analyze this huge dataset for future predictions. Thus, one of the software called "Radiant software of DICOM viewer, is utilized to test different CT images. In this research work, it mainly focuses on the data analytics for the dataset of corona affected COVID-19 infection. By using the transfer learning approach on lungs CT scan, differentiating between COVID and non-COVID affected patients.

II.METHODOLOGY

A. DATASET COLLECTION:

The main aim of this paper is to introduce a calculation mold based on COVID-19 lung CT scan. Thus, firstly we have collected 369 CT scan images from non-COVID and 353 COVID CT scan images from 216 COVID patients [2]. Totally for the purpose of training, dataset of 1342 COVID images and dataset of 691 non-COVID images were available. And for the purpose of testing, dataset of 498 COVID and dataset of 705 non-COVID images were available. To be noted is, for studying there is no requirement of separate validation- dataset, as there is a limited dataset. Thus, instead of this, another strategy called double annoyed legalization is functional, so that examination information can be predict in every of the double. The next process after preparation is; evaluation of averaged data of every prediction test with respect to the ground truth, by using two datasets, 3000 images are prepared and thereafter divided to two parts for training 1800 images and for testing 1200 images.

eISSN 1303-5150

it is necessary to the number of images and valid dataset is required to predict the disease accurately, because small dataset may lead to over-fitting. a new technique introduced called learning of transfer the images from the consideration of large amount of dataset using images of CT having different classes by considering many images in the ImageNetdataset [7]. the advantages of new technique is mainly without training the images of CT possible to predict the problem. There are many transformation techniques to synthesize data for augmentation, like;

i. The flip: here the descriptions are rotate straight and also up and down.

ii. The turn: at defined angle descriptions be turned.

iii. The move: here the pictures are shifted left, right, upward and downward

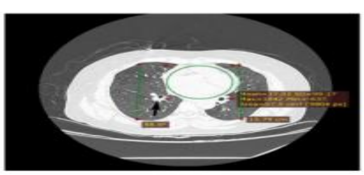

Figure 1 : axial lung CT appeared lesions in the right lung

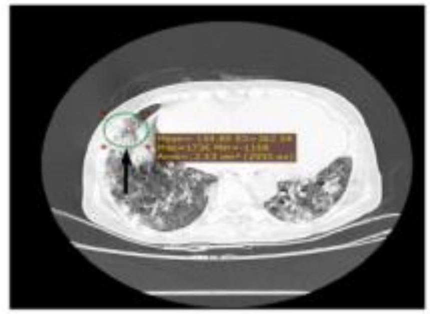

Figure 2: axial lung CT appeared lesions in the left lung

The figure 1 shows the results of image processing and also features analyzing COVID-19 which is depicting 91.67% of accuracy as depicted in figure 2.

Table 1: quantity of pictures in training and valid testing data set

	Covid	Mean	Max	Area
Image 1	253	37.02	1842	57.5cm
Image 2	98	154.69	1736	2.63cm
Total	351	191.71	3578	60.13

The dataset of image distribution in testing and training is as shown in table 1. Then flipping of images in horizontal and vertical way is performed wherein we will get 1088 descriptions for valid training, between which 544 are valid pictures there will be about 544 augmented images. Such that, this flipping of images results in another 203 more test images. The obtained images are considered as separate images after the transformation. As it fetches better results, only the flipping is carried out. Here, the rotation of images is best suitable for circular images, and shifting can cause some loss of information.

B.DATA AUGEMENTATION:

The image augmentation techniques are not a best idea for our dataset as the obtained dataset is not sufficient, means still bigger dataset is required. Thus, using this technique, our collected dataset of images becomes superior exclusive of receiving any other novel descriptions. Henceforth 1744 lung CT scanned images of COVID 19 was obtained from augmentation method. The figure 3 shows, class wise distribution of frames for training and testing process. And for training, it was observed that 1273 non COVID CT frames and 1392 COVID-19 CT frames and for testing 315 non COVID-19 CT frames and 352 COVID 19 CT frames are observed. In accordance with the width dimensions of the images, the transformation is composed of rotation, horizontal and vertical translations, scaling and shearing. The brightness and contrast have been adjusted using color jittering. Table II shows, the parameters that we have used in image augmentation process.

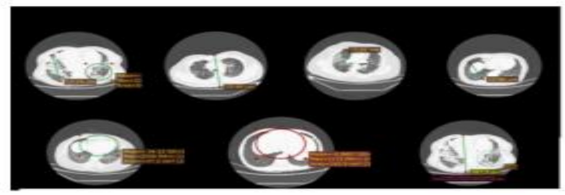

Figure 3: axial lung CT appeared in different COVID positive images

The figure 3 is the image of CT scan of COVID- 19 and intestinal lung diseases to the left and right column respectively. The one which is indicated with red arrows of eclipse are the main lesion regions which are having interclass similarity and intra-class variation. This is the main challenge in COVID-19 screening task. And also, the lung image which is infected with pneumonia contains a very large sector of non-lesion regions that is having a complicated variation of tissues.

Table II: Image for augmentation

Sl No	Parameter	Value
1	Rotation range	22
2	Zooming	0.4
3	Width shifting	0.3
4	Height shifting	0.3

eISSN 1303-5150

5	Shearing	0.2

Table II shows the different parameters like rotation range, zooming, width shifting and height shifting and shearing.

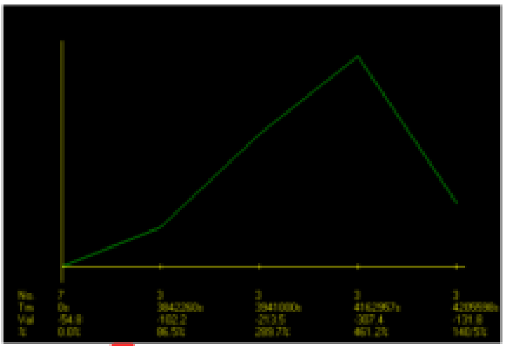

Figure 4: graphical representation of lung CT images of various subjects

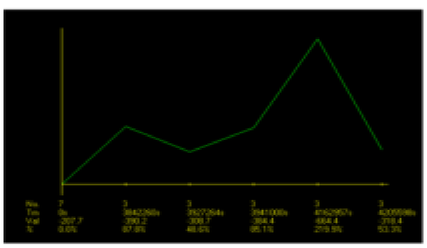

Figure 5: graphical representation of lung CT images of various subjects

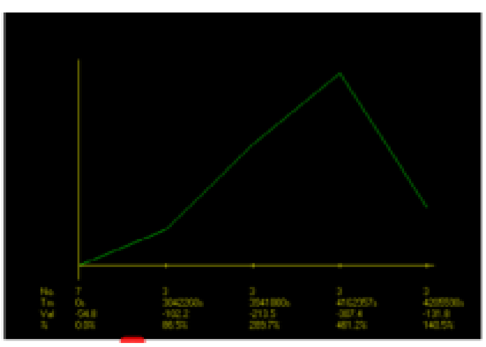

Figure 6: graphical representation of lung CT images of various subjects

The time intensity curve is as shown in above figure 4 to 6. These curves allow us to visualize the lesions clearly through scheming the signal strength standards over time after the organization of difference fabric. Many forms of curves such as straight, curved and plateau can be obtained. Next, we performed normalization where CT images to grey-scale and further pixel values are re-scaled from 0 to 1. Due to the perilous problem, we have to set the input CT image's size to 32 X 32 X1, thus we can accelerate our model.

III. EVALUATION CRITERIA

For the estimate of valid parameters, we have considered accuracy, recall, precession, AUC score and F1 score. In all these predictions "Accuracy" is the total number of corrected predictions. Between precision and recall, F1 score is considered as harmonic mean. And AUC is considered as the area under ROC curve which gives the combined measurement of presentation across every the potential categorization thresholds. In equation 1, 2, 3 and 4, accuracy recall, precession and F1 score are shown respectively.

$$Accuracy = \frac{TP + TN}{FP + TP + FN + TN}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FN}$$

$$F1\ score = 2\frac{* Precision * Recall}{Precision + Recall}$$

We used Radiant licensed software to implement the proposed system.

Table III: categorization feat contrast on examination in formation based on entity classes.

LR Methodology	Cases	Precision	Recall	F1 Score
Platue 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

The identification of COVID 19 based on CT images which has reached maximum accuracy of 83.3%. By using different methodologies like classes in the COVID and Non- COVID, Precision, Recall and F1 score is measured. And by utilizing cyclic learning rate strategy, the highest precision of 0.98 is obtained in COVID class and by using reduces on plateau strategy, the highest precision of 0.93 is obtained in non-COVID class. Few researchers persuaded research on different medical images by deep learning and achieved accuracy of 82% [8] and precision of 79%. The proposed model overcome all the deficiency of all performance parameters.

Table iv: Model performance on test data using different learning rate strategies

	Plateau LR	Cyclic LR	Constant LR
Macro average precession	0.91	0.90	0.92
Macro average recall	0.92	0.88	0.89
Macro average F1 score	0.91	0.91	0.92
Accuracy	0.94	0.90	0.91
AUC score	0.95	0.92	0.93

The accuracy, precision, recall, F1 score and AUC score of the proposed model is tabulated in table IV. The most excellent concert on valid test set viz., accuracy 0.397, F1 score 0.896, AUC score 0.895. Constant learning rate strategy performed poor in accuracy.

The proposed method is tested for 1200 COVID-19 images. Where in which 1164 images are analyzed for positive and remaining images are tested for negative. Thus, this technique detection rate is 97% accuracy.

IV. CONCLUSION

The COVID-19 primarily affect the lungs and leads to breath problem and finally ends in death. This makes the changes in lungs CT scan images of affected patients also. We have analyzed the detection of COVID based on lungs CT scanned images. We obtained dataset that contain 349 COVID-19 lungs CT image frames. The selected information growth technique to differentiate COVID positive images from negative images. After applying this technique, we get accuracy 36.06%, F1 score 87%, precision 85%. ROC curve for COVID-19 detection. This paper has visual exploratory data analysis of CT images based on COVID cases.

References

[1] Stones Analyses in Computed Tomography Images". 'IEEE International Conference on Innovations in Engineering, Technology and Sciences' (ICIETS), NIE Institute of Technology Mysore,20th and 21st September 2018.

[2] Aravind Jadhav, Sanjay Pujari. "Radiological image Quality enhancement and analysis". 'IEEE International Conference on Advance in Information Technology (ICAIT-2019), Adichunhanagiri Institute of Technology Chikkamagaluru, 26th and 27th July 2019.Page no: 3119

[3] Aravind Jadhav, Sanjay Pujari. "Resolution Enhancement of CT Images Based on Histogram Equalization". 'International Journal of Engineering and Advanced Technology (IJEAT)' Volume-6 Issue-ICDSIP17, Page No.: 129-132, March 2017.

[4] Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, Cai M, Yang J, Li Y, Meng X, Xu B. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). medRxiv; 2020.

[5] J. Pu, J. K. Leader, A. Bandos et al., "Automated quantification of COVID-19 severity and progression using chest CT images," European Radiology, 2020.

[6] L. Xiao, P. Li, F. Sun et al., "Development and Validation of a Deep Learning-Based Model Using Computed Tomography Imaging for Predicting Disease Severity of Coronavirus Disease 2019," Frontiers in Bioengineering and Biotechnology, vol. 8, 2020.

[7] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing," Radiology, pp. 200343-200343, 2020.

[8] A. Jacobi, M. Chung, A. Bernheim, and C. Eber, "Portable chest x-ray in coronavirus disease-19 (covid-19): A pictorial review," Clinical Imaging, 2020.

[9] E. Grøvik, D. Yi, M. Iv, E. Tong, D. Rubin, and G. Zaharchuk, "Deep learning enables automatic detection and segmentation of brain metastases on multisequence mri," Journal of Magnetic Resonance Imaging, vol. 51, no. 1, pp. 175–182, 2020.

[10] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, and P. Niu, "A novel coronavirus from patients with pneumonia in China, 2019,"New England Journal of Medicine, 2020.

[11] W. Tan, X. Zhao, X. Ma, W. Wang, P. Niu, W. Xu, G.F. Gao, and G. Wu,"A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019 2020,". China CDC Weekly, vol. 2, no. 4, pp. 61-62, 2020.

[12] E. Mahase," China coronavirus: WHO declares international emergency as death toll exceeds 200,"*BMJ: British Medical Journal (Online)*, vol. 368, 2020.

[13] D.S. Hui, E.I. Azhar, T.A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T.D. Mchugh, Z.A. Memish, C. Drosten, and A. Zumla,"The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China," *International Journal of Infectious Diseases*, vol. 91, pp. 264-266, 2020.

[14] Y. Pan, H. Guan, S. Zhou, Y. Wang, Q. Li, T. Zhu, Q. Hu, and L. Xia, "Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China," *European radiology*, pp. 1-4, 2020.

[15] Y. Fang, , H. Zhang, , Y. Xu, , J. Xie, , P. Pang, and W. Ji," CT manifestations of two cases of 2019 novel coronavirus (2019-nCoV) pneumonia," *Radiology*, vol. 295, no. 1, pp. 208-209, 2020.

[16] M.J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, and N. Shukla,"COVID-19 detection through transfer learning using multimodal imaging data," *IEEE Access*, vol. 8, pp. 149808-149824, 2020

[17] M. Loey, G. Manogaran, and N.E.M. Khalifa,"A deep transfer learning model with classical data augmentation and cgan to detect covid- 19 from chest ct radiography digital images," *Neural Computing and Applications*, pp. 1-13, 2020.

[18] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing," *Radiology*, pp. 200343-200343, 2020.

[19] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z.A. Fayad, N. Zhang, K. Diao, B. Lin, X. Zhu, K. Li, and S. Li," Chest CT findings in coro- navirus disease-19 (COVID-19): relationship to duration of infection," *Radiology*, pp. 200463.

[20] X. Yang, X. He, J. Zhao, Y. Zhang, S. Zhang, and P. Xie," COVID- CT-dataset: a CT scan dataset about COVID-19," *arXiv, pp.arXiv-2003*, 2020.

[21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, " Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[22] J. Premaladha, and K.S. Ravichandran,"Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algo- rithms," *Journal of medical systems*, vol. 40, no. 4, pp. 96, 2016.

[23] C. Goutte, and E. Gaussier,"A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *European confer- ence on information retrieval*, Berlin, Heidelberg, Mar. 2005, pp. 345- 359.