COMPARATIVE STUDY OF COVID AND PULMONARY FIBROTIC CT LUNG IMAGES USING SIAMESE NETWORKS WITH VGG16

Sergey. V. Aksenov,
N.J. Francis, N.S. Francis, group No. A7-39
Tomsk Polytechnic University
E-mail: natzina_92@yahoo.com

Abstract

In this research an algorithm is proposed to produce comparative results between Pulmonary Fibrosis of the lungs and COVID computer tomography lung images for the purpose of research to aid in the field of medical science. The Siamese Network which is based on parallel tandem operation to produce comparative results, is altered by changing or altering the implementation function using the VGG16 neural network. The input data set in the method uses a variation of healthy lung CT images along with CT images of cases with pulmonary fibrosis and COVID. The main aim is to produce a comparative study on the textural variation of the CT images under study to further enhance research outputs in the future with accuracy and less time consumption.

Introduction

The widespread Coronavirus disease (COVID-19) has become a topic of high research [1, 2] since its fast and immediate spread globally which has become a serious concern since the beginning of the year 2020. Artificial Intelligence [3] and improved deep learning techniques [4] over the years with the help of Computer Tomographic (CT) scans to detect diseases has proved major success in the field of medical science. The AlexNet [5] in 2012 and GoogleNet [5,6] which is used to detect cancer are examples of CNN models producing accurate results in medical science. In the case of COVID and other related lung fibrosis CT images, it can be easily misclassified predicting incorrect diagnosis [7]. The textural pattern between the CT images can be invariantly studied simultaneously to produce differential results. By classifying and identifying various texture features [8,9] in these images early detection is possible without misclassification.

Hence, with the help of a Siamese deep learning network [10], COVID abrasive manifestations in CT scans is distinguished from Pulmonary Fibrosis CT scans. The network is based on a divided infrastructure where the inputs of two variants are input simultaneously with shared weights between them. The main aim is produce results to aid radiologists and thereby help in future related diagnostic research.

Methods and Technology

The overall working architecture is based on the Siamese network [10] which is commonly known as the ‘twin network’. The concept of this neural network is defined by two neural networks working parallelly in tandem by using a shared weights [11]. The comparative study of the varying CT lung scans consisting of Healthy, COVID and pulmonary fibrosis CT scans requires a modified version of the existing model. The VGG16 model [12] is implemented as the twin network here for the simultaneous feature detection in the Siamese Model.

The parallel network Siamese model is used for presenting two input CT scans into the network and to verify whether the images belong to the same define classified class [13]. The model proposed is a 3D framework using deep learning to detect misclassified CT scans of COVID and Pulmonary Fibrosis. The model helps in extracting both 2D local and 3D global features [14,15]. The Siamese network in this case uses the VGG16 with ReLu operator. A series of CT scans are input into the model and fibrotic features are detected simultaneously with the help of the k-means and the region growing algorithm. The region growing algorithm in this case helps in detection of nodular and non-nodular regions. This is then compared with manually annotated images by the VGG annotator. The CT scan have now defined features which are the max pooled [16]. These features are then sent to the fully connected layer of the Siamese network and by the Euclidean distance [16] operator the features are compared and then set to a probability score by the softmax activation function for each classified image such as the healthy, COVID-19 and the pulmonary fibrosis CT scan. Thus, the final output predicts the comparative analysis. This output is further tested with a validation test dataset that consists of images that are already defined and accurately diagnosed. The Final output from the model is the verified CT image scan with classified diagnoses, eliminating the misclassified ones. . During the pre-processing, feature extraction takes place where the fibrotic regions are identified by the K-mean algorithm. This unsupervised algorithm [17] helps in locating pixels of varying differential texture and clustering them by identifying the nearest pixels by basing it on the K-centroids. It is given as follows:

$$H_f = \sum_{m=1}^{K} \sum_{n=1}^{i} || x_n^{(m)} - y_m||^2$$

Where, H_f is the objective function
K is the number of clusters
i is the number of cases
x_n is a selected case
y_m is the centroid for the cluster m

Once the fibrotic abrasive regions have been identified, the region growing algorithm is further applied to differentiate the nodules from non-nodular detection. This algorithm is similar to the clustering algorithm where the image is partitioned into regions. The discontinuity between the greyscale property helps in determining any direct region. The final fully connected layer [16,17] in the VGG16 consists of 4096 channels in total and after the softmax activation function is applied the final output is that consisting of 1000 classes. This is then compared with a pretrained data set consisting of images with a healthy CT scan, COVID19 and pulmonary fibrosis CT scan which finally predicts the output with a true classified CT scan. After the final layer as depicted in the image with 1000 classes, the output is compared with the defined dataset to produce the required output.

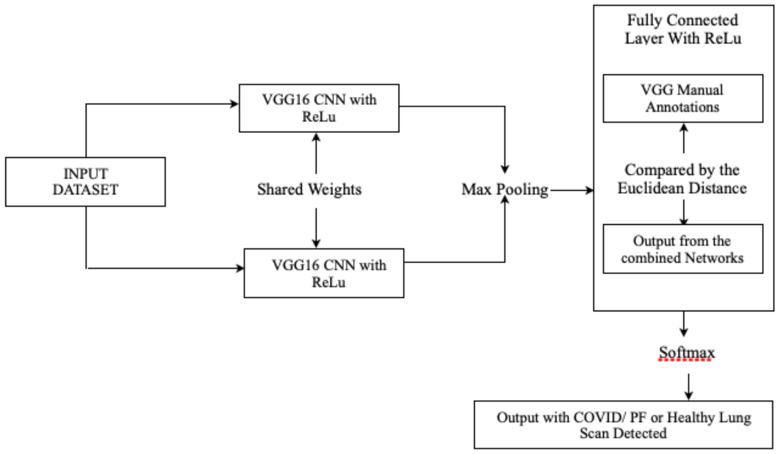

Fig. 1. The architectural structure of the proposed model to identify healthy, COVID and Pulmonary Fibrosis CT Lung Scan

The training data set was randomly split into the data set for training the model and the comparative dataset for manual annotation with a ratio of 8:2. The manual annotated dataset was not include in any internal model training and validation [18] and only used for the purpose of comparison to produce the relative output. The process is carried out on an 8GB GPU processor. The dataset was divided into 6 groups and sent for testing. The sensitivity, specificity and the AUC are recorded to identify the distinction between COVID and fibrotic images. The average sensitivity and specificity recorded for COVID images equal to 0.97 and 0.94. The average sensitivity and specificity recoded for fibrotic images approximates to 0.93 and 0.99.

Conclusion

The robust model of the Siamese deep learning network, helps in distinguishing CT scans into a healthy one, COVID-19 affected and those with Pulmonary fibrosis. The VGG16 network incorporated into the Siamese model helps in identifying features accurately and distinctively. The CT scans on comparison with the ground truth classified images that were produced manually along with the k-means and the region growing algorithm helps in defining the features more distinctively.

The limitations faced by the system are due to the lack of an available large dataset containing COVID-19 CT scans, predictions of only true or false case scenarios could be predicted. The future of the research would include to diagnose its severity and distinguish the cases from existing pulmonary fibrotic cases from post COVID-19 pulmonary fibrotic cases. However with the acquired dataset, the model produces results with accuracy based on the statistical analysis with efficiency.

References

1. Gennaro F.D., Pizzol D., Marotta C., Antunes M., Racalbuto V., Veronese N., Smith L. Coronavirus Diseases (COVID-19) Current Status and Future Perspectives: A Narrative Review. International Journal of Environmental Research and Public Health. – 2020. – vol. 17. – №. 8. – P. 2690.
2. Wiersinga W.J., Rhodes A., Cheng A.C., Peacock S.J., Prescott H.C. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19). JAMA. – 2020. – vol. 324. – №. 8. – P. 782-789.
3. Kim E.E. Artificial Intelligence and Computer-aided Diagnosis in Medicine. Current Medical Imaging Formerly Current Medical Imaging Reviews. – 2020. – vol. 16. – №. 1. – P. 3–11.
4. De Bruijne M. Machine learning approaches in medical image analysis: From detection to diagnosis. Medical Image Analysis. – 2016. – vol. 33. – P. 94–97.
5. Almas B., Sathesh K., Rajasekaran S. A Deep Analysis of Google Net and Alex Net for Lung Cancer Detection. Regular Issue. –2019. – vol. 9. – №. 2. – P. 395–399.
6. Fang T. A Novel Computer-Aided Lung Cancer Detection Method Based on Transfer Learning from GoogLeNet and Median Intensity Projections. IEEE International Conference on Computer and Communication Engineering Technology (CCET). – 2018.
7. Orish N., Lopez D.M. Misclassification of Asymptomatic Infections of COVID-19. Asian Journal of Medicine and Health. – 2020. – P. 119–122.
8. Francis N.S., Francis N.J., Xu Y., Saqib M., Aljasar S.A. Identify Cancer in Affected Bronchopulmonary Lung Segments Using Gated-SCNN Modelled with RPN. 2020 IEEE 6th International Conference on Control Science and Systems Engineering (ICCSSE). – 2020.
9. Aljasar S.A., Xu Y., Saqib M. Anaylsis and detection of community-acquired pneumonia using PSPNet with complex daubechies wavelets. Indian Journal of Computer Science and Engineering. – 2020. – vol. 11. – №. 3. – P. 217–225.
10. Chicco. Siamese Neural Networks: An Overview. Artificial Neural Networks. – 2020. – P. 73–94.
11. Takahashi R., Matsubara T., Uehara K. A Novel Weight-Shared Multi-Stage CNN for Scale Robustness. IEEE Transactions on Circuits and Systems for Video Technology. – 2019. – vol. 29. – №. 4. – P. 1090–1101.
12. Geng L., Zhang S., Tong J., Xiao Z. Lung segmentation method with dilated convolution based on VGG-16 network. Computer Assisted Surgery. – 2019. – vol. 24. – №. 2. – P. 27–33.
13. Francis N.J., Francis N.S., Axyonov S.V., Aljasar S.A., Xu Y., Saqib M. Diagnostic of Cystic Fibrosis in Lung Computer Tomographic Images using Image Annotation and Improved PSPNet Modelling. Journal of Physics: Conference Series. –2020. – vol. 1611. – P. 120–128.
14. Al-Shabi M., Lan B.L., Chan W.Y., Ng K.H., Tan M. Lung nodule classification using deep Local–Global networks. International Journal of Computer Assisted Radiology and Surgery. –2019. – vol. 14. – №. 10. – P. 1815–1819.
15. Francis N.S., Francis N.J., Axyonov S.V., Saqib M., Xu Y., Aljasar S.A. Applying Ternary Net Weights to the R-Mask Neural Network to Identify Bronchopulmonary Lung Segments. Journal of Physics: Conference Series. – 2020. – vol. 1611, P. 120-126.
16. Wang W., Taft D.A., Chen Y.J., Zhang J., Wallace C.T., Xu M., Watkins S.C., Xing J. Learn to segment single cells with deep distance estimator and deep cell detector. Computers in Biology and Medicine. – 2019. – vol. 108. – P. 133–141.
17. Lee W.M. Unsupervised Learning-Clustering Using K-Means. Python® Machine Learning. – 2019. – P. 221–242.
18. Kollias D., Zafeiriou S. Training Deep Neural Networks with Different Datasets In-the-wild: The Emotion Recognition Paradigm. International Joint Conference on Neural Networks (IJCNN). – 2018.