ASIAN JOURNAL OF INFORMATION TECHNOLOGY
MEDWELL PUBLICATIONS

COVID-19 Segmentation and Classification from CT Scan Images

Sridevi Ramachandra Rathod and Harmeet Kaur Khanuja
Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, India

Key words: COVID-19, segmentation, k-means, convolutional neural network, pneumonia

Corresponding Author:
Sridevi Ramachandra Rathod
Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, India

Page No.: 168-173
Volume: 20, Issue 6, 2021
ISSN: 1682-3915
Asian Journal of Information Technology

Abstract: The pandemic coronavirus disease-2019 (COVID-19) has infected millions of people in over 200 countries and territories as of 2021. It is very necessary to detect COVID-19 in the initial stage to provide appropriate medical treatment to patients and also to protect the uninfected people. For this reason, we develop a framework to automatically segment COVID-19 CT images using k-means clustering and use them to train proposed convolutional neural network to classify COVID-19 from normal CT images. Rapid growth in machine learning and deep learning has been doing great work to reduce time of radiologists by assisting them in the diagnosis of COVID-19. Our framework is evaluated upon 349 positive and 397 negative CT scans to detect COVID-19 and help in taking appropriate diagnostic decisions. To evaluate the performance of proposed approach, we compared our results with pre-trained models such as VGG19, InceptionV3 and ResNet50.

INTRODUCTION

The rapidly spreading Coronavirus Disease-2019 (COVID-19) has become a worldwide pandemic, since, the year 2019, originated in Wuhan in China. The coronavirus is a kind of virus that leads to illness such as common cold, Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). The coronavirus disease affects different people in different ways. Some people get cured without hospitalization and in some cases it leads to death due to critical complications. Along with the most usual symptoms of COVID-19 such as cough, shortness of breath and fever, the COVID-19 infected patients will also suffer from pneumonia.

The characteristics of pneumonia found in COVID-19 patients may be similar to characteristics of pneumonia obtained by other viruses or bacteria. Therefore, it is difficult to find the source of the disease without being tested for COVID-19 or any other respiratory infections. According to some researchers, it is said that patients with COVID-19 pneumonia have their both lungs affected by pneumonia but not just one lung and the pneumonia has pattern of ground glass opacity via. CT scan. In clinic practice, errors in segmentation can cause critical wrong results. Therefore, in medical image analysis, segmentation of lesions or infection portions require greater level of efficiency compared to natural or common images.

As per records of World Health Organization (WHO), globally COVID-19 as of 19 July 2021, confirmed cases are 192,284,207, confirmed deaths are 4,136,518[1]. The CT (Computed Tomography) imaging technology plays an important role in the identification of characteristics in the lung occurred by COVID-19. To diagnose the disease COVID-19 accurately and to assist in follow up analysis

of the disease, segmentation of lesions known as infected portions from CT scans is very necessary. To find quantifiable value of disease progression, the imaging techniques such as CT scan or Chest X-ray are being used as a part in the diagnosis of COVID-19. Doctors can easily visualize the changes that occur due to COVID-19 pneumonia in our lungs with the help of the imaging techniques. The testing of this coronavirus disease manually, needs a large number of testing kits which are very costly and inefficient blood tests, where a blood test needs some hours of time to arrive at the testing result. Therefore, to overcome these situations, in order to provide better and efficient treatment, deep learning techniques are being used in the diagnostic process of COVID-19.

In this study, we aim to build an automatic segmentation model of COVID-19 pneumonia lesions from CT images using k-means clustering and its classification using convolutional neural network to achieve better performance.

Literature review: Existing works performed on segmentation and classification of COVID-19 pneumonia from the given CT or X-ray images to help in diagnosing of COVID-19 are presented below.

The study[2], presents a novel Coronavirus (2019-nCoV) which belongs to the category of betacoronavirus. The virus 2019-nCoV is found in patients of Wuhans's hospitals and is recognized with the help of techniques such as molecular and un-biased DNA sequencing. The drawback of this study is that it does not satisfy Koch's postulates. By Benvenuto *et al*.[3] report epidemic transmission of the virus and its history, by using the approach of epidemiology in combination with molecular process to focus on genetic variation in pathogens. This study helps in preventing future epidemics like SARS-CoV and 2019-CoV.

Huang *et al*.[4], determined a parameter from CT image known as percentage of lung opacification (QCT-PLO) by using a deep learning tool. The quantitative parameter QCT-PLO is greatly increased between baseline CT and 1st follow-up CT. Limitation of this study is that it needs supervision of radiologists as there is no clarification of pulmonary opacities. This study[5] has shown how convolutional neural network is used to visualize required features to determine whether a CT scan is COVID-19 or not with the help of heatmaps. Even though the model does not show the classification of the disease into different severities but it can clearly do the prediction as COVID-19, community acquired pneumonia (CAP) or non-pneumonia. It is very difficult to determine which features are used to distinguish CAP from COVID-19.

Shen *et al*.[6], presented convolutional neural network approach made up of multiple layers. This multi-layered approach is used to analyse medical images, where as the model's first layer is used to detect motifs by applying filters. Limitation of this approach is that there is no fixed number of layers and filters. In the work[7], deep separable convolution kernels are used along with convolution model. This proposed model used lesser number of parameters and filters, replacing few convolutional layers. The model is tested on X-ray scan of chest images of around 6000 and the experimental results has shown that number of iterations needs to be increased to improve accuracy.

In the work[8], full convolution neural network is trained with the proposed spatially adaptive reweighting technique where weights are generated based on local texture of the image. The approach is applied to skin images dataset in order to segment skin lesions. One of the merits of this model is that it uses lesser number of clean and larger number of noisy data. The study[9], uses bidirectional chain coding technique with support vector machine as classifier. Over-segmentation is reduced by implementing point pairs selection technique. Experimenting on juxtapleural nodules, the proposed model achieves good overlap ratio. The approach can deal with low quality images but there is no re-inclusion of juxtapleural nodules in consolidation areas. In our previous study[10], we presented a brief survey of works carried out on segmentation and classification of COVID-19. The study presents comparative study of various techniques used in the process of COVID-19 diagnosis along with their merits and demerits. And it has shown that CNN and UNet++ have achieved better accuracy comparatively.

This study presents an overview of different techniques used for segmentation and classification in the field of COVID-19 prediction. Existing studies reveal that there are certain limitations concerning incomplete, inaccurate images, labelled data and require more number of iterations. Based on the above, it has been found that not much work has been carried out for training the network using segmented images. Therefore, we aim at developing an automatic model for segmentation and classification of COVID-19 using deep learning framework to achieve higher performance.

MATERIALS AND METHODS

In this work, we aim at developing an automatic segmentation and classification model of COVID-19 pneumonia lesions from CT scan images using k-means clustering and proposed convolutional neural network to achieve better performance.

The system architecture diagram of our model is as shown in Fig. 1. The model consists of four modules such as image pre-processing, segmentation using k-means, feature extraction and classification using CNN. The

model receives CT image as input and produces output as COVID-19 or non-COVID-19. The proposed system, respective algorithms and mathematical model are introduced below:

- Architecture
- Algorithms

k-means clustering: The segmentation phase is performed using k-means clustering algorithm. The input for segmentation phase is pre-processed CT scan image of COVID-19. Segmentation is the classification of an image into many parts or regions. Here, segmentation of pneumonia lesions from CT scan images of COVID-19 is done. k-means clustering is used to segment the image into different clusters. Clustering the intensity values of the image into different regions or clusters.

k-value is chosen based on desired number of clusters. Then randomly choosing the centers of k clusters and calculating mean of the cluster. Based on Euclidean distance between each pixel and each cluster, every pixel is assigned to a cluster. Re-estimating the center and repeating the process until the center doesn't move. Finally, we get the clusters containing similar data points and the output of the segmentation phase is the image used to train the proposed Convolutional Neural Network for classification task (Fig. 2).

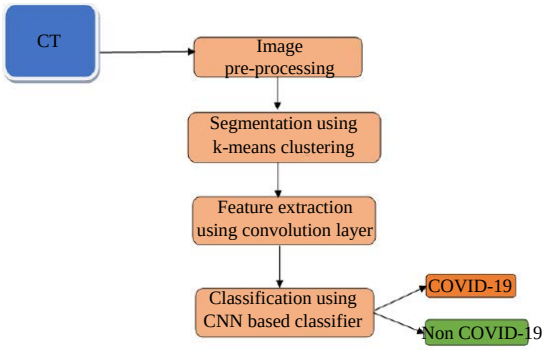

Fig. 1: Proposed system architecture

Proposed convolutional neural network: After the segmentation phase, features are extracted from the segmented image using convolution layer. Based on the features learned, classification is done using CNN based classifier.

As shown in Fig. 3., the building blocks of CNN architecture are convolutional layers, pooling layers and fully connected layers. Let the input be a 2D CT image, then 2D convolutional layer is applied to extract the local features from the input image, using a fixed size filter called kernel. The pooling layers perform sub-sampling to reduce the size of feature maps. The features from the convolutional layers are propagated till the last fully connected layer which uses sigmoid activation function for prediction as 0 or 1.

The output of the classification phase is in two categories. One class is consisting images of COVID-19 detection and another class consisting of images of non-COVID-19 detection.

Mathematical model: The mathematical model for automatic segmentation of COVID-19 pneumonia lesion from CT Images is as:

$$S = \{I, F, O\}$$

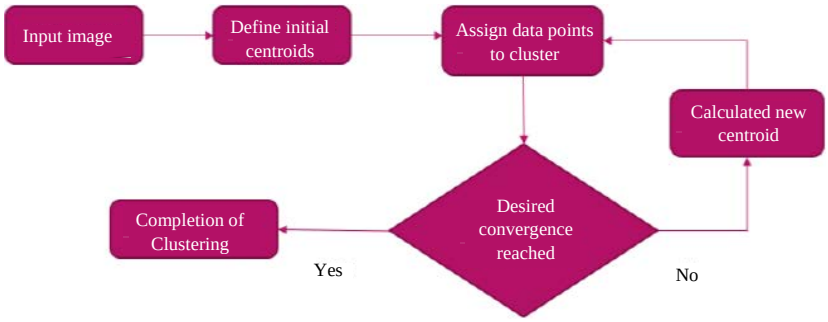

Fig. 2: Flow of k-means algorithm

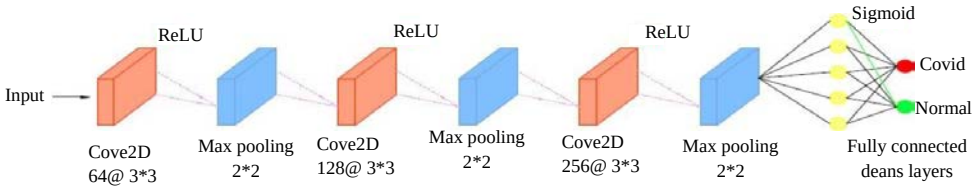

Fig. 3: Proposed convolutional neural network

Where:
I = Set of inputs. The input consists of set of images. It uses CT images
F = Set of functions F = {F1, F2, F3, F4}

F1: image preprocessing: OpenCV for pre-processing and median filter for noise removal.

F2: image segmentation: Using k-means clustering:

$$M = \frac{\sum_{i:c(i)=k} Xi}{Nk},\ k = 1, 2, ..., k$$

$$D(i) = \arg\min||X_i - M_k||^2,\ i = 1, 2, ..., N$$

F3: image feature extraction: Convolution layer is used to extract features from an input image. Operations such as edge detection, blur and sharpen is done by applying filters.

F4: image classification: Using convolutional neural network made up of:

- The 3 convolution layers of filter (3*3) and ReLu activation function
- The 3 max pooling layers of filter (2*2)
- The 1 flatten layer
- The 2 Dense layers
- Sigmoid activation function: $\sigma(x) = 1/1+e^{-x}$

O: Class 1 = COVID-19 or Class 2 = non-COVID-19.

RESULTS AND DISCUSSION

For our model, we use the data set consisting of 349 CT scan images of COVID-19 from 216 patients and 397 non-COVID- 19 CTs. Experimental evaluation is done to compare the results for evaluating the performance. For experimental result evaluation, we use the confusion matrix with the following notations:

- TP: True Positive (correctly predicted number of instances)
- FP: False positive (incorrectly predicted number of instances)
- TN: True Negative (correctly predicted number of instances as not required)
- FN: False Negative (incorrectly predicted number of instances as not required)

On the basis of these parameters, we can evaluate the accuracy as follows:

- Accuracy = (TP+TN)/(TP+TN+FP+FN)
- Sensitivity = TP/(TP+FN)
- Specificity = TN/(TN+FP)
- Precision = TP/(TP+FP)
- F1-score = 2*((precision*recall)/(precision+recall))

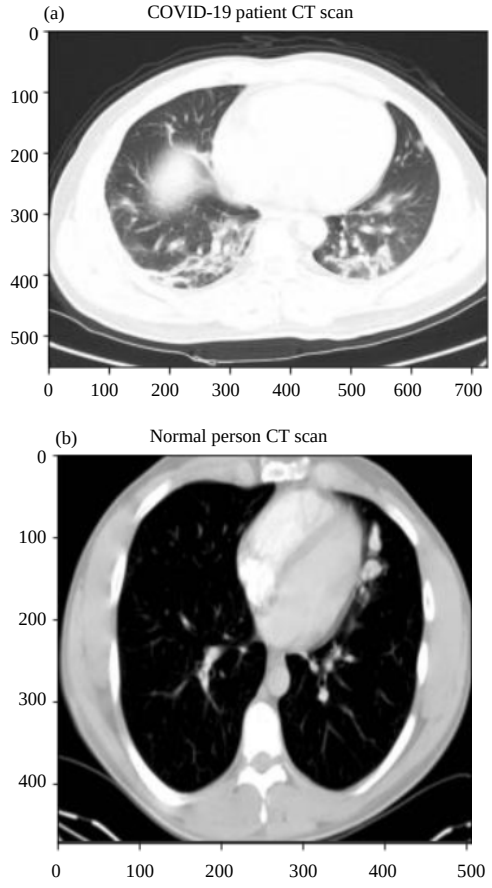

Fig. 4(a, b): Sample original images in dataset, (a) COVID CT and (b) Normal CT

The dataset consists a total of 746 CT scan images. 80% of the data is used for training and remaining 20% of the data is used for testing. Figure 4 shows sample COVID positive and normal CT scan from the dataset.

After applying k-means algorithm for the COVID positive CT scan with value of k = 4, following results are obtained as shown in Fig. 5. The cluster 3 or extracted image 3 contains the COVID-19 pneumonia lesions.

Following, in Fig. 6, are some original COVID CT scans along with their segmented images using k-means clustering with k = 4. The original image is segmented in four clusters with black, white, dark gray and light gray colors.

The proposed model is trained once with un-segmented (before applying k-means) images (CNN) and once with segmented (after applying k-means) images (CNN_Seg). The accuracy of the model is improved with segmented images, i.e., CNN_Seg. Performance of the proposed model is compared with pre-trained networks such as VGG19, InceptionV3 and ResNet50. Table 1 shows the training and validation accuracies, precisions, recalls and losses of various models. Proposed CNN_Seg

Table 1: Comparative results based on training and validation

Model v/s metrics	Accuracy	Precision	Recall	Loss	Val_Accuracy	Val_Precision	Val_Recall	Val_Loss
VGG_19	70.45	70.45	70.45	55.61	70.00	70.00	70.00	56.46
ResNet50	59.04	59.04	59.04	77.21	65.33	65.33	65.33	64.48
InceptionV3	68.74	68.74	68.74	98.65	68.00	68.00	68.00	89.63
Proposed CNN	78.81	86.48	73.58	46.69	67.86	63.23	86.73	56.17
Proposed CNN_Seg	83.52	84.19	82.07	35.61	70.09	71.63	78.91	62.04

Table 2: Comparative results based on evaluation metrics

Model v/s metrics	Image_size	Sensitivity	Specificity	F1-score
VGG_19	224*224	68.11	71.60	67.62
ResNet50	224*224	71.42	62.96	53.57
InceptionV3	224*224	64.10	72.22	67.56
Proposed CNN	50*50	78.26	63.22	73.13
Proposed CNN_Seg	50*50	67.46	71.63	75.09

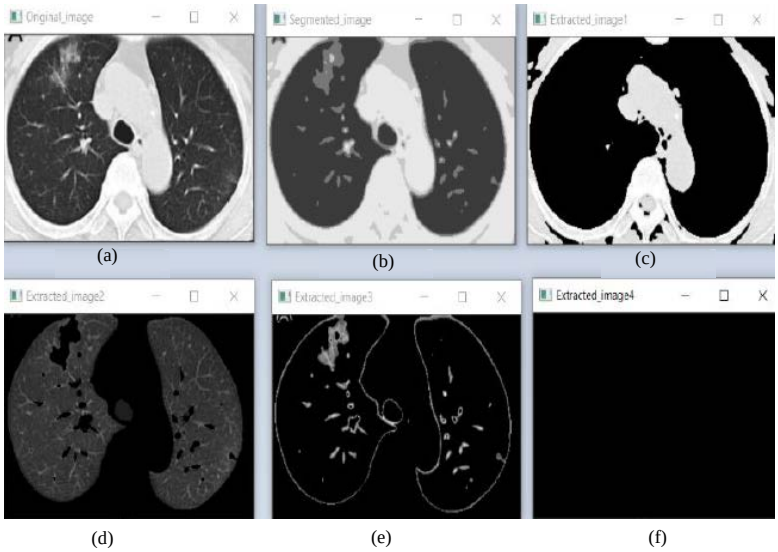

Fig. 5(a-f): k-means output images, (a) Original, (b) Segmented, (c) Cluster 1, (d) Cluster 2, (e) Cluster 3 and (f) Cluster 4

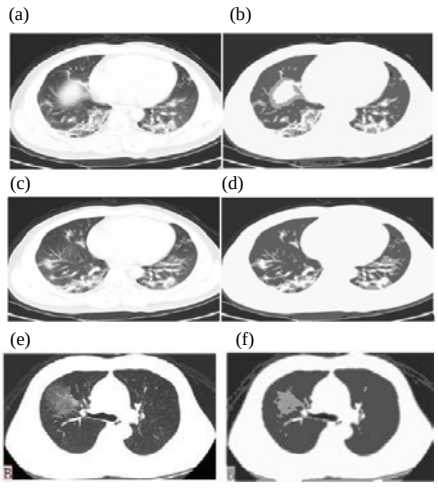

Fig. 6(a-f): Sample original and segmented results, (a-c) Original COVID images and (d-f) Segmented images

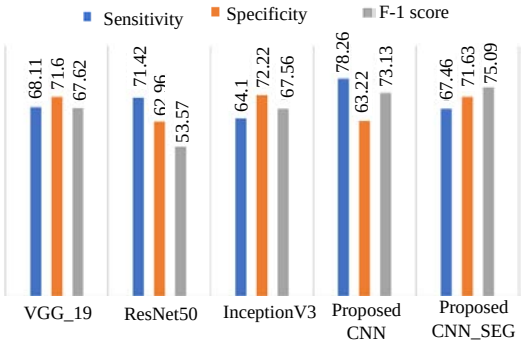

Fig. 7: Comparison of different networks with proposed CNN for classification of COVID-19

has achieved accuracy of 83.52% which is higher than proposed CNN and other models. As shown in Table 2, proposed CNN_Seg has achieved F1-score of 75.09% which is greater than proposed CNN and other models (Fig. 7).

CONCLUSION

The coronavirus disease (COVID-19) has become a worldwide pandemic since the year 2019. There are millions of COVID-19 cases reported worldwide with over millions of deaths (as of July 2021). Along with the common symptoms in patients of COVID-19 such as shortness of breath, cough and fever, the patients also suffer from pneumonia. In this study, we develop a framework to automatically segment and classify COVID-19 CT images using k-means clustering and proposed convolutional neural network. After training the proposed network with segmented images using k-means clustering, the accuracy is high compared with proposed network trained with original images. The accuracy of the proposed CNN trained with segmented images is also high when compared with pre-trained models such as VGG19, InceptionV3 and ResNet50.

REFERENCES

01. Anonymous, 2019. Coronavirus disease (COVID-19) pandemic. World Health Organization, Geneva, Switzerland.
02. Zhu, N., D. Zhang, W. Wang, X. Li and B. Yang *et al*., 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med., 382: 727-733.
03. Benvenuto D., Giovanetti M., Salemi M., Prosperi M., Cecilia De Flora, Junior Alcantara L. C., Angeletti S. and Ciccozzi M., 2020. The global spread of 2019-nCoV: a molecular evolutionary analysis. Pathog., Glob., Health 114: 64-67.
04. Huang, L., R. Han, T. Ai, P. Yu, H. Kang, Q. Tao and L. Xia, 2020. Serial quantitative chest CT assessment of COVID-19: A deep learning approach. Radiol. Cardiothorac. Imaging, Vol. 2, No. 2. 10.1148/ryct.2020200075
05. Li, L., L. Qin, Z. Xu, Y. Yin and X. Wang *et al*., 2020. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology, Vol. 296, No. 2. 10.1148/radiol.2020200905
06. Shen, D., G. Wu and H.I. Suk, 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng., 19: 221-248.
07. Sarada, N. and K.T. Rao, 2021. A neural network architecture using separable neural networks for the identification of pneumonia in digital chest radiographs. Int. J. e-Collaboration (IJeC.), 17: 89-100.
08. Mirikharaji, Z., Y. Yan and G. Hamarneh, 2019. Learning to Segment Skin Lesions from Noisy Annotations. In: Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data, Wang, Q., F. Milletari, H.V. Nguyen, S. Albarqouni and M.J. Cardoso, *et al*. (Eds.)., Springer, Cham, Switzerland, pp: 207-215.
09. Shen, S., A.A. Bui, J. Cong and W. Hsu, 2015. An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. Comput. Biol. Med., 57: 139-149.
10. Rathod, S.R. and H.K. Khanuja, 2021. Automatic segmentation of COVID-19 pneumonia lesions and its classification from CT images: A survey. Proceedings of the 2021 International Conference on Intelligent Technologies (CONIT), June 25-27, 2021, IEEE, Hubli, India, pp: 1-8.