Advances and Applications in Mathematical Sciences
Volume 20, Issue 12, 2021, Pages 3309-3317

AN ABNORMAL AREA IDENTIFICATION IN LUNGS FOR COVID-19 PANDEMIC USING AN IMPROVED HYBRID ABNORMAL DATA DETECTION ALGORITHM (HADDA)

APPARNA ALLADA, R. BHAVANI, KAVITHA CHADUVULA and R. PRIYA

Research Scholar
Department of Computer Science and Engineering
Annamalai University, India

Department of Information Technology
Gudlavalleru Engineering College, India

Abstract

Covid-19 is very dangerous pandemic that is firstly identified in Wuhan city, china. This disease spreads very rapidly due to its contagious characteristics, hence identified as world pandemic by WHO in march 2020. It is a severe acute respiratory syndrome. Though many countries implemented lockdown to stop spreading of this pandemic, it resulted in economical effect of many organizations and institutions. Initially, the symptoms of this pandemic were reported as cold, dry cough, fever and body pains which are identified in the affected person. Day-to-day, WHO is releasing newly identified symptoms such as loss of taste and smell. It is a big challenge to rely on the external symptoms to detect the disease. Though RNA or PCR test and rapid antigen tests (sometimes known as a rapid diagnostic test-RDT) are used to detect viral proteins (known as antigens), these tests sometimes fail eventually and virus cannot be detected. Swab samples are collected from the nose and/or throat in the above tests. The more appropriate test would be is a CT scan for the lungs that identifies virus affected portions. Many deep learning and machine learning algorithms are being used to increase the accuracy of identifying abnormal portions. The proposed algorithm mainly concentrates on people who may not be tested positive but have the symptoms, because this set of people are the main sources of virus spread. For those patients, CT scan is the better option. In this paper, the abnormal area is identified in lungs using the proposed Hybrid Abnormal data detection Algorithm (HADDA). The experiments are carried on 200 CT scan lung images and the performance is evaluated; sensitivity, specificity and accuracy are observed to be better than compared algorithms.

2020 Mathematics Subject Classification: 68.
Keywords: Covid-19, RNA, PCR, Deep Learning, Machine learning, HADDA.
Received November 20, 2020; Accepted December 20, 2020

I. Introduction

Corona virus has spread in such a way that it happened like we have to walk along with it in our day-to-day life, and move on with it. Human species is greatly affected by this pandemic. In 2020 March-11, WHO (World health organization announced the virus called COVID-19, a new pandemic and advised all the countries to lockdown [1]. The virus belongs to the RNA family and it is more compatible of causing remarkable viral bacteria in humans and animals. The size of the corona virus is medium size which affects not only humans but also birds and mammals, where bats are hosts of this virus and bats consist of huge number of viral genotype of corona virus. It is spreading very rapidly all over the world. More than 22.5 million people are affected with this virus from all over the world, up to 24th of August. Approximately 1 million infected people resulted in death with this disease. Around 216 countries are affected with this virus, which led to many economic flaws. More than 7 million people recovered worldwide [2]. India is second place with 8.9 millions of people affected and 8.1 million people recovered. USA is in first place counting around 110 millions of infected people with this virus.

COVID-19 testing is involved in analyzing the samples which indicate the present or past presence of severe acute respiratory syndrome-associated corona virus 2 (SARS-CoV-2). This test can be done to detect whether the virus is infected to patient or not and check the human blood whether the patient have developed any antibodies in the human body. The diagnosis process for this disease is mainly based on two types- (i) Laboratory testing approach which collects the patient swab present in nose or mouth. (ii) Real-time reverse transcription-polymerase chain reaction (RT-PCR) is other approach to check the COVID-19 status (positive or negative).

In this paper, the comparative results are shown by using various deep learning techniques and performance of proposed system. The proposed system works on the basis of finding the abnormal affected area within the lungs CT scan images.

II. Literature Survey

In this section, the performance of various deep learning and machine learning techniques are discussed.

The author Ahuja et al. [4] explained about the data augmentation and existed networks for COVID-19 images that are ready to classify. Constant wavelets, abnormal pivot, interpretation, and shear sports have been implemented to the CT scan images and these come under data augmentation. Among all the existing systems, the better one that is proposed by the author for the classification is ResNet18, which performed better when compared with existing techniques and the performance of proposed system obtains 0.9965 AUC rating.

The author Liu et al. [5] proposed the new system called deep neural network technique, which is more compatible to check this dataset images and this achieved 0.94 AUC rating and uses the VGG16 due to the fact of the backbone version.

One of the author considered machine learning (ML) algorithms in place of deep learning algorithms, Barstugan et al. [6] took ML algorithms into consideration to arrange 150 COVID-19 and non-COVID-19 images. Various detail extraction strategies, for instance, grey-level size zone matrix (GLSZM) and discrete wavelet transform (DWT) have been taken into consideration in the issue of extraction measure, and consequently eliminated highlights that have been ordered using an assist vector machine. K-fold cross-approvals have been acted in the analyses with 2, 5, and 10 folds. Barstugan et al. inferred that exactness become carried out with the aid of using SVM using the GLSZM encompass extraction method.

Bai et al. [7] achieved the profound getting to know engineering Efficient Net B4 [8] to organization COVID-19 and pneumonia cuts of CT checks. The end of the six radiologists at the evaluating sufferers has been applied to evaluate the effectiveness of the results received with the aid of using an AI version. The AI version carried out more precision, whilst the conventional exactness of the locating of radiologists becomes received at low precision.

III. Dataset Description (SARS-CoV-2 CT-scan dataset)

This dataset consists of 2482 CT scan images and an overall count of 120 patients. In this, 60 patients are positive with 1252 CT scans, out of which 32 are males and 28 are females; 60 patients are negative with 1230 scans, among which 30 are males and 30 females. These samples are collected from hospitals of Sao Paulo, Brazil. The below table gives the consolidated report of the dataset used.

Table 1. Details of data samples collected.

<table>
<tr><td></td><td>Total</td><td colspan="2">Positive</td><td colspan="2">Negative</td></tr>
<tr><td>Scans:</td><td>2482</td><td colspan="2">1252</td><td colspan="2">1230</td></tr>
<tr><td>Patients:</td><td>120</td><td rowspan="2">60</td><td>Male 32</td><td rowspan="2">60</td><td>Male 30</td></tr>
<tr><td></td><td></td><td>Female 28</td><td>Female 30</td></tr>
</table>

The following images shown in Figure 2 are the sample CT scan images of affected and asymptotic patients, which are given as input to the preprocessor. The main idea of the proposed work is to detect positive patients who are asymptotic.

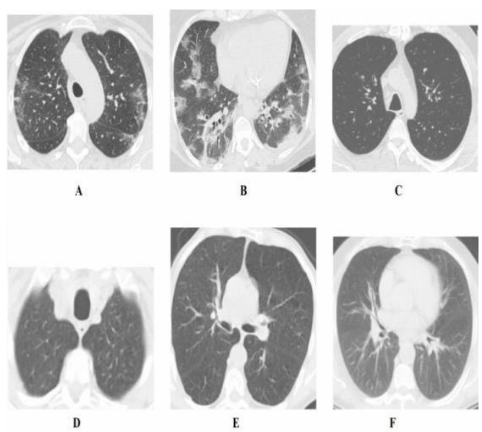


Figure 2. (A, B, C) are COVID-19 affected patient CT scan samples,

Figure 2. (D, E, F) are CT scan samples of negative patients.

IV. Hybrid Abnormal Data Detection Algorithm (HADDA)

The proposed work mainly focuses on the detection of virus in an asymptotic person, using CT scan image of lung. CT scan images are considered as input as they can be better analyzed than X-ray images. In this work, Recurrent Neural Networks (RNN) is used, which is one of the deep neural networks algorithms. In some of the traditional neural networks, the inputs and outputs belong to each other and these are independent, in some cases the next area of image is needed to predict the abnormal area of the image. RNN calculates the overall image size and area; this is accomplished by the hidden layer. To improve the performance of the RNN, the canny edge detection algorithm is adopted. This helps to detect edge boundaries of the affected area, more accurately. Figure 3 shows the architecture of proposed methodology.

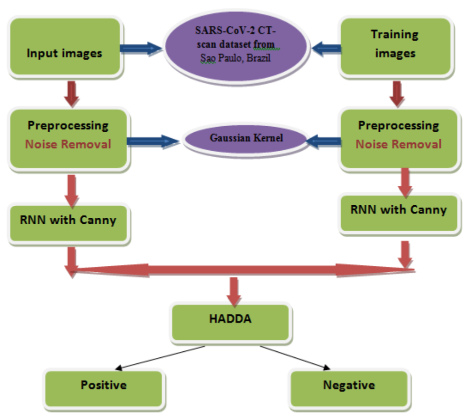


Figure 3. Architecture of Proposed methodology.

In this proposed algorithm, RNN is calculated based on the three layers that are combined together and the weights that are biased to all the hidden layers is same, then these are converted to single recurrent layer. The three states that are present in this RNN are given below.

$$h_t = f(h_{t-1}, x_t)$$

where $h_t \rightarrow$ present state

$h_{t-1} \rightarrow$ past state

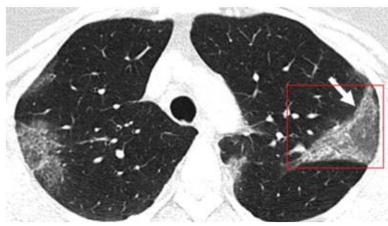
$x_t \rightarrow$ initialize state

Noise Reduction: Noise of the image to be reduced to avoid un-necessary calculations and processing. To accomplish this task, we have used

Gaussian blur to smooth it. The image convolution method is applied with a Gaussian Kernel. The kernel size depends on the expected blurring effect. The equation for a Gaussian filter kernel of size $(2k+1)\times(2k+1)$ is given by:

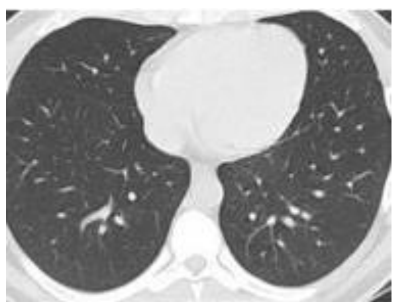
$$H_{ab} = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(a-(k+1))^2+(b-(k+1))^2}{2\sigma^2}\right); 1 \leq a, b \leq (2k+1).$$

The following are the images of COVID-19 positive (Figure 4(a)) and negative patient's (Figure 4(b)) CT scan images after noise reduction using Gaussian kernel. The abnormal detected is represented by arrow mark. No such abnormal area is detected in negative patient's CT scan.



(a)

Figure 4(a). COVID-19 positive patient's CT scan image (Figure 2(a)) after applying the noise reduction.



(b)

Figure 4(b). COVID-19 negative patient's CT scan image (Figure 2(f)) after applying the noise reduction.

V. Experimental Results

The performance is calculated based on the following metrics: Sensitivity, Specificity and Accuracy. The performance evaluation and comparison is also done using these metrics. This can be calculated for every input image and also to estimate the overall performance of the proposed system. The measurements are given below.

False Positive Rate (FPR)

This calculates the percentage of normal classified image.

$$FNR = \frac{FP}{FP + TN}.$$

False Negative Rate (FNR)

This calculates the abnormal area of the image.

$$FNR = \frac{FN}{FN + TN}.$$

Sensitivity

From all the actual positives (normal) calculates the sensitivity.

$$\text{Sesitivity} = \frac{No.\,of\ TP}{No.\,of\ TN + No.\,of\ FP}.$$

Specificity

From all the actual negatives (abnormal) calculates the specificity.

$$\text{Specificity} = \frac{No.\,of\ TN}{No.\,of\ TN + No.\,of\ FP}$$

Accuracy: This will calculate the overall accuracy of the overall normal and abnormal affected area.

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

VI. Results

The implementation is done with python programming language; Win-python is used to load the required libraries such as matplotlib, sci-kit learns, numpy, scipy for image manipulation and processing etc. System configuration consists of 4GB ram and 1TB hard disk is to be used.

Three performance metrics were taken into consideration and the comparison is done accordingly:

Table 2. Performance metrics.

Method	Sensitivity	Specificity	Accuracy
Soares et al. [8]	95.53	99.16	97.38
HADDA	97.80	99.60	98.78

The sensitivity, specificity and accuracy were improved with the proposed method. The same is shown below in Figure 5:

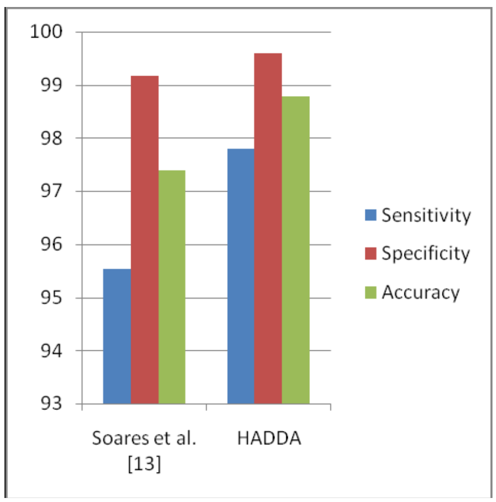


Figure 5. Performance Evaluation.

VI. Conclusion

This paper mainly focuses on detecting the abnormality on given input CT scan images that are taken from the [13] datasets. The algorithms divide the input images as layers and states. This can dynamically calculate the overall area of the image and radius of the every region in the image. Canny edge detection is adopted to find the edges of the input image and to find the accurate abnormality and decide whether it is serious or not. Thus the performance of detecting abnormality in the images has been evaluated and increased using HADDA.

Future scope

The future enhancement to the proposed methodology can be done by using CNN, instead of RNN and check for better accuracy. Feature Selection can also be added to improve the performance. More sophisticated and better training algorithms can be used to yield the accurate detection of abnormal area.

References

[1] World Health Organization, August 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[2] World meter, August 2020, https://www.worldometers.info/coronavirus/.

[3] F. Shi, J. Wang and J. Shi et al., Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19, IEEE Reviews in Biomedical Engineering, 2020.

[4] S. Ahuja, B. K. Panigrahi, N. Dey, T. Gandhi and V. Rajinikanth, Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices, Applied Intelligence, 2020.

[5] B. Liu, X. Gao, M. He, L. Liu and G. Yin, A fast online COVID-19 diagnostic system with chest CT scans, Proceedings of KDD 2020, New York, NY, USA, 2020.

[6] Barstugan, U. Ozkaya and S. Ozturk, Coronavirus (COVID-19) classification using CT images by machine learning methods, ArXiv, 2020, https://arXiv:2003.09424.

[7] H. X. Bai, R. Wang and Z. Xiong et al., AI augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT, Radiology, vol. 296, no. 3, pp. E156-E165, 2020.

[8] E. Soares, P. Angelov, S. Biaso, M. H. Froes and D. K. Abe, Sars-cov-2 CT-scan dataset: a large dataset of real patients CT scans for sars-cov-2 identification. medRxiv; 2020.