Diagnosis of Coronavirus Disease 2019 (COVID-19) using CT images

Chandragiri Charan

Dept of Computer Science, SV University, Tirupati

***Abstract*—** *Early diagnosis of the coronavirus disease in 2019 (COVID-19) is essential for controlling this pandemic. COVID-19 has been spreading rapidly all over the world. There is no vaccine available for this virus yet. Fast and accurate COVID-19 screening is possible using computed tomography (CT) scan images. The deep learning techniques used in the proposed method is based on a convolutional neural network (CNN). Our manuscript focuses on differentiating the CT scan images of COVID-19 and non-COVID 19 CT using different deep learning techniques. A self-developed model named CTnet-10 was designed for the COVID-19 diagnosis, having an accuracy of 82.1%. Also, other models that we tested are DenseNet-169, VGG-16, ResNet-50, InceptionV3, and VGG-19. The VGG-19 proved to be superior with an accuracy of 94.52% as compared to all other deep learning models. Automated diagnosis of COVID-19 from the CT scan pictures can be used by the doctors as a quick and efficient method for COVID-19 screening.*

I. INTRODUCTION

The new coronavirus infection was first reported in Wuhan, China, and since then it has strongly spread out since January 2020 worldwide. The World Health Organization (WHO) declared the outbreak from the Coronavirus disease 2019 (COVID-19) to be a public health emergency of international concern on the 30th of January, 2020. COVID-19 is a respiratory ailment caused by the coronavirus. The most common symptoms include fever, fatigue, dry cough, loss of appetite, body aches, and mucus. Some non-specific symptoms may include sore throat, headache, chills with shaking sometimes, loss of smell or taste, running nose, vomiting, or diarrhea. Symptoms may usually take 5 to 6 days to show after a person comes in contact with the virus. People with mild symptoms may recover on their own. People suffering from other health conditions such as diabetes or heart problems may suffer from serious symptoms. Animals are also able to transmit this infection while getting affected themselves. Mainly two similar viruses were reported earlier, which were severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS coronavirus). These viruses caused major respiratory problems and are zoonotic in nature.

In comparison to RT-PCR, the thorax computer tomography (CT) is possibly more reliable, useful, and quicker technology for the classification and assessment of COVID-19, in particular to the epidemic region. Almost all hospitals have CT image screening; hence, the thorax CT pictures can be used for the early detection of COVID-19 patients. However, the COVID-19 classification based on the thorax CT requires a radiology expert, and a lot of valuable time is lost. In the current scenario, the COVID-19 test results take more than 24 h to detect the virus in the human body. There is an urgent need to recognize the illness in the early stage and to put the infected immediately under quarantine because no specific drugs are available for COVID-19. The Chinese government reported that the diagnosis for confirmation of COVID-19 is done with the help of the real-time polymerase chain reaction (RT-PCR). RT-PCR suffers from high false-negative rates and utilizes a lot of time since the machine which is used for the test takes around 4–8 h to process the samples of the patients. The low sensitivity RT-PCR test is not satisfactory in the present pandemic situation. In some cases, the infected are not possibly recognized on time and do not receive suitable treatment. The infected can be assigned sometimes as COVID-19 negative because of a false-negative result. Hence, automated analysis of the thorax CT pictures is desirable to save the valuable time of the medical specialist staff. This will also avoid delays in starting treatment.

Deep learning is the most efficient technique that can be used in medical science. It is a fast and efficient method for the diagnosis and prognosis of various illnesses with a good accuracy rate. There are specifically trained models to classify the inputs into different categories desired by the programmers. In the medical field, they are used to detect heart problems, tumour's using image analysis, diagnosing cancer, and many other applications. It is also used to differentiate the CT scan images of the patients infected with COVID-19 as positive or not infected, i.e., negative. A self-developed model CTnet-10 was created having an accuracy of 82.1%. To improve the accuracy, we had also passed the CT scan image through multiple pre-existing models. We found that the VGG-19 model is best to classify the images as COVID-19 positive or negative as it

gave a better accuracy of 94.52%. A graphical representation of our proposed model is demonstrated. The CT scan image is passed through a VGG-19 model that categorizes the CT scan into COVID-19 positive or COVID-19 negative.

Coronaviruses have been around for many decades, and it has affected many animals/mammal species and human being. By March 11, 2020, the World Health Organization (WHO) declared the new coronavirus called the COVID-19, a pandemic, and it has brought the entire globe into a compulsory lockdown. Coronavirus is a family of RNA viruses that is capable of causing significant viral pathogens in humans and animals. Corona is medium-sized viruses with the largest viral RNA genome known. Coronavirus infects both birds and mammals, but the bat is host to the largest number of the viral genotype of coronavirus. So, the bat is the host and does not get infected. It can, however, spread the virus to a human. As of 24th of August 2020, there have been more than 23 million confirmed cases of coronavirus worldwide, with about 800,000 of such cases resulting in the death of the infected patient. This is spread around 216 countries, areas, or territories. However, around five million infected patients have recovered worldwide. The USA, Brazil, India, and Russia are the top four countries with the highest number of cases. Around 90 million tests have conducted in China, followed by the USA, Russia, and India, with 72 million, 33 million, and 32 million tests, respectively.

Testing COVID-19 involves analyzing samples that indicate the present or past presence of severe acute respiratory syndrome-associated coronavirus 2 (SARS-CoV-2). The test is done to detect either the presence of the virus or of antibodies produced in response to infection. COVID-19 diagnostic approach is mainly divided into two broad categories, a laboratory-based approach, which includes point of care-testing, nucleic acid testing, antigens tests, and serology (antibody) tests. The other approach is using medical imaging diagnostic tools such as X-ray and computed tomography (CT).

The laboratory-based tests are performed on samples obtained via nasopharyngeal swab, throat swabs, sputum, and deep airway material. The most common diagnostic approach is the nasopharyngeal swab, which involves exposing a swab to paper strips containing artificial antibodies designed to bind to coronavirus antigens. Antigens bind to the strips and give a visual readout [4]. The process is pretty fast and is employed at the point of care. The nucleic acid test has low sensitivity between 60-71%. On the other hand, Fang et al. showed that radiologic methods could provide higher sensitivity than that of lab tests.

The use of medical imaging tools is the second approach of COVID-19 virus detection. These tools are playing an important role in the management of patients that are confirmed or suspected to be infected with the virus. It is worthy of note that without clinical suspicion, findings from X-ray, or CT images are nonspecific as many other diseases could have a similar pattern.

Thoracic CT scan is the imaging modality of choice that plays a vital role in the management of COVID-19. Thoracic CT has a high sensitivity for diagnosis of COVID-19 which makes it a primary tool for COVID-19 detection. CT scan involves transmitting X-rays through the patient's chest, which are then detected by radiation detectors and reconstructed into high-resolution medical images. There are certain patterns to look out for in a chest CT scans which present themselves in different characteristic manifestations. The potential findings with 100% confidence for COVID-19 in thoracic CT images are and consolidation, air bronchograms, reverse halo, and peri lobular pattern.

The abovementioned findings are reports presented by a radiologist who specializes in interpreting medical images. Interpretation of these findings by expert radiologists does not have a very high sensitivity. Artificial intelligence (AI) has been employed as it plays a key role in every aspect of COVID-19 crisis management. AI has proven to be useful in medical applications since its inception, and it became widely accepted due to its high prediction and accuracy rates. In the diagnosis stage of COVID-19, AI can be used to recognize patterns on medical images taken by CT. Other applications of AI include, but not limited to, virus detection, diagnosis and prediction, prevention, response, recovery, and to accelerate research. AI can be used to segment regions of interest and capture fine structures in chest CT images, self-learned features can easily be extracted for diagnosis and other applications as well. A recent study showed that AI accurately detected COVID-19 and was also able to differentiate it from other lung diseases and community-acquired pneumonia. In this study, we review the diagnosis of COVID-19 by using chest CT toward AI.

II. PROBLEM STATEMENT

In Existing system, deep learning approaches such as CNN, which performed the feature extraction process automatically, were widely used in these researches.

Besides, pretrained networks were commonly used for the segmentation, feature extraction, and classification stages. Especially DenseNet121, ResNet50, ShuffleNet V2 were successfully reported by the researchers in the classification stages, while successful results were obtained with the images produced by UNet ++ at the segmentation stage. It was pointed out by

the researchers that many of the developed systems were modeled using the modifications or improvements pretrained networks to improve the classification accuracy of COVID-19 in CT images after preprocessing and segmentation stages. This has shown that widely used pretrained networks can be used very successfully at every stage of image classification. Some researchers classified COVID-19 cases using machine learning techniques instead of using deep learning approaches by extracting the features from the images and achieved high recognition results. This brings essential advantages in terms of learning speed.

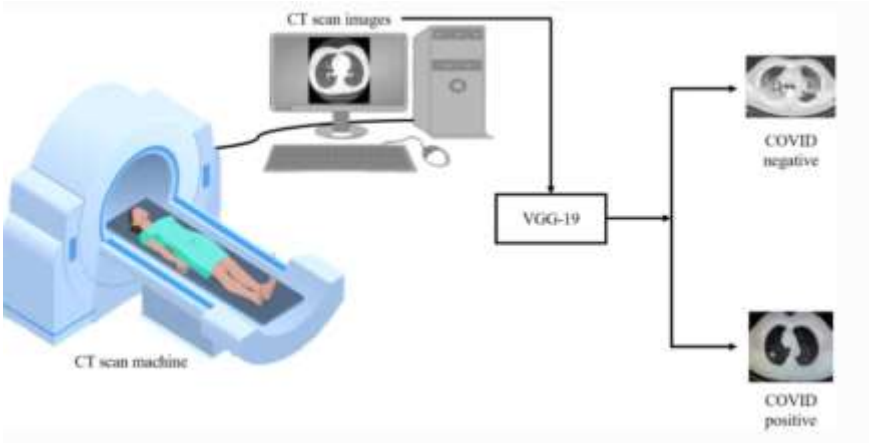
However, while the images used are not standard and performing experiments on different image databases in each research does not make it possible to make a comprehensive comparison, it contributes to deduce general opinion. While the -fold cross-validation is time-consuming, a few of the researches used it, and most of the researchers performed experiments using a hold-out method, which is based on dividing the dataset into training and testing set with defined percentages. However, this makes it challenging to analyze the consistency of the models, but it does not reduce the importance of performed experiments, obtained results, and the role of artificial intelligence in the fight against COVID-19.

III. METHODOLOGY

The COVID-19 CT dataset consisted of the images of patients that had tested positive for COVID-19 and the subsequent was also confirmed by the RT-PCR method. From a total of 738 CT scan images, 349 images from 216 patients were confirmed to have COVID-19 whereas 463 images were of the non-COVID-19 patients. These images were split into a training set, validation set, and test set with a split of 80%, 10%, and 10% respectively. The workflow diagram of the proposed system. The CT scan procedure starts by either walk-in or getting an appointment. It is then followed by the registration and the filling of the prep or consent form by the patient. The procedure for the examination of the CT scan by the radiologist can be done in two ways. The first way consists of getting a wet film. After making the payment, the wet film is handed to the patient. In the second way, the wet film captured by the radiographer is given to the radiologist for preparing a report. The patient then collects the report. The CT scan images are then fed to the deep learning models for detecting COVID-19. After the examination, the CT scan images can be directly fed to the deep learning model to classify the CT scan images as COVID-19 positive or COVID-19 negative.

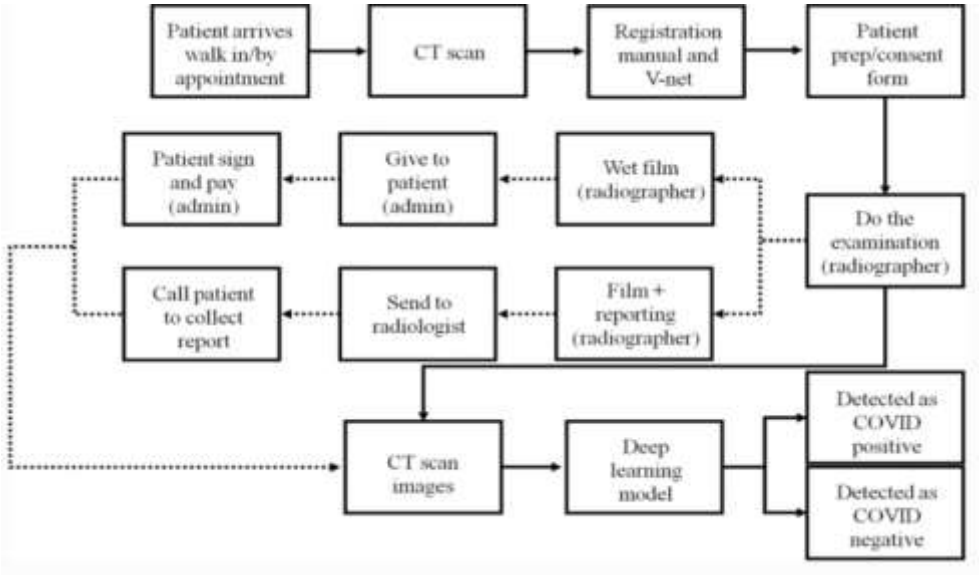
Our self-developed network (CTnet-10) was fed with an input image of dimension 128×128×3 and at every layer, the dimensions of the activations were changed. Convolutional blocks are major building blocks of neural networks. Pooling layers are helpful in reducing the number of computations to be performed. Both the convolutional block I and II consisted of 2 layers and a pooling layer with a flattening layer to convert 2D to 1D layer, which was fed to a dense layer of 256 neurons. A dropout of value 0.3 had been added in which the output layer contained 1 neuron and then the result had been predicted to classify the CT scan images.

IV. ARCHITECTURE



System flow diagram. The CT scan machine gives the CT scan image of a patient for the screening of COVID-19. The input image is passed through a VGG-19 model that categorizes the image as COVID-19 positive or negative.

4.1. WORKFLOW



Workflow diagram of the proposed system. The procedure of getting a CT scan starts by either walk-in or getting an appointment. It is followed by the registration and the filling of the prep or consent form by the patient. The actual procedure of getting a CT scan by the radiographer starts in two ways: one being the wet film and the other film and getting a report. The reports are then collected by the patients and the CT scan images can be fed to the deep learning models. After the examination, the CT scan images can be directly fed to the deep learning model to classify the CT scan images as COVID-19 positive or COVID-19 negative.

MODULE 1: Data Collection

Datasets are collected from Kaggle website. Google Scholar for AI for COVID-19 diagnosis with chest CT. Therefore, we categorized the studies into four main tasks as follows: COVID-19/normal, COVID-19/non-COVID-19, there are 747 images, 350 images are covid and 397 images are non-covid.

MODULE 2: Pre-Processing

Once the data is extracted from the medical source as the datasets, this information has to be passed to the classifier. The classifier cleans the dataset by removing redundant data like noise in order to make sure that CT Scan image is identified and removed before the analysis.

MODULE 3: VGG-19

For the VGG-19 model, the image dimensions used were 224×224×3, and the output was a number between 0 and 1. For this case, less than 0.5 corresponds to COVID-19 positive and greater than or equal to 0.5 implies COVID-19 negative. As mentioned above, we used VGG-19 architecture with pre-trained weights of imagenet. It is a 24- layer model (as shown in Fig. 4) which consists a total of 5 convolutional blocks, 3 max pool layers, and 3 FC layers, but we did a fine-tuning by using pre-trained weights for all convolutional blocks, removing the last two fully connected (FC) layer and then adding 2 FC layer with 4096 neurons. Dropout was used with each of these layers for regularization with a rate of 0.3. The final binary classification layer of single-neuron governed by sigmoid activation was added. The model was compiled with ADAM optimization with the default learning rate; the loss function used was binary cross-entropy. The model was trained on a batch size of 32 and Early Stoping was used to prevent overfitting. First, it was trained on 30 epochs without Early Stoping, and then on 20 epochs with Early Stoping, the model stopped at epoch no. 10.

Configuration of the model of VGG-19. The model was fed with an input image of size 224×224×3. There are a total of 5 convolutional blocks. It passes through one of the convolutional layers and a ramp of dimension 224×224×64. Then it passes through a next pooling and convolutional layer of dimension 112×112×128. Then it again passes through two pooling layers of dimensions 56×56×128, 28×28×512 respectively. This is further passed through the consecutive layer of dimensions 14×14×512, and a pooling layer of 7×7×512. It is then passed through 25088 neurons of the flattened layer, which is consecutively passed through an FC layer of 4096 neurons, in which the dropout layer was used in each of these. After passing it through a single neuron sigmoid and linear, the CT scan images are classified as COVID-19 positive or negative.

MODULE 4: **Model Fitting**

Detection of COVID-19 in CT scans continue, the researchers who take into account the peer-review period in the journals share the results they obtained in their studies with other researchers and scientists as preprints in different publication environments. Machine learning is used to make decisions on tasks that people have difficulty making decisions or problems that require more stable decisions using both numerical and image-based data. A deep convolutional neural network (CNN) is the most widely used among machine learning methods. It is one of the first preferred neural networks, especially in image-based problems, since it contains both feature extraction and classification stages and produces very effective results. In image-based COVID-19 researches, the CNN model or different models produced from CNN are widely encountered. In the researches, a generally hold-out method and a few -fold cross-validation were used during the training phase. In the hold-out method, while training is done by dividing the data into two parts as test and train, in -fold cross-validation, the data is divided into -folds, and the folds are trained -times by shifting the testing fold in each training so that each fold is used in the test phase. It is used as a better method for model evaluation.

MODULE 5: Detection

The performance of different models used for the study was evaluated using the confusion matrix The confusion matrix is extremely useful for calculating the accuracy of the models and represents the information in a better and understandable format. Amongst all the models that we studied, VGG-19 gave the least error in classifying the CT scans into COVID-19 and non-COVID-19. In the case of the Inception V3 model, all the COVID-19 CT scan images were wrongly classified into non-COVID-19, thus giving the least accuracy for the classification.

V. CONCLUSION

COVID-19 continues to spread around the globe. New classification and prediction models using AI, together with more publicly available datasets, have been arising increasingly. However, the majority of the studies are from the preprint literature and have not peer-reviewed. Furthermore, many of them have different classification tasks. Some of the studies have been conducted with very limited data. The data used in the studies might have come from different institutions and different scanners. Therefore, pre-processing of the data to make the radiographic images more similar and uniform is important in terms of providing more efficient analysis and consistency. The lack of demographic and clinical information of the patients is another limitation of these studies. We believe as the more dataset on COVID-19 with are available, the more accurate studies will be conducted. These findings are promising for AI to be used in the clinic as a supportive system for physicians in the detection of COVID-19.

REFERENCES

[1] Zhang W, Du RH, Li B, et al. Molecular and serological investigation of 2019-nCoV infected patients: implication of multiple shedding routes. *Emerg Microbes Infect*. 2020;9(1):386-389.

[2] Pan Y, Zhang D, Yang P, Poon LLM, Wang Q. Viral load of SARS-CoV-2 in clinical samples. *Lancet Infect Dis*. Published online February 24, 2020.

[3] Dalva-Baird N, Alobuia WM, Bedavid E, et al. Racial and ethnic inequities in the early distribution of U.S. COVID-19 testing sites and mortality. *European Journal of Clinical Investigation*. 2021;51(11)

[4] Economic Policy Institute. Black Workers Face Two of the Most Lethal Preexisting Conditions for Coronavirus—Racism and Economic Inequality