2020

Using AI to Diagnose COVID-19 from Patient Chest CT Scans

Samuel Arellano  
*Southern Methodist University*, capstonesmu2020@gmail.com

Liang Huang  
*Southern Methodist University*, lianghuang.ca@gmail.com

Joe Jiang  
*Southern Methodist University*, yanliu3936@gmail.com

Kenneth Richardson  
*Southern Methodist University*, klrichardso1@gmail.com

Omar Yasser  
*Columbus State University*, yasser_omar@columbusstate.edu

Follow this and additional works at: [https://scholar.smu.edu/datasciencereview](https://scholar.smu.edu/datasciencereview)

Part of the Other Analytical, Diagnostic and Therapeutic Techniques and Equipment Commons

**Recommended Citation**

Arellano, Samuel; Huang, Liang; Jiang, Joe; Richardson, Kenneth; and Yasser, Omar (2020) "Using AI to Diagnose COVID-19 from Patient Chest CT Scans," *SMU Data Science Review*: Vol. 3: No. 2, Article 18. Available at: [https://scholar.smu.edu/datasciencereview/vol3/iss2/18](https://scholar.smu.edu/datasciencereview/vol3/iss2/18)

This Article is brought to you for free and open access by SMU Scholar. It has been accepted for inclusion in SMU Data Science Review by an authorized administrator of SMU Scholar. For more information, please visit [http://digitalrepository.smu.edu](http://digitalrepository.smu.edu).
Using AI to Diagnose COVID-19 from Patient Chest CT Scans

Kenneth Richardson, Liang Huang, Joe Jiang, Sam Arellano, and Omar Youssef
Master of Science in Data Science, Southern Methodist University, Dallas, TX 75275 USA

Abstract. Rapid and accurate detection of COVID-19 remains the best weapon to control and prevent the spread of this pandemic, at least before a vaccine or treatment is available. In this study, we trained our custom computer vision models to predict COVID-19 from patients’ CT scans. We trained one model using Google’s AutoML Vision platform and achieved comparable accuracy with previously reported models. We also trained several custom models using transfer learning by taking advantage of several well-known pre-trained computer vision models, including Resnet and Inception models. The models are fine-tuned with a relatively large dataset and their high accuracy should make them more generalizable for potential clinical application.

1 Introduction

Shortly after being reported in Wuhan China in early December 2019, the World Health Organization (WHO) declared COVID-19 a global health emergency on January 30, 2020, and subsequently a global pandemic on March 11, 2020 [1]. Since then, SARS-CoV-2, the virus that causes the disease has proved to be highly contagious and reported to be capable of asymptomatic transmission. According to figures from John Hopkins University, as of November 12, 2020, the number of confirmed COVID-19 cases has reached 52.8 million worldwide with over 10.6 million of the cases in the US alone [2]. Currently, there are no Food and Drug Administration (FDA) approved vaccines or therapeutics for the treatment of COVID-19 patients. However, there have been varying degrees of social distancing measures implemented to limit the spread of the virus. The Centers for Disease Control and Prevention (CDC) recommends that people wear a face mask to cover their nose and mouth. The intention of recommending a face covering in communities is not only for the wearer’s protection but also to minimize the spread of the virus from the wearer to others. This is especially important in the case of asymptomatic individuals that are infected but show no flu-like symptoms [3]. Despite stringent public health measures and efforts to encourage social distancing, the virus appears to still be spreading rapidly around the world.

As COVID-19 is caused by the SARS-CoV-2 retrovirus, the common method to diagnose patients is through the use of RNA detection methods that are based on reverse transcription followed by polymerase chain reaction (RT-PCR). This PCR-based test is relatively unreliable, and as the stock of tests is still in fairly short supply
and increasingly high demand, they are not always readily available to healthcare agencies. Furthermore, these tests can have a low sensitivity ranging from 65%-95% giving them false-negative rates of up to 35%, meaning that in some cases up to a third of the tests indicated a patient was COVID-19 negative even though the patient was infected with the disease [4].

Although several approaches to diagnose COVID-19 using patient’s CT scans have been reported, they are not without challenges. Given the high demand, the number of CT scans needed to be examined, and the limited number of radiologists available to examine and diagnose them; manual diagnosis of CT scans has presented a considerable bottleneck. Therefore, an artificial intelligence-based approach is an appealing alternative for rapidly diagnosing patients.

Multiple studies have attempted to use machine learning and artificial intelligence to help diagnose the CT scans retrospectively. The number of images in the datasets used by these studies ranges from 106 to 5941 [5]. The sensitivity and accuracy of the reported models range from 0.8-1.0 and 0.65-1.0 respectively [6-8]. However, the conclusions are not without controversy with multiple professional medical societies strongly opposed to the use of AI to diagnose CT scans [9]. While our primary hypothesis is to test the feasibility of using AI to diagnose COVID-19 patients, we are equally interested in identifying the potential pitfalls of using this approach in the medical field.

The literature review section of this paper discusses the related published work. It includes early COVID-19 symptoms, primary methods of detection, and previously attempted uses of machine learning and artificial intelligence to diagnose COVID-19. Section 2.1 (SARS-CoV-2) describes in detail the symptoms of COVID-19, Reverse Transcription and Polymerase Chain Reaction (RT-PCR) method to detect the viruses, and the most common feature in COVID-19 infected lung CT scans, Ground Glass Opacity (GGO). Section 2.2 (Deep Learning vs. Traditional ML) introduces the history of CAD systems and expounds upon recent advances using deep learning to detect COVID-19. Section 3 (Data) describes the dataset acquisitions. Section 4 (Methods) describes the advantages of using the Auto ML Platform. Section 5 (Results) provides results of Auto ML and Pytorch Custom Models. Lastly, Section 6 (Discussion) summarizes our findings and conclusions.

## 2 Related Work

### 2.1 SARS-CoV-2

A growing body of research has identified bats as the evolutionary source of the SARS-CoV-2 virus. The current hypothesis is that the first transmission of the SARS-CoV-2 virus occurred when the virus jumped from bats to humans [10]. The virus was then transmitted from human to human thus becoming a global health crisis. The reported early symptoms of COVID-19 are cough, shortness of breath, fever, and other related flu-like symptoms with some infected patients never experiencing any of these and remaining asymptomatic. Unfortunately, significant portions of the infected
patients develop a lower respiratory tract infection (LRTI) that may require hospitalization. Furthermore, with an average death rate of 2%-5%, the SARS-CoV-2 virus has a much higher mortality rate than the coronaviruses that result in the common cold. The disease also spreads much faster than the coronavirus that caused the prior SARS outbreak. One key to prevent the further spread of COVID-19 hinges on finding reliable ways to diagnose patients rapidly.

The current primary method of diagnosing patients is based on the detection of viral RNA by Reverse Transcription and Polymerase Chain Reaction method (RT-PCR), commonly known as PCR testing. The upper respiratory samples, including nasopharyngeal swabs, oropharyngeal swabs, nasal aspirates, and nasopharyngeal washes, are broadly recommended. For patients exhibiting productive cough, lower respiratory samples, including sputum, BAL fluid, and tracheal aspirates, are recommended [11]. Unfortunately, if the virus is not accumulated in the swabbed areas, the tests may fail to diagnose a COVID-19 patient correctly. According to one study, SAR-CoV-2 could be reliably detected in sputum followed by nasal swabs in the first 14 days after onset, whereas throat swabs were unreliable 8 days after onset. The reliability problem presents a challenge to correctly identifying COVID patients with earlier tests yielding only about 70% accuracy. Given the variability in viral loads, a negative test result from a respiratory sample does not rule out the disease. Furthermore, in some cases, the false-negative rates can be up to 35%, rendering the test useless [11]. While some of these false negatives could stem from improper sampling techniques, the false negatives can also arise from where the viral loads are low, or mutations in the viral genome. Another big challenge is the lack of sufficient diagnostic kits at medical facilities in many countries. Even with government intervention, the stock of PCR tests is still in relatively short supply. This has resulted in a backlog of tests in some cases. The problem is further compounded by the long wait time for test results to come back from the lab. As a result, urgent treatments for patients with acute illness may be delayed. In addition, several PCR tests might be needed to correctly diagnose this disease, resulting in delayed patients treatment and quarantine.

Among the clinical manifestations of virus infection, ground-glass opacity (GGO) pattern is the most common feature in COVID-19 infected lung CT scans. These patterns are usually multifocal, bilateral, and peripheral, but may present as a unifocal lesion in the early phase of the disease, commonly located in the lobe of the right lung. In the later stage of the disease where there is a thickened interlobular and intralobular lines in combination with a GGO pattern called “crazy paving” [12]. Given these prominent features, computed tomography (CT) lung scans have become a fast alternative method to diagnose patients with COVID-19 [13]. Chest CT scans are non-invasive and are used to take X-ray measurements at different angles across a patient’s chest to produce cross-sectional images. Radiologists who analyze these images will look for abnormal features that can help lead to a COVID diagnosis. Several retrospective studies have shown that the CT scans have a higher sensitivity and better false-negative rates diagnosing COVID-19 when compared to PCR testing. However, as imaging features overlap with other viral pneumonia, the specificity can be low [14]. Below are sample chest CT scan images of confirmed COVID-19 patients.
Fig 1. 59-year male with confirmed COVID-19 [14]. However, the PCR-test result was a false-negative. The arrow indicates areas of ground-glass opacity (GGO) and massive consolidation in the posterior parts of the lower lobes.

2.2 Deep Learning vs. Traditional ML

Computer vision using transfer learning for deep learning has become one of the most popular applications in the world of AI for medical imaging. Transfer learning aims to optimize the pre-trained CNN models to the custom datasets and objectives. In practice, the last layer of the pre-trained models is replaced by a custom layer suited for the specific task. The advantages of transfer learning are two-fold, first, the model training is fast because the majority of the model parameters are frozen and unchanged during training; second, perhaps more importantly, because the frozen layers have already learned to recognize features, training a custom model only requires several hundred to several thousand images, instead of the millions of images required for training a CNN model from scratch.

An early use of traditional machine learning was introduced in scientific literature a half-century ago. Lodwick, a medical doctor from Iowa first introduced the term CAD, “Computer-Aided Diagnosis” this term in 1966. He developed a system for predicting the prognosis of lung cancer patients by their posterior, anterior, and lateral chest radiographs. His general approach was to convert visual images on roentgenograms into numerical sequences that could be manipulated and evaluated by the digital computer [15]. Today we call this method “training a classifier” with feature vectors. Although serious attempts at CAD development began in the 1980s, the CAD systems have only recently become an inseparable part of routine clinical practice. In particular, these systems have assisted radiologists in detecting lung abnormalities. The predictions from these systems are not only reproducible but also capable of detecting subtle changes that are easily missed by visual inspection [16].

In this study, a proposed CAD system based on deep learning was implemented to distinguish COVID-19 from non-COVID-19 lung infections. 510 COVID-19 and 510
non-COVID-19 image patches were extracted from CT slices by a radiologist and were used to train on ten different pre-trained computer vision models, including AlexNet, VGG-16, VGG-19, SqueezeNet, GoogleNet, MobilNetV2, ResNet-18, ResNet-50, ResNet-101, and Xception. Five performance indices, including Accuracy, Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and the SUC scores were used to compare the diagnosis from the models and the radiologists themselves. The overall AUC scores range from 0.894-0.994. The AUC score from the radiologist's diagnosis was 0.873 with 89.21% sensitivity, 83.33% specificity, 86.27% accuracy, 84.25% PPV, and 88.54% NPV. Based on the image patches, the radiologist had worse performance with a 0.603 AUC, 61.76% sensitivity, 58.82% specificity, 60.29% accuracy, 60.00% PPV, and 60.61% NPV [13], suggesting that there is great potential for computer vision model to improve the current COVID-19 diagnosis through chest CT scans [16]. Among the pre-trained CNN models, ResNet-101 and Xception have been shown to have the best performance in training CT images to diagnose COVID-19 patients.

Another recent paper reported a Lesion-Attention Deep Neural Network (LA-DNN) model developed by researchers from Hong Kong University to predict COVID-19 positive or negative from a richly annotated chest CT image dataset. The CT image dataset contains 746 public chest CT images of COVID-19 patients collected from over 760 preprints, and the data annotations are accompanied by the textual radiology reports. Two types of information were extracted from the images, the first is whether an image indicates a positive or negative case of COVID-19, and the other is the description of five types of lesions on the CT images associated with the positive cases. The proposed data-driven LA-DNN model focuses on the primary task of binary classification for COVID-19 diagnosis, while an auxiliary multi-label learning task is implemented simultaneously to draw the model's attention to the five types of lesions during the training. The joint task learning process makes it a highly sample-efficient deep learning model that can learn COVID-19 radiology features effectively with very limited samples [17].

The Dutch Radiological Society developed the Coronavirus Disease 2019 (COVID-19) Reporting and Data System (CO-RADS) based on other efforts for standardization, such as the Lung Imaging Reporting and Data System or Breast Imaging Reporting and Data System. CO-RADS assesses the suspicion for pulmonary involvement of COVID-19 on a scale from 1 (very low) to 5 (very high). The system is meant to be used in patients with moderate to severe symptoms of COVID-19. The system was evaluated using 105 chest CT scans of patients admitted to the hospital with clinical suspicion of COVID-19 and in whom PCR test was performed (mean, 62 years ± 16 [standard deviation]; 61 men, 53 with positive PCR results). Eight observers used CO-RADS to assess the scans. Fleiss $k$ value was calculated, and scores of individual observers were compared with the median of the remaining seven observers. The resulting area under the receiver operating characteristics curve (AUC) was compared with results from the PCR test and clinical diagnosis of COVID-19. CO-RADS performs very well in predicting COVID-19 in patients with moderate to severe symptoms and has a substantial interobserver agreement, especially for categories 1 and 5 [18].
3 Data

We identified a number of potential datasets that are available from different resources [19]. After research and trials, we allocated 3 datasets for our study these are summarized in (Table 1).

Table 1. Sources of CT Scans for custom COVID-19 prediction models

<table>
<thead>
<tr>
<th>Resource</th>
<th>Description</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIMCV- COVID19+ Dataset</td>
<td>This dataset contains 314 samples of COVID-19 positive and COVID-19 negative CT scans from patients. [20]</td>
<td>Total of 314 Images 135 COVID 179 Non-COVID</td>
</tr>
<tr>
<td>MosMed COVID 19 CT Scans</td>
<td>This dataset contains anonymized human lung computed tomography (CT) scans with COVID-19 related findings, as well as without such findings. In total, there are 1000 CT scans each from a unique patient. [21]</td>
<td>20 Patients 1000 CT images per patients</td>
</tr>
<tr>
<td>Sao Paulo SARS-CoV-2 CT-scan Dataset</td>
<td>This dataset was collected from real patients in hospitals from Sao Paulo, Brazil[22].</td>
<td>Total 2482 CT Scans, 1252 positives and 1230 negative for COVID-19</td>
</tr>
</tbody>
</table>

One of the datasets is a blend of 2D images from Tongji Hospital, Wuhan, China [20]. A senior radiologist released this dataset and has addressed questions and comments from researchers around the world. The major concerns are summarized as follows. First, when the original CT images are put into papers, the quality of these images is degraded, which may render the diagnosis decisions less accurate. The quality degradation includes: the Hounsfield Unit (HU) values are lost; the number of bits per pixel is reduced; the resolution of images is reduced. Second, the original CT scan contains a sequence of CT slices, but when put into papers, only a few key slices are selected, which may have a negative impact on the diagnosis as well.

However, the issues raised in these concerns do not appear to significantly affect the accuracy of diagnosis. First, experienced radiologists are able to make an accurate diagnosis from low-quality CT images. For example, given a photo taken by a smartphone of the original CT image, experienced radiologists can make an accurate diagnosis by just looking at the photo, though the CT image in the photo has much lower quality than the original CT image. Likewise, the quality gap between CT images in papers and original CT images does not appear to hurt the accuracy of diagnosis. Second, while it is preferable to read a sequence of CT slices, oftentimes a slice of CT contains enough clinical information for accurate decision-making.
4 Methods

This study plans to take advantage of transfer learning in computer vision and train a custom COVID-19 classifier using several pre-trained models in the PyTorch torchvision library; including resnet50, resnet101, and inception_v3. Fine-tuning these pre-trained models will be part of the process by freezing all the layers except the last fully connected layer (fc), which will then be replaced with a sigmoid layer for binary classification. Google AutoML vision will be used as a benchmark to compare to our custom model to make predictions from CT scans. This platform automatically reshuffles the images and assigns them into train, validation, and test datasets. The system will then output the most suitable/accurate model. For model evaluation, not only the overall accuracy (precision and recall) are provided, but also a confusion matrix will be generated in a visual format that provides insights into the model performance. One of the most appealing aspects of the AutoML platform is its ease to deploy the final trained model and make predictions on test images right away. However, you can also download the model and similarly deploy it locally or on other platforms. The added benefit is that we’re able to design a custom user interface for the application.

Fig 2. Sketch of Google AutoML Vision. The platform allows direct model training from images and convenient deployment on Google Cloud Platform.

Following are methods used to process original CT images before feeding to the designed DNN pipelines. U-Net is a convolutional neural network that was developed for biomedical image segmentation at the Computer Science Department of the University of Freiburg, Germany. U-Net is used to do lung segmentation. Pre masked lesion images are used to train neural networks to do lesion masking. Once we get the cropped lesion image, we use image augmentation techniques on them.
Fig 3. Pre masked lesion images are used to train neural networks to do lesion masking.

5 Results

5.1 AutoML Model

Even without the labeling of lesions and with limited data, our current models trained on the Google AutoML vision platform have already reached a similar level of precision and recall rate in cited published research papers. These results from Google AutoML will be used as a benchmark to compare with our custom trained models described in detail in the following section. Briefly, the process of training an AutoML model first involves uploading the images to Google Cloud Platform (GCP) in a data storage bucket, including both COVID and Non-COVID CT images in two different folders. Then using a Jupyter notebook, a data index CSV file was created to link the images and their label (folder name) to their google bucket path.
The next step is to use the index file to import the images as a dataset [20] into the Google AutoML platform (Figure 4). The platform automatically divides the images into training testing datasets to allow direct model training. Depending on the selected parameters, such as the desired precision level, the training itself can take as long as 16 node hours. After the model training has finished, a simple metric summarizes the model’s precision and recall scores, which help us understand how accurate the model is. With an average precision score of 0.97, our AutoML model is already on par with the published models.

Another useful model metric is the area under the precision-recall curve. It measures how well your model performs across all score thresholds. In AutoML Vision, this metric is called Average Precision (Figure 5).
Fig 5. The precision and recall scores of our AutoML model.

A more direct evaluation of the model performance is provided by a confusion matrix (Table 2). From the confusion matrix, we can see that the model performs extremely well on recognizing the COVID-19 CT images, but does poorly on Non-COVID CT scans, only correctly recognizing 78% of the images as Non-COVID.

Table 2. AutoML model confusion matrix

<table>
<thead>
<tr>
<th>True Label</th>
<th>Predicted Label</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-COVID-19</td>
<td>78%(14/18)</td>
<td>22%(4/18)</td>
</tr>
<tr>
<td>COVID-19</td>
<td>0%(0/13)</td>
<td>100%(13/13)</td>
</tr>
</tbody>
</table>

5.2 Pytorch Custom Models

We also took advantage of transfer learning in computer vision and trained our custom COVID-19 classifier using several pre-trained models in the torchvision library, including resnet50, resnet101, and inception_v3. We fine-tuned these models by freezing all the layers except the last fully connected layer (fc), which we replaced with a binary classification layer. We used a large dataset from Sao Paulo that contains around 2500 CT scans and trained the models for 200 epochs and the Resnet-50 model’s training and validation loss and accuracy curves are shown (Fig. 6).
Fig 6. The training and validation loss (left) and accuracy (right) of Resnet-50 models. The model was trained for 200 epochs on Google Collab.

The Resnet-50 model confusion matrix is shown (Table 3). The model has a high precision rate of 96.8% but suffers from a relatively high false-positive rate at 10.4%. Given the potential impact of misdiagnosed asymptomatic carriers on the continued spread of COVID-19, reducing the false-negative rate is critical.

Table 3. Resnet-50 model confusion matrix. The model has a high recall rate but suffers from a quite high false-positive rate.

<table>
<thead>
<tr>
<th>True Label</th>
<th>Predicted Label</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COVID-19</td>
</tr>
<tr>
<td>COVID-19</td>
<td>96.8%(122/126)</td>
</tr>
<tr>
<td>Non-COVID-19</td>
<td>10.4%(11/106)</td>
</tr>
</tbody>
</table>

The architecture of the Resnet-50 model and test accuracies of the three custom models are summarized in Fig. 8. Among the three models, we have similar test accuracies. Though Resnet50 had the highest model test accuracy, Resnet-101 exhibited the lowest false positive rate at 7.5%.
6 Discussion

6.1 Limitations

Both our AutoML vision and custom models’ performance metrics were on par with that of the published COVID-19 prediction models. However, there are at least three potential limitations concerning the results of this study. The first limitation is that our model training was completely dependent on publicly available datasets. There are conflicting claims about whether these datasets reflect the reality of all potential COVID-19 infected patients. Many infected COVID-19 patients are asymptomatic, and it is unclear to what degree these patients share the symptoms seen in the CT lung scans of hospitalized patients. This complicates the practical usage of CT lung scans and these diagnostic models as a frontline tool for diagnosing COVID-19 as we lack comparative scans of COVID-positive asymptomatic patients. A second limitation of this study is that there were no criteria provided for which lung position CT scanned image should be used for training the model and for predicting the patient’s COVID-19 status. Though, the study did attempt to synthesize all the images from each patient’s CT scans and extract features from those scans to train a prediction model. A third limitation is the overlapping symptoms that COVID-19 patients share with lung infections such as bronchitis and pneumonia. This may explain why models that focused on the binary classification of COVID-19 tended to have higher false-positive rates than models that focused on differentiating between the three conditions.

Fig. 8. Training custom models using transfer learning. A. The architecture of pre-trained Resnet-50 vision model with the last layer modifier to diagnose COVID-19 from CT scans [16]. B. The test accuracies of three different custom-trained models.
Despite these limitations, our findings highlight that trained models can achieve levels of precision that match and exceed those of experienced radiologists. Using these models alongside the established work of the Dutch Radiological Society could lead to computer vision assisted CO-RADS. This would allow for the training models that can determine the severity of COVID-19 using a standardized scale. Thus allowing allow health care professionals to accurately diagnose the degree of infection so that the appropriate level of care could be administered in a timely manner, increasing the chances of a patient’s recovery.

6.2 Future Research

As the spread of COVID-19 continues into winter, hospitals will likely find themselves treating a variety of respiratory illnesses concurrently. While our models focused primarily on determining whether a patient had COVID-19 or not, our models could be trained to differentiate between different types of respiratory diseases.

In addition to expanding the predictive capability of our models, it would be useful to extend the current findings by examining an integrated deep learning algorithms pipeline. The pipeline would encapsulate the best practices of producing transfer learning image classification models that include lung segmentation, lesion detection, masking, and lesion image augmentation. Furthermore, this pipeline would simulate the processing of CT scans with automated image data streaming in a sequential end-to-end workflow.

COVID-19 is still a relatively new disease and though a vaccine seems ever more likely, the high rate at which SARS-CoV-2 has been observed to mutate and its accelerated spread around the world may be indicators that COVID-19 will be pervasive if not perennial disease. As such, continued research into the viability of computer-aided diagnosis using CT scans along with the development of standardized metrics and processes may be a key to easing the burden of already taxed medical infrastructures.

7 Conclusion

The results of our research suggest that while the application of these models as frontline diagnostic tools may not be practical or effective, their use as diagnostic tools for hospitalized patients has great potential. Even with a vaccine, we anticipate that COVID-19 will be a concern into the near future and may potentially evolve into a seasonal if not perennial health risk. The implementation of these models as part of computer-aided diagnosis pipelines and their integration into current diagnostic practices may help make the identification of COVID-19 more efficient. This study shows that custom trained computer vision models can accurately predict COVID-19 infections and suggests that if implemented alongside a standardized workflow, they could greatly benefit overburdened radiology departments resulting in improved diagnosis of respiratory diseases and subsequently, better quality of care for patients.
References


19. The AIMI Center https://aimi.stanford.edu/resources/covid19/data (Date: Aug 10, 2020)