Automated COVID-19 Detection and Diagnosis Framework Based on Severity Assessment

Xiaohan Chen¹, ², Hongqiao Wen¹*, Shengzong Wu¹, ², Weishun Lan³ and Rong Zheng³*

¹National Engineering Laboratory for Fiber Optic Sensing Technology, Wuhan University of Technology, Wuhan, Hubei, 430070, China
²School of information engineering, Wuhan University of Technology, Wuhan, Hubei, 430070, China
³Maternal and Child Hospital of Hubei Province, Huazhong University of Science and Technology, Wuhan, Hubei, 430070, China
*Corresponding author’s e-mail: whq@whut.edu.cn and zhengrong617@163.com

Abstract. Computed tomography (CT) has been an important complementary indicator in the diagnosis of coronavirus disease 2019 (COVID-19). The pandemic of COVID-19 has led to a sharp increase in the number of suspected cases, which puts great pressure on radiologists. A computer-supported assisting methodology is essential to get the preliminary diagnosis regarding the pneumonia infection. In this paper, we proposed a deep learning framework for COVID-19 diagnosis and severity assessment using chest CT. The framework can not only distinguish COVID-19 patients from healthy people, but also assess the severity of patients as early or progressive stage, which makes patients with different conditions in baseline test get reasonable allocation of medical resources. The framework is composed of two modules: segmentation module and diagnosis module. Segmentation module is designed to extract the regions of interest and calculate the opacity percentage, while diagnosis module is utilized to identify suspect cases and divide them into three categories: health, early stage, and progressive stage. A total of 150 CT exams were used to train and test. An F1 score of 95.44% for COVID-19 detection and an F1 score of 90.87% for severity assessment are obtained. We also evaluated the influence of the opacity percentage calculated by the segmentation module on the classification results. By using the opacity percentage characteristic, the accuracy is improved from 94.16% to 97.42%.

1. Introduction
Coronavirus disease 2019 (COVID-19) caused by Severe Acute Respiratory Syndrome Corona-Virus 2 (SARS-CoV-2) has widely spread all over the world since the beginning of 2020 [1]. The World Health Organization (WHO) announced on January 30, 2020, that the outbreak was a Public Health Emergency of International Concern (PHEIC), and confirmed COVID-19 as a pandemic on March 11, 2020 [2]. The disease is rapidly affecting worldwide population with statistics quickly falling out of date. According to statistics of WHO, the COVID-19 cases between the days of December 31th, 2019, and August 18th, 2020 is reported 21,732,472 confirmed cases and 77,036 confirmed death all over the world, and 216 countries have been affected.

Real-time polymerase chain reaction (RT-PCR), as the golden standard for COVID-19 diagnosis, was used to confirm each suspect case. But confirming COVID-19 patients using serial RT-PCR tests...
is time-consuming and has been reported to suffer from high false negative rates on initial test [3][4]. On the other hand, because chest computed tomography (CT) scans collected from COVID-19 patients frequently show bilateral patchy shadows or ground glass opacity (GGO) in the lung [5][6], it has been used as an important complementary indicator in COVID-19 screening due to high sensitivity. However, the pandemic of COVID-19 has led to a sharp increase in the number of suspected cases, which improves the diagnostic burden of the radiologists. A computer-supported assisting methodology is essential to get the preliminary diagnosis regarding the pneumonia infection.

Artificial Intelligence (AI) based chest CT analysis systems have been proposed to detect COVID-19 from chest CT. Li et al. [7] proposed a deep learning model called COVNet which can accurately detect COVID-19 and differentiate it from community acquired pneumonia and other lung diseases. The COVNet framework consists of a RestNet50 [8] as the backbone, which takes a series of CT slices as input and generates features for the corresponding slices. The extracted features from all sliced are then passed to max-pooling layer and fully-connected layer to generate a probability score for each type. Butt el al. [9] reviewed a study, which compared multiple convolutional neural network (CNN) models to classify CT samples with COVID-19, Influenza viral pneumonia, or no-infection. The candidate infection regions were first segmented out using a three dimensional deep learning model from pulmonary CT image set. These separated images were then categorized into COVID-19, Influenza-A viral pneumonia and irrelevant to infection groups, together with the corresponding confidence scores using a location-attention classification model. Gozes et al. [10] proposed a rapidly developed AI-based automated CT image analysis tools which can achieve high accuracy in detection of coronavirus positive patients as well as visualization of infection fields. However, in combating COVID-19, not only detection tools are needed, but also assessment of patient severity is needed. Based on the severity level, the appropriate treatment procedure is to be planned which includes the need for the assisting devices, monitoring devices, and the choice of the drug and its dosage level [11].

In this study, an automated COVID-19 detection and diagnosis framework based on severity assessment is proposed using chest CT. Our contribution can be attributed to the following: (1) an end-to-end identification structure of COVID-19 is proposed, which can not only be used as an independent COVID-19 detection system, but also be connected with any existing model to evaluate the patient severity; (2) A COVID-19 quantitative indicator based on CT slice is proposed and evaluated. By using the indicator, the accuracy is improved from 94.16% to 97.42%.

2. Materials and Methods

2.1. Overview
The architecture we proposed is shown in Figure 1. The structure can be divided into two parts: segmentation module and diagnosis module. Segmentation module was designed to extract regions of interest (ROI). The diagnosis module was utilized to analysis the CT slices in baseline test and generate a confidence score of three categories: health, early stage, and progressive stage.

Figure 1. Overview of proposed framework
Given a CT slice, the segmentation module was firstly used to segment lung mask and infection mask. Both masks were utilized to calculate the percentage of opacity (PO) indicator. Meanwhile, lung crop was obtained by pixel-level combination of lung mask and original CT slice. Then lung crop features extracted by a CNN and PO value were integrated at a fully-connection layer. Finally, the predict results were obtained using a softmax activation.

2.2. Dataset
In this study, we retrospectively collected 150 CT exams (100 CT exams from COVID-19 patients, 50 CT exams from health people). Patients with COVID-19 underwent baseline test of chest CT and several follow ups. At baseline test, all COVID-19 patients were classified into four clinical types: mild, moderate, severe and critical type, based on the Diagnosis and Treatment Protocol of Novel Coronavirus (trial version 5th) from the National Health Commission of the People’s Republic of China. On this basis, we reclassified the clinical types. Mild was reclassified as early stage, moderate, severe and critical type as progressive stage. A total of 52 CT exams were labeled as early stage, and the rest of 48 CT exams were labeled as progressive stage. Every COVID-19 patient was confirmed with RT-PCR testing kit. The dataset structure is shown in Table 1.

| Dataset      | Number of CT Exams | Number of Selected CT Images |
|--------------|---------------------|-----------------------------|
| Segmentation | Early stage 7       | Health 5                    | 3462 |
|              | Progressive stage 3 |                             |      |
| Diagnosis    | Early stage 45      | Health 45                   | 3584 |
|              | Progressive stage 48|                             |      |
| Total        | Early stage 52      | Health 50                   |      |
|              | Progressive stage 48|                             |      |

A group of CT examination usually contains different number of images due to the difference of spatial resolution and personal difference. Moreover, not all images contain valuable information for diagnosis. It is necessary to select the appropriate images to train and test deep learning model. In addition, all three categories of CT exams should be covered to ensure the comprehensiveness of selected image. Three experienced radiologists participated in this work.

In ROI extraction step, we sampled 3462 images from 15 CT exams, including 7 CT exams of early stage, 3 CT exams of progressive stage, and 5 CT exams of healthy people, to take all situations into consideration and get more accurate segmentation results. For each image, the radiologists manually annotated ground truth of infections and lungs.

In COVID-19 diagnosis step, a total of 135 CT exams were used, including 45 CT exams of early stage, 45 CT exams of progressive stage, and 45 CT exams of healthy people. 3584 images were selected according to corresponding characteristics: (1) early stage, GGO was the main radiologic demonstration (2) progressive stage, the main radiologic demonstration included diffuse GGO, crazy-paving pattern and consolidation (3) health, no abnormal opacities related with COVID-19 should be found in selected images.

2.3. Data preprocess
The original CT slices were enhanced by performing Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm. CLAHE algorithm can effectively enhance the local contrast of image and suppress the noise amplification. Then bilinear interpolation was applied to rescale the image from 512x512 to 224x224. After that, image normalization was performed as the Equation (1) shows:

\[
X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}
\]  

Where \(X_{\text{norm}}\) is the normalized data, \(X\) is the original data, \(X_{\text{max}}\) and \(X_{\text{min}}\) are maximum and minimum values in the raw data, respectively.

Since data augmentation alleviates the over fitting by adding variants to the dataset, we applied data augmentation to enhance the generalization ability of our models. The augmented data were generated...
for each training patch using random image translation, rotation and horizontal or vertical flip. The translation step was selected from [0, 20] voxels, and the rotation angle was randomly selected from [0°, 20°]. Besides, the mode for filling new pixels as ‘nearest’.

2.4. ROI Extraction

In developing solutions for detecting COVID-19 disease, the segmentation of lung ROIs is a foundational step. It depicts the ROIs in lung CT images (such as lungs, lung lobes, bronchopulmonary segments, and infected regions or lesions) for further evaluation and quantification, while effectively filtering useless background information. Traditional thoracic CT ROIs extraction is normally done based on hounsfield unit (HU) value and digital image processing algorithms. In AI-based biomedical image segmentation, many excellent deep learning models have been proposed and applied to COVID-19 detection. For this study, two segmentation models were trained separately. U-Net [12], shown in Figure 2, is performed to segment the lung pixels. And U-Net++ [13], shown in Figure 3, is used to intercept the infection pixels from the background. It is important to guard further steps against irrelevant features that could severely affect reliable decision-making.

A total of 3462 image were used for segmentation. All training images and their segmentation maps, defined as the intersection of the areas marked by radiologists, were used to train the network in an end-to-end manner to minimize a combined loss. The combined loss function, as the following shows, consists of binary crossentropy loss and Dice loss [14] with a balance weight of 0.5.

\[
\text{Loss} = 0.5 \times \text{diceLoss} + 0.5 \times \text{bceLoss}
\]  

(2)
Where $\text{Loss}$ is the combined loss function, $\text{bceLoss}$ is the binary crossentropy loss, $\text{diceLoss}$ is the Dice loss.

Generally, Dice loss, which is used to measure the similarity between predicted value and true value, is an appropriate choice in medical image segmentation task. However, due to the confusion of dice loss optimization process, we added binary crossentropy loss to make the training process easier to optimize. Dice coefficient was used instead of accuracy to evaluate the overlap ratio between an automatically segmented infection region and the corresponding reference region provided by radiologist. A Laplace smoothing is joined into Dice coefficient to avoid denominator 0 and mitigate over fitting. Equation (3), (4), and (5) show Dice loss, Dice coefficient and binary cross entropy loss, respectively.

\[
diceLoss = 1 - \text{diceCoeff} \\
\text{diceCoeff} = 2 \times \frac{(X \cap Y) + 1}{X + Y + 1} \\
bceLoss = -\sum_{i=1}^{n} y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)
\]

Where $X$ is the predicted mask, $Y$ is the true mask, $\cap$ is the intersection operator, $y_i$ and $\hat{y}_i$ are predicted probability and true value.

The Adam optimizer [15] with a batch size of 32 was adopted as the optimizer. The maximum epoch number was set to 200 and the learning rate was set to 0.0005. Moreover, we randomly chose 20% of the training patches to form a validation set and terminate the training process even before reaching the maximum epoch number, if the error on the other 80% of training patches continues to decline but the error on the validation set stops decreasing.

### 2.5. Percentage of opacity

After admitting the patient, the imaging procedures, such as CT/X-ray data are considered to identify infected section and its level of severity. Generally, the percentage of opacity is an important factor to describe severity level. The opacity percentage goes from low to high, which indicates that the disease goes from mild to severe. The PO calculated from three dimensional data is usually more accurate than two dimensional data. Limited by the size of dataset, we proposed a slice-based PO. Firstly, the lung masks and the infection masks predicted by segmentation models were translated into binary image. Then, corresponding binary images were employed to calculated lung area and infection area separately. Finally, the PO achieved by the infection area divided by the lung area. The process can be expressed as follows:

\[
PO = \frac{S_{\text{infection}}}{S_{\text{lung}}} \times 100\%
\]

Where $PO$ is the percentage of opacity, $S_{\text{infection}}$ is the infection area, and $S_{\text{lung}}$ is the lung area.

Considering the differences of opacity area between left and right lungs, we calculated PO of left lung and PO of right lung individually. And the maximum value was took as the PO of this section. The process can be expressed as follows:

\[
PO_{\text{max}} = \text{MAX}(PO_{\text{left\_lung}}, PO_{\text{right\_lung}})
\]

Where $PO_{\text{max}}$ is the maximum value of left lung and right lung, $PO_{\text{left\_lung}}$ is the PO value of left lung, and $PO_{\text{right\_lung}}$ is the PO value of right lung, and $\text{MAX}(\cdot)$ is an function to get the maximum value.

### 2.6. COVID-19 Diagnosis

In the analysis of medical data, one of the biggest difficulties faced by researchers is the limited number of available datasets. Labeling this data by experts is both costly and time consuming. An effective way to achieve significant results in classification problems with limited data size is transfer learning. The
most prominent practice to perform transfer learning is exploiting the CNNs participated and stood out in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [16], which evaluates algorithms for object detection and image classification at large scale. In this work, we developed a Xception [17] Model using transfer learning with ImageNet weights. Xception consists of 36 convolution layers, which can be regarded as a linear stack of depthwise separable convolution layers with residual connections. The architecture is shown in Figure 4.

Figure 4. Xception structure.

The classical Xception network structure was used for image feature extraction. Global Average Pooling layer was followed for the dimensional reduction of data. The output of the Global Average Pooling layer was flattened to a 2048-dimensional feature vector and then converted into a 16-dimensional feature vector using a full-connect layer. The PO value related to severity assessment was first normalized to the same order of magnitude and then concatenated to this full-connect layer. The number of output layer neurons was set to 3 to obtain the final classification result together with the confidence score.

The mini-batch stochastic gradient descent [18] algorithm with a batch size of 32 was adopted as the optimizer. The maximum epoch number was set to 200 and the learning rate was set to 0.001. Moreover, an early stopping strategy was utilized to reduce overfitting during training stage. We randomly chose 20% of the training patches to form a validation set and terminate the training process even before reaching the maximum epoch number, if the error on the other 80% of training patches continues to decline but the error on the validation set stops decreasing.

It should be noted that, although we used the Xception network for this study, a DCNN of any arbitrary structure can be embedded into our work for feature extraction.

3. Results and Discussion

The program was written in Python® under a Windows® 10 operating system on a workstation with Nvidia Geforce RTX 2080Ti. A 5-fold cross validation was performed to evaluate the performance of our work. The training dataset was randomly divided into five parts. During each training, one of them was used as the test set, and the rest as the training set to form a 5-fold cross validation. The final score would then be the mean of five scores.

As shown in Table 2, the performance of segmentation model was assessed by Dice coefficient. Dice coefficient was used instead of accuracy to evaluate the overlap ratio between an automatically segmented infection region and the corresponding reference region provided by radiologist. It can be seen that lung segmentation task was more accurate than infection segmentation task, which is also proved in Figure 5.
Table 2. Comparison of segmentation results

| Model       | Task                      | Dice Coefficient (%) |
|-------------|---------------------------|----------------------|
| U-Net      | Lung Segmentation         | 98.126               |
| U-Net      | Infection Segmentation    | 84.15                |
| U-Net++    | Lung Segmentation         | 98.22                |
| U-Net++    | Infection Segmentation    | 86.47                |

Figure 5. Segmentation results.

The performance of classifier was assessed by F1 score, area under curve (AUC) and receiver operating characteristic (ROC) curve. F1 score takes into account both precision and recall of the classification model. It can be regarded as a harmonic average of model precision and recall. AUC is defined as the area under the ROC curve, and a classifier with a larger AUC value is better generally. As shown in Table 3 and Figure 6, we evaluated the performance in two condition. It is obviously that the classifier has more accurate results in distinguishing COVID-19 from health. This may be due to the imprecise segmentation results of infected regions. A finer segmentation result can improve this situation.

Table 3. Comparison of binary classification results

| Label                      | F1 Score (%) | AUC (%) |
|----------------------------|--------------|---------|
| health, COVID19            | 95.44        | 99.3    |
| Early stage, progressive stage | 90.87        | 98.5    |
Figure 6. ROC curve. Figure (a) shows the ROC curve under binary label of health and COVID-19. Figure (b) shows the ROC curve under binary label of early stage and progressive stage.

We also tested the effect of PO indicator on the classification results. The performance was evaluated by accuracy in three cases. All methods ran on the same computer, using the same dataset. As shown in Table 4, the model using PO has better classification ability. The accuracy was improved from 94.16% to 97.42% by using $P_{max}$ indicator, due to fully consideration of different infection severity of left and right lung.

| Model   | Indicator | Accuracy (%) |
|---------|-----------|--------------|
| Xception | None      | 94.16        |
| Xception | $PO$      | 95.31        |
| Xception | $P_{max}$ | 97.42        |

4. Conclusion
This paper proposed a deep learning framework for COVID-19 diagnosis and severity assessment using chest CT. The structure can be divided into two parts: segmentation module and diagnosis module. Segmentation module was designed to extract ROI. The diagnosis module was utilized to analysis the CT slices in baseline test and generate a confidence score of three categories: health, early stage, and progressive stage. We used 150 CT exams to train the model. An F1 score of 95.44% for COVID-19 dection and an F1 score of 90.87% for severity assessment are obtained. In future work, a finer segmentation model for infection segmentation will be our main concern.

References
[1] Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G. et al. (2020) A new coronavirus associated with human respiratory disease in China. Nature, 579(7798): 265-269.
[2] World Health Organization. (2020) World Health Coronavirus disease (covid-2019) situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus2019/situation-reports.
[3] Ai, T., et al. (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology, 200642.
[4] Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., and Ji, W. (2020) Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology, 200432.
[5] Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet, 395(10223): 497-506.
[6] Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., et al. (2020) Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. Jama, 323(11): 1061-1069.

[7] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., et al. (2020) Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology.

[8] He, K., Zhang, X., Ren, S., and Sun, J. (2016) Deep residual learning for image recognition. In IEEE conference on computer vision and pattern recognition. pp. 770-778.

[9] Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., et al. (2020) A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering.

[10] Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P. D., Zhang, H., Ji, W., et al. (2020) Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. arXiv preprint arXiv:2003.05037.

[11] Chen, J., Li, K., Zhang, Z., Li, K., & Yu, P. S. (2020). A survey on applications of artificial intelligence in fighting against covid-19. arXiv preprint arXiv:2007.02202.

[12] Ronneberger, O., Fischer, P., and Brox, T. (2015) U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention. Springer, Cham. pp. 234-241.

[13] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2018) Unet++: A nested u-net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham. pp. 3-11.

[14] Milletari, F., Navab, N., and Ahmadi, S. A. (2016) V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 fourth international conference on 3D vision (3DV). IEEE. pp. 565-571.

[15] Kingma, D. P., and Ba, J. (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[16] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition. IEEE. pp. 248-255.

[17] Chollet, F. (2017) Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1251-1258.

[18] Cherry, J. M., Adler, C., Ball, C., Chervitz, S. A., Dwight, S. S., Hester, E. T., et al. (1998) SGD: Saccharomyces genome database. Nucleic acids research, 26(1): 73-79.