COVID-FACT: A Fully-Automated Capsule Network-Based Framework for Identification of COVID-19 Cases from Chest CT Scans

Shahin Heidarian[1], Parnian Afshar[2], Nastaran Enshaei[2], Farnoosh Naderkhani[2], Moezedin Javad Rafiee[3], Faranak Babaki Fard[4], Kaveh Samimi[5], S. Farokh Atashzar[6,7], Anastasia Oikonomou[8], Konstantinos N. Plataniotis[9] and Arash Mohammadi[2]*

[1]Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada, [2]Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC, Canada, [3]Department of Medicine and Diagnostic Radiology, McGill University Health Center-Research Institute, Montreal, QC, Canada, [4]Biomedical Sciences Department, Faculty of Medicine, University of Montreal, Montreal, QC, Canada, [5]Department of Radiology, Iran University of Medical Science, Tehran, Iran, [6]Department of Electrical and Computer Engineering, New York University, New York, NY, United States, [7]Department of Mechanical and Aerospace Engineering, New York University, New York, NY, United States, [8]Department of Medical Imaging, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON, Canada, [9]Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada

The newly discovered Coronavirus Disease 2019 (COVID-19) has been globally spreading and causing hundreds of thousands of deaths around the world as of its first emergence in late 2019. The rapid outbreak of this disease has overwhelmed health care infrastructures and arises the need to allocate medical equipment and resources more efficiently. The early diagnosis of this disease will lead to the rapid separation of COVID-19 and non-COVID cases, which will be helpful for health care authorities to optimize resource allocation plans and early prevention of the disease. In this regard, a growing number of studies are investigating the capability of deep learning for early diagnosis of COVID-19. Computed tomography (CT) scans have shown distinctive features and higher sensitivity compared to other diagnostic tests, in particular the current gold standard, i.e., the Reverse Transcription Polymerase Chain Reaction (RT-PCR) test. Current deep learning-based algorithms are mainly developed based on Convolutional Neural Networks (CNNs) to identify COVID-19 pneumonia cases. CNNs, however, require extensive data augmentation and large datasets to identify detailed spatial relations between image instances. Furthermore, existing algorithms utilizing CT scans, either extend slice-level predictions to patient-level ones using a simple thresholding mechanism or rely on a sophisticated infection segmentation to identify the disease. In this paper, we propose a two-stage fully automated CT-based framework for identification of COVID-19 positive cases referred to as the "COVID-FACT". COVID-FACT utilizes Capsule Networks, as its main building blocks and is, therefore, capable of capturing spatial information. In particular, to make the proposed COVID-FACT independent from sophisticated segmentations of the area of infection, slices demonstrating infection are detected at the first stage and the second stage is responsible for classifying patients into COVID and non-COVID cases. COVID-FACT detects slices with infection, and identifies positive

COVID-19 cases using an in-house CT scan dataset, containing COVID-19, community acquired pneumonia, and normal cases. Based on our experiments, COVID-FACT achieves an accuracy of 90.82%, a sensitivity of 94.55%, a specificity of 86.04%, and an Area Under the Curve (AUC) of 0.98, while depending on far less supervision and annotation, in comparison to its counterparts.

Keywords: capsule networks, COVID-19, computed tomography scans, fully automated classification, deep learning

1 INTRODUCTION

The recent outbreak of the novel coronavirus infection (COVID-19) has sparked an unforeseeable global crisis since its emergence in late 2019. Resulting COVID-19 pandemic is reshaping our societies and people's lives in many ways and caused more than half a million deaths so far. In spite of the global enterprise to prevent the rapid outbreak of the disease, there are still thousands of reported cases around the world on daily bases, which raised the concern of facing a major second wave of the pandemic. Early diagnosis of COVID-19, therefore, is of paramount importance, to assist health and government authorities with developing efficient resource allocations and breaking the transmission chain.

Reverse Transcription Polymerase Chain Reaction (RT-PCR), which is currently the gold standard in diagnosing COVID-19, is time-consuming and prone to high false-negative rate (Fang et al., 2020). Recently, chest Computed Tomography (CT) scans and Chest Radiographs (CR) of COVID-19 patients, have shown specific findings, such as bilateral and peripheral distribution of Ground Glass Opacities (GGO) mostly in the lung lower lobes, and patchy consolidations in some of the cases (Inui et al., 2020). Diffuse distribution, vascular thickening, and fine reticular opacities are other commonly observed features of COVID-19 reported in (Bai et al., 2020; Chung et al., 2020; Ng et al., 2020; Shi et al., 2020). Although imaging studies and their results can be obtained in a timely fashion, such features can be seen in other viral or bacterial infections or other entities such as organizing pneumonia, leading to misclassification even by experienced radiologists.

With the increasing number of people in need of COVID-19 examination, health care professionals are experiencing a heavy workload reducing their concentration to properly diagnose COVID-19 cases and confirm the results. This arises the need to distinguish normal cases and non-COVID infections from COVID-19 cases in a timely fashion to put a higher focus on COVID-19 infected cases. Using deep learning-based algorithms to classify patients into COVID and non-COVID, health care professionals can exclude non-COVID cases quickly in the first step and allow for paying more attention and allocating more medical resources to COVID-19 identified cases. It is worth mentioning that although the RT-PCR, as a non-destructive diagnosis test, is commonly used for COVID-19 detection, in some countries with high number of COVID-19 cases, CT imaging is widely used as the primary detection technique. Therefore, there is an unmet need to develop advanced deep learning-based solutions based on CT images to speed up the diagnosis procedure.

1.1 Literature Review

Convolutional Neural Networks (CNNs) have been widely used in several studies to account for the human-centered weaknesses in detecting COVID-19. CNNs are powerful models in related tasks and are capable of extracting distinguishing features from CT scans and chest radiographs (Yamashita et al., 2018). In this regard, many studies have utilized CNNs to identify COVID-19 cases from medical images. The study by (Wang and Wong, 2020), is an example of the application of CNN in COVID-19 detection, where CNN is first pre-trained on the ImageNet dataset (Krizhevsky et al., 2017). Fine-tuning is then performed using a CR dataset. Results show an accuracy of 93.3% in distinguishing normal, non-COVID-19 pneumonia (viral and bacterial), and COVID-19 infection cases (Sethy et al., 2020). have also explored the same problem with the difference that the CNN is followed by a Support Vector Machine (SVM), to identify positive COVID-19 cases. Their obtained results show an overall accuracy of 95.38%, sensitivity of 97.29% and specificity of 93.47%. Another study by (Mahmud et al., 2020) proposed a CNN-based model utilizing depth-wise convolutions with varying dilation rates to extract more diversified features from chest radiographs. They used a pre-trained model on a dataset of normal, viral, and bacterial pneumonia patients followed by additional fine-tuned layers on a dataset of COVID-19 and other pneumonia patients, obtaining an overall accuracy of 90.2%, sensitivity of 89.9%, and specificity of 89.1%.

Chest radiograph acquisition is relatively simple with less radiation exposure than CT scans. However, a single CR image fails to incorporate details of infections in the lung and cannot provide a comprehensive view for the lung infection diagnosis. CT scan, on the other hand, is an alternative imaging modality that incorporates the detailed structure of the lung and infected areas. Unlike CR images, CT scans generate cross-sectional images (slices) to create a 3D representation of the body. Consequently, there has been a surge of interest on utilizing 2D and 3D CT images to identify COVID-19 infection. For instance (Yang et al., 2020), proposed a DenseNet-based model to classify manually selected slices with COVID-19 manifestations and pulmonary parenchyma into COVID-19 and normal classes. The underlying study achieved an accuracy of 92% on the patient-level classification by averaging slice-level probabilities followed by a threshold of 0.8 on the averaged values. Furthermore, the dataset used to train and test the model does not include other types of pneumonia. *Identified Drawback 1*: Such methods require manual selecting slices demonstrating infection to feed the model, which makes the overall process time-consuming and only partially automated. To extract features from all CT slices (Li et al., 2020), first segmented

the lung regions using a U-net based segmentation method (Ronneberger et al., 2015), and then used them to fine-tune a ResNet50 model, which was pre-trained on natural images from the ImageNet dataset (Deng et al., 2009). Extracted features are then combined using a max-pooling operation followed by a fully connected layer to generate probability scores for each disease type. Their proposed model achieved sensitivities of 90%, 87%, and 94% for COVID-19, Community Acquired Pneumonia (CAP), and non-pneumonia cases respectively. *Identified Drawback 2:* Such methods combine extracted features from all slices of a patient, with or without infection, which potentially results in lower accuracy as there are numerous slices without evidence of infection in a volumetric CT scan of an infected patient.

In the study by (Hu et al., 2020), segmented lungs are fed into a multi-scale CNN-based classification model, which utilizes intermediate CNN layers to obtain classification scores, and aggregates scores generated by intermediate layers to make the final prediction. Their proposed method achieves an overall accuracy of 87.4% in the three-way classification (Zhang et al., 2020). proposed a two-stage method consisting of a Deeplabv3-based lung-lesion segmentation model (Chen et al., 2017) followed by a 3D ResNet18 classification model (Hara et al., 2017) to identify lung lesions and abnormalities and use them to classify patients into COVID-19, community acquired pneumonia, and normal findings. They manually annotated chest CT scans into seven regions to train their lung segmentation model, which is a time-consuming and sophisticated task requiring high level of thoracic radiology expertise to accomplish. Their proposed method achieves the overall accuracy of 92.49% in both three-way and binary (COVID-19 vs. others) classifications.

1.2 Problem Statement

At one hand, we aim to address the two identified drawbacks of the aforementioned methods. More specifically, existing solutions either require a precise annotation/labeling of lung images, which is time-consuming and error-prone, especially when we are facing a new and unknown type of disease such as COVID-19, or assign the patient-level label to all the slices. On the other hand, CNN, which is widely adopted in COVID-19 studies, suffers from an important drawback that reduces its reliability in clinical practice. CNNs are required to be trained on different variations of the same object to fully capture the spatial relations and patterns. In other words, CNNs, commonly, fail to recognize an object when it is rotated or transformed. In practice, extensive data augmentation and/or adoption of huge data resources are needed to compensate for the lack of spatial interpretation. As COVID-19 is a relatively new phenomenon, large datasets are not easily accessible, especially due to strict privacy preserving constraints. Furthermore, most COVID-19 cases have been reported with a specific infection distribution in their image (Bai et al., 2020; Chung et al., 2020; Ng et al., 2020; Shi et al., 2020), which makes capturing spatial relations in the image highly important.

1.3 Contributions

As stated previously, structure of infection spread in the lung for COVID-19 is not yet fully understood given its recent and abrupt emergence. Furthermore, COVID-19 has a particular structure in affecting the lung, therefore, picking up those spatial structures are significantly important. Capsule Networks (CapsNets) (Hinton et al., 2018), in contrast to CNNs, are equipped with routing by agreement process enabling them to capture such spatial patterns. Even without a large dataset, capsules interpret the object instantiation parameters, besides its existence, and by reaching a mutual agreement, higher-level objects are developed from lower-level ones. The superiority of Capsule Networks over their counterparts has been shown in different medial image processing problems (Afshar et al., 2018; Afshar et al., 2019a; Afshar et al., 2019b; Afshar et al., 2020b; Afshar et al., 2020d; Afshar et al., 2020c). Recently, we proposed a Capsule Network-based framework (Afshar et al., 2020a), referred to as the COVID-CAPS, to identify COVID-19 cases from chest radiographs, which achieved an accuracy of 98.3%, a specificity of 98.6%, and a sensitivity of 80%. As stated previously, CT imaging is superior for COVID-19 detection and diagnosis purposes when compared to chest radiographs. However, as in the case of CT imaging, we are dealing with 3D inputs and several slices per patient (compared to one chest radiograph per patient), the learning process is significantly more challenging. As such, accuracies of deep models trained over CT scans cannot be directly compared with those obtained based on chest radiographs.

Following our previous study on chest radiographs, in the present study, we take one step forward and propose a fully automated two-stage Capsule Network-based framework, referred to as the COVID-FACT, to identify COVID-19 patients using chest CT images. Based on our in-house dataset, COVID-FACT achieves an accuracy of 90.82%, sensitivity of 94.55%, specificity of 86.04%, and Area Under the Curve (AUC) of 0.98. We developed two variants of the COVID-FACT, one of which is fed with the whole chest CT image, while the other one utilizes the segmented lung area as the input. In the latter case, instead of using an original chest CT image, first a segmentation model (Hofmanninger et al., 2020) is applied to extract the lung region, which is then provided as input to the COVID-FACT. This will be further clarified in **Section 3**. Experimental results show that the model coupled with lung segmentation achieves the same overall accuracy compared to the other COVID-FACT variation working with original images. However, using the segmented lung regions increases the sensitivity and AUC from 92.72% and 0.95 to 94.55% and 0.98, respectively, while slightly decreasing the specificity from 88.37% to 86.04%.

COVID-FACT benefits from a two-stage design, which is of paramount importance in COVID-19 detection using CT scans, as a CT examination is typically associated with hundreds of slices that cannot be analyzed at once. At the first stage, the proposed COVID-FACT detects slices demonstrating infection in a 3D volumetric CT scan to be analyzed and classified at the next stage. At the second stage, candidate slices detected at the previous stage are classified into COVID and non-COVID (community acquired pneumonia and normal) cases and a voting mechanism is applied to generate the classification scores in the patient level. COVID-FACT's two-stage architecture has the advantage of being trained by even weakly labeled dataset,

as errors at the first stage can be compensated at the second stage. As a result, COVID-FACT does not require any infection annotation or a very precise slice labeling, which is a valuable asset due to the limited knowledge and experience on the novel COVID-19 disease. In fact, manual annotation is completely removed from the COVID-FACT. The only information required from the radiologists to train the first stage is the slices containing evidence of infection. In other words, COVID-FACT is not dependent on the manual delineation of specific infected regions in the slices, which is a complicated and time-consuming task compared to only identifying slices with the evidence of infection. This issue is more critical in the case of a novel disease such as COVID-19, which requires comprehensive research to identify the disease manifestations. It is worth noting that the pre-trained lung segmentation model used as the pre-processing step in our study is related to the well-studied lung segmentation task, which is totally different from the infection segmentation. As a final note, we would like to mention that the radiologist's input is not required in the test phase of the COVID-FACT and the trained framework is fully automated.

The reminder of the paper is organized as follows: **Section 2** describes the dataset and imaging protocol used in this study. **Section 3** presents a brief description of Capsule Networks and explains the proposed COVID-FACT in details. Experimental results and model evaluation are presented in **Section 4**. Finally, **Section 5** concludes the work.

2 MATERIALS AND EQUIPMENT

In this section, we will explain the in-house dataset used in this study, along with the associated imaging protocol.

2.1 Dataset

The dataset used in this study, referred to as the "COVID-CT-MD" Afshar et al. (2021), contains volumetric chest CT scans of 171 patients positive for COVID-19 infection, 60 patients with Community Acquired Pneumonia (CAP), and 76 normal patients acquired from April 2018 to May 2020. The average age of patients is 50 ± 16 including 183 men and 124 women. This dataset and the related annotations are publicly available through Figshare at https://figshare.com/s/c20215f3d42c98f09ad0.

Diagnosis of COVID-19 infection is based on positive real-time reverse transcription polymerase chain reaction (rRT-PCR) test results, clinical parameters, and CT scan manifestations by a thoracic radiologist, with 20 years of experience in thoracic imaging. CAP and normal cases were included from another study and the diagnosis was confirmed using clinical parameters, and CT scans. A subset of 55 COVID-19, and 25 community acquired pneumonia cases were analyzed by the radiologist to identify and label slices with evidence of infection as shown in **Figure 1**. This labeling process focuses more on distinctive manifestations rather than slices with minimal findings. The labeled subset of the data contains $4,962$ number of slices demonstrating infection and $18,447$ number of slices without infection. The data is then used to train and validate the first stage of our proposed COVID-FACT model to extract slices
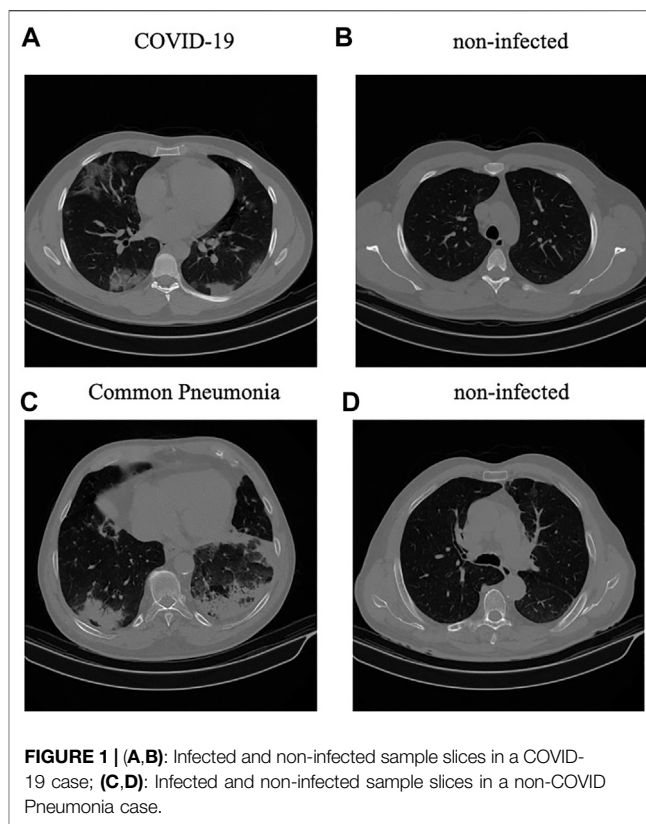

FIGURE 1 | (A,B): Infected and non-infected sample slices in a COVID-19 case; **(C,D)**: Infected and non-infected sample slices in a non-COVID Pneumonia case.

demonstrating infection from volumetric CT scans to be used in the second classification stage. We have randomly divided this subset into three separate components for training, validation, and testing. 60% of the labeled data is used for training, 10% for validation, and 30% for the test. The unlabeled subset is also randomly divided with the same proportion and used along with the labeled data to develop the second stage of the COVID-FACT model and evaluate the overall method. The data leakage between the train and test sets has been prevented. In other words, all slices related to a patient are included either in the train or the test dataset. This research work is performed based on the policy certification number 30013394 of Ethical acceptability for secondary use of medical data approved by Concordia University. Furthermore, informed consent is obtained from all the patients. Finally, the dataset is complied with the DICOM supplement 142 (Clinical Trial De-identification Profiles) DICOM Standards Committee, Working Group 18 Clinical Trials (2011), indicating that all CT studies are de-identified by either removing or obfuscating the patient and center-related information such as names, UIDs, dates, times, and comments based on the directions specified in DICOM Standards Committee, Working Group 18 Clinical Trials (2011).

2.2 Imaging Protocol

All CT examinations have been acquired using a single CT scanner with the same acquisition setting and technical parameters, which are presented in **Table 1**, where kVP (kiloVoltage Peak) and Exposure Time affect the radiation exposure dose, while Slice Thickness and Reconstruction

Scanner manufacturer and model	Slice thickness (mm)	Image type	kVP (kV)	Exposure time (ms)	Reconstruction matrix	Window center	Window width
SIEMENS, SOMATOM scope	2	Axial	110	600	512×512	(50, −600)	(350, 1200)

Matrix represent the axial resolution and output size of the images, respectively Raman et al. (2013). Next, we describe the proposed COVID-FACT framework followed by the experimental results.

3 METHODS

The COVID-FACT framework is developed to automatically distinguish COVID-19 cases from other types of pneumonia and normal cases using volumetric chest CT scans. It utilizes a lung segmentation model at a pre-processing step to segment lung regions and pass them as the input to the two-stage Capsule Network-based classifier. The first stage of the COVID-FACT extracts slices demonstrating infection in a CT scan, while the second stage uses the detected slices in first stage to classify patients into COVID-19 and non-COVID cases. Finally, the Gradient-weighted Class Activation Mapping (Grad-CAM) localization approach (Selvaraju et al., 2017) is incorporated into the model to highlight important components of a chest CT scan, that contribute the most to the final decision.

In this section, different components of the proposed COVID-FACT are explained. First, Capsule Network, which is the main building block of our proposed approach, is briefly introduced. Then the lung segmentation method is described, followed by the details related to the first and second stages of the COVID-FACT architecture. Finally, the Grad-CAM localization mapping approach is presented.

3.1 Capsule Networks

A Capsule Network (CapsNet) is an alternative architecture for CNNs with the advantage of capturing hierarchical and spatial relations between image instances. Each Capsule layer utilizes several capsules to determine existence probability and pose of image instances using an instantiation vector. The length of the vector represents the existence probability and the orientation determines the pose. Each Capsule i consists of a set of neurons, which collectively create the instantiation vector $\boldsymbol{u}_i$ for the associated instance. Capsules in lower layers try to predict the output of Capsules in higher levels using a trainable weight matrix $\boldsymbol{W}_{ij}$ as follows

$$\widehat{u}_{j|i} = \boldsymbol{W}_{ij}\boldsymbol{u}_i, \tag{1}$$

where $\widehat{u}_{j|i}$ is the predicted output of Capsule j in the next layer by the Capsule i in the lower layer. The association between the prediction $\widehat{u}_{j|i}$ and the actual output of Capsule j, denoted by $\mathbf{v}_j$, is determined by taking the inner product of $\widehat{u}_{j|i}$ and $\boldsymbol{v}_j$. The higher the inner product, the more contribution of the lower level capsules to the higher level one. The contribution of Capsule i

to the output of the Capsule j in the next layer is determined by a coupling coefficient c_{ij}, trained over a course of few iterations known as the "Routing by Agreement" given by

$$a_{ij} = \boldsymbol{v}_j \cdot \widehat{\boldsymbol{u}}_{j|i}, \tag{2}$$

$$b_{ij} = b_{ij} + a_{ij}, \tag{3}$$

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}, \tag{4}$$

$$s_j = \sum_i c_{ij}\widehat{u}_{j|i}, \tag{5}$$

and

$$\boldsymbol{v}_j = \frac{\|\boldsymbol{s}_j\|^2}{1 + \|\boldsymbol{s}_j\|^2} \frac{\boldsymbol{s}_j}{\|\boldsymbol{s}_j\|}, \tag{6}$$

where a_{ij} is referred to as the agreement coefficient between the prediction and actual output, and b_{ij} denotes the log prior of the coupling coefficient c_{ij}. Vector $\boldsymbol{s}_j$ denotes the Capsule output before applying the squashing function. As the length of output vectors represents probabilities, the ultimate output of Capsule j ($\boldsymbol{v}_j$) is obtained by squashing $\boldsymbol{s}_j$ between 0 and 1 using the squashing function defined in **Eq. 6**. In order to update weight matrix $\boldsymbol{W}_{ij}$ through a backward training process, the loss function is calculated for each Capsule k as follows

$$l_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k)\max(0, \|v_k\| - m^-)^2, \tag{7}$$

where T_k is 1 when the class k is present and 0 otherwise. m^+, m^-, and λ are hyper parameters of the model and are originally set to 0.9, 0.1, and 0.5, respectively. The overall loss is the summation of all losses calculated for all Capsules.

3.2 Proposed COVID-FACT

The overall architecture of the COVID-FACT is illustrated in **Figure 2**, which consists of a lung segmentation model at the beginning followed by two Capsule Network-based models and an average voting mechanism coupled with a thresholding approach to generate patient-level classification results. The three components of the COVID-FACT are as follows:

- Lung Segmentation: The input of the COVID-FACT is the segmented lung regions identified by a U-net based segmentation model (Hofmanninger et al., 2020), referred to as the "U-net (R231CovidWeb)", which has been initially trained on a large and diverse dataset including multiple pulmonary diseases, and fine-tuned on a small dataset of COVID-19 images. The Input of the U-net (R231CovidWeb) model is a single slice with the size of 512×512. The output is the lung tissues, which will be normalized between 0 and 1 to
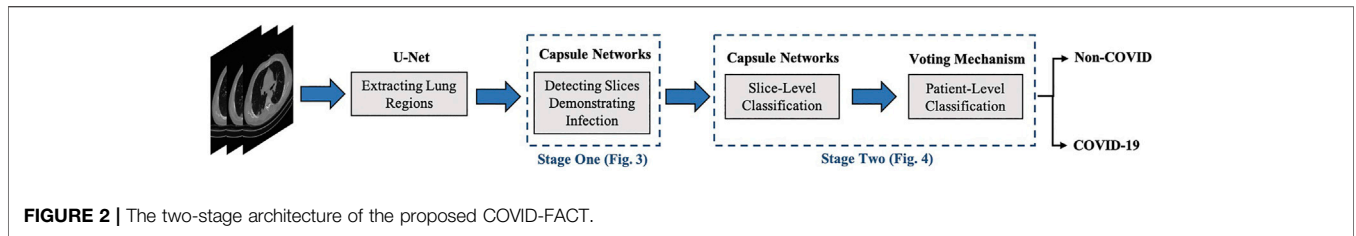
FIGURE 2 | The two-stage architecture of the proposed COVID-FACT.

generalize the features and help the model to perform more effectively. Following the literature (Hu et al., 2020; Zhang et al., 2020), we down-sampled the output from 512×512 to 256×256 size to reduce the complexity and memory allocation without losing any significant information. Slices with no detected lung regions are removed and the remaining are fed into the first stage of the COVID-FACT model.

- COVID-FACT's Stage One: The first stage of the COVID-FACT, shown in **Figure 3** is responsible to identify slices demonstrating infection (by COVID-19 or other types of pneumonia). Using this stage, we discard slices without infection and focus only on the ones with infection. Intuitively speaking, this process is similar in nature to the way that radiologists analyze a CT scan. When radiologists review a CT scan containing numerous consecutive cross-sectional slices of the body, they identify the slices with an abnormality in the first step, and analyze the abnormal ones to diagnose the disease in the next step. Existing CT-based deep learning processing methods either use all slices as a 3D input to a classifier, or classify individual slices and transform slice-level predictions to the patient-level ones using a threshold on the entire slices (Rahimzadeh et al., 2021). Determining a threshold on the number or percentage of slices demonstrating infection over the entire slices is not precise, as most pulmonary infections have different stages with involvement of different lung regions (Yu et al., 2020). Furthermore, a CT

scan may contain different number of slices depending on the acquisition setting, which makes it impossible to find such a threshold. In most methods passing all slices as a 3D input to the model, the input size is fixed and the model is trained to assign higher scores to slices demonstrating infection. However, the performance of such models will be reduced when testing on a dataset other than the dataset on which they are originally trained (Zhang et al., 2020).

The model used in stage one of the proposed COVID-FACT is adapted from the COVID-CAPS model presented in our previous work (Afshar et al., 2020a), which was developed to identify COVID cases from chest radiographs. The first stage consists of four convolutional layers and three capsule layers. The first and second layers are convolutional ones followed by a batch-normalization. Similarly, the third and fourth layers are convolutional ones followed by a max-pooling layer. The fourth layer, referred to as the primary Capsule layer, is reshaped to form the desired primary capsules. Afterwards, three capsule layers perform sequential routing processes. Finally, the last Capsule layer represents two classes of infected and non-infected slices. The input of stage one is set of CT slices corresponding to a patient, and the output is slices of the volumetric CT scan demonstrating infection. The output of stage one may vary in size for each patient due to different areas of lung involvement and phase of infection.
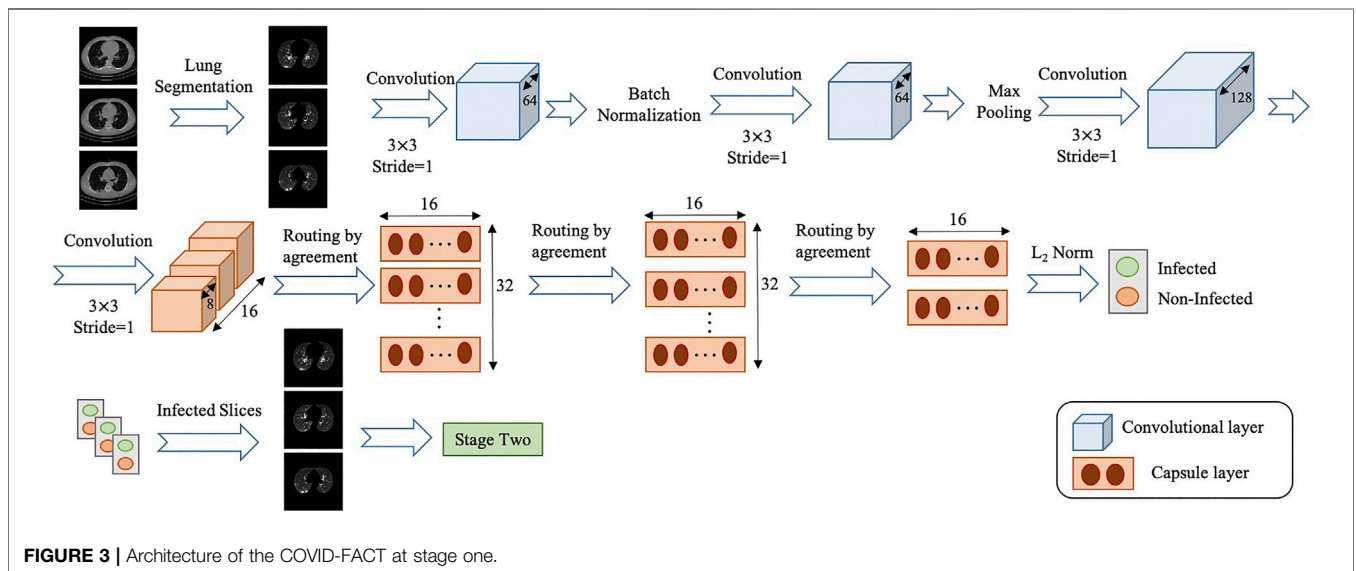

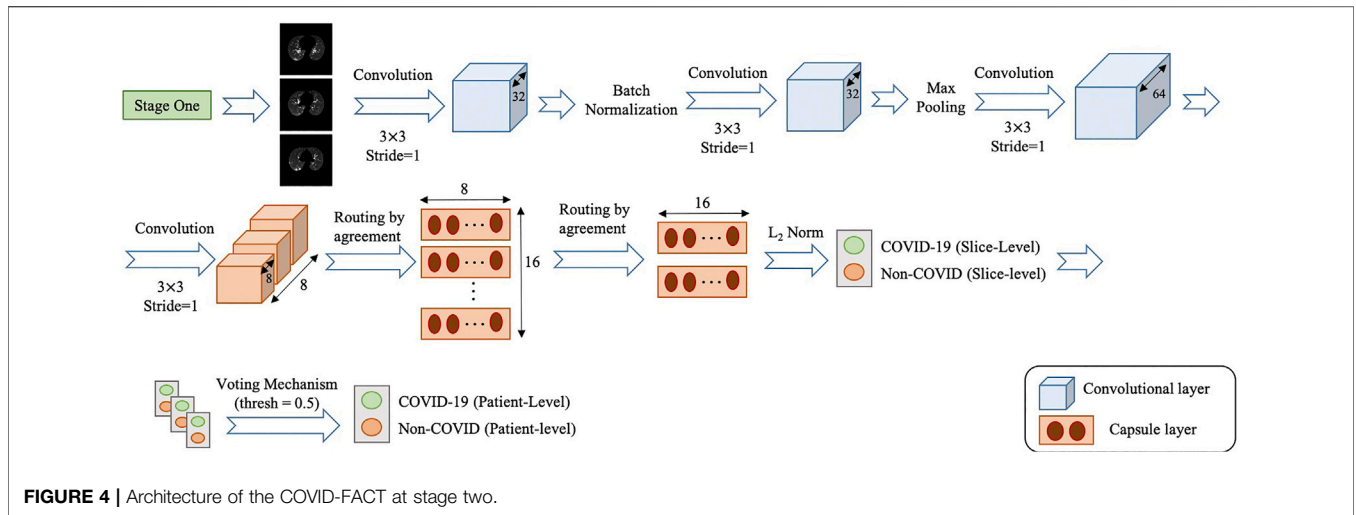FIGURE 3 | Architecture of the COVID-FACT at stage one.

FIGURE 4 | Architecture of the COVID-FACT at stage two.

In order to cope with our imbalanced training dataset, we modified the loss function, so that a higher penalty rate is given to the false positive (infected slices) cases. The loss function is modified as follows

$$loss = \frac{N^+}{N^+ + N^-} \times loss^- + \frac{N^-}{N^+ + N^-} \times loss^+, \quad (8)$$

where N^+ is the number of positive samples, N^- is the number of negative samples, $loss^+$ denotes the loss associated with positive samples, and $loss^-$ denotes the loss associated with negative samples.

• COVID-FACT's Stage Two: As mentioned earlier, we need to apply classification methods on a subset of slices demonstrating infection rather than on the entire slices in a CT scan. It is worth noting that, lung segmentation (i.e., extracting lung tissues) is performed in one of the variants of the COVID-FACT as a pre-processing step. The first stage of the COVID-FACT, on the other hand, is tasked with this specific issue of extracting slices demonstrating infections.

The second stage of the COVID-FACT takes candidate slices of a patient detected in stage one as the input, and classifies them into one of COVID-19 or non-COVID (including normal and pneumonia) classes, i.e., we consider a binary classification problem. Stage two is a stack of four convolutional and two capsule layers shown in **Figure 4**. The output of the last capsule indicates classification probabilities in the slice-level. An average voting function is applied to the classification probabilities, in order to aggregate slice-level values and find the patient-level predictions as follows

$$P(p_k \in c) = \frac{1}{L_k} \sum_{i=1}^{L_k} P(s_i^k \in c), \quad (9)$$

where $P(p_k \in c)$ refers to the probability that patient k belongs to the target class c (e.g., COVID-19), L_k is the total number of slices detected in stage one for patient k, and $P(s_i^k \in c)$ refers to the probability that the ith slice detected for patient k belongs to the

target class c. It is worth noting that while, initially, the COVID-FACT performs slice-level classification in its second stage, the output is patient-level classification (through its voting mechanism), which is on par with other works that COVID-FACT is compared with. As a final note to our discussion, we would like to add that, corona virus infection is, typically, distributed across the lung volume as such manifests itself in several CT slices. Therefore, having a single slice identified as COVID-19 infection can not necessarily lead to a positive COVID-19 detection.

Similar to stage one, the loss function modification in **Eq. 8** is used in the training phase of Stage two. The default cut-off probability of 0.5 is chosen in Stage two to distinguish COVID-19 and non-COVID cases. However, it is worth mentioning that the main concern in the clinical practice is to have a high sensitivity in identifying COVID-19 positive patients, even if the specificity is not very high. As such, the classification cut-off probability can be modified by physicians using the ROC curve shown in **Figure 5** in order to provide a desired balance between the sensitivity and the specificity (e.g., having a high sensitivity while the specificity is also satisfying). In other words, physicians can decide how much certainty is required to consider a CT scan as a COVID-19 positive case. By choosing a cut-off value higher than 0.5, we can exclude those community acquired pneumonia cases that contain highly overlapped features with COVID-19 cases. On the other hand, by selecting a lower cut-off value, we will allow more cases to be identified as a COVID-19 case.

To further improve the ability of the proposed COVID-FACT model to distinguish COVID-19 and non-COVID cases and attenuate effects of errors in the first stage, we classify all patients with less than 3% of slices demonstrating infection in the entire volume as a non-COVID case. These cases are more likely normal cases without any slices with infection. The few slices with infection identified for these cases might be due to the model error in the first stage, non-infectious abnormalities such as pulmonary fibrosis, or motion artifacts in the original images, which will be covered by this threshold. Based on (Yu et al., 2020), it can be interpreted that 4% lung involvement is the
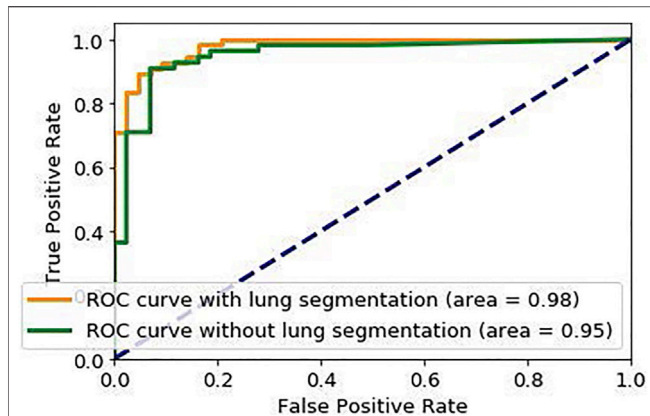
FIGURE 5 | ROC curve of the proposed COVID-FACT.

4 EXPERIMENTAL RESULTS

The proposed COVID-FACT is tested on the in-house dataset described earlier in **Section 2**. The testing set contains 53 COVID-19 and 43 non-COVID cases (including 19 community acquired pneumonia and 24 normal cases). We used the Adam optimizer with the initial learning rate of $1e-4$, batch size of 16, and 100 epochs. The model with the minimum loss value on the validation set was selected to evaluate the performance of the model on the test set. The proposed COVID-FACT method achieved an accuracy of 90.82%, sensitivity of 94.55%, specificity of 86.04%, and AUC of 0.97. The obtained ROC curve is shown in **Figure 5**. The training and validation loss curves are also illustrated in **Figure 6**.

In a second experiment, we trained our model using the complete CT images without segmenting the lung regions. The obtained model reached an accuracy of 90.82%, sensitivity of 92.72%, specificity of 88.37%, and AUC of 0.95. The corresponding ROC curve is shown in **Figure 5**. This experiment indicates that segmenting lung regions in the first step will increase the sensitivity and AUC from 92.72% and 0.95 to 90.82% and 0.98 respectively, while slightly decreases the specificity from 88.37% to 86.04%. Although the numerical results show a slight improvement achieved by segmenting the lung regions, further investigating the sources of errors demonstrates the superiority of using segmented lung regions over the original CT scans. In the COVID-FACT model using lung segmented regions, none of COVID-19 and community acquired pneumonia cases have been mis-classified as a normal case by the 3% thresholding after the first stage, and 95.84% (23/24) of normal cases have been identified correctly using this threshold, while for the model without the lung segmentation, there is one mis-classification of a COVID-19 case by the 3% thresholding, and 91.66% (22/24) of normal cases were identified correctly using this threshold.

Furthermore, we compared performance of the Capsule Network-based framework of COVID-FACT with a CNN-based alternative to demonstrate the effectiveness of Capsule Networks and their superiority over CNN in terms of number of trainable parameters and accuracy. In other words, the CNN-based alternative model has the same front-end (convolutional layers) as that of COVID-FACT in both stages. However, the Capsule layers are replaced by fully connected layers including 128 neurons for intermediate layers and two neurons for the last layer at each stage. The last fully connected layer in each stage is followed by a sigmoid activation function and the remaining modifications and hyper-parameters are kept the same as used in COVID-FACT. The CNN-based COVID-FACT achieved an accuracy of 71.43%, sensitivity of 81.82%, and specificity of 58.14%. The COVID-FACT performance, and number of trainable parameters for examined models are presented in **Table 2**. It is worth noting that in designing the CNN-based COVID-FACT described above, the complexity and structure have been kept similar to its capsule-based version. The goal is to evaluate and illustrate potential advantages of capsule network design over its CNN-based counterpart. Alternative models using CNN architecture and fully connected layers such as the DenseNet model (Yang et al., 2020), however, consist of several convolutional layers and a high degree of complexity,
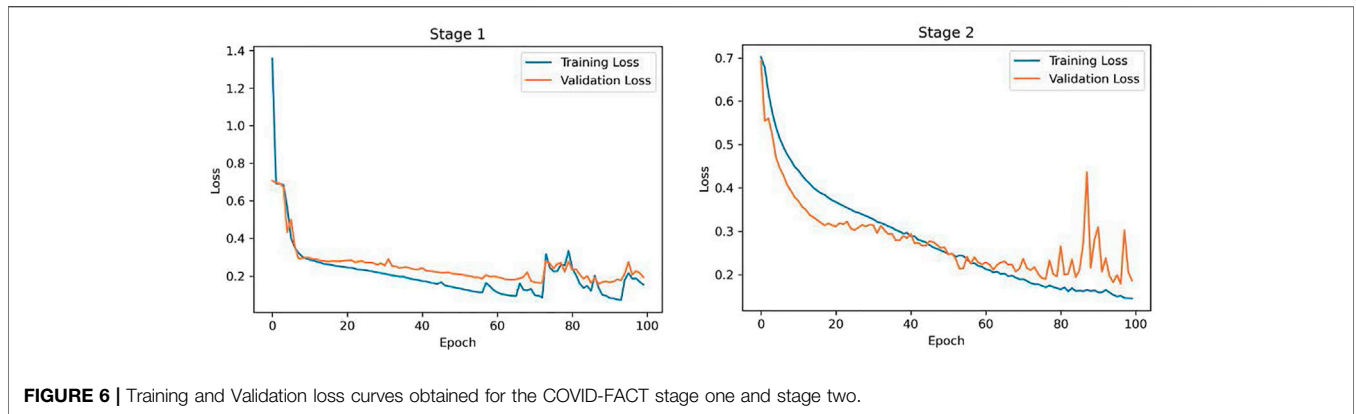
minimum percentage for COVID-19 positive cases. In addition, the minimum percentage of slices demonstrating infection detected by the radiologist in our dataset is 7%, and therefore 3% would be a safe threshold to prevent mis-classifying infected cases as normal.

As a final note, it is worth mentioning that the role of Stage 1 is critical to achieving a fully automated framework, which does not require any input from the radiologists, especially when an early and fast diagnosis is desired. However, the COVID-FACT framework is completely flexible and Stage 1 can be skipped if the slices demonstrating infections have already been identified by the radiologists, meaning that the normal cases are already identified in this case and Stage 2 merely separates COVID-19 and CAP cases.

- Grad-CAM: Using the Grad-CAM approach, we can visually verify the relation between the model's prediction and the features extracted by the intermediate convolutional layers, which ultimately leads to a higher level of interpretability of the model. Grad-CAM's outcome is a weighted average of the feature maps of a convolutional layer, followed by a Rectified Linear Unit (ReLU) activation function, i.e.,

$$L_{Grad-CAM}^{c} = ReLU\left(\sum_{k} \alpha_{k}^{c} A^{k}\right), \qquad (10)$$

where $L_{Grad-CAM}^{c}$ refers to the Grad-CAM's output for the target class c; α_{k}^{c} is the importance weight for the feature map k and the target class c, and; A^{k} refers to the feature map k of a convolutional layer. The weights α_{k}^{c} are obtained based on the gradients of the probability score of the target class with respect to an intermediate convolutional layer followed by a global average pooling function as follows

$$\alpha_{k}^{c} = \frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^{c}}{\partial A_{ij}^{k}}, \qquad (11)$$

where y^{c} is the prediction value (probability) for target class c, and Z refers to the total number of feature maps in the convolutional layer.

FIGURE 6 | Training and Validation loss curves obtained for the COVID-FACT stage one and stage two.

TABLE 2 | Results obtained from COVID-FACT and the alternative CNN-based model.

Method	Accuracy	Sensitivity	Specificity	AUC	Trainable parameters
COVID-FACT with lung segmentation	90.82	94.55	86.04%	0.98	406,880
COVID-FACT without lung segmentation	90.82	92.72%	88.37	0.95	406,880
CNN-based COVID-FACT	71.43%	81.82%	58.14%	0.67	365,806,660

as such it is expected from such complex models to outperform the CNN-based COVID-FACT.

As mentioned earlier, the ROC curve provides physicians with a precious tool to modify the sensitivity/specificity balance based on their preference by changing the classification cut-off probability. To elaborate this point, we changed the default cut-off probability from 0.5 to 0.75 and reached an accuracy of 91.83%, a sensitivity of 90.91%, and a specificity of 93.02%. Further increasing the cut-off probability to 0.8 results in the same accuracy of 91.83%, a lower sensitivity of 89.01%, and a higher specificity of 95.34%. On the other hand, decreasing the cut-off probability from 0.5 to 0.35 will increase the accuracy and the sensitivity to 91.83% and 98.18% respectively, while slightly decreases the specificity to 83.72%. The performance of the COVID-FACT for different values of cut-off probability are presented in **Table 3**.

While performance of the COVID-FACT is evaluated by its final decision made in the second stage, the first stage plays a crucial role in the overall accuracy of the model. As such, performance of the COVID-FACT in the first stage is also reported in **Table 4**. As shown in **Table 4**, ~91% of the slices demonstrating infection are identified correctly by the COVID-FACT at the first stage, while there are some mis-classified slices that will be passed to the next stage as the infectious slices. It is also evident that the CNN-based model cannot properly identify infectious slices, which in turn led to the low performance of the second stage. It is worth mentioning that stage one is only responsible to detect candidate slices, while stage two classifies the slices into COVID and non-COVID categories. The second stage is followed by an aggregation mechanism, which takes all the slices of a patient into account and consequently decreases the impact of mis-classified slices at the first stage. We have also investigated the performance of the model when the commonly used focal loss function (Lin et al., 2017) is utilized to train the model. The

TABLE 3 | Performance of COVID-FACT for different values of cut-off probability.

Cut-off probability	0.35	0.5	0.6	0.7	0.75	0.8
Accuracy (%)	91.83	90.82	91.83	90.82	91.83	91.83
Sensitivity (%)	98.18	94.55	92.73	90.91	90.91	89.01
Specificity (%)	83.72	86.04	90.70	90.70	93.02	95.34

COVID-FACT framework trained by the focal loss function ($\gamma = 2$, $\alpha = 0.25$) achieved the same patient-level performance compared to our proposed model while the performance of the first stage was lower with the accuracy of 92.79%, sensitivity of 87.69%, and the specificity of 97.03%. The lower sensitivity in the first stage shows benefits of using the modified loss function as the role of the first stage in the pipeline is to detect slices with the evidence of infection to be analyzed in the second stage. As such, the model, which is trained using our modified loss function has been selected as the final model due to its higher accuracy and sensitivity in detecting slices demonstrating infection.

As another experiment, performance of stage two is evaluated without applying the first stage to provide a better comparison of the models used in the second stage. More specifically, the stage two model is trained based on the infectious slices identified by the radiologist and evaluated on the labeled test set including 17 COVID-19 and 8 CAP cases. The numbers of correctly predicted cases in this experiment are presented in **Table 5**. The experimental results obtained by the COVID-FACT framework using the lung segmentation achieved quite a similar performance compared to the case in which the model was trained based on the outputs of stage one. This result further demonstrates that the Capsule Network and the aggregation mechanism used in stage two can cope with errors in the previous stage and achieve desirable performance. It is worth

TABLE 4 | The performance of stage one in diagnosis of slices demonstrating infection.

Method (stage one)	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
COVID-FACT with lung segmentation	93.14	90.75	94.01%	0.96
COVID-FACT without lung segmentation	92.78%	87.59%	94.36	0.96
CNN-based COVID-FACT	79.74%	33.00%	91.28%	0.64

mentioning that this experiment was performed using only the labeled dataset, which consequently provided a smaller dataset to train the model.

The localization maps generated by the Grad-CAM method are illustrated in **Figure 7** for the second and fourth convolutional layers in the first stage of the COVID-FACT. It is evident in **Figure 7** that the COVID-FACT model is looking at the right infectious areas of the lung to make the final decision. Due to the inherent structure of the Capsule layers, which represent image instances separately, their outputs cannot be superimposed over the input image. Consequently, in this study, the Grad-CAM localization maps are presented only for convolutional layers.

4.1 K-Fold Cross-Validation

We have evaluated the performance of the COVID-FACT and its variants based on the 5-fold cross-validation (Stone, 1974) to provide more objective assessments. In this experiment, the COVID-FACT achieves the accuracy of $87.61 \pm 2.00\%$, the sensitivity of $88.30 \pm 3.22\%$, and specificity of $86.75 \pm 1.91\%$. Using the same 5-fold cross-validation technique, the COVID-FACT without using the segmented lung areas achieves the accuracy of $87.31 \pm 3.37\%$, sensitivity of $88.32 \pm 5.00\%$, and specificity of $86.03 \pm 3.18\%$. Finally, the CNN-based COVID-FACT achieves the accuracy of $64.49 \pm 1.61\%$, sensitivity of $79.58 \pm 6.61\%$, and specificity of $46.67 \pm 8.48\%$. The results confirm the superiority of the COVID-FACT using the segmented lung areas over its variants as was demonstrated in the previous experiments based on randomly selected test dataset. Moreover, similar to the previous experiments, modifying the cut-off probability is beneficial in the cross-validation case to adjust the capability of the model to focus on COVID or non-COVID cases depending on radiologists' priorities. More specifically, in the aforementioned 5-fold cross-validation, decreasing the cut-off probability to 0.35 increases the sensitivity to $92.97 \pm 2.96\%$ while the overall accuracy remains the same. Increasing the cut-off probability to 0.6, on the other hand, increases the specificity to $91.16 \pm 3.73\%$ and provides the same accuracy similar to the previous case.

5 DISCUSSION

In this study, we proposed a fully automated Capsule Network-based framework, referred to as the COVID-FACT, to diagnose COVID-19 disease based on chest CT scans. The proposed framework consists of two stages, each of which containing several layers of convolutional and Capsule layers. COVID-FACT is augmented with a thresholding method to classify CT scans with zero or very few slices demonstrating infection as non-COVID patients, and an average voting mechanism coupled with a thresholding approach is embedded to extend slice-level classification into patient-level ones. Experimental results indicate that the COVID-FACT achieves a satisfactory performance, in particular a high sensitivity with far less trainable parameters, supervision requirements, and annotations compared to its counterparts.

We further investigated mis-classified cases to determine the limitations and possible improvements. **Table 6** shows the number of the mis-classified cases for each type of the input disease (COVID-19, CAP, normal) obtained at stage two, as well as the number of normal cases that were not identified correctly by the 3% threshold after the first stage. The low rate of errors obtained by the 3% threshold in the first stage demonstrates the capability of COVID-FACT to identify normal cases in the first stage, which is very helpful for physicians and radiologists to exclude normal cases at the very beginning of their study.

As in the case of highly contagious diseases such as COVID-19, the False-Negative-Rate (FNR) is of utmost importance, we have further analyzed such errors to explore the possible sources of the mis-classification. As shown in **Table 6** there are 3/55 COVID-19 cases that are mis-classified by the COVID-FACT framework. We found that one mis-classified COVID-19 case contains unifocal infection manifestation with consolidation predominance rather than GGO, which are more common in CAP cases rather than COVID-19 ones. One other case of error was identified as an incomplete CT scan with missing slices, which has consequently made the correct identification difficult for the framework. In addition, we have reviewed the aforementioned errors in the case of image quality and lung segmentation as other potential causes of the error. The assessment results showed that the image qualities are adequate and the segmentation model performed well without removing or cropping the infection manifestations. Therefore, some errors are likely to be caused by the similarities between the infection patterns in CAP and COVID-19 cases. It is worth noting that decreasing the cut-off probability from 0.5 to 0.35, as shown in **Table 3**, will result in the correct classification of the two false-negative cases, which contain similar characteristics to other
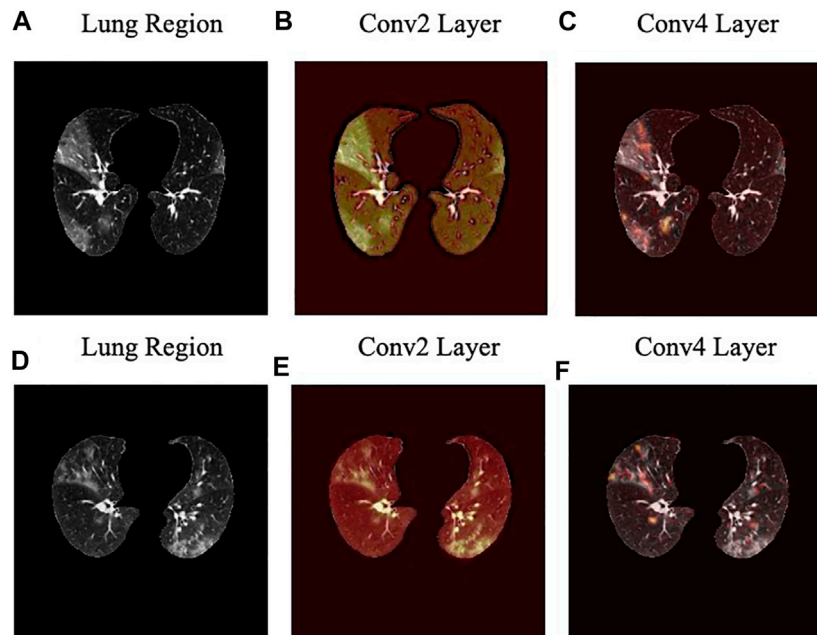
TABLE 5 | Correctly predicted cases using only stage two without applying the first stage.

Model	COVID-19	CAP
Stage 2 with lung segmentation	94.1% (16/17)	87.5% (7/8)
Stage 2 without lung segmentation	88.2% (15/17)	62.5% (5/8)
Stage 2 CNN-based	82.4% (14/17)	25% (2/8)

FIGURE 7 | Localization heatmaps for the second and forth convolutional layers of the first stage obtained by the Grad-CAM for two slices.

infections. This can be considered as a remedy, when FNR is of the main concern.

We also identified that errors in stage one are mainly caused by non-infectious abnormalities such as pulmonary fibrosis and artifacts. In this regard, we have further explored slices with the evidence of artifact where no infection manifestation presents. In some cases, the motion artifact or the artifacts caused by the presence of metallic components inside the body have generated some components in the image that were mis-classified as infectious slices. **Figure 8** illustrates 4 samples of such slices in which images A) and B) belong to a mis-classified normal case while images C) and D) are related to two CAP cases, where classified correctly in the second stage. It is worth mentioning that, the number of such slices is negligible especially when they appear in cases that have multiple infectious slices (caused by CAP or COVID-19). In those cases, the influence of such slices with the evidence of artifact will be diminished by the second stage and the following aggregation mechanism. Motion artifact reduction algorithms can be investigated as a future work to cope with undesired impacts of the artifacts on the final result. It is worth mentioning that during the labeling process accomplished by the radiologist to detect slices demonstrating infection, we

noticed that in some cases the abnormalities are barely visible with the standard visualization setting (window center and window width). Those abnormalities have been detected by changing the image contrast (by adjusting the window center and width) manually by the radiologist. This limitation will arise the need to research on the optimal contrast and window level use in future studies. As another limitation, we can point to the retrospective study used in the data collection part of this research. Although the provided dataset is acquired with the utmost caution and inspection, a retrospective data collection might add inappropriate cases to the study at hand. The potential improvement to address this limitation could be the collaboration of more radiologists in analyzing and labeling the data to assess if the interobserver agreement is satisfying or not.

As a side note to our discussion, we would like to mention that while both CT and CR can decrease the false negative rate at the admission and discharge times, the CR is less sensitive, and less specific compared to CT. Some studies such as Reference (Wong et al., 2020) report that CR often shows no lung infection in COVID-19 patients at early stages resulting in a low sensitivity of 69% for diagnosis of COVID-19. Therefore, chest CT has a key role for diagnosis of COVID-19 in the early stages of the infection and also to set up a prognosis. Furthermore, a single CR image fails to incorporate details of infections in the lung and cannot provide a comprehensive view for the lung infection diagnosis. Unlike CR images, CT scans generate cross-sectional images (slices) and create a 3D representation of the body (i.e., each patient is associated with several 2D slices). As a result, CT images can show detailed structure of the lung and infected areas. Consequently, CT is considered as the preferred modality for grading and evaluation of imaging manifestations for COVID-19 diagnosis. It is worth adding that as CT scans are 3D images, as

TABLE 6 | The number of the mis-classified cases for each type of the input disease and the number of cases that were not identified correctly by. the 3% threshold.

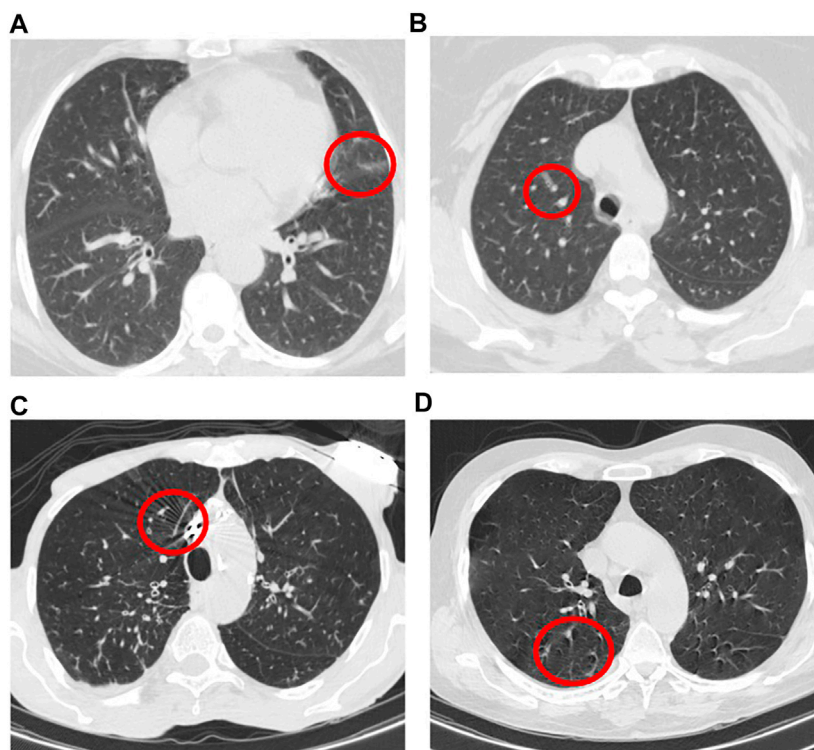| Input	Errors (thresholding)	Errors (stage two)
COVID-19	0/55	3/55
CAP	0/19	5/19
Normal	1/24	1/24

FIGURE 8 | Example of slices with the evidence of artifact where no infection manifestation presents.

opposed to 2D chest radiographs, they are more difficult to be processed using ML and DL techniques, as the currently available resources cannot efficiently process the whole volume at once. As such, slice-level and thresholding techniques are utilized to cope with such limitations, leading to a reduced performance compared to the models working with CR (e.g., the COVID-CAPS (Afshar et al., 2020d), which deals with 2D chest radiographs). The focus of our ongoing research is to further enhance performance of CT-based COVID-19 diagnosis models to fill the gap between the radiologists' performance and that of volumetric-based DL techniques.

As a final note, unlike our previous work on the chest radiographs (Afshar et al., 2020a), where we used a more imbalanced public dataset, the dataset used in this study contains a substantial number of COVID-19 confirmed cases making our results more reliable. Upon receiving more data from medical centers and collaborators, we will continue to further modify and validate the COVID-FACT by incorporating new datasets.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: Figshare, https://figshare.com/s/c20215f3d42c98f09ad0.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Concordia University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SH, PA, and NE implemented the deep learning models. SH and PA drafted the manuscript jointly with AM and FN. FB, KS, and MR supervised the clinical study and data collection. FB and MR annotated the CT images. AO contributed to interpretation analysis and edited the manuscript. SA and KP edited the manuscript, AM, FN, and MR directed and supervised the study. All authors reviewed the manuscript.

FUNDING

REFERENCES

Afshar, P., Heidarian, S., Enshaei, N., Naderkhani, F., Rafiee, M. J., Oikonomou, A., et al. (2021). *COVID-CT-MD: COVID-19 Computed Tomography (CT) Scan Dataset Applicable in Machine Learning and Deep Learning*. Nature Scientific Data. In Press.

Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K. N., and Mohammadi, A. (2020a). COVID-CAPS: A Capsule Network-Based Framework for Identification of COVID-19 Cases from X-Ray Images. *Pattern Recognition Lett.* 138, 638–643. doi:10.1016/j.patrec.2020.09.010

Afshar, P., Mohammadi, A., and Plataniotis, K. N. (2020c). "BayesCap: A Bayesian Approach to Brain Tumor Classification Using Capsule Networks," in Submitted to IEEE International Conference on Image Processing (ICIP) (IEEE), 2024–2028. doi:10.1109/LSP.2020.3034858

Afshar, P., Mohammadi, A., and Plataniotis, K. N. (2018). "Brain Tumor Type Classification via Capsule Networks," in 2018 25th IEEE International Conference on Image Processing (ICIP) (IEEE), 3129–3133. doi:10.1109/ICIP.2018.8451379

Afshar, P., Oikonomou, A., Naderkhani, F., Tyrrell, P. N., Plataniotis, K. N., Farahani, K., et al. (2020b). 3D-MCN: A 3D Multi-Scale Capsule Network for Lung Nodule Malignancy Prediction. *Sci. Rep.* 10, 7948. doi:10.1038/s41598-020-64824-5

Afshar, P., Plataniotis, K. N., and Mohammadi, A. (2020d). "BoostCaps: A Boosted Capsule Network for Brain Tumor Classification," in *Accepted in IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE), 20–24. doi:10.1109/EMBC44109.2020.9175922

Afshar, P., Plataniotis, K. N., and Mohammadi, A. (2019a). "Capsule Networks for Brain Tumor Classification Based on MRI Images and Coarse Tumor Boundaries," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE), 1368–1372. doi:10.1109/ICASSP.2019.8683759

Afshar, P., Plataniotis, K. N., and Mohammadi, A. (2019b). "Capsule Networks' Interpretability for Brain Tumor Classification via Radiomics Analyses," in 2019 IEEE International Conference on Image Processing (ICIP) (IEEE), 3816–3820. doi:10.1109/ICIP.2019.8803615

Bai, H. X., Hsieh, B., Xiong, Z., Halsey, K., Choi, J. W., Tran, T. M. L., et al. (2020). Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT. *Radiol.* 296, E46–E54. doi:10.1148/radiol.2020200823

Chen, L. C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation *ArXiv: 1706.05587*.

Chung, M., Bernheim, A., Mei, X., Zhang, N., Huang, M., Zeng, X., et al. (2020). CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). *Radiol.* 295, 202–207. doi:10.1148/radiol.2020200230

Deng, J., Dong, W., Socher, R., Li, L. J., Kai Li, Kai., and Li Fei-Fei, Li. (2009). "ImageNet: A Large-Scale Hierarchical Image Database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), 248–255. doi:10.1109/CVPR.2009.5206848

DICOM Standards Committee, Working Group 18 Clinical Trials (2011). *Supplement 142: Clinical Trial De-identification Profiles*. Rosslyn, VI, United States: DICOM Standard, 1–44.

Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., et al. (2020). Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiol.* 296, E115–E117. doi:10.1148/radiol.2020200432

Hara, K., Kataoka, H., and Satoh, Y. (2017). "Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition," in Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, Venice, Italy (ICCVW) (ICCVW), 3154–3160. doi:10.1109/ICCVW.2017.373

Hinton, G., Sabour, S., and Frosst, N. (2018). "Matrix Capsules with EM Routing," in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 1–29.

Hofmanninger, J., Prayer, F., Pan, J., Rohrich, S., Prosch, H., and Langs, G. (2020). *Automatic Lung Segmentation in Routine Imaging Is a Data Diversity Problem, Not a Methodology Problem*, 1–10.

Hu, S., Gao, Y., Niu, Z., Jiang, Y., Li, L., Xiao, X., et al. (2020). Weakly Supervised Deep Learning for COVID-19 Infection Detection and Classification from CT Images. *IEEE Access* 8, 118869–118883. doi:10.1109/ACCESS.2020.3005510

Inui, S., Fujikawa, A., Jitsu, M., Kunishima, N., Watanabe, S., Suzuki, Y., et al. (2020). Chest CT Findings in Cases from the Cruise Ship Diamond Princess with Coronavirus Disease (COVID-19). *Radiol. Cardiothorac. Imaging* 2, e200110. doi:10.1148/ryct.2020200110

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 60, 84–90. doi:10.1145/3065386

Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., et al. (2020). Using Artificial Intelligence to Detect COVID-19 and Community-Acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiol.* 296, E65–E71. doi:10.1148/radiol.2020200905

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). *Focal Loss for Dense Object Detection*. doi:10.1109/iccv.2017.324

Mahmud, T., Rahman, M. A., and Fattah, S. A. (2020). CovXNet: A Multi-Dilation Convolutional Neural Network for Automatic COVID-19 and Other Pneumonia Detection from Chest X-Ray Images with Transferable Multi-Receptive Feature Optimization. *Comput. Biol. Med.* 122, 103869. doi:10.1016/j.compbiomed.2020.103869

Ng, M.-Y., Lee, E. Y. P., Yang, J., Yang, F., Li, X., Wang, H., et al. (2020). Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review. *Radiol. Cardiothorac. Imaging* 2, e200034. doi:10.1148/ryct.2020200034

Rahimzadeh, M., Attar, A., and Sakhaei, S. M. (2021). A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset *Biomed. Signal Process. Control.* 68, 102588. doi:10.1016/j.bspc.2021.102588

Raman, S. P., Mahesh, M., Blasko, R. V., and Fishman, E. K. (2013). CT Scan Parameters and Radiation Dose: Practical Advice for Radiologists. *J. Am. Coll. Radiol.* 10, 840–846. doi:10.1016/j.jacr.2013.05.032

Ronneberger, O., Fischer, P., and Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*, 234–241. doi:10.1007/978-3-319-24574-4_28

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in 2017 IEEE International Conference on Computer Vision (ICCV) (IEEE), 618–626. doi:10.1109/ICCV.2017.74

Sethy, P. K., Behera, S. K., Ratha, P. K., and Biswas, P. (2020). Detection of Coronavirus Disease (COVID-19) Based on Deep Features. *Int. J. Math. Eng. Manag. Sci.* 5, 643–651. doi:10.20944/preprints202003.0300.v1

Shi, H., Han, X., Jiang, N., Cao, Y., Alwalid, O., Gu, J., et al. (2020). Radiological Findings from 81 Patients with COVID-19 Pneumonia in Wuhan, China: a Descriptive Study. *Lancet Infect. Dis.* 20, 425–434. doi:10.1016/S1473-3099(20)30086-4

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc. Ser. B (Methodological)* 36, 111–133. doi:10.1111/j.2517-6161.1974.tb00994.x

Wang, L., and Wong, A. (2020). *COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images*.

Wong, H. Y. F., Lam, H. Y. S., Fong, A. H.-T., Leung, S. T., Chin, T. W.-Y., Lo, C. S. Y., et al. (2020). Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19. *Radiol.* 296, E72–E78. doi:10.1148/radiol.2020201160

Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional Neural Networks: an Overview and Application in Radiology. *Insights Imaging* 9, 611–629. doi:10.1007/s13244-018-0639-9

Yang, S., Jiang, L., Cao, Z., Wang, L., Cao, J., Feng, R., et al. (2020). Deep Learning for Detecting Corona Virus Disease 2019 (COVID-19) on High-Resolution Computed Tomography: a Pilot Study. *Ann. Transl Med.* 8, 450. doi:10.21037/atm.2020.03.132

Yu, N., Shen, C., Yu, Y., Dang, M., Cai, S., and Guo, Y. (2020). Lung Involvement in Patients with Coronavirus Disease-19 (COVID-19): a Retrospective Study Based on Quantitative CT Findings. *Chin. J. Acad. Radiol.* 3, 102–107. doi:10.1007/s42058-020-00034-2

Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., et al. (2020). Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* 181, 1423–1433. doi:10.1016/j.cell.2020.04.045