SKINN: Semantic Knowledge Inference Neural Network for COVID-19 Classification

Chongyang Gao, Pengze Liu

***Abstract*—The COVID-19 caused by a novel coronavirus has infected more than 50 million people globally in 2020. Due to the long incubation period for COVID-19, a fast and accurate COVID-19 detection method becomes an urgent demand. The real-time polymerase chain reaction (RT-PCR) is widely used as the standard COVID-19 detection approach. However, the availability of vast-scale RT-PCR is limited in certain countries and regions. Computed Tomography (CT) scan has been widely used for pneumonia detection. Recent advances in deep learning-based computer vision provide powerful tools to perform feature extraction and disease detection. Recent research works in COVID-19 classification demonstrated the effectiveness of Convolutional Neural Networks (CNNs) based on CT scans. However, due to the limited amount of annotated data and the oversimplified problem formulation, these methods failed to outline the lesion areas while making predictions. The lesion areas are informative to both the radiologists and the classification model. In this paper, we propose a novel Semantic-Knowledge Inference Neural Network (SKINN) that fully exploits the semantic information of the lesion for more accurate COVID-19 classification. We introduce a two-step transfer learning method to train our SKINN. By combining the predicted lesion masks with the CT scans, our SKINN outperforms prior state-of-the-art methods by accuracy, F-1 score, and AUC score.**

***Index Terms*—Data analysis; Public healthcare; Image classification;**

I. Introduction

THE outbreak of COVID-19 in September 2019 has cost more than one million death all over the world, and the recorded rate is above all infectious diseases before [39]. The COVID-19 has spread rapidly and widely because it spreads by respiratory droplets, and the symptoms of COVID-19 may appear 2-14 days after exposure to the virus [60]. Additionally, compared with viral pneumonia, which is mild most of the time, COVID-19 may result in severe symptoms. Cutting off transmission is proven to be one of the most effective actions in virus control, which requires collective effort in testing and self-isolation. There are increasing demands on available data collection devices and diagnosis technologies. The real-time polymerase chain reaction (RT-PCR) is currently used as a standard method to detect viral nucleic acid [9], [27]. However, many rural regions and countries cannot provide RT-PCR tests, and the RT-PCR test is not the most sensitive testing method [3]. Furthermore, RT-PCR cannot detect the evolved coronavirus because of the change of its DNA sequence [58]. The recent development in Computer Vision and Deep Learning provides new alternatives for assisted medical imaging recognition and diagnosis. In this paper, we propose SKINN, a novel segmentation network as an assistant for CT based COVID-19 diagnosis.

Image techniques, *e.g.*, Chest X-rays and lung Computed Tomography (CT-scans) play an important role in detecting the characters of corona virus [37]. There is a strong correlation between chest CT and RT-PCR testing in coronavirus disease [3], and in some cases, screening chest CT provides higher sensitivity [14]. As a result, it is feasible and effective to screen the radiography to diagnose the coronavirus disease. The key point to detect coronavirus symptoms by radiography is to find the subtle differences between healthy people and patients, including the ground-glass area, consolidation area, and pleural effusion area. It is stressful for radiologists to screen all day with full attention during the pandemic, and the accuracy of screening may be affected. CT-scan processing is demanded to assist from automatic diagnosis applications to ease the burden of the radiologists.

With the blooming of deep learning in recent days, we have developed many powerful techniques to solve medical image processing problems. The most widely used method is the convolutional neural networks (CNNs). CNNs can help diagnose many diseases related to lung [2], [22], [32], [44], [46], [53]. However, most of the previous models cannot be directly projected to the COVID-19 CT scans with high confidence. The visualization or highlighting of the lesion area is also missing in most of the previous CNN networks. The CNN networks for COVID-19 diagnosis are expected to present good visualization of the parts where the radiologists need to pay more attention. However, the previously proposed CNN networks focused more on generating high accuracy of classification result, which is necessary but insufficient to convince the professionals in radiology. Also, previous CNN models can hardly be applied in practice because they are black boxes, which may raise concerns regarding false assertions. Furthermore, most of the previous model only fits a specific dataset and learns the irrelevant biases. Consequently, the classifiers are not robust to the high variation in real life.

This paper proposes a semantic knowledge inference neural network (SKINN) to distinguish the healthy people and the patients infected by COVID-19. Our SKINN is more robust and can provide segmentation of the CT-scans, including ground-glass area, consolidation area, and pleural effusion area simultaneously, which helps the clinic staff find subtle insights to decide more confidently. Specifically, we use the Fully Convolutional Networks (FCN) [30] pretrained on ImageNet dataset [43] to train the segmentation branch of our model. After training on segmentation data, we incorporate the segmentation branch into the classification part of the model. This procedure is called two-step transfer learning. In both the training and inference stage, the SKINN generates the segmentation for the input CT-scans firstly. Then we combine the input CT-scans and the segmentation inference as the input of the classification module of the SKINN. There are two advantages of using segmentation knowledge to guide the COVID-19 classification. Firstly, the training and inference of the model fit the way of human reasoning. When humans try to find the subtle difference between the two images, We humans typically find the possible regions first and detail them. Our SKINN produces segmentation inference first, which highlights the possible region, and then the classification network gets benefits from segmentation knowledge to boost the performance. Secondly, the semantic knowledge inference is improved by training classification tasks because of the back-propagation mechanism. The radiologists can get initial intuition from the inferred segmentation produced by our SKINN.

The rest parts are organized as follows. We introduce the related work in Section II and then illustrate our semantic knowledge inference neural network (SKINN) in Section III. The experiments and human studies are presented in Section IV. In Section V, we discuss the limitations and future work that could improve the existing method. Section VI further introduced the possible extensions of SKINN for multiple datasets. In Section VII we conclude our methods and contributions.

In conclusion, the main contributions of this paper are as follows.

- We proposed a two-step transfer learning method to train our SKINN segmentation branch, which alleviates the problem of the scarce data with segmentation labeled and verified by radiologists.
- We manage to combine the CT-scan image and its segmentation inference knowledge to boost our model in the training stage, which leverages the semantic information thoroughly and improves both the classification and segmentation performance.
- Our SKINN can infer the segmentation for the CT-scans and determine whether COVID-19 infects people.
- The classification results outperform state-of-the-art models. The inferred segmentation of CT-scans is accurate and can help the radiologists to find subtle insights.

II. Related work

In this section, we will discuss three categories of research that are related to our proposed method, including deep learning for medical image processing in general, transfer learning in healthcare, and deep learning specifically for pneumonia diagnosis.

A. Deep Learning for Medical Images Recognition

Medical image recognition has long been a problem because of its high demand for domain knowledge. More efficient expertise training and recognition on a large scale can be achieved by computer vision algorithms. Computer vision techniques are widely applied in assist the gene structures visualization [42], medical image registration [20], the detection of biological structures [48] and so on.

Benefiting from the increasing data, boosting of computational power [26], and more advanced frameworks [1], [23], [38], deep learning has achieved considerable successes in computer vision and image processing. Deep learning-based computer vision achieved super-human performance in image recognition [18]. As a subcategory of deep learning models, Convolutional Neural Networks (CNNs) are suitable for feature extraction from visual input, which can serve as a powerful tool to tackle medical image processing, visualization, and recognition problems. In recent years, More and more researchers manage to use deep learning models to learn the feature representation for medical images [16], [59]. The learned features boost medical image retrieval [40], multimodal medical record analysis [49]. Similar to the previous works, the problem of COVID-19 diagnosis can also be formulated as computer vision problems. In our SKINN, we formulate the problem as a cascade of semantic segmentation and image classification problem and approach the problem with a hybrid of two Convolutional Neural Networks for feature extraction as well as feature decoding.

B. Transfer Learning in Healthcare

Despite the massive successes of deep learning in many computer vision tasks, the training and optimization of deep learning models require a considerable amount of labeled data, which may not necessarily be available in certain scenarios. To avoid massive data annotation, researchers introduced transfer learning that leverage knowledge learned from one domain to another task. In transfer learning, the model is trained in two stages, which are the pretraining stage and the fine-tuning stage. In the pretraining stage, the model is trained on an auxiliary task. The knowledge learned from the auxiliary task, represented by the parameters of the model, is applied to the fine-tuning stage as the initialization. In the fine-tuning stage, the data of interest is used to fine-tune the model adapted from the pretraining stage.

Transfer learning is widely applied to disease predictions and diagnosis. Dawud *et al.* [10] proposed a transfer learning method based on AlexNet-SVM for brain hemorrhage classification, which achieves a reasonably promising result. De *et al.* [11] addressed the glucose forecasting problem for diabetic people with a multi-source adversarial transfer learning framework. Muhammad *et al.* [35] reviews major deep learning models for brain tumor classification (BTC), denoting that the state-of-the-art performance is highly dependent on extensive experiments using transfer learning. Farhadi *et al.* [15] grouped an enormous number of breast cancer datasets for the pretraining stage model training and transferred the knowledge to a specific breast cancer dataset. The transfer learning led to a huge success in the targeted dataset.

Transfer learning is also used for wearable and mobile device data analysis [4], [5], [8]. With transfer learning, the privacy of the user is protected [34]. Martinez *et al.* [33] proposes a two-stage transfer learning protocol that fully leverages knowledge from general tasks and pneumonia imaging tasks for COVID-19 classification, which successfully addressed the problem of lacking positive training samples at the early stage of COVID-19 research.

In our proposed method, we use a two-stage transfer learning that fully exploits the low-level features in CT scans and the semantic information regarding the lesion areas. The knowledge is transferred to the COVID-19 CT dataset for pneumonia classification. We also perform ablative studies to demonstrate the effectiveness of transfer learning for COVID-19 diagnosis.

C. Pneumonia Diagnosis with Deep Learning

Computer vision methods are applied widely in pneumonia diagnosis. X-ray images and CT scans both provide useful visual features that serve as the input to the classifiers. Previous researches show some promising results using neural networks for classifying diseases [55] and getting segmentation [56]. Deep learning becomes the most popular way of detecting diseases [41], [47].

The pandemic is causing more attention to deep learning aided COVID-19 diagnosis since 2020. Among the COVID-oriented methods, CT-scans and X-ray images are the most popular medical images [3], [29] that recent research works tend to rely on. The researchers formulate the COVID-19 detection problem into two categories, including disease classification [7], [28], [57] and lesion segmentation [13], [31]. Narin *et al.* [36] suggests that ResNet-50 [18] has superior performance over other CNN frameworks in COVID-19 image classification. Li *et al.* [28] tries to craft visual features for COVID-19 classification with a hybrid of Neural Networks and Support Vector Machine (SVM). These classification frameworks fit well with the COVID-19 classification dataset, suggesting that the preliminary goal of COVID-19 diagnosis can be solved by deep learning. However, many of the classification networks failed to highlight the lesions, which makes it hard for the radiologist to be informed of the visual information.

To overcome the visualization problem, more researchers began to formulate the COVID-19 diagnosis as a semantic segmentation problem. U-Net++ [62] is proven to be a reliable framework for COVID-19 related segmentation works. Ai *et al.* [3] and Chen *et al.* [7] fully leverage the U-Net++ framework for highlighting suspicious lesions. In addition to single image classification, Butt *et al.* [6] try to exploit spatial information among CT slices with 3D Convolutional Neural Network [54]. Gozes *et al.* [17] also leverage 2D and 3D features for COVID-19 visual feature extraction.

III. Proposed Method

In this section, we introduce the problem definition and a formalized solution firstly. Then, we present our SKINN architecture and detail two sub-networks of SKINN, segmentation network and classification network, separately. Thirdly, we introduce the datasets we used. Finally, we show the two-stage training procedure of our SKINN using transfer learning.

A. Problem and method Definition

Given a set of CT scans, our ultimate goal is to classify whether the patient is tested positive for COVID-19. However, a binary label denoting the diagnostic result is less informative for the healthcare professionals to justify the classification result. To provide semantic information and highlight the lesions, we define an intermediate goal to predict the COVID-19 lesions by pixel-level classification. The intermediate goal here is to infer the segmentation knowledge, *e.g.*, the area of ground-glass in CT-scans. It can be formulated as a semantic segmentation problem, which outputs a binary mask denoting the lesion areas in the CT-scan. Since the lesion segmentation is highly correlated with our ultimate goal of COVID-19 diagnosis, we will leverage both the parameter and the pixel-level mask of the segmentation network to perform image-level classification.

Formally, given a set of CT scans, denoted as $X = \{x_i | i \in 0, 1, ..., n-1\}$, where $x_i \in \mathbb{R}^{w \times h}$ is the i-th CT scan, we build a cascade of two networks, i.e. lesion segmentation network G and COVID-19 classification network F, to perform semantic knowledge information and image classification respectively. To simplify the definition, we assume that the CTs are single-slice. Then we have

$$M_i = G(x_i, \theta_{seg}) \in \mathbb{B}^{w \times h} \tag{1}$$

where M_i is the binary mask with the same size as the input image, denoting the lesion areas based on the pixel-wise classification, and θ_{seg}, denoting the parameters in the segmentation network G, which we will optimize in the training stage. After obtaining the binary mask, we use the concatenation of the mask and the CT slice as the input for the COVID-19 classification network, as shown in Equation 2:

$$s_i = F(x_i \oplus M_i, \theta_{cls}), i \in \{0, 1\} \tag{2}$$

where s_i denotes the network output for class i given a CT scan x_i as input, and θ_{cls} is the parameter of the classification network that we will optimize. We use $\oplus$ to denote matrix concatenation.

Since our end task is formulated as a binary classification problem, we may intuitively define our loss function as

$$\mathcal{L}(s_i) = -\Sigma_{i=0}^{C} \hat{y}_i log \frac{e^{s_i}}{\Sigma_{j=0}^{C} e^{s_j}} \tag{3}$$

where C denotes the total number of classes. Since our model will perform binary classification, $C = 2$ in our experiment. $\hat{y}_i$ denotes the ground-truth label of the corresponding CT scan x_i.

Specifically, we will optimize the parameters of the segmentation network and the classification network separately in the first stage and altogether in the second stage to minimize the error. We instantiate this formulation as a convolutional neural network, SKINN, which will be introduced in the following subsection.

B. SKINN Architecture

In this paper, we propose Semantic Knowledge Inference Neural Network (SKINN), a multi-stage transfer learning network that performs COVID-19 classification and pixel-wise lesion segmentation in parallel. Inspired by CRNet [19] and Inf-Net [13], the SKINN can classify CT scans and generate

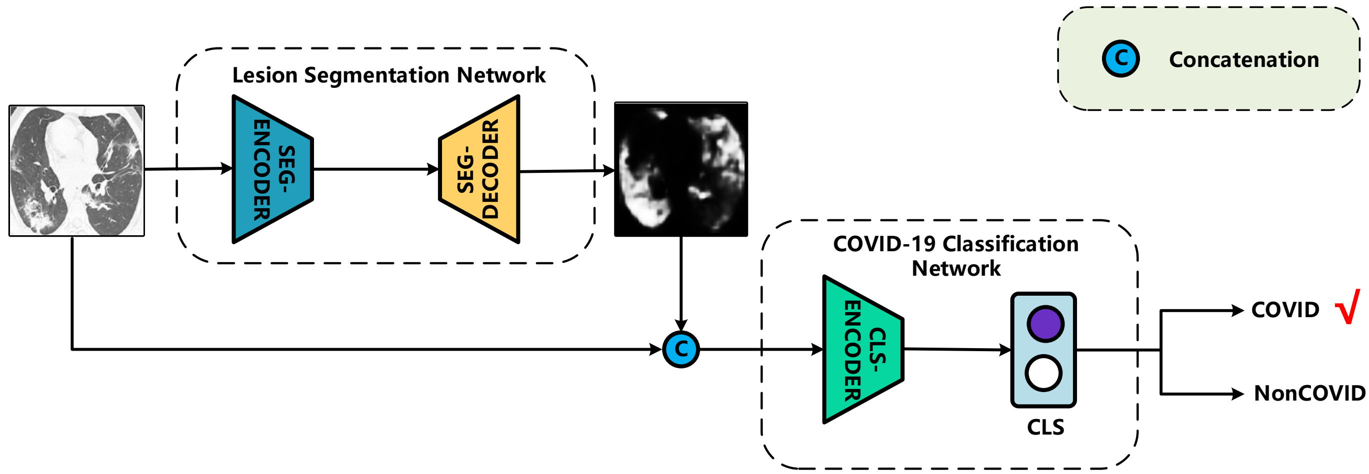

Figure 1. The overview of our SKINN. The SKINN is a hybrid of two sub-networks: the Lesion Segmentation Network (SEG) and the COVID-19 Classification Network (CLS). The Lesion Segmentation Network consists of the SEG-ENCODER and SEG-DECODER. The input of the Segmentation Network is the CT scans. The COVID-19 Classification Network contains the CLS-ENCODER and a Fully Connected layer. The input of the COVID-19 Classification Network is the concatenation of the CT scan and its segmentation mask map.

lesion masks altogether. The overall network architecture is shown in Fig 1. The network comprises two sub-networks, i.e. *Lesion Segmentation Network* and *COVID-19 Classification Network*. The Lesion Segmentation Network consists of the SEG-ENCODER and SEG-DECODER. The COVID-19 Classification Network contains the CLS-ENCODER and a Fully Connected layer. We will introduce these two sub-networks explicitly in the following paragraphs.

1) Overview of SKINN: The SKINN consists of two sub-networks, Lesion Segmentation Network and COVID-19 Classification Network, as shown in Fig 1. The inference workflow of our SKINN is that the CT scan is input into the Lesion Segmentation Network firstly. Passing the SEG-ENCODER and SEG-DECODER, sequentially, the CT scan is transformed into a mask map, where lesion parts are highlighted. With the help of the Lesion Segmentation Network, the semantic knowledge is inferred. Secondly, we concatenate the semantic mask map and the original CT scan and then put it into the COVID-19 Classification Network. The CLS-ENCODER extracts the essential features with the help of inferred semantic knowledge and a followed fully connected layer utilize the features to determine whether it is infected by COVID-19. During the inference stage, our SKINN produces both the Lesion segmentation map and the classification, which helps the clinicians to analyze the CT scans and come up with their decisions.

2) Lesion Segmentation Network: The Lesion Segmentation Network is inspired by Inf-Net [13], but we design a new structure and abandon the semi-supervised part. Our lesion segmentation network consists of Seg-Encoder and Seg-decoder, which achieves better segmentation results.

Seg-Encoder: The encoder works as a feature extractor in SKINN, and it is demonstrated in Figure 2. The encoder takes a batch of CT scan slices in each iteration and feed the image slices into a down-sampling Convolutional Neural Network. The encoder contains two types of layers, i.e., convolution layers and pooling layers. Each convolution layer is a stack of filters that captures different features from the input feature maps. The pooling layers are parameter-less, which perform max-pooling to reduce the size of each feature map. Theoretically, the encoder can be implemented as any CNN architectures. DenseNet [21], a CNN architecture with dense inter-layer connections, is recently attracting more attention because of its highly efficient parameters and extraordinary performance on many computer vision tasks. Therefore, we use a compressed DenseNet with three dense blocks as our encoder in SKINN. The final layer of the encoder, referred to as the ”bottleneck layer”, outputs a stack of feature maps with minimum height and width. The output feature map will be fed into the *seg-decoder* for mask reconstruction.

Seg-Decoder The decoder is a symmetrical version of the *seg-encoder*. The convolution layers in the encoder are mirrored as deconvolution layers in the decoder. The pooling layers are mapped to up-pooling. The aim of introducing the decoder architecture is to reconstruct the input-size mask from the bottleneck layer. Feature maps from the bottleneck layer will be fed into the decoder as an up-sampling process. The output size of the decoder is the same as the input size, i.e. $w \times h$. We will train our segmentation model with the decoder output together with the ground truth maps. Considering the information loss in the feature extraction stage, we will concatenate the multi-level feature maps in the decoder with the corresponding feature map in the encoder. This operation improves the segmentation result by leveraging low-level features extracted from the *seg-encoder*.

3) COVID-19 Classification Network: We designed our COVID-19 Classification Network, as presented in Fig 4, following the analysis of deep learning for COVID-19 diagnosis [19]. Notably, the classification network's input has four channels: RGB channels and a segmentation mask channel. Although four-channel input leads to more parameters, we propose a smaller but more efficient network consisting of

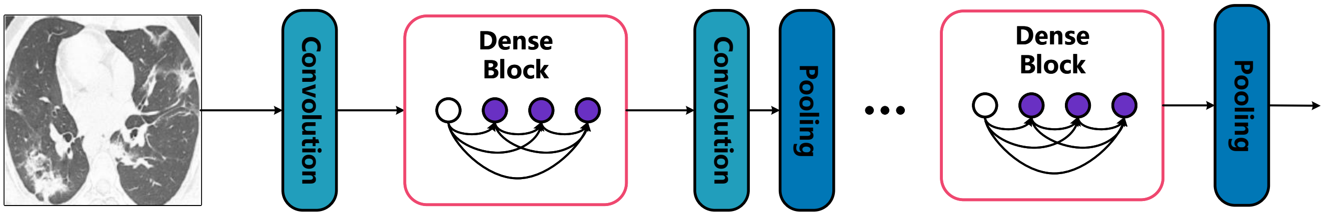

Figure 2. The encoder structure for both the classification network and the segmentation network.

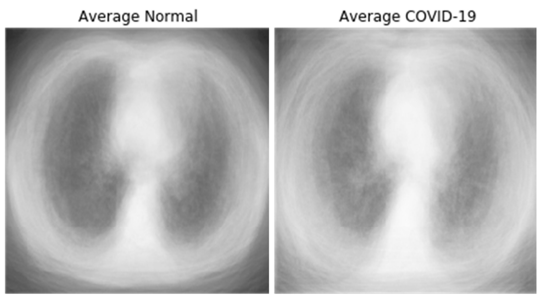

Figure 3. The by-pixel average image of normal people and COVID-19 patients in *COVID-CT Dataset*. COVID-19 patients tend to have more blurred lung contours. Additionally, their pulmonary cavity is less transparent in the CT scan as compared to normal people.

a CLS-ENCODER and a fully connected layer.

Cls-Encoder We construct our Cls-Encoder follows the structure in Figure 2. The input of Cls-Encoder is the concatenation of CT-scans and segmentation maps. The input is feed into a Convolution layer and follows with a Dense Block, which consists of six stacked composed dense layers with 1×1 kernel size and 3×3 kernel size. Convolution and Pooling layers are used as transition layers between the dense blocks. Then three dense blocks are stacked, with 12, 32, 32 dense layers, respectively. Between every two dense blocks, there are transition layers. The output of the Cls-Encoder is a vector. The dimensionality of this vector is $Batchsize \times 1000$.

Classifier The output of the *cls-encoder* is fed into the classifier to perform per-sample COVID-19 classification. We use a linear layer on top of the *cls-encoder* output to stretched the feature maps into a one-dimensional feature vector. This vector is deconvoluted by the deconvolution layer and then follows convolution layers and pooling layers. The feature vector is typically regarded as the input of the classifier. There are two design considerations while choosing the classifier. Firstly, we need to enforce end-to-end training. To utilize the advantage of multi-task learning, we need to guarantee that the classifier can be trained end-to-end with the encoder and the decoder, so traditional machine learning classifiers that require offline training, e.g. Support Vector Machine (SVM), shall be abandoned by SKINN. Secondly, we shall reduce the number of parameters in the classifier. The output of the encoder is a high-dimensional feature map, depending on the number of kernels in the bottleneck layer. Earlier CNN architectures [25], [45] prefer using fully-connected layers as the classifier. But the dense connections are proved to be unnecessary and parameter inefficient. In SKINN, we use a mini fully convolutional network to generate a 2-dimensions result, which will be optimized with Softmax loss.

C. Datasets

To generate the classification and segmentation results at the same time, we trained our model on two datasets: COVID-19 CT segmentation dataset [52] and COVID-CT Dataset [61].

1) COVID-19 CT segmentation dataset: This dataset contains 100 CT images from more than 40 patients with COVID-19. There are 100 training slices images and 100 slices of training COVID-19 masks. The test images include ten slices.

2) COVID-CT Dataset: The COVID-CT Dataset contains 349 CT images, which are labeled as being infected by COVID-19 from 216 patients, and 463 images are labeled as negative in the training set. In the test set, there are 173 images labeled as positive and 168 images labeled as negative. The validation set contains 88 images labeled as being infected by COVID-19 and 64 images labeled as not being infected.

Figure 3 displays the average image for both positive samples of COVID-19 patients and negative samples of normal people. The difference between the two average images suggests that it is possible to classify the CT scans with visual features from the images. But since the difference is subtle, considering CT images from both classes are lung CT scans, it might be insufficient to classify the image well with pure visual information and depend purely on naive search by the Convolutional Neural Networks. In our SKINN, we leverage the semantic information together with the visual features to perform classification.

D. Training Strategy

We train our model using transfer learning in two stages, including separately training and jointly training, for inferring the semantic knowledge and use the knowledge to guide the COVID-19 classification. And the algorithms for the two-stage transfer learning are illustrated in Algorithm 1 and in Algorithm 2, respectively.

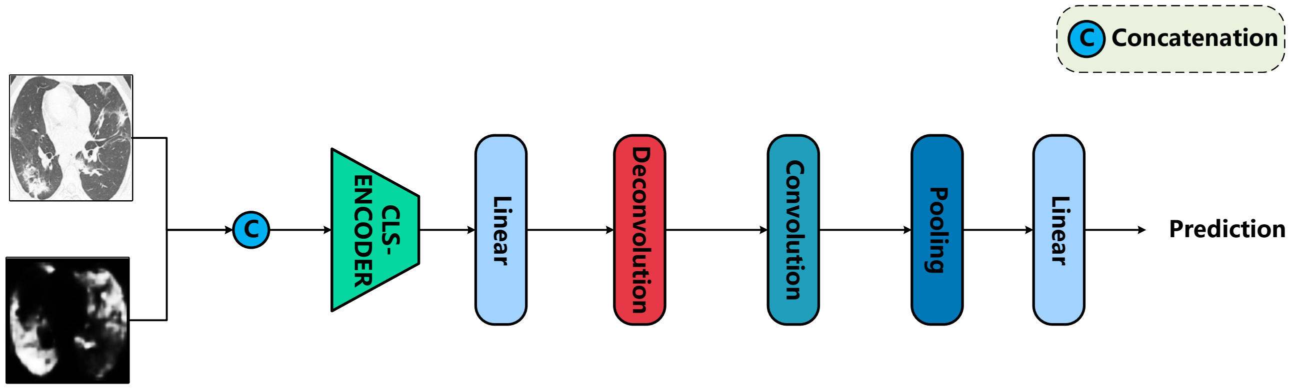

Figure 4. The COVID-19 classifier of classification network. The classifier consists of the first five blocks, and the last block is the classifier.

1) Separately Training: Firstly, we trained the lesion segmentation network solely with the COVID-19 CT segmentation dataset. In this stage, our goal is that, given a CT image, the segmentation network generates the mask, which contains the essential semantic information, the lesion areas. To train the segmentation network, We use the Binary Cross Entropy (BCE) loss function, as shown in Equation 4:

$$\mathcal{L}_{BCE} = \Sigma_{i=0}^{w}\Sigma_{j=0}^{h}|\mathcal{T}\log(\mathcal{S}) + (1-\mathcal{T})log(1-\mathcal{S})| \quad (4)$$

where $\mathcal{S}$ is the predicted map and $\mathcal{T}$ is the ground-truth map. COVID classification network is fixed during the separate training stage.

Algorithm 1 Pretraining algorithm

θ_S, θ_C ▷ The parameters of segmentation and classification network
X_I, Y_I ▷ The data, labels of ImageNet Dataset.
procedure PRETRAINING(θ_S, θ_C, X_I, Y_I) ▷ The parameters of networks and pre-training data
  $iter$ ▷ The number of training iterations
  while $i < iter$ **do**
    $G_S \leftarrow Calculate\_Gradients(X_I, Y_I, \theta_S)$
    $G_C \leftarrow Calculate\_Gradients(X_I, Y_I, \theta_C)$
    $\theta_S \leftarrow Backpropagation(G_S, \theta_S)$ ▷ Updating the parameters of segmentation network
    $\theta_C \leftarrow Backpropagation(G_C, \theta_C)$ ▷ Updating the parameters of classification network
    $i \leftarrow i + 1$
  end

2) Jointly Training: In the jointly training, we train the two sub-networks simultaneously with COVID-CT Dataset. The binary classification loss in Equation 4 is used. Thankfully, because we use the concatenation of the output of the segmentation network and the CT image, both sub-networks can be updated by backpropagation. Hence, the model is trained in an end-to-end way in the jointly training stage.

Algorithm 2 Training algorithm

θ_{SKINN} ▷ The parameters of KINN network
θ_S; $\theta_S \in \theta_{SKINN}$ ▷ The parameters of segmentation network
X_S, Y_S ▷ The data, labels of Segmentation Dataset.
X_C, Y_C ▷ The data, labels of Classification Dataset.
procedure TRAINING(θ_{SKINN}, θ_S, θ_C, X_S, Y_S, X_C, Y_C)
  $iter$ ▷ The number of training iterations
  j ▷ Starting jointly training from j-th iteration.
  while $i < iter$ **do**
    if $i < j$ **then**
      $G_S \leftarrow Calculate\_Gradients(X_S, Y_S, \theta_S)$
      $\theta_S \leftarrow Backpropagation(G_S, \theta_S)$ ▷ Updating the parameters of segmentation network
    else
      $G_{SKINN} \leftarrow Calculate\_Gradients(X_C, Y_C, \theta_{SKINN})$
      $\theta_{SKINN} \leftarrow Backpropagation(G_{SKINN}, \theta_{SKINN})$ ▷ Updating the parameters of SKINN network
    $i \leftarrow i + 1$
  end

IV. RESULTS

A. Settings

We used DenseNet-169 [21] as the base architecture for our SKINN. As a hybrid of segmentation network and classification network, the whole SKINN is finetuned end-to-end with COVID-19 binary labels. In our experiments, we tested multiple hyperparameter sets and selected the parameters with the best performance. We use Adam [24] Solver to optimize our model, with an initial learning rate of 1×10^{-5}. We reduced the learning rate to 0.5 of its original value after each epoch.

B. Metrics

To evaluate the performance of our model, we use three metrics: Accuracy, F1-score, and Area Under the ROC Curve

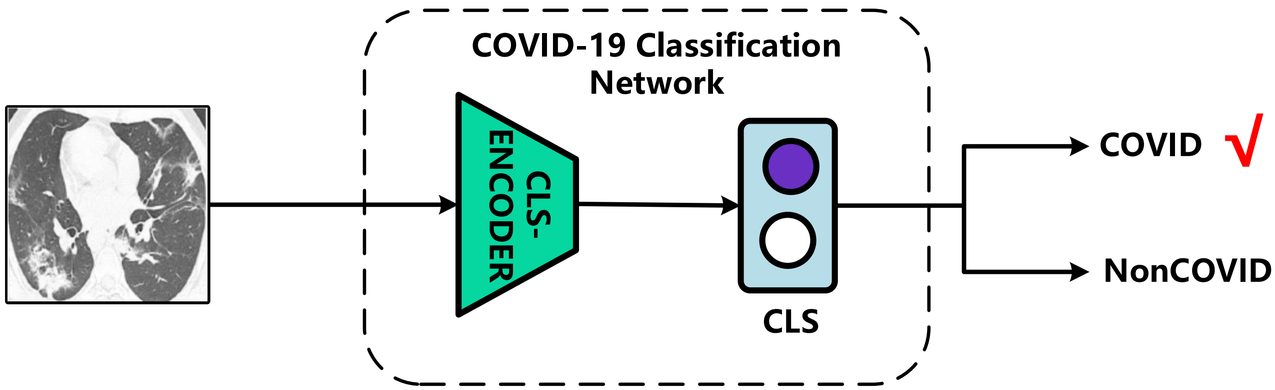

Figure 5. The **CLS** network for ablative study. We keep the COVID-19 Classification Network and remove the lesion segmentation network. The input to the network is the original CT scan. The CLS network is trained end-to-end on *COVID-19 CT dataset* to perform COVID-19 classification.

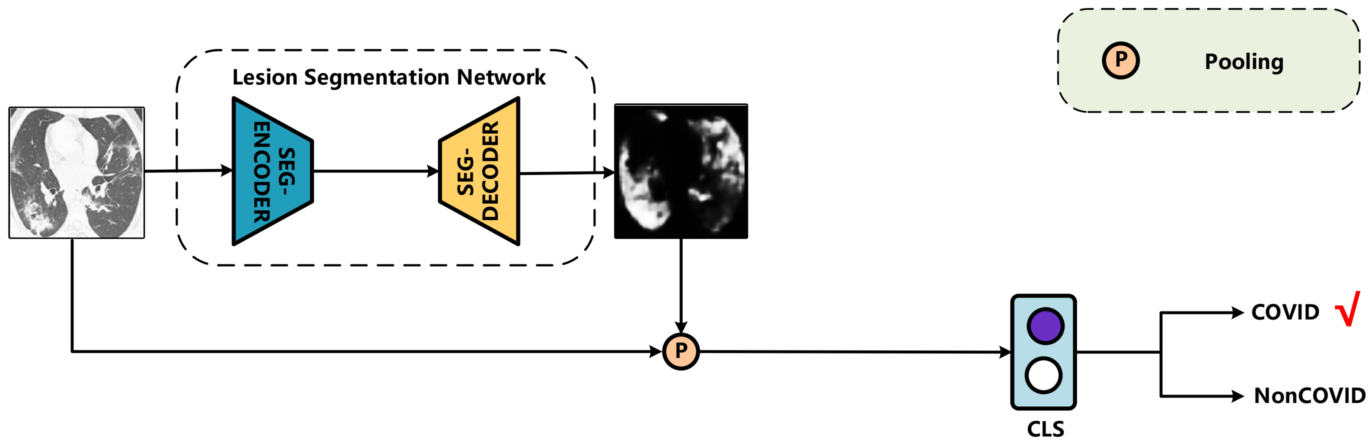

Figure 6. The **SEG** network for ablative study. We keep the Lesion Segmentation Network and remove the COVID-19 Classification Network. The predicted full-size lesion mask is feed into a pooling layer for downsampling and a small fully-connected network to generate binary output. The SEG network is also trained on *COVID-19 CT dataset* for COVID-19 classification.

(AUC) [50]. We detail each metric as follows.

Accuracy: Accuracy is the ratio of number of correct predictions to the total number of predictions that are made, as illustrated in Equation 5,

$$Accuracy = \frac{\Sigma_{i=0}^{N_{pred}} \mathbb{1}_{y_i = \hat{y}_i}}{N_{pred}} \tag{5}$$

where N_{pred} is the number of predictions that have been made by the model.

Area Under the Curve: The Area under the Curve is the area under the curve of plot False Positive Rate and True Positive Rate as X-axis and Y-axis, respectively, at different points in [0, 1].

F1-score: F1-score is the is the harmonic mean of the precision and recall, as shown in Equation 6,

$$F1 = \frac{2}{recall^{-1} + precision^{-1}} \tag{6}$$

where

$$recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

and

$$precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

.

C. Ablative Studies

1) Comparing Methods: To quantify the importance of the proposed segmentation network and classification network, we ablated our SKINN with baselines: **CLS**: We only used the classification network to predict, as shown in Fig 5. **SEG**: We fed the output of the segmentation network into a pooling layer and then used a fully connected layer to get the predictions, as shown in Fig 6. Since the output of the lesion segmentation network is a binary mask of the same w and h as the input CT scan, we need to further map the lesion mask into a binary


(a)

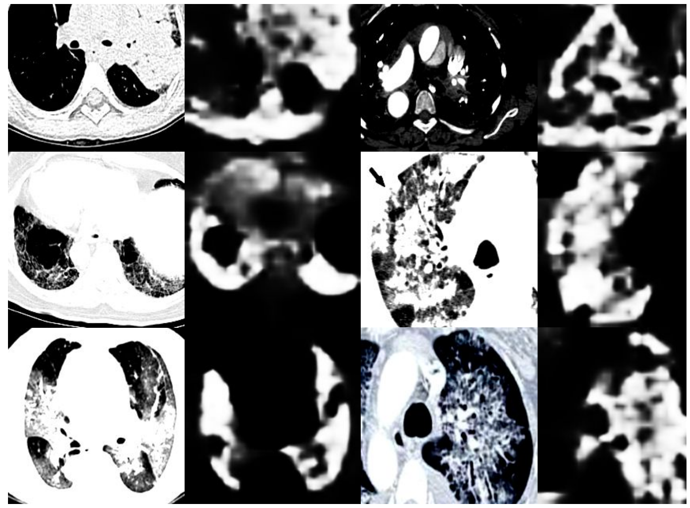
(b)

Figure 7. Examples of the dataset and the covid lesion mask predicted by the SEG Network. For each sample, the images on the left side are the CT scans, and the images on the right side are the lesion masks predicted by the SEG Network. Figure (a) includes negative samples (No COVID-19). Figure (b) contains positive samples (COVID-19). The predicted masks of positive samples contain highlighted regions, denoting lesion areas, are captured from the COVID-19 patients. The apparent visual difference between positive and negative samples can serve as the basis for classification models.

Table I
THE PERFORMANCES OF ABLATIVE BASELINES.

Models	Accuracy	F1-score	AUC
CLS	0.79	0.76	0.90
SEG	0.63	0.57	0.69
SKINN	**0.95**	**0.95**	**0.99**

result. In our experiment, we firstly feed the mask into a pooling layer for downsampling. The size of the new mask is 8×8. Then we flatten the mask into a single-dimensional feature vector and feed the vector into a densely connected network with an output channel of 2. With these adaptations, we could train the SEG network with our binary classification labels end-to-end. **SKINN**: We used our SKINN model with both the segmentation network and classification network, as shown in Fig 1.

2) Analysis: Table I shows that the performance of SKINN is more than 10% and 20% higher than **CLS** and **SEG**, respectively. Moreover, the **SKINN** achieves the value of 0.95, 0.95, and 0.99 scores on Accuracy, F1-score, and AUC, respectively. On the one hand, the results illustrate that both SKINN subnetworks play an important role in good classification predictions. On the other hand, our proposed way of combining the segmentation network and classification network is effective. It shows state-of-the-art performance, proving our assumption that the highlighted lesion segmentation benefits the COVID-19 classification.

D. Comparisons with State-of-The-Arts

1) Comparing Methods: We compare our result with prior arts on *COVID-19 CT Dataset* by three metrics: Accuracy, F1, and AUC. We compare our result with three widely used general-purpose CNN architectures, i.e., ResNet-50 [18], DenseNet-169 [21], and two versions of EfficientNet [51], as the baseline methods. We also compare our SKINN with two recently proposed networks customized specifically for CT-based COVID-19 classification, i.e., CRNet [19] and DenseNet-169 with Self-Supervised Learning [19]. Among the methods introduced above, the DenseNet-169 with Self-Trans was the state-of-the-art method for COVID-19 classification by the time our SKINN is introduced.

Table II
THE PERFORMANCES OF VARIOUS STATE-OF-THE-ART METHODS.

Models	Accuracy	F1-score	AUC
ResNet-50 [18]	0.77	0.74	0.86
DenseNet-169 [21]	0.79	0.76	0.90
EfficientNet-b0 [51]	0.77	0.78	0.89
EfficientNet-b1 [51]	0.79	0.79	0.84
CRNet [19]	0.73	0.76	0.79
DenseNet-169 (Trans) [19]	0.83	0.81	0.87
DenseNet-169 (Self-Trans) [19]	0.86	0.85	0.94
SKINN	**0.95**	**0.95**	**0.99**

2) Analysis: Table II suggests that the widely used baseline Convolutional Neural Network architectures [18], [21], [51] could only achieve an accuracy of less than 80% in the *COVID-19 CT Dataset*. The failure of these baseline models can be explained from two perspective. Firstly, these models are pretrained on ImageNet [12] general classification dataset with millions of images, and finetuned on the COVID-19 CT dataset with only hundreds of samples. The domain adaptation from general images to the CT scans is huge, suggesting the knowledge learned from the ImageNet may not be informative for the CT based task. Secondly, These baseline methods rely solely on the feature extraction by the convolution kernels, which may potentially lose some information. Similar to the baseline methods, CRNet [19] is a reduced version of the

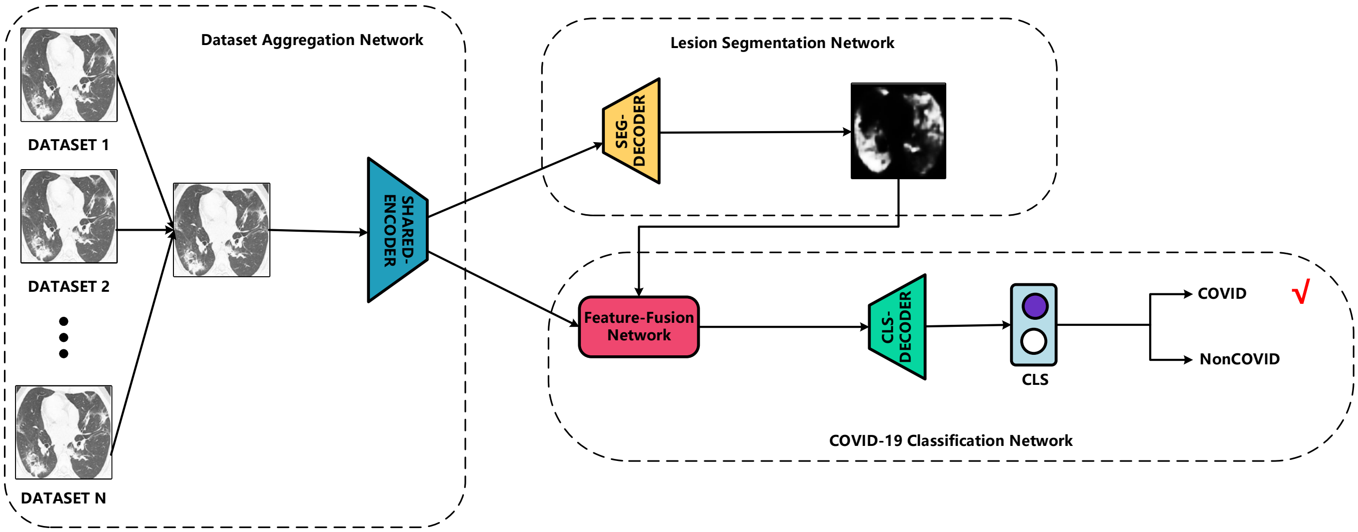

Figure 8. The future architecture of our proposed Multi-datasets SKINN. It includes three sub-networks: the Dataset Aggregation Network, the Lesion Segmentation Network, and COVID-19 Classification Network. The dataset Aggregation network contains a shared encoder. The Lesion Segmentation Network consists of the SEG-DECODER. The input of the Segmentation Network is the features of CT scans from different datasets. The COVID-19 Classification Network contains the Feature-Fusion Network, CLS-ENCODER, and a classifier. The input of the COVID-19 Classification Network is the concatenation of the features of CT scans and CT-scans' segmentation mask maps.

baseline architectures. It also relies on the feature searching by the network itself, and achieved less promising result due to its shallowness and limited model size. He *et al.* [19] introduced a self-supervised learning method, named as Self-Trans, by introducing auxiliary tasks with model-labeled data to enhance the classification result. The information of the Self-Trans is also originated from the CT scans.

In Section IV-C, we discussed the advantage of leveraging semantic information to boost the performance of classification. As shown in Table II, our SKINN achieves state-of-the-art results, which significantly surpassed all the previous approaches by all three evaluation metrics. As compared with the previous state-of-the-art work [19], SKINN shares the same base architecture, i.e. DenseNet-169 [21], as the Self-Trans model [19]. SKINN achieved 95% accuracy and 95% F-1 score, both 9% above the prior art. The Area Under the ROC Curve of SKINN is 0.99, which is close to a perfect solution.

E. Qualitative results

We showed the segmentation maps generated by our segmentation network in Figure 7. As Figure 7 (a) shows, for healthy people, there are little highlighted areas in CT-scans, which means the little regions are considered as lesion areas by the SEG network. Compared with the highlighted areas in the negative samples, the highlighted areas in the CT-scans, which are diagnosed as COVID-19, are much larger, as illustrated in Figure 7 (b). The significant difference in lesion masks between positive and negative samples is the key to the success of our **CLS** model, which relies on both the visual features from the CT scans and the lesion masks for COVID-19 classification. Furthermore, the promising correspondence of lesion areas between the original CT-scans and segmentation masks demonstrates that our segmentation network can find the lesion area accurately, which shows the interpretability of our SKINN. Most importantly, the segmentation map provides an overview of the lung condition, which will help the clinicians with their diagnosis.

V. Discussion and Limitations

There are three limitations in our research that we should focus on in our future works.

Firstly, since most previous COVID-19 classification networks use weakly supervised or semi-supervised methods to highlight the lesion areas, we did not compare our fully supervised segmentation result with the other methods. Qualitative result analysis of the segmentation result should be considered in our future works. We could use pixel-wise accuracy or Intersection over Union (IoU) to compare with other segmentation methods.

Secondly, the combination method for CT scans and the lesion masks can be further improved. Currently, we directly concatenate the lesion mask into the CT scan image as a new channel in addition to the RGB channels. By concatenation, the semantic information is indirectly projected to the visual information, and the feature extraction is highly dependent on the convolution kernels in the SEG network. To combat this, we may also test element-wise operations, adding semantic information as a weight to the CT scans.

Thirdly, for SKINN training, we train the whole network in an end-to-end manner. We did not explicitly differentiate between the SEG network and the CLS network. By the existing design, we assume the classification-segmentation multi-task learning will be beneficial for both tasks. Nonetheless, negative transfer in the back-propagation may result in worse performance for specific tasks. Specifically, the segmentation result might be negatively affected during the training phase for the classification network. This characteristic can be avoided by fixing the segmentation network's parameters or setting a small weight value for segmentation network back-propagation while training the classification network. Further ablative studies need to be conducted to justify the design choices.

VI. SKINN for Multi-datasets

In future work, we try to incorporate more modules for dealing with various CT-scans dataset into our SKINN. As shown in Figure 8, we aggregated the CT-scans in various CT-scans dataset. After pre-processing, the size of all the input images is the same, and we used the same normalization and crop way as our plain SKINN.

Dataset Aggregation Network contains a shared encoder. The input of the Dataset Aggregation Network is the resized CT-scans. It is also trained by two-step transfer learning. Specifically, it is pretrained on ImageNet and finetuned on all the collected datasets. The shared encoder could learn from both the segmentation and classification dataset, which improves its ability to extract key information.

Lesion Segmentation Network contains Seg-Decoder, which has a similar structure as mentioned in Section III. It contains more layers to achieve equivalent performance as the plain SKINN because the Seg-Encoder is replaced by the shared encoder.

COVID-19 Classification Network contains a feature fusion network, CLS Decoder, and a classifier. The feature fusion net is used to fuse the information from the features produced by the shared encoder and the segmentation maps generated by the Lesion Segmentation network because the features are not corresponding to the segmentation. Our plain SKINN uses CT-scans and segmentation that can be aligned with each other. The CLS decoder and classifier follow the same structure as explained in Section III.

With our Multi-datasets SKINN, we can adapt our method on the aggregation of datasets, and the performance of our Multi-datasets SKINN on classification tasks will be improved. The designed Shared Encoder and feature-fusion network help the Multi-datasets SKINN to mitigate the data shifts among different datasets, and we don't need to train an encoder for each dataset, so our multi-datasets SKINN can utilize the data with different size and condition in a reasonably concise model.

VII. Conclusion

In this paper, we propose SKINN, a novel Convolutional Neural Network that fully exploits the semantic information about pneumonia lesions. SKINN uses the predicted lesion mask in conjunction with the CT scans to perform feature extraction and COVID-19 classification. Our SKINN performs better than the prior state-of-the-art methods from all three metrics. We also demonstrate that benefiting from the hybrid architecture, SKINN could generate lesion scans while performing COVID-19 classification. The highly accurate COVID-19 prediction result, together with the lesion mask, will potentially contribute to the automation of CT-based COVID-19 diagnosis.

References

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016.

[2] Asmaa Abbas, Mohammed M Abdelsamea, and Mohamed Medhat Gaber. Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network. *arXiv preprint arXiv:2003.13815*, 2020.

[3] Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Liming Xia. Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology*, page 200642, 2020.

[4] Ahmed Alghamdi, Mohamed Hammad, Hassan Ugail, Asmaa Abdel-Raheem, Khan Muhammad, Hany S Khalifa, and Ahmed A Abd El-Latif. Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. *arXiv preprint arXiv:1906.09358*, 2019.

[5] Musaed Alhussein and Ghulam Muhammad. Voice pathology detection using deep learning on mobile healthcare framework. *IEEE Access*, 6:41034–41041, 2018.

[6] Charmaine Butt, Jagpal Gill, David Chun, and Benson A Babu. Deep learning system to screen coronavirus disease 2019 pneumonia. *Applied Intelligence*, page 1, 2020.

[7] Jun Chen, Lianlian Wu, Jun Zhang, Liang Zhang, Dexin Gong, Yilin Zhao, Shan Hu, Yonggui Wang, Xiao Hu, Biqing Zheng, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. *MedRxiv*, 2020.

[8] Yiqiang Chen, Xin Qin, Jindong Wang, Chaohui Yu, and Wen Gao. Fedhealth: A federated transfer learning framework for wearable healthcare. *IEEE Intelligent Systems*, 2020.

[9] Victor M Corman, Olfert Landt, Marco Kaiser, Richard Molenkamp, Adam Meijer, Daniel KW Chu, Tobias Bleicker, Sebastian Brünink, Julia Schneider, Marie Luisa Schmidt, et al. Detection of 2019 novel coronavirus (2019-ncov) by real-time rt-pcr. *Eurosurveillance*, 25(3):2000045, 2020.

[10] Awwal Muhammad Dawud, Kamil Yurtkan, and Huseyin Oztoprak. Application of deep learning in neuroradiology: Brain haemorrhage classification using transfer learning. *Computational Intelligence and Neuroscience*, 2019, 2019.

[11] Maxime De Bois, Mounîm A El Yacoubi, and Mehdi Ammi. Adversarial multi-source transfer learning in healthcare: Application to glucose prediction for diabetic people. *arXiv preprint arXiv:2006.15940*, 2020.

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[13] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Inf-net: Automatic covid-19 lung infection segmentation from ct images. *IEEE Transactions on Medical Imaging*, 2020.

[14] Yicheng Fang, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang, and Wenbin Ji. Sensitivity of chest ct for covid-19: comparison to rt-pcr. *Radiology*, page 200432, 2020.

[15] Akram Farhadi, David Chen, Rozalina McCoy, Christopher Scott, John A Miller, Celine M Vachon, and Che Ngufor. Breast cancer classification using deep transfer learning on structured healthcare data. In *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 277–286. IEEE, 2019.

[16] Eli Gibson, Wenqi Li, Carole Sudre, Lucas Fidon, Dzhoshkun I Shakir, Guotai Wang, Zach Eaton-Rosen, Robert Gray, Tom Doel, Yipeng Hu, et al. Niftynet: a deep-learning platform for medical imaging. *Computer methods and programs in biomedicine*, 158:113–122, 2018.

[17] Ophir Gozes, Maayan Frid-Adar, Hayit Greenspan, Patrick D Browning, Huangqi Zhang, Wenbin Ji, Adam Bernheim, and Eliot Siegel. 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*, 2020.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[19] Xuehai He, Xingyi Yang, Shanghang Zhang, Jinyu Zhao, Yichen Zhang, Eric Xing, and Pengtao Xie. Sample-efficient deep learning for covid-19 diagnosis based on ct scans. *medRxiv*, 2020.

[20] Derek LG Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. Medical image registration. *Physics in medicine & biology*, 46(3):R1, 2001.

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[22] Md Zabirul Islam, Md Milon Islam, and Amanullah Asraf. A combined deep cnn-lstm network for the detection of novel coronavirus (covid-19) using x-ray images. *Informatics in Medicine Unlocked*, 20:100412, 2020.

[23] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678, 2014.

[24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[26] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[27] Lan Lan, Dan Xu, Guangming Ye, Chen Xia, Shaokang Wang, Yirong Li, and Haibo Xu. Positive rt-pcr test results in patients recovered from covid-19. *Jama*, 323(15):1502–1503, 2020.

[28] Lin Li, Lixin Qin, Zeguo Xu, Youbing Yin, Xin Wang, Bin Kong, Junjie Bai, Yi Lu, Zhenghan Fang, Qi Song, et al. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology*, 2020.

[29] Yan Li and Liming Xia. Coronavirus disease 2019 (covid-19): role of chest ct in diagnosis and management. *American Journal of Roentgenology*, 214(6):1280–1286, 2020.

[30] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2015.

[31] Jun Ma, Yixin Wang, Xingle An, Cheng Ge, Ziqi Yu, Jianan Chen, Qiongjie Zhu, Guoqiang Dong, Jian He, Zhiqiang He, et al. Towards efficient covid-19 ct annotation: A benchmark for lung and infection segmentation. *arXiv preprint arXiv:2004.12537*, 2020.

[32] Taban Majeed, Rasber Rashid, Dashti Ali, and Aras Asaad. Covid-19 detection using cnn transfer learning from x-ray images. *medRxiv*, 2020.

[33] Alejandro R Martinez. Classification of covid-19 in ct scans using multi-source transfer learning. *arXiv preprint arXiv:2009.10474*, 2020.

[34] Minghui Min, Xiaoyue Wan, Liang Xiao, Ye Chen, Minghua Xia, Di Wu, and Huaiyu Dai. Learning-based privacy-aware offloading for healthcare iot with energy harvesting. *IEEE Internet of Things Journal*, 6(3):4307–4316, 2018.

[35] Khan Muhammad, Salman Khan, Javier Del Ser, and Victor Hugo C de Albuquerque. Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[36] Ali Narin, Ceren Kaya, and Ziynet Pamuk. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, 2020.

[37] Ming-Yen Ng, Elaine YP Lee, Jin Yang, Fangfang Yang, Xia Li, Hongxia Wang, Macy Mei-sze Lui, Christine Shing-Yen Lo, Barry Leung, Pek-Lan Khong, et al. Imaging profile of the covid-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*, 2(1):e200034, 2020.

[38] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pages 8026–8037, 2019.

[39] Lan T Phan, Thuong V Nguyen, Quang C Luong, Thinh V Nguyen, Hieu T Nguyen, Hung Q Le, Thuc T Nguyen, Thang M Cao, and Quang D Pham. Importation and human-to-human transmission of a novel coronavirus in vietnam. *New England Journal of Medicine*, 382(9):872–874, 2020.

[40] Adnan Qayyum, Syed Muhammad Anwar, Muhammad Awais, and Muhammad Majid. Medical image retrieval using deep convolutional neural network. *Neurocomputing*, 266:8–20, 2017.

[41] Zhi Zhen Qin, Melissa S Sander, Bishwa Rai, Collins N Titahong, Santat Sudrungrot, Sylvain N Laah, Lal Mani Adhikari, E Jane Carter, Lekha Puri, Andrew J Codlin, et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Scientific reports*, 9(1):1–10, 2019.

[42] Davide Rambaldi and Francesca D Ciccarelli. Fancygene: dynamic visualization of gene structures and protein domain architectures on genomic loci. *Bioinformatics*, 25(17):2281–2282, 2009.

[43] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge, 2015.

[44] Kabid Hassan Shibly, Samrat Kumar Dey, Md Tahzib Ul Islam, and Md Mahbubur Rahman. Covid faster r-cnn: A novel framework to diagnose novel coronavirus disease (covid-19) in x-ray images. *medRxiv*, 2020.

[45] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[46] Dilbag Singh, Vijay Kumar, and Manjit Kaur. Classification of covid-19 patients from chest ct images using multi-objective differential evolution–based convolutional neural networks. *European Journal of Clinical Microbiology & Infectious Diseases*, pages 1–11, 2020.

[47] Ramandeep Singh, Mannudeep K Kalra, Chayanin Nitiwarangkul, John A Patti, Fatemeh Homayounieh, Atul Padole, Pooja Rao, Preetham Putha, Victorine V Muse, Amita Sharma, et al. Deep learning in chest radiography: detection of findings and presence of change. *PloS one*, 13(10):e0204155, 2018.

[48] Korsuk Sirinukunwattana, Shan E Ahmed Raza, Yee-Wah Tsang, David RJ Snead, Ian A Cree, and Nasir M Rajpoot. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging*, 35(5):1196–1206, 2016.

[49] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer's Disease Neuroimaging Initiative, et al. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 101:569–582, 2014.

[50] Mary M Tai. A mathematical model for the determination of total area under glucose tolerance and other metabolic curves. *Diabetes care*, 17(2):152–154, 1994.

[51] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

[52] The MedSeg team. "covid-19 ct segmentation dataset". In *https://medicalsegmentation.com/covid19/*, 2020.

[53] Suat Toraman, Talha Burak Alakus, and Ibrahim Turkoglu. Convolutional capsnet: A novel artificial neural network approach to detect covid-19 disease from x-ray images using capsule networks. *Chaos, Solitons Fractals*, 140:110122, 2020.

[54] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 4489–4497, 2015.

[55] Muhammad Usman, Siddique Latif, Muhammad Asim, Byoung-Dai Lee, and Junaid Qadir. Retrospective motion correction in multishot mri using generative adversarial network. *Scientific Reports*, 10(1):1–11, 2020.

[56] Muhammad Usman, Byoung-Dai Lee, Shi-Sub Byon, Sung-Hyun Kim, Byung-il Lee, and Yeong-Gil Shin. Volumetric lung nodule segmentation using adaptive roi with multi-view residual learning. *Scientific Reports*, 10(1):1–15, 2020.

[57] Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, et al. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). *MedRxiv*, 2020.

[58] Jessica Watson, Penny F Whiting, and John E Brush. Interpreting a covid-19 test result. *Bmj*, 369, 2020.

[59] Yan Xu, Tao Mo, Qiwei Feng, Peilin Zhong, Maode Lai, I Eric, and Chao Chang. Deep learning of feature representation with multiple instance learning for medical image analysis. In *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1626–1630. IEEE, 2014.

[60] Pan Zhai, Yanbing Ding, Xia Wu, Junke Long, Yanjun Zhong, and Yiming Li. The epidemiology, diagnosis and treatment of covid-19. *International journal of antimicrobial agents*, page 105955, 2020.

[61] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. Covid-ct-dataset: a ct scan dataset about covid-19. *arXiv preprint arXiv:2003.13865*, 2020.

[62] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. Springer, 2018.