

Semi-supervised Learning for COVID-19 Image Classification via ResNet

Lucy Nwosu, Xiangfang Li, Lijun Qian, *Senior Member, IEEE*, Seungchan Kim, *Member, IEEE*, Xishuang Dong, *Member, IEEE*

Abstract—Coronavirus disease 2019 (COVID-19) is an ongoing global pandemic in over 200 countries and territories, which has resulted in a great public health concern across the international community. Analysis of X-ray imaging data can play a critical role in timely and accurate screening and fighting against COVID-19. Supervised deep learning has been successfully applied to recognize COVID-19 pathology from X-ray imaging datasets. However, it requires a substantial amount of annotated X-ray images to train models, which is often not applicable to data analysis for emerging events such as COVID-19 outbreak, especially in the early stage of the outbreak. To address this challenge, this paper proposes a two-path semi-supervised deep learning model, *ssResNet*, based on Residual Neural Network (ResNet) for COVID-19 image classification, where two paths refer to a supervised path and an unsupervised path, respectively. Moreover, we design a weighted supervised loss that assigns higher weight for the minority classes in the training process to resolve the data imbalance. Experimental results on a large-scale of X-ray image dataset *COVIDx* demonstrate that the proposed model can achieve promising performance even when trained on very few labeled training images.

Index Terms—COVID-19 Image Classification, Semi-supervised Learning, Residual Neural Network, Joint Optimization, Data Imbalance

I. INTRODUCTION

Coronavirus disease 2019 (COVID-19) outbreak has led to the heavy losses of the world's economy and life. To reduce the spread of COVID-19 and the death rate, it is essential to detect the disease at the early stage with effective and timely screening/testing and place COVID-19 infected patients

L. Nwosu is with the Center of Computational Systems Biology, Department of Electrical and Computer Engineering, Prairie View A&M University, Texas A&M University System, Prairie View, TX 77446, USA (e-mail: lnwosu@student.pvamu.edu).

X. Li is with the Center of Excellence in Research and Education for Big Military Data Intelligence (CREDIT Center) and the Center of Computational Systems Biology, Department of Electrical and Computer Engineering, Prairie View A&M University, Texas A&M University System, Prairie View, TX 77446, USA (e-mail: xili@pvamu.edu).

L. Qian is with the Center of Excellence in Research and Education for Big Military Data Intelligence (CREDIT Center) and the Center of Computational Systems Biology, Department of Electrical and Computer Engineering, Prairie View A&M University, Texas A&M University System, Prairie View, TX 77446, USA (e-mail: liqian@pvamu.edu).

S. Kim is with the Center of Computational Systems Biology, Department of Electrical and Computer Engineering, Prairie View A&M University, Texas A&M University System, Prairie View, TX 77446, USA (e-mail: sekim@pvamu.edu).

X. Dong is with the Center of Computational Systems Biology and the Center of Excellence in Research and Education for Big Military Data Intelligence (CREDIT Center), Department of Electrical and Computer Engineering, Prairie View A&M University, Texas A&M University System, Prairie View, TX 77446, USA (e-mail: xidong@pvamu.edu).

in quarantine immediately [1], [2]. Artificial intelligence (AI), an emerging technology for medical imaging processing, has actively contributed to the fight against COVID-19 [3]. Compared to the traditional imaging workflow that heavily relies on human interpretation, AI enables more safe, accurate and efficient imaging solutions.

Recent AI-empowered applications in COVID-19 detection include the dedicated imaging platform, the lung and infection region segmentation, as well as the clinical assessment and diagnosis [4], [5], [6]. Moreover, commercial products integrate AI to combat COVID-19 and demonstrate the capability of the AI technology [4]. All of these examples show the tremendous enthusiasm cast by the public for AI-empowered progress in the medical imaging field, especially during the ongoing COVID-19 pandemic.

Regarding the COVID-19 research based on AI, COVID-19 image classification becomes more and more attractive, which is to separate COVID-19 patients from non-COVID-19 subjects using the features extracted from medical images. Specially, supervised deep learning such as convolutional neural networks (CNN) has been very popular in this research area. For example, Wang *et al.* proposed a 2D CNN supervised model to analyze delineated region patches to accomplish classification between COVID-19 and typical viral pneumonia [7]. Similarly, Xu *et al.* utilized candidate infection regions to complete COVID-19 classification via supervised ResNet-18 [8].

In addition, as a powerful deep learning model for medical image analysis, UNet [9] was employed for COVID-19 image classification and segmentation. For example, Zheng *et al.* employed UNet to obtain lung segmentation and predicted the probability of COVID-19 with 3D CNN on segmentation features [6]. Jin *et al.* proposed a UNet++ based segmentation model for locating lesions and built a ResNet-50 based classification model for COVID-19 diagnosis [10]. Chen *et al.* implemented COVID-19 classification with the patterns of segmented lesions extracted by supervised UNet++ [11], [12]. Moreover, they employed a 2D Deeplab model for the lung segmentation and a 2D ResNet-152 model for lung-mask slice based identification of positive COVID-19 cases [13]. Although supervised deep learning presents impressive performance on COVID-19 image classification, it requires a large amount of annotated medical images to train models, which is not practical with respect to limited data resources related to COVID-19, due to huge costs of labeling medical images, and labeling noise [14].

To reduce the efforts on labeling medical images for

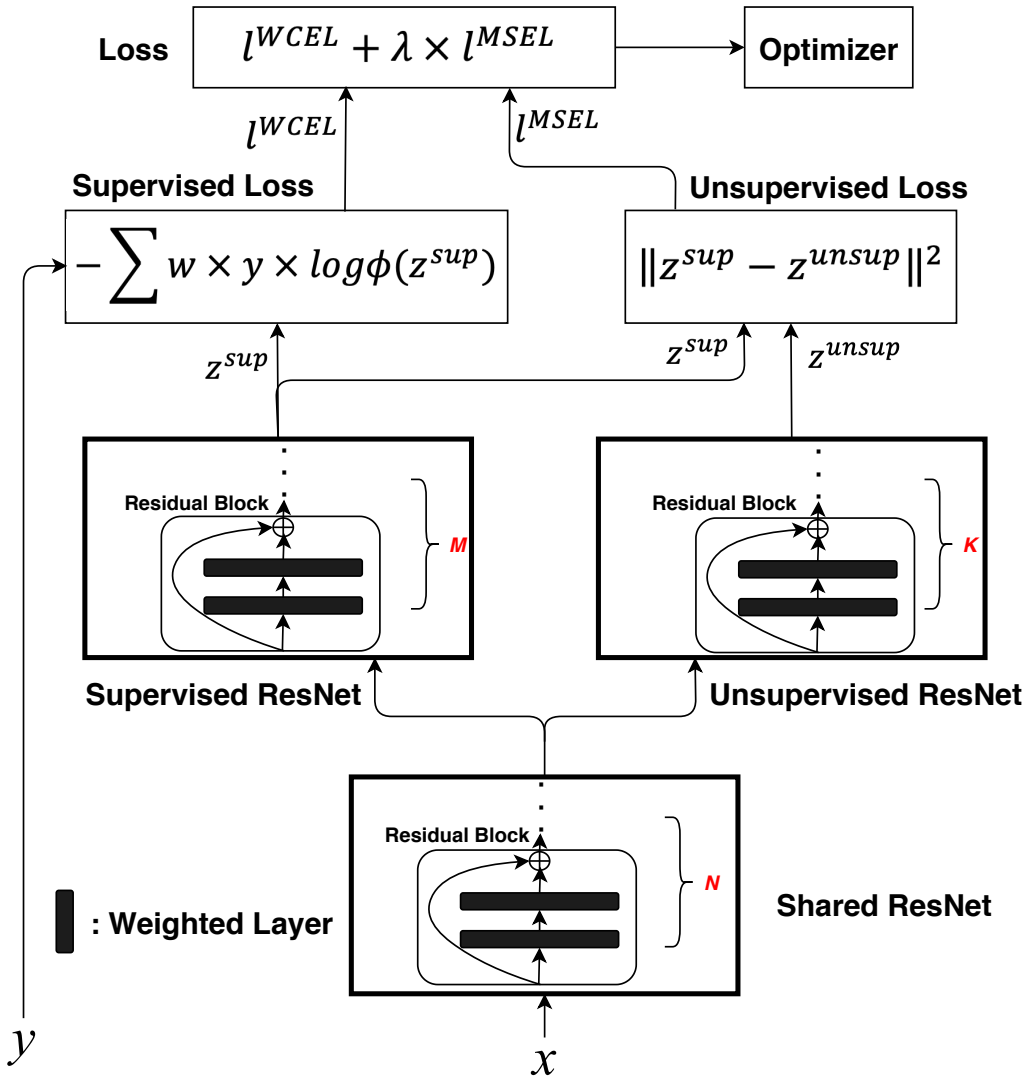


Fig. 1. Framework of the proposed semi-supervised learning. Input x is the medical image. Labels such as y are available only for the labeled inputs. Shared ResNet will evaluate the input to obtain the low-level representations as inputs to supervised ResNet and unsupervised ResNet, where these three ResNets are built with residual blocks and N , M , and K are numbers of residual blocks for these three ResNets. Then z^{sup} and z^{unsup} are outputs from the supervised ResNet and the unsupervised ResNet, respectively. Moreover, z^{sup} and y will be applied to calculate a weighted cross entropy loss l^{WCEL} whereas z^{sup} and z^{unsup} are used to calculate a mean squared error loss l^{MSEL} , where w the weight to different classes of samples. We jointly optimize the combined losses, where λ is the weight for unsupervised loss. \oplus is the short-cut connection in the residue operation.

COVID-19 image classification, we build a two-path semi-supervised deep learning model that is able to learn on both labeled and unlabeled medical images, based on residual neural networks (ResNet) [15]. ResNet is an artificial neural network developed by mimicking pyramidal cells in the cerebral cortex. It is to introduce a so-called “identity shortcut connection” that skips one or more layers since stacking layers should not degrade the network performance. With ResNet, we implement a two-path semi-supervised learning model that is composed of three components, namely, shared ResNet, supervised ResNet, and unsupervised ResNet.

Framework of the proposed model is shown in Fig. 1. One path is composed of a shared ResNet and a supervised ResNet while the other path consists of the shared ResNet and an unsupervised ResNet. All data (labeled and unlabeled data) will be evaluated to calculate the unsupervised loss that

is the mean squared error loss (MSEL), while only labeled data will be used to calculate the supervised loss that is the cross entropy loss (CEL). Specifically, we design a weighted cross entropy loss (WCEL) that assigns more weight to the COVID-19 class for addressing the data imbalance. Reducing MSEL is to enhance the image representation while decreasing WCEL is to enhance classification performance. We validate the proposed model on a large-scale of X-ray image dataset *COVIDx* and experimental results demonstrate the proposed model can accomplish COVID-19 image classification with promising performance even when trained on the extremely limited amount of labeled X-ray images.

The contributions in this study are below.

- We propose a semi-supervised deep learning model with ResNet through jointly training a supervised ResNet and an unsupervised ResNet. We observed that the proposed

model can learn on both unlabeled images and labeled images jointly for COVID-19 image classification with high performance.

- The proposed model is validated on a large-scale COVID-19 image dataset. Experimental results indicate the proposed model is able to effectively recognize COVID-19 images by learning on very few labeled medical images, for example, less than 10% samples in the training data, which meets the requirement of few available labeled data from the medical domain for real applications [14], especially for the cases at the early stage of such global pandemic.

II. PROPOSED METHODOLOGY

We propose a semi-supervised ResNet to address the challenge of lacking of labeled data for COVID-19 image classification, where the detailed framework is shown in Fig. 1. The shared ResNet will generate a new representation of input x , where the new representation z is given by

$$z = f_{pooling}(f_{Resblock_N} \cdots f_{Resblock_1}(x')). \quad (1)$$

where

$$x' = f_{conv}(x). \quad (2)$$

$$f_{Resblock}(x') = x + f_{conv}(f_{conv}(x')). \quad (3)$$

$f_{cov}(\cdot)$ is the convolutional operation. $f_{Resblock}(\cdot)$ is the residual operation [15] and $f_{Resblock_N} \cdots f_{Resblock_1}(\cdot)$ refers to N sequencing residual operations. $f_{pooling}(\cdot)$ is the pooling operation. Introducing this shared ResNet to the proposed model is inspired by deep multi-task learning [16], [17], since different tasks share a low-level feature representation extracted from the input x . In addition, the reason for learning low-level feature representations instead of directly using x is that the original representation may not have enough expressive power for multiple tasks [18]. With the training data in all tasks, a more powerful representation can be learned for all tasks and this representation will improve performance. As shown in Fig. 1, we have two ‘‘tasks’’ in our proposed model, namely, a supervised task and an unsupervised task, which is similar to the framework of deep multi-task learning. Therefore, the shared ResNet is necessary to feed the low-level representations to these two tasks.

The output z from the shared ResNet is evaluated by two ResNets, namely, a supervised ResNet and an unsupervised ResNet. For the supervised ResNet, it is to learn the deep features of labeled samples. The output z^{sup} of the supervised ResNet is given by

$$z^{sup} = f_{pooling}^{sup}(f_{Resblock_M}^{sup} \cdots f_{Resblock_1}^{sup}(z')). \quad (4)$$

where

$$z' = f_{conv}^{sup}(z). \quad (5)$$

$$f_{Resblock}^{sup}(z') = z' + f_{conv}^{sup}(f_{conv}^{sup}(z')). \quad (6)$$

We employ the same operations including the pooling operation $f_{pooling}^{sup}(\cdot)$, the convolutional operation $f_{conv}^{sup}(\cdot)$, and

M sequencing residual operations $f_{Resblock}^{sup}(\cdot)$. Moreover, we build the unsupervised ResNet to generate another representation of all inputs including labeled data and unlabeled data. This representation z^{unsup} is given by

$$z^{unsup} = f_{pooling}^{unsup}(f_{Resblock_K}^{unsup} \cdots f_{Resblock_1}^{unsup}(z'')). \quad (7)$$

where

$$z'' = f_{conv}^{unsup}(z). \quad (8)$$

$$f_{Resblock}^{unsup}(z'') = z'' + f_{conv}^{unsup}(f_{conv}^{unsup}(z'')). \quad (9)$$

Similarly, we employ the pooling operation $f_{pooling}^{unsup}(\cdot)$, the convolutional operation $f_{conv}^{unsup}(\cdot)$, and K sequencing residual operation $f_{Resblock}^{unsup}(\cdot)$ to build the unsupervised ResNet. Then, we utilize those two vectors z^{sup} and z^{unsup} to calculate the weighted cross entropy loss (WCEL) and mean squared error loss (MSEL) for supervised and unsupervised paths, respectively. They are given by

$$l^{WCEL} = - \sum w \times y \times \log \phi(z^{sup}). \quad (10)$$

$$l^{MSEL} = \|z^{sup} - z^{unsup}\|^2. \quad (11)$$

where y is the ground truth of the input and w is corresponding weight. $\phi(\cdot)$ is the softmax activation function. l^{WCEL} is the weighted cross entropy loss to account for the loss of labeled inputs. To enhance classification performance for the minority class (COVID-19 class), we assign more weight to COVID-19 class, where during the learning procedure the classifier will pay more attentions to the COVID-19 class so as to reduce the learning bias that is caused by data imbalance.

l^{MSEL} is to measure the differences between z^{sup} and z^{unsup} . Since training ResNets with dropout regularization and gradient-based optimization is a stochastic process, the two ResNets will have different link weights after training. In other words, there will be differences between the two prediction vectors z^{sup} and z^{unsup} that are from these two ResNets (the supervised ResNet and the unsupervised ResNet). These differences can be treated as an error in the classification and thus minimizing this loss is a goal in the proposed model, which is inspired by Π model [19].

Based on these two losses, the total loss is defined by

$$Loss = l^{WCEL} + \lambda \times l^{MSEL}. \quad (12)$$

where λ is the weight for l^{MSEL} . Training the proposed model is to optimize $Loss$ on the training data. At the beginning of training, the total loss and the learning gradients are dominated by the supervised loss component, i.e., the labeled data only. At later stage of training, unlabeled data will contribute more than labeled data. These processes are controlled by fine-tuning λ [19]. The detailed steps for learning of the proposed model is shown in Algorithm 1. $f_{\theta_{shared}}(\cdot)$ is to learn the low-level features from the medical images. Parameters of the shared ResNet θ_{shared} include weights learned for the operations, namely, pooling operation $f_{pooling}(\cdot)$, convolutional operation $f_{conv}(\cdot)$, and residual operation $f_{Resblock}(\cdot)$.

Algorithm 1 Learning of Semi-supervised ResNet (SSResNet)

Require: training sample x_i , the set of training samples S , label y_i for x_i ($i \in S$)

```

1: for  $t$  in [1, num epochs] do
2:   for each minibatch  $B$  do
3:      $z_{i \in B} \leftarrow f_{\theta_{shared}}(x_{i \in B}) \triangleright$  shared representation
4:      $z_{i \in B}^{sup} \leftarrow f_{\theta_{sup}}(z_{i \in B}) \triangleright$  supervised representation
5:      $z_{i \in B}^{unsup} \leftarrow f_{\theta_{unsup}}(z_{i \in B}) \triangleright$  unsupervised representation
6:      $l_{i \in B}^{WCEL} \leftarrow -\frac{1}{|B|} \sum_{i \in B \cap S} \log \phi(z_{i \in B}^{sup})[y_i w_i] \triangleright$  supervised loss component
7:      $l_{i \in B}^{MSEL} \leftarrow \frac{1}{C|B|} \sum_{i \in B} \|z_{i \in B}^{sup} - z_{i \in B}^{unsup}\|^2 \triangleright$  unsupervised loss component
8:      $Loss \leftarrow l_{i \in B}^{WCEL} + \lambda \times l_{i \in B}^{MSEL} \triangleright$  total loss
9:     update  $\theta_{shared}, \theta_{sup}, \theta_{unsup}$  using optimizer, e.g.,
       ADAM
       return  $\theta_{shared}, \theta_{sup}, \theta_{unsup}$ 

```

After extracting low-level feature representations from inputs, we use $f_{\theta_{sup}}(\cdot)$ and $f_{\theta_{unsup}}(\cdot)$ to obtain higher level representations z^{sup} and z^{unsup} , where z^{sup} is used to complete COVID-19 classification. In addition, z^{sup} and z^{unsup} are employed to enhance the image representations. Parameters of the supervised ResNet θ_{sup} include weights learned for the operations, namely, pooling operation $f_{pooling}^{sup}(\cdot)$, convolutional operation $f_{conv}^{sup}(\cdot)$, and residual operation $f_{Resblock}^{sup}(\cdot)$ while those of the unsupervised ResNet θ_{unsup} consist of weights learned for the operations, namely, pooling operation $f_{pooling}^{unsup}(\cdot)$, convolutional operation $f_{conv}^{unsup}(\cdot)$, and residual operation $f_{Resblock}^{unsup}(\cdot)$. Specially, in the training procedure, we overcome the data imbalance by assigning more weight w_i to the minority class (COVID-19 class) of samples. Finally, we employ ADAM optimizer to jointly optimize the total loss.

III. EXPERIMENT

A. Dataset

We employ a large-scale of chest X-ray dataset *COVIDx* [20] to validate the proposed model. It is comprised of 18,543 chest radiography images across 13,725 cases. Example chest X-ray images belonging to normal, pneumonia, and COVID-19 classes from *COVIDx* dataset are shown in Figure 2. When we examine these examples, we can differentiate these images in terms of features shown within areas marked by the blue circle since we can observe some lighter areas indicating COVID-19 infected regions in the blue circle.

Additionally, when examining the class distribution between training and testing data, we noticed that class distribution of the training set is significantly different from that of testing set. Hence we rebuild the data by splitting the dataset into training and testing datasets that share similar class distributions, where 70% and 30% of data are used for training and testing datasets, respectively. The detailed information of the rebuilt dataset is shown in Table I for sample distribution.

We can observe that the sample distribution is extremely unbalanced regarding the number of samples of COVID-19

TABLE I
SAMPLE DISTRIBUTION IN DIFFERENT CLASSES FOR TRAINING AND TESTING DATASETS

Dataset	Normal	Pneumonia	COVID-19	Total
Training	6,195	6,708	75	12,978
Testing	2,656	2,876	33	5,565
Total	8,851	9,584	108	18,543

class. This poses a great challenge for obtaining a classifier with high performance. We overcome this challenge by the weighted cross entropy loss in the proposed model that is to assign more weight to the minority class (COVID-19 class) during the training, where the details are presented in section two.

B. Experimental settings

In this experiment, our proposed model performed COVID-19 classification. The key hyper parameters for training the proposed model are: Minibatch size: 256, Number of epoch: 50, Optimizer: Adam optimizer, and Initial Learning rate: 0.1. They are determined by trial and error. Moreover, the details of the model architecture is illustrated in Table II, where the residual block is the standard one [15]. Specifically, the output of the proposed model contains two parts: image class $\phi(z^{sup})$ and a new representation z^{unsup} . We employ COVID-Net¹ [20] as a baseline supervised model to present the state-of-the-art performance of COVID-19 image classification for comparison. Furthermore, we compared the proposed model with SRC-MT [21] that is the state-of-the-art of semi-supervised learning since it outperformed Π model [19] and mean teacher model [22] in the area of medical image classification.

TABLE II
THE PROPOSED NETWORK ARCHITECTURE.

Name	Description
Input	Medical Images
Shared ResNet	one convolutional layer, 2 residual block, batch normalization, one pooling layer
Supervised ResNet	one convolutional layer, 2 residual block, batch normalization, one pooling layer
Unsupervised ResNet	one convolutional layer, 2 residual block, batch normalization, one pooling layer
Output	image class $\phi(z^{sup})$ and a new representation z^{unsup}

C. Evaluation metric

We applied different evaluation metrics to evaluate the performance of our proposed model. Since our task is a multi-class classification problem, we use accuracy, macro-average Precision (MacroP), macro-average Recall (MacroR), and macro-average Fscore (MacroF) [23], [24], [25]. Accuracy is calculated by dividing the number of medical images identified correctly over the total number of testing medical images.

¹<https://github.com/lindawang/COVID-Net>

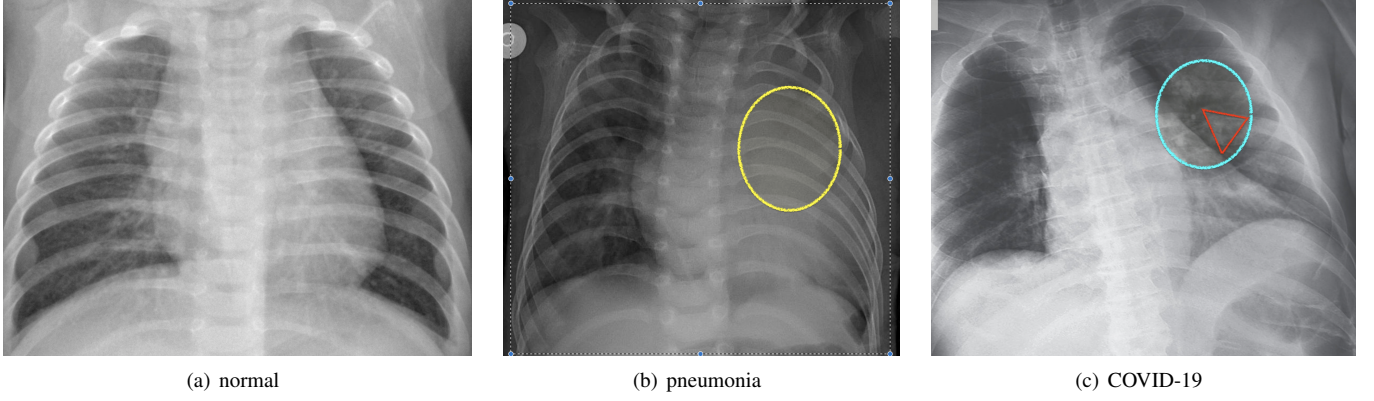


Fig. 2. Examples of chest radiography images belonging to normal, pneumonia, and COVID-19 classes are shown in (a), (b) and (c), respectively. Yellow circle locates infected regions of pneumonia for subfigure (b) while in subfigure (c) the red rectangle shape of region in the blue circle shows the potential infected areas of COVID-19.

$$Accuracy = \frac{N_{correct}}{N_{total}}. \quad (13)$$

Macro-average [26] is to calculate the metrics such as Precision, Recall and F-scores independently for each image class and then utilize the average of these metrics. It is to evaluate the whole performance of classifying image classes.

$$MacroF = \frac{1}{C} \sum_{c=1}^C Fscore_c. \quad (14)$$

$$MacroP = \frac{1}{C} \sum_{c=1}^C Precision_c. \quad (15)$$

$$MacroR = \frac{1}{C} \sum_{c=1}^C Recall_c. \quad (16)$$

where C denotes the total number of image classes and $Fscore_c$, $Precision_c$, $Recall_c$ are $Fscore$, $Precision$, $Recall$ values in the c^{th} image class which are defined by

$$Fscore = \frac{2 \times Precision \times Recall}{Precision + Recall}. \quad (17)$$

where $Precision$ defines the capability of a model to represent only correct image classes and $Recall$ computes the aptness to refer all corresponding correct image classes:

$$Precision = \frac{TP}{TP + FP}. \quad (18)$$

$$Recall = \frac{TP}{TP + FN}. \quad (19)$$

whereas TP (True Positive) counts total number of medical images matched the annotated images. FP (False Positive) measures the number of recognized classes does not match the annotated images. FN (False Negative) counts the number of medical images that does not match the predicted medical images. The ideal case of learning from imbalanced datasets such as *COVIDx* is to improve the recall without hurting the precision. However, recall and precision goals are often conflicting, since when increasing the true positive (TP) for

the minority class (True), the number of false positives (FP) can also be increased; this will reduce the precision [27]. In addition, we employ confusion matrix to check the detailed performance for each class, especially on COVID-19 class.

D. Experimental results

We evaluated the proposed model performance in four steps. The first step is to examine the performance of supervised learning baselines, which is to prove if ResNet is a reasonable supervised model for COVID-19 image classification. A competitive supervised baseline is useful to compare the proposed semi-supervised model in order to present the effectiveness of the proposed model. Furthermore, we will check whether fewer labeled data will lead to lower performance. The second step is to comparing the proposed model with state-of-the-art semi-supervised learning. The third step is to examine whether the hyper-parameter setting will affect the performance of the proposed model significantly. Finally, we will discuss why the proposed model cannot classify certain COVID-19 cases.

1) *Supervised learning for COVID-19 classification*: Table III presents the comparison of supervised baselines built with ResNet. We observed that ResNet (100%) can outperformed COVID-Net (100%) when comparing accuracy, macro-average precision, and macro-average Fscore. It means that ResNet is a competitive supervised baseline for COVID-19 image classification. Additionally, for learning on fewer labeled data, we only focus on the cases of 5%, 7%, and 9% labeled data since the labeled data will be very scarce in medical domain [14] during the early stage of a global pandemic such as COVID-19 outbreak. We observed that the classification accuracy can be improved by increasing the labeled data to train ResNet. Meanwhile, the performance such as accuracy and MacroF is reduced significantly when comparing with ResNet (100%), which demonstrates that more labeled data is imperative for building high-performance supervised models. Moreover, we observed that weighted ResNet cannot improve the performance since we might assign inappropriate weight to different classes.

On the other side, we compare their confusion matrix to examine the performance details in Fig. 3. It indicates that

TABLE III

COMPARISON ON SUPERVISED BASELINE PERFORMANCE. RESNET IS TRAINED ON DIFFERENT RATIOS (%) OF LABELED X-RAY IMAGES. WEIGHTED RESNET IS BUILT BY ASSIGNING MORE WEIGHT TO COVID-19 CLASS DURING TRAINING FOR OVERCOMING THE CHALLENGE OF DATA IMBALANCE.

DL	Accuracy (%)	MacroP (%)	MacroR (%)	MacroF (%)
COVID-Net (100%)	93.98	72.70	96.05	77.33
ResNet (100%)	94.68	90.52	78.87	83.04
ResNet (5%)	87.59	58.23	58.50	58.30
ResNet (7%)	89.27	59.51	59.91	59.68
ResNet (9%)	90.10	60.05	60.44	60.24
Weighted ResNet (5%)	86.45	57.63	57.95	57.78
Weighted ResNet (7%)	88.70	59.15	59.54	59.31
Weighted ResNet (9%)	89.36	66.16	60.46	60.78

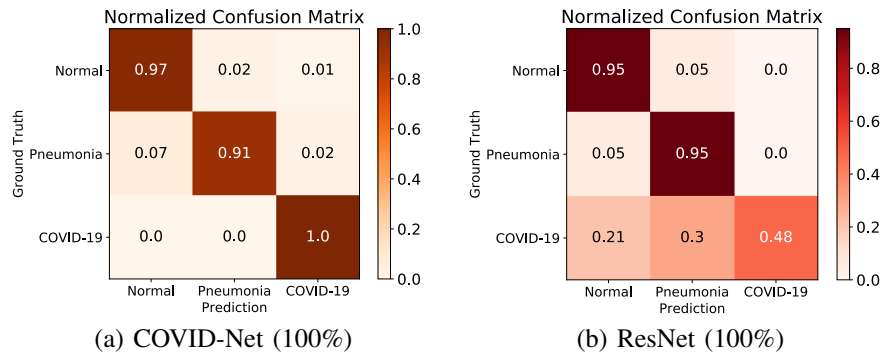


Fig. 3. Comparison of confusion matrix generated by COVID-Net (100%) and ResNet (100%).

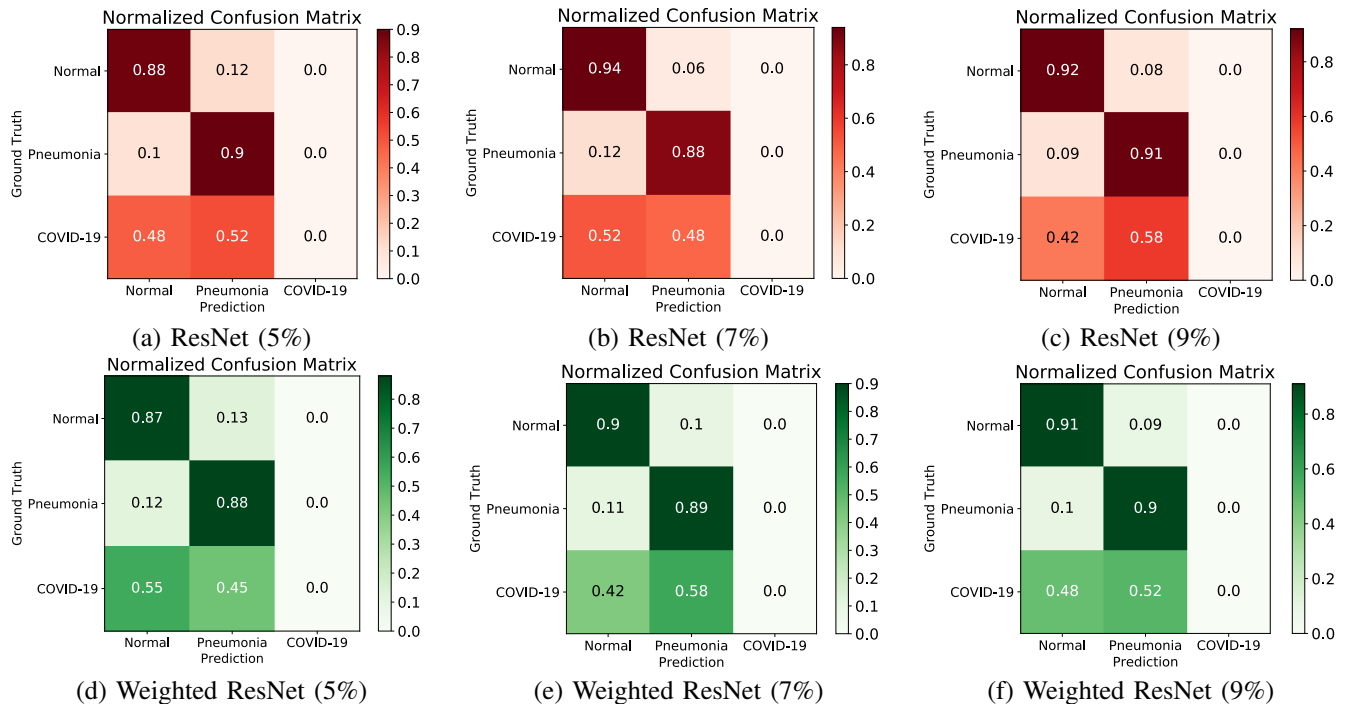


Fig. 4. Comparison of confusion matrix generated by different ResNets training on different ratios of labeled data. Weighted ResNet is built by assigning more weight to COVID-19 class during training for overcome the challenge of data imbalance.

ResNet (100%) can be a promising supervised baseline model when compared to COVID-Net in terms of the accuracies on the normal and pneumonia classes. For the COVID-19 class, ResNet is lower than COVID-Net since COVID-Net employed transfer learning to enhance performance.

To check the performance for each class when learning on

fewer labeled data, we present the detailed performance with confusion matrix shown in Fig. 4. When we use low ratios of labeled training data to train models, ResNet cannot recognize COVID-19 images effectively, which is duo to insufficient COVID-19 labeled samples. In the training sets of these cases, only few of images are for the COVID-19 class. For example,

in the case of ResNet (5%), we only have three images for COVID-19 class in the training data, which means most of training images are for the classes of Normal and Pneumonia. Learning on this data will lead to classification bias. Weighted ResNet was not sufficient to enhance the performance, which means even more weight assigned to COVID-19 class is not enough to overcome the lack of labeled samples to learn distinguish features to differentiate COVID-19 patients from Non-COVID-19 patients on X-ray images with supervised learning.

2) *Comparing the proposed model with state-of-the-art semi-supervised learning*: In this section, we will examine if the proposed model is able to effectively identify COVID-19 samples by training on very limited amount of annotated images. Table IV presents the comparison of classification performance between SRC-MT and the proposed model (SS-ResNet). Overall accuracies of SRC-MT are better than those of the proposed model. However, when only 5% labeled samples were used for training, MacroF of our proposed model is higher than that of SRC-MT, which indicates that the proposed model is more effective in detecting COVID-19 samples. can detect COVID-19 samples with higher performance. It means that compared to SRC-MT, the unsupervised path could enhance the data representation for improving COVID-19 classification more effectively.

In addition, we examine detailed performance of each class with confusion metrics shown in Fig. 5. We observed that the accuracy of recognizing COVID-19 by the proposed model is higher than that of SMC-TC, which means SSResNets can learn more effective features from unlabeled data to recognize COVID-19 samples. Furthermore, with the increased ratios of labeled data, the accuracies of recognizing COVID-19 is enhanced significantly. It means that the unsupervised path can enhance the representations of images to improve the classification. In other words, unlabeled data contributed to increasing the COVID-19 classification performance significantly by enhancing the image representations with the unsupervised path of the SSResNet.

3) *Hyper-parameter setting*: In addition to examining the performance comparison between the proposed models and baselines, we have to figure out whether the proposed model is sensitive to the hyper-parameters. There are various hyper-parameters involved in the learning procedure of the proposed model. Here, we choose class weight to check since different weights would lead to different performance of recognizing COVID-19 samples. Table V shows the comparison results for different weights of three classes. We observe that different weights will result in significant differences of the performance when examining the values of accuracy. On the other hand, compared to accuracy and MacroP, MacroR and MacroF are less sensitive to the weight of COVID-19 class. Generally, we have to delicately select the weight for COVID-19 class to obtain the optimal performance.

4) *Error Analysis*: Fig. 6 presents three COVID-19 samples that are classified into Normal, COVID-19, and Pneumonia classes, respectively. X-ray images of COVID-19 patients shows various features for different stages of COVID-19

patients². At the early stage of COVID-19 patients, X-ray images cannot present significant features (Fig. 6 (a)) that can be used to differentiate COVID-19 and Non COVID-19 patients, which leads to the incorrect classification result for the sample. It is consistent with the expectation that X-ray images are not ideal evidences to support diagnosis of COVID-19 for the patients at the early stage.

However, with development of COVID-19, X-ray images are able to present obvious features such as multifocal lung airspace opacities, nodules and consolidation (Fig. 6 (b)), which contributes to the correct classification result. Unfortunately, if the patients are at the late stage of COVID-19, X-ray images presents lobar diffused consolidation (See Fig. 6 (c)) that is similar to features of pneumonia. These features will be confusing to the proposed model and lead to the incorrect result for the sample shown in Fig. 6 (c). In summary, in terms of samples shown in Fig. 6, the proposed model will be effective for the patients who are in the development of COVID-19 rather than those at the early stage or late stage of such disease.

IV. RELATED WORK

Deep learning technique has shown its power on classification of COVID-19. Ghoshal *et al.* [28] proposed a Bayesian convolutional neural network to estimate the diagnosis uncertainty in COVID-19 prediction, where the dataset includes 70 lung X-ray images of patients with COVID-19 from an online COVID-19 dataset [29], and non-COVID-19 images from Kaggle's Chest X-Ray data (Pneumonia). Narin *et al.* [30] is to detect COVID-19 infection from X-ray images through comparing three different deep learning models, namely, ResNet50, InceptionV3, and Inception-ResNetV2. The evaluation results show that the ResNet50 model outperformed other two models. Zhang *et al.* [31] also utilized ResNet to complete COVID-19 classification on X-ray images and estimated an anomaly score to optimize the COVID-19 score for the classification. In addition, Wang *et al.* [20] propose COVID-Net to detect COVID-19 cases using X-ray images. In general, most current studies use X-ray images to differentiate between COVID-19 and other pneumonia and healthy subjects.

In addition to COVID-19 image classification, it is imperative to figure out the regions of infection of COVID-19 since it will provide detailed information on COVID-9 for diagnosis. Semantic segmentation is able to help us recognize the regions and corresponding patterns to assess and quantify COVID-19, where the regions of interest (ROIs) contains those of lung, lobes, bronchopulmonary segments, and infected regions or lesions, in the chest X-ray or CT images. Moreover, segmented regions could be further used to extract handcrafted or self-learned features for diagnosis and other applications. Deep learning has promoted the development of semantic segmentation of images significantly [9], [32]. To segment ROIs in CT, the segmentation networks for COVID-19 include classic U-Net [6], [33], [5], UNet++ [12], and VB-Net [34]. The segmentation methods related to COVID-19 can be classified

²<https://www.uclahealth.org/radiology/covid-19-chest-x-ray-guideline>

TABLE IV
COMPARING PERFORMANCE BETWEEN SRC-MT AND OUR MODEL (SEMI-SUPERVISED RESNET (SSRESNET)).

Semi-supervised Model	Accuracy (%)	MacroP (%)	MacroR (%)	MacroF (%)
SRC-MT (5%)	90.67	61.08	60.75	60.59
SRC-MT (7%)	89.82	89.92	74.13	78.95
SRC-MT (9%)	92.79	93.61	79.15	84.15
Our model	Accuracy (%)	MacroP (%)	MacroR (%)	MacroF (%)
SSResNet (5%)	84.95	61.18	66.76	62.41
SSResNet (7%)	84.21	63.67	67.85	62.83
SSResNet (9%)	81.79	59.34	70.99	59.19

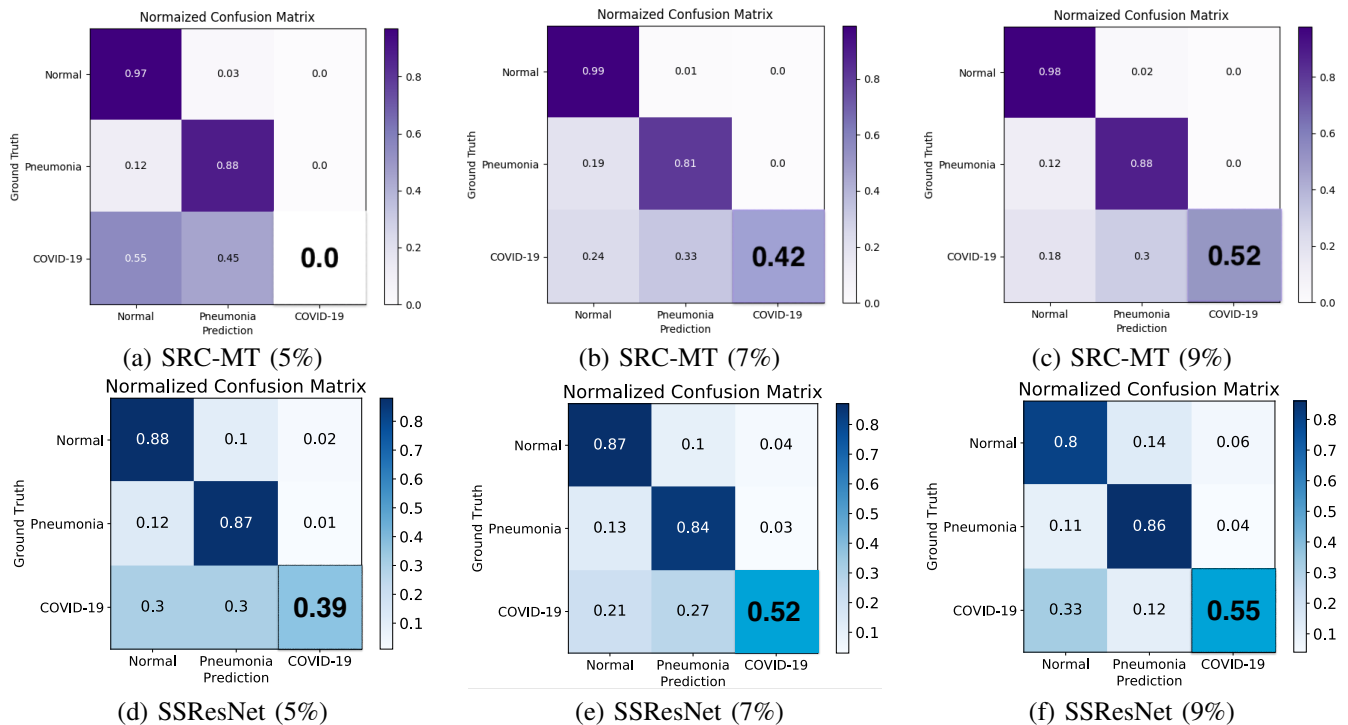


Fig. 5. Comparison of confusion matrix generated with SRC-MT and SSResNets trained on different ratios of labeled data.

TABLE V
COMPARING PERFORMANCE WITH DIFFERENT CLASS WEIGHTS. c_1 , c_2 , AND c_3 ARE THE WEIGHTS OF NORMAL CLASS, PNEUMONIA CLASS, AND COVID-19 CLASS, RESPECTIVELY.

Class Weight	5% Labeled Data			
	Accuracy	MacroP	MacroR	MacroF
$c_1 = 1, c_2 = 1, c_3 = 2$	84.95	61.18	66.76	62.41
$c_1 = 1, c_2 = 1, c_3 = 5$	78.82	58.18	67.38	56.79
$c_1 = 1, c_2 = 1, c_3 = 10$	66.78	57.24	66.70	50.77
Class Weight	7% Labeled Data			
	Accuracy	MacroP	MacroR	MacroF
$c_1 = 1, c_2 = 1, c_3 = 2$	85.90	65.75	64.62	64.84
$c_1 = 1, c_2 = 1, c_3 = 5$	84.21	63.67	67.85	62.83
$c_1 = 1, c_2 = 1, c_3 = 10$	79.31	58.57	67.71	59.39
Class Weight	9% Labeled Data			
	Accuracy	MacroP	MacroR	MacroF
$c_1 = 1, c_2 = 1, c_3 = 2$	87.28	70.32	62.93	64.65
$c_1 = 1, c_2 = 1, c_3 = 5$	84.69	60.91	66.79	61.86
$c_1 = 1, c_2 = 1, c_3 = 10$	81.79	59.34	70.99	59.19

into two groups: 1) the lung-region-oriented methods and 2) the lung-lesion-oriented methods. The first group aims at separating lung regions, i.e., whole lung and lung lobes, from other (background) regions in CT or X-ray images [10], [35].

For example, Jin *et al.* [10] is to detect the whole lung region with UNet++. The second group is to detect lesions (or metal and motion artifacts) in the lung from lung regions [36], [37]. The experimental results indicate that the segmentation

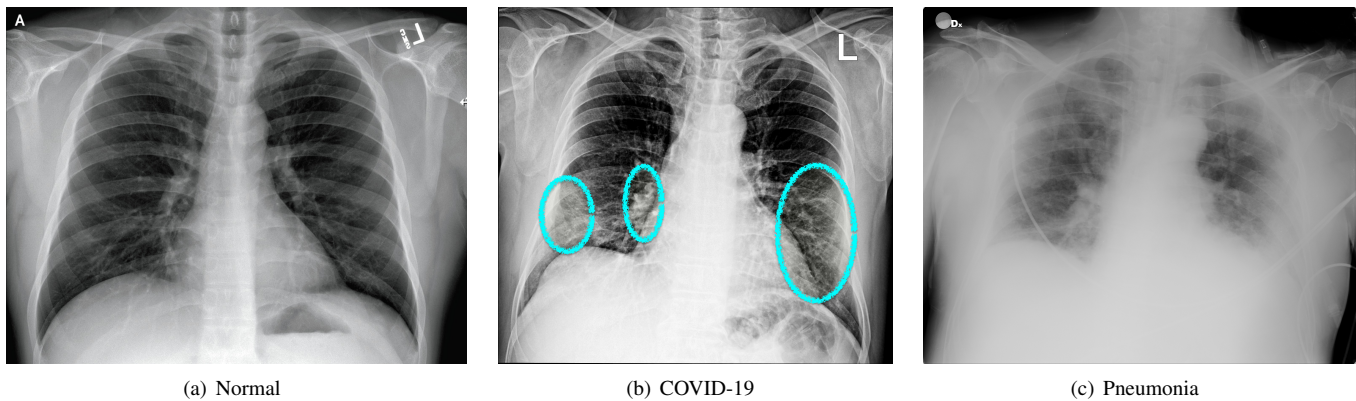


Fig. 6. COVID-19 samples classified into Normal, COVID-19, and Pneumonia classes, are shown in (a), (b), and (c) respectively. The blue circles locate the infected regions of COVID-19.

of X-ray images is even more challenging because of the ribs projected onto soft tissues in 2D. Although supervised deep learning outperforms other models on these two tasks, it requires substantial amount of labeled data to train the model, which is not practical in real applications.

Semi-supervised deep learning has attracted lots of attentions since it has the strong ability to generalize the model performance through learning on labeled data and unlabeled data [38], [39], [40], [19]. Generally, it is to train the deep neural networks by jointly optimizing the standard supervised classification loss on labeled samples and an unsupervised loss on unlabelled data [38], [19]. The rationale of these semi-supervised learning models is to enrich the supervision signals by exploiting the knowledge learned on unlabeled data [41], or regularize the network by enforcing smooth and consistent classification boundaries [40]. Regarding COVID-19 research such as COVID-19 image classification and image segmentation, semi-supervised learning is employed to resolve the lacking of labeled data [42], [43], [44], [45], [46], [47]. However, for COVID-19 image classification, these studies [42], [43], [44] have not comprehensively examined the model performance on a large-scale of X-ray image dataset such as *COVIDx* [20] by comparing with the state-of-the-art, especially for the case of very few labeled data such as less than 10% labeled data. This paper proposed a semi-supervised deep learning model for COVID-19 image classification and checked out the model performance systematically on the *COVIDx* [20] dataset.

V. CONCLUSION AND FUTURE WORK

In this paper, a novel framework of semi-supervised deep learning is proposed for COVID-19 image classification on chest X-ray images. Supervised learning based COVID-19 classification on X-ray datasets could provide useful information to medical staff for facilitating a diagnosis of COVID-19 in an effective and efficient manner. Unfortunately, it relies on the availability of large amount of labeled medical images, which are not available in practice in the early outbreak of such global pandemic. Hence, we propose a semi-supervised learning model based on ResNet that can utilize unlabeled images to enhance classification performance. There are two

paths in the model for reducing supervised cross entropy loss and unsupervised mean squared error loss, respectively. Then training is performed by jointly optimizing these two losses, which allows the proposed scheme to take advantage of the information from both labeled and unlabeled images. Experimental results demonstrate that the proposed model could recognize COVID-19 lung pathology effectively by learning on very limited labeled images and substantial unlabeled images. For the future work, we plan to extend the proposed model for other tasks such as COVID-19 image segmentation.

ACKNOWLEDGMENT

This research work is supported in part by the Texas A&M Chancellor's Research Initiative (CRI), the U.S. National Science Foundation (NSF) award 1464387 and 1736196, and by the U.S. Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) under agreement number FA8750-15-2-0119. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. National Science Foundation (NSF) or the U.S. Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) or the U.S. Government. In addition, Y. Zhao, in the nursing program of Houston Community College, helped complete the error analysis of COVID-19 classification by using her medical imaging expertise.

REFERENCES

- [1] "NY Times: Germany got testing right, what can we learn?" accessed: 2020-04-29.
- [2] "CDC: Testing for COVID-19," [urlhttps://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html](https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html), accessed: 2020-04-19.
- [3] J. Bullock, K. H. Pham, C. S. N. Lam, M. Luengo-Oroz *et al.*, "Mapping the landscape of artificial intelligence applications against covid-19," *arXiv preprint arXiv:2003.11336*, 2020.
- [4] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," *IEEE reviews in biomedical engineering*, 2020.

- [5] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. 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.
- [6] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for covid-19 from chest ct using weak label," *medRxiv*, 2020.
- [7] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng *et al.*, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *MedRxiv*, 2020.
- [8] C. Butt, J. Gill, D. Chun, and B. A. Babu, "Deep learning system to screen coronavirus disease 2019 pneumonia," *Applied Intelligence*, p. 1, 2020.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [10] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng *et al.*, "Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system in four weeks," *medRxiv*, 2020.
- [11] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. 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*. Springer, 2018, pp. 3–11.
- [12] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng *et al.*, "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study," *medRxiv*, 2020.
- [13] C. Jin, W. Chen, Y. Cao, Z. Xu, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, and J. Feng, "Development and evaluation of an ai system for covid-19 diagnosis," *medRxiv*, 2020.
- [14] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [16] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *European conference on computer vision*. Springer, 2014, pp. 94–108.
- [17] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.
- [18] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [19] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," *arXiv preprint arXiv:1610.02242*, 2016.
- [20] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *arXiv preprint arXiv:2003.09871*, 2020.
- [21] Q. Liu, L. Yu, L. Luo, Q. Dou, and P. A. Heng, "Semi-supervised medical image classification with relation-driven self-ensembling model," *IEEE Transactions on Medical Imaging*, 2020.
- [22] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Advances in neural information processing systems*, 2017, pp. 1195–1204.
- [23] V. Van Asch, "Macro-and micro-averaged evaluation measures [[basic draft]]," *Belgium: CLiPS*, vol. 49, 2013.
- [24] F. Zhu, H. Li, W. Ouyang, N. Yu, and X. Wang, "Learning spatial regularization with image-level supervisions for multi-label image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5513–5522.
- [25] H. Borgli, V. Thambawita, P. H. Smedsrud, S. Hicks, D. Jha, S. L. Eskeland, K. R. Randel, K. Pogorelov, M. Lux, D. T. D. Nguyen *et al.*, "Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Scientific Data*, vol. 7, no. 1, pp. 1–14, 2020.
- [26] Y. Yang, "A study of thresholding strategies for text categorization," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, 2001, pp. 137–145.
- [27] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in *Data mining and knowledge discovery handbook*. Springer, 2009, pp. 875–886.
- [28] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection," *arXiv preprint arXiv:2003.10769*, 2020.
- [29] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection," *arXiv preprint arXiv:2003.11597*, 2020.
- [30] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *arXiv preprint arXiv:2003.10849*, 2020.
- [31] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "Covid-19 screening on chest x-ray images using deep learning based anomaly detection," *arXiv preprint arXiv:2003.12338*, 2020.
- [32] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.
- [33] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, and L. Xia, "Serial quantitative chest ct assessment of covid-19: Deep-learning approach," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200075, 2020.
- [34] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, and Y. Shi, "Lung infection quantification of covid-19 in ct images with deep learning," 2020.
- [35] L. Tang, X. Zhang, Y. Wang, and X. Zeng, "Severe covid-19 pneumonia: Assessing inflammation burden with volume-rendered chest ct," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200044, 2020.
- [36] Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, and H. Shi, "Longitudinal assessment of covid-19 using a deep learning-based quantitative ct pipeline: Illustration of two cases," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200082, 2020.
- [37] G. Gaál, B. Maga, and A. Lukács, "Attention u-net based adversarial architectures for chest x-ray lung segmentation," *arXiv preprint arXiv:2003.10304*, 2020.
- [38] J. Weston, F. Ratle, and R. Collobert, "Deep learning via semi-supervised embedding," in *Proceedings of the 25th international conference on machine learning*, 2008, pp. 1168–1175.
- [39] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Advances in neural information processing systems*, 2014, pp. 3581–3589.
- [40] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," in *Advances in neural information processing systems*, 2015, pp. 3546–3554.
- [41] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, vol. 3, no. 2, 2013.
- [42] J. Zhou, B. Jing, and Z. Wang, "Soda: Detecting covid-19 in chest x-rays with semi-supervised open set domain adaptation," *arXiv preprint arXiv:2005.11003*, 2020.
- [43] S. Calderon-Ramirez, A. Moemeni, D. Elizondo, S. Colreavy-Donnelly, L. F. Chavarria-Estrada, M. A. Molina-Cabello *et al.*, "Correcting data imbalance for semi-supervised covid-19 detection using x-ray chest images," *arXiv preprint arXiv:2008.08496*, 2020.
- [44] A. Patricchio, T. Scarlatti, M. Mattheakis, P. Protopapas, and M. Brambilla, "Semi-supervised neural networks solve an inverse problem for modeling covid-19 spread," *arXiv preprint arXiv:2010.05074*, 2020.
- [45] J. Ma, Z. Nie, C. Wang, G. Dong, Q. Zhu, J. He, L. Gui, and X. Yang, "Active contour regularized semi-supervised learning for covid-19 ct infection segmentation with limited annotations," *Physics in Medicine & Biology*, 2020.
- [46] D. Yang, Z. Xu, W. Li, A. Myronenko, H. R. Roth, S. Harmon, S. Xu, B. Turkbey, E. Turkbey, X. Wang *et al.*, "Federated semi-supervised learning for covid region segmentation in chest ct using multi-national data from china, italy, japan," *arXiv preprint arXiv:2011.11750*, 2020.
- [47] A. D. Berenguer, H. Sahli, B. Joukovsky, M. Kvasnytsia, I. Dirks, M. Alioscha-Perez, N. Deligiannis, P. Gonidakis, S. A. Sánchez, R. Brahimetaj *et al.*, "Explainable-by-design semi-supervised representation learning for covid-19 diagnosis from ct imaging," *arXiv preprint arXiv:2011.11719*, 2020.