

# 4S-DT: Self-Supervised Super Sample Decomposition for Transfer Learning With Application to COVID-19 Detection

Asmaa Abbas<sup>1</sup>, Mohammed M. Abdelsamea<sup>2</sup>, and Mohamed Medhat Gaber<sup>3</sup>

**Abstract**—Due to the high availability of large-scale annotated image datasets, knowledge transfer from pretrained models showed outstanding performance in medical image classification. However, building a robust image classification model for datasets with data irregularity or imbalanced classes can be a very challenging task, especially in the medical imaging domain. In this article, we propose a novel deep convolutional neural network, which we called self-supervised super sample decomposition for transfer learning (4S-DT) model. The 4S-DT encourages a coarse-to-fine transfer learning from large-scale image recognition tasks to a specific chest X-ray image classification task using a generic self-supervised sample decomposition approach. Our main contribution is a novel self-supervised learning mechanism guided by a super sample decomposition of unlabeled chest X-ray images. 4S-DT helps in improving the robustness of knowledge transformation via a downstream learning strategy with a class-decomposition (CD) layer to simplify the local structure of the data. The 4S-DT can deal with any irregularities in the image dataset by investigating its class boundaries using a downstream CD mechanism. We used 50 000 unlabeled chest X-ray images to achieve our coarse-to-fine transfer learning with an application to COVID-19 detection, as an exemplar. The 4S-DT has achieved a high accuracy of 99.8% on the larger of the two datasets used in the experimental study and an accuracy of 97.54% on the smaller dataset, which was enriched by augmented images, out of which all real COVID-19 cases were detected.

**Index Terms**—Chest X-ray image classification, convolutional neural network (CNN), data irregularities, self-supervision, transfer learning.

## I. INTRODUCTION

**D**IAGNOSIS of COVID-19 is associated with the symptoms of pneumonia and chest X-ray tests [1]. Chest X-ray is the essential imaging technique that plays an important role in the diagnosis of COVID-19 disease. Fig. 1(a) and (b) shows examples of a normal chest X-ray, a positive one with

COVID-19, a positive image with the severe acute respiratory syndrome (SARS), and some examples of other unlabeled chest X-ray images used in this work.

Several statistical machine learning methods have been previously used for automatic classification of digitized lung images [2], [3]. For instance, in [4], a small set of three statistical features were calculated from lung texture to distinguish between malignant and benign lung nodules using a support vector machine (SVM) classifier. A statistical co-occurrence matrix method was used with backpropagation network [5] to classify samples from being normal or cancerous. With the high availability of enough annotated image data, deep learning approaches [6]–[8] usually provide a superiority performance over the statistical machine learning approaches. Convolutional neural network (CNN) is one of the most commonly used deep learning approaches with superior achievements in the medical imaging domain [9]. The primary success of CNN is due to its capability to learn local features automatically from domain-specific images, unlike the statistical machine learning methods. One of the popular strategies for training a CNN model is to transfer learned knowledge from a pretrained network that fulfilled one generic task into a new specific task [10]. Transfer learning is faster and easy to apply without the need for a huge annotated dataset for training; therefore, many scientists tend to adopt this strategy, especially with medical imaging. Transfer learning can be accomplished with three main scenarios [11]: 1) “shallow tuning” that adapts only the classification layer in a way to cope with the new task and freezes the weights of the remaining layers without updating; 2) “deep tuning” that aims to retrain all the weights of the adopted pretrained network from end-to-end; and 3) “fine-tuning” that aims to gradually train layers by tuning the learning parameters until a significant performance boost is achieved. Transfer knowledge via fine-tuning scenario demonstrated outstanding performance in chest X-ray and computed tomography (CT) image classification [12], [13].

The emergence of COVID-19 as a pandemic disease dictated the need for faster detection methods to contain the spread of the virus. As aforementioned, chest X-ray imaging comes in as a promising solution, particularly when combined with an effective machine learning model. In addition to data irregularities that can be dealt with through class-decomposition (CD), scarcity of data, especially in the early months of the pandemic, made it hard to realize the adoption of chest X-ray images as a means for detection. On the other hand, self-supervised learning is being popularized recently

Manuscript received September 15, 2020; revised January 23, 2021 and April 1, 2021; accepted May 17, 2021. Date of publication May 26, 2021; date of current version July 7, 2021. (Corresponding author: Mohammed M. Abdelsamea.)

Asmaa Abbas is with the Department of Mathematics, University of Assiut, Assiut 71515, Egypt.

Mohammed M. Abdelsamea is with the School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7AP, U.K., and also with the Department of Computer Science, Faculty of Computers and Information, University of Assiut, Assiut 71515, Egypt (e-mail: Mohammed.abdelsamea@bcu.ac.uk).

Mohamed Medhat Gaber is with the School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7AP, U.K.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TNNLS.2021.3082015>.

Digital Object Identifier 10.1109/TNNLS.2021.3082015

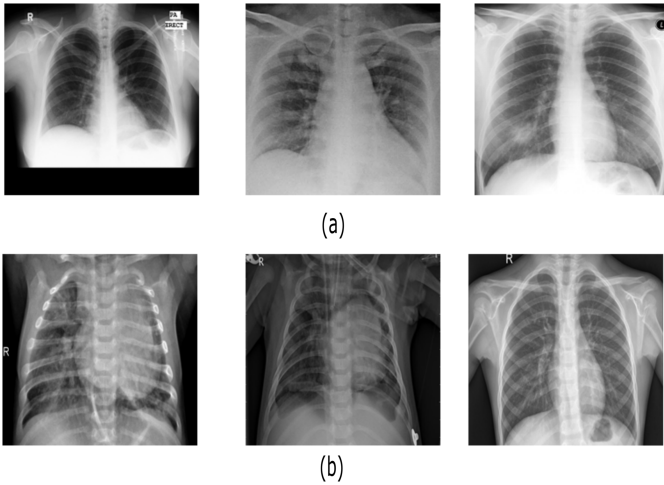


Fig. 1. Examples of (a) labeled chest X-ray images (from left to right: normal, COVID-19, and SARS images) and (b) unlabeled chest X-ray images used in this work for self-supervision learning.

to address the expensive labeling of data acquired at an unprecedented rate. In self-supervised learning, unlabeled data are used for feature learning by assigning each example a pseudo label. In the case of CNNs applied on image data, each image is assigned a pseudo label, and CNN is trained to extract visual features of the data. The training of a CNN by pseudo-labeled images as input is called pretext task learning, whereas the training of the produced CNN from the pretext training using labeled data is called downstream task training. Inherently, such a pipeline allows for effective utilization of large unlabeled datasets. The success of the pretext task learning relies on pseudo labeling methods. In [14], four categories of methods were identified. Context-based image feature learning by means of context similarity has demonstrated a particularly effective pseudo labeling mechanism. DeepCluster [15] is the state-of-the-art method under this category. DeepCluster is a super sample decomposition method that generates pseudo labels through the clustering of CNN features. Sample decomposition is the process of applying clustering on the whole training set as a step for improving supervised learning performance [16]. When the clustering is performed on a larger data sample, we refer to this process as super sample decomposition. However, we argue that the coupling of the pretext task and the pseudo labeling can limit the effectiveness of the pretext task in the self-supervised learning process. In our proposed super sample decomposition, the pretext task training uses cluster assignments as pseudo labels, where the clustering process is decoupled from the pretext training. We propose the clustering of encoded images through an autoencoder (AE) neural network, allowing flexibility of utilizing different features and clustering methods, as appropriate. We argue that this can be most effective in medical image classification, evident by the experimentally validated use of CD for transfer learning in a method coined as DeTraC [17].

In this article, we propose a novel deep CNN, we term self-supervised super sample decomposition for transfer learning (4S-DT) model for the detection of COVID-19 cases.<sup>1</sup> 4S-DT

has been designed in a way to encourage coarse-to-fine transfer learning based on a self-supervised sample decomposition approach. 4S-DT can deal with any irregularities in the data distribution and the limited availability of training samples in some classes. The contributions of this article can be summarized as follows.

- 1) We provide a novel mechanism for self-supervised sample decomposition using a large set of unlabeled chest X-ray images for a pretext training task.
- 2) We provide a generic coarse-to-fine transfer learning strategy to gradually improve the robustness of knowledge transformation from large-scale image recognition tasks to a specific chest X-ray image classification task.
- 3) We provide a downstream CD layer in the downstream training phase to cope with any irregularities in the data distribution and simplify its local structure.
- 4) We provide a thorough experimental study on COVID-19 detection, pushing the boundaries of state-of-the-art techniques in terms of accuracy and robustness of the proposed model.

This article is organized as follows. In Section II, we review the state-of-the-art methods for COVID-19 detection. Section III discusses the main components of our proposed 4S-DT model. Section IV describes our experiments on several chest X-ray images collected from different hospitals. In Section V, we discuss our findings and conclude the work.

## II. PREVIOUS WORK ON COVID-19 DETECTION FROM CHEST X-RAY

In February 2020, the World Health Organization (WHO) has declared that a new virus called COVID-19 has started to spread aggressively in several countries [18]. Diagnosis of COVID-19 is typically associated with pneumonia-like symptoms, which can be revealed by both genetic and imaging tests. Fast detection of the virus will directly contribute to managing and controlling its spread. Imaging tests, especially chest X-rays, can provide fast detection of COVID-19 cases. The historical conception of medical image diagnostic systems has been comprehensively explored through an enormous number of approaches ranging from statistical machine learning to deep learning. A CNN is one of the most effective approaches in the diagnosis of lung diseases, including COVID-19 directly from chest X-ray images. Several recent reviews have been carried out to highlight significant contributions to the detection of COVID-19 [19]–[21]. For instance, in [22], a modified version of ResNet-50 pretrained CNN model has been used to classify CT images into three classes: healthy, COVID-19, and bacterial pneumonia. In [23], a CNN model, called COVID-Net, based on transfer learning was used to classify chest X-ray images into four classes: normal, bacterial infection, non-COVID, and COVID-19 viral infection. In [24], a weakly supervised approach has been proposed using the 3-D chest CT volumes for COVID-19 detection and lesion localization relying on ground-truth masks obtained by an unsupervised lung segmentation method and a 3-D ResNet pretrained model. In [25], a dataset of chest X-ray images from patients with pneumonia, confirmed COVID-19 disease, and

<sup>1</sup>The developed code is available at <https://github.com/asmaa4may/4S-DT>

normal incidents was used to evaluate the performance of the state-of-the-art CNN models based on transfer learning. The study suggested that transfer learning can provide important biomarkers for the detection of COVID-19 cases. It has been experimentally demonstrated that transfer learning can provide a robust solution to cope with the limited availability of training samples from confirmed COVID-19 cases [26].

In [27], self-supervised learning using context distortion is applied for classification, segmentation, and localization in different medical imaging problems. When used in classification, the method was applied for scan plane detection in fetal 2-D ultrasound images, showing classification improvement in some settings. However, we argue that our proposed method is more effective because context distortion is able to generate localized features, instead of global image features that can be more effective for classification tasks.

Having reviewed the related work, it is evident that despite the notable success of deep learning in the detection of COVID-19 cases from chest X-ray images, data scarcity and irregularities have not been explored explicitly. It is common in medical imaging in particular that datasets exhibit different types of irregularities (e.g., overlapping classes with imbalance problems) that affect the resulting accuracy of deep learning models. With the unfolding of COVID-19, chest X-ray images are rather scarce. Thus, this work focuses on coping with data irregularities through CD and data scarcity through super sample decomposition, as detailed in Section III.

### III. 4S-DT MODEL

This section describes, in sufficient details, our proposed deep CNN, 4S-DT model for detecting COVID-19 cases from chest X-ray images. Starting with an overview of the architecture through to the different components of the model, the section discusses the workflow and formalizes the method.

The 4S-DT model consists of three training phases (see Fig. 2). In the first phase, we adapted a self-supervised super sample decomposition mechanism to learn the local structure of generic chest X-ray images, with the aim of overcoming the limited availability of labeled chest X-ray images by leveraging the huge availability of unlabeled chest X-ray images. This is by training an AE model to extract deep local features from each sample (i.e., image) in a super large set of unlabeled generic chest X-ray images. Once the feature space of the unlabeled image dataset is constructed, a sample decomposition approach is adopted to create pseudo labels for the generic chest X-ray images. In the second phase, we use the pseudo labels to achieve coarse transfer learning using an ImageNet pretrained CNN model for the classification of pseudo-labeled chest X-ray images (as a pretext training task), resulting in chest X-ray-related convolutional features. This pretext training task aims at fine-tuning parameters of an ImageNet pretrained CNN model to cope with a generic chest X-ray image recognition task. In the last phase, we use trained convolutional features to achieve downstream training. The downstream training task is more task-specific by adapting a fine transfer learning from generic chest X-ray recognition to COVID-19 detection. In this stage, we also adapt a CD layer to simplify the local structure of the image data distribution,

where a sophisticated gradient descent optimization method is used. Finally, we apply a class composition to refine the final classification of the images.

#### A. Super Sample Decomposition

Given a set of unlabeled images  $X = \{x^1, x^2, \dots, x^{n'}\}$ , our super sample decomposition component aims to find and use pseudo labels during the pretext training task of 4S-DT. To this end, an AE is first used to extract deep features associated with each image. For each input image  $x$ , the representation vector  $h^d$  and the reconstructed image  $\hat{x}$  can be defined as

$$h^d = f(W^{(1)}x + b^{(1)}) \quad (1)$$

$$\hat{x} = f(W^{(2)}h^d + b^{(2)}) \quad (2)$$

where  $W^{(1)}$  and  $W^{(2)}$  are the weight matrices,  $b^{(1)}$  and  $b^{(2)}$  are the bias vectors, and  $f$  is the active function. The reconstruction error  $L(x, \hat{x})$  between  $\hat{x}$  and  $x$  is defined as

$$L(x, \hat{x}) = \frac{1}{2} \|x - \hat{x}\|^2. \quad (3)$$

The overall cost function of the  $n'$  unlabeled images,  $E_{\text{AE}}(W, b)$ , can be defined as

$$E_{\text{AE}}(W, b) = \left[ \frac{1}{n'} \sum_{i=1}^{n'} L(x^i, \hat{x}^i) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} \left( W_{ji}^{(l)} \right)^2 \quad (4)$$

where the first term denotes the reconstruction error of the whole datasets and the second term is the regularization weight penalty term, which aims to prevent overfitting by restraining the magnitude of the weights.  $\lambda$  is the weight decay parameter,  $n_l$  is the layer number of the network,  $s_l$  denotes the neuron number in layer  $l$ , and  $W_{ji}^{(l)}$  is the connecting weight between neuron  $i$  in layer  $l+1$  and neuron  $j$  in layer  $l$ .

Once the training of the AE has been accomplished, density-based spatial clustering of applications with noise (DBSCAN) is used to cluster the image data distribution  $X$  into a number of classes  $c$  based on the extracted features  $h^d$ . DBSCAN is an unsupervised clustering algorithm, which is a considerably representative density-based clustering algorithm that defines clusters as the largest set of points connected by density.

Let the image dataset  $X$  be mapped into a low-dimensional feature space denoted by  $H \in R^{n' \times d}$ , where  $H = (h_1, h_2, \dots, h_{n'})$ . An image  $x^i$  (represented by  $h_i$ ) is density-connected to image  $x^j$  (represented by  $h_j$ ) with respect to Eps (i.e., neighborhood radius) and MinPts (i.e., the minimum number of objects within the neighborhood radius of core object) if there exists a core object  $x^k$  such that both  $x^i$  and  $x^j$  are directly density-reachable from  $x^k$  with respect to Eps and MinPts. An image  $x^i$  is directly density-reachable from an image  $x^j$  if  $x^i$  is within the Eps-neighborhood of  $N_{\text{Eps}}(x^j)$  and  $x^j$  is a core object, where Eps-neighborhood can be defined as

$$N_{\text{Eps}}(x_j) = \{x_i \in X | \text{dis}(x_i, x_j) \leq \text{Eps}\}. \quad (5)$$

DBSCAN results in  $C$  clusters, where each cluster is constructed by maximizing the density reachability relationship

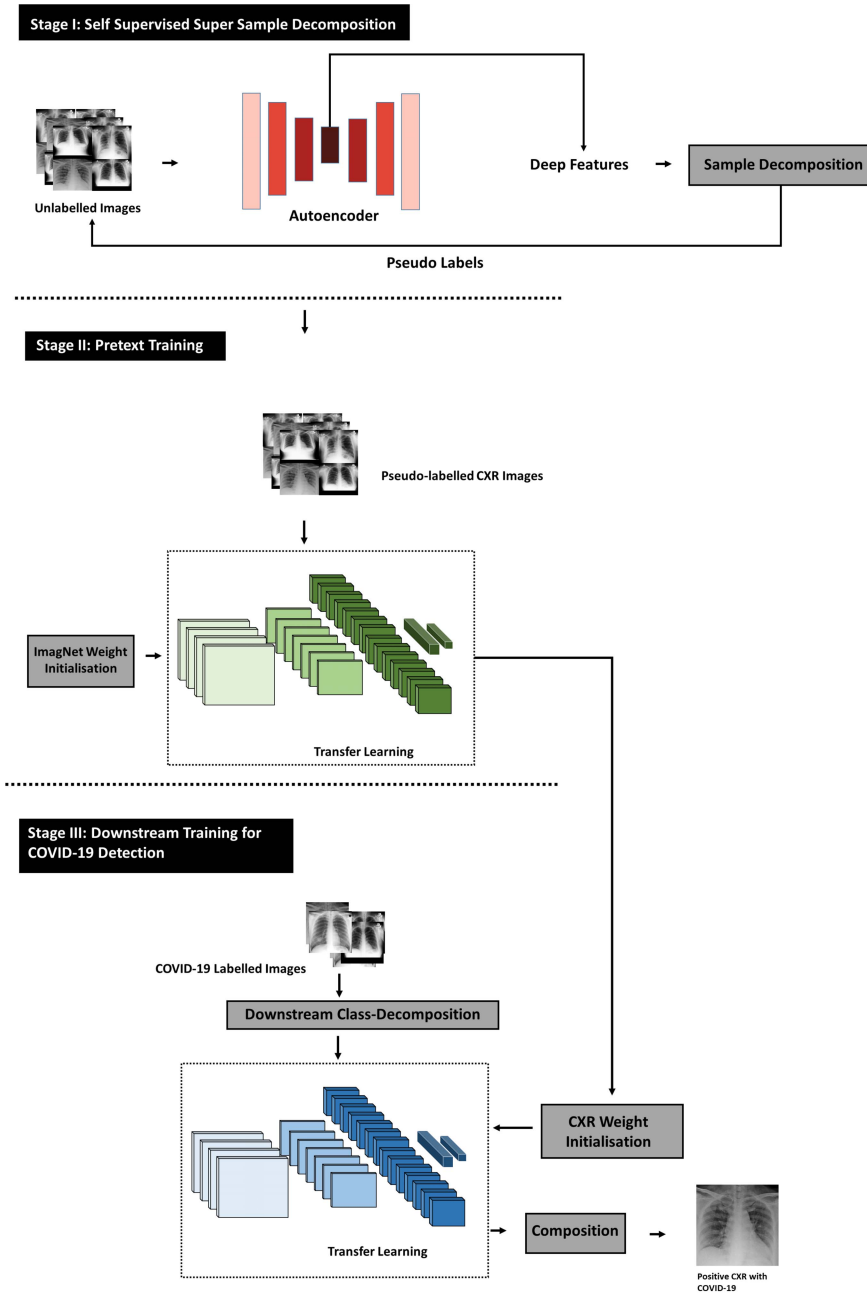


Fig. 2. Graphical representation of 4S-DT model. In stage I, a self-supervised sample decomposition mechanism has been designed to generate pseudo labels for unlabeled chest X-ray images. Then, in stage II, the pseudo labels are used to achieve coarse transfer learning, resulting in chest X-ray-related deep features. Finally, in stage III, a downstream training has been accomplished by adapting a fine transfer learning from chest X-ray recognition (achieved by the pretext training of stage II) to COVID-19 detection.

among images of the same cluster. The  $C$  cluster labels will be assigned to the  $n'$  unlabeled images and will be presented as pseudo labels for the pretext training task and, hence, the downstream training task. The pseudo-labeled image dataset can then be defined as  $X' = \{(x^i, y^c) | c \in C\}$ .

**B. Pretext Training**

With the high availability of large-scale annotated image datasets, the chance for the different classes to be well-represented is high. Therefore, the learned in-between class boundaries are most likely to be generic enough to new samples. On the other hand, with the limited availability of annotated medical image data, especially when some classes

are suffering more compared to others in terms of the size and representation, the generalization error might increase. This is because there might be a miscalibration between the minority and majority classes. Large-scale annotated image datasets (such as ImageNet) provide effective solutions to such a challenge via transfer learning where tens of millions of parameters (of CNN architectures) are required to be trained.

A shallow-tuning mode was used during the adaptation and training of an ImageNet pretrained CNN model using the collected chest X-ray image dataset. We used the off-the-shelf CNN features of pretrained models on ImageNet (where the training is accomplished only on the final classification layer) to construct the image feature space.

Mini-batch of stochastic gradient descent (mSGD) was used to minimize the categorical cross-entropy loss function,  $E_{\text{coarse}}(\cdot)$

$$E_{\text{coarse}}(y^c, z'(x^j, W')) = - \sum_{c=1}^C y^c \ln z'(x^j, W') \quad (6)$$

where  $x^j$  is the set of self-labeled images in the training,  $y^c$  is their associated self-labels, and  $z'(x^j, W')$  is the predicted output from a softmax function, with  $W'$  the converged weight matrix associated with the ImageNet pretrained model (i.e., we used  $W'$  of ImageNet pretrained CNN model for weight initialization to achieve a coarse transfer learning).

### C. Downstream Training

A fine-tuning mode was used during the adaptation of a pretrained CNN network (e.g., ResNet18) model using feature maps from the coarse transfer learning stage. However, due to the high dimensionality associated with the images, we applied principle component analysis (PCA) to project the high-dimensional feature space into a lower dimension, where highly correlated features were ignored. This step is important for the downstream CD process in the downstream training phase to produce more homogeneous classes, reduce the memory requirements, and improve the efficiency of the framework.

Now, assume that our feature space (PCA's output) is represented by a 2-D matrix (denoted as dataset  $A$ ), and  $\mathbf{L}$  is a class category.  $A$  and  $\mathbf{L}$  can be rewritten as

$$A = \begin{bmatrix} a_{11} & a_{11} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \quad \mathbf{L} = \{l_1, l_2, \dots, l_{c'}\} \quad (7)$$

where  $n$  is the number of images,  $m$  is the number of features, and  $c'$  is the number of classes. For downstream CD, we used  $k$ -means clustering [28] to further divide each class into homogeneous subclasses (or clusters), where each pattern in the original class  $\mathbf{L}$  is assigned to a class label associated with the nearest centroid based on the squared Euclidean distance (SED)

$$\text{SED} = \sum_{j=1}^k \sum_{i=1}^n \|a_i^{(j)} - c_j\| \quad (8)$$

where centroids are denoted as  $c_j$ . Once the clustering is accomplished, each class in  $\mathbf{L}$  will further be divided into  $k$  subclasses, resulting in a new dataset (denoted as dataset  $B$ ). Accordingly, the relationship between dataset  $A$  and  $B$  can be mathematically described as

$$A = (A|\mathbf{L}) \mapsto B = (B|\mathbf{C}') \quad (9)$$

where the number of instances in  $A$  is equal to  $B$  while  $\mathbf{C}'$  encodes the new labels of the subclasses (e.g.,  $\mathbf{C}' = \{l_{11}, l_{12}, \dots, l_{1k}, l_{21}, l_{22}, \dots, l_{2k}, \dots, l_{c'k}\}$ ).

For transfer learning, we used the ResNet [29] model, which showed excellent performance with only 18 layers. Here, we consider freezing the weights of low-level layers and update weights of high-level layers. With the limited

availability of training data, stochastic gradient descent (SGD) can heavily be fluctuating the objective/loss function, and hence, overfitting can occur. To improve convergence and overcome overfitting, the mSGD was used to minimize the objective function,  $E_{\text{fine}}(\cdot)$ , with categorical cross-entropy loss

$$E_{\text{fine}}(g^{l^i}, z(o^j, \hat{W})) = - \sum_{i=1}^{c'k} g^{l^i} \ln z(o^j, \hat{W}) \quad (10)$$

where  $o^j$  is the set of input labeled images in the training,  $g^{l^i}$  is the ground-truth labels, and  $z(o^j, \hat{W})$  is the predicted output from a softmax function, with  $\hat{W}$  the converged weight matrix associated with the coarse transfer learning model.

### D. Performance Evaluation

In the downstream CD layer of 4S-DT, we divide each class within the image dataset into several subclasses, where each subclass is treated as a new independent class. In the composition phase, those subclasses are assembled back to produce the final prediction based on the original image dataset. For performance evaluation, we adopted accuracy (ACC), specificity (SP), and sensitivity (SN) metrics for multi-classes confusion matrix, and the input image can be classified into one of ( $c'$ ) nonoverlapping classes. As a consequence, the confusion matrix would be a ( $N_{c'} \times N_{c'}$ ) matrix and the metrics are defined as

$$\text{Accuracy(ACC)} = \frac{1}{c'} \sum_{i=1}^{c'} \frac{\text{TP}_i + \text{TN}_i}{\text{TP}_i + \text{TN}_i + \text{FP}_i + \text{FN}_i} \quad (11)$$

$$\text{Sensitivity(SN)} = \frac{1}{c'} \sum_{i=1}^{c'} \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \quad (12)$$

$$\text{Specificity(SP)} = \frac{1}{c'} \sum_{i=1}^{c'} \frac{\text{TN}_i}{\text{TN}_i + \text{FP}_i} \quad (13)$$

where  $c'$  is the original number of classes in the dataset, TP is the true positive in case of COVID-19 case, and TN is the true negative in case of normal or other disease, while FP and FN are the incorrect model predictions for COVID-19 and other cases. Also, the TP, TN, FP, and FN for a specific class  $i$  are defined as

$$\text{TP}_i = \sum_{i=1}^n x_{ii} \quad (14)$$

$$\text{TN}_i = \sum_{j=1}^c \sum_{k=1}^c x_{jk}, \quad j \neq i, \quad k \neq i \quad (15)$$

$$\text{FP}_i = \sum_{j=1}^c x_{ji}, \quad j \neq i \quad (16)$$

$$\text{FN}_i = \sum_{j=1}^c x_{ij}, \quad j \neq i \quad (17)$$

where  $x_{ii}$  is an element in the diagonal of the matrix. Having discussed and formalized the 4S-DT model in this section in detail, Section III-E validates the model experimentally. The model establishes the effectiveness of self-supervised super sample decomposition in detecting COVID-19 from chest X-ray images.

**Algorithm 1** Procedural Steps of the 4S-DT

1: **procedure**

- Input:
  - A large set of an unlabeled *CXR* image set.
  - *CXR* image set divided into training and testing sets.
  - Ground-truth labels.
- Output:
  - Predicted labels.

**Stage I. Self Supervised Super Sample Decomposition:**

- 2: Training an AE model to extract deep local features from the unlabeled *CXR*.
- 3: Apply an unsupervised clustering algorithm (e.g. DBSCAN) for constructing the pseudo-labels.

**Stage II. Pretext Training**

- 4: Use an ImageNet pre-trained *CNN* model (e.g. ResNet18) for the classification of the pseudo-labeled *CXR*.
- 5: Fine-tune parameters of the adopted pre-training *CNN* model.

**Stage III. Downstream Task:**

- 6: Class Decomposition:
  - Use an ImageNet pre-trained *CNN* model (e.g. AlexNet) as a feature extractor to extract discriminative features from input *CXR* images.
  - Apply *PCA* on the deep feature space for dimension reduction.
  - Use reduced feature space of the input *CXR* images to decompose original classes into a number of sub(or decomposed) classes.
- 7: Fine transfer learning:
  - Adapt the final classification layer of the pretext *CNN* model to the decomposed classes.
  - Fine-tune parameters of the pretext training *CNN* model.
- 8: Class Composition:
  - Calculate the predicted labels associated with the decomposed classes.
  - Refine the final classification using error-correction criteria.

9: **end procedure**

*E. Procedural Steps of 4S-DT Model*

Having discussed the mathematical formulations of 4S-DT model, in the following, the procedural steps of 4S-DT model is demonstrated and summarized in Algorithm 1.

IV. EXPERIMENTAL RESULTS

This section presents the datasets used in training and evaluating our 4S-DT model and discusses the experimental results.

*A. Datasets*

In this work, we used three datasets of labeled and unlabeled chest X-ray images, defined, respectively as follows.

- 1) *Unlabeled Chest X-Ray Dataset*: A large set of chest X-ray images used as an unlabeled dataset: A set

TABLE I  
DISTRIBUTION OF CLASSES IN COVID-19 DATASET-A

Type	COVID-19	SARS	Normal	Total
Training set	74	8	56	138
Augmented training set	662	69	504	1235
Testing set 1	31	3	24	58
Testing set 2	283	30	216	529

TABLE II  
DISTRIBUTION OF CLASSES IN COVID-19 DATASET-B

Class Name	Train	Test	Total
COVID-19	460	116	576
Normal	1266	317	1583
Pneumonia	3418	855	4273

of 50000 unlabeled chest X-ray images collected from three different datasets: first, 336 cases with a manifestation of tuberculosis and 326 normal cases from [30] and [31], 5863 chest X-Ray images with two categories: pneumonia and normal from [32], and a set of 43475 chest X-ray images randomly selected from a total of 112120 chest X-ray images, including 14 diseases, available from [33].

- 2) *COVID-19 Dataset-A*: An imbalanced set of labeled chest X-ray with COVID-19 cases: 80 normal cases from [31] and [30] and chest X-ray dataset from [34], which contains 105 and 11 cases of COVID-19 and SARS, respectively. We divided the dataset into two groups: 70% for training and 30% for testing. Due to the limited availability of training images, we applied different data augmentation techniques (such as flipping up/down and right/left, translation, and rotation using random five different angles) to generate more samples (see Table I).
- 3) *COVID-19 Dataset-B*: We used a public chest X-ray dataset that already divided into two sets (training and testing), and each set consists of three classes (e.g., COVID-19, Normal, and Pneumonia), see Table II. The dataset is available for download at <https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia>.

Note that chest X-ray images of dataset-A and dataset-B are progressively updated and the distributions of images in these datasets (see Tables I and II) can be considered as a snapshot at the time of submitting this article. Therefore, any attempt to compare the performance of methods on such datasets at different points in time would be misleading. Moreover, the performance of the methods reported in this article is expected to improve in the future with the growing availability of labeled images.

All the experiments in our work have been carried out on a PC with the following configuration: 3.70-GHz Intel Core i3-6100 Duo, NVIDIA Corporation with the donation of the Quadra P5000GPU, and 8.00-GB RAM.

*B. Self-Supervised Training of 4S-DT*

We trained our AE with 80 neurons in the first hidden layer and 50 neurons in the second hidden layer for the reconstruction of input unlabeled images (see Fig. 3).

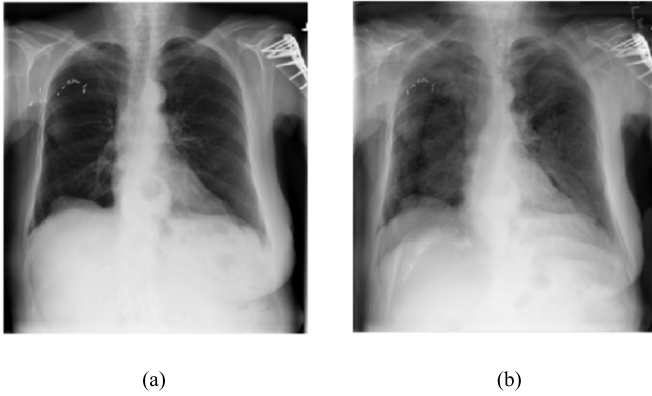


Fig. 3. Example of a reconstructed chest X-ray image by our AE. (a) Original image. (b) Reconstructed image.

The trained AE is then used to extract a set of deep features from the unlabeled chest X-ray images. The extracted features were fed into the DBSCAN clustering algorithm for constructing the clusters (and hence the pseudo labels). Since DBSCAN is sensitive to the neighborhood radius, we employed a  $k$ -nearest-neighbor ( $k$ -NN) [35] search to determine the optimal (Eps) value. As shown in Fig. 4, the optimal value for Eps was 1.861. MinPts parameter has been derived from the number of features ( $d$ ) such that  $\text{MinPts} \geq d + 1$ . Consequently, we used and tested different values for MinPts parameter such as 51, 54, and 56 resulting in 13, 6, and 4 clusters, respectively. For the coarse transfer learning, we used the ResNet18 pretrained CNN model. The classification performance, on the pseudo-labeled samples, associated with the 13, 6, and 4 clusters were 48.1%, 53.26%, and 64.37%, respectively. Therefore, we fix the number of clusters (and hence the number of pseudo labels) to be 4 in all experiments in this work.

1) *Downstream CD of 4S-DT*: We used AlexNet [36] pretrained network based on a shallow learning mode to extract discriminative features of the labeled dataset. We set a value of 0.0001 for the learning rate, except the last fully connected layer (was 0.01), the mini-batch size was 128 with the minimum of 256 epochs, 0.001 was set for the weight decay to prevent the overfitting through training the model, and 0.9 for the momentum speed. At this stage, 4096 attributes were obtained, and therefore, we used PCA to reduce the dimension of feature space. For the CD step, we used  $k$ -means clustering [28], where  $k$  has been selected to be 2, and hence, each class in  $\mathbf{L}$  has been further divided into two subclasses, resulting in a new dataset with six classes. The adoption of  $k$ -means for CD with  $k = 2$  is based on the results achieved by the DeTraC model in [17].

### C. Effectiveness of 4S-DT

We first validate the performance of 4S-DT with ResNet18 (as the backbone network) on the 58 test images (i.e., testing set 1), where the augmented training set is used for training (see Table I). Our ResNet architecture consists of residual blocks and each block has two  $3 \times 3$  Conv layers, where each layer is followed by batch normalization and a rectified linear unit (ReLU) activation function. Our ResNet architecture

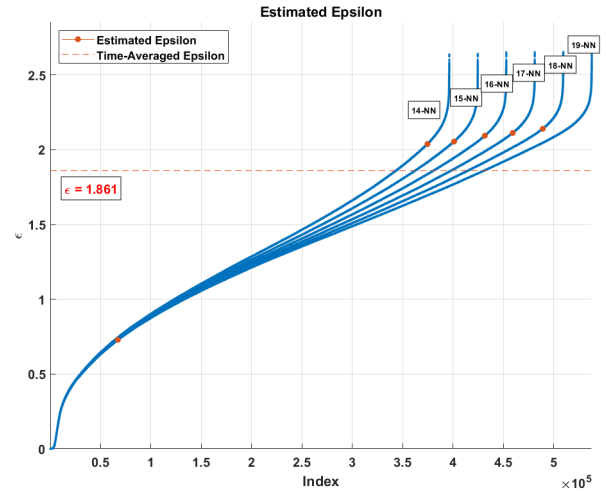


Fig. 4. Estimation of the optimal Eps.

TABLE III  
ADOPTED RESNET18 ARCHITECTURE USED IN THE FINE-TUNING STUDY IN OUR EXPERIMENTS

Layer Name	ResNet18	Output Size
Conv1	$7 \times 7$ , 64, stride (2)	$112 \times 112 \times 64$
Conv2	Layer-Res2a	$56 \times 56 \times 64$
	Layer-Res2b	
Conv3	Layer-Res3a	$28 \times 28 \times 128$
	Layer-Res3b	
Conv4	Layer-Res4a	$14 \times 14 \times 256$
	Layer-Res4b	
Conv5	Layer-Res5a	$7 \times 7 \times 512$
	Layer-Res5b	

consists of residual blocks and each block has two  $3 \times 3$  Conv layers, where each layer is followed by batch normalization and a ReLU activation function. Table III shows the adopted architecture used in our experiment.

During the training of the backbone network, the learning rate for all the CNN layers was fixed to 0.0001 except for the last fully connected layer (was 0.01) to accelerate the learning. The mini-batch size was 256 with a minimum of 200 epochs, 0.0001 was set for the weight decay to prevent the overfitting through training the model, and the momentum value was 0.95. The schedule of drop learning rate was set to 0.95 every five epochs. 4S-DT was trained in a shallow and a fine-tuning mode (see Table IV). Moreover, we also compare the performance of the proposed model without the self-supervised sample decomposition component (i.e., w/o 4S-D) and without both 4S-D and CD (w/o 4S-D + CD) on the 58 testing set. 4S-DT has achieved 100% accuracy in the detection of COVID-19 cases with 100% (95% confidence interval (CI): 96.4% and 98.7%) for sensitivity and specificity (95% CI: 94.5% and 100%), see Fig. 5(a). As shown in Fig. 5(a) and Table IV, 4S-DT shows a significant contribution in improving the transfer learning process with both the self-supervised sample decomposition and downstream CD components. Also, we applied 4S-DT based on ResNet18 pretrained network on the original classes of COVID-19 dataset with an imbalance class after eliminating the samples from the training set. As shown in Fig. 5(b), 4S-DT has achieved 96.43% accuracy (95% CI: 92.5% and 98.6%) in the detection of COVID-19 cases with sensitivity

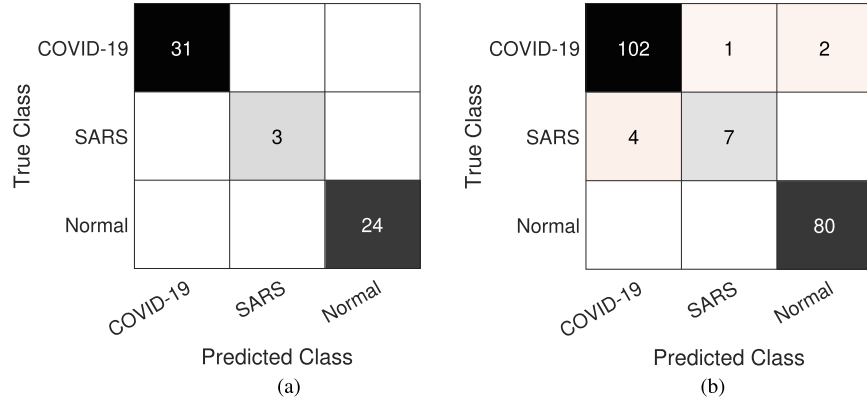


Fig. 5. Confusion matrix obtained by (a) 4S-DT on the 58 test set and (b) 4S-DT when trained on augmented images only and tested on the 196 cases of (COVID-19 dataset-A).

TABLE IV

OVERALL CLASSIFICATION PERFORMANCE OF 4S-DT MODEL OBTAINED WITHOUT BOTH 4S-D AND CD [w/o (4S-D + CD)], WITHOUT 4S-D ONLY (w/o 4S-D), AND WITH BOTH 4S-D AND CD (4S-D + CD)

Tuning Mode	COVID-19 dataset-A	w/o (4S-D + CD)				w/o 4S-D				with (4S-D + CD)			
		ACC (%)	SN (%)	SP (%)	Epochs #	ACC (%)	SN (%)	SP (%)	Epochs #	ACC (%)	SN (%)	SP (%)	Epochs #
Shallow-tuning	Testing set 1	77.35	69.23	78.50	87	91.38	85.19	92.01	65	94.83	96.77	92.59	54
Fine-tuning: Res5b	Testing set 1	79.13	69.02	71.50	92	93.10	92.50	85.9	61	94.83	96.77	92.59	46
Fine-tuning: Res5a	Testing set 1	78.42	79.0	81.67	105	93.10	92.50	85.9	61	94.83	93.67	94.06	35
Fine-tuning: Res4b	Testing set 1	78.40	79.0	81.67	84	93.10	92.50	85.90	61	96.55	92.59	95.8	58
Fine-tuning: Res4a	Testing set 1	80.87	78.2	79.81	103	93.10	92.50	85.90	61	94.83	93.55	96.30	51
Fine-tuning: Res3b	Testing set 1	80.87	78.2	79.81	46	93.10	92.50	85.90	61	94.83	94.21	96.5	38
Shallow-tuning	Testing set 2	79.61	62.87	79.20	60	92.12	64.13	94.2	61	97.48	88.64	98.01	29
Fine-tuning: Res5b	Testing set 2	79.61	62.87	79.20	96	93.84	64.18	94.06	75	97.66	93.08	98.41	42
Fine-tuning: Res5a	Testing set 2	81.41	75.40	89.62	103	93.84	64.18	94.06	83	97.23	87.33	97.73	33
Fine-tuning: Res4b	Testing set 2	81.41	75.40	85.0	63	93.96	64.33	94.20	112	96.8	86.3	97.91	32
Fine-tuning: Res4a	Testing set 2	84.57	88.52	89.62	72	94.04	64.52	94.37	82	97.99	87.15	98.10	25
Fine-tuning: Res3b	Testing set 2	81.42	85.36	89.88	57	94.34	64.16	94.25	73	97.99	87.71	97.42	32

97.1% (95% CI: 92.24% and 97.76%) and specificity 95.60 % (95% CI: 93.41% and 96.5%).

To allow for further investigation and make testing of COVID-19 detection more challenging, we applied the same data augmentation techniques (used for the training samples) to the small testing set to increase the number of testing samples. The new test sample distribution, we called testing set 2, contains 283 COVID-19 images, 30 SARS images, and 216 normal images (see Table I). Consequently, we used testing set 2 for testing and augmented training set for training (see Table I), unless otherwise mentioned, for the performance evaluation of all methods in the experiments described in the following. We validated the performance of: 1) the full version of 4S-DT with 4S-D component and 2) without 4S-D. For a fair comparison, we used the same backbone network (i.e., ResNet18) with the downstream CD component, where both versions have been trained in shallow and fine-tuning modes. As shown in Table IV, 4S-D and CD components show significant improvement in all cases. Moreover, 4S-DT demonstrates better performance, in all cases, with a less number of epochs, confirming its efficiency and robustness at the same time.

We also measured the statistical significance using the Wilcoxon signed-rank test [37] with continuity correction to establish the statistical significance of the results. At 0.05 significance level, the  $p$ -value is equal to 0.002516 [with (4S-D + CD) versus w/o (4S-D + CD)]. Consequently, the effect of the 4S-D + CD is statistically significant. Having

established the statistical significance of the accuracy boost of the proposed method with its two main components (4S-D and CD), it is important to establish the statistical significance of each of the two components. Consequently, the  $p$ -value is equal to 0.002516 (w/o (4S-D + CD) versus w/o 4S-D) to confirm that CD has a statistically significant accuracy boost on its own. In addition, the  $p$ -value is equal to 0.002421 (with (4S-D + CD) versus w/o 4S-D), establishing the statistical significance of the accuracy boost of the 4S-D component.

#### D. Classification Performance of 4S-DT

To evaluate the performance of 4S-DT on COVID-19 dataset-A and COVID-19 dataset-B, we applied different ImageNet pretrained CNN networks, such as VGG19 [38], ResNet18 [39], and GoogleNet [40]. The classification performance of 4S-DT, on COVID-19 cases, was reported in Table V. Transfer learning has been accomplished via a deep tuning scenario (with 15 epochs and SGD was the optimizer), where the same parameter settings were used for the pretrained CNN networks (as described in Table VI).

As shown in Table V, 4S-DT has achieved a high accuracy in the detection of COVID-19 cases (on the COVID-19 dataset-A) with 97.54% (95% CI: 96.22% and 98.91%), with a specificity of 97.15% (95% CI: 94.23% and 98.85%) and sensitivity of 97.88% (95% CI: 95.46% and 99.22%) on the 529 test chest X-ray images of test set 2. Moreover, 4S-DT has achieved a high accuracy of 99.8% (95% CI: 99.44% and 99.98%), with sensitivity of 99.3% (95% CI: 93.91% and



TABLE V

CLASSIFICATION PERFORMANCE OF 4S-DT AND OTHER MODELS (IN A DEEP-TUNING MODE) ON COVID-19 DATASET-A AND COVID-19 DATASET-B

Model	Backbone Network	COVID-19 dataset-A			COVID-19 dataset-B		
		ACC (%)	SN (%)	SP (%)	ACC (%)	SN (%)	SP (%)
<i>4S-DT</i>	ResNet18	97.54	97.88	97.15	99.6	96.5	99.9
<i>DeTraC</i>		95.12	97.91	91.87	97.5	95.5	98.2
ResNet18		92.5	65.01	94.3	94.74	93.3	97.84
<i>4S-DT</i>	GoogleNet	94.15	97.07	93.08	99.2	93.9	99.7
<i>DeTraC</i>		94.71	97.8	95.76	97.10	97.41	99.48
GoogleNet		93.68	92.59	91.52	94.43	88.17	91.15
<i>4S-DT</i>	VGG19	95.28	93.66	97.15	99.8	99.7	100
<i>DeTraC</i>		97.35	98.23	96.34	98.2	96.5	99.4
VGG19		94.59	91.64	93.08	95.28	92.74	94.23

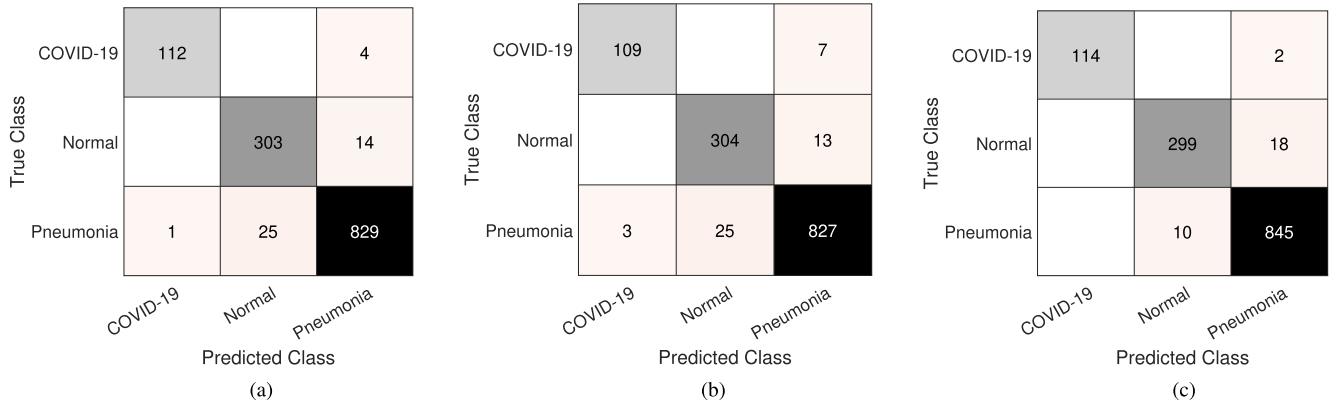


Fig. 6. Confusion matrix results of COVID-19 dataset-B obtained by 4S-DT, using different pretrained networks: (a) ResNet18, (b) GoogleNet, and (c) VGG19.

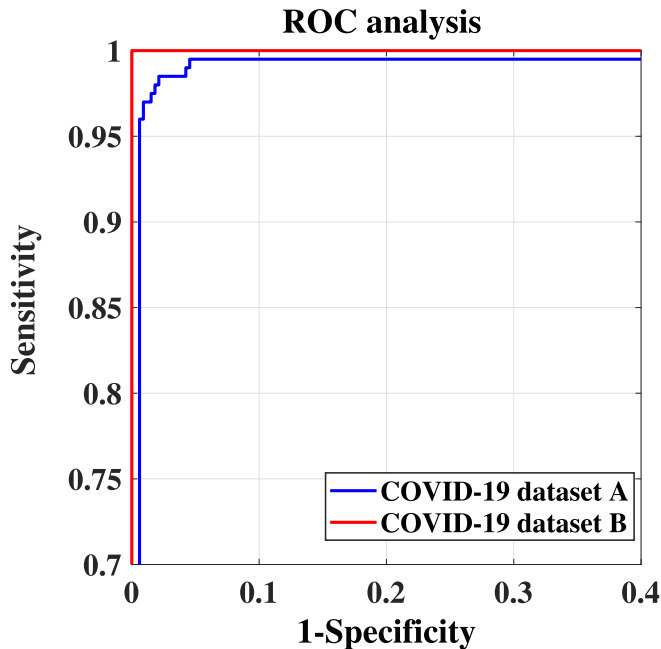


Fig. 7. ROC curve obtained during the training of 4S-DT on the COVID-19 dataset-A and COVID-19 dataset-B.

99.79%) and specificity of 100% (95% CI: 99.69% and 100%) in the detection of COVID-19 cases of COVID-19 dataset-B. Moreover, Fig. 6 shows the confusion matrix obtained by each pretrained network for each class in the COVID-19 dataset-B.

Finally, Fig. 7 shows the area under the curve [AUC; receiver operating characteristic (ROC)] between the true

TABLE VI

PARAMETERS SETTINGS FOR EACH PRETRAINED MODEL USED FOR TRAINING 4S-DT MODEL ON COVID-19 DATASET-A AND COVID-19 DATASET-B

Pre-trained model	Learning rate	MB-Size	Weight decay	Learning rate-decay
ResNet18	0.001	256	0.0001	0.9 every 3 epochs
GoogleNet	0.0001	128	0.001	0.95 every 2 epochs
VGG19	0.01	32	0.0001	0.9 every 2 epochs

positive rate and false positive rate obtained by 4S-DT with ResNet18 pretrained model on COVID-19 dataset-A (in blue) with an AUC value of 99.58% (95% CI: 99.01% and 99.95%) and with VGG19 pretrained model on COVID-19 dataset-B (in red) with an AUC value of 100% CI: 99.33% and 100%), to confirm the robustness of 4S-DT behaviors during the training process.

## V. CONCLUSION AND DISCUSSION

Diagnosis of COVID-19 disease can be accomplished by genetic and imaging tests. Chest X-ray imaging test provides a promising fast detection of COVID-19 cases and consequently can contribute to controlling the spread of the virus. In chest X-ray diagnosis, data irregularities still remain a challenging problem, especially with the limited availability of confirmed samples with some diseases such as COVID-19 and SARS, which usually results in miscalibration between the different classes in the dataset. Consequently, COVID-19 detection from chest X-ray images presents a challenging problem due to the irregularities and the limited availability of annotated cases.

In transfer learning, an ablation study [41] has been previously conducted to demonstrate the generality/specificity of

CNN's features by quantifying their transferability. It demonstrated that features' transferability was negatively affected by two experimentally validated observations: 1) splitting the network at two adjacent layers, where coadapted neurons exist, makes the optimization process difficult, and 2) when the target domain (e.g., COVID-19 detection) is significantly dissimilar to the source domain (e.g., large-scale image classification), the number of layers with generalized features can be small. Fine-tuning allows features to be adapted to the domain-specific task with a significant performance (based on how much training samples are available) by gradually training layers of the CNN model. However, due to the limited availability of training COVID-19 cases coupled by irregularity problems with other cases such as SARS (see Table I), traditional fine-tuning mode fails to provide domain-specific features (as shown in Table IV). Consequently, the 4S-DT model has adopted a deep fine-tuning learning mode to overcome the problem of observation 1) by allowing for the coadaptation among CNN's layers during the optimization process, resulting in more specialized features and hence better classification, as shown in Tables IV and V. The proposed 4S-DT model has the ability to overcome the problem that has been pointed out in observation 2) by adopting a self-supervised sample decomposition (4S-D) approach to generate pseudo labels for the classification of unlabeled chest X-ray images as a pretext learning task. The 4S-D leads to a better performance (see Table IV) by allowing for a coarse feature transformation from a generic task of ImageNet recognition to a domain-specific task of chest X-ray image recognition, taking advantage of the huge availability of chest X-ray images. 4S-DT has also the ability to deal with data irregularities by a CD utilized in its downstream learning component. Moreover, with the CD component of our 4S-DT, the features' transferability has been improved using a coarse-to-fine approach that has been designed to cope with more challenging problems such as data irregularity (e.g., class overlap in COVID-19 and SARS cases) resulting in higher performance, as shown in Tables IV and V.

To conclude, the proposed 4S-DT model has been designed to cope with data scarcity through its self-supervised sample decomposition approach. 4S-DT has also the ability to deal with data irregularities by a CD adapted in its downstream learning component. 4S-DT has demonstrated its effectiveness and efficiency in coping with the detection of COVID-19 cases in a dataset with irregularities in its distribution. In this work, we used 50000 unlabeled chest X-ray images for the development of our self-supervised sample decomposition approach to perform transfer learning with an application to COVID-19 detection. We achieved an accuracy of 97.54% with a specificity of 97.15% and sensitivity of 97.88% on 529 test chest X-ray images (of COVID-19 dataset-A), i.e., testing set 2, with 283 COVID-19 samples. We also achieved a high accuracy of 99.8% with a specificity of 100% and sensitivity of 99.7% in the detection of COVID-19 cases of COVID-19 dataset-B.

With the continuous collection of data, we aim in the future to extend the development and validation of 4S-DT with multimodality datasets, including clinical records, and developing an explainability component to increase the trustworthiness

and usability of 4S-DT. Explainable surrogate models could be used to improve the usability and trustworthiness of 4S-DT as a diagnostic tool. This is by adapting a machine teaching mechanism, aiming at guiding the attention of 4S-DT to the diagnostic relevant features during the training process. Medical professionals can act as the main source of feedback given to 4S-DT to run the machine teaching session. This machine teaching mechanism will not only help in better understanding why 4S-DT makes a particular decision, but more importantly, it will correct and improve its learning by focusing on the usage of diagnostic relevant features that are clinically valid. As another future development, we aim to further support COVID-19 diagnosis from CXR/CT images and provide a fast and easy-to-use solution using ensemble learning and network pruning. Ensemble learning is a promising and experimentally validated technique that can exploit diversity using multiple models. On the other hand, network pruning has been proposed with the aim of removing redundant structures of the neural network for accelerating its running-time inference speed. Removing or pruning these redundant parameters will result in a network with lower complexity. One can employ network pruning on 4S-DT to obtain a small and efficient classification model to consume much less energy and consequently provide an easy-to-deploy solution.

## REFERENCES

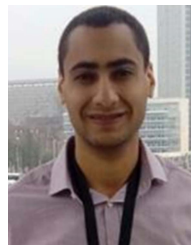
- [1] H. Shi *et al.*, "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: A descriptive study," *Lancet Infectious Diseases*, vol. 20, no. 4, pp. 425–434, Apr. 2020.
- [2] E. Dandil, M. Cakiroglu, Z. Eksi, M. Ozkan, O. K. Kurt, and A. Canan, "Artificial neural network-based classification system for lung nodules on computed tomography scans," in *Proc. 6th Int. Conf. Soft Comput. Pattern Recognit. (SoCPar)*, Aug. 2014, pp. 382–386.
- [3] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 202–209, Jan. 2014.
- [4] T. Manikandan and N. Bharathi, "Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier," *J. Med. Syst.*, vol. 40, no. 7, p. 181, Jul. 2016.
- [5] P. B. Sangamithraa and S. Govindaraju, "Lung tumour detection and classification using EK-mean clustering," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, Mar. 2016, pp. 2201–2206.
- [6] E. Pesce, S. J. Withey, P.-P. Ypsilantis, R. Bakewell, V. Goh, and G. Montana, "Learning to detect chest radiographs containing pulmonary lesions using visual attention networks," *Med. Image Anal.*, vol. 53, pp. 26–38, Apr. 2019.
- [7] Y. Xie, J. Zhang, and Y. Xia, "Semi-supervised adversarial model for benign–malignant lung nodule classification on chest CT," *Med. Image Anal.*, vol. 57, pp. 237–248, Oct. 2019.
- [8] A. Abbas and M. M. Abdelsamea, "Learning transformations for automated classification of manifestation of tuberculosis using convolutional neural network," in *Proc. 13th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2018, pp. 122–126.
- [9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
- [10] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [11] Q. Li, W. Cai, X. Wang, Y. Zhou, D. Dagan Feng, and M. Chen, "Medical image classification with convolutional neural network," in *Proc. 13th Int. Conf. Control Autom. Robot. Vis. (ICARCV)*, Dec. 2014, pp. 844–848.
- [12] R. Joyseere, S. Otálora, H. Müller, and A. Depeursinge, "Fusing learned representations from riesz filters and deep CNN for lung tissue classification," *Med. Image Anal.*, vol. 56, pp. 172–183, Aug. 2019.
- [13] M. Gao *et al.*, "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks," *Comput. Methods Biomech. Biomed. Eng., Imag. Vis.*, vol. 6, no. 1, pp. 1–6, Jan. 2018.

- [14] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, May 4, 2020, doi: [10.1109/TPAMI.2020.2992393](https://doi.org/10.1109/TPAMI.2020.2992393).
- [15] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 132–149.
- [16] L. Rokach, O. Maimon, and O. Arad, "Improving supervised learning by sample decomposition," *Int. J. Comput. Intell. Appl.*, vol. 5, no. 1, pp. 37–53, Mar. 2005.
- [17] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "DeTrac: Transfer learning of class decomposed medical images in convolutional neural networks," *IEEE Access*, vol. 8, pp. 74901–74913, 2020.
- [18] *Coronavirus Disease (COVID-2019) Situation Reports*, World Health Organization, Geneva, Switzerland, 2020.
- [19] F. Shi *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 4–15, 2021.
- [20] D. Dong *et al.*, "The role of imaging in the detection and management of COVID-19: A review," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 16–29, 2021.
- [21] S. A. Harmon *et al.*, "Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets," *Nature Commun.*, vol. 11, no. 1, pp. 1–7, Dec. 2020.
- [22] Y. Song *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, doi: [10.1109/TCBB.2021.3065361](https://doi.org/10.1109/TCBB.2021.3065361).
- [23] L. Wang, Z. Q. Lin, and A. Wong, "COVID-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Dec. 2020.
- [24] X. Wang *et al.*, "A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2615–2625, Aug. 2020.
- [25] I. D. Apostolopoulos and T. A. Mpesiana, "COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, Jun. 2020.
- [26] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020.
- [27] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "Self-supervised learning for medical image analysis using image context restoration," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101539.
- [28] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [30] S. Jaeger *et al.*, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 233–245, Feb. 2014.
- [31] S. Candemir *et al.*, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 577–590, Feb. 2014.
- [32] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [33] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2097–2106.
- [34] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 image data collection," 2020, *arXiv:2003.11597*. [Online]. Available: <http://arxiv.org/abs/2003.11597>
- [35] S. A. Dudani, "The distance-weighted  $k$ -nearest-neighbor rule," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, no. 4, pp. 325–327, Apr. 1976.
- [36] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [37] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in Statistics*. New York, NY, USA: Springer, 1992, pp. 196–202.
- [38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [39] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1–7.
- [40] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
- [41] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" 2014, *arXiv:1411.1792*. [Online]. Available: <http://arxiv.org/abs/1411.1792>



**Asmaa Abbas** was born in Asyut, Egypt. She received the B.Sc. and M.Sc. degrees in computer science from the Department of Mathematics, University of Assiut, Asyut, Egypt, in 2004 and 2020, respectively.

She is currently a Researcher with the Department of Mathematics, University of Assiut. Her research interests include deep learning, medical image analysis, and data mining.



**Mohammed M. Abdelsamea** received the Ph.D. degree (with Doctor Europaeus) in computer science and engineering from the IMT Institute for Advanced Studies, Lucca, Italy, in 2015.

Before moving to U.K., he worked as a Lecturer of computer science at the University of Assiut, Asyut, Egypt. He was a Visiting Researcher at Robert Gordon University, Aberdeen, U.K. He is currently a Senior Lecturer in data and information science at the School of Computing and Digital Technology, Birmingham City University (BCU), Birmingham, U.K.

Before joining BCU, he worked at the School of Computer Science, University of Nottingham, Nottingham, U.K., the Centre for Mechanochemical Cell Biology, Warwick University, Coventry, U.K., and Nottingham Molecular Pathology Node (NMPN) and the Division of Cancer and Stem Cells, Nottingham Medical School, Nottingham, as a Research Fellow. In 2016, he was a Marie Curie Research Fellow at the School of Computer Science, University of Nottingham. His main research interests are in computer vision, including image processing, deep learning, data mining and machine learning, pattern recognition, and image analysis.



**Mohamed Medhat Gaber** received the Ph.D. degree from Monash University, Clayton, VIC, Australia, in 2006.

He is currently a Professor of data analytics at the School of Computing and Digital Technology, Birmingham City University, Birmingham, U.K. He then held appointments at the University of Sydney, Sydney, NSW, Australia, CSIRO, Canberra, ACT, Australia, and Monash University. Prior to joining Birmingham City University, he worked at Robert Gordon University, Aberdeen, U.K., as a

Reader in computer science and the University of Portsmouth, Portsmouth, U.K., as a Senior Lecturer in computer science. He has published over 200 articles, coauthored three monograph-style books, and edited/co-edited six books on data mining and knowledge discovery. His work has attracted well over six thousand citations, with an H-index of 40.

Dr. Gaber is recognized as a fellow of the British Higher Education Academy (HEA). He is also a member of the International Panel of Expert Advisers for the Australasian Data Mining Conferences. He has served in the program committee of major conferences related to data mining. He has co-chaired numerous scientific events on various data mining topics. In 2007, he received the CSIRO Teamwork Award.