Potential Diagnosis of COVID-19 from Chest X-Ray and CT Findings Using Semi-Supervised Learning

Pracheta Sahoo (pracheta.sahoo@utdallas.edu)
The University of Texas at Dallas https://orcid.org/0000-0002-7999-4882
Indranil Roy
Northwestern University
Randeep Ahlawat
The University of Texas at Dallas
Saquib Irtiza
The University of Texas at Dallas
Latifur Khan
The University of Texas at Dallas

Research Article

Keywords: Deep Neural Networks, Semi-supervised learning, Diagnosis, COVID-19, medical imaging

Posted Date: May 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-319686/v1

Noname manuscript No.
(will be inserted by the editor)

Potential Diagnosis of COVID-19 from Chest X-Ray and CT Findings Using Semi-Supervised Learning

Pracheta Sahoo [1], Indranil Roy[2], Randeep Ahlawat[1], Saquib Irtiza[1], and Latifur Khan[1]

[1]Department of Computer Science, The University of Texas at Dallas, Richardson, Texas 75080, United States
[2]Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3113, United States

Received: date / Accepted: date

Abstract COVID-19 is an infectious disease, which has adversely affected public health and the economy across the world. On account of the highly infectious nature of the disease, rapid automated diagnosis of COVID-19 is urgently needed. A few recent findings suggest that chest X-rays and CT scans can be used by machine learning for the diagnosis of COVID-19. Herein, we employed Semi-Supervised Learning (SSL) approaches to detect COVID-19 cases accurately by analyzing digital chest X-rays and CT scans. On a relatively small COVID-19 radiography dataset, which contains only 219 COVID-19 positive images, 1341 normal and 1345 viral pneumonia images, our algorithm, COVIDCon, which takes advantage of data augmentation, consistency regularization, and multicontrastive learning, attains 97.07% average class prediction accuracy, with 1000 labeled images, which is 7.65% better than the next best SSL method, Virtual Adversarial Training. COVIDCon performs even better on a larger COVID-19 CT Scan dataset that contains 82767 images. It achieved an excellent accuracy of 99.13%, at 20000 labels, which is 6.45% better than the next best Pseudo-Labeling approach. COVIDCon outperforms other state-of-the-art algorithms at every label that we have investigated. These results demonstrate COVIDCon as the benchmark SSL algorithm for potential diagnosis of COVID-19 from chest X-rays and CT-Scans. Furthermore, COVIDCon performs exceptionally well in identifying COVID-19 positive cases from a completely unseen repository with a confirmed COVID-19 case history. COVIDCon, may provide a fast, accurate, and reliable method for screening COVID-19 patients.

Keywords Deep Neural Networks · Semi-supervised learning · Diagnosis · COVID-19 · medical imaging

Pracheta Sahoo : pracheta.sahoo@utdallas.edu

Indranil Roy : indranil@northwestern.edu

Randeep Ahlawat : randeep.ahlawat@utdallas.edu

Saquib Irtiza : saquib.irtiza@utdallas.edu

Latifur Khan : lkhan@utdallas.edu

Address(es) of author(s) should be given

1 Introduction

Novel Coronavirus disease 2019 (COVID-19) [1] is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). On account of the highly infectious nature of the disease, it has spread worldwide in a short period of time, adversely affecting public health and global economy. In order to reduce the rapid spread of COVID-19, fast and accurate diagnosis of patients is of primary interest.

The definitive test for COVID-19 is the Real-time Reverse Transcriptase Polymerase chain reaction (RT-PCR) [26] test, which has a sensitivity [2] between 51 and 94% and takes around 24 hours to obtain the results. The long waiting time may increase the chances of spreading the disease to others. CT scans and standard chest X-rays [65]—on the other hand, may save time for the diagnosis of COVID-19. Cases have been reported [23,57] where RT-PCR gave false-negative results, but radiographic techniques could detect COVID-19. Chest X-rays and chest CT scans are, therefore, being used to get more information and are being considered [31] a screening tool alongside the RT-PCR test [10]. In fact, a study conducted by Ai et al. [7] shows that CT Scan has better sensitivity compared to RT-

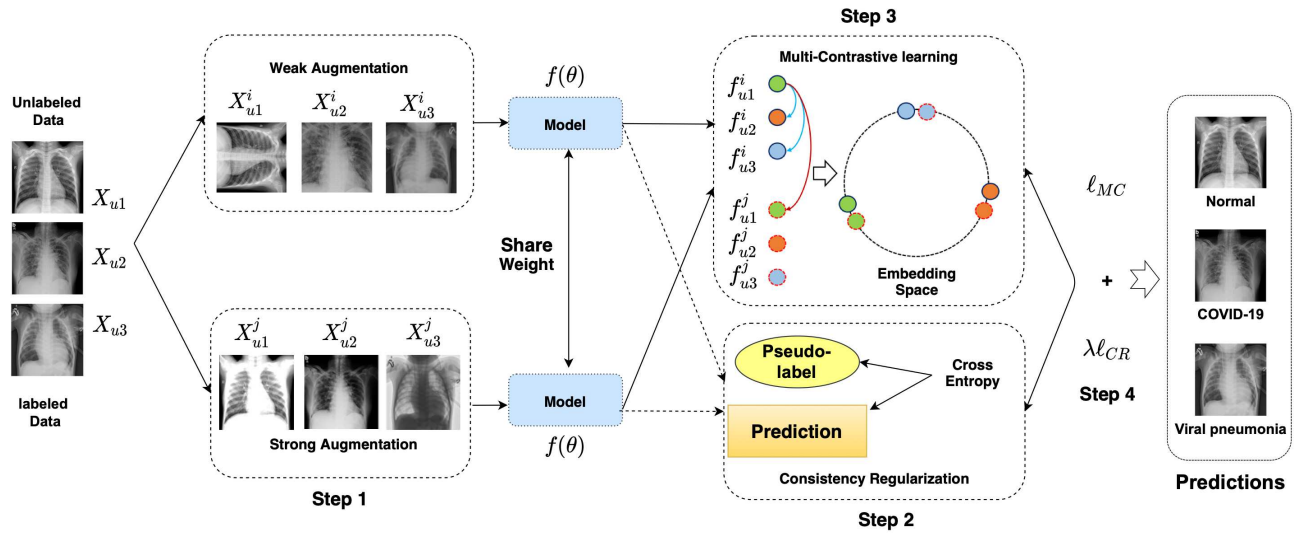

Fig. 1: The framework of COVIDCon. Different input chest X-rays with corresponding (weak and strong) augmented transformations are projected into embedding features. For the consistency regularization, the model prediction works as pseudo-label, ℓ_{CR} aims to make the output from strong augmentation match of the pseudo-label. The model is trained under the combination of ℓ_{CR} and ℓ_{MC}. The colored solid lined circles represent strong augmented data points, the dotted circles represent weak augmented data points, and same color represents same class

PCR in detecting COVID-19 and can be used for the diagnosis of the disease. CT-Scans can also provide information about the severity of COVID-19, which cannot be obtained by standard RT-PCR testing.

Chest CT scans and X-rays of COVID-19 patients often display [61] abnormal patterns, for example, bilateral, multilobar ground glass opacities (GGO) with a peripheral or posterior distribution, mainly in the lower lobes and on occasion in the middle lobe. The visual differences in chest CT scan and X-ray images between COVID-19, pneumonia, and normal patients are subtle and require expert radiologists, which causes a bottleneck as their number is limited. In this context, artificial intelligence/machine learning can help to identify positive cases, find abnormalities and provide aid to medical personnel.

For the past few years, deep learning models have demonstrated their potentials to be useful to radiologists and medical imaging experts for various disease detection and classification tasks such as skin cancer [17], brain disease [32], arrhythmia [58], breast cancer [35], pneumonia [43], etc. A few state-of-the-art supervised models have even achieved performance equivalent to experienced radiologists [43]. Recently, deep learning-based approaches [51,62,63,15,39,59,40,52,20,41] have also been applied successfully in diagnosis of COVID-19. The extraordinary performance of such deep learning models based on supervised learning [14], however, requires large amounts of labeled data. A situation that may be challenging in medical image analysis, where data collection and annotation are time-intensive tasks and increase the workload on the radiologists.

Semi-supervised learning (SSL) [64] curbs the disadvantages of supervised learning methods by utilizing a small amount of labeled and a larger amount of unlabeled data. Data augmentation, together with transfer learning, can produce powerful, more robust models that require less training time. In medical imaging, large unlabeled datasets are readily available along with smaller high-quality labeled datasets. Therefore, SSL methods can be an excellent option for automated medical image diagnosis.

Up until now, many SSL models, such as Pseudo-labelling, [30] VAT, [36] π-Model, [28] Mean Teacher, [47] MixMatch, [13] and FixMatch [46] have been implemented successfully for analyzing two-dimensional image data. Recently, we also reported a semi-supervised image classification algorithm, MultiCon [44], which uses multi-contrastive learning alongside consistency regularization to outperform other semi-supervised approaches in image classification.

Consistency regularization is a semi-supervised learning component that adds a supplementary loss function to a network such that the predictions of the network remain consistent even after the inputs are per-

turbed. This is done by making use of the unlabeled data to find latent features for the additional loss function. Phillip et. al. [11] developed the first variant of this component which was later made popular by Samuli et. al. [29] and Mehdi et. al. [45] . Some of the recent variations include replacing parts of the loss function [37], using this component in larger SSL pipeline [13,27] and replacing ℓ^2 loss with cross-entropy [55] etc.

Contrastive Learning is the process that allows models to learn high level features about the dataset by finding how similar or different a pair of data points are from each other. This is a unsupervised learning techniques that is implemented before any segmentation or classification task. Despite being a relatively new area of study, it has been the basis for many excellent works such as contrastive predictive coding [21,38], representation learning using Deep InfoMax (DIM) [22,49,8] or momentum contast [19], learning invariances using Augmented Multiscale DIM [12] or Contrastive Multi-View Coding [48] etc.

Previous deep learning approaches [51,62,63,15,52, 20,41,53,9,16,54] have shown promises in identifying COVID-19 cases from chest radiography images. But as the performance of these methods depend on large labeled dataset, researchers have also tried to explore the performance of semi-supervised learning methods in the diagnosis of COVID-19 cases. [24,18,25,33,6]. Jun et. al. [33] used active contour regularization on a region scalable fitting (RSF) model to further tune the pseudo-labels of the unlabeled CT images. Mohammed et. al. [6] proposed a dual path few shot semi-supervised segmentation approach that uses only a few labeled CT images to accurately segment COVID-19 infection. A semi-supervised shallow framework was proposed by Debanjan et. al. [25] that diagnosed segmentation of CT images produced by parallel quantum-inspired self-supervised Network (PQIS-Net). Shahin et. al. [24] used autoencoder based semi supervised approach to first extract regions of interest from chest X-Ray images which are then fed to a deep architecture to classify them.

Herein, we use the state-of-the-art SSL algorithms and our algorithm, MultiCon, [44] for the classification of grey-scale X-ray and CT scan images as COVID-19, pneumonia, or normal cases from a publicly available COVID-19 radiography dataset [42] and a COVID-19 CT Scan dataset [62]. In the present work, we refer MultiCon as COVIDCon for better understanding to the readers. The loss function of COVIDCon is constituted using two loss terms—consistency regularization and multi-contrastive learning 1. The consistency regularization component keeps the probability distributions of weakly augmented and strongly augmented dataset similar to each other. The multi-contrastive learning part keeps the data points of the same class together and instances of different classes further apart. The notion behind using these two components together is the complementary nature of these components. Contrastive learning maximizes the mutual information between the differently augmented views of the data, whereas consistency regularization minimizes the cross entropy between class predictions of augmented data. The model learns about the different features of COVID-19, viral pneumonia and normal cases from the supervised part of the data in consistency regularization. Without using this component, the algorithm would have been an unsupervised algorithm having vague idea about the subtle feature differences about the individual classes—COVID-19, viral pneumonia and normal. COVIDCon achieves 97.07% accuracy on the COVID-19 radiography dataset using just 1000 labeled data and outperforms all other state-of-the-art SSL algorithms by a large margin. In the case of CT scan dataset, COVIDCon achieves an impressive accuracy of 99.13%, at 20000 labels, whereas the next best Mean Teacher model turns out to be 6.45% less accurate. Overall, COVIDCon provides exemplary benchmark results on the semi-supervised classification scenario for COVID-19 diagnosis.

2 Methodology

This section presents the methods and materials used in this study. Sections 2.1 and 2.2 refer to the descriptions of COVID-19 Radiography and CT Scan datasets used for training and testing the SSL methods.

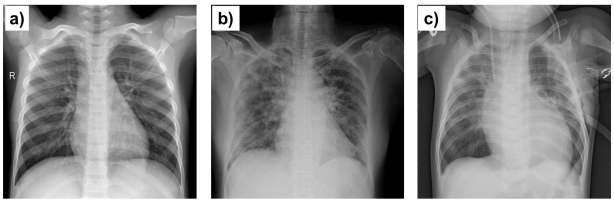

Fig. 2: Sample chest X-rays taken from the COVID-19 Radiography dataset. (a) Normal case, (b) COVID-19 case showing bilateral ground-glass opacities with prominent peripheral, perihilar and basal distribution within a multilobar involvement, and (c) viral pneumonia case with visible left basilar opacity.

Approach	Artificial augmentation	Artificial post-processing	Contrastive learning
Π-model [28]	Weak	None	None
Pseudo-label [30]	Weak	None	None
Mean Teacher [47]	Weak	None	None
VAT [36]	None	Adversarial	None
ICT [50]	Weak	Sharpening	None
MixMatch [13]	Multiple Weak	Sharpening	None
FixMatch [46]	Weak+Strong	Pseudo-labeling	None
COVIDCon [44]	Weak+Strong	Pseudo-labeling	Multi Contrastive Loss

Table 1: Comparison between different SSL algorithms. Unlike other SSL algorithms, MultiCon works with all three components—(i) artificial augmentation, (ii) artificial post-processing, and (iii) contrastive learning.

2.1 COVID-19 Radiography Dataset

The COVID-19 radiography dataset [42] has been developed by a team of researchers from Qatar University, the University of Dhaka and their collaborators from Pakistan and Malaysia with the help of medical doctors. The dataset consists mostly of posterior-to-anterior images of chest X-rays with 1024 x 1024 resolution from COVID-19 positive cases, viral pneumonia cases, as well as normal cases. COVID-19 Radiography dataset is available publicly, and it gets updated continuously with new X-ray images. As of 14th June 2020, the dataset contains a total of 2905 unique images, which are distributed over 3 imbalanced classes, namely COVID-19, Normal and Viral Pneumonia. The COVID-19 class contains 219 unique X-ray images, while viral pneumonia and normal classes contain 1341, 1345 images, respectively. A few sample chest X-rays of the COVID-19, viral pneumonia, and normal cases are listed in Figure 2.

2.2 COVID-19 CT Scan Dataset

The China Consortium of Chest CT Image Investigation (CC-CCII) compiled a dataset [62] of CT images, with 512 x 512 resolution, from cohorts of their patients. The dataset consists of 617,775 images from 4154 patients divided into three classes of novel coronavirus pneumonia (NCP), common pneumonia (CP), and normal cases. The images in the common pneumonia class are a mixture of some of the common classes of pneumonia in China, such as viral pneumonia and bacterial pneumonia. The version used for this experiment (version 2.2) is publicly available for download and was released on 24th April 2020. Image slices containing lesions were only used to train the model, resulting in 21777 images in the coronavirus pneumonia class and 36894 images in the common pneumonia class. In order to maintain a balance among the data in the three classes, a subset of 24096 images from the normal class was chosen randomly. A few sample CT scans of the COVID-19, common pneumonia and normal cases are listed in Figure 3.

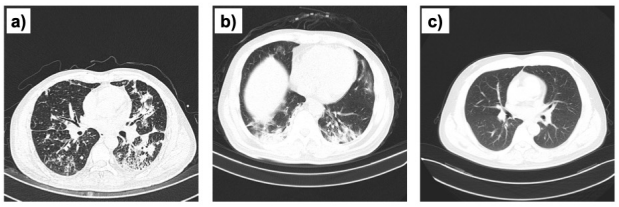

Fig. 3: CT images taken from COVID-19 CT Scan dataset. Typical examples showing (a) common pneumonia (CP), (b) COVID-19 (NCP), and (c) normal CT scan images.

2.3 Our algorithm:COVIDCon

For the detection of COVID-19 from the digital chest X-Rays and CT scans, we have used our algorithm, COVIDCon, which is based on our previously developed algorithm [44]. We have compared the performance of COVIDCon with other state-of-the-art SSL methods including MixMatch [13], Virtual Adversarial Training (VAT) [36], Pseudo-labeling [30], Mean Teacher [47], Π model [28], Interpolation Consistency Training (ICT) [50] in (Table 1).

Though we applied MultiCon [44] in the field of drug classification, the use of contrastive learning is still relatively unexplored in the field of medical imaging. We observed that, similar to drug classification,

x-ray and CT image classification, especially in the case of COVID-19, is mostly dependent on discerning subtle features. MultiCon's impressive performance in drug classification inspired us to apply it to the medical imaging domain as COVIDCon.

COVIDCon is a Semi-supervised Learning (SSL) technique that combines consistency regularization and multi-contrastive learning approach to learn a feature embedding, where the augmented views of the same data maintain a minimum distance between them. The main inspiration behind such a technique is to allow deep neural network models to train with the minimal number of labeled data and a large unlabeled dataset. This is especially advantageous since unlabeled data is cheap and easy to obtain, and it also relieves experts from manually labeling them.The proposed algorithm is a three-step process which comprises of first augmenting the data, followed by pseudo-labelling the unlabeled data and then finally obtaining a feature representation using a contrastive prediction task. Unlike other methods, COVIDCon combines the loss terms from both consistency regularization and multi-contrastive learning so that the feature embedding can satisfy the positive concentrated and negative separated properties.

Data augmentations used in this technique are of two types: weak augmentation ($\mathcal{W}_a$) and strong augmentation $\mathcal{S}_a(x_u)$. Weak augmentation corresponds to a flip-shift strategy, which randomly flips images with a probability of 50% and translates them vertically and horizontally up to 12.5%. Strong augmentation strategies are based on RandAugment. We color inversion, contrast adjustment, translation, etc transformsations and RandAugment selects a subset of these and randomly assigns how intense they will be within a mini-batch.

A side effect of using strong augmentations, is that the image might differ too much from the original image. For this reason, the weakly augmented version of each unlabeled image is assigned a pseudo label and we try to assign this pseudo label to the corresponding strongly augmented version using Cross Entropy loss. Consistency Regularization is a combination of supervised Cross Entropy loss of labeled images and unsupervised Cross Entropy with pseudo labels and unlabeled data. This ensures even after the perturbations caused by strong augmentation, the labeled and unlabeled follow a similar distribution.

After assigning a pseudo label to strongly augmented unlabeled images, we select those images in the mini-batch whose pseudo labels are predicted with a probability greater than a threshold. Of the selected images, we extract their feature vectors and reduce distances between vectors of images with same class while increasing distances between vectors of images with different classes. This strategy results in clustering of images with same classes and results in a more generalized model.

$$\ell_{CR} = \frac{1}{M}\sum_{u=1}^{M} [\mathbb{1}(\max(q_u) \geq \tau)]\, L_{CE}(\hat{y}_u, P(y|\mathcal{S}_a(x_u)) \\ \hat{y}_u = \arg\max(P(y|\mathcal{W}_a(x_u)) \tag{1}$$

In the case of multi-contrastive learning, the augmented dataset is clustered with the goal of grouping data from the same class and pushing the data from different classes further away from each other. Given the similarity measurement $\mathcal{S}(x_i, x_j) = \mathcal{S}(f(x_i;\theta), f(x_j;\theta))$ of the sample pair (x_i, x_j), the multi-contrastive loss ℓ_{MC} is:

$$\ell_{MC} = \sum_{i=1}^{m}\Big\{\frac{1}{m^+}\sum_{y_i=y_j} g(\lambda_1[\omega - S(x_i,x_j)]) \\ +\frac{1}{m^-}\sum_{y_i\neq y_j} g(\lambda_2[S(x_i,x_j)-\omega])\Big\} \\ s.t \quad g(x) = \log(1+\exp(x)) \tag{2}$$

where ω is the hyperparameter in the Binomial deviance[60]. Finally, the consistency regularization component and multi-contrastive loss sums together and gives the overall loss term ℓ: $\ell = \ell_{CR} + \lambda\ell_{MC}$ Where λ is a fixed scalar hyperparameter denoting the relative weight of different objective function.

COVIDCon is a specialized algorithm developed for two-dimensional image analysis. It is especially suitable for identifying subtle visual differences in images and therefore has the potential in clinical diagnosis, for example, detection of COVID-19 by X-ray image analysis.

2.4 Implementation Details

We utilized Pytorch [3] to implement COVIDCon on the radiography and CT Scan datasets. The X-ray and CT Scan images were resized into 84×84 and pretrained ResNeXt-101 [56] was used as the deep learning model. The ResNeXt model was pretrained in a weakly-supervised fashion on 940 million public images with 1.5K hashtags matching with 1000 ImageNet1K synsets, followed by fine-tuning on ImageNet1K dataset. [4] For the implementation of COVIDCon, we randomly initialized and added 3 fully connected layers to ResNeXt. COVIDCon was trained for 10 epochs, the learning rate was set at 0.01, the size

of the mini-batch was 128, and the weight decay was 0.001.

We used an identical set of hyperparameters ($\lambda_1 = 2$, $\lambda_2 = 40$, $\omega = 1$ and $\eta = 0.1$, $\lambda = 0.8$ during the training step) for all experiments. For training and testing, we split the dataset into 80:20 ratio, respectively. Overall, we repeat each experiment five times independently and report the average result.

2.5 Baseline Methods

We consider recent state-of-the-art methods, such as Π-Model [28], Mean Teacher [47], ICT [50], Virtual Adversarial Training [36], Pseudo-Label [30], MixMatch [13], and FixMatch, [46] as the baseline methods. We use the official codes from the original papers and implement the same network architecture ResNeXt-101, [56] training epoch, and initialized the hyperparameters in each method based on the author's recommendations. We retuned the hyperparameters for each baseline method to ensure a fair comparison with COVIDCon.

3 Results

3.1 Results on COVID-19 Radiography Dataset

We first employed COVIDCon on the COVID-19 Radiography dataset and obtained the average class prediction accuracies (Table 2) of the model with different numbers of labeled data. COVIDCon outperformed other state-of-the-art SSL methods at all labeled data (Figure 4). With just 50 labels, COVIDCon achieved an average accuracy of 92.71%, which is slightly better than the next best state-of-the-art FixMatch model. The performance of FixMatch, however, decreased with more labeled data on account of overfitting, an issue, which was not observed in the case of COVIDCon. COVIDCon reached the highest class prediction accuracy of 97.07% on average with 1000 labeled data, which is 7.65% better than the next best VAT method.

In order to better understand the performance of COVIDCon, we analyzed (Figure 5) the confusion matrix. For 98.73% of cases, COVIDCon correctly predicted the true class for COVID-19 X-ray images with 1000 labeled data. The prediction accuracies were 96.1% and 97.6% for normal and viral pneumonia cases, respectively. Given the small and imbalanced nature of the dataset, COVIDCon performed extremely well in predicting all the three classes. It successfully differentiated all COVID-19 cases from viral pneumonia cases and confused only 1.27% COVID-19 cases as normal. Given the similarities of the image features in some of

Table 2: Comparison of accuracy achieved by COVIDCon and other state-of-the-art methods on COVID-19 Radiography dataset. COVIDCon outperforms other methods at every label.

Methods/Labels	50	200	1000
Π-model	79.82±0.61	70.04±0.31	73.24±0.15
Mean Teacher	67.46±0.72	73.55±0.45	85.04±0.22
ICT	78.93±0.57	79.28±0.36	80.71±0.12
VAT	77.6±0.69	79.55±0.47	89.42±0.25
MixMatch	61.51±0.62	74.75±0.34	85.62±0.12
Pseudo-label	72.35±0.56	77.2±0.41	80.88±0.18
COVIDCon	**92.71±0.57**	**94.22±0.36**	**97.07±0.14**

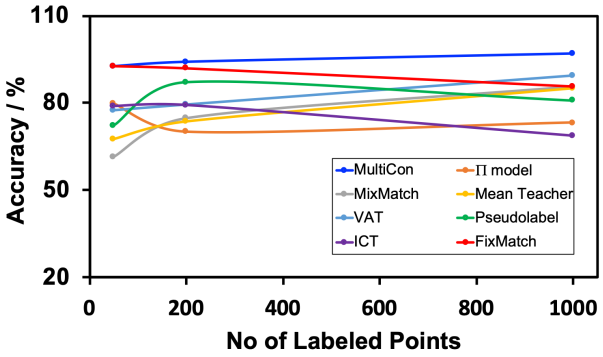

Fig. 4: Comparison between the true class prediction accuracies of COVIDCon, and other standard state-of-the-art SSL methods employed on the COVID-19 radiography dataset: COVIDCon (blue trace) outperformed all other SSL methods at every label, and its accuracy reached 97.07% on average at 1000 labels.

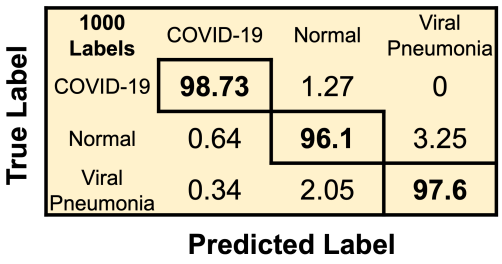

Fig. 5: Confusion matrices from COVIDCon showing the proportion of each predicted class (x-axis) for chest X-ray images in each true class (y-axis) with 1000 labels on the COVID-19 Radiography dataset. True class prediction accuracies are highlighted in bold. All numbers are rounded to two decimal places.

the chest X-rays of COVID-19 and viral pneumonia cases in the working dataset, COVIDCon performed exceptionally well, and a few mispredictions are well within the limit of diagnostic inaccuracy in actual clinical settings.

The class prediction accuracies are improved with increasing epoch numbers. At 6 epoch COVIDCon reached the accuracy of 94.22%, which is slightly decreased at 10 epoch with 50 labeled data.

3.2 Results on COVID-19 CT Scan Dataset

The class prediction accuracies are improved with increasing epoch numbers. At 6 epoch COVIDCon reached the accuracy of 98.56%, which increased further to 99.13% at 10 epoch with 20000 labeled data. The confusion matrices in Figure 6 show that COVIDCon achieved excellent accuracies to identify all of the three classes. With increasing labeled data, independent class prediction accuracies have also increased. For example, with 5000 labeled data, COVIDCon misclassified 3 out of 100 COVID-19 cases as viral pneumonia, whereas just 1.6% COVID-19 cases were assigned wrongly as viral pneumonia.

3.3 Receiver Operator Characteristics Curve

Receiver operator characteristic areas under the curves (ROC-AUC) are obtained (Figure 7) on the COVID-19 Radiography and COVID-19 CT scan datasets at 1000 and 20000 labels, respectively. The microaverage Area Under the Curve (AUC) is found to be 0.99 and 1, respectively, for both the datasets. The ROC-AUC values for individual classes lie in the range of 0.98 and 1, proving the efficacy of our method.

Consistent with the results obtained on the radiography dataset, COVIDCon outperformed other state-of-the-art SSL methods at all numbers of labeled data. The performance of COVIDCon on the CT scan dataset is recorded in Table 3. With 5000 labeled data, COVIDCon achieved an average accuracy of 98.30%, which is 6% better than the next best state-of-the-art Pseudo Label model. COVIDCon reached the highest-class prediction accuracy of 99.13% on average, with 20000 labeled data, which is 6.45% better than the next best state of the art model, Mix Match (Figure 8).

3.4 t-Distributed Stochastic Neighbor Embedding (t-SNE) Analysis

In order to understand the discrimination of learned embedding from our approach, we visualize the final embedding using t-SNE [34] implementation, which shows a snapshot of the COVID-19 Radiography dataset projected into a 2-dimensional feature space. We mimic the learning procedure by randomly selecting 1000 labeled samples under 10 epochs training. Then we test the learned embedding on the test set and show (Figure 9) the visualizations. The results demonstrate the fact that with the assistance of multi-contrastive loss, COVIDCon could reduce the intra-class variances much better in comparison to the next best FixMatch approach, and produce well-separated feature embeddings, resulting in high prediction accuracy for all of the three classes as can be seen from the confusion matrices in Figure 5 and 6. Similarly, COVIDCon produces much better well-separated feature embeddings (Figure 9) on CT scan datasets as well. All these results demonstrate the importance of the multi-contrastive loss component (Table 1) of COVIDCon.

3.5 Testing COVIDCon on a New Unseen COVID-19 Dataset

Oftentimes machine learning model that works almost perfectly using cross-validation fails miserably when tested to new unseen data. Therefore, we tested COVIDCon on a small repository [5] of confirmed COVID-19 cases with a case history. COVIDCon performed exceptionally well in identifying COVID-19 positive cases except the case number 59638 (Figure 10), which is predicted as normal from chest X-ray analysis. Furthermore, two unknown cases, number 59554 and 56442, from the repository are also identified as COVID-19 by COVIDCon. Interestingly, the case 56442 case has COVID-suspected CT features (Figure 10), and is also identified successfully as COVID-19 positive by COVIDCon. These results demonstrate the accuracy of our model and it's potential use in the clinical settings.

3.6 Ablation Study

In order to understand the effect of each loss component of COVIDCon, we performed an ablation study. We isolated different components of our loss function and investigated their impacts on the model's performance. We evaluated Supervised Cross-Entropy (CE), Consistency Regularization (CR), Cross-Entropy and Multi-Contrastive (MC) loss, and finally COVIDCon. From the prediction accuracies for the COVID-19 Radiography and CT Scan datasets, as recorded in table 4, we observed that the contrastive loss in conjunction with supervised cross-entropy performed quite well on its

Methods/Labels	5000	10000	20000	30000
Π-model	82.78±0.53	86.67±0.47	87.06±0.32	84.72±0.19
Mean Teacher	91.37±0.51	92.7±0.43	92.46±0.29	93.54±0.42
ICT	89.22±0.18	88.6±0.73	89.18±0.45	89.27±0.26
VAT	90.15±0.43	89.42±0.55	89.07±0.31	92.66±0.28
MixMatch	85.41±0.82	86.93±0.45	89.64±0.11	86.11±0.21
Pseudo-label	92.3±0.67	92.85±0.41	92.68±0.24	91.79±0.14
FixMatch	89.35±0.36	78.8±0.98	88.32±0.38	81.29±0.87
COVIDCon	**98.30±0.77**	**98.79±0.25**	**99.13±0.29**	**99.06±0.37**

Table 3: Comparison of accuracy achieved by COVIDCon and other state-of-the-art SSL methods on CT scan dataset. COVIDCon outperforms others methods at every label.
.

5000 Labels	COVID-19	Normal	Viral Pneumonia
COVID-19	**96.61**	0.43	2.96
Normal	0.19	**99.55**	0.26
Viral Pneumonia	0.69	0.82	**98.49**

20000 Labels	COVID-19	Normal	Viral Pneumonia
COVID-19	**98.12**	0.28	1.60
Normal	0.19	**99.66**	0.15
Viral Pneumonia	0.15	0.45	**99.40**

Fig. 6: Confusion matrices from COVIDCon showing the proportion of each predicted class (x-axis) for chest CT scan images in each true class (y-axis) with 5000, and (d) 20000 labels on the COVID-19 CT Scan dataset, and True class prediction accuracies are highlighted in bold. All numbers are rounded to two decimal places.

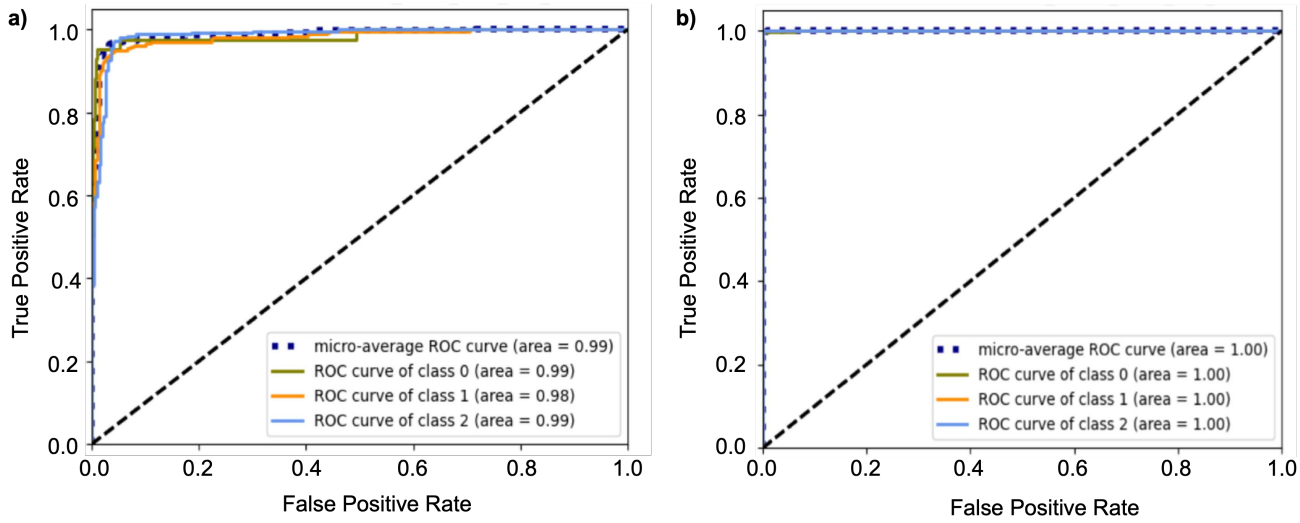

Fig. 7: ROC-AUC curves on Radiography and CT scan Datasets, class 0: COVID-19, class 1: normal, class 2: viral pneumonia

own. But the composition of all components, proposed in COVIDCon demonstrated the best performance.

4 Discussion

Our investigations demonstrate demonstrates the use of state-of-the-art benchmark SSL methods for the potential diagnosis of COVID using X-ray and CT-Scan for the first time. We observe that COVIDCon exhibits significantly improved performance compared to other state-of-the-art methods in all settings that we have studied. For limited labeled data, COVIDCon performs exceptionally well. For instance, at 1000 labels, on the COVID-19 Radiography Dataset, it achieves an accuracy of 97.07%, which is 7.65% better than the next best VAT model [36]. The improved performance of

Loss Function	COVID-19 Radiography (1000 labels)	COVID-19 CT Scan (30000 labels)
Supervised CE	55.21	96.93
Semi-Supervised CR	96.13	98.4
CE + MC	96.34	98.51
COVIDCon	**97.07**	**99.06**

Table 4: Ablation study of COVIDCon on COVID-19 Radiography and CT scan dataset.

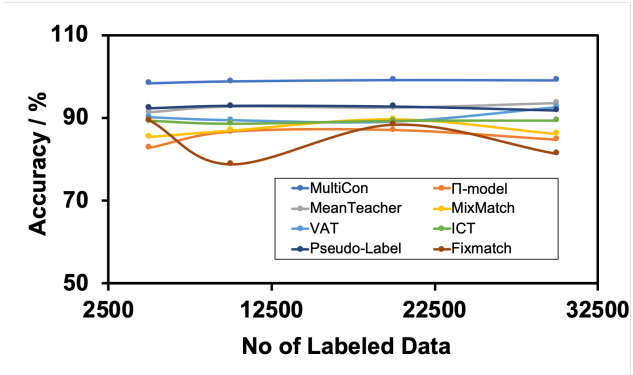

Fig. 8: Comparison between the true class prediction accuracies of COVIDCon, and other standard state-of-the-art SSL methods employed on the CT Scan dataset: COVIDCon (blue trace) outperformed all other SSL methods at every label, and its accuracy reached 99.13% on average at 20000 labels.

COVIDCon on such a small dataset signifies its use in the medical diagnosis domain, where it is often difficult to obtain large annotated datasets. COVIDCon also works extremely well on a larger COVID-19 CT Scan Dataset. It achieves an accuracy of 99.13%, which is 6.5% superior to its closest competitor, Pseudo-label. These results demonstrate COVIDCon as the benchmark SSL algorithm for potential diagnosis of COVID-19 from chest X-rays and CT-Scans.

Furthermore, COVIDCon performs exceptionally well in identifying COVID-19 positive cases from completely unseen chest X-rays and CT scans; therefore has potentials to be used in clinical settings. In that context, additional attributes, such as demographic information, race, etc. can also be included in COVIDCon for strengthening the ground for classification. It will be also important to understand the possibilities of identifying asymptomatic COVID-19 cases and differentiate them from normal cases from X-rays and CT scans screening using COVIDCon. For that, a large amount of data from normal, asymptomatic, as well as symptomatic COVID-19 cases will be required and collaborative work between hospitals and machine learning scientists will be necessary. Future studies will be focused on making the proposed approach available to the society for use in early and efficient detection of COVID-19.

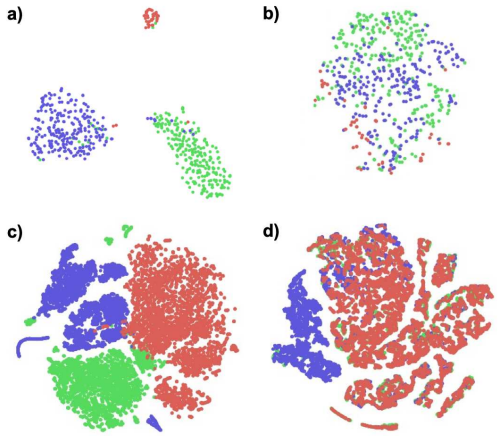

Fig. 9: Visualization of learned feature embeddings on COVID-19 Radiography and CT Scan datasets. a) COVIDCon, b) FixMatch feature embedding with 1000 labels on COVID-19 Radiography dataset. c) COVIDCon, d) FixMatch feature embedding with 20000 labels on CT Scan dataset. Red, blue, and green dots are COVID-19, pneumonia, and normal cases.

5 Acknowledgement

(Pracheta Sahoo and Indranil Roy contributed equally to this work.). The authors thank Zhuoyi Wang for his valuable insights and support.

6 Funding

The research reported herein was supported in part by NSF awards (DMS-1737978, DGE-2039542, OAC1828467, OAC-1931541, DGE-1906630) and an IBM faculty award (re-search).

7 Conflict of Interest

Authors declared no conflict of interest.

8 Availability of data and material

All the data and materials are available on request to the authors (L.K., P.S.).

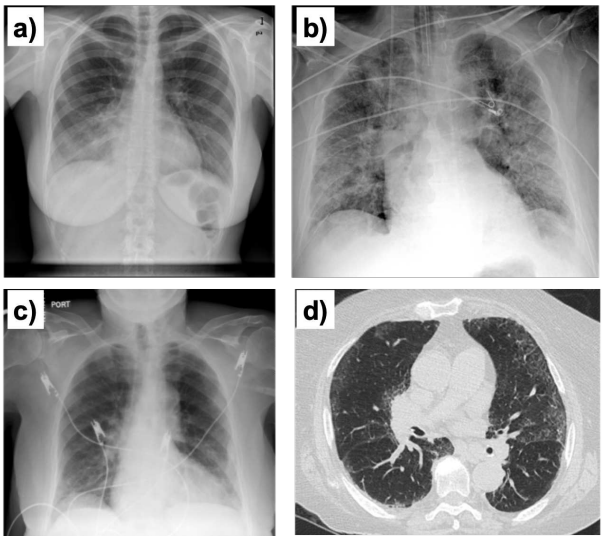

Fig. 10: X-Rays and CT scan cases from a new Unseen COVID-19 Repository [1]. a) Case number 59638: Chest X-ray of a 29-year-old immunocompromised female patient with a 3-day history of cough and fever. Past medical history includes severe ulcerative colitis treated with tofacitinib. b) Chest X-ray of case number 59554 presented to emergency department with respiratory distress and dyspnea. C) Chest X-ray of the case number 56442 with a history of shortness of breath and cough, and recent international travel with someone who tested positive for COVID-19. d) CT scan of the case number 56442.

9 Code availability

Not applicable.

10 Ethics approval

This article does not contain any studies with human participants performed by any of the authors.

11 Consent to participate

Not applicable.

12 Consent for publication

The authors give consent to publish the paper.

References

1. https://www.who.int/health-topics/coronavirus#tab=tab_1
2. https://www.healthnewsreview.org/toolkit/tips-for-understanding-studies/understanding-medical-tests-sensitivity-specificity-and-positive-predictive-value/
3. https://pytorch.org/
4. http://www.image-net.org
5. https://cortex.acr.org/COVID19/
6. Abdel-Basset, M., Chang, V., Hawash, H., Chakrabortty, R.K., Ryan, M.: Fss-2019-ncov: A deep learning architecture for semi-supervised few-shot segmentation of covid-19 infection. Knowledge-Based Systems **212**, 106647 (2021)
7. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., Xia, L.: Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. Radiology p. 200642 (2020)
8. Anand, A., Racah, E., Ozair, S., Bengio, Y., Côté, M.A., Hjelm, R.D.: Unsupervised state representation learning in atari. arXiv preprint arXiv:1906.08226 (2019)
9. Apostolopoulos, I.D., Mpesiana, T.A.: Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine **43**(2), 635–640 (2020)
10. Axell-House, D.B., Lavingia, R., Rafferty, M., Clark, E., Amirian, E.S., Chiao, E.Y.: The estimation of diagnostic accuracy of tests for COVID-19: A scoping review. J. Infect. (2020)
11. Bachman, P., Alsharif, O., Precup, D.: Learning with pseudo-ensembles. arXiv preprint arXiv:1412.4864 (2014)
12. Bachman, P., Hjelm, R.D., Buchwalter, W.: Learning representations by maximizing mutual information across views. arXiv preprint arXiv:1906.00910 (2019)
13. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. In: NIPS, pp. 5050–5060 (2019)
14. Caruana, R., Niculescu-Mizil, A.: An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd international conference on Machine learning, pp. 161–168 (2006)
15. Cohen, J.P., Morrison, P., Dao, L.: COVID-19 image data collection. arXiv preprint arXiv:2003.11597 (2020)
16. Das, D., Santosh, K., Pal, U.: Truncated inception net: Covid-19 outbreak screening using chest x-rays. Physical and engineering sciences in medicine **43**(3), 915–925 (2020)
17. Esteva, A., Kuprel, B., Novoa, R.A., et al.: Correction: Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. Nature **546**(7660), 686–686 (2017)
18. Fan, D.P., Zhou, T., Ji, G.P., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L.: Inf-net: Automatic covid-19 lung infection segmentation from ct images. IEEE Transactions on Medical Imaging **39**(8), 2626–2637 (2020)
19. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9729–9738 (2020)
20. Hemdan, E.E.D., Shouman, M.A., Karar, M.E.: Covidx-net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images. arXiv preprint arXiv:2003.11055 (2020)
21. Henaff, O.: Data-efficient image recognition with contrastive predictive coding. In: International Conference on Machine Learning, pp. 4182–4192. PMLR (2020)

22. Hjelm, R.D., Fedorov, A., Lavoie-Marchildon, S., Grewal, K., Bachman, P., Trischler, A., Bengio, Y.: Learning deep representations by mutual information estimation and maximization. arXiv preprint arXiv:1808.06670 (2018)
23. Kanne, J.P., Little, B.P., Chung, J.H., Elicker, B.M., Ketai, L.H.: Essentials for radiologists on COVID-19: An update—radiology scientific expert panel (2020)
24. Khobahi, S., Agarwal, C., Soltanalian, M.: Coronet: A deep network architecture for semi-supervised task-based identification of covid-19 from chest x-ray images. MedRxiv (2020)
25. Konar, D., Panigrahi, B.K., Bhattacharyya, S., Dey, N., Jiang, R.: Auto-diagnosis of covid-19 using lung ct images with semi-supervised shallow learning network. IEEE Access **9**, 28716–28728 (2021)
26. Kubista, M., Andrade, J.M., Bengtsson, M., Forootan, A., et al.: The real-time polymerase chain reaction. Mol. Asp. Med. **27**(2-3), 95–125 (2006)
27. Kurakin, A., Raffel, C., Berthelot, D., Cubuk, E.D., Zhang, H., Sohn, K., Carlini, N.: Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring (2020)
28. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)
29. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)
30. Lee, D.H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML, vol. 3, p. 2 (2013)
31. Lee, E.Y., Ng, M.Y., Khong, P.L.: COVID-19 pneumonia: What has CT taught us? Lancet Infect. Dis. **20**(4), 384–385 (2020)
32. Li, R., Zhang, W., Suk, H.I., Wang, L., Li, J., Shen, D., Ji, S.: Deep learning based imaging data completion for improved brain disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 305–312. Springer (2014)
33. Ma, J., Nie, Z., Wang, C., Dong, G., Zhu, Q., He, J., Gui, L., Yang, X.: Active contour regularized semi-supervised learning for covid-19 ct infection segmentation with limited annotations. Physics in Medicine & Biology **65**(22), 225034 (2020)
34. Maaten, L.v.d., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)
35. McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.C., Darzi, A., et al.: International evaluation of an ai system for breast cancer screening. Nature **577**(7788), 89–94 (2020)
36. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. TPAMI **41**(8), 1979–1993 (2018)
37. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE transactions on pattern analysis and machine intelligence **41**(8), 1979–1993 (2018)
38. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
39. Ozturk, T., Talo, M., Yildirim, E.A., Baloglu, U.B., Yildirim, O., Acharya, U.R.: Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput. Biol. Med. p. 103792 (2020)
40. Panwar, H., Gupta, P., Siddiqui, M.K., Morales-Menendez, R., Singh, V.: Application of deep learning for fast detection of COVID-19 in X-rays using nCOVnet. Chaos, Solit Fractals p. 109944 (2020)
41. Pathak, Y., Shukla, P.K., Tiwari, A., Stalin, S., Singh, S., Shukla, P.K.: Deep transfer learning based classification model for COVID-19 disease. IRBM (2020)
42. Rahman, T., et al.: COVID-19 radiography database (2020). URL https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
43. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M.P., Ng, A.Y.: Chexnet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225 (2017)
44. Sahoo, P., Roy, I., Wang, Z., Mi, F., Yu, L., Balasubramani, P., Khan, L., Stoddart, J.F.: Multicon: A semi-supervised approach for predicting drug function from chemical structure analysis. J. Chem. Inf. Model. **60**, 5995–6006 (2020)
45. Sajjadi, M., Javanmardi, M., Tasdizen, T.: Mutual exclusivity loss for semi-supervised deep learning. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 1908–1912. IEEE (2016)
46. Sohn, K., Berthelot, D., Li, C.L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., Raffel, C.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. arXiv preprint arXiv:2001.07685 (2020)
47. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Advances in neural information processing systems, pp. 1195–1204 (2017)
48. Tian, Y., Krishnan, D., Isola, P.: Contrastive multiview coding. arXiv preprint arXiv:1906.05849 (2019)
49. Velickovic, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D.: Deep graph infomax. In: ICLR (Poster) (2019)
50. Verma, V., Lamb, A., Kannala, J., Bengio, Y., Lopez-Paz, D.: Interpolation consistency training for semi-supervised learning. IJCAI (2019)
51. Wang, L., Wong, A.: 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)
52. Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., Cai, M., Yang, J., Li, Y., Meng, X., et al.: A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). MedRxiv (2020)
53. Wehbe, R.M., Sheng, J., Dutta, S., Chai, S., Dravid, A., Barutcu, S., Wu, Y., Cantrell, D.R., Xiao, N., Allen, B.D., et al.: Deepcovid-xr: An artificial intelligence algorithm to detect COVID-19 on chest radiographs trained and tested on a large us clinical dataset. Radiology p. 203511 (2020)
54. Whitaker, M., Kron, T., Sobolewski, M., Dove, R.: Covid-19 pandemic planning: considerations for radiation oncology medical physics. Physical and Engineering Sciences in Medicine **43**(2), 473–480 (2020)
55. Xie, Q., Dai, Z., Hovy, E., Luong, M.T., Le, Q.V.: Unsupervised data augmentation for consistency training. arXiv preprint arXiv:1904.12848 (2019)
56. Xie, S., Girshick, R., He, K., et al.: Aggregated residual transformations for deep neural networks. In: CVPR 2017, pp. 1492–1500

57. Xie, X., Zhong, Z., Zhao, W., Zheng, C., Wang, F., Liu, J.: Chest CT for typical 2019-ncov pneumonia: Relationship to negative RT-PCR testing. Radiology p. 200343 (2020)
58. shensheng Xu, S., Mak, M.W., Cheung, C.C.: Deep neural networks versus support vector machines for ecg arrhythmia classification. In: 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 127–132. IEEE (2017)
59. Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., Yu, L., Chen, Y., Su, J., Lang, G., et al.: Deep learning system to screen coronavirus disease 2019 pneumonia. arxiv 2020. arXiv preprint arXiv:2002.09334
60. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Deep metric learning for person re-identification. In: 2014 22nd International Conference on Pattern Recognition, pp. 34–39. IEEE (2014)
61. Yoon, S.H., Lee, K.H., Kim, J.Y., Lee, Y.K., Ko, H., Kim, K.H., Park, C.M., Kim, Y.H.: Chest radiographic and CT findings of the 2019 novel coronavirus disease (COVID-19): analysis of nine patients treated in korea. Korean J Radiol **21**(4), 494–500 (2020)
62. Zhang, K., Liu, X., Shen, J., Li, Z., Sang, Y., Wu, X., Zha, Y., Liang, W., Wang, C., Wang, K., et al.: Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. Cell (2020)
63. Zhao, J., Zhang, Y., He, X., Xie, P.: COVID-CT-dataset: a CT scan dataset about COVID-19. arXiv preprint arXiv:2003.13865 (2020)
64. Zhu, X., Goldberg, A.B.: Introduction to semi-supervised learning. Synth. Lect. Artif. Intell. Mach. Learn. **3**(1), 1–130 (2009)
65. Zu, Z.Y., Jiang, M.D., Xu, P.P., Chen, W., Ni, Q.Q., Lu, G.M., Zhang, L.J.: Coronavirus disease 2019 (COVID-19): A perspective from China. Radiology p. 200490 (2020)

Figures

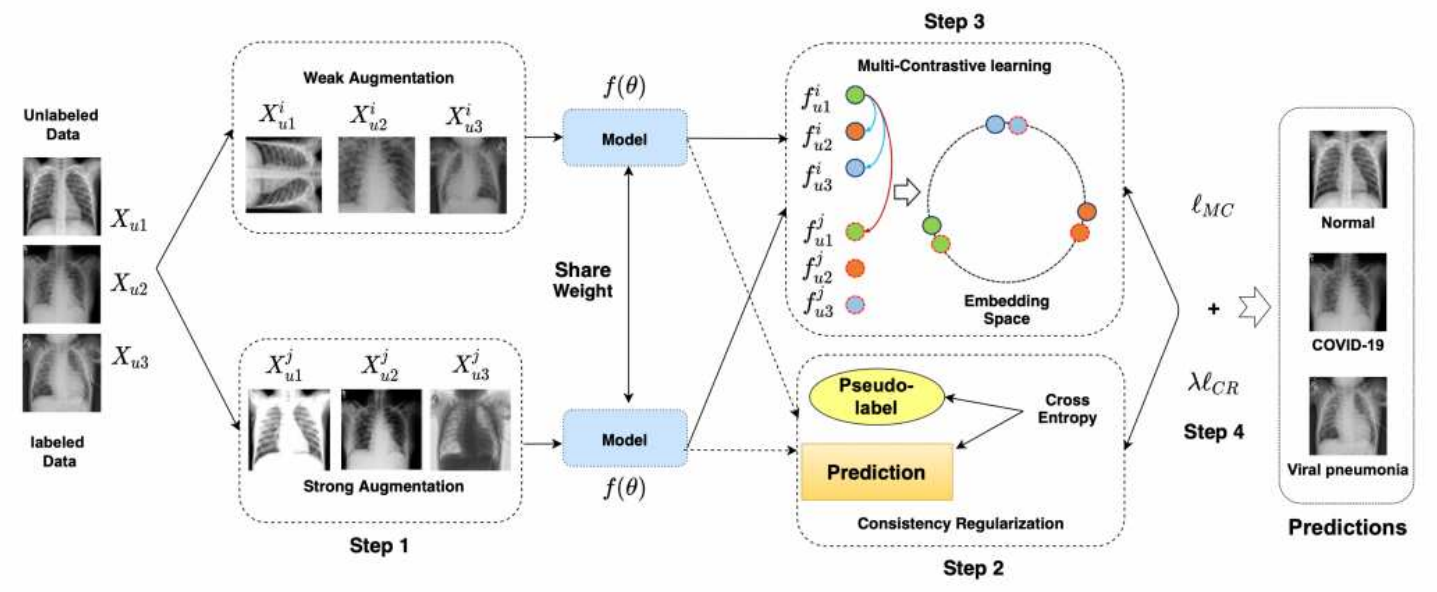

Figure 1

The framework of COVIDCon. Different input chest X-rays with corresponding (weak and strong) augmented transformations are projected into embedding features. For the consistency regularization, the model prediction works as pseudo-label, ℓCR aims to make the output from strong augmentation match of the pseudo-label. The model is trained under the combination of ℓCR and ℓMC . The colored solid lined circles represent strong augmented data points, the dotted circles represent weak augmented data points, and same color represents same class

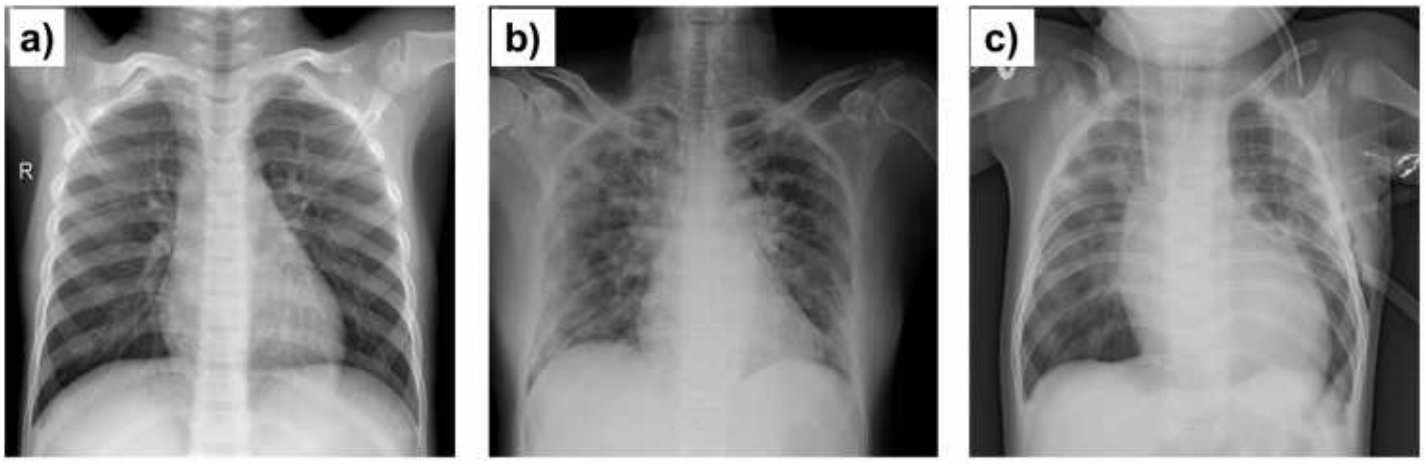

Figure 2

Sample chest X-rays taken from the COVID-19 Radiography dataset. (a) Normal case, (b) COVID-19 case showing bilateral ground-glass opacities with prominent peripheral, perihilar and basal distribution within a multilobar involvement, and (c) viral pneumonia case with visible left basilar opacity

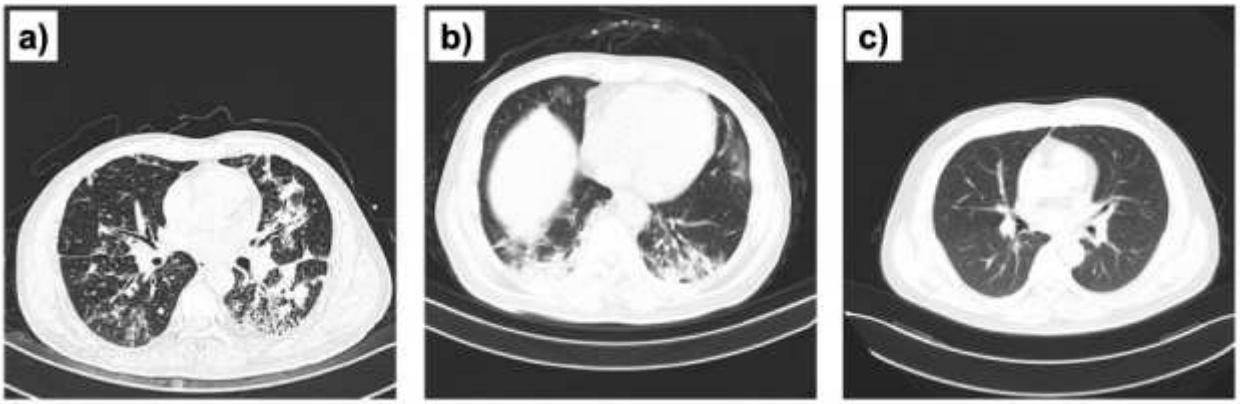

Figure 3

CT images taken from COVID-19 CT Scan dataset. Typical examples showing (a) common pneumonia (CP), (b) COVID-19 (NCP), and (c) normal CT scan images.

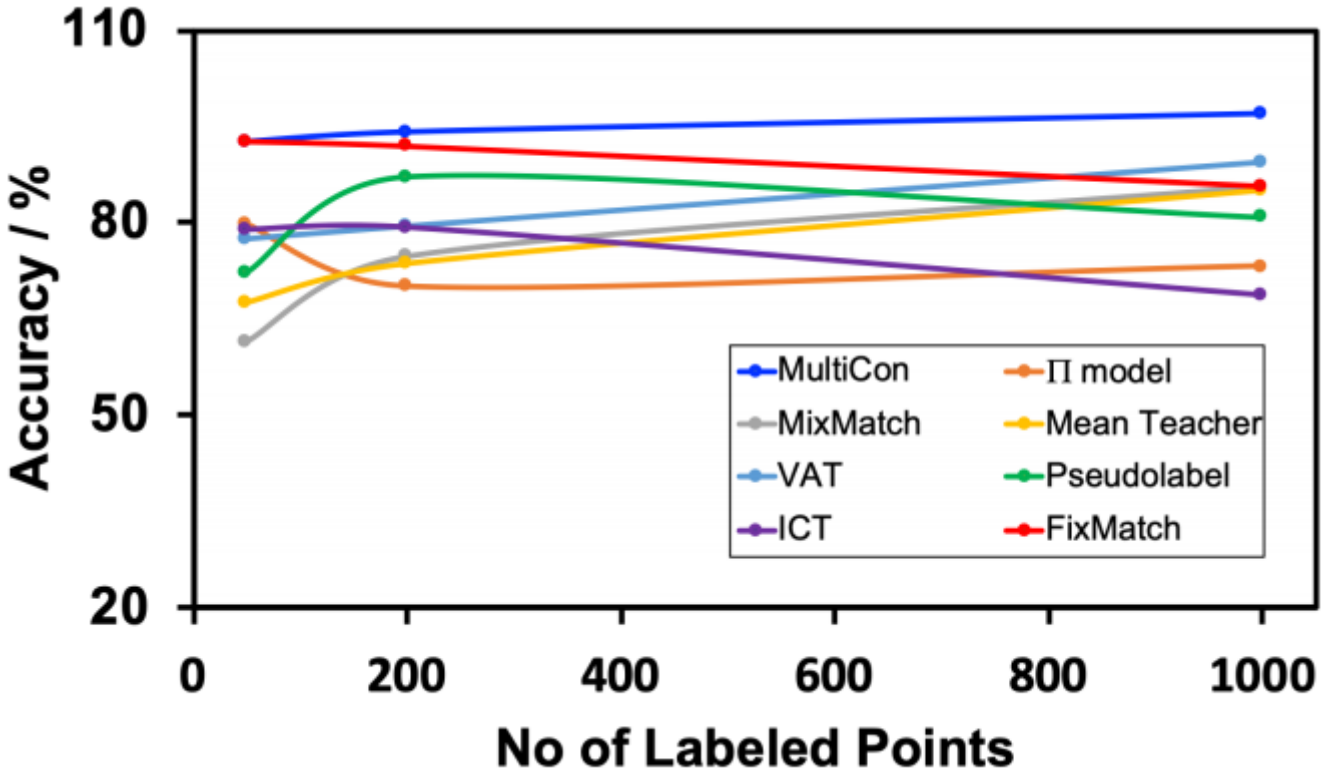

Figure 4

Comparison between the true class prediction accuracies of COVIDCon, and other standard state-of-the-art SSL methods employed on the COVID-19 radiography dataset: COVIDCon (blue trace) outperformed all other SSL methods at every label, and its accuracy reached 97.07% on average at 1000 labels.

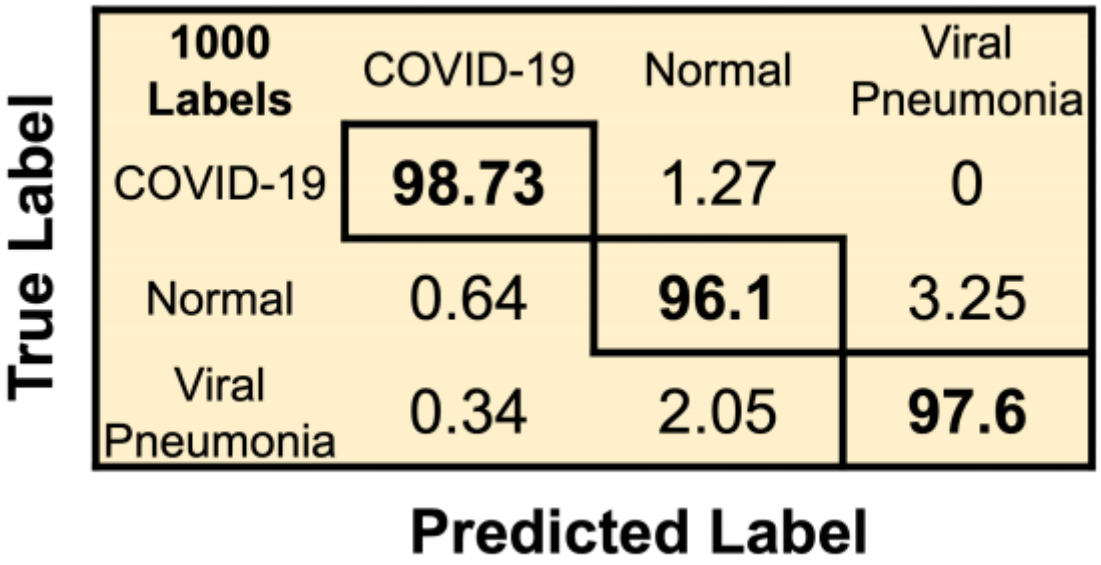

Figure 5

Confusion matrices from COVIDCon showing the proportion of each predicted class (x-axis) for chest X-ray images in each true class (y-axis) with 1000 labels on the COVID-19 Radiography dataset. True class prediction accuracies are highlighted in bold. All numbers are rounded to two decimal places

5000 Labels	COVID-19	Normal	Viral Pneumonia
COVID-19	**96.61**	0.43	2.96
Normal	0.19	**99.55**	0.26
Viral Pneumonia	0.69	0.82	**98.49**

20000 Labels	COVID-19	Normal	Viral Pneumonia
COVID-19	**98.12**	0.28	1.60
Normal	0.19	**99.66**	0.15
Viral Pneumonia	0.15	0.45	**99.40**

Figure 6

Confusion matrices from COVIDCon showing the proportion of each predicted class (x-axis) for chest CT scan images in each true class (y-axis) with 5000, and (d) 20000 labels on the COVID-19 CT Scan dataset, and True class prediction accuracies are highlighted in bold. All numbers are rounded to two decimal places

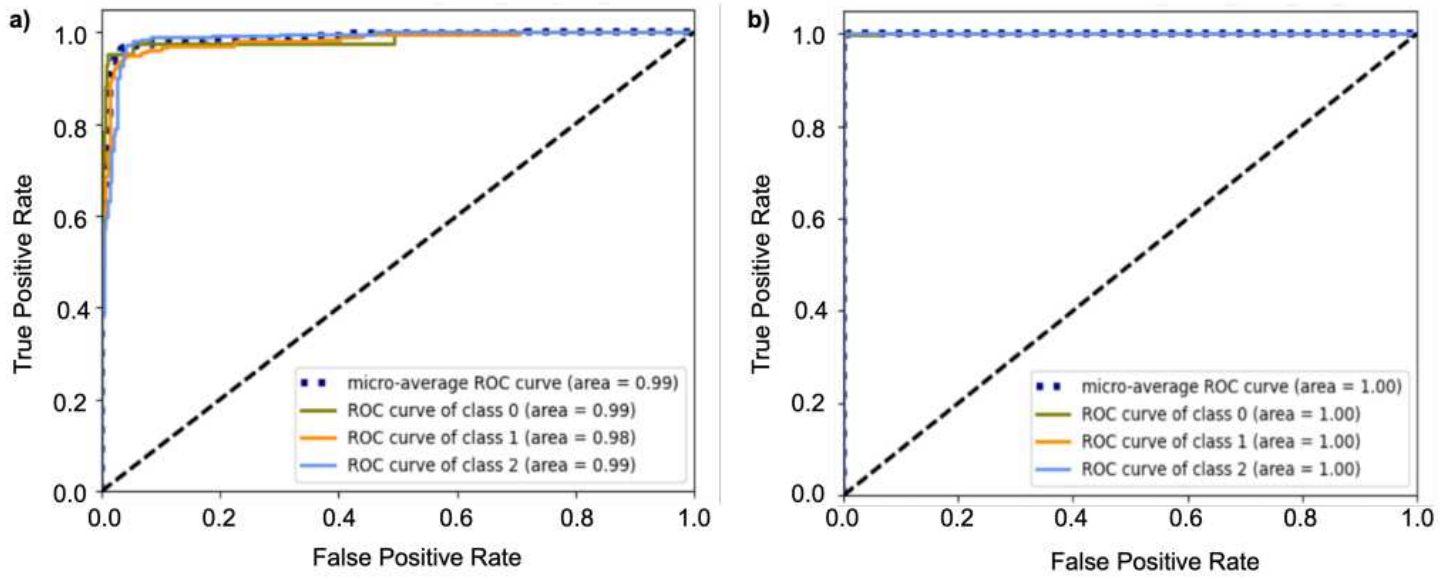

Figure 7

OC-AUC curves on Radiography and CT scan Datasets, class 0: COVID-19, class 1: normal, class 2: viral pneumonia

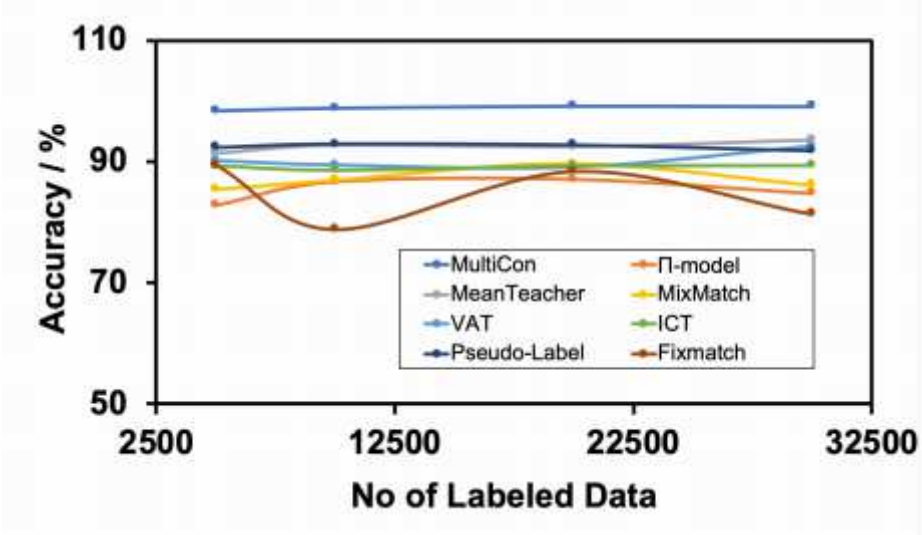

Figure 8

Comparison between the true class prediction accuracies of COVIDCon, and other standard state-of-the-art SSL methods employed on the CT Scan dataset: COVIDCon (blue trace) outperformed all other SSL methods at every label, and its accuracy reached 99.13% on average at 20000 labels

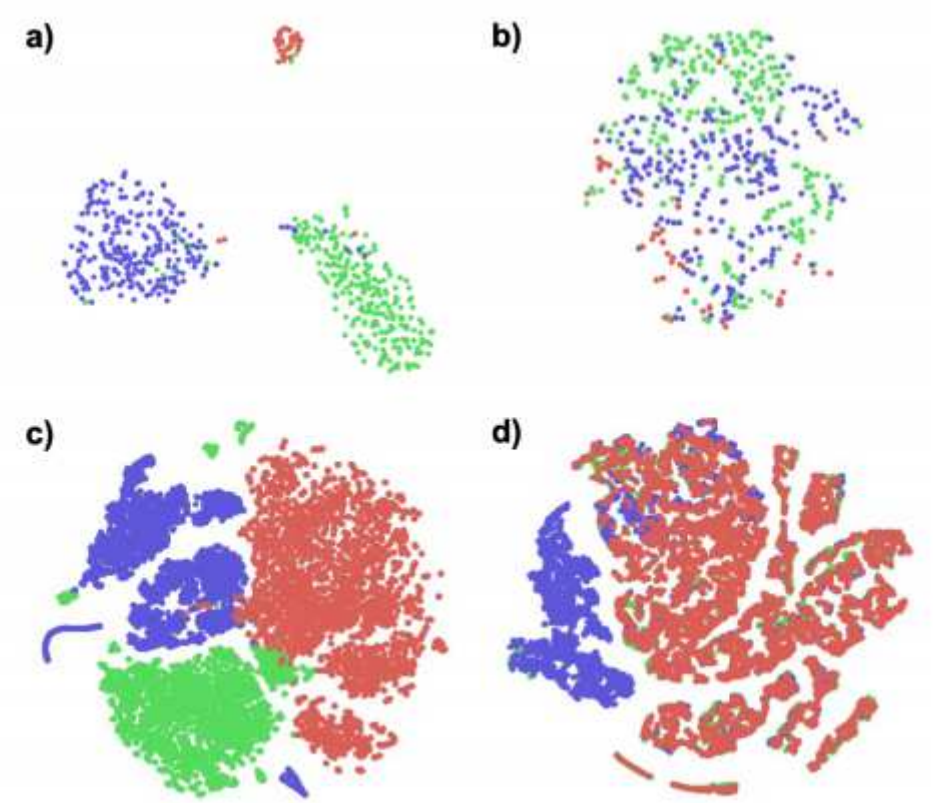

Figure 9

Visualization of learned feature embeddings on COVID-19 Radiography and CT Scan datasets. a) COVIDCon, b) FixMatch feature embedding with 1000 labels on COVID-19 Radiography dataset. c) COVIDCon, d) FixMatch feature embedding with 20000 labels on CT Scan dataset. Red, blue, and green dots are COVID-19, pneumonia, and normal cases.

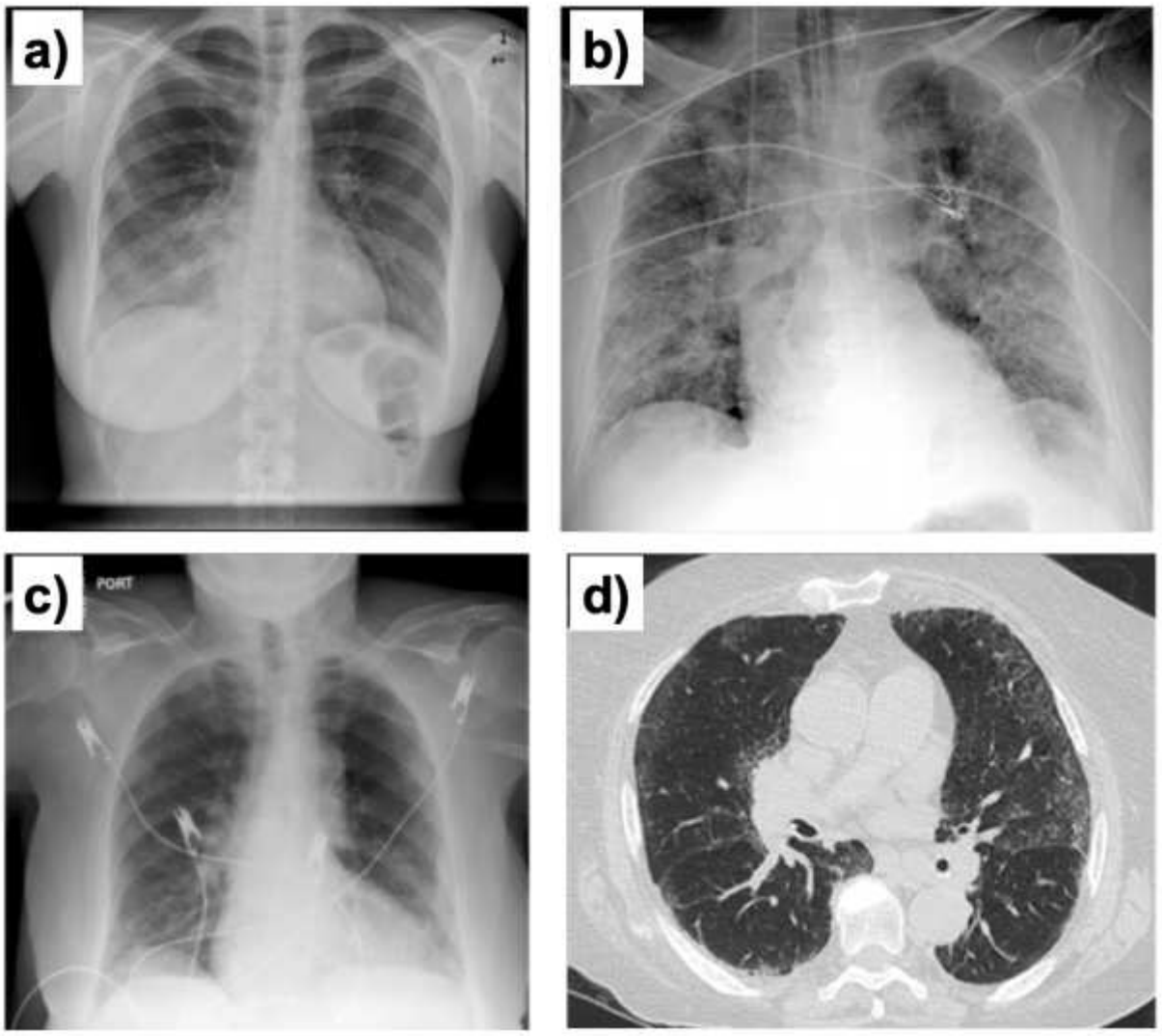

Figure 10

X-Rays and CT scan cases from a new Unseen COVID-19 Repository 1 . a) Case number 59638: Chest X-ray of a 29-year-old immunocompromised female patient with a 3-day history of cough and fever. Past medical history includes severe ulcerative colitis treated with tofacitinib. b) Chest X-ray of case number 59554 presented to emergency department with respiratory distress and dyspnea. C) Chest X-ray of the case number 56442 with a history of shortness of breath and cough, and recent international travel with someone who tested positive for COVID-19. d) CT scan of the case number 56442.