HOKKAIDO UNIVERSITY

Title	Boosting automatic COVID-19 detection performance with self-supervised learning and batch knowledge ensembling
Author(s)	Li, Guang; Togo, Ren; Ogawa, Takahiro et al.
Citation	Computers in biology and medicine, 158, 106877 https://doi.org/10.1016/j.compbiomed.2023.106877
Issue Date	2023-04-03
Doc URL	https://hdl.handle.net/2115/91735
Rights	© 2023. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/
Rights(URL)	https://creativecommons.org/licenses/by-nc-nd/4.0/
Type	journal article
File Information	Guang_CBM_2023.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Boosting Automatic COVID-19 Detection Performance with Self-Supervised Learning and Batch Knowledge Ensembling

Guang Li[a], Ren Togo[b], Takahiro Ogawa[b], Miki Haseyama[b]

[a]*Graduate School of Information Science and Technology, Hokkaido University, N-14, W-9, Kita-Ku, Sapporo, 060-0814, Japan*

[b]*Faculty of Information Science and Technology, Hokkaido University, N-14, W-9, Kita-Ku, Sapporo, 060-0814, Japan*

Abstract

Problem: Detecting COVID-19 from chest X-Ray (CXR) images has become one of the fastest and easiest methods for detecting COVID-19. However, the existing methods usually use supervised transfer learning from natural images as a pretraining process. These methods do not consider the unique features of COVID-19 and the similar features between COVID-19 and other pneumonia.

Aim: In this paper, we want to design a novel high-accuracy COVID-19 detection method that uses CXR images, which can consider the unique features of COVID-19 and the similar features between COVID-19 and other pneumonia.

Methods: Our method consists of two phases. One is self-supervised learning-based pertaining; the other is batch knowledge ensembling-based fine-tuning. Self-supervised learning-based pretraining can learn distinguished representations from CXR images without manually annotated labels. On the other hand, batch knowledge ensembling-based fine-tuning can utilize category knowledge of images in a batch according to their visual feature similarities to improve detection performance. Unlike our previous implementation, we introduce batch knowledge ensembling into the fine-tuning phase, reducing the memory used in self-supervised learning and improving COVID-19 detection accuracy.

Results: On two public COVID-19 CXR datasets, namely, a large dataset and an unbalanced dataset, our method exhibited promising COVID-19 detection performance. Our method maintains high detection accuracy even when annotated CXR training images are reduced significantly (e.g., using only 10% of the original dataset). In addition, our method is insensitive to changes in hyperparameters.

Conclusion: The proposed method outperforms other state-of-the-art COVID-19 detection methods in different settings. Our method can reduce the workloads of healthcare providers and radiologists.

Keywords: COVID-19, CXR images, Self-supervised learning, Batch knowledge ensembling.

1. Introduction

As a pandemic, COVID-19 has rapidly spread worldwide, affecting the health and lives of billions of people [1]. There have been 632,533,408 confirmed cases of COVID-19, including 6,592,320 deaths reported in the world as of November 15, 2022.[1] Vaccines have played a critical role in preventing the spread of new infections, but the situation remains critical because some countries and special populations are still unable to receive the vaccine. Recently, the number of infected people has been increasing drastically since the high variability of COVID-19. Identifying infected patients early and separating them from the population is crucial to controlling the infection. In line with WHO reports, the gold standard for COVID-19 detection is currently real-time reverse transcription polymerase chain reaction (RT-PCR) [2]. Although RT-PCR has many advantages, it also has some shortcomings, including a high false-negative rate and being a lengthy process [3]. Furthermore, RT-PCR is typically inadequate in many hard-hit and underdeveloped areas, which is detrimental to stop the continued spread of COVID-19 worldwide [4]. The use of computed tomography (CT) and X-ray have been complementary to RT-PCR for COVID-19 detection [5].

The common presence of radiologic findings of pneumonia in patients with COVID-19 makes radiologic examinations valuable for COVID-19 detection [6]. The chest X-ray (CXR) is particularly beneficial because it cost less, takes less time, and exposes you to less radiation than CT [7]. The detection of COVID-19 through chest radiography offers significant potential for screening and analyzing the health of patients. However, manual detection of COVID-19 from CXR images has several problems. For example, requiring radiologists to detect COVID-19 infection from several CXR images is challenging as it is time-consuming and easily has human errors [8]. Furthermore, it is difficult to distinguish COVID-19 from other vi-

Email addresses: guang@lmd.ist.hokudai.ac.jp (Guang Li[a]), togo@lmd.ist.hokudai.ac.jp (Ren Togo[b]), ogawa@lmd.ist.hokudai.ac.jp (Takahiro Ogawa[b]), mhaseyama@lmd.ist.hokudai.ac.jp (Miki Haseyama[b])

[1] https://covid19.who.int

ral pneumonia cases because of their similar appearances [9].

To overcome the above problems, several researchers have used deep learning (DL) to detect COVID-19 infection from CXR images [10, 11]. A study [12] has proposed using transfer learning with dominant convolutional neural networks (CNNs) such as ResNet [13], SqueezeNet [14], and DenseNet [15] to perform COVID-19 detection from CXR images, achieving good detection performance on an unbalanced small COVID-19 dataset. Moreover, another study [16] has proposed the ensemble fine-tuning of CNNs, end-to-end training of CNNs, and deep feature extraction for an additional support vector machine (SVM) classifier to obtain better COVID-19 detection performance. However, these methods used supervised transfer learning from natural images as a pretraining process and did not consider the unique features of COVID-19 and the similar features between COVID-19 and other pneumonia. Furthermore, these studies typically evaluated small COVID-19 CXR image datasets, which may have limitations when used in a real-world clinical situation.

In this study, we propose a novel automatic COVID-19 detection method with self-supervised learning and batch knowledge ensembling using CXR images. Self-supervised learning can learn distinguished representations from CXR images without manually annotated labels as a pretraining phase, which is more suitable for complex COVID-19 visual features than supervised learning. Furthermore, batch knowledge ensembling can utilize category information of images in a batch according to their visual feature similarities to boost detection performance. The proposed method achieved promising COVID-19 detection performance on a large COVID-19 CXR dataset and an unbalanced COVID-19 CXR dataset. Our method also maintains high detection accuracy even when the annotated CXR training images are reduced significantly (e.g., using only 10% of the original dataset). In addition, our method is insensitive to changes in hyperparameters. As a result, the proposed method outperforms state-of-the-art (SOTA) COVID-19 detection methods in different settings. Our method can reduce the workloads of healthcare providers and radiologists.

Our contributions are summarized as follows.

- We propose performing self-supervised learning to learn distinguished representations without manually annotated labels from CXR images as a pretraining phase, which is more suitable for complex COVID-19 visual features than supervised learning.
- We propose batch knowledge ensembling-based fine-tuning, which can utilize category information of images in a batch according to their visual feature similarities to boost the detection performance.
- The proposed method achieves high detection accuracy on a large COVID-19 CXR dataset and an unbalanced dataset; it is insensitive to changes in hyperparameters.
- The proposed method can maintain high detection accuracy even when the annotated CXR training images are reduced significantly.

A preliminary report of this study was published in our previous study [17]. Our previous work is extended in this paper in the following ways. First, we introduce batch knowledge ensembling into the fine-tuning phase, reducing the memory used in self-supervised learning and achieving higher COVID-19 detection accuracy. Second, we evaluate the performance of the proposed method in two network structures compared with other SOTA methods. Third, we use a new unbalanced COVID-19 dataset to evaluate the robustness of the proposed method. Finally, we discuss the effects of hyperparameters on the proposed method.

The remainder of this paper is organized as follows. In Section 2, we introduce related works. In Section 3, we describe the details of the proposed method. Then, we present the experiments, discussion, and conclusions in Sections 4, 5, and 6, respectively.

2. Related Works

2.1. Automatic COVID-19 Detection

Numerous studies have been conducted with DL to perform COVID-19 detection using CXR [18] or CT images [19]. Different neural network structures, transfer learning techniques, and ensemble methods have been proposed to improve the performance of automatic COVID-19 detection. For example, several studies [20, 21, 22, 23, 24, 25] have proposed transfer learning methods with the most used deep CNNs such as AlexNet [26], VGGNet [27], InceptionNet [28], ResNet, DenseNet, CheXNet [29], and MobileNet [30] to perform COVID-19 detection from CXR images. Furthermore, some researchers [31, 32, 33, 34, 35] have proposed integrating DCNN feature extraction with machine learning algorithms, such as SVM [36], k-nearest neighbour [37], naive bayes [38], decision tree [39], reliefF [40], and sparse autoencoder [41]. In addition, several studies have used capsule networks or attention-based networks to perform COVID-19 detection from CXR images, which can consider more spatial relationships than traditional DCNNs [42, 43, 44, 45]. Furthermore, some studies have proposed using confidence-aware or uncertainty-aware learning for robust COVID-19 detection [46, 47, 48]. Although these methods achieved good COVID-19 detection performance, they were typically tested on small COVID-19 datasets and applied only to supervised learning.

2.2. Self-Supervised Learning

In the past few years, self-supervised learning has attracted widespread attention in the field of machine learning [49]. In contrast to supervised learning, self-supervised learning takes advantage of image characteristics (such as position, color, and texture) without manually annotating labels [50]. Studies have demonstrated that good representations can be learned from predicting context or playing a jigsaw game on images [51, 52]. Furthermore, some contrastive-based self-supervised learning methods are effective on natural image datasets [53, 54, 55]. A Siamese network is used to maximize the similarity between the representations of views, with inputs defined as two augmented

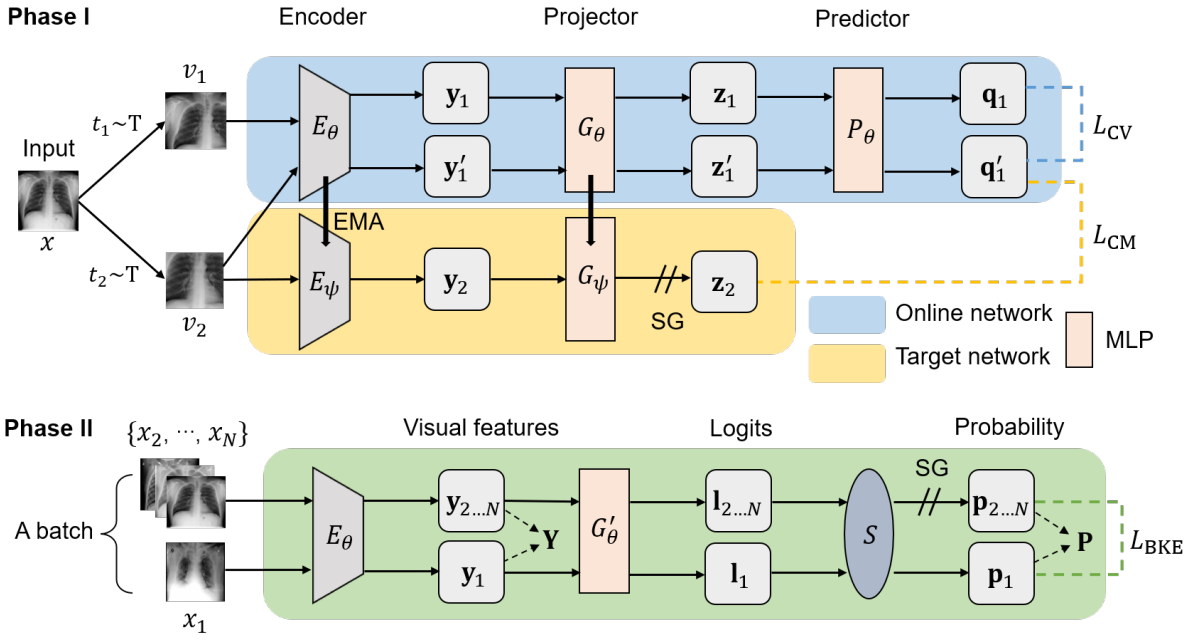


Figure 1: Overview of the proposed method. Our method consists of two phases, Phase I is self-supervised learning-based pretraining, and Phase II is batch knowledge ensembling-based fine-tuning. EMA represents exponential moving average, SG represents stop-gradient, MLP represents multilayer perceptron, and S represents softmax function.

views from one image [56, 57]. Compared with supervised learning, self-supervised learning without manually labeled annotations can learn fine-grained representations and is suitable for high-complexity CXR images [58, 59]. Therefore, we propose learning distinguished representations from COVID-19 CXR images using contrastive-based self-supervised learning as a pretraining phase.

2.3. Knowledge Ensembling

Generally, an ensemble of multiple networks produces better predictions than a single network [60]. By aggregating different networks, knowledge ensembling technologies generate robust supervision signals [61]. Knowledge ensembling even has been applied to semi-supervised [62] and self-supervised scenarios [54, 63]. However, the existing knowledge ensembling is limited to outputs of multi-networks, which is not applicable in low computation resource situations such as clinical use. Hence, we proposed a novel batch knowledge ensembling method, which utilizes the knowledge in the data of a training batch with only one network. With batch knowledge ensembling-based fine-tuning, our method can boost COVID-19 detection accuracy with a low-computing source.

3. COVID-19 Detection with Self-Supervised Learning and Batch Knowledge Ensembling

Figure 1 shows an overview of the proposed method. Our method consists of two phases: the first is a self-supervised learning-based fine-tuning phase for learning distinguished representations from CXR images, and the second is a batch knowledge ensembling-based fine-tuning phase for the accurate automatic detection of COVID-19. We show the details of the first and second phases in subsections 3.1 and 3.2, respectively.

3.1. Phase I: Self-Supervised Learning-based Pretraining

First, we introduce the self-supervised learning-based pretraining phase. Our self-supervised learning method uses an online network and a target network to learn distinguished representations from CXR images. Encoder E_θ, projector G_θ, and predictor P_θ belong to the online network. Encoder E_ψ and projector G_ψ belong to the target network. The transformations t_1 and t_2 for an input CXR image x are chosen at random from the distribution T to obtain a pair of views $v_1 = t_1(x_1)$ and $v_2 = t_2(x_2)$. Various augmentation methods are used in these transformations, including cropping, resizing, flipping, color jittering, and Gaussian blur.

Online network encoder E_θ and projector G_θ process the view v_1. Accordingly, target network encoder E_ψ and projector G_ψ process the view v_2. As a result, the output $\mathbf{z}_2$ represents the feature of v_2 in the target network. For cross-view loss calculations, we use v_2 to obtain a copy and input it into the online network. Subsequently, we transform two views into features $\mathbf{q}_1$ and $\mathbf{q}'_1$ using the predictor P_θ of the online network. The cross-view loss L_{CV} can be calculated as follows:

$$\begin{aligned} L_{\text{CV}} &= \|\hat{\mathbf{q}}_1 - \hat{\mathbf{q}}'_1\|_2^2 \\ &= 2 - 2 \cdot \frac{\langle \mathbf{q}_1, \mathbf{q}'_1 \rangle}{\|\mathbf{q}_1\|_2 \cdot \|\mathbf{q}'_1\|_2}, \end{aligned} \tag{1}$$

where $\hat{\mathbf{q}}_1 = \mathbf{q}_1/\|\mathbf{q}_1\|_2$ and $\hat{\mathbf{q}}'_1 = \mathbf{q}'_1/\|\mathbf{q}'_1\|_2$ represent the normalized features of v_1 and v_2 processed by the online network, respectively. The cross-model loss L_{CM} can be calculated as

follows:

$$L_{\mathrm{CM}} = \|\hat{\mathbf{q}}_1' - \hat{\mathbf{z}}_2\|_2^2 = 2 - 2 \cdot \frac{\left\langle \mathbf{q}_1', \mathbf{z}_2 \right\rangle}{\|\mathbf{q}_1'\|_2 \cdot \|\mathbf{z}_2\|_2}, \tag{2}$$

where $\hat{\mathbf{z}}_2 = \mathbf{z}_2/\|\mathbf{z}_2\|_2$ represents the normalized features of v_2 processed by the target network. The predictor P_θ is used only in the online network to prevent learning collapse [54]. By minimizing the total loss, the weights of the online network (θ) are updated. The total loss $\mathcal{L}_{\theta,\psi}$ and optimizing process are defined as follows:

$$\mathcal{L}_{\theta,\psi} = \mathcal{L}_{\mathrm{CV}} + \mathcal{L}_{\mathrm{CM}}, \tag{3}$$

$$\theta \leftarrow \mathrm{Opt}(\theta, \nabla_\theta \mathcal{L}_{\theta,\psi}, \alpha), \tag{4}$$

where Opt represents the optimizer and α represents the learning rate. Based on the weights of the online network (θ), the target network weights (ψ) are derived by multiplying them by the exponential moving average.

$$\psi \leftarrow \zeta\psi + (1-\zeta)\theta, \tag{5}$$

where ζ represents the degree of moving average. The gradient is not backpropagated through the target network for stable training [55].

After the self-supervised learning-based pretraining phase, the online network encoder E_θ learns distinguished representations from CXR images, and the parameters are saved for the next fine-tuning phase. Compared with supervised learning, self-supervised learning without manually labeled annotations can learn fine-grained representations and is suitable for high-complexity CXR images.

3.2. Phase II: Batch Knowledge Ensembling-based Fine-tuning

Next, we introduce the batch knowledge ensembling-based fine-tuning phase of our method. Combined knowledge from images with similar visual features could provide better classification performance because they tend to have similar predicted probabilities. The above concept can be used to improve COVID-19 detection performance according to the similarity of visual features between different CXR images in a batch.

First, we can obtain the visual feature similarity matrix $\mathbf{Y} \in \mathbb{R}^{N\times N}$ using the encoded visual features $\{\mathbf{y}_1, ..., \mathbf{y}_N\}$ in a batch (N images) as follows:

$$\mathbf{Y}_{i,j} = (\hat{\mathbf{y}}_i^\top \hat{\mathbf{y}}_j), \tag{6}$$

where $\hat{\mathbf{y}}_i = \mathbf{y}_i/\|\mathbf{y}_i\|_2$ represents the normalized features, and i, j represents the batch indices. To avoid self-knowledge reinforcement, we eliminate diagonal entries with the identity matrix $\mathbf{I}$ by $\mathbf{Y} = \mathbf{Y} \odot (1 - \mathbf{I})$. Then, we normalize the similarity matrix for visual features as follows:

$$\hat{\mathbf{Y}}_{i,j} = \frac{\exp(\mathbf{Y}_{i,j})}{\sum_{j\neq i} \exp(\mathbf{Y}_{i,j})}, \forall i \in \{1, ..., N\}. \tag{7}$$

By applying a projector G_θ' and a softmax function S to the output logits $\{\mathbf{l}_1, ..., \mathbf{l}_N\}$, we can obtain the predictive probabilities $\{\mathbf{p}_1, ..., \mathbf{p}_N\}$ as follows:

$$\mathbf{p}_{(k)} = \frac{\exp(\mathbf{l}_k/\tau)}{\sum_{i=1}^{K} \exp(\mathbf{l}_i/\tau)}, \tag{8}$$

where K represents the number of classes, and τ represents a temperature hyperparameter for the soft degree. The probability matrix of a batch of CXR images is predicted as $\mathbf{P} = [\mathbf{p}_1, ..., \mathbf{p}_N]^\top \in \mathbb{R}^{N\times K}$. We generate the soft targets $\mathbf{Q}$ as a weighted sum of the initial probability matrix $\mathbf{P}$ and the propagated probability matrix $\hat{\mathbf{Y}}\mathbf{P}$ to prevent propagating noisy predictions as follows:

$$\mathbf{Q} = \omega\hat{\mathbf{Y}}\mathbf{P} + (1-\omega)\mathbf{P}. \tag{9}$$

Furthermore, we propagate and ensemble multiple times to generate better soft targets $\mathbf{Q}$ for batch knowledge ensembling as follows:

$$\begin{aligned}\mathbf{Q}_{(t)} &= \omega\hat{\mathbf{Y}}\mathbf{Q}_{(t-1)} + (1-\omega)\mathbf{P}, \\ &= (\omega\hat{\mathbf{Y}})^t\mathbf{P} + (1-\omega)\sum_{i=0}^{t-1}(\omega\hat{\mathbf{Y}})^i\mathbf{P},\end{aligned} \tag{10}$$

where ω represents a weight factor, and t represents the t-th iteration. As the number of iterations approaches infinite, we obtain $\lim_{t\to\infty}(\omega\hat{\mathbf{Y}})^t = 0$ and $\lim_{t\to\infty}\sum_{i=0}^{t-1}(\omega\hat{\mathbf{Y}})^i = (\mathbf{I} - \omega\hat{\mathbf{Y}})^{-1}$. Based on the above observation, an approximate inference formulation can be calculated as follows:

$$\mathbf{Q} = (1-\omega)(\mathbf{I} - \omega\hat{\mathbf{Y}})^{-1}\mathbf{P}. \tag{11}$$

Finally, we define the batch knowledge ensembling loss L_{BKE} as follows:

$$L_{\mathrm{BKE}} = L_{\mathrm{CE}} + \lambda \cdot \tau^2 \cdot D_{\mathrm{KL}}(\mathbf{Q}\|\mathbf{P}), \tag{12}$$

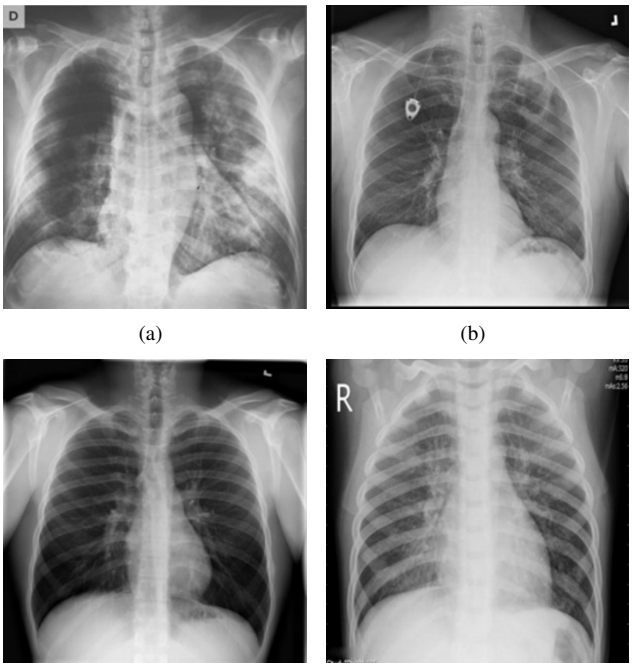
where L_{CE} represents the ordinary cross-entropy loss, D_{KL} represents Kullback-Leibler divergence, and λ represents a balance hyperparameter. There is no backpropagation of gradients through soft targets for stable training [17].

Our previous method [17] introduces batch knowledge ensembling into the self-supervised learning phase, making it memory costly and sensitive to hyperparameters. Unlike the previous implementation, we introduce batch knowledge ensembling into the fine-tuning phase, reducing the memory used in self-supervised learning and improving COVID-19 detection accuracy. According to the similarities of visual features in the different CXR images, the encoder E_θ can learn better representations and use them for the final COVID-19 detection.

4. Experiments

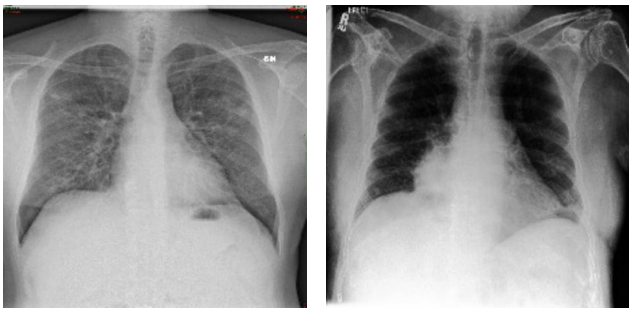
4.1. Dataset and Settings

The datasets used in our study include the large COVID-19 CXR dataset [64] and COVID5K dataset [12], with COVID5K



(a) (b) (c) (d)

Figure 2: Examples of CXR images in the large COVID-19 CXR dataset [64]: (a) COVID-19, (b) Lung Opacity (c) Normal, and (d) Viral Pneumonia.



(a) (b)

Figure 3: Examples of CXR images in the COVID5K dataset [12]: (a) COVID-19 , (b) Normal.

being an unbalanced dataset. Table 1 shows that the COVID-19 CXR dataset has four categories, and the ratio of data in the training and test sets is 8:2.[2] As presented in Table 2, the unbalanced dataset has two classes with a total number of 5,520 images, where only 520 are COVID-19 images.[3] Figures 2 and 3 show some examples of CXR images from the two datasets. Each image is grayscale and resized to a resolution of 224 pixels by 224 pixels. The areas under the receiver operating characteristic curve (AUC), sensitivity (Sen), specificity (Spe), harmonic mean (HM) of Sen and Spe, and classification accuracy (Acc) were used as evaluation metrics. COVID-19 was considered positive for Sen, Spe, HM, and AUC, whereas the others were considered negative.

[2]https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

[3]https://github.com/shervinmin/DeepCovid

Table 1: Details of the large COVID-19 CXR dataset [64].

Class	Full	Training set	Test set
COVID-19	3,616	2,893	723
Lung Opacity	6,012	4,810	1,202
Normal	10,192	8,154	2,038
Viral Pneumonia	1,345	1,076	269

Table 2: Details of the COVID5K dataset [12].

Class	Full	Training set	Test set
COVID-19	520	420	100
Normal	5,000	2,000	3,000

We used ResNet18 or ResNet50 [13] as an encoder and stochastic gradient descent as an optimizer. The projectors and predictor are two-layer MLPs with the same structure as that used in [54]. After 40 epochs of self-supervised learning, 30 epochs of fine-tuning were performed on the datasets. The results are the average and variance of the last 10 fine-tuning epochs. In the self-supervised learning-based pretraining phase, the batch size, the generated view size, and the degree of moving average ζ were set to 256, 112, and 0.996, respectively [65]. Data augmentation methods such as cropping, resizing, flipping, and Gaussian blurring are used to generate random views. In the batch knowledge ensembling-based fine-tuning phase, the hyperparameters ω, N, λ, and τ were set to 0.5, 128, 8.0, and 1.0, respectively [17].

Several contrastive-based self-supervised learning methods were used, including BKE [17], Cross [66], BYOL [54], SimSiam [55], PIRL [67], and SimCLR [53]. Note that Cross [66] is our previous work that only considered self-supervised learning for COVID-19 detection. As masked image modeling-based self-supervised learning methods using vision transformer [68] have recently become a new trend, we also use two SOTA methods RGMIM [69] and MAE [70] for comparison. In our experiments, RGMIM and MAE used the ViT-Base model. In addition, we trained from scratch and used transfer learning methods as baselines. To test COVID-19 detection accuracy using a small amount of annotated data, we selected 1%, 10%, and 50% of the training set for the fine-tuning process, respectively. Note that we use the same selection ratio in each category.

4.2. Test Accuracy on the Large COVID-19 CXR Dataset

The test accuracy of COVID-19 detection on the training data is presented in Table 3. Specifically, when applying the ResNet50 model to all training data, transfer learning achieved HM, AUC, and Acc scores of 0.968, 0.997, and 0.936, respectively, and the best comparison method BKE [17] achieved HM, AUC, and Acc scores of 0.988, 0.999, and 0.957, respectively. On the other hand, our method achieved HM, AUC, and Acc scores of 0.994, 1.000, and 0.966, respectively, on the large COVID-19 CXR dataset.

Figure 4 depicts the best performance confusion matrix of our method. Our method not only discriminates well between patients with COVID-19 and normal patients but also

Table 3: Test accuracy on the large COVID-19 CXR Dataset.

Method	Structure	Sen	Spe	HM	AUC	Acc
Ours	ResNet50	**0.989±0.000**	**1.000±0.000**	**0.994±0.000**	**1.000±0.000**	**0.966±0.000**
BKE		0.980±0.004	0.997±0.001	0.988±0.002	0.999±0.000	0.957±0.001
Cross		0.972±0.003	0.997±0.001	0.985±0.001	0.999±0.000	0.953±0.001
BYOL		0.973±0.004	0.996±0.001	0.985±0.002	0.999±0.000	0.954±0.001
SimSiam		0.974±0.004	0.995±0.001	0.984±0.002	0.998±0.000	0.950±0.001
PIRL-Jigsaw		0.977±0.003	0.997±0.001	0.987±0.001	0.999±0.000	0.951±0.001
PIRL-Rotation		0.973±0.002	0.997±0.001	0.985±0.001	0.999±0.000	0.951±0.001
SimCLR		0.913±0.006	0.994±0.001	0.952±0.003	0.996±0.000	0.936±0.001
Transfer		0.944±0.004	0.994±0.001	0.968±0.002	0.997±0.000	0.936±0.001
From Scratch		0.665±0.013	0.954±0.003	0.783±0.008	0.935±0.001	0.774±0.002
Ours	ResNet18	**0.982±0.000**	**1.000±0.000**	**0.994±0.000**	**1.000±0.000**	**0.960±0.001**
BKE		0.972±0.004	0.998±0.000	0.985±0.002	1.000±0.000	0.951±0.001
Cross		0.944±0.003	0.990±0.001	0.967±0.001	0.996±0.000	0.934±0.002
BYOL		0.934±0.007	0.990±0.002	0.961±0.003	0.995±0.000	0.932±0.001
SimSiam		0.940±0.002	0.988±0.001	0.963±0.001	0.996±0.000	0.929±0.001
PIRL-Jigsaw		0.931±0.004	0.992±0.001	0.961±0.002	0.997±0.000	0.930±0.001
PIRL-Rotation		0.936±0.007	0.994±0.001	0.964±0.003	0.997±0.000	0.930±0.001
SimCLR		0.806±0.012	0.982±0.001	0.886±0.007	0.978±0.000	0.903±0.002
Transfer		0.900±0.008	0.981±0.003	0.939±0.003	0.993±0.000	0.909±0.001
From Scratch		0.849±0.010	0.958±0.004	0.900±0.004	0.974±0.000	0.831±0.001

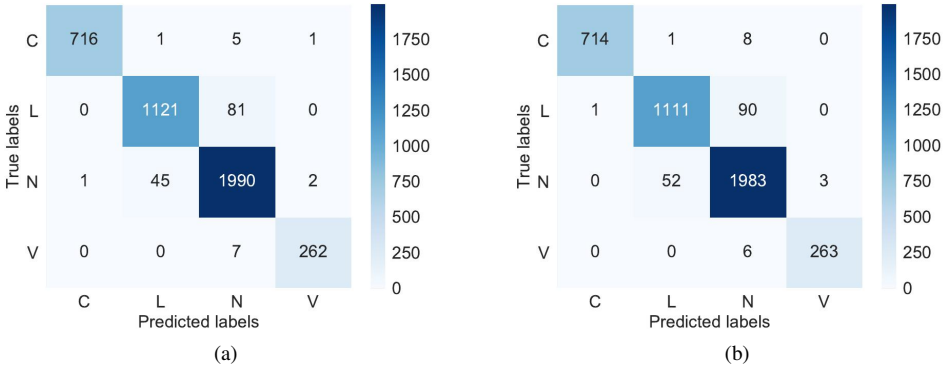


Figure 4: Best performance confusion matrix of our method. (a): ResNet50, (b): ResNet18.

achieves very high accuracy in identifying COVID-19 and other pneumonia. The test results in different settings show that our method achieves promising detection results on the large COVID-19 CXR dataset, and our method significantly outperforms other comparison methods, improving COVID-19 detection performance. Different from the previous method BKE [17], we introduce batch knowledge ensembling into the fine-tuning phase, reducing the computation cost and required memory. Therefore, our method can run on a single NVIDIA Tesla P100 GPU with a memory of 16G, whereas the previous method necessitates two GPUs. In addition, the training times of our method and BKE are approximately 97 and 124 min, respectively. Our method is faster and more efficient than the previous method.

The COVID-19 detection results for different annotated data volumes are presented in Table 4 and Fig. 5. The table and figure show that, compared with other comparison methods, our method vastly improved COVID-19 detection in a small amount of annotated data situations, such as 1% and 10% of the training set (169 and 1,693 images), and achieved promising detection performance even for 10% of the training set. Compared with the vision transformer-based methods RGMIM [69] and MAE [70], although we use the traditional and straightforward ResNet model, our method outperformed them, especially when the amount of annotated data was significantly reduced. In the real world, COVID-19 may have limited annotated training data because of the varying infection status, medical resources, and data-sharing policies of different countries [71]. However, the proposed method can still be applied to this case for high-performance automatic COVID-19 detection.

Table 4: Test accuracy in different annotated data volumes when compared with vision transformer-based methods.

Method	Structure	1%	10%	50%	100%
Ours	ResNet50	**0.859**	**0.934**	**0.960**	**0.966**
Ours	ResNet18	0.811	0.925	0.952	0.960
RGMIM	ViT-Base	0.771	0.919	0.957	0.962
MAE	ViT-Base	0.754	0.903	0.948	0.956
Transfer	ViT-Base	0.689	0.893	0.940	0.953
From Scratch	ViT-Base	0.413	0.645	0.810	0.848

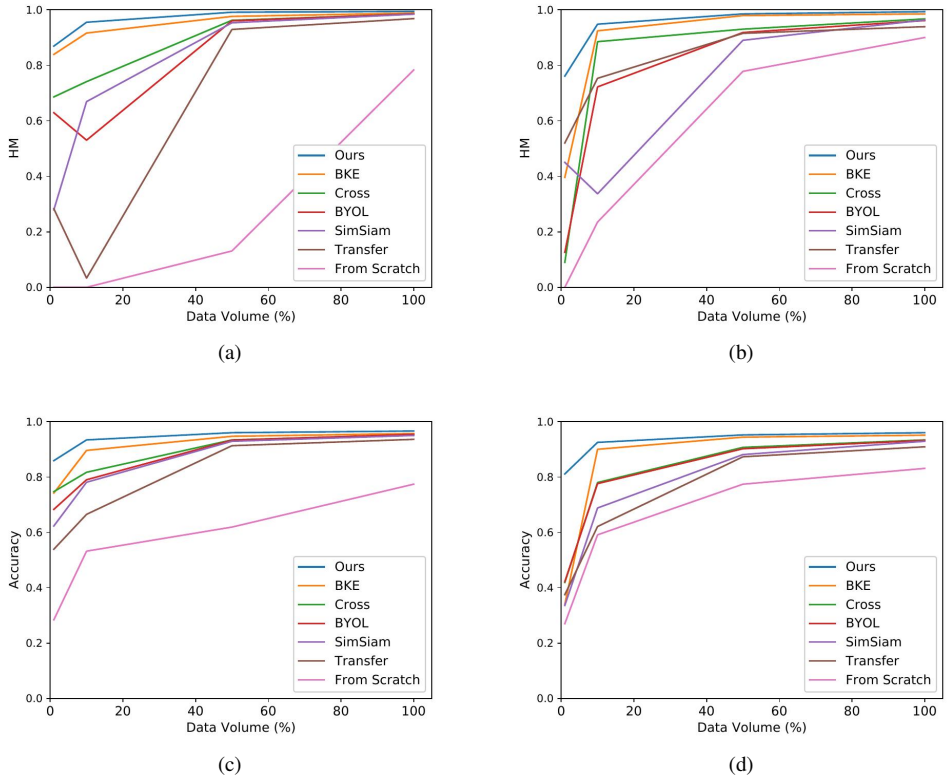


Figure 5: Test accuracy in different annotated data volumes: (a) HM of ResNet50, (b) HM of ResNet18, (c) Accuracy of ResNet50, and (d) Accuracy of ResNet18.

4.3. Test Accuracy on the COVID5K Dataset

The test accuracy of the COVID-19 detection on the training data and confusion matrices of our method are presented in Table 5 and Fig. 6. The results show the average and variance of the last 10 fine-tuning epochs. Specifically, when using all training data and the ResNet50 structure, transfer learning achieved Sen, Spe, and HM scores of 0.961, 0.908, and 0.934, respectively, and the best comparison method Cross [66] achieved Sen, Spe, and HM scores of 0.999, 0.925, and 0.960, respectively. On the other hand, our method achieved Sen, Spe, and HM scores of 0.990, 0.971, and 0.980, respectively, on the unbalanced COVID-19 CXR dataset. Our method achieved promising detection results on the unbalanced COVID-19 CXR dataset, which shows its robustness and that it can be used in extreme data situations in the real world.

4.4. Exploring the Impact of Hyperparameters on Experimental Results

The evaluation results of different hyperparameters in the batch knowledge ensembling-based fine-tuning phase are shown in Tables 6 and 7. For ensembling, ω accounts for the propagation of knowledge between the anchor CXR image and other images within the same batch. As ω increases, the refined soft targets gain more information from other samples. We studied the effects of the ensembling weight ω by changing it from 0.1 to 0.9. We observed that our method is insensitive to the ensembling weight. Furthermore, when ω became larger, the HM scores decreased, but the accuracy increased, which indicated a greater bias toward the correct detection of normalcy and other pneumonia.

Because our method uses the category information of different CXR images in a batch, we investigated the effects of the batch size by changing it from 32 to 512. As shown in the table, our method achieved the best results when the batch size

Table 5: Test accuracy on the COVID5K dataset.

Method	Structure	Sen	Spe	HM	AUC
Ours	ResNet50	0.990±0.000	0.971±0.004	**0.980±0.002**	**0.997±0.000**
BKE		0.926±0.013	**0.989±0.001**	0.957±0.007	0.995±0.000
Cross		**0.999±0.005**	0.925±0.016	0.960±0.008	0.995±0.000
Transfer		0.961±0.010	0.908±0.013	0.934±0.005	0.984±0.001
From Scratch		0.818±0.026	0.916±0.016	0.864±0.009	0.930±0.002
Ours	ResNet18	0.958±0.004	**0.988±0.002**	**0.973±0.002**	**0.989±0.000**
BKE		0.939±0.003	0.973±0.002	0.955±0.002	**0.989±0.000**
Cross		**0.970±0.000**	0.946±0.007	0.958±0.004	0.987±0.000
Transfer		0.910±0.016	0.987±0.002	0.947±0.008	0.976±0.000
From Scratch		0.895±0.007	0.978±0.002	0.935±0.003	0.956±0.001

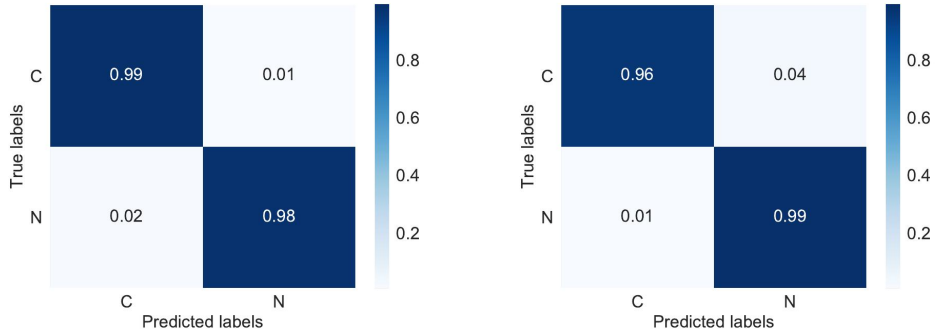


Figure 6: Best performance confusion matrix of our method. (a): ResNet50, (b): ResNet18.

Table 6: Evaluation results on the changes of the ensembling weight ω and the batch size N.

ω	HM	Acc
0.1	0.996	0.964
0.3	0.996	0.967
0.5	0.994	0.966
0.7	0.994	0.967
0.9	0.993	0.967

N	HM	Acc
32	0.993	0.961
64	0.994	0.963
128	0.994	0.966
256	0.993	0.962
512	0.992	0.964

Table 7: Evaluation results on the changes of the temperature τ and the weighting factor λ.

τ	HM	Acc
2	0.994	0.964
4	0.994	0.965
8	0.994	0.966
16	0.994	0.964

λ	HM	Acc
0.5	0.994	0.963
1	0.994	0.966
2	0.994	0.963
4	0.992	0.960

was 128. The predicted logits and soft targets are scaled using τ. By increasing τ, the probability distribution over classes becomes smoother. A temperature change of τ from 2.0 to 16.0 was used to investigate the effect of the temperature value. Our method is insensitive to temperature and achieved the best results when τ was set to 8.0. A λ value is adopted to balance the cross-entropy loss and batch knowledge ensembling loss, which is generally set to 1.0. We investigate the effects of the λ value by varying it from 0.5 to 4.0. As shown in the table, our method achieved the best results when λ was set to 1.0.

In the real world, because there are differences in shooting equipment and patients in different regions or countries, if the model is very sensitive to changes in hyperparameters, a lot of time and money will likely be wasted on hyperparameter adjustments [72]. However, from the evaluation results of different hyperparameters, the proposed method is insensitive to changes in hyperparameters, which shows its potential to be used in a real-world clinical situation.

4.5. Performance Comparison with Existing Methods

The performance comparison with existing methods for COVID-19 detection from CXR images is presented in Table 8. As shown in Table 8, although these methods have achieved relatively high detection accuracy [73, 74, 75, 76, 77, 78, 79, 80], these studies have typically been evaluated on small COVID-19 CXR image datasets with only two or three classes and may have limitations when used in real clinical situations. However, our method was evaluated on a large COVID-19 CXR dataset with 4 classes and 3,616 COVID-19 images and achieved promising detection performance. In addition, our method uses only the most used simple structure, ResNet50, which has advantages in terms of reliability and practicality.

Table 8: Performance comparison with the existing methods.

Method	Structure	Dataset	Accuracy
Narin et al. [73]	Inception-ResNetV2	COVID-19: 50, Normal: 50	Two-class: 0.980
Waheed et al. [74]	Auxiliary Classifier Generative Adversarial Network	COVID-19: 403, Normal: 721	Two-class: 0.950
Ozturk et al. [75]	DarkCovidNet	COVID-19: 127, Normal: 500	Two-class: 0.981
Zhang et al. [76]	ResNet34	COVID-19: 189, Normal: 235, Viral Pneumonia: 63	Three-class: 0.911
Togacar et al. [77]	Stacked models: MobileNetV2, SqueenzeNet, SVM	COVID-19: 295, Normal: 65, Viral Pneumonia: 98	Three-class: 0.993
Gianchandani et al. [78]	Ensemble models: VGG16, ResNet152, DenseNet201	COVID-19: 423, Normal: 1,579, Viral Pneumonia: 1,485	Three-class: 0.962
Wang et al. [79]	COVID-Net	COVID-19: 358, Normal: 8,066, Viral Pneumonia: 5,538	Three-class: 0.933
Gour et al. et al. [80]	UA-ConvNet	COVID-19: 219, Normal: 1,341, Viral Pneumonia: 1,345	Three-class: 0.988
Ours	ResNet50	COVID-19: 3,616, Normal: 10,192, Viral Pneumonia: 1,345, Lung Opacity: 6,012	Four-class: 0.966

5. Discussion

Using DL for computer-aided detection can reduce the burden on healthcare systems [81, 82, 83]. On the large and unbalanced COVID-19 CXR datasets, our method exhibited good automatic COVID-19 detection performance. Significantly, when using a small amount of annotated training data for fine-tuning, our method outperformed the other SOTA methods. Because of the entirely different infection status, the number of medical resources, and data sharing policies of COVID-19 in other countries and cities, there is a likelihood of limited annotated training data [84, 85]. Nevertheless, the proposed method can still be applied to this case for high-performance COVID-19 detection, which makes our method stands out above the pool of studies on DL for COVID-19 detection.

Our method also has limitations. For example, because our method is contrastive-based self-supervised learning and is designed on CNNs, we have not yet figured out how to migrate it to new vision transformer-based structures [86, 87, 88], which will be one of our future studies. In our experiments, we want to explore the impact and robustness of the initial parameters on different fine-tuning stages and data volumes. Hence, we use the average of the last 10 fine-tuning epochs to test the performance. Also, since we evaluate model performance in different subsets of settings (i.e., 1%, 10%, and 50%), it is expensive to perform N-fold cross-validation in all settings. However, N-fold cross-validation is a more common method, which we will consider in our future work. For the medical AI field, since there is already a trend to shift from only using AI-driven models to the Internet of Medical Things (IoMT) enabled systems [89], one of our future directions is to apply our algorithms to such systems, which can perform high-efficiency real-time COVID-19 detection. Furthermore, the ethical and privacy issues related to medical data sharing have been the main challenges for computer-aided detection systems [90]. However, our related studies about medical data distillation [91, 92, 93, 94] can improve the effectiveness and security of medical data sharing among different medical facilities, which fits well with the proposed method and is expected to be applied in clinical situations.

6. Conclusion

We have proposed a novel automatic COVID-19 detection method with self-supervised learning and batch knowledge ensembling using CXR images. Self-supervised learning-based pretraining can learn distinguished representations from CXR images without manually annotated labels. Furthermore, batch knowledge ensembling-based fine-tuning can utilize category knowledge of images in a batch according to their visual feature similarities to boost detection performance. On two public COVID-19 CXR datasets, including a large dataset and an unbalanced dataset, our method exhibited promising COVID-19 detection performance.

In the real world, COVID-19 may have limited annotated

training data because of the varying infection status, medical resources, and data-sharing policies of different countries. The proposed method can still be applied to this case for high-performance automatic COVID-19 detection. Also, there are differences in shooting equipment and patients in different regions or countries, if the model is very sensitive to changes in hyperparameters, a lot of time and money will likely be wasted on hyperparameter adjustments. The proposed method is insensitive to changes in hyperparameters, which shows its potential to be used in a real-world clinical situation. Our method can reduce the workloads of healthcare providers and radiologists.

Despite the promising experimental results, the proposed method should be tested on other COVID-19 CXR image datasets and different image modalities (*e.g.*, CT and Ultrasound) for any potential bias. Furthermore, comparing with traditional diagnostic methods, such as PCR tests and clinical assessments will be one of our future works. Also, exploring the utilization of the proposed method on transformer-based structures is a potential research direction.

Declaration of competing interest

None declared.

Acknowledgments

This study was partly supported by the MEXT Doctoral program for Data-Related InnoVation Expert Hokkaido University (D-DRIVE-HU) program, the Hokkaido University-Hitachi Collaborative Education and Research Support Program and AMED Grant Number JP21zf0127004. This study was conducted at the Data Science Computing System of Education and Research Center for Mathematical and Data Science, Hokkaido University.

References

[1] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of sars-cov-2, Nature Medicine 26 (4) (2020) 450–452 (2020).

[2] E. Pujadas, N. Ibeh, M. M. Hernandez, A. Waluszko, T. Sidorenko, V. Flores, B. Shiffrin, N. Chiu, A. Young-Francois, M. D. Nowak, et al., Comparison of sars-cov-2 detection from nasopharyngeal swab samples by the roche cobas 6800 sars-cov-2 test and a laboratory-developed real-time rt-pcr test, Journal of Medical Virology 92 (9) (2020) 1695–1698 (2020).

[3] M. Dramé, M. T. Teguo, E. Proye, F. Hequet, M. Hentzien, L. Kanagaratnam, L. Godaert, Should rt-pcr be considered a gold standard in the diagnosis of covid-19?, Journal of Medical Virology (2020).

[4] Q. Li, H. Wang, X. Li, Y. Zheng, Y. Wei, P. Zhang, Q. Ding, J. Lin, S. Tang, Y. Zhao, et al., The role played by traditional chinese medicine in preventing and treating covid-19 in china, Frontiers of Medicine 14 (5) (2020) 681–688 (2020).

[5] G. D. Rubin, C. J. Ryerson, L. B. Haramati, N. Sverzellati, J. P. Kanne, S. Raoof, N. W. Schluger, A. Volpi, J.-J. Yim, I. B. Martin, et al., The role of chest imaging in patient management during the covid-19 pandemic: a multinational consensus statement from the fleischner society, Radiology 296 (1) (2020) 172–180 (2020).

[6] H. Shi, X. Han, N. Jiang, Y. Cao, O. Alwalid, J. Gu, Y. Fan, C. Zheng, Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study, The Lancet Infectious Diseases 20 (4) (2020) 425–434 (2020).

[7] Y. Chen, Y. Lin, X. Xu, J. Ding, C. Li, Y. Zeng, W. Liu, W. Xie, J. Huang, Classification of lungs infected covid-19 images based on inception-resnet, Computer Methods and Programs in Biomedicine 225 (2022) 107053 (2022).

[8] A. Shoeibi, M. Khodatars, R. Alizadehsani, N. Ghassemi, M. Jafari, P. Moridian, A. Khadem, D. Sadeghi, S. Hussain, A. Zare, et al., Automated detection and forecasting of covid-19 using deep learning techniques: A review, arXiv preprint arXiv:2007.10785 (2020).

[9] A. Chaddad, L. Hassan, C. Desrosiers, Deep radiomic analysis for predicting coronavirus disease 2019 in computerized tomography and x-ray images, IEEE Transactions on Neural Networks and Learning Systems 33 (1) (2021) 3–11 (2021).

[10] S. Kumar, M. K. Chaube, S. H. Alsamhi, S. K. Gupta, M. Guizani, R. Gravina, G. Fortino, A novel multimodal fusion framework for early diagnosis and accurate classification of covid-19 patients using x-ray images and speech signal processing techniques, Computer methods and programs in biomedicine 226 (2022) 107109 (2022).

[11] R. Gulakala, B. Markert, M. Stoffel, Rapid diagnosis of covid-19 infections by a progressively growing gan and cnn optimization, Computer Methods and Programs in Biomedicine (2022) 107262 (2022).

[12] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G. J. Soufi, Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning, Medical Image Analysis 65 (2020) 101794 (2020).

[13] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778 (2016).

[14] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size, in: Proceedings of the International Conference on Learning Representations (ICLR), 2018, pp. 1–13 (2018).

[15] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708 (2017).

[16] A. M. Ismael, A. Şengür, Deep learning approaches for covid-19 detection based on chest x-ray images, Expert Systems with Applications 164 (2021) 114054 (2021).

[17] G. Li, R. Togo, T. Ogawa, M. Haseyama, Self-knowledge distillation based self-supervised learning for covid-19 detection from chest x-ray images, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 1371–1375 (2022).

[18] N. Subramanian, O. Elharrouss, S. Al-Maadeed, M. Chowdhury, A review of deep learning-based detection methods for covid-19, Computers in Biology and Medicine (2022) 105233 (2022).

[19] M. A. Mohammed, B. Al-Khateeb, M. Yousif, S. A. Mostafa, S. Kadry, K. H. Abdulkareem, B. Garcia-Zapirain, Novel crow swarm optimization algorithm and selection approach for optimal deep learning covid-19 diagnostic model, Computational Intelligence and Neuroscience 2022 (2022).

[20] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, M. A. Rahman, Q. Wang, S. Qi, F. Kong, X. Zhu, X. Zhao, Identification of covid-19 samples from chest x-ray images using deep learning: A comparison of transfer learning approaches, Journal of X-ray Science and Technology 28 (5) (2020) 821–839 (2020).

[21] M. M. Taresh, N. Zhu, T. A. A. Ali, A. S. Hameed, M. L. Mutar, Transfer learning to detect covid-19 automatically from x-ray images using convolutional neural networks, International Journal of Biomedical Imaging 2021 (2021).

[22] W. Jin, S. Dong, C. Dong, X. Ye, Hybrid ensemble model for differential diagnosis between covid-19 and common viral pneumonia by chest x-ray radiograph, Computers in Biology and Medicine 131 (2021) 104252 (2021).

[23] A. U. Ibrahim, M. Ozsoz, S. Serte, F. Al-Turjman, P. S. Yakoi, Pneumonia classification using deep learning from chest x-ray images during covid-19, Cognitive Computation (2021) 1–13 (2021).

[24] A. Umar Ibrahim, M. Ozsoz, S. Serte, F. Al-Turjman, S. Habeeb Kolapo, Convolutional neural network for diagnosis of viral pneumonia and covid-19 alike diseases, Expert Systems 39 (10) (2022) e12705 (2022).

[25] A. T. Nagi, M. J. Awan, M. A. Mohammed, A. Mahmoud, A. Majumdar, O. Thinnukool, Performance analysis for covid-19 diagnosis using custom and state-of-the-art deep learning models, Applied Sciences 12 (13)

(2022) 6364 (2022).

[26] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), 2012, pp. 1097–1105 (2012).

[27] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Proceedings of the International Conference on Learning Representations (ICLR), 2015, pp. 1–14 (2015).

[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818–2826 (2016).

[29] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, et al., Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, arXiv preprint arXiv:1711.05225 (2017).

[30] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, Q. V. Le, Mnasnet: Platform-aware neural architecture search for mobile, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2820–2828 (2019).

[31] J. N. Hasoon, A. H. Fadel, R. S. Hameed, S. A. Mostafa, B. A. Khalaf, M. A. Mohammed, J. Nedoma, Covid-19 anomaly detection and classification method based on supervised machine learning of chest x-ray images, Results in Physics 31 (2021) 105045 (2021).

[32] M. A. Mohammed, K. H. Abdulkareem, B. Garcia-Zapirain, S. A. Mostafa, M. S. Maashi, A. S. Al-Waisy, M. A. Subhi, A. A. Mutlag, D.-N. Le, A comprehensive investigation of machine learning feature extraction and classification methods for automated diagnosis of covid-19 based on x-ray images, Computers, Materials & Continua 66 (2021) 3289–3310 (2021).

[33] J. Gayathri, B. Abraham, M. Sujarani, M. S. Nair, A computer-aided diagnosis system for the classification of covid-19 and non-covid-19 pneumonia on chest x-ray images by integrating cnn with sparse autoencoder and feed forward neural network, Computers in Biology and Medicine 141 (2022) 105134 (2022).

[34] M. F. Aslan, K. Sabanci, A. Durdu, M. F. Unlersen, Covid-19 diagnosis using state-of-the-art cnn architecture features and bayesian optimization, Computers in Biology and Medicine (2022) 105244 (2022).

[35] M. A. Mohammed, M. S. Maashi, M. Arif, M. K. Nallapaneni, O. Geman, Intelligent systems and computational methods in medical and healthcare solutions with their challenges during covid-19 pandemic, Journal of Intelligent Systems 30 (1) (2021) 976–979 (2021).

[36] A. Ben-Hur, J. Weston, A user's guide to support vector machines, in: Data Mining Techniques for the Life Sciences, Springer, 2010, pp. 223–239 (2010).

[37] J. Gou, H. Ma, W. Ou, S. Zeng, Y. Rao, H. Yang, A generalized mean distance-based k-nearest neighbor classifier, Expert Systems with Applications 115 (2019) 356–372 (2019).

[38] G. I. Webb, E. Keogh, R. Miikkulainen, Naïve bayes., Encyclopedia of Machine Learning 15 (2010) 713–714 (2010).

[39] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, S. D. Brown, An introduction to decision tree modeling, Journal of Chemometrics: A Journal of the Chemometrics Society 18 (6) (2004) 275–285 (2004).

[40] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of relieff and rrelieff, Machine Learning 53 (1) (2003) 23–69 (2003).

[41] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, A. Madabhushi, Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images, IEEE Transactions on Medical Imaging 35 (1) (2015) 119–130 (2015).

[42] S. Tiwari, A. Jain, Convolutional capsule network for covid-19 detection using radiography images, International Journal of Imaging Systems and Technology 31 (2) (2021) 525–539 (2021).

[43] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, A. Mohammadi, Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images, Pattern Recognition Letters 138 (2020) 638–643 (2020).

[44] S. Toraman, T. B. Alakus, I. Turkoglu, Convolutional capsnet: A novel artificial neural network approach to detect covid-19 disease from x-ray images using capsule networks, Chaos, Solitons & Fractals 140 (2020) 110122 (2020).

[45] Z. Lin, Z. He, S. Xie, X. Wang, J. Tan, J. Lu, B. Tan, Aanet: Adaptive attention network for covid-19 detection from chest x-ray images, IEEE Transactions on Neural Networks and Learning Systems 32 (11) (2021) 4781–4792 (2021).

[46] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen, et al., Viral pneumonia screening on chest x-rays using confidence-aware anomaly detection, IEEE Transactions on Medical Imaging 40 (3) (2020) 879–890 (2020).

[47] A. Shamsi, H. Asgharnezhad, S. S. Jokandan, A. Khosravi, P. M. Kebria, D. Nahavandi, S. Nahavandi, D. Srinivasan, An uncertainty-aware transfer learning-based framework for covid-19 diagnosis, IEEE Transactions on Neural Networks and Learning Systems 32 (4) (2021) 1408–1417 (2021).

[48] S. Dong, Q. Yang, Y. Fu, M. Tian, C. Zhuo, Rconet: Deformable mutual information maximization and high-order uncertainty-aware learning for robust covid-19 detection, IEEE Transactions on Neural Networks and Learning Systems 32 (8) (2021) 3401–3411 (2021).

[49] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, J. Tang, Self-supervised learning: Generative or contrastive, IEEE Transactions on Knowledge and Data Engineering (2021).

[50] M. Minderer, O. Bachem, N. Houlsby, M. Tschannen, Automatic shortcut removal for self-supervised representation learning, in: International Conference on Machine Learning, PMLR, 2020, pp. 6927–6937 (2020).

[51] C. Doersch, A. Gupta, A. A. Efros, Unsupervised visual representation learning by context prediction, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2015, pp. 1422–1430 (2015).

[52] M. Noroozi, P. Favaro, Unsupervised learning of visual representations by solving jigsaw puzzles, in: Proceedings of the European Conference on Computer Vision (ECCV), 2016, pp. 69–84 (2016).

[53] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: Proceedings of the International Conference on Machine Learning (ICML), 2020, pp. 1597–1607 (2020).

[54] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al., Bootstrap your own latent-a new approach to self-supervised learning, in: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), 2020, pp. 21271–21284 (2020).

[55] X. Chen, K. He, Exploring simple siamese representation learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 15750–15758 (2021).

[56] Y. Tian, L. Yu, X. Chen, S. Ganguli, Understanding self-supervised learning with dual deep networks, arXiv preprint arXiv:2010.00578 (2020).

[57] O. Pantazis, G. J. Brostow, K. E. Jones, O. Mac Aodha, Focus on the positives: Self-supervised learning for biodiversity monitoring, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 10583–10592 (2021).

[58] Z. Zhou, V. Sodha, M. M. R. Siddiquee, R. Feng, N. Tajbakhsh, M. B. Gotway, J. Liang, Models genesis: Generic autodidactic models for 3d medical image analysis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2019, pp. 384–393 (2019).

[59] S. Azizi, B. Mustafa, F. Ryan, Z. Beaver, J. Freyberg, J. Deaton, A. Loh, A. Karthikesalingam, S. Kornblith, T. Chen, et al., Big self-supervised models advance medical image classification, arXiv preprint arXiv:2101.05224 (2021).

[60] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network, arXiv preprint arXiv:1503.02531 (2015).

[61] J. Gou, B. Yu, S. J. Maybank, D. Tao, Knowledge distillation: A survey, International Journal of Computer Vision 129 (2021) 1789–1819 (2021).

[62] A. Tarvainen, H. Valpola, Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, in: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 1–10 (2017).

[63] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 9729–9738 (2020).

[64] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan, et al., Exploring the effect of image enhancement techniques on covid-19 detec-

tion using chest x-ray images, Computers in Biology and Medicine 132 (2021) 104319 (2021).
[65] G. Li, R. Togo, T. Ogawa, M. Haseyama, Self-supervised learning for gastritis detection with gastric x-ray images, arXiv preprint arXiv:2104.02864 (2021).
[66] G. Li, R. Togo, T. Ogawa, M. Haseyama, Covid-19 detection based on self-supervised transfer learning using chest x-ray images, International Journal of Computer Assisted Radiology and Surgery (2022) 1–8 (2022).
[67] I. Misra, L. v. d. Maaten, Self-supervised learning of pretext-invariant representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 6707–6717 (2020).
[68] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: Transformers for image recognition at scale, in: ICLR, 2021, pp. 1–21 (2021).
[69] G. Li, R. Togo, T. Ogawa, M. Haseyama, Rgmim: Region-guided masked image modeling for covid-19 detection, arXiv preprint arXiv:2211.00313 (2022).
[70] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, R. Girshick, Masked autoencoders are scalable vision learners, in: CVPR, 2022, pp. 16000–16009 (2022).
[71] Y. N. Alhalaseh, H. A. Elshabrawy, M. Erashdi, M. Shahait, A. M. Abu-Humdan, M. Al-Hussaini, Allocation of the "already" limited medical resources amid the covid-19 pandemic, an iterative ethical encounter including suggested solutions from a real life encounter, Frontiers in Medicine 7 (2021) 1076 (2021).
[72] D. Zimmerer, F. Isensee, J. Petersen, S. Kohl, K. Maier-Hein, Unsupervised anomaly localization using variational auto-encoders, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2019, pp. 289–297 (2019).
[73] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, Pattern Analysis and Applications 24 (2021) 1207–1220 (2021).
[74] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, P. R. Pinheiro, Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection, IEEE Access 8 (2020) 91916–91923 (2020).
[75] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, U. R. Acharya, Automated detection of covid-19 cases using deep neural networks with x-ray images, Computers in Biology and Medicine 121 (2020) 103792 (2020).
[76] R. Zhang, Z. Guo, Y. Sun, Q. Lu, Z. Xu, Z. Yao, M. Duan, S. Liu, Y. Ren, L. Huang, et al., Covid19xraynet: a two-step transfer learning model for the covid-19 detecting problem based on a limited number of chest x-ray images, Interdisciplinary Sciences: Computational Life Sciences 12 (2020) 555–565 (2020).
[77] M. Toğaçar, B. Ergen, Z. Cömert, Covid-19 detection using deep learning models to exploit social mimic optimization and structured chest x-ray images using fuzzy color and stacking approaches, Computers in Biology and Medicine 121 (2020) 103805 (2020).
[78] N. Gianchandani, A. Jaiswal, D. Singh, V. Kumar, M. Kaur, Rapid covid-19 diagnosis using ensemble deep transfer learning models from chest radiographic images, Journal of Ambient Intelligence and Humanized Computing (2020) 1–13 (2020).
[79] L. Wang, Z. Q. Lin, A. Wong, Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, Scientific Reports 10 (1) (2020) 1–12 (2020).
[80] M. Gour, S. Jain, Uncertainty-aware convolutional neural network for covid-19 x-ray images classification, Computers in Biology and Medicine 140 (2022) 105047 (2022).
[81] S. Bhattacharya, P. K. R. Maddikunta, Q.-V. Pham, T. R. Gadekallu, C. L. Chowdhary, M. Alazab, M. J. Piran, et al., Deep learning and medical image processing for coronavirus (covid-19) pandemic: A survey, Sustainable Cities and Society 65 (2021) 102589 (2021).
[82] I. Feki, S. Ammar, Y. Kessentini, K. Muhammad, Federated learning for covid-19 screening from chest x-ray images, Applied Soft Computing 106 (2021) 107330 (2021).
[83] Q. Dou, T. Y. So, M. Jiang, Q. Liu, V. Vardhanabhuti, G. Kaissis, Z. Li, W. Si, H. H. Lee, K. Yu, et al., Federated deep learning for detecting covid-19 lung abnormalities in ct: a privacy-preserving multinational validation study, NPJ digital medicine 4 (1) (2021) 1–11 (2021).
[84] N. Peiffer-Smadja, R. Maatoug, F.-X. Lescure, E. D'ortenzio, J. Pineau, J.-R. King, Machine learning for covid-19 needs global collaboration and data-sharing, Nature Machine Intelligence 2 (6) (2020) 293–294 (2020).
[85] S. Latif, M. Usman, S. Manzoor, W. Iqbal, J. Qadir, G. Tyson, I. Castro, A. Razi, M. N. K. Boulos, A. Weller, et al., Leveraging data science to combat covid-19: A comprehensive review, IEEE Transactions on Artificial Intelligence 1 (1) (2020) 85–103 (2020).
[86] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022 (2021).
[87] S. Mehta, M. Rastegari, Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer, in: Proceedings of the International Conference on Learning Representations (ICLR), 2022, pp. 1–26 (2022).
[88] Y. Li, G. Yuan, Y. Wen, E. Hu, G. Evangelidis, S. Tulyakov, Y. Wang, J. Ren, Efficientformer: Vision transformers at mobilenet speed, in: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), 2022, pp. 1–19 (2022).
[89] A. Ghubaish, T. Salman, M. Zolanvari, D. Unal, A. Al-Ali, R. Jain, Recent advances in the internet-of-medical-things (iomt) systems security, IEEE Internet of Things Journal 8 (11) (2020) 8707–8718 (2020).
[90] T. Dhar, N. Dey, S. Borra, R. S. Sherratt, Challenges of deep learning in medical image analysis-improving explainability and trust, IEEE Transactions on Technology and Society (2023).
[91] G. Li, R. Togo, T. Ogawa, M. Haseyama, Soft-label anonymous gastric x-ray image distillation, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), 2020, pp. 305–309 (2020).
[92] G. Li, R. Togo, T. Ogawa, M. Haseyama, Compressed gastric image generation based on soft-label dataset distillation for medical data sharing, Computer Methods and Programs in Biomedicine 227 (2022) 107189 (2022).
[93] G. Li, R. Togo, T. Ogawa, M. Haseyama, Dataset distillation using parameter pruning, arXiv preprint arXiv:2209.14609 (2022).
[94] G. Li, R. Togo, T. Ogawa, M. Haseyama, Dataset distillation for medical dataset sharing, in: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), Workshop, 2023, pp. 1–6 (2023).