IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 15, 2021, accepted October 31, 2021, date of publication November 4, 2021, date of current version November 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3125324

Self-Supervised Deep Convolutional Neural Network for Chest X-Ray Classification

MATEJ GAZDA[1,2], JÁN PLAVKA[2], JAKUB GAZDA[3], AND PETER DROTÁR[1], (Member, IEEE)

[1]Intelligent Information Systems Laboratory, Faculty of Electrical Engineering and Informatics, Technical University of Košice, 04201 Košice, Slovakia
[2]Department of Mathematics and Theoretical Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, 04201 Košice, Slovakia
[3]2nd Department of Internal Medicine, Pavol Jozef Šafárik University and Louis Pasteur University Hospital, 04011 Košice, Slovakia

Corresponding author: Peter Drotár (peter.drotar@tuke.sk)

This work was supported in part by the Slovak Research and Development Agency under Contract APVV-16-0211, and in part by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under Contract VEGA 1/0327/20.

ABSTRACT Chest radiography is a relatively cheap, widely available medical procedure that conveys key information for making diagnostic decisions. Chest X-rays are frequently used in the diagnosis of respiratory diseases such as pneumonia or COVID-19. In this paper, we propose a self-supervised deep neural network that is pretrained on an unlabeled chest X-ray dataset. Pretraining is achieved through the contrastive learning approach by comparing representations of differently augmented input images. The learned representations are transferred to downstream tasks – the classification of respiratory diseases. We evaluate the proposed approach on two tasks for pneumonia classification, one for COVID-19 recognition and one for discrimination of different pneumonia types. The results show that our approach yields competitive results without requiring large amounts of labeled training data.

INDEX TERMS Self-supervised learning, contrastive learning, deep learning, convolutional neural network, chest X-ray, COVID-19.

I. INTRODUCTION

Medical imaging utilization has increased rapidly in recent decades, increasing by more than 50% for some modalities [1]. Although the rate of increase has slowed in recent years [2], medical imaging is still considered a significant diagnostic method.

Among all medical imaging modalities, radiography is cost effective and is frequently employed by hospitals, emergency services, and other medical facilities. Chest radiography (or chest X-ray (CXR)) is a painless, noninvasive, and powerful investigatory method that conveys crucial respiratory disease information. For respiratory diseases, CXR is a basal diagnostic tool. In CXR, pulmonary opacification (''the white lung field'') is the result of a decrease in the ratio of gas to soft tissue in the lung. Pulmonary opacification has several likely causes, including atelectasis, bronchogenic carcinoma, pleural effusion, tuberculosis, or bacterial or viral pneumonia.

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

A diagnosis of pneumonia is usually made after considering a combination of clinical symptoms (cough, fever, pathological respiratory sounds), laboratory results (white blood cell count, C-reactive protein and procalcitonin levels, blood gas analysis, and sputum culture), and the presence of pulmonary opacification on CXR. Although the diagnosis and treatment of pneumonia are straightforward in most cases, rapid and accurate diagnosis is specifically required in uncertain cases because complications resulting from an initial misdiagnosis may lead to prolonged hospitalization and resource-draining medical care. Pneumonia is one of the most frequent causes of death in patients of all ages [3] and accounts for a significant number of hospital admissions [4].

The recent outbreak of coronavirus disease 2019 (COVID-19) has ushered in unprecedented challenges for most medical facilities. The enormous number of infections calls not only for prevention but also for early diagnosis followed by effective treatment. In this scenario, chest radiography has proven to be one of the most time- and cost-effective tools for COVID-19 diagnosis [5]. Upon CXR, patients who suffer from COVID-19 pneumonia present a combination of different multifocal patterns of pulmonary opacification.

However, in contrast to community-acquired bacterial pneumonia, these changes are frequently bilateral. Furthermore, while the distribution of these changes is initially peripheral, during the course of the disease, they usually spread to other parts of the lung parenchyma as well. A recent study [6] showed that the correct diagnosis of mild and moderate COVID-19 from CXR is challenging, even for experienced radiologists. A shortage of medical personnel, the necessity of large numbers of diagnostic decisions even under unfavorable conditions, and the demand for quick and accurate medical decisions all mean that the need for computer-aided diagnostics is greater than ever.

In this study, we addressed the two most pronounced use cases of the CXR classification. Pneumonia is one of the most frequent and serious inflammatory conditions, and COVID-19 is currently inflicting a devastating impact on healthcare services and even economies in many states worldwide.

This study makes several contributions. First, we trained a deep convolutional neural network (CNN) in a self-supervised fashion using pseudolabels generated through random data augmentation on a large dataset of unlabeled CXR images. This goes beyond current state-of-the art techniques that rely on a large corpus of labeled data. Second, we proposed utilizing this pretrained CNN as a feature extractor for several downstream tasks aimed at CXR classification. Third, even though the proposed CXR classification network does not require large amounts of labeled data, it achieves performance levels competitive to those of supervised counterparts. Extensive experiments on COVID-19 and pneumonia detection tasks validate that the proposed model obtains very reasonable CXR feature representations; thus, it enables accurate CXR classifications on the four evaluated datasets.

The remainder of this paper is organized as follows. In Section II, we provide a brief overview of related works on CNN utilization for CXR classification and contrastive learning. In Section III, we introduce the proposed approach for a self-supervised CNN and its architecture. In Sections IV and V, we describe the datasets used in this study, present the experiments and report their results. Finally, after providing a discussion in Section VI, we draw conclusions in Section VII.

II. RELATED WORKS

In this section, we present related works on CXR classification and the methods involved in our work.

A. CONVOLUTIONAL NEURAL NETWORKS FOR CXR CLASSIFICATION

Two main enablers have emerged in the CXR classification domain in recent years. The most frequent limitation for successful pattern recognition in the biomedical imaging domain has always been (and still is) a scarcity of data. This problem was partially overcome with the introduction of transfer learning for CNNs. Fine-tuned CNNs have shown enormous potential and have even outperformed fully trained CNNs in many applications [7]. The second enabler involves data: several large CXR datasets have been made publicly available, permitting the utilization of many new methodologies for CXR classification [8], [9]. Most of the forthcoming works have taken advantage of one or both of these enablers.

CXR classification has drawn attention from the research community for several years, but the arrival of the COVID-19 pandemic has boosted interest in this topic. The majority of recent works on CXR classification focus solely or partially on COVID-19 classification.

Given recent findings, the most straightforward approach to diagnosing COVID-19 from CXR images is to use existing CNN architectures pretrained on ImageNet and fine-tune them on a target COVID-19 dataset. This was the approach employed by the authors of [10]. They fine-tuned four state-of-the-art convolutional networks (ResNet18, ResNet50, SqueezeNet, and DenseNet-121) to identify COVID-19. Similarly, Apostolopoulos and Mpesiana [11] evaluated five other CNN architectures pretrained on ImageNet and found that the most promising results were achieved by the VGG-19 architecture and the compact MobileNet network. Instead of utilizing a single CNN, some authors have proposed ensembles of CNNs for COVID-19 detection. Guarrasi *et al.* [12] employed Pareto-based multiobjective optimization to build a CNN ensemble, and Rajaraman *et al.* [13] showed that iterative pruning of the task-specific models not only improved prediction performance on the test data but also significantly reduced the number of trainable parameters.

Other authors have tried to optimize performance by designing a CNN architecture tailored to CXR classification. These models are inspired by existing architectures such as CoroNet [14], which was inspired by the Xception design, DarkCovidNet [15], which is based on the DarkNet [16] CNN, and COVID-CAPS [17], which capitalizes on the capsule networks that preserve the spatial information between images. Instead of using an existing architecture, Wang *et al.* [18] employed generative synthesis to develop COVID-Net—a machine-designed deep CNN. An interesting approach combining CNN with graph CNN was presented by Kumar *et al.* [19] and achieved classification accuracy as high as 97% on the COVIDx dataset. Other approaches rely on adding a class decomposition layer [20] to a pretrained CNN or a specific domain adaptation module to a fully convolutional network [21].

In addition to using only CNN to classify chest CXR, specific features have been proposed and extracted to boost the classification performance [22]. Moreover, the extracted features can be combined with CNN to provide more robust prediction [23].

Methodologies that do not focus on architecture improvements but rather attempt to improve pre- or postprocessing include the work of Heidari *et al.* on preprocessing X-ray images [24] or Morís *et al.* [25] on advanced data augmentation for improving COVID-19 screening.

Many of the above approaches have shown very promising results and achieved high classification accuracies. However, these methods must be considered with caution

because several additional aspects must be considered before accepting a particular design as a production-ready solution. First, many of the aforementioned studies combined the COVID-19 dataset for experiments with other publicly available datasets to create a dataset used for model training and testing. This increases the chance that the model provides an output based on not only disease-related features but also dataset-specific aspects, such as contrast and saturation. Second, some studies, such as [17] and [18], utilized only simple hold-out model validation. Criticisms of previous studies are detailed in [26], where the authors propose a more robust solution called COVID-SDNet and utilize a new dataset for model validation. Biases in released datasets used to train diagnostic systems are also discussed in [27].

Some authors have considered other practical aspects of CXR classification, such as the limited available datasets. Oh *et al.* [28] proposed a solution for overcoming the lack of sufficient training datasets based on a pretrained ResNet-18, which processed CXR images through smaller patches. The authors of [29] approached viral pneumonia detection as an anomaly detection problem. Using this approach, they were able to avoid training the model on large numbers of different pneumonia cases and focus solely on viral pneumonia. Recently, Luo *et al.* [30] proposed a framework to integrate the knowledge from different datasets and effectively trained a neural network to classify thoracic diseases.

To date, all previous approaches have relied on backbone networks pretrained on ImageNet. Transfer learning makes CNNs trained on large-scale natural images suitable for medical images. However, the disparities between natural images and X-ray images are quite significant. Training a CNN from scratch on a large X-ray dataset can further boost performance. Some early papers utilized self-supervised learning (SSL) [31], [32], confirming that this is a viable approach. More recently, the authors of [33] proposed a self-supervised approach guided by super sample decomposition and reported 99.8% accuracy. Additionally, Aviles-Rivero *et al.* [34] designed a graph-based deep semisupervised approach that needs only a very small labeled dataset and provides results competitive with supervised approaches.

COVID-19 detection from CXR images is a very hot research area, and new papers are appearing continuously. Covering all recent advances is outside the scope of this work. We have tried to mention different approaches and some representative cases of CXR classification; however, for more comprehensive reviews, we refer the interested reader to [35]–[40]. These reviews summarize recent works and provide multiple perspectives on the most recent advances in COVID-19 detection from CXRs.

B. CONTRASTIVE LEARNING OF VISUAL REPRESENTATIONS

Self-supervised neural networks provide unprecedented performance in computer vision tasks. Generative models operate mostly in the pixel space, which is computationally expensive and unsustainable on larger scales. On the other hand, contrastive discriminative methods operate on the augmented views of the same image, thus avoiding the computationally costly generation of the pixel space. In addition, contrastive discriminative methods currently achieve state-of-the-art performance on SSL tasks [41], [42]. Various approaches for self-supervised model training exist. The main paradigm has shifted towards instance discriminative models, where similar contrastive learning (SimCLR) [43], momentum contrast for unsupervised visual representation learning (MoCo) [44] and bootstrap your own latent architecture (BYOL) [45] have demonstrated as-yet-untapped potential. The representations learned by these architectures are on par with those of their supervised counterparts [46], [47].

From the point of view of pretext task selection, contrastive learning can be divided into context-instance contrast and context-context contrast [48]. The former tries to find relations between the local features and the global representation of an instance (i.e., wheels and windows to a car). We believe that the learned local features help to distinguish between the target classes. Some examples of pretext tasks working in the context-instance principle are a jigsaw puzzle [49] and rotation angle detection [50].

Context-context contrast architectures focus on the relationships between the global representations of different samples. CMC [41], MoCo [44], and SimCLR [43] contrast between positive and negative pairs, where the positive pairs constitute the same image augmented in different ways, while the negative pairs constitute all remaining images. The number of negative and positive pairs depends solely on the type of self-supervised architecture.

SimCLR and MoCo share the idea of using positive and negative pairs, but they differ in how the pairs are handled. In SimCLR, [44], negative pairs are processed within the batch; thus, SimCLR requires a larger batch size. MoCo's representations of negative keys are maintained in a separate queue encoded by a momentum encoder.

BYOL claims to achieve better results than SimCLR and MoCo without using negative samples in its loss function. Different from SimCLR and MoCo, BYOL employs an L_2 error loss function instead of contrastive loss while using a principle similar to the momentum encoder introduced in MoCo.

BYOL takes advantage of two neural networks called ''online'' and ''target'' networks that learn by interactions between each other. BYOL initializes the optimization step by including one augmented view of a single image. It teaches the online network to correctly predict the representation of a differently augmented view of the same image produced by the target network.

III. METHODS

A. SELF-SUPERVISED LEARNING

SSL is a subset of unsupervised learning methods that aim to learn meaningful representations from unlabeled data. The representations can then be reused for downstream (target) tasks as either a base for fine-tuning or as a fixed feature

extractor for models such as logistic regression, SVM, and many others. Because manually annotated labels are not available in the training data, the SSL's first step is to generate pseudolabels automatically through carefully selected pretext tasks.

Formally, SSL can be defined as the minimization of an objective function $J(\boldsymbol{\theta})$ parameterized by parameters $\boldsymbol{\theta} \in \mathbb{R}^d$, which represents the mean loss over all training samples:

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x}\sim\hat{p}_{\text{data}}}\mathcal{L}(m(\boldsymbol{x};\boldsymbol{\theta}), \pi(\boldsymbol{x})), \tag{1}$$

where $\hat{p}_{\text{data}}$ is an empirical distribution, $\mathcal{L}$ is the per-example loss function, $m(\cdot,\cdot)$ is the model prediction when the input is $\boldsymbol{x}$, and $\pi(.)$ is a function that returns pseudolabels for input $\boldsymbol{x}$ based on the pretext task.

Optimization of such a neural network is accomplished similarly to supervised learning – by updating the parameters $\boldsymbol{\theta}$ in the direction of the antigradient using methods based on stochastic gradient descent:

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta\frac{1}{B}\sum_{i=Bt+1}^{B(t+1)}\frac{\partial\mathcal{L}(m(\boldsymbol{x}_i;\boldsymbol{\theta}), \pi(\boldsymbol{x}_i))}{\partial\boldsymbol{\theta}}, \tag{2}$$

where $\mathcal{L}$ is a loss function of the i-th example from the batch sampled at time t, B stands for the batch size, and η stands for a hyperparameter called the learning rate.

Modern SSL designs decouple the neural network architecture from downstream tasks, which makes the transfer of knowledge more straightforward. State-of-the art SSLs such as SimCLR [43] and MoCo [44] use the ResNet50 [51] architecture on datasets such as CIFAR-10 [52] and ImageNet [53] just as their supervised counterparts do.

B. TRANSFER LEARNING

Despite recent advances in deep learning and hardware accessibility, neural network training still tends to be slow and resource intensive. The transfer of knowledge from one domain to another reduces these burdens and has proved effective in numerous applications [54], [55].

Transfer learning is applicable to tasks with different degrees of label availability. The knowledge extracted from a base domain can be acquired in an unsupervised, semisupervised, or supervised fashion. For unsupervised pretraining, transfer learning is defined as follows. Let $\mathcal{D}_S = (\mathcal{X}_S, \mathcal{P}_S)$ be a pretext dataset consisting of a set of samples $\mathcal{X}_S$ with corresponding pseudolabels $\mathcal{P}_S$ generated by the underlying pretext task and a downstream dataset $\mathcal{D}_T = (\mathcal{X}_T, \mathcal{Y}_T)$, where $\mathcal{X}_T$ denotes the set of training samples and $\mathcal{Y}_T$ denotes the set of true labels. Given example source datasets $\mathcal{D}_S$, pretext tasks $\mathcal{T}_S$, downstream datasets $\mathcal{D}_T$, and downstream tasks $\mathcal{T}_T$, transfer learning aims to reduce the loss function $\mathcal{L}$ of the model used for downstream tasks ($\mathcal{T}_T$) using the knowledge acquired from pretext tasks $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$.

C. PROPOSED APPROACH

We propose a neural network that solves pretext tasks based on contrastive instance discrimination similar to those used in SimCLR [43] and MoCo [56]. The data augmentation module creates two versions of each image in the current batch, thus creating positive and negative pairs.

A positive pair is formed by two versions of one image that were augmented differently. Conversely, two images are denoted as a negative pair when they are augmented from different base images.

A neural network learns to solve this task by comparing representations of positive and negative pairs. It discriminates between them by maximizing the agreement between representations of two instances of the same image.

The representations are compared by cosine similarity, which is defined for two vectors $\boldsymbol{u}, \boldsymbol{v}$ as follows:

$$\text{sim}(\boldsymbol{u}, \boldsymbol{v}) = \frac{\boldsymbol{u}^T\boldsymbol{v}}{||\boldsymbol{u}||_2 \cdot ||\boldsymbol{v}||_2}. \tag{3}$$

Other similarity functions, such as Euclidean distance or dot product, could also be employed.

The proposed learning architecture consists of three parts: a backbone neural network $m(.)$, a projection head $n(.)$, and a stochastic data augmentation module $\mathcal{A}$. The backbone network $m(.)$ is a ResNet50 Wide network that extracts representations from the augmented data examples. The projection head $n(.)$ (see Table 1) transforms the output from the backbone network into a latent space where contrastive loss is applied. The size of the output vector is a hyperparameter allowing the final size to be adjusted to properly reflect the size of the original image. The data augmentation module $\mathcal{A}$ is a module that returns random augmentations as follows: resized crop, horizontal flip, rotation, Gaussian blur and color jitter. A resized crop involves a random crop of the image followed by a resize back up to the original image size. The entire learning architecture is depicted in Fig. 1.

The optimization step is performed as follows. First, for a minibatch sampled at time t and consisting of images $\mathcal{X}_t = \{\boldsymbol{x}_{Bt+1}, \boldsymbol{x}_{Bt+2}, \ldots, \boldsymbol{x}_{Bt+B}\}$ of size B is drawn from the dataset samples $\mathcal{X}$, $\mathcal{X}_t \subseteq \mathcal{X}$, similar to supervised learning. Then, for each image in the minibatch, a positive pair is formed by augmenting the image twice with random augmentations, from which $2B$ images are obtained. The images are then encoded via the backbone network $m(.)$ to obtain the representation vectors. The representations are passed through the projection head $n(.)$ to obtain projection vectors. The set of projection vectors is denoted as $\mathcal{Z}_t$. To calculate the model error, we apply the NT-Xent loss (*normalized temperature-scaled cross entropy loss*) introduced in [57]. For a positive pair (z_i, z_j) drawn from the set of projections of augmented images, $\mathcal{Z}_t$ is loss calculated as follows:

$$l(z_i, z_j) = -\log\frac{e^{\text{sim}(z_i,z_j)/\tau}}{\sum_{z_k\in\mathcal{Z}_t-\{z_i\}} e^{\text{sim}(z_i,z_k)/\tau}}, \tag{4}$$

where sim is a similarity function, $\mathcal{Z}_t - \{z_i\}$ are the $2B$ projections of augmented images (with the exception of the projection z_i), and τ is a hyperparameter called temperature.

TABLE 1. Projection head *n*(.).

Layer	Size	Bias
Global Average Pooling Layer	-	-
Dense Layer	2048	True
Batch Normalization Layer & ReLU	-	-
Dense Layer	128	False

After the loss is calculated, we backpropagate the errors to optimize the weights of the backbone neural network $m(.)$ and the projection head $n(.)$. At the end of the training process, we extract the features from the last layer of the backbone neural network $m(.)$ and discard the projection head $n(.)$.

The convergence criterion of the loss function can be approached similarly to supervised methods by looking at the validation loss curve. Most state-of-the-art SSL methods, such as MoCo [56] and SimCLR [43], do not employ an early stopper but set the number of epochs to a fixed number.

To mitigate the error from prolonged training and possible overfitting, a cosine annealing learning rate scheduler and weight decay are utilized.

IV. DATA

In this study, we utilized several CXR datasets. First, a large-scale dataset is required for network pretraining. Therefore, to formulate the pretext task, we utilized the CheXpert dataset [8], which contains 224,316 chest radiographs from 65,420 patients. The samples were labeled by extracting data from radiology reports from October 2002 to July 2017 for 14 commonly observed conditions, such as pneumonia, pneumothorax, and cardiomegaly. It should be noted that even though labels are available, we do not use these during pretraining because the proposed model is unsupervised.

To evaluate the transferability of the model to an external target dataset, we acquired four public datasets. First, the Cell dataset [58] comprises 5,323 X-ray images from children, including 3,883 cases of viral and bacterial pneumonia and 1,349 normal images. The labels were provided by two expert physicians and verified by a third physician. The second dataset is the ChestX-ray14 [9] dataset, which comprises 112,120 X-ray images with eight disease labels from 30,805 patients. We used only a subset of this dataset by selecting only patients with pneumonia and a matched number of healthy controls. The other two datasets were compiled only recently and were intended for COVID-19 detection. The C19-Cohen dataset [59] (accessed 21.10.2020) is a collection of different types of pneumonia (viral, bacterial, and fungal). We selected two classes: 304 patients with COVID-19 and 114 patients with other types of pneumonia. Finally, we also evaluated the proposed model on the COVIDGR dataset [26], which contains 426 CXR images of patients with COVID-19 of four different severity levels and the same number of control subjects. Note that while 76 of these 426 COVID-19 patients were diagnosed as positive by PCR, their CXRs were evaluated as normal, making the classification task more challenging. A brief summary of all the datasets utilized in this study is presented in Tab. 2.

TABLE 2. Datasets used in this study.

Dataset	# samples	# class 0	# class 1	source
pretext				
CheXpert	224,316	na	na	[8]
target				
Cell	5,323	3,883	1,349	[58]
ChestX-ray-14	2,706	1,353	1,353	[9]
C19-Cohen	807	564	243	[59]
COVIDGR	852	426	426	[26]

V. EXPERIMENTS AND RESULTS

In this section, we analyze some aspects of the network pretraining task and report the experimental results.

To demonstrate the generalizability of our approach, we formulated four classification tasks on four publicly available CXR datasets. We avoided combining datasets for a particular classification task and used only one dataset for a specific classification task. In this manner, we tried to avoid the bias and criticism outlined in [26]. The datasets details are presented in Tab. 2. For the Cell dataset and ChestX-ray-14, we classified subjects as having pneumonia or as healthy. Similarly, for COVIDGR, we discriminated between patients with and without COVID-19 disease. Finally, because the C19-Cohen dataset does not include healthy controls, we discriminated between COVID-19 and other types of pneumonia.

A. NETWORK TRAINING AND FEATURE EXTRACTION

We pretrained a ResNet-50 Wide model in the self-supervised task-agnostic way on the large CheXpert dataset of CXR images. The effective batch size was set to 128, and the temperature value was 0.5. We used the Adam optimizer [60] with a learning rate of 0.0005. The model was trained for 100 epochs without any stopping criteria. We used cosine annealing to change the learning rate during the training. The training process convergence is depicted in Fig. 2 (a). The loss on the pretext task plateaus at approximately the 25th epoch and does not decrease further but instead oscillates around some value. This result raises the question of whether it is necessary to train the model beyond the 25th epoch. However, the relationship between the loss on the pretext task and the model's performance on the target task has not yet been established. Here, we analyze the performance in terms of the prediction accuracy on the two datasets (the Cell dataset and the COVIDGR dataset). The prediction accuracy for each epoch is depicted in Fig. 2 (b) and Fig. 2 (c). Although the loss on the pretext task does not improve significantly beyond the 25th–30th epochs, the accuracy achieved for predictions on the Cell dataset increases when we employ models that have been trained with a larger number of epochs. This indicates an interesting phenomenon. During pretraining, even after the loss on the pretext task is no longer improving, the model is still learning. In contrast to the prediction accuracy obtained on the Cell dataset, for the COVIDGR dataset, the accuracy does not gradually improve for models trained beyond the 25th epoch. However, this can

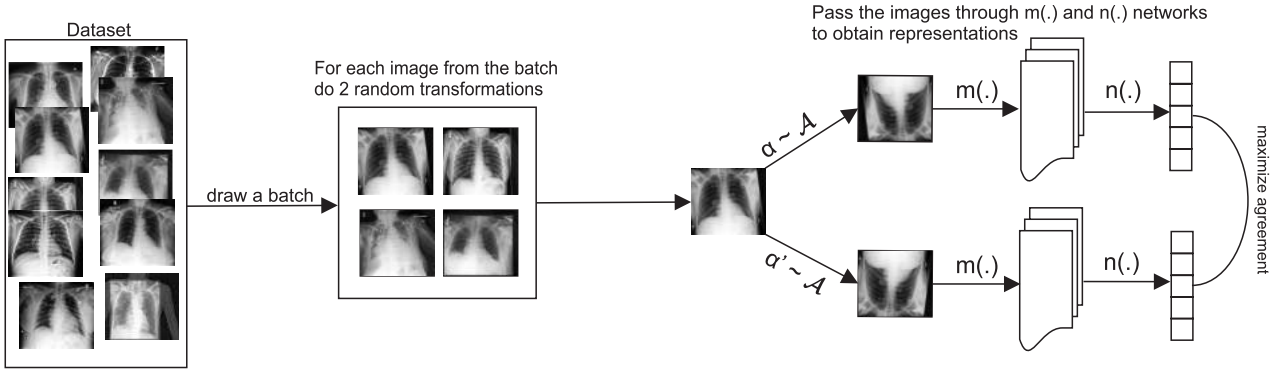

FIGURE 1. Self-supervised architecture.

be explained by the composition of the images in the dataset. As explained before, COVIDGR also contains CXR images from COVID-19 patients in which expert radiologists did not find any pathological changes. The highest reported result so far on this dataset is 76.18% [26]. Our model achieved this accuracy level quite early; thus, we hypothesize that there was no room for further improvement.

To visualize the learned representations, we chose models from four different checkpoints and extracted features. The models were trained for 10, 25, 50 and 100 epochs. Fig. 3 shows t-SNE visualizations of features extracted from the Cell and COVIDGR datasets by these four models. While the Cell dataset exhibits a noticeable but slight improvement in the separability of two classes, the two classes for the COVIDGR dataset seem to be interlaced through all four images.

B. NUMERICAL RESULTS

To examine the predictive performance of the proposed approach, we employ transfer learning and use the pretrained network as a fixed feature extractor. The size of the extracted feature vector is determined by the last dense layer in the projection head, which was 128 in our network. We adopted logistic regression as a classifier and evaluated the model on four different CXR datasets. We also evaluated other linear classifiers, but the results and general trend were very similar, so we omit those results for the sake of readability. To ensure the model's generalizability and avoid overfitting, we used stratified 5-fold cross-validation. The datasets were divided into training, validation, and testing subsets. If the original paper that introduced the datasets also provided the specification of training/validation/test subsets of data, we used that division to achieve fair comparisons of model accuracy with the published results (COVIDGR, ChestX-ray-14 and Cell). Otherwise, we divided the data as follows: 70% as training samples, 10% for validation and 20% as test data (C19-Cohen). Furthermore, we ensured that the CXR images of a particular patient were present only within the same data subset to prevent data leakage that would cause positive bias. All CXR images were resized to 224 × 224 pixels.

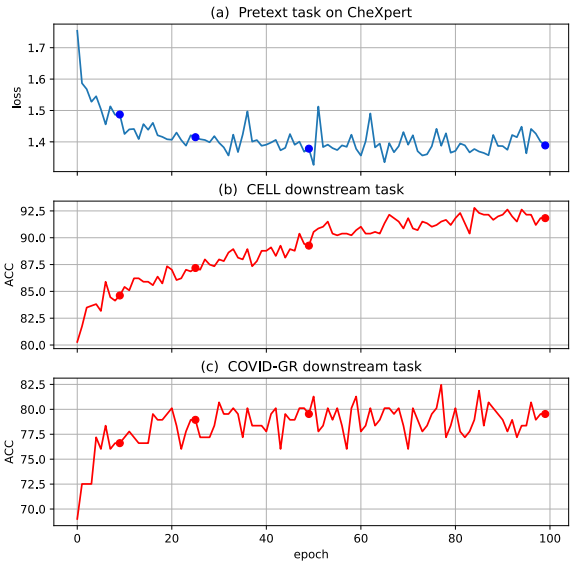

FIGURE 2. Loss and accuracy in each epoch. (a) Loss on the pretext task and accuracy on the (b) Cell dataset and (c) COVIDGR dataset.

To determine the optimal logistic regression parameters, we searched through the parameter space C = {0.01, 0.05, 0.1, 0.2, 0.5, 1}. The logistic regression weights were automatically adjusted to be inversely proportional to the class frequencies in the input data. The other parameters were set to their default values. The best model was adopted after a grid search based on the area under the receiver operating characteristic curve (AUC) metric.

We also evaluated the amount of data required for the pretext task to correctly identify relevant features that are beneficial for the downstream task by testing with three different data fractions: 1%, 10% and 100%.

The results of the trained models are depicted in Table 3. To provide a better overview of model performance, we calculated several metrics: accuracy (ACC), AUC, sensitivity (SEN) and specificity (SPE).

Our first observation is that the prediction accuracy differs between the datasets. The model prediction accuracy varies more in the pneumonia classification task (Cell, ChestX-ray-14) than in the COVID-19 classification task

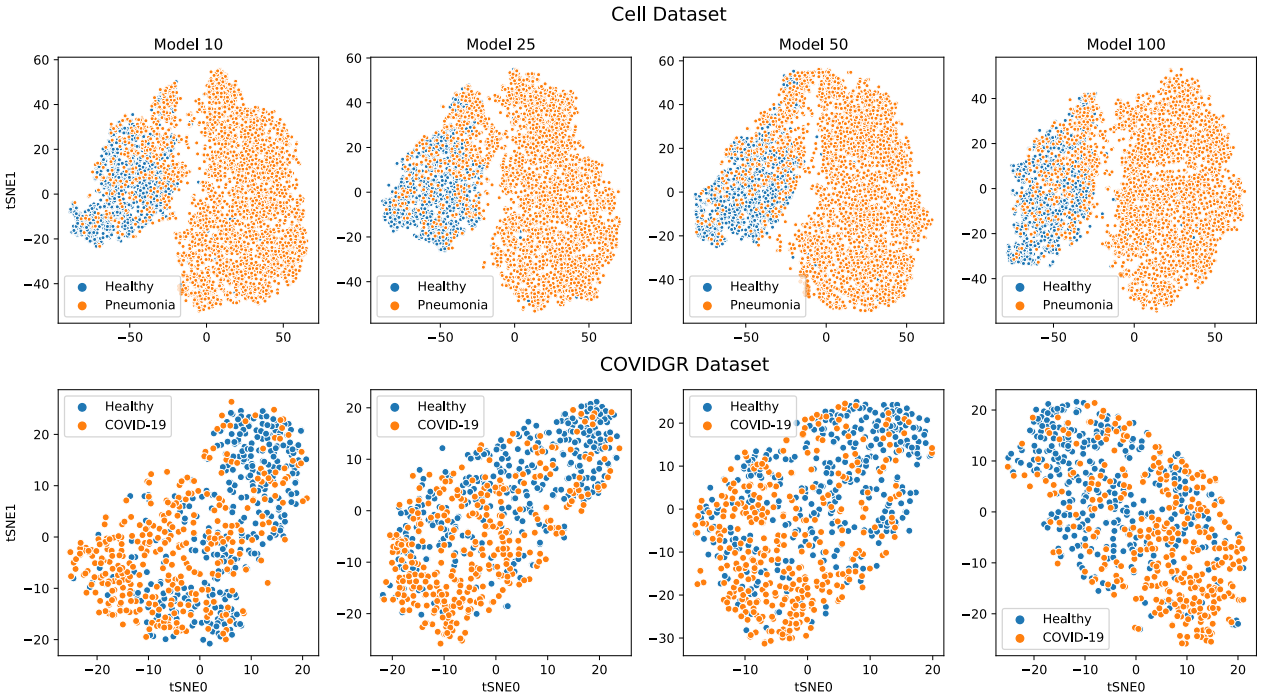

FIGURE 3. t-SNE visualization of the features extracted by models selected in different stages of the pretraining process (10th, 25th, 50th and 100th epochs). The features shown are from the Cell and COVIDGR datasets.

(C19-Cohen, COVIDGR). This variation is caused by the different dataset compositions. The datasets were compiled from different sources and acquired by different devices, which influences the image characteristics. However, more importantly, the Cell dataset is composed of CXR children aged four to six years, making classification a specific type of task.

On the Cell dataset, the AUC and ACC tend to increase as the dataset fraction size increases for the pretext task. The highest AUC 97.7% of our model is higher than the 96.6% reported in [58], which used transfer learning based on ImageNet. This result shows that representations learned in a self-supervised fashion on smaller datasets with semantically closer domains are more beneficial than supervised pretraining on large but semantically very different datasets such as ImageNet.

The model evaluated on the ChestX-ray-14 dataset yielded significantly lower results than the model evaluated on the Cell dataset. In this case, the results are independent of the fraction of the dataset used for training the pretext task. Tab. 3 shows that the proposed model achieves a higher score than the results published in [9]. However, the comparison is not entirely fair because the authors of [9] were solving a multiclass problem, which could have had a negative impact on the model accuracy compared to binary classification.

Models trained on the COVIDGR and Cohen-19 datasets achieved comparable results. One model trained on C19-COHEN achieved AUCs up to 91.5% when using a 10% fraction of the dataset in the pretext task. Surprisingly, it outperformed another model trained on the pretext task with the entire dataset. We hypothesize that this may have been caused by the better performance of the logistic regression model due to the hyperparameters found in the grid search. Some of the previously published papers combined the C19-Cohen dataset with CXRs of healthy controls obtained from other datasets. We intentionally avoided such combinations, and to the best of our knowledge, no published paper has conducted classification only on the C19-Cohen dataset; thus, we cannot directly compare the performance of our model with those of others on the C19-Cohen dataset.

Encouraging results were achieved on the COVIDGR dataset in differentiating between healthy and COVID-19 CXRs. Our CNN pretrained in a self-supervised fashion was able to outperform the supervised COVID-SDNet [26] model by a few percent. Although this difference is not large, it should be noted that SSL does not require a large, labeled training dataset, which can save substantial human resources.

C. EXPLAINING CNN DECISIONS

To shed some light on the CNN decisions, we employ gradient-weighted class activation mapping (Grad-CAM) to highlight the important regions of the CXR image corresponding to a decision. Grad-CAM is a class-discriminative localization technique that generates visual explanations for CNN decisions [61]. Fig. 4 shows CXRs of six different patients correctly classified as pneumonia cases. Here, images of both ground-glass opacities and consolidations are present together with air bronchograms. An air bronchogram is a dark radiographic appearance of an air-filled bronchus (dark thread-like line) made visible by the opacification of

the surrounding alveoli (''white lung field''). An air bronchogram is another pathological sign frequently associated with pneumonia. The more edematous the lung tissue is, the easier it is to spot; however, this applies only to those bronchi that have at least some residual air inside them. In this instance, the areas in the lungs highlighted by the attention map cover the regions with visible pulmonary opacification and air bronchograms, which provides the grounds for a correct diagnosis. In clinical practice, a radiologist looks for the same radiological signs as the neural network used here for decision making. Other areas highlighted by the attention map are outside the lung region. These areas cannot be linked to any pathology caused by pneumonia and probably reflect zones incorrectly evaluated by the visualization algorithm or by the CNN itself.

VI. DISCUSSION

We proposed an approach based on a self-supervised convolutional network and evaluated it on four datasets. To avoid the critiques presented by [26] and others [62], we did not combine existing datasets, which clearly limits the available training and testing data size. On the other hand, it helps ensure that the model learns only parameters related to specific aspects of the disease pathology and not differences arising from different devices or acquisition procedures. Here, it should be noted that we avoided combining datasets from different sources in that the disease class samples were provided by one source and the control subject samples were extracted from different sources. To increase the confidence of our approach, we evaluated the trained model on four different datasets focused on two different diseases. We differentiated between CXR with no findings and CXR-containing pathologies and between CXRs of patients with different pneumonia types and those with COVID-19-induced pneumonia.

One issue we touched on briefly in Section V-A is the selection of the optimal model for transfer learning. Fig. 2 clearly shows that the relationship between model performance on the pretext task and the subsequent downstream task is not straightforward. The question that arises is how to select the optimal checkpoint for the pretrained model. According to the traditional training view, the user may be tempted to stop the training at approximately the 40th epoch based on the loss curve in Fig. 2(a) because the loss is no longer improving. However, at this point, the model is far from optimal in the sense of the prediction accuracy on the downstream task (at least for the Cell dataset 2(b)). Further research is needed to establish the relationship between model performance on pretext tasks and downstream tasks.

A. LIMITATIONS OF THE STUDY

We provide decision support for the classification of CXR images; however, the output should be taken with caution. The final diagnosis should always be based on the combination of clinical symptoms (cough, fever, pathological respiratory sounds) and laboratory results. However, in critical pandemic situations such as the one we are experiencing currently, medical staff are extremely busy, and a solution that provides rapid automated information could help to reduce the burden on medical personnel.

Some datasets, such as C19-Cohen, are compiled from X-ray images from different institutions, so this could result in some bias. However, the images from different resources are distributed in both classes, so potential bias should be quite limited. This is a dataset frequently utilized in similar studies, so we included it for comparison even though the results should be interpreted carefully.

Clearly, there is a strong need for further validation and detailed assessment led by transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) [63] before an automated approach can be used in clinical practice. Additional datasets from different types of devices need to be included in testing and evaluation. However, this study has proven that it is possible to train the model in a self-supervised fashion and apply it successfully to medical imaging tasks without the need for large amounts of labeled data. This may open new research horizons because similar approaches can be examined for other types of medical imaging, such as computer tomography and retinal imaging.

The proposed approach for CXR classification is based on a deep CNN, and a fundamental principle of these models is that they work in a black-box manner. The ability to explain the decisions of deep learning models is still in its early stages and is a hot research topic. In this paper, we utilize grad-CAM to provide not only prediction but also some reasoning for the network decision. Grad-CAM shows the region that is most relevant for the prediction. The explanations are based on the learned attention regions in Fig. 4 and are the same as areas that doctors review. However, these explanations do not cover all aspects and peculiarities included in the final decision. Because explainability is of crucial importance for medical applications, this and other limitations will be addressed in future work.

B. FAULT ANALYSIS

We also investigated some misclassifications to obtain a better understanding of CNN decisions. Fig. 5 shows four CXRs that were incorrectly classified. Cases (a) and (b) were both misclassified as pneumonia. The patient in Fig. 5(a) has a small consolidation-like area (''white lung field'') in the (right) middle lobe, and Case (b) shows a distinct diffuse reticular interstitial pattern. Both patterns may resemble pneumonia findings, which may have led the model to incorrect classification. It is likely that a radiology specialist would make the same mistake if the only available information was the radiograph. The true cause of these patterns would have to be determined by the patient's clinical history and additional radiological examinations (such as computed tomography scan). On the other hand, the CXR images in 5(c) and 5(d) were also misclassified as pneumonia, but these do not present any structural changes that could be associated

TABLE 3. Prediction performance of the proposed approach on four CXR datasets. We provide previous published results for comparison purposes. * There is no available result for comparison because previously published results combine the C19-Cohen dataset with some other CXR dataset. Moreover, data are continuously added to the 19-Cohen dataset. ** Result of pneumonia prediction as part of a multiclass classification.

Dataset	Fraction of dataset												ImageNet pretraining				Published result
	1%				10%				100%								ACC/AUC
	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE	
Cell	85.6	96.6	99.5	62.4	86.9	96.9	99.2	66.2	91.5	97.7	98.7	79.5	83.2	94.9	98.2	58.1	92.8/96.8 [58]
C19-Cohen	84.9	89.2	90.6	73.6	81.8	91.5	85.8	73.6	81.1	88.2	83	77.4	69.8	77.0	80.2	49.1	na*
COVIDGR	79.5	86.6	83.5	75.6	77.8	86.0	80	75.6	78.4	87.1	83.5	73.3	71.3	77.7	69.4	73.2	76.16/na [26]
ChestX-ray-14	71.5	79.1	72.8	71.4	71.2	78.1	72.0	71.2	71.4	78.4	71.3	71.5	69.6	75.1	67.0	69.6	na/65.8** [9]

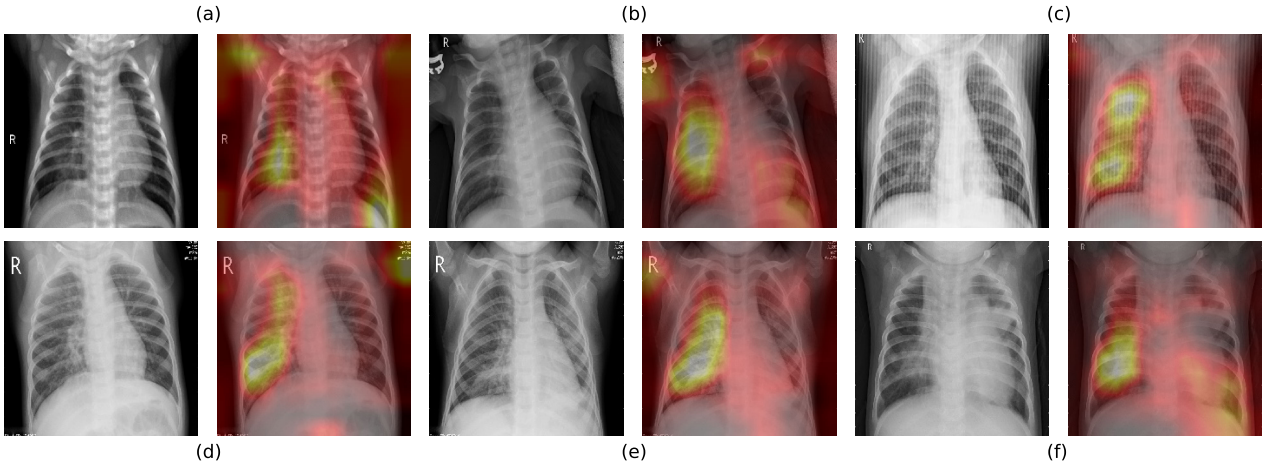

FIGURE 4. Visualizations of 6 CXR images from the Cell dataset and their Grad-CAM attention maps. These images were taken of patients with pneumonia and were diagnosed with pneumonia by our model.

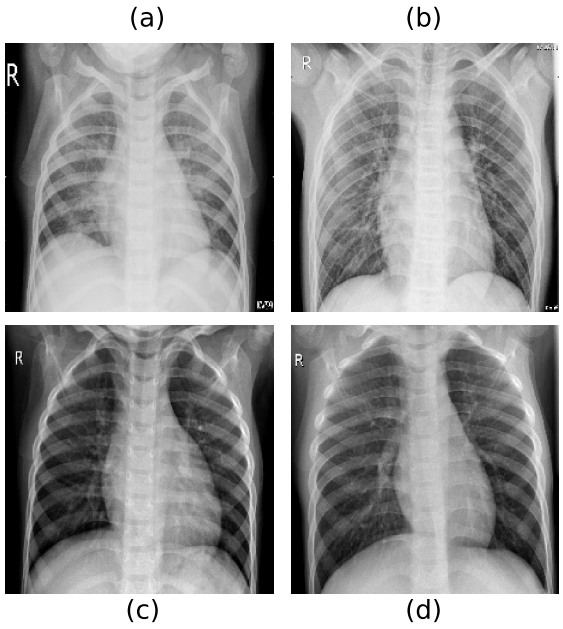

FIGURE 5. CXR images of four healthy subjects classified as pneumonia.

with the incorrect classification and should have been classified as healthy patients. It is difficult to determine the cause of these particular misinterpretations. This demonstrates the disadvantage of a CNN behaving as a black box.

VII. CONCLUSION

The current pandemic further highlights the need to include diagnostic decision support systems in clinical decision making. The successful incorporation of these systems into contemporary medical devices could automate certain tasks and reduce medical personnel workloads. Automated solutions also contribute strongly during noncritical times by giving medical specialists more time for tasks and duties that require a more careful or specific approach.

As a contribution to medical expert systems, we introduced a solution that classifies CXR images. The proposed approach utilizes a CNN pretrained on an unlabeled dataset of CXR images. By avoiding the need for labeled data, which are both scarce and expensive in the medical domain, our approach offers new possibilities for CNN utilization by demonstrating that CNN networks do not need to be trained on only natural images (such as the ImageNet dataset), as in the majority of approaches today; they can instead be trained on images that are semantically closer to the target task. In our case, a network pretrained on the ChestXpert dataset was able to learn meaningful representations and extract relevant features for pneumonia and COVID-19 detection. The obtained results of our unsupervised model are competitive with their supervised counterparts. Considering that self-supervised contrastive learning for visual representations is a very new topic, this approach has huge potential. Later methodological improvements may further boost the performance.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Buša and D. Hubáček, M.D., for their valuable comments.

REFERENCES

[1] R. Smith-Bindman, D. L. Miglioretti, E. Johnson, C. Lee, H. S. Feigelson, M. Flynn, R. T. Greenlee, R. L. Kruger, M. C. Hornbrook, D. Roblin, L. I. Solberg, N. Vanneman, S. Weinmann, and A. E. Williams, "Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996–2010," *J. Amer. Med. Assoc.*, vol. 307, no. 22, pp. 2400–2409, Jun. 2012, doi: 10.1001/jama.2012.5960.

[2] A. S. Hong, D. Levin, L. Parker, V. M. Rao, D. Ross-Degnan, and J. F. Wharam, "Trends in diagnostic imaging utilization among medicare and commercially insured adults from 2003 through 2016," *Radiology*, vol. 294, no. 2, pp. 342–350, Feb. 2020, doi: 10.1148/radiol.2019191116.

[3] O. Ruuskanen, E. Lahti, L. C. Jennings, and D. R. Murdoch, "Viral pneumonia," *Lancet*, vol. 377, no. 9773, pp. 1264–1275, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673610614596

[4] J. S. Brown, "Community-acquired pneumonia," *Clin. Med.*, vol. 12, no. 6, pp. 538–543, 2012. [Online]. Available: https://www.rcpjournals.org/content/12/6/538

[5] H. Y. F. Wong, H. Y. S. Lam, A. H.-T. Fong, S. T. Leung, T. W.-Y. Chin, C. S. Y. Lo, M. M.-S. Lui, J. C. Y. Lee, K. W.-H. Chiu, T. W.-H. Chung, E. Y. P. Lee, E. Y. F. Wan, I. F. N. Hung, T. P. W. Lam, M. D. Kuo, and M.-Y. Ng, "Frequency and distribution of chest radiographic findings in patients positive for COVID-19," *Radiology*, vol. 296, no. 2, pp. E72–E78, Aug. 2020.

[6] J. Russell, A. Echenique, S. R. Daugherty, and M. Weinstock, "Chest X-ray findings among urgent care patients with COVID-19 are not affected by patient age or gender: A retrospective cohort study of 636 ambulatory patients," *J. Urgent Care Med.*, vol. 14, no. 10, pp. 13–18, 2020.

[7] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[8] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, J. Seekins, D. A. Mong, S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungren, and A. Y. Ng, "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 590–597.

[9] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.

[10] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101794. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841520301584

[11] I. D. Apostolopoulos and T. Bessiana, "COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, 2020, doi: 10.1007/s13246-020-00865-4.

[12] V. Guarrasi, N. C. D'Amico, R. Sicilia, E. Cordelli, and P. Soda, "Pareto optimization of deep networks for COVID-19 diagnosis from chest X-rays," *Pattern Recognit.*, vol. 121, Jan. 2022, Art. no. 108242. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320321004234

[13] S. Rajaraman, J. Siegelman, P. O. Alderson, L. S. Folio, L. R. Folio, and S. K. Antani, "Iteratively pruned deep learning ensembles for COVID-19 detection in chest X-rays," *IEEE Access*, vol. 8, pp. 115041–115050, 2020.

[14] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105581.

[15] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103792. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482520301621

[16] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[17] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," *Pattern Recognit. Lett.*, vol. 138, pp. 638–643, Oct. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865520303512

[18] L. Wang, Z. Q. Lin, and A. Wong, *Sci. Rep.*, no. 1, Art. no. 19549.

[19] A. Kumar, A. R. Tripathi, S. C. Satapathy, and Y.-D. Zhang, "SARS-Net: COVID-19 detection from chest X-rays by combining graph convolutional network and convolutional neural network," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108255. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320321004350

[20] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *Int. J. Speech Technol.*, vol. 51, no. 2, pp. 854–864, Feb. 2021, doi: 10.1007/s10489-020-01829-7.

[21] P. Zhang, Y. Zhong, Y. Deng, X. Tang, and X. Li, "DRR4COVID: Learning automated COVID-19 infection segmentation from digitally reconstructed radiographs," *IEEE Access*, vol. 8, pp. 207736–207757, 2020.

[22] S. Varela-Santos and P. Melin, "A new approach for classifying coronavirus COVID-19 based on its manifestation on chest X-rays using texture features and neural networks," *Inf. Sci.*, vol. 545, pp. 403–414, Feb. 2021. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025520309531

[23] P. Soda *et al.*, "AIforCOVID: Predicting the clinical outcomes in patients with COVID-19 applying AI to chest-X-rays. An Italian multicentre study," *Med. Image Anal.*, vol. 74, Dec. 2021, Art. no. 102216. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841521002619

[24] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, "Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms," *Int. J. Med. Informat.*, vol. 144, Dec. 2020, Art. no. 104284. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S138650562030959X

[25] D. I. Morís, J. J. de Moura Ramos, J. N. Buján, and M. O. Hortas, "Data augmentation approaches using cycle-consistent adversarial networks for improving COVID-19 screening in portable chest X-ray images," *Expert Syst. Appl.*, vol. 185, Dec. 2021, Art. no. 115681. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421010666

[26] S. Tabik, A. Gómez-Ríos, J. L. Martín-Rodríguez, I. Sevillano-García, M. Rey-Area, D. Charte, E. Guirado, J. L. Suárez, J. Luengo, M. A. Valero-González, P. Garcia-Villanova, E. Olmedo-Sánchez, and F. Herrera, "COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 12, pp. 3595–3605, Dec. 2020.

[27] O. D. T. Catala, I. S. Igual, F. J. Perez-Benito, D. M. Escriva, V. O. Castello, R. Llobet, and J.-C. Perez-Cortes, "Bias analysis on public X-ray image datasets of pneumonia and COVID-19 patients," *IEEE Access*, vol. 9, pp. 42370–42383, 2021.

[28] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020.

[29] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen, and Y. Xia, "Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 879–890, Mar. 2021.

[30] L. Luo, L. Yu, H. Chen, Q. Liu, X. Wang, J. Xu, and P.-A. Heng, "Deep mining external imperfect data for chest X-ray disease screening," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3583–3594, Nov. 2020.

[31] H. Sowrirajan, J. Yang, A. Y. Ng, and P. Rajpurkar, "MoCo pretraining improves representation and transferability of chest X-ray models," in *Proc. 4th Conf. Med. Imag. With Deep Learn.*, 2020, pp. 728–744.

[32] Z. Wang, Y. Xiao, Y. Li, J. Zhang, F. Lu, M. Hou, and X. Liu, "Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays," *Pattern Recognit.*, vol. 110, Feb. 2021, Art. no. 107613.

[33] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "4S-DT: Self-supervised super sample decomposition for transfer learning with application to COVID-19 detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 2798–2808, Jul. 2021.

[34] A. I. Aviles-Rivero, P. Sellars, C.-B. Schönlieb, and N. Papadakis, "GraphXCOVID: Explainable deep graph diffusion pseudo-labelling for identifying COVID-19 on chest X-rays," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108274. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320321004544

[35] S. Nabavi, A. Ejmalian, M. E. Moghaddam, A. A. Abin, A. F. Frangi, M. Mohammadi, and H. S. Rad, "Medical imaging and computational image analysis in COVID-19 diagnosis: A review," *Comput. Biol. Med.*, vol. 135, Aug. 2021, Art. no. 104605. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521003991

[36] A. Ulhaq, J. Born, A. Khan, D. P. S. Gomes, S. Chakraborty, and M. Paul, "COVID-19 control by computer vision approaches: A survey," *IEEE Access*, vol. 8, pp. 179437–179456, 2020.
[37] Y. Boucharеb, P. M. Khaniabadi, F. Al Kindi, H. Al Dhuhli, I. Shiri, H. Zaidi, and A. Rahmim, "Artificial intelligence-driven assessment of radiological images for COVID-19," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104665. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521004595
[38] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 4–15, 2021.
[39] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, and T. M. Islam, "Can AI help in screening viral and COVID-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132665–132676, 2020.
[40] D. Dong, Z. Tang, S. Wang, H. Hui, L. Gong, Y. Lu, Z. Xue, H. Liao, F. Chen, F. Yang, R. Jin, K. Wang, Z. Liu, J. Wei, W. Mu, H. Zhang, J. Jiang, J. Tian, and H. Li, "The role of imaging in the detection and management of COVID-19: A review," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 16–29, 2021.
[41] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.
[42] O. Henaff, "Data-efficient image recognition with contrastive predictive coding," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4182–4192.
[43] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020, *arXiv:2002.05709*.
[44] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9729–9738.
[45] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. H. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. D. Guo, M. G. Azar, B. Piot, K. Kavukcuoglu, R. Munos, and M. Valko, "Bootstrap your own latent: A new approach to self-supervised learning," 2020, *arXiv:2006.07733*.
[46] I. Misra and L. van der Maaten, "Self-supervised learning of pretext-invariant representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6707–6717.
[47] A. Newell and J. Deng, "How useful is self-supervised pretraining for visual tasks?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7345–7354.
[48] X. Liu, F. Zhang, Z. Hou, Z. Wang, L. Mian, J. Zhang, and J. Tang, "Self-supervised learning: Generative or contrastive," 2020, vol. 1, no. 2, *arXiv:2006.08218*.
[49] M. Noroozi and P. Favaro, "Unsupervised learning of visual representations by solving jigsaw puzzles," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 69–84.
[50] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," 2018, *arXiv:1803.07728*.
[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[52] A. Krizhevsky, V. Nair, and G. Hinton. *CIFAR-10 (Canadian Institute for Advanced Research)*. [Online]. Available: http://www.cs.toronto.edu/~kriz/cifar.html
[53] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[54] A. van Opbroek, M. A. Ikram, M. W. Vernooij, and M. de Bruijne, "Transfer learning improves supervised image segmentation across imaging protocols," *IEEE Trans. Med. Imag.*, vol. 34, no. 5, pp. 1018–1030, May 2015.
[55] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
[56] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3733–3742.
[57] K. Sohn, "Improved deep metric learning with multi-class n-pair loss objective," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 1857–1865.
[58] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
[59] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q Duong, and M. Ghassemi, "COVID-19 image data collection: Prospective predictions are the future," 2020, *arXiv:2006.11988*. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset
[60] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[61] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
[62] G. Maguolo and L. Nanni, "A critic evaluation of methods for COVID-19 automatic detection from X-ray images," *Inf. Fusion*, vol. 76, pp. 1–7, Dec. 2021.
[63] G. S. Collins and K. G. M. Moons, "Reporting of artificial intelligence prediction models," *Lancet*, vol. 393, no. 10181, pp. 1577–1579, Apr. 2019.

MATEJ GAZDA received the M.Sc. degree in applied computer science from the Faculty of Electrical Engineering and Informatics, Technical University of Košice, Košice, Slovakia, in 2019. His research interests include biomedical image analysis, feature selection, and biomedical decision support systems.

JÁN PLAVKA graduated in discrete mathematics from Pavol Jozef Šafarik University, Košice. He received the Ph.D. degree in the field of steady-state discrete event dynamic systems, in 1991. He is currently a Professor with the Department of Mathematics and Theoretical Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice. His scientific research focuses on computer science, the complexity of algorithms, and discrete event dynamic systems. In addition, he investigates questions related to neuron networks.

JAKUB GAZDA received the M.D. degree from the Medical Faculty of Pavol Jozef Šafarik University, Košice, Slovakia, in 2018. He currently works at the Department of Internal Medicine. His research interests include autoimmune liver diseases and medical imaging.

PETER DROTÁR (Member, IEEE) received the M.Sc. and Ph.D. degrees in electronics from the Faculty of Electrical Engineering and Informatics, Technical University of Košice, Košice, Slovakia, in 2007 and 2010, respectively. From 2010 to 2012, he was with Honeywell International, Advanced Technology Europe, as a Scientist for communication and surveillance systems. From 2012 to 2015, he was with the SIX Research Centre, Brno University of Technology, Brno, Czech Republic, as a Postdoctoral Research Assistant. He is currently an Associate Professor with the Department of Computers and Informatics, Technical University of Košice. He leads research and development projects concerning biomedical decision support systems. His research interests include biomedical signal and image processing, feature selection, and pattern recognition.