Automatic Quantification of COVID-19 Pulmonary Edema by Self-supervised Contrastive Learning

Zhaohui Liang[1[0000-0002-9361-5535]], Zhiyun Xue[1[0000-0003-0644-385X]], Sivaramakrishnan Rajaraman[1[0000-0003-0871-8634]], Yang Feng[1[0000-0002-8334-7450]], and Sameer Antani[1[0000-0002-0040-1387]]

[1] Computational Health Research Branch, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA
sameer.antani@nih.gov

Abstract. We proposed a self-supervised machine learning method to automatically rate the severity of pulmonary edema of the frontal chest X-ray radiographs (CXR) of COVID-19 viral pneumonia with the modified radiographic assessment of lung edema (mRALE) scoring system. The new model was first optimized with the simple Siamese network (SimSiam) architecture where a ResNet-50 pretrained by ImageNet database was used as the backbone. The encoder projected a 2048-dimension embedding as representation features to a downstream fully connected deep neural network for mRALE score prediction. A 5-fold cross-validation with 2,599 frontal CXRs was used to examine the new model's performance with comparison to a non-pretrained SimSiam encoder and a ResNet-50 trained from scratch. The mean absolute error (MAE) of the new model is 5.05 (95%CI 5.03-5.08), the mean squared error (MSE) is 66.67 (95%CI 66.29-67.06), and the Spearman's correlation coefficient (Spearman ρ) to the expert-annotated scores is 0.77 (95%CI 0.75-0.79). All the performance metrics of the new model are superior to the two comparators (P<0.01), and the scores of MSE and Spearman ρ of the two comparators have no statistical difference (P>0.05). We conclude that the self-supervised contrastive learning method is an effective strategy for mRALE automated scoring. It provides a new approach to improve machine learning performance and minimize the expert knowledge involvement in quantitative medical image pattern learning.

Keywords: Self-supervised Learning, Contrastive Learning, COVID-19, Pulmonary Edema, Deep Learning.

1 Introduction

1.1 Lung Edema and COVID-19 Prognosis

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the main cause of death of coronavirus disease (COVID-19) that imposed a massive global pandemic since 2020. According to World health Organization, the confirmed cases of COVID-19 have reached 767,518,723 and the death toll has accumulated to 6,947,192[1]. As we are approaching the end of the pandemic, new studies not only concentrate on the

detection and prevention of COVID-19, but also on the evaluation and prediction of the prognosis of SARS-CoV-2. For example, a cohort study in Japan found that the chance of the SARS-CoV-2 patients for prolonged mechanical ventilation (PMV) was quantitatively relevant to their lung ultrasound score (LUS) and radiographic assessment of the lung edema (RALE) score [2]. An international multicenter study found that the progressive increase of RALE is associated to the mortality of acute respiratory distress syndrome (ARDS) cases by COVID-19 [3]. RALE introduced by Warren MA et. al. is a quantitative tool to assess the extent and density of alveolar opacities on chest radiographs that reflect the degree of pulmonary edema [4]. As a key indicator of acute respiratory distress syndrome (ARDS), pulmonary edema is an indicator of SARS-CoV-2 severity and a predictor of prognosis [5]. The modified RALE scoring system (mRALE) is a non-invasive measure to evaluate the severity of pulmonary edema on chest X-ray radiographs (CXR). It renders a total score ranging from 0 the 24 to respectively assess the extent and density of alveolar opacities of different regions of lung assessed on frontal CXRs. A retrospective cohort study on refractory cardiogenic shock and cardiac arrest patients receiving veno-arterial extracorporeal membrane oxygenation (VA-ECMO) support concluded that RALE is a discriminator for mortality and the progressive monitoring of RALE scores is a good predictor for therapeutic effects [6]. RALE and mRALE are also applicable to SARS-CoV-2. An international multicenter study in Europe observed 350 CXRs from 139 COVID-19 ARDS patients in intensive care units (ICU) revealed that the progressive increase of RALE score in CXRs was associated to higher mortality and longer use time of ventilator [7].

1.2 Deep Learning for COVID-19 Images

Since the continuous monitoring of pulmonary edema through CXRs is an effective measure for ARDS severity and a predictor for prognosis, it can serve as an important parameter for clinical decision making, which provides a handy method for COVID-19 patient administration, particularly when CXRs are one of the most common and low-cost examinations for COVID-19 severity assessment at the peak of pandemic. Meanwhile, machine learning (ML), especially deep learning (DL), has been widely studied at the beginning of the pandemic to enhance disease detection and management with fruitful outcomes, but it also faces multiple challenges in these critical use cases. A review on the application of DL for COVID-19 image processing found that convolutional neural networks (CNNs) have gained the most popularity for image classifications of COVID-19. Meanwhile, but the authors pointed out the current challenges to DL for COVID-19 image analysis applications include the lack of image quality assurance, unbalance and diversity of the image datasets, and the generalization and reproducibility of the available models and algorithms, etc. [8]. A survey based on over 100 published papers on DL for COVID-19 explored the three common imaging modalities of COVID-19: CXR, CT and ultrasound images. It concluded that the most effective and time-efficient method to detect COVID-19 is CXR imaging for its ease access, low cost and low exposure to radiation compared to CT scans, and higher sensitive com-

pared to ultrasound images. It also pointed out future research should enhance the reliability and robustness of DL applications such as using explainable AI to reduce the uncertainty and improving the accessibility of large, well-annotated image datasets [9].

One latest finding by Xie et al. is the introduction of the dense regression activation maps (dRAMs) to segment the lesion lung region from COVID-19 CT scans with the Dice coefficient of 70.2% [10]. Another is the bilateral adaptive graph-based (BA-GCN) model by Meng et al. with 2-D segmentation function for the 3-D CT scans to detect early infection of COVID-19 with significant improvement on both learning ability and generalization ability [11]. Signoroni et al. presented the BS-Net for weakly supervised learning to perform both classifications, scoring and segmentation simultaneously [12]. However, all the high performance of these new methods is based on the availability of large well-annotated datasets. If the models are applied to a new use case, they must be retrained with new data, or the performance will inevitably decrease. The research on adversarial network attacks supports this idea because the well-trained DL models for COVID-19 detection will become vulnerable from simple network attacks [13-14]. This phenomenon can be explained by the randomness of pattern capturing of DL. It means that the seemingly high performance of DL in fact relies on the random combination of insignificant patterns which cannot be mapped to human expertise. This assumption is supported by quantitative measures such as RALE or mRALE which are confirmed as important predictors for the severity of ARDS and SARS-CoV-2. Thus, The automated quantification for COVID-19 is a solution to these challenges.

1.3 Deep Learning for COVID-19 Severity Quantification

The mRALE score prediction model proposed by Li et al. in 2020 is considered as a success for COVID-19 automated quantification, where a Siamese Network model was first pretrained by over 160,000 reference CXR images, then it was fine-tuned by 314 frontal CXR images to predict the mRALE score for pulmonary edema severity. The model achieved high consistency with the medical expert annotation with the Spearman's correlation coefficient of 0.86 and the area under the receiver operating characteristic curve (ROC AUG) of 0.80 [15]. Another similar research by Horng et al. used semi-supervised learning to train a DenseNet architecture with a CXR dataset with 369,071 from 64,581 patients to learn the cardiopulmonary context, then the model was fine-tuned to learn the severity of chronic heart failure (CHF) in four ordinal levels based on alveolar edema [16].

1.4 Study Motivation

The current success in DL mainly relies on large CXR datasets as context to guide the DL models to acquire necessary prior knowledge such as the modality significant visual patterns for the downstream tasks. Our study aims to solve the difficulty of data accessibility of DL for COVID-19 severity quantification. A new ML technology called contrastive learning is applied to learn the cardiopulmonary context patterns by comparing the similarity of identical image pairs with random augmentation. An encoder for extracting meaningful features from CXRs can be trained by this strategy for mRALE score prediction. We believe this method can reduce the data-greedy limitation of DL.

2 Material and Methods

2.1 Contrastive Learning

Contrastive learning is a machine learning approach to initialize the model training in the unsupervised learning manner to acquire meaningful representations of the training data patterns. Then the training switches to the downstream supervised learning for explicit tasks. It is considered as a transfer learning strategy to improve ML performance. The classic contrastive learning uses both positive and negative data pairs as inputs, where the positive pairs are pulled closer in the latent space while the negative pairs are pushed apart during ML optimization. For the single input image example i, its contrastive loss is defined as:

$$L_i = -\frac{e^{s_{i,i'}}}{\sum_j e^{s_{i,j}}} = -s_{i,i'} + log\left(\sum_j e^{s_{i,j}}\right) \tag{1}$$

where $-s_{i,i'}$ is the loss term for the positive pair and $log\left(\sum_j e^{s_{i,j}}\right)$ is the loss term for the negative pair. In Eq. (1), the computation cost for the positive pair loss is obviously lower than the negative pair because the former is simply the negative cosine similarity while the latter one is the summation of all the difference of negative pairs. Note that the negative pairs loss is used to maintain the training mode, if a new method can be found to prevent mode collapse, we can remove the negative pairs loss to simplify the computation as shown by the bootstrap your own latent (BYOL) method by Tian et al., where two identical networks were paralleled to learn the image similarity in the teacher-and student mode [17]. The loss objective of BYOL is revised as:

$$L_i = -\frac{e^{s_{i,i'}}}{\sum_j e^{s_{i,j}}} = -s_{i,i'} + \beta \cdot log\left(\sum_j e^{s_{i,j}}\right) \tag{2}$$

where β is a tunable hyperparameter ranging from 0 to 1. When β=0, the loss is only determined by the positive pair part. The BYOL method also uses the stop-gradient (stop-grad) method to prevent the teacher network from being updated by backpropagation, and a predictor is added to the student network as the learning reference for the teacher network's update using exponential moving average (EMA) weighted by β. This asymmetric architecture is effective to prevent mode collapse. The teacher network finally learns the feature similarity representations and it can act as the target encoder for any downstream tasks such as classification and regression.

The BYOL is further simplified by removing the identical paralleled architecture and replacing with a shared network for both encoder arms as the implementation of the Simple Siamese network (SimSiam) [18]. The SimSiam first used a single Siamese architecture as the backbone shared by the two sets of identical input images respectively with random augmentation. Taking the advantage that the Siamese network naturally produces "inductive biases for modeling invariance", the contrastive learning mode can be kept and progress stably. The SimSiam loss function is defined as:

$$L = \frac{1}{2}\mathcal{D}(p_1, z_2) + \frac{1}{2}\mathcal{D}(p_2, z_1) = \frac{1}{2}\mathcal{D}(p_1, stopgrad(z_2)) + \frac{1}{2}\mathcal{D}(p_2, stopgrad(z_1)) \tag{3}$$

where z_1 and z_2 are the outputs of the projector of the p_1 and p_2 are the outputs of the predictor. Note that the stop-gradient mechanism plays a crucial role to prevent mode collapse in SimSiam optimization as mentioned in the original paper [18].

In our study, we use the pretrained ResNet-50 by ImageNet as the backbone of the SimSiam encoder and the projector output is a 2,048-dimensional embedding, which serves as the feature map for the downstream regression tasks. The whole architecture of the SimSiam based self-supervised model is illustrated in Fig. 1 in the next section.

2.2 Regression based on Self-supervised Learning Features

Self-supervised learning (SSL) formulates vision patterns through the contrastive learning process. It can be used as the pattern extractor for a regressor model composed of two parts: the SimSiam network for SSL contrastive learning, and the regressor for mRALE prediction by supervised learning.

As discussed in the previous section, the SimSiam can be further divided into the encoder and the predictor, forming a structure like an autoencoder (AE). In our model, the encoder is a ResNet50 (backbone) pretrained by the ImageNet dataset. The pretrained model is easier to train to guide the filters focus on the meaningful region such as the alveoli of the lungs. It can be verified by the gradient-weighted class activation mapping (Grad-CAM). Our experiments showed that the combined strategy of transfer learning and contrastive learning can effectively keep a stable training mode compared to training from scratch. The regressor is a multi-layer perceptron (MLP) composed of two fully connected dense layers respectively with 256 neurons and followed by batch normalization. The whole architecture is illustrated in Fig. 1.

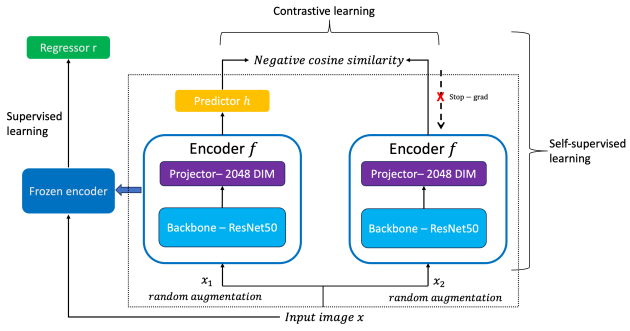

Fig. 1. Architecture of the self-supervised, contrastive learning regression model

2.3 Chest X-ray Radiograph Dataset and Experiment Setting

The CXR dataset used in our study is from the 2023 MIDRC mRALE Mastermind Challenge [19]. It contains 2,599 portable CXR images in the frontal anterior-posterior (AP) view with the mRALE scores annotated by expert radiologists and four quantitative scores representing the extent of involvement and the degree of density of each side of the lung. The images are stored in the DICOM format and accessible on the MIDRC mRALE Mastermind GitHub:
https://github.com/MIDRC/COVID19_Challenges.

The pre-processed CXR images are illustrated in Fig. 2. The images received augmentations including random flipping, color jittering, and color dropping to form two identical datasets where the index orders of each training image are the same. Note that we omitted the random cropping augmentation as recommended in the original SimSiam method because random cropping is likely to remove the alveolar opacities which are the crucial patterns for the final mRALE score.

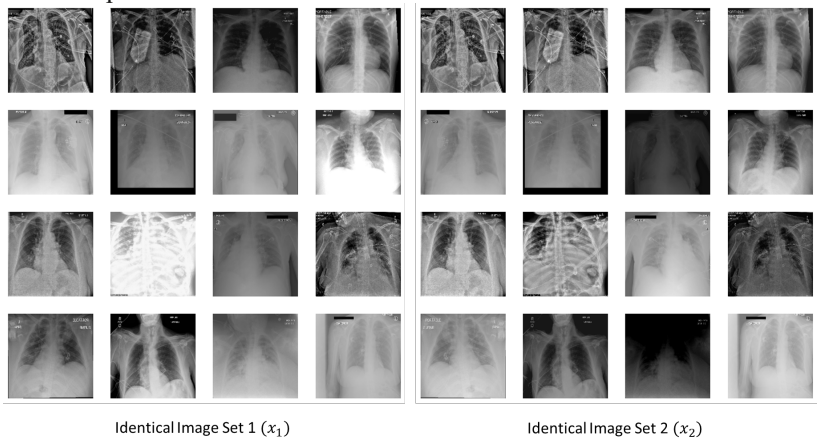

Fig. 2. Two identical image sets with random augmentations.

The retrieved images are resized to 224-by-224, 3 channels to match the input size of the pretrained ResNet-50 backbone. We split the dataset into five folds (520 images for each fold) for the cross-validation evaluation. The self-supervised contrastive learning regression model was implemented with TensorFlow version 2.12 in Python. The experiments were performed on the Amazon SageMaker Studio on a g4dn.2xlarge instance with a single Nvidia T4 tensor core GPU.

3 Results

3.1 Contrastive Learning

In the self-supervised contrastive learning phase, we used the pre-defined ResNet-50 network from the Keras API and removed the top classification layers as the backbone of the encoder. The output activation map is connected to a multi-layer perceptron (MLP) with L2 regularization with a weight decay rate of 1×10^{-3} and layer-wise batch normalization to stabilize the contrastive learning mode. We respective trained an encoder starting with the pretrain weights by ImageNet and another encoder starting from scratch. We first froze all the layers of the backbone and trained the encoders by stochastic gradient descent (SGD) with momentum=0.6 and cosine decay with initial learning rate of 0.02 with 15 epochs, then we unfroze the topmost 20 layers of the backbone except for the batch normalization layers and tuned the encoders with initial learning rate of 0.02 with another 20 epochs. The training process is shown in Fig. 3.

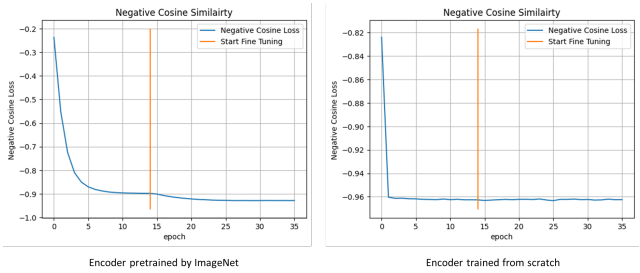

Fig. 3. Encoder optimized by self-supervised contrastive learning

Fig. 3 shows that the encoder starting with the pretrained weights by ImageNet (hot start) has smooth loss converge compared to the encoder trained from scratch (cold start) where the loss drop sharply to -0.96 after the second epoch and remained close to the minimum. This observation reflects the contrastive learning mode is successfully kept with the hot start while the training of the model with cold start was trapped at some local saddle points according to the original design of the SimSiam model [18].

3.2 Quantitative Prediction of mRALE

The weights of the encoders were frozen after optimized by the contrastive learning and they were connected to regressors and trained with the images and annotated scores to predict the CXR mRALE based on the lung edema patterns captured by the encoders. The regressors are MLPs with two fully connected layers with batch normalization. The regressor models were first trained for 20 epochs. Then we unfroze the topmost 20 layers of the encoder (except from the batch normalization layers) and fine- tuned the whole regressor model with the initial learning rate of 1×10^{-5} for 15 epochs. We also trained a ResNet-50 model with the identical regressor head from scratch as performance comparator to the new model. The gradient-weighted class activation mapping (Grad-CAM) is used to indicate the regression models' focus on the images. It helps to explain the regressor behavior when combined with the performance metrics. (Fig. 4)

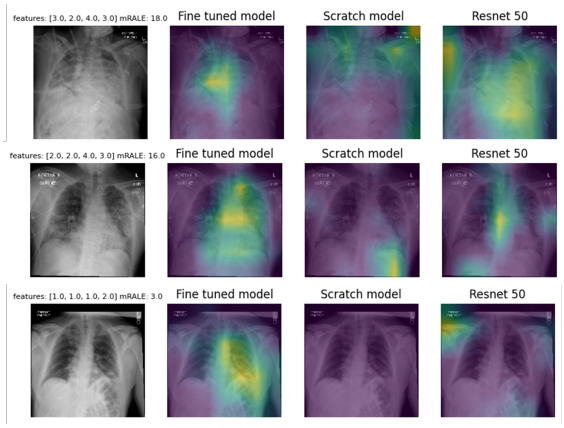

Fig. 4. Network focus regions indicated by Grad CAM

We performed a 5-fold cross validation to evaluate the overall performance of the three regressors with the metrics of mean squared error (MSE), mean absolute error (MSE) and the Spearman's correlation coefficient (Spearman ρ) reflecting the consistency of DL prediction to expert annotations. The mean scores with the 95% confidential interval (95% CI) are listed in Table 1. It indicates that the proposed regressor using an encoder with pretrained weights by ImageNet and trained by self-supervised contrastive learning has the best prediction performance, but the regressor trained from cold start shows no superiority compared to directly training a ResNet-50 for the mRALE regression task. The result is also supported by the Grad CAM where the self-supervised contrastive regressor with ImageNet pretrained weights can better focus on the pulmonary regions of the CXRs (see Fig. 4).

Table 1. Table captions should be placed above the tables.

Regressor	MSE (95% CI)	MAE (95% CI)	Spearman ρ (95% CI)
Self-supervised contrastive regressor (pretrain)	5.05 (5.03-5.08) *	66.67 (66.29-67.06) *	0.77 (0.75-0.79) *
Self-supervised contrastive regressor (scratch)	5.71 (5.68-5.76) *	71.25 (70.66-71.84) *	0.52 (0.45-0.59) *
ResNet-50 regressor	5.49 (5.39-5.58) *	70.09 (69.00-71.18) *▼	0.57 (0.51-0.65) *▼

*$p < 0.01$ in t-test for mean comparison, ▼ $p > 0.05$ in t-test for mean comparison

4 Conclusion and Discussion

We propose to use the self-supervised contrastive learning strategy to train a SimSiam based encoder as feature extractor for the regression model to predict mRALE score directly from the visual pattern of frontal CXRs. The results show that the self-supervised contrastive learning strategy combined with the pretrained weights by ImageNet can achieve convincing prediction performance with a small image dataset with limit training samples.

Future work should focus on using more feature localization methods such as vision transformer (ViT) with attention to further separate the meaningful vision patterns from the background by dividing the image into small patches and using multiple segmentation models to extract the region of interest (ROI) from the medical images to filter out the noise pattern. In addition, explainable AI technology such as Grad CAM is helpful to track and verify the DL model behaviors such that the randomness of DL can be minimized. We believe DL is a promising tool for automated medical image analysis and its performance can significantly improve when the knowledge acquired by AI is consistent to the long-term accumulative human expertise.

5 Funding

This research is supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health.

References

1. WHO: WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/. Accessed on July 2, 2023.
2. Taniguchi, H., Ohya, A., Yamagata, H., Iwashita, M, Abe, T, Takeuchi, I.: Prolonged mechanical ventilation in patients with severe COVID-19 is associated with serial modified-lung ultrasound scores: A single-centre cohort study. PLoS One 17(7), e0271391 (2022).
3. Valk, C.M.A., Zimatore, C., Mazzinari, G., Pierrakos, C., Sivakorn, C., Dechsanga, J., et al.: The Prognostic Capacity of the Radiographic Assessment for Lung Edema Score in Patients With COVID-19 Acute Respiratory Distress Syndrome-An International Multicenter Observational Study. Front Med (Lausanne) 8, 772056 (2021).
4. Warren, M.A., Zhao, Z., Koyama, T., Bastarache, J.A., Shaver, C.M., Semler, M.W., et al.: Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. Thorax 73(9), 840-6 (2018).
5. Matthay, M.A., Ware, L.B., Zimmerman, G.A.: The acute respiratory distress syndrome. J Clin Invest 122(8), 2731-40 (2012).
6. Voigt, I., Mighali, M., Manda, D., Aurich P, Bruder O.: Radiographic assessment of lung edema (RALE) score is associated with clinical outcomes in patients with refractory cardiogenic shock and refractory cardiac arrest after percutaneous implantation of extracorporeal life support. Intern Emerg Med 17(5), 1463-70 (2022).
7. Valk, C.M.A., Zimatore, C., Mazzinari, G., Pierrakos, C., Sivakorn, C., Dechsanga, J., et al.: The Prognostic Capacity of the Radiographic Assessment for Lung Edema Score in Patients With COVID-19 Acute Respiratory Distress Syndrome-An International Multicenter Observational Study. Front Med (Lausanne) 8, 772056 (2021).
8. Aggarwal, P., Mishra, N.K., Fatimah, B., Singh, P., Gupta, A., Joshi, S.D.: COVID-19 image classification using deep learning: Advances, challenges and opportunities. Comput Biol Med 144, 105350 (2022).
9. Khattab, R., Abdelmaksoud, I.R., Abdelrazek, S.: Deep Convolutional Neural Networks for Detecting COVID-19 Using Medical Images: A Survey. New Gener Comput 41(2), 343-400 (2023).
10. Xie, W., Jacobs, C., Charbonnier, J.P., van Ginneken, B.: Dense regression activation maps for lesion segmentation in CT scans of COVID-19 patients. Med Image Anal 86, 102771 (2023).
11. Meng, Y., Bridge, J., Addison, C., Wang, M., Merritt, C., Franks, S., et al.: Bilateral adaptive graph convolutional network on CT based Covid-19 diagnosis with uncertainty-aware consensus-assisted multiple instance learning. Med Image Anal 84, 102722 (2023).
12. Signoroni, A., Savardi, M., Benini, S., Adami, N., Leonardi, R., Gibellini, P., et al.: BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset. Med Image Anal 71, 102046 (2021).
13. Rahman, A., Hossain, M.S., Alrajeh, N.A., Alsolami, F.: Adversarial Examples-Security Threats to COVID-19 Deep Learning Systems in Medical IoT Devices. IEEE Internet Things J 8(12), 9603-10 (2021).
14. Li, Y., Liu, S.: The Threat of Adversarial Attack on a COVID-19 CT Image-Based Deep Learning System. Bioengineering (Basel) 10(2), 194 (2023).
15. Li, M.D., Arun, N.T., Gidwani, M., Chang, K., Deng, F., Little, B.P., et al.: Automated Assessment and Tracking of COVID-19 Pulmonary Disease Severity on Chest Radiographs using Convolutional Siamese Neural Networks. Radiol Artif Intell 2(4), e200079 (2020).
16. Horng, S., Liao, R., Wang, X., Dalal, S., Golland, P., Berkowitz, S.J.: Deep Learning to Quantify Pulmonary Edema in Chest Radiographs. Radiol Artif Intell 3(2), e190228 (2021).

17. Tian, Y., Chen, X., Ganguli, S.: Understanding self-supervised learning dynamics without contrastive pairs. In Proceedings of the 38th International Conference on Machine Learning, vol. 139, pp. 10268-10278. MLR Press, (2021).
18. Chen, X., He, K.: Exploring Simple Siamese Representation Learning. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville (2021).
19. MIDRC. MIDRC mRALE Mastermind Challenge: AI to predict COVID severity on chest radiographs. https://www.midrc.org/mrale-mastermind-2023. Accessed on July 2, 2023.