A Comparison Study to Detect COVID-19 Chest X-Ray Images with SOTA Deep Learning Models

Qingzhong Liu [1] **Zhongxue Chen** [2] **Hnery C. Liu** [3]

Abstract

By using a recently released chest X-ray (CXR) image database for COVID-19 positive cases along with Normal, Lung Opacity (Non-COVID lung infection), and Viral Pneumonia images, this study compares the performance of SOTA deep learning models in detecting COVID-19 CXR images. Pre-trained deep learning models are re-trained under several combinations of optimizers, learning rate schedulers, and loss functions. Our study shows that these SOTA deep learning models perform well if the models and parameters are selected meticulously. Overall, EfficientNet is superior to others especially across different optimizers. Regarding the loss function, the integration of cosine embedding similarity and cross entropy is slightly better than cross entropy itself while we adopt the SGD optimizer. In terms of optimizer, SGD constantly performs well while Adam and AdamW are unstable across different models.

1. Introduction

Coronavirus disease 2019 (COVID-19) is a contagious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) and it was first identified in Wuhan, China in 2019 (Page et al., 2021). The virus spreads worldwide, leading to the COVID-19 pandemic. As of May 27th, 2022, it has confirmed over 500 million COVID-19 cases and over 6 million deaths (ArcGIS, 2022).

People with COVID-19 may be asymptomatic or experience one or more of the following symptoms: fever or chills, cough, myalgia, headache, fatigue, breathing difficulties, loss of smell or taste, sore throat, congestion or runny nose, nausea or vomiting, diarrhea (Stokes et al., 2020). At least a third of infected people do not develop noticeable symptoms (Oran and Topol, 2021). In an analysis of more than 1.3 million laboratory-confirmed cases of COVID-19 that were reported in the United States between January and May 2020, 14% of patients required hospitalization, 2% were admitted to the intensive care unit, and 5% died (Stokes et al., 2020).

Testing is important to identify and help reduce the spread of COVID-19. Viral tests, including a nucleic acid amplification test (NAAT) and antigen tests, are used to diagnose COVID-19. Antibody tests (serology) are not indicated to diagnose a current infection. NAATs that use reverse transcription-polymerase chain reaction (RT-PCR) technology to detect SARS-CoV-2 ribonucleic acid (RNA) are highly sensitive and specific to detect SARS-CoV-2 RNA in respiratory specimens. Clinical RT-PCR tests for SARS-CoV-2 that determine the cycle threshold (Ct) value are not validated to determine viral load, and the NIH recommends that Ct values should be used clinically in consultation (COVID-19 Treatment, 2022).

Chest radiographs of patients with severe COVID-19 may demonstrate bilateral air-space consolidation (Sadiq et al., 2021). Chest computed tomography (CT) images from patients with COVID-19 may demonstrate bilateral, peripheral ground glass opacities and consolidation (Kanne et al., 2021; Doerschug and Schmidt, 2022). Less common CT findings can include intra- or inter-lobular septal thickening with ground glass opacities (crazy paving pattern) or focal and rounded areas of ground glass opacity surrounded by a ring or arc of denser consolidation (reverse halo sign).

In the setting of the COVID-19 pandemic, chest imaging plays a very important role in the early diagnosis and the treatment planning for patients with suspected or confirmed COVID-19 chest infections, as it is readily available in the community physician offices, urgent care clinics and hospital emergency departments. The authors (Cleverley et al., 2020) found that no single feature of covid-19 pneumonia on a chest radiograph is specific or diagnostic, but a combination of multifocal peripheral lung changes of ground glass opacity and/or consolidation, which are most com-

*Equal contribution [1]Department of Computer Science, Sam Houston State University, Huntsville, TX 77382, U.S.A. [2]Department of Epidemiology and Biostatistics, Indiana University Bloomington, Bloomington, IN 47405, U.S.A. [3]College of Natural Science, University of Texas at Austin, Austin, TX 78712, U.S.A.. Correspondence to: Qingzhong Liu <liu@shsu.edu>.

Proceedings of the 1st Workshop on Healthcare AI and COVID-19, ICML 2022.

monly bilateral, may be present, and the diagnosis might be complicated as covid-19 pneumonia may or may not be visible on chest radiograph while considering other causes for patients' respiratory symptoms. The authors (Smith et al., 2020) suggested that the presence of patchy and/or confluent, bandlike ground-glass opacity or consolidation in a peripheral and mid to lower lung zone distribution on a chest radiograph obtained in the setting of pandemic COVID-19 was highly suggestive of severe acute respiratory syndrome coronavirus 2 infection and should be used in conjunction with clinical judgment to make a diagnosis. In the study (Yasin and Gouda, 2020), there were 350 patients proven with positive COVID-19 disease; 220 patients (62.9%) had abnormal baseline chest X-ray (CXR) and 130 patients (37.1%) had normal baseline CXR. Their study demonstrates that radiographic findings are very good predictors for assessing the course of COVID-19 disease and it could be used as long-term consequences monitoring.

2. Related Study

To address the challenges caused by the pandemic, artificial intelligence has been widely used to help provide solutions to save lives and to stop the pandemic outbreak. In detecting COVID-19 from CXR images, Guefrechi et al. utilize fine-tuned deep learning models ResNet50, InceptionV3, and VGG15 to distinguish COVID-19 CXR images from normal CXR images (Guefrechi et al., 2021). The results show that transfer learning is effective, showing strong performance and easy-to-deploy COVID-19 detection, which may be used in cases wherein the materials and RT-PCR tests are limited. Unfortunately, the dataset contains only COVID-19 and normal CXR images, other viral Pneumonia images are not involved. Second, the latest deep learning models are missing in the study. The third, the hyper parameters including optimizers and learning rate schedulers, which may be very important to the detection performance, are not discussed.

The authors in the article (Akter et al., 2021) applied a modified MobileNetV2 to COVID-19 CXR images. The resulting model produced the highest accuracy of 98% in classifying COVID-19 and healthy chest X-rays among all the implemented CNN models. The results suggest that the proposed method can efficiently identify the symptoms of infection from chest X-ray images better than existing methods. Similarly, the data only contains two classes: COVID-19 and healthy chest X-rays. Nayak et al. compared eight pre-trained deep CNN models, namely, VGG-16, Inception-V3, ResNet-34, MobileNet-V2, AlexNet, GoogleNet, ResNet-50, and SqueezeNet for COVID-19 CXR images, and they claimed that ResNet-34 outperformed other competitive networks (Nayak et al., 2021).

Hussain et al. designed a 22-layer CNN architecture, which has achieved an accuracy of 99.1% for 2 class classification, 94.2% for 3 class classification, and 91.2% for 4 class classification. The numbers of images used to train the model for 4-class classification are 500 COVID-19, 800 normal, 400 pneumonia-viral, and 400 pneumonia-bacteria (Hussain et al., 2021). The authors in the article(Okolo et al., 2021) applied 11 CNN models for the task of classifying COVID-19, normal and viral pneumonia X-ray images (3-class problem). The authors claimed that that the EfficientNetB4- and the Xception-based models perform the best.

In the article (Kumar et al., 2022), modified versions of VGG16, VGG19, ResNet50, and InceptionV3 were applied to a total of 720 CXR images, containing 540 normal and 180 COVID-19, and the authors claimed that the modified InceptionV3 performed the best. The authors (Sampen and Lavarello, 2022) compared three deep learning architectures (COVID-net, CovXNet and DarkCovidNet) with CXR images with COVID-19, pneumonia and healthy (3305 images for each class), the study claims that DarkCovidNet achieved 94.04% and CovXNet obtained 92.02%. Recently, Li and Li compared 17 different deep learning models and applied the ensemble of these 17 models. Among viral pneumonia, diagnostic accuracy for Covid-19 reaches 99.95%. The authors claimed that high diagnostic accuracy was achieved for distinguishing Covid-19 pneumonia from bacterial pneumonia (Li and Li, 2022).

Although many studies have been conducted for the CXR-based diagnosis of COVID-19, unfortunately, the latest deep learning models have not been thoroughly investigated yet. Additionally, most studies are constrained in the limited CXR data or just limited to the binary classification of COVID-19 and healthy CXR images. Regarding deep learning itself, different hyper-parameters, optimizers, learning rate schedulers, and the loss functions may have significant impacts on the performance, which have not been carefully investigated. In this study, we aim to adopt the STOA deep learning models that have been proven in the ImageNet benchmark testing, transfer these pre-trained models with fine-tuning to COVID-19 CXR images, and compare the performance with different optimizers, learning rate schedulers, and loss functions.

3. Proposed Study

3.1. Deep Learning Models

The ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al., 2015) evaluates algorithms for object detection and image classification at large scale. It contains hundreds and thousands of images, which has been instrumental in advancing computer vision and deep learning research. PyTorch Image Models named timm is a library for state-of-the-art image classification, containing

a collection of image models, optimizers, schedulers, augmentations, and training/validating scripts with ability to reproduce ImageNet training result (Wightman, 2019). It contains validation and benchmark results for the models in this collection. We select the following top-ranking models in the classification of ImageNet-1k dataset, listed in the Table 1, for our study in COVID-19 CXR image recognition. These SOTA models are selected from Bidirectional Encoder representation from Image Transformers (BEiT) (Bao et al., 2021), a family of pure ConvNet models dubbed ConvNeXt, ConvNeXts (Liu et al., 2022), a new vision Transformer, called Swin Transformer (Liu et al., 2021), EfficientNet (Xie et al., 2020) and EfficientNetV2 (Tan and Le, 2021), and Vision Transformer (Dosovitskiy et al., 2020).

Table 1. Deep learning models used in our study.

MODEL_NO	MODEL_NAME
1	$beit\_large\_patch16\_512$
2	$beit\_large\_patch16\_384$
3	$beit\_large\_patch16\_224$
4	$beit\_base\_patch16\_384$
5	$convnext\_large\_384\_in22ft1k$
6	$convnext\_xlarge\_in22ft1k$
7	$convnext\_large\_in22ft1k$
8	$convnext\_base\_384\_in22ft1k$
9	$swin\_large\_patch4\_window12\_384$
10	$swin\_base\_patch4\_window12\_384$
11	$swin\_large\_patch4\_window7\_224$
12	$tf\_efficientnet\_b6\_ns$
13	$tf\_efficientnetv2\_xl\_in21ft1k$
14	$tf\_efficientnetv2\_l\_in21ft1k$
15	$vit\_large\_patch16\_384$
16	$vit\_large\_r50\_s32\_384$

3.2. Selection of Optimizers and Learning Rate Schedulers

In deep learning, optimizers can be explained as a mathematical function to modify the weights of the network given the gradients and additional information, depending on the formulation of the optimizer. Optimizers are built upon the idea of gradient descent, the greedy approach of iteratively decreasing the loss function by following the gradient. The pre-trained pytorch image models (Wightman, 2019) were trained with stochastic gradient descent (SGD) optimizer. Therefore, we adopt SGD as a primary optimizer in out study, and briefly introduce below.

Consider the object function

$$J(\omega) = -\frac{1}{n}\sum_{i=1}^{n} J^{(i)}(\omega) \tag{1}$$

Where the parameter ω that minimizes $J(\omega)$ is to be estimated. Each summand function $J^{(\mathrm{i})}(\omega)$ is typically associated with the i-th observation in the dataset for training.

To minimize the object function, a standard gradient descent method is performed in the following iterations:

$$\omega := \omega - \eta\nabla J(\omega) = \omega - \frac{\eta}{n}\sum_{i=1}^{n}\nabla J^{(i)}(\omega) \tag{2}$$

Where η is a step size or called the learning rate.

Stochastic gradient descent with momentum remembers the update $\Delta\omega$ at each iteration, and determines the next update as a linear combination of the gradient and the pervious update:

$$\Delta\omega := \alpha\Delta\omega - \eta\nabla J^{(i)}(\omega) \tag{3}$$

$$\omega := \omega + \Delta\omega \tag{4}$$

That leads to:

$$\omega := \omega - \eta\nabla J^{(i)}(\omega) + \alpha\Delta\omega \tag{5}$$

Where the parameter ω which minimize $J(\omega)$ is to be estimated, η is a step size or called the learning rate, and α is an exponential decay factor between 0 and 1 that determines the relative contribution of the current gradient and earlier gradients to the weight change.

Adaptive optimizers like Adam (Kingma and Ba, 2015) have become a popular choice for training neural networks, however it is observed to not generalize as well. Loshchilov and Hutter (Loshchilov and Hutter, 2019) demonstrate that L2 regularization is significantly less effective for adaptive algorithms than for SGD. They propose an improved version of Adam called AdamW that yields much better model, and it can compete with SGD while training faster. In this study, we consider SGD as a primary optimizer, also include Adam and AdamW on the optimizer list.

Recently, a study is conducted to investigate the liver segmentation task with two popular learning rate schedulers, Pytorch OneCycleLR and Pytorch ReduceLROnPlateau. The study shows that both schedulers perform well and Pytorch ReduceLROnPlateau is slightly better (Akter et al., 2022). In general, it reduces learning rate when a metric has stopped improving. This scheduler reads a metrics quantity and if no improvement is seen for a 'patience' number of epochs, the learning rate is reduced. We will adopt the ReduceLROnPlateau as our first learning rate scheduler.

Loshchilov and Hutter propose a simple warm restart technique for stochastic gradient descent. Set the learning rate of each parameter group using a cosine annealing schedule, where $\eta_{\max}$ is set to the initial rate, T_{cur} is the number of epochs since the last restart and T_{i} is the number of epochs between two warm restarts (Loshchilov and Hutter, 2017).

$$\eta_{\mathrm{t}} = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})(1 + cos(\frac{T_{\mathrm{cur}}}{T_{\mathrm{t}}}\pi)) \tag{6}$$

When $T_{cur} = T_i$, set $\eta_t = \eta_{min}$. When $T_{cur} = 0$, set $\eta_t = \eta_{max}$.

In this study, we select Pytorch cosineAnnealingWarmRestarts as our second learning rate scheduler.

3.3. Loss Functions

Cross entropy is frequently used for loss function. For a multiple-class classification, the loss function is given by

$$L(\hat{y}, y) = -\sum_{c=1}^{M} y^{(c)} \log(\hat{y}^{(c)}) \tag{7}$$

Where $y^{(c)}$ is the 0 or 1, indicating whether class label c is the correct classification and $\hat{y}^{(c)}$ is the prediction on the class label c.

Cosine embedding loss measures the loss given inputs x_1, x_2 and a label tensor y containing values 1 or -1. It is used for measuring whether two inputs are similar or dissimilar, using the cosine distance, and is typically used for learning nonlinear embeddings or semi-supervised learning. The loss function for each sample is

$$L(x, y) = \begin{cases} 1 + \cos(x_1, x_2) & \text{if} \quad y = 1 \\ max(0, \cos(x_1, x_2) - margin) & \text{if} \quad y = -1 \end{cases} \tag{8}$$

3.4. Dataset and Experimental Setup

We adopt the COVID-19 CXR image dataset (Chowdhury et al., 2020; Rahman et al., 2021), which contains 3616 COVID-19, 6012 lung opacity, 10192 normal, and 1345 viral Pneumonia image files. We randomly select one CXR image from each class, shown in Figure 1.

We randomly select 72% from each class for training and 13% for validation, and the remaining 15% for testing. In each experiment, the same training, validation, and testing are applied to each fine-tuning deep learning model. We select the initial learning rate individually from 0.0005, 0.001, 0.005, 0.01, and 0.05, overall, the models with the initial learning rate 0.001 are better than the models on other initial learning rates. We compare the cross-entropy loss and a combination of cross entropy and cosine embedding loss with the weight ratio of cross entropy to cosine embedding is 1:10. The dataset contains imbalanced classes, while we randomly selected samples for training, different weights are assigned to the four classes, computing by the total number of all class images over the number of each class. The total number is 21165, the weights are 5.85 for COVID-19, 3.52 for lung opacity, 2.08 for normal, and 15.74 for viral Pneumonia.

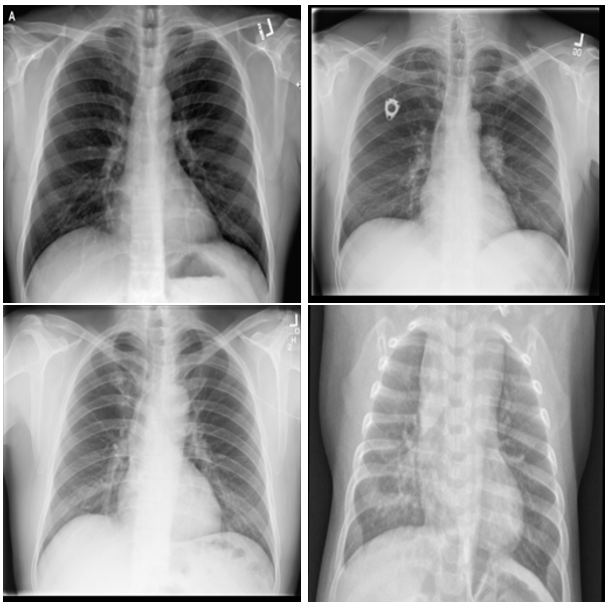

Figure 1. Image examples: COVID-19 (upper left), Lung opacity (upper right), normal (lower left), and viral pneumonia (lower right)

4. Experimental Results

Table 2 shows the mean testing accuracy of the 16 models with the six combinations of optimizer, loss function, and learning rate scheduler, wherein cosine_entropy is the loss of cosine embedding loss +0.1 cross entropy, noted as cosine_entropy in this paper, and ReduceLROnPlateau is named Plateau in Table 2.

It shows that the combination of Adam and AdamW with cosineAnnealing does not perform well and the performance varies across different models. Besides the optimizer, the testing accuracy results in Table 2 show that the loss function has significant impact on some models, e.g., BEiT models. On average, the testing accuracy with the cosine_entropy loss is better than entropy loss function over different deep learning models. The results in Table 2 also demonstrate that EfficientNet family generally performs well across the six combinations. In each combination, the highest testing accuracy is associated with EfficientNet models.

Figure 2 shows the boxplots to compare the four combinations with the SGD optimizer. It indicates that the integration of the loss function cosine embedding similarity with cross entropy is slightly better than cross entropy alone. ReduceLROnPlateau is comparable to consineannealing learning rate scheduler in our experiments.

Tables 3 to 18 present the normalized confusion matrix (NCM) of the testing results with SGD, cosine_entropy, and cosineannealing learning rate scheduler, wherein pre-

Table 2. Testing accuracy (%)with differnt optimizers, loss functions, and learning rate schedulers.

Model_No	SGD Cosine_Entropy Plateau	SGD Cosine_Entropy CosineAnnealing	SGD Entropy Plateau	SGD Entropy CosineAnnealing	Adam Cosine_Entropy CosineAnnealing	AdamW Cosine_Entropy CosineAnnealing
1	94.5	94.6	81.4	82.3	49.1	45.9
2	93.2	93.0	85.7	90.9	46.7	85.8
3	94.5	93.2	94.1	94.2	38.2	70.3
4	95.1	93.6	85.0	92.7	36.3	87.7
5	95.0	94.7	95.1	95.2	27.9	27.2
6	94.5	94.9	94.4	93.5	16.9	85.2
7	93.5	94.6	95.0	93.3	49.0	47.4
8	95.3	94.2	95.1	94.6	49.0	17.1
9	95.7	95.2	95.2	94.0	49.0	49.4
10	95.7	95.7	95.3	93.4	6.2	16.6
11	94.6	94.8	94.5	94.7	6.2	44.6
12	94.9	96.3	95.3	95.6	92.8	93.0
13	95.9	95.7	96.9	95.3	80.5	92.5
14	95.9	95.9	95.9	96.1	85.1	93.6
15	95.6	95.0	95.5	95.8	22.1	55.3
16	95.6	95.1	94.6	93.4	27.9	17.8

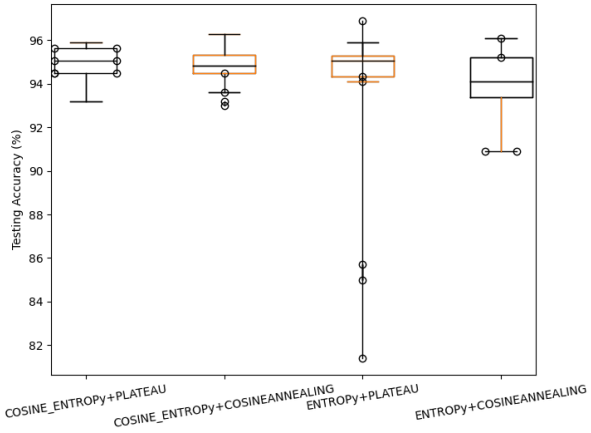

Figure 2. Boxplots of the testing results on the four combinations with SGD optimizer

diction is abbreviated as 'pred'. The results from Tables 3-18 show that all SOTA models perform well by using the SGD optimizer. Regarding the prediction of each class of CXR images, we have the following observations from the 16 models: the top models among the 16 models are convnext_base_384_in22ft1k (99.4%) for detecting COVID-19, beit_large_patch16_512 (95.4%) in predicting lung opacity, swin_large_patch4_window12_384 (96.7%) in detecting normal, and tf_efficientnetv2_xl_in21ft1k (100%) in identifying viral pneumonia images. Overall, all five SOTA deep learning architectures perform well. If we combine all results with Table 2, EfficientNet is slightly superior to others, especially with Adam and AdamW optimizers, the advantage of EfficientNet is noticeable.

Table 3. NCM with beit_large_patch16_512.

Accuracy (%)		Truth COVID	Opacity	Normal	Viral
Pred	COVID	98.9	0.5	0.4	0.0
	Opacity	0.7	95.4	6.8	0.0
	Normal	0.2	4.0	92.2	2.5
	Viral	0.2	0.1	0.5	97.5

Table 4. NCM with beit_large_patch16_384.

Accuracy (%)		Truth COVID	Opacity	Normal	Viral
Pred	COVID	96.5	0.5	0.9	1.4
	Opacity	1.2	90.4	6.2	0.0
	Normal	1.9	9.1	92.0	4.2
	Viral	0.3	0.0	0.9	94.4

Table 5. NCM with beit_large_patch16_224.

Accuracy (%)		Truth COVID	Opacity	Normal	Viral
Pred	COVID	95.7	2.4	2.0	1.5
	Opacity	3.4	80.7	11.6	0.0
	Normal	0.6	16.8	84.5	4.5
	Viral	0.4	0.1	2.0	94.0

Table 6. NCM with beit_base_patch16_384.

Accuracy (%)		Truth COVID	Opacity	Normal	Viral
Pred	COVID	95.9	0.9	0.3	0.5
	Opacity	0.2	92.0	7.1	0.0
	Normal	3.0	7.1	91.9	0.5
	Viral	0.9	0.1	0.7	99.1

Table 7. NCM with convnext_large_384_in22ft1k.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	97.6	0.3	0.3	0.0
	Opacity	0.5	91.0	2.6	0.0
	Normal	1.6	8.7	96.6	3.2
	Viral	0.2	0.0	0.5	96.8

Table 8. NCM with convnext_xlarge_in22ft1k.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	97.6	0.9	0.4	0.0
	Opacity	1.1	92.8	4.5	0.0
	Normal	1.1	6.3	94.5	2.7
	Viral	0.2	0.0	0.6	97.3

Table 9. NCM with convnext_large_in22ft1k.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	98.4	1.9	1.7	0.5
	Opacity	1.0	93.6	5.1	0.0
	Normal	0.5	4.5	92.4	0.5
	Viral	0.0	0.0	0.9	99.0

Table 10. NCM with convnext_base_384_in22ft1k.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	99.4	1.2	0.3	1.1
	Opacity	0.6	92.7	5.1	0.0
	Normal	0.0	6.1	93.9	0.5
	Viral	0.0	0.0	0.6	98.4

Table 11. NCM with swin_large_patch4_window12_384.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	98.2	1.0	0.3	1.0
	Opacity	0.5	89.5	2.4	0.0
	Normal	1.2	9.5	96.7	3.4
	Viral	0.0	0.0	0.6	95.6

Table 12. NCM with swin_base_patch4_window12_384.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	99.2	0.2	0.3	0.0
	Opacity	0.4	93.5	3.9	0.0
	Normal	0.2	6.2	95.3	1.4
	Viral	0.2	0.0	0.6	98.6

Table 13. NCM with swin_large_patch4_window7_224.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	99.3	0.4	0.5	0.5
	Opacity	0.2	93.5	2.9	0.0
	Normal	0.5	6.0	95.8	1.5
	Viral	0.0	0.0	0.8	98.0

Table 14. NCM with tf_efficientnet_b6_ns.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	98.5	0.6	0.4	0.5
	Opacity	0.4	92.4	4.0	0.0
	Normal	1.1	7.1	94.8	1.4
	Viral	0.0	0.0	0.9	98.1

Table 15. NCM with tf_efficientnetv2_xl_in21ft1k.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	97.6	0.3	0.2	0.0
	Opacity	1.4	93.6	5.1	0.0
	Normal	0.8	6.1	93.9	0.0
	Viral	0.2	0.0	0.8	100.0

Table 16. NCM with tf_efficientnetv2_l_in21ft1k.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	97.1	0.8	0.3	0.9
	Opacity	0.6	90.6	2.5	0.0
	Normal	1.6	8.7	96.2	1.4
	Viral	0.8	0.0	1.0	97.7

Table 17. NCM with vit_large_patch16_384.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	98.6	0.2	1.0	0.0
	Opacity	0.6	92.5	3.8	0.0
	Normal	0.4	7.3	94.7	9.6
	Viral	0.4	0.0	5.8	99.0

Table 18. NCM with vit_base_patch16_384.

Accuracy (%)		Truth			
		COVID	Opacity	Normal	Viral
Pred	COVID	90.5	5.0	3.8	2.4
	Opacity	4.7	79.1	10.0	0.5
	Normal	4.1	15.3	82.1	2.9
	Viral	0.7	0.6	4.2	94.3

5. Conclusion

In this study, we compare the 16 SOTA deep learning models for detecting COVID-19 CXR images for four-class classification. Experimental results show that the hyper-parameters, the optimizers and the loss function have significant impact on the detection performance. The SGD optimizer and the integration of cosine embedding similarity and cross entropy loss are ideal for different models. Both learning rate schedulers ReduceLRonPlateau and cosineAnnealingWarmRestarts perform well. Overall, if parameters are selected meticulously, all models are effective with the SGD optimizer. While Adam and AdamW are selected as the optimizer, the performances of most models except EfficientNet are noticeably unstable.

Acknowledgements

We are thankful to anonymous reviewers for their insightful comments and revising suggestions.

References

Akter, S., Shamrat, F., Chakraborty, S., Karim, A., and Azam, S. (2021). COVID-19 Detection Using Deep Learning Algorithm on Chest X-ray Images. *Biology*, 10(11), 1174. https://doi.org/10.3390/biology10111174.

Al-Kababji, A., Bensaali, F., Dakua, S. P. (2022). Scheduling Techniques for Liver Segmentation: ReduceLRonPlateau Vs OneCycleLR. arXiv preprint arXiv:2202.06373.

ArcGIS. Johns Hopkins University (2022). COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Retrieved 27 May 2022.

Bao, H., Dong, L., and Wei, F. (2021). BEiT: BERT Pre-Training of Image Transformers. ArXiv, abs/2106.08254.

Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Iqbal, A., Al-Emadi, N., Reaz, M. B. I/, Islam, M. T. (2020). Can AI help in screening Viral and COVID-19 pneumonia? *IEEE Access*, Vol. 8, pp. 132665 - 132676.

Cleverley, J., Piper, J., Jones, M. M. (2020). The role of chest radiography in confirming covid-19 pneumonia. *BMJ* 2020; 370:m2426.

COVID-19 Treatment Guidelines Panel. (2022). Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. *National Institutes of Health*, https://www.covid19treatmentguidelines.nih.gov/. Accessed on May 28, 2022.

Doerschug, K. C., Schmidt, G. A. (2022). Pulmonary Aspects of COVID-19. *Annual Review of Medicine*. 73(1):81-93. doi:10.1146/annurev-med-042220-014817external icon.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929v2.

Guefrechi, S., Jabra, M. B., Ammar, A., Koubaa, A. and Hamam, H. (2021). Deep learning based detection of COVID-19 from chest X-ray images. *Multimedia tools and applications*. 80(21-23), 31803–31820.

Hussain, E., Hasan, M., Rahman, M. A., Lee, I., Tamanna, T., Parvez, M. Z. (2021). CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. *Chaos, Solitons & Fractals*. Volume 142, 2021, 110495, ISSN 0960-0779.

Kanne, J. P., Bai, H., Bernheim, A., Chung, M., Haramati, L. B., Kallmes, D. F., Little, B. P., Rubin, G. D., and Sverzellati, N. (2021). COVID-19 Imaging: What We Know Now and What Remains Unknown. *Radiology*. 299(3), E262–E279. https://doi.org/10.1148/radiol.2021204522.

Kingma, D. P., Ba, J. (2015). Adam: A Method for Stochastic Optimization. Proc. 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015.

Kumar, V., Zarrad, A., Gupta, R., Cheikhrouhou, O. (2022). COV-DLS: Prediction of COVID-19 from X-Rays Using Enhanced Deep Transfer Learning Techniques. *Journal of Healthcare Engineering*. 2022:6216273. doi: 10.1155/2022/6216273. PMID: 35422979; PMCID: PMC9002900.

Li, D., Li, S. (2022). An artificial intelligence deep learning platform achieves high diagnostic accuracy for Covid-19 pneumonia by reading chest X-ray images. *iScience*. 25(4):104031. doi: 10.1016/j.isci.2022.104031. Epub 2022 Mar 6. PMID: 35280932; PMCID: PMC8898091.

Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., Xie, S. (2022). A ConvNet for the 2020s. arXiv:2201.03545v2.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. and Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 9992-10002.

Loshchilov, I. and Hutter, F. (2019). Decoupled Weight Decay Regularization. CLR (Poster) 2019.

Loshchilov, I. and Hutter, F. (2017). SGDR: Stochastic Gradient Descent with Warm Restarts. CLR (Poster) 2017.

Nayak, S. R., Nayak, D. R., Sinha, U., Arora, V., Pachori, R. B. (2021). Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study. *Biomedical Signal Processing and Control*. Volume 64, 102365, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2020.102365.

Okolo, G. I., Katsigiannis, S., Althobaiti, T., Ramzan, N. (2021). On the Use of Deep Learning for Imaging-Based COVID-19 Detection Using Chest X-rays. *Sensors*. 21(17):5702.

Oran, D. P., Topol, E. J. (2021). The Proportion of SARS-CoV-2 Infections That Are Asymptomatic: A Systematic Review. *Annals of Internal Medicine*. 174 (5): M20-6976. doi:10.7326/M20-6976. PMC 7839426. PMID 33481642.

Page, J., Hinshaw, D., McKay, B. (2021). In Hunt for Covid-19 Origin, Patient Zero Points to Second Wuhan Market – The man with the first confirmed infection of the new coronavirus told the WHO team that his parents had shopped there. The Wall Street Journal. Retrieved 27 February 2021.

Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S. B. A., Islam, M. T., Maadeed, S. A., Zughaier, S. M., Khan, M. S. and Chowdhury, M. E. (2021). Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-ray Images. *Computers in Biology and Medicine*.Volume 132, 104319, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2021.104319.

Russakovsky, O., Deng, J., Su, H. et al. (2015). ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis*. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Sadiq, Z., Rana, S., Mahfoud, Z., Raoof, A. (2021). Systematic review and meta-analysis of chest radiograph (CXR) findings in COVID-19. *Clinical Imaging*. 80, 229–238. https://doi.org/10.1016/j.clinimag.2021.06.039.

Sampén, D. and Lavarello, R. (2022). Comparison of deep learning architectures for COVID-19 diagnosis using chest X-ray images. Proc. SPIE 12035, Medical Imaging 2022: Image Perception, Observer Performance, and Technology Assessment, 120350Z (4 April 2022); https://doi.org/10.1117/12.2613002.

Smith, D. L., Grenier,J. P., Batte, C., Spieler, B. (2020). A Characteristic Chest Radiographic Pattern in the Setting of the COVID-19 Pandemic. *Radiol Cardiothorac Imaging*. 2(5):e200280. doi: 10.1148/ryct.2020200280. PMID: 33778626; PMCID: PMC7605076.

Stokes, E. K., Zambrano, L. D., Anderson, K. N., et al. (2020). Coronavirus Disease 2019 Case Surveillance — United States, January 22–May 30, 2020. *MMWR Morb Mortal Wkly Rep 2020*. 69:759–765. DOI: http://dx.doi.org/10.15585/mmwr.mm6924e2.

Tan, M. and Le, Q . (2021). EfficientNetV2: Smaller Models and Faster Training. ArXiv, abs/2104.00298.

Wightman, R. (2019). *PyTorch Image Models*. GitHub repository, doi :10.5281/zenodo.4414861.

Xie, Q., Luong, M. T., Hovy, E. and Le, Q. V. (2020). Self-Training with Noisy Student Improves ImageNet Classification. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10684-10695, doi: 10.1109/CVPR42600.2020.01070.

Yasin, R. and Gouda, W. (2020). Chest X-ray findings monitoring COVID-19 disease course and severity. *Egypt J Radiol Nucl Med* 51, 193. https://doi.org/10.1186/s43055-020-00296-x.