

Received 30 August 2022, accepted 14 September 2022, date of publication 20 September 2022,
date of current version 28 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3208138

SURVEY

A Comprehensive Review of Deep Learning-Based Methods for COVID-19 Detection Using Chest X-Ray Images

SAEED S. ALAHMARI¹, (Member, IEEE), BADERALDEEN ALTAZI^{2,5}, JISOO HWANG³,
SAMUEL HAWKINS⁴, (Member, IEEE), AND TAWFIQ SALEM³, (Member, IEEE)

¹Department of Computer Science, Najran University, Najran 66462, Saudi Arabia

²Batterjee Medical College, College of Medicine, Jeddah 21442, Saudi Arabia

³Department of Computer and Information Technology, Purdue University, West Lafayette, IN 47907, USA

⁴Department of Computer Science and Information Systems, Bradley University, Peoria, IL 61625, USA

⁵Department of Medical Physics, King Abdullah Medical City at The Holy Capital, Mecca 24246, Saudi Arabia

Corresponding author: Saeed S. Alahmari (ssalahmari@nu.edu.sa)

This work was supported by Najran University through the Deanship of Scientific Research under Grant NU-/SERC/10/590.

ABSTRACT The novel coronavirus disease 2019 (COVID-19) added tremendous pressure on healthcare services worldwide. COVID-19 early detection is of the utmost importance to control the spread of the coronavirus pandemic and to reduce pressure on health services. There have been many approaches to detect COVID-19; the most commonly used one is the nasal swab technique. Before that was available chest X-ray radiographs were used. X-ray radiographs are a primary care method to reveal lung infections, which allows physicians to assess and plan a course of treatment. X-ray machines are prevalent, which makes this method a preferable first approach for the detection of new diseases. However, this method requires a radiologist to assess each chest X-ray image. Therefore, different automated methods using machine learning techniques have been proposed to assist in speeding up diagnoses and improving the decision-making process. In this paper, we review deep learning approaches for COVID-19 detection using chest X-ray images. We found that the majority of deep learning approaches for COVID-19 detection use transfer learning. A discussion of the limitations and challenges of deep learning in radiography images is presented. Finally, we provide potential improvements for higher accuracy and generalisability when using deep learning models for COVID-19 detection.


INDEX TERMS Machine learning, pneumonia, radiology, diagnostic imaging, COVID-19.

I. INTRODUCTION

The novel coronavirus disease 2019 (COVID-19) rapidly spread quickly causing a global pandemic. The first incidence of the virus was identified in early December 2019 in Wuhan, the People's Republic of China. Strong measures were applied to control the spread of infection. Some of these measures included shutdowns, isolation, and close monitoring of contacts, which have caused economic crisis, recession, and affected the mental well-being of many individuals around the world [1], [2], [3]. The World Health Organization (WHO) declared the COVID-19 outbreak as a global pandemic on March 11th, 2020 [4]. The death toll of the

resulting pandemic was 5.04 million confirmed death and 249.54 million confirmed cases of infection worldwide as of November 6, 2021 [5].

The first few months of the pandemic were challenging for hospitals, medical teams, and governments to control and test millions of infected people. Healthcare systems could not keep up with the significant number of infected cases, and in some countries, the healthcare system collapsed under the COVID-19 surge. The rapid spread of the virus increased the need for early detection of positive COVID-19 cases and faster diagnosis [6] to help control and better understand the pandemic. Reverse Transcription Polymerase Chain Reaction test (RT-PCR) is the current standard tool used to detect COVID-19 infection. However, RT-PCR is a time-consuming and expensive process and it has reported

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

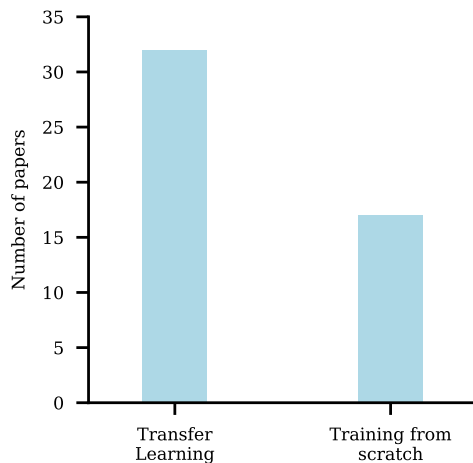


FIGURE 1. The total number of papers reviewed in this manuscript for transfer learning and training from scratch.

fairly high false-negative rates, which may cause treatment inefficiency or failure [7], [8]. Other methods, including Computed Tomography (CT) scans have been proposed to assist, expedite, and increase the accuracy of the testing process. However, methods based on RT-PCR and CT scans can be expensive in many countries because of insufficient facilities. Other new methods based on chest X-rays, which are less expensive and more widely available, have shown potential improvements in detecting COVID-19.

Inspecting a chest X-ray to diagnose COVID-19 add to the burden of radiologists as they review and interpret lung X-ray images. Although this approach requires time from a radiologist to interpret the images, X-ray evidence may be more accurate as opposed to reverse transcription polymerase chain reaction (RT-PCR) [9]. To aid the diagnosis of COVID-19, researchers have proposed several automated methods based on machine learning algorithms to analyze X-rays of COVID-19 cases. Automating the COVID-19 diagnostic process using X-ray images acts as a decision-supporting tool assisting radiologists as well as promoting the early detection and treatment of COVID-19.

There have been previous papers that have reviewed machine learning techniques for COVID-19 detection [10], [11], [12], [13], [14], [15]. However, these previous reviews covered older research papers. In this comprehensive review article, we review recent peer-reviewed articles published from January 2020 through the end of 2021. We classify the reviewed papers into two categories: transfer learning and training from scratch. A graph of the reviewed papers in each category is shown in Figure 1. An appendix is provided to explain the different deep learning architectures used in the reviewed papers.

II. DEEP LEARNING

Recently, deep learning-based approaches became one of the most popular algorithms in machine learning. These approaches have outperformed and achieved state-of-the-

art performance in many learning-based research problems [16], [17], [18]. The popularity of deep learning started in late 2012, when a deep-learning approach based on convolutional neural networks (CNNs) outperformed all other methods in the best-known computer-vision competition, ImageNet [19]. Such networks (i.e., CNNs) are designed to take advantage of a two-dimensional input, employing a series of convolutional layers for extracting features at different spatial locations. They have achieved cutting-edge results for many vision tasks, including object recognition [19], [20], scene classification [21], [22], [23], and 3D image understanding [24], [25].

The unique ability of deep learning methods to automatically learn a hierarchical feature representation of input data, make them an excellent choice when compared with traditional machine learning methods that depend on hand-engineering features [26]. CNNs are a powerful tool for extracting features from the input images and differentiating the importance among the features. It can also process not only 2-dimensional (2D) images but also 3-dimensional (3D) images. Recently, CNNs have been used in many fields, including medical imaging analysis.

Since medical images comprise both 2-D (e.g., X-rays) and 3-D (e.g., MRI) images, CNNs have a significant advantage in medical purposes over other various machine learning models [27]. While other machine learning algorithms based on vectorizations cannot use spatial information of an image effectively, CNNs can preserve spatial and structural information of an image; this is one of the key factors when analyzing medical images. For instance, CNNs can precisely recognize high-level and low-level features compared to other machine learning techniques. Therefore, CNN methods have been actively investigated for analyzing medical images [28].

The general architecture of CNNs model comprises convolutional layers, activation functions (e.g. rectified linear unit (RELU)), pooling layers, and fully connected layers as shown in Figure 2. When an input image is fed into a convolutional layer, the layer detects raw pixels (i.e., low-level features: lines and edges) in the image. Using these low-level features, a ReLU layer produces a feature map that consists of higher features, such as a cell or cytoplasm. A convolutional layer facilitates three major mechanisms (sparse connection, weight sharing, and sub-sampling) which reduce the degrees of freedom in a model. A sparse connection means that it only connects some inputs to the next layer. Weight sharing allows the network to reduce the number of weights updated within a convolutional layer and reduces the training time. Sub-sampling is performed through a pooling layer to reduce the number of parameters and the size of the image. This process lessens over-fitting and increases efficiency. The pooling layer computes a single value for each part in a grid within a filter. The values are replaced with the maximum number (max pooling) or the average number (average pooling) in the grid. Finally, the output from the last pooling layer or convolutional layer is passed to the fully connected layer. The input to the fully connected layer is flattened. In other

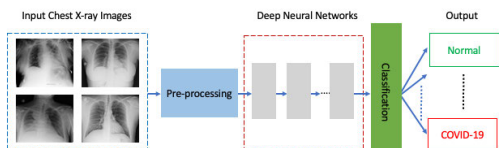


FIGURE 2. An overview of COVID-19 classification with a generic deep CNN.

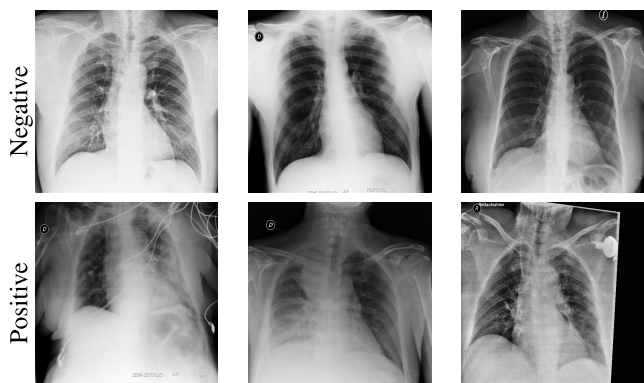


FIGURE 3. Chest X-ray samples from datasets in the reviewed papers. The top row shows COVID-19 negative samples and the bottom row shows positive ones.

words, a 3D matrix becomes ‘flatten’ as vector values. Then, using an activation function such as Softmax, the last layer calculates the probability of an object in the input image being in a particular class.

Deep learning approaches can be categorized into transfer learning and training from scratch. In the following subsections, we highlight the advantages and disadvantages of those approaches.

A. TRANSFER LEARNING

Transfer learning is the process of transferring knowledge learned from a source task to a destination task of the same or different domain [29]. This knowledge transferring process allows for improved deep learning performance, especially in the case of limited labeled data to train a model for a destination task. Transfer learning reduces the training time and power consumption by using knowledge already learned from the source task [30]. Appendix A provides more details about transfer learning.

B. TRAINING FROM SCRATCH

In contrast to transfer learning, training deep learning models from scratch requires starting from random sampled initial weights. However, this approach requires a large labeled dataset for learning high parameterized deep neural networks. Additionally, training from scratch requires time and large computation resources.

III. CHEST X-RAY DATASETS

Chest X-ray images are used by physicians to quickly and reliably diagnose a variety of diseases. Moreover, chest X-ray

TABLE 1. Chest X-ray images datasets summary including reference, hyperlinks for the datasets, categories of chest X-ray images, and the total number of images. To access the datasets click the hyperlink “Available online”.

Dataset Reference	chest X-ray image category				Total number of images
	Pneumonia	Normal	COVID19	Other	
[[31] Available online	✓				79
[[32] Available online		✓		✓	247
[[33] Available online			✓		468
[[34] Available online		✓		✓	1107
[[35] Available online		✓	✓	✓	5381
[[36] Available online	✓	✓			5856
[[37] Available online	✓	✓			5856
[[38] Available online	✓		✓		5933
[[39] Available online	✓	✓	✓		13975
[[40] Available online	✓	✓	✓	✓	21173
[[41] Available online	✓	✓			29700
[[42] Available online	✓	✓		✓	112120
[[43] Available online		✓	✓		13609
[[44] Not Available online		✓	✓		610

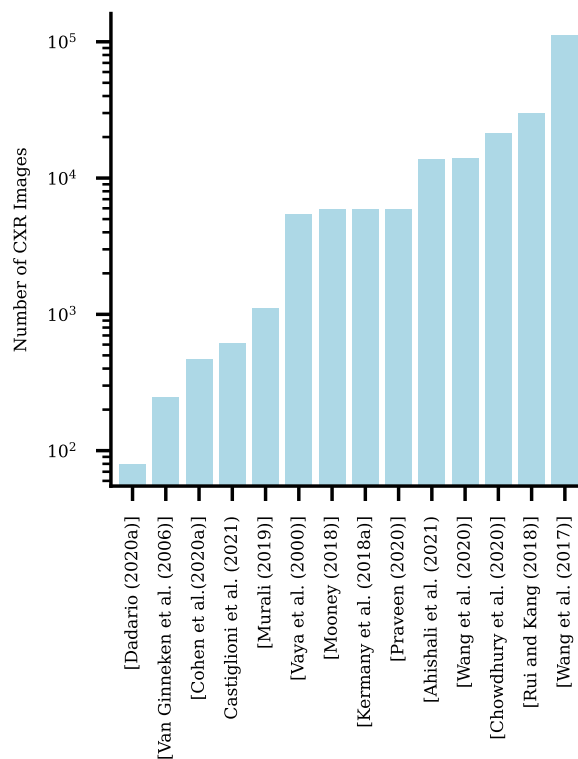


FIGURE 4. Total number of chest X-ray images for each dataset.

images are important for training learning algorithms to learn features for detecting pulmonary diseases. The datasets of COVID-19 chest X-ray images and non-COVID-19 chest X-ray images which have been used in the reviewed literature

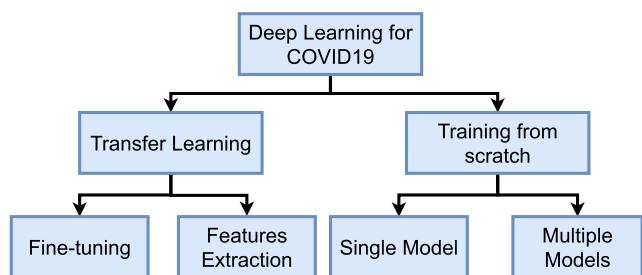


FIGURE 5. The main tasks of deep learning applied to chest X-ray images for COVID-19 detection.

for this manuscript are described in Table 1 and Figure 4. These imaging datasets serve the research community as a fertile environment for investigation and research [37], [42], [45]. However, to determine the optimum dataset, we performed a thorough statistical and imaging analysis of several open-source datasets (Table 1). The criteria we considered in assessing the datasets were based on simple statistical parameters, such as patient demographics, kurtosis, and skewness of the dataset in terms of outcomes bias. Additionally, common medical physics-based criteria may apply to all datasets. The quality of the images in terms of scan protocol such as kVp and mAs, the field of view, angle of images along with different machine manufacturers may have a prominent effect on the quality of the image and the reproducibility of the scans. Figure 3, shows some examples of COVID-19 positive and negative X-ray images.

CoronaHack Chest X-ray dataset [38] contains 58 (1%) patients with confirmed COVID-19 patients in the training set, so the data is skewed toward non-positive outcomes. The rest of the infected patients had pneumonia in either viral or bacterial causes. This may lead to bias in classification outcomes. A much larger COVID-19 outcome data is needed for this dataset. The data does not show any patient demographics, especially whether all COVID-19 patients in the dataset had pneumonia symptoms.

In another dataset presented in [36], some chest X-ray images (Pneumonia) dataset show inconsistency in terms of age, the field of view, and image parameters. Some images have artifacts because of the placement of cables and devices during the scan. Such artifacts may affect the validity of the classifier models during training and testing.

The NIH dataset [42] is a well-established and documented pneumonia imaging data that may be used as a preliminary check for researchers to test their approach. Some concerns raised in the previous two datasets are the type of X-ray used; single or dual-energy and image parameter.

IV. DETECTION OF COVID-19 USING DEEP LEARNING

Different approaches for COVID-19 detection using chest radiology images and deep learning have been proposed. In this section, we review the proposed methods for COVID-19 detection using chest X-ray images. Moreover, we provide a classification of deep learning approaches

applied to chest X-ray images for COVID-19 detection, as shown in Figure 5. From the reviewed papers, we found that there are two main approaches applied for COVID-19 detection using chest X-ray images: transfer learning and training from scratch. For transfer learning, some approaches used fine-tuning of the trained deep learning models using chest X-ray datasets for COVID-19 detection, whereas the other approaches extracted features from the trained models for applying classification algorithms, such as Support Vector Machine (SVM). With training from scratch, some approaches have trained a single model using convolutional neural networks (CNNs), while other methods trained multiple deep learning models for COVID-19 detection. In Tables 2, 3, 4, 5 and 6, a summary of each reviewed article is provided including performance, method, datasets, and availability of source code.

A. TRANSFER LEARNING-BASED APPROACHES

Transfer learning approaches can be divided into two categories: 1) fine-tuning-based methods and 2) features extraction followed by classification methods. In this section, we review approaches that used transfer learning for detecting COVID-19 infection using chest X-ray images. We first review fine-tuning-based approaches, followed by features extraction-based classification. Appendix A provides a visual illustration of transfer learning, including fine-tuning and features extraction approaches.

1) FINE-TUNING-BASED APPROACHES

Fine-tuning is the process of using knowledge transferred from a different domain as initial weights for training a deep learning model for a given task. This process takes advantage of models trained on large labeled datasets to enable learning models effectively for smaller labeled datasets [46]. In this section, we review deep learning approaches that used pre-trained deep learning models (mostly pre-trained on ImageNet) for fine-tuning using chest X-ray images to learn models for COVID-19 detection.

Some approaches have used off-shelf neural networks for fine-tuning the trained models on chest X-ray images. In [47], a transfer learning approach of a pre-trained deep learning model on ImageNet was proposed for early detection of COVID-19 using chest X-ray images. The pre-trained models were: VGG16, VGG19 [48], InceptionResNetV2 [49], Xception [50], InceptionV3 [51], MobileNet [52], and DenseNet121 [16]. The authors used a public dataset of chest X-ray images for fine-tuning the pre-trained models. The datasets were: 1) 164 X-ray images of COVID-19 were obtained from [33]. 2) A random selection of 210 images from each class was done using a dataset available on the Kaggle website which contains 5856 X-ray images (JPEG) of two classes normal and pneumonia (viral pneumonia/bacterial pneumonia)¹ [36]. The proposed approach

¹<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>

classified each image into one of three classes: normal, COVID-19, or pneumonia. The authors applied an ensemble of the predictions from the seven fine-tuned deep learning models. Another fine-tuning approach was presented by El-Gannour *et al.* in [53]. The authors evaluated transfer learning of models pre-trained on the ImageNet dataset for COVID-19 classification using a chest X-ray images dataset. The models evaluated were: VGG16, VGG19, InceptionV3, Xception, ResNet50v2, and mobileNetv2. The COVID-19 dataset used in this paper was from [40] and [54]. The best performing model was Xception with an accuracy of 98% and precision of 100%.

Some studies focused on the evaluation of existing deep learning models by fine-tuning those models using X-ray images. An evaluation of different fine-tuning deep learning models was done in [55]. The pre-trained deep learning models include: AlexNet [19], SqueezeNet [56], GoogleNet [57], ResNet-50 [18], DarkNet-53 [58], DarkNet-19 [58], ShuttleNet [59], NasNet-Mobile [60], Xception [50], Plae365-GoogLeNet [57], MobileNet-v2 [52], DenseNet-201 [16], ResNet-18 [18], Inception-ResNet-v28 [49], Inception-v3 [51], ResNet-101 [18], and VGG19 [48]. The training set for fine-tuning the pre-trained models was taken from two publicly available datasets: 1) the CoronaHack-Chest dataset, which randomly selected 50 images from COVID-19 images and 50 random images from bacterial, viral pneumonia, and normal classes [33]. This dataset was split for training and validation using an 80:20 split, where 80% of the dataset was used for training and 20% was used for testing. 2) Vancouver General Hospital (VGH), British Columbia, Canada dataset which contains 85 COVID-19 X-ray images. The images for the other classes were taken from CoronaHack-Chest dataset. The results showed that DarkNet-19 had the best performance of accuracy, about 94.28%, whereas ResNet-50 accuracy was 93.69% on the test set.

A comparative study of different deep learning models for COVID-19 detection was done by Shazia *et al.* [61]. This study compared the performance of VGG16, VGG19, DenseNet121, Inception-ResNet-v2, Inception-v3, ResNet50, and Xception neural networks. The best result was found using DenseNet121 where the accuracy was 99.48%.

Because of the lack of enough training examples for fine-tuning state-of-the-art deep learning models, some approaches combined multiple datasets together to get larger training data. An experimental study of different state-of-the-art deep learning models on a combined dataset from multiple publicly available imaging datasets was done in [62]. The authors used the dataset provided in [33] and [37]. The total number of COVID-19 images in the combined dataset was 224, the total number of bacterial pneumonia was 700 images, whereas the total number of normal condition chest X-ray images was 504. Another combined dataset was created where viral pneumonia was added as a class. The total number of images for viral and bacterial pneumonia was 714. The evaluation was done for two classes classification (COVID-19 vs. Non-COVID-19) and three classes

classification (COVID-19 vs. pneumonia vs. normal). Pre-trained deep learning models on ImageNet were used for fine-tuning where lower layers of the deep learning model were kept frozen, and the remaining layers were set as trainable. The deep learning networks used in this experimental study are VGG19 [48], MobileNet v2 [63], Inception [49], Xception [50], and Inception ResNet v2 [49]. The best results were obtained by the VGG19 network where the accuracy of two-class classification and three-class classification was 98.75% and 93.48% respectively. The model evaluation was done using tenfold cross-validation. A drawback of this study is the lack of enough COVID-19 images and the data imbalance.

Fine-tuning of Inceptionv3 model pre-trained on ImageNet dataset was done in [64]. The head of Inceptionv3 was replaced with four fully connected layers. The last fully connected layer outputs classification for three classes: COVID-19, pneumonia, and normal. This fine-tuning was done using 193 images randomly selected for pneumonia and normal classes using the X-ray images dataset [36], and 163 COVID-19 images from the COVID-19 dataset [33]. This approach showed an accuracy of 98% on the test set. Another transfer learning and fine-tuning of the VGG16 model was done by [65] using chest X-ray images of COVID-19 from [33]. The normal X-ray images were chosen from [36]. The total COVID-19 positive images used was 141, whereas the total number of normal X-ray images was 1341. VGG16 was fine-tuned after freezing all the layers except for the last three layers of the VGG16 architecture. Data were split into training and validation, the performance on the validation was high at 99.45% (accuracy) using 5-fold cross-validation; however, the data was imbalanced.

Data imbalance is a critical challenge for training and fine-tuning deep learning models. An approach that balances the data before fine-tuning deep learning was proposed in [66]. This approach fine-tunes Xception neural network trained on ImageNet. The proposed method is called CoroNet, where the head of the neural network was replaced with two fully connected layers. This model was used for binary classification, three classes based classification, and four classes based classification. The classes for the binary classification were COVID-19 and normal, whereas the three classes classification were: COVID-19, normal, and pneumonia. The four classes classification were: COVID-19, normal, pneumonia-bacterial, and pneumonia-viral. This approach was trained and tested using 4 folds cross-validation using two publicly available datasets. The first dataset contains 290 COVID-19 chest radiology images [33]. The second dataset contains pneumonia bacterial, pneumonia viral, and normal chest X-ray images [36]. The total images from these two datasets were 1300 X-ray images. To avoid the imbalanced data issue, random under-sampling was done from the majority class until the dataset became balanced. Therefore, the total number of COVID-19 images was 290, the total number of normal X-ray images was 310, the total number of pneumonia-bacterial was 330, and the

total number of pneumonia-viral images was 327. The results were 89.6% (accuracy for four classes classification), 95% (accuracy for three classes classification), and 99% (accuracy for binary classification).

Another approach that balances the dataset before fine-tuning deep learning models was proposed in [67]. This approach uses an integrated stacking technique and fine-tuning of deep learning models to detect COVID-19 using chest X-ray images [67]. This approach used pre-trained deep learning models on ImageNet for fine-tuning using chest X-ray images. The pre-trained deep learning models were: ResNet101 [18], Xception [50], Inception-v3 [51], MobileNet [63], and NASNet [60]. The last convolution layer from each model was fine-tuned, whereas the other layers were kept frozen. This approach used two publicly available datasets for fine-tuning, where the learned representations from fine-tuning were integrated using two fully connected layers for classification. The first dataset was the COVID-19 radiography dataset [40] which is comprised of 219, 1345, and 1341 COVID-19, pneumonia, and normal chest X-ray images, respectively. The second dataset was the Chest X-ray dataset [45] which contains 142 COVID-19 chest X-ray images. Because of the imbalance of the dataset, random sampling of each class was done to balance the dataset. The total images per class after the random sampling was 361, 365, and 362 for COVID-19, normal, and pneumonia, respectively. The accuracy for three classes classifications (i.e., COVID-19 vs. pneumonia vs. normal) was 99.08% whereas the accuracy for two classes classifications (i.e., COVID-19 vs. non-COVID-19) was 99.52%

A transfer learning and fine-tuning approach was done in [68]. The authors used the Inception v3 network [51] pre-trained on ImageNet [69]. Since the dataset for COVID-19 is small, and to avoid over-fitting, the Inception v3 neural network was truncated to reduce the number of parameters. The dataset used for this work was a combination of three datasets: COVID-19 collection dataset [33] which consists of 162 COVID-19 chest X-ray images, pneumonia collection dataset [70], which contains 5863 chest X-ray images of viral and bacterial pneumonia, and Tuberculosis collection [71] that has 820 chest X-ray images. The authors tried six experiments where the data combination from different datasets was considered. The best result was obtained using training and testing on a subset of the data, which includes 162 COVID-19 cases and 1583 healthy chest X-ray images. The accuracy was 100% and the AUC was 1.0. However, the data imbalance is a serious problem with this approach.

An approach to detect COVID-19 using chest X-ray images was proposed in [72]. This approach was based on fine-tuning of SqueezeNet neural network pre-trained on ImageNet [56]. The authors used offline augmentation to balance the dataset. The network was fine-tuned with the hyperparameter Bayesian optimization technique [73]. The dataset used in this approach was based on [33] and [37], where the best accuracy was 98.26%. The issue with this approach

is that the authors reported results on augmented test data, which does not represent real-life test examples.

A three steps approach was proposed in [74] to detect COVID-19 in chest X-ray images. The approach aims to first classify a chest X-ray image as healthy (for a healthy patient) versus infected (for an infected patient with pulmonary diseases, which include COVID-19). If the chest X-ray image was classified as for an infected patient with pulmonary disease, then the second step was applied, which aims to detect if the pulmonary infection was pneumonia or COVID-19. If the image was classified as COVID-19, then a Grad-CAM visualization approach was applied to show which part of the image the neural network was used for the decision. The dataset for this approach was a combination of multiple publicly available datasets: the COVID-19 dataset from [33], [75], and the National Research of Health Chest X-ray dataset [42]. Fine-tuning of the VGG16 model pre-trained on ImageNet [48] was done, where the head of the neural network was replaced for the classification of chest X-ray images into two classes. The accuracy of the model in the first step was 96% and the accuracy of the model in the second step was 98%.

To learn recognition of certain diseases based on medical images, a combination of multiple fine-tuned deep learning models where each model learns from different classes of medical images can help improve disease detection. In [76], a transfer learning approach was proposed, where three ResNet models were trained to classify chest X-ray images. The first model was trained to classify normal vs. diseased. The second model was trained to classify pneumonia and non-pneumonia. The third model was trained to classify COVID-19 vs. non-COVID-19. After training the models, concatenation of the three models was done where all the layers were frozen except the classification layers were set as trainable. New layers for concatenation and classification of the three models were added. Then fine-tuning of the models to classify chest X-ray images of classes: COVID-19, normal, and pneumonia were done. The datasets for this approach were from the RSNA Pneumonia Detection challenge dataset [41] and the COVID-19 X-ray image dataset [33]. The total number of images of normal, pneumonia, and COVID-19 X-ray images were 1579, 4245, and 184, respectively. This approach had the best accuracy of 95.5%.

Li *et al.* proposed the COVID-MobileXpert approach using a knowledge transfer and distillation framework [77]. A DenseNet-121 architecture was pre-trained and fine-tuned using an attending physician network and resident fellow network, respectively. The pre-training was done on the chest X-ray image dataset [42]. MobileNetv2 and SqueezeNet were used as medical student networks for on-device COVID-19 case screening. Resident fellow network was used to train lightweight medical student network using knowledge distillation. The datasets used for training, validation, and testing this approach were from [41] and [33].

A neural network called COVID-Net was proposed in [39]. This deep learning network is inspired by ResNet architecture

and the design of the network was done by generative synthesis [78]. The authors also created a benchmark dataset called COVIDx, comprising 13,975 chest X-ray images collected from multiple publicly available datasets. COVID-Net was trained on ImageNet [69], then fine-tuned on the COVIDx dataset. Compared to ResNet and VGG19, the COVID-Net has a low number of parameters. The accuracy of COVID-Net was 93.3% on the COVIDx test set. The dataset and source code are publicly available.²

Fine-tuning of deep learning models trained on similar types of images such as X-rays could yield better detection performance compared to models trained on a different dataset such as ImageNet. A transfer learning approach was done in [79], where a designed deep learning network called CovXNet was trained on X-ray images of viral pneumonia, bacterial pneumonia, and normal images. Then the trained model was fine-tuned on a smaller dataset of COVID-19. The CovXNet deep learning network comprises multiple residual blocks, where each residual block has stacked depth-wise dilated convolution layers. Stacking of multiple neural networks trained on different resolutions of input images was done, where the prediction was used to learn a meta-learner algorithm such as random forests and XGBoost. The authors trained the deep learning network on two publicly available datasets: 1) pneumonia dataset chest X-ray images collected in Guangzhou Medical Center, China [37], and 2) COVID-19 dataset collected from Sylhet Medical College, Bangladesh. The results showed an accuracy of 96.9% for COVID-19 vs. viral pneumonia, 94.7% for COVID-19 vs. bacterial pneumonia, and 90.2% for the multi-class task of COVID-19, normal, viral, and bacterial pneumonia.

Learning data augmentation using Bayesian optimization and autoencoder to create augmented data was proposed in [80]. This approach was used to fine-tune pretrained neural networks such as AlexNet, ShuffleNet, ResNet18, and GoogleNet. This approach showed superior results compared to state-of-the-art augmentation startiges. Another approach that uses transfer learning by fine-tuning DenseNet-121 for COVID-19 diagnosis and discrimination of COVID-19 from other type of pneumonia was proposed in [81]. The trained deep learning model was compared to radiologist performance and showed same performance as senior radiologists. Another transfer learning approach was done using neutrosophic domain [82]. The authors converted RGB images to neutrosophic domain which gives three type of images: true image, indeterminacy image, and falsity image. Then fine-tuning of three deep learning models including AlexNet, GoogleNet, and ResNet18 was done. The best result was obtained using chest X-ray images in indeterminacy domain.

A convolutional support estimator network (CSEN) was proposed for early detection of COVID-19 infections using X-ray images [43], [83]. This proposed classifier was compared to state-of-the-art neural networks such as DenseNet-

121, where CSEN showed lower accuracy for COVID-19 detection compared to DenseNet-121. A study of supplemental training of deep learning fine-tuning for COVID-19 classification using X-ray images was performed in [84]. The authors evaluated performance of transfer learning of models (DenseNet-121) trained on a similar task to classify COVID-19. Then fine-tuning of pre-trained ImageNet model (DenseNet-121) was done sequentially on pneumonia and COVID-19 datasets.

A comparison between an ensemble of ten convolutional neural networks (CNN) and radiologists was performed in [44]. The authors used TRACE4 system³ to fine-tune ResNet-50 pre-trained neural network to classify X-ray images into COVID-19 and non-COVID-19. The X-ray images were collected by the authors from two hospitals in Italy. The results showed that the ensemble of CNNs had a superior results over radiologists. Another approach for assessing the severity of COVID-19 progression was proposed in [85]. This approach uses transfer learning of VGG16 pre-trained on pneumonia dataset [41] for fine-tuning. This approach classifies X-ray images into normal, mild, moderate, and severe.

2) FEATURES EXTRACTION

In this section, we review papers that extract deep features or traditional features from chest X-ray images to apply classification algorithms for COVID-19 detection. The extracted deep features were from ImageNet-based pre-trained deep learning models and traditional features, such as co-occurrence matrix features. The following paragraphs explain the proposed methods for COVID-19 detection using extracted features from chest X-ray images.

A cascaded deep learning approach for COVID-19 detection in chest X-ray images was proposed in [86]. This approach used VGG16 [48] and Capsule-Net [87]. The deep features were extracted from the last convolution layer of the pre-trained VGG16 model on ImageNet. The extracted features were fed to Capsule-Net, followed by classification layers for two classes and multi-class classification. The proposed method was evaluated on the chest X-ray images dataset [31]. The two classes classification were: COVID-19 positive vs. COVID-19 negative, whereas the classes for multi-class classification were: viral pneumonia, normal, and COVID-19 positive. The average precision, recall, and F1-score for two classes classification were 0.97 using the VGG-CapsNet method.

A comparison study of transfer learning of deep learning models pre-trained on ImageNet was done in [88]. The authors extracted features from each model and passed the features to a perceptron neural network to classify each chest X-ray image into one of three classes: normal, pneumonia, and COVID-19. The pre-training models used in this study were: VGG16, VGG19, DenseNet201, Inception_ResNet_v2, Inception_v3, ResNet50, MobileNet_v2.

²<https://github.com/lindawangf/COVID-Net>

³http://www.deeptracetechnology.com/files/TechnicalSheet__TRACE4.pdf

The dataset for this study was based on two publicly available datasets: chest X-ray & CT dataset [33], [37], and [33]. The best accuracy was obtained by Inception_Resnet_v2 extracted features, where the accuracy was 92.18% and the F1 score was 92.07%.

In [89], a hybrid approach for COVID-19 detection using chest X-ray images was proposed. This approach is based on AlexNet [19] pre-trained on ImageNet for COVID-19 detection using three methods. 1) feature extraction followed by feature selection and SVM classification algorithm. 2) Fine-tuning of AlexNet, where the classification layer was replaced with a fully connected layer with three nodes to classify chest X-ray images into three classes: normal, pneumonia, and COVID-19. 3) AlexNet improved version was proposed where two fully connected layers were added while replacing the classification layer from the original AlexNet. The dataset for this approach was from publicly available datasets: 1) NIH chest X-ray dataset,⁴ 2) COVID-19 chest X-ray images dataset [42] and [33], where random images were selected from each class: normal, pneumonia, and COVID-19, for training the model on a balanced dataset. The best result of this approach was obtained using improved AlexNet + ReliefF + SVM (where total selected features were 40), and the accuracy was 98.64% and the F1-score was 98.63%.

An approach for detecting COVID-19 in a chest X-ray images dataset was proposed in [90]. This method created a dataset from two publicly available datasets [33] and [42]. The dataset was called RYDLS-20 which was made available by the authors.⁵ This approach extracts hand-crafted texture features and deep features from chest X-ray images for training several classifiers. The methods for hand-crafted features extraction were: Local binary pattern (LBP) [91], Elongated quinary patterns (EQP) [92], binarized statistical image features (BSIF) [93], local phase quantization (LPQ) [94], local directional number (LDN) [95], oriented basic image features (oBIFs) [96], Locally encoded transform feature histogram (LETRIST) [97]. The deep features were extracted from the pre-trained Inception-v3 model [57]. After the fusion of the features, the authors trained the classifiers. Moreover, the authors tried late fusion of the prediction of each classifier. The classification was done for the flat multi-class classification task and for the hierarchical classification task, where the hierarchical classification task was done using Clus-HMC.⁶ The best result was obtained using an early fusion of BSIF and EQP and LPQ features with SMOTE over-sampling for data balancing [98] where the F1-score was 0.889. Another approach was proposed in [99], where chest X-ray images were used to extract texture features using a Co-occurrence matrix and local binary patterns, where the total number of extracted features was 129 features. These features were used

to train a feed-forward neural network comprising 4 hidden layers. Moreover, the authors experimented with flattening the chest X-ray images to a vector and feeding the flattened vector to a feed-forward neural network of four hidden layers. Additionally, the authors experimented with CNNs of four convolution layers followed by two fully connected layers for classification. This method was trained and validated using publicly available datasets from [100] and [33]. The results for the trained feed-forward neural network to classify chest X-ray images as normal, pneumonia, or COVID-19 was 98.82% (accuracy), whereas CNN accuracy was 95.48%, and texture feature extracted based feed-forward neural network accuracy was 98.56%. Although this method showed high performance, the data is imbalanced, and the feed-forward neural network had numerous parameters which can lead to over-fitting. Moreover, the extracted features were from the entire chest X-ray images which cover a larger area beyond the lung. This feature extraction could cause learning from non-relevant information for the target task.

In [101], the authors proposed a deep learning approach called ConStacknet which is based on StackNet meta-modeling in combination with CNNs to learn discriminative features from X-ray images. Two datasets were used to evaluate this approach: 1) A publicly available dataset comprising 5216 X-ray images which include 4273 pneumonia cases, and 1583 normal cases [36]. 2) Kaggle dataset of COVID-19 X-ray images [102], and 3) a publicly available dataset that contains X-ray images of patients having pneumonia infections including COVID-19 disease [33]. The authors used VGG16 architecture pre-trained on ImageNet to extract deep features from chest X-ray images. Then, some processing of the features was done, such as standardization (removing the mean and scaling to unit variance). After that, Stacknet was applied, which consisted of multiple classifiers for ensemble [103] and [104]. The accuracy of the Stacknet model was 97% on the test set.

An approach for features fusion was proposed in [105]. This approach incorporate traditional features (histogram oriented gradient) and deep features extracted from X-ray images for learning to detect COVID-19. The authors showed that fusion of features has a superior results over classifiers that uses deep features and traditional features independently. Another approach for evaluation and assessment of COVID-19 based models was proposed in [13]. The authors used Inception-v3 features extraction and selection followed by classification using SVM and assessment using entropy.

B. TRAINING FROM SCRATCH

In this section, we review the papers which presented deep learning approaches trained from scratch for COVID-19 detection using chest X-ray images datasets. These approaches can be categorized into single model-based approaches and multiple model-based approaches as detailed in the following subsections.

⁴<https://www.kaggle.com/nih-chest-xrays/data>

⁵<https://tinyurl.com/metka4ck>

⁶Available for download at <https://dtai.cs.kuleuven.be/clus/>

1) SINGLE MODEL-BASED APPROACHES

Convolutional Neural Networks (CNNs) are a type of neural network that enable learning spatial features and weight sharing using kernels for grid-like data topology. The details of CNNs have been discussed in Section II. In this subsection, we review the papers that presented a single model-based CNN, which was trained for detecting COVID-19 using chest X-ray images.

A simple architecture that comprises two convolution layers was designed, trained, and presented in [106]. This neural network was trained using data from three publicly available datasets: 1) the Joseph Paul dataset, which contains 542 frontal chest X-ray images from 262 patients [33] and [33]. 2) the COVID-19 Radiography dataset created by researchers from Qatar University [40], [54]. 3) Dataset created by Kermany *et al.* [100]. The combined dataset was split into training, validation, and testing sets, where the accuracy of the test set was 99.2%.

The quality of X-ray images is essential for high-performance deep learning models. Therefore, an approach that studied the pre-processing of chest X-ray images for performance improvement was done in [107]. This approach has a pre-processing step, which includes an adaptive median filter and histogram equalization. The CNNs trained to classify the chest X-ray images had four layers of convolution, followed by two fully connected layers. Training CNN was done using two datasets: 1) COVID-19 Radiography dataset which had 219 positive COVID-19 chest X-ray images [40], and 2) chest X-ray dataset created by Murali Kummitha which contained 107 chest X-ray images of effusion disease and 1000 chest X-ray images of normal cases [34]. The results showed higher performance when using the pre-processed images compared to non-processed images for both tasks: two classes classification (i.e., COVID-19 vs normal) and three classes classification (i.e., COVID-19 vs normal vs effusion). The accuracy of the deep learning model when applying histogram equalization pre-processing step for two classes and three classes classification was 98.62% and 95.77% respectively.

Since there are limited COVID-19 datasets of chest X-ray images, and the chest X-ray images cover a wide range of areas beyond the lung such as the shoulders and neck, segmentation of the lung lobes from chest X-ray images is important for the convolutional neural network models to learn effective and discriminative features and to extract patches of chest X-ray for training a neural network. To address the shortage of training data, an approach was proposed in [108] to extract patches from chest X-ray images. This method used FC-DenseNet103 for semantic segmentation of the lung, then K patches were extracted from the images to train a classification architecture (ResNet18). The final classification was an ensemble of K patches predictions of a chest X-ray image, where $K = 100$. The method was trained on chest X-ray images from [32], [70], [109] for lung segmentation. For the classification task, the training and testing datasets were combined from multiple publicly available

datasets [32], [33], [70], [109]. The labels for the classification task were normal, bacterial pneumonia, tuberculosis, and viral pneumonia (which include COVID-19 cases). The results using the ensemble of patches (local appearance) are better than the models trained with the entire images (global appearance), where the accuracy of the ensemble model was 88.9%. Another approach that used segmented lung lobes for learning attention-based approach called MANet (mask attention network) was proposed in [110]. This approach avoids the problem of heavy computations using a soft attention approach by using the mask. This approach used two stages: segmentation of the lung region stage, and the classification stage. The segmentation stage used U-Net with ResNet backbone for obtaining lung region masks. Then, the classification stage used chest X-ray images and the segmentation masks to classify each chest X-ray image into five classes: normal, tuberculosis (TB), bacterial pneumonia (BP), viral pneumonia (VP), and COVID-19. Multiple deep learning architectures were used for the classification task, including ResNet34, ResNet50, VGG16, and Inception-v3. The dataset for this approach came from multiple publicly available datasets: The first dataset is Montgomery County and Shenzhen No.3 People's Hospital [70], the second dataset is an image dataset [37], and the third dataset is a publicly available dataset of COVID-19 chest X-ray images available on GitHub [33]. The total number of chest X-ray images for normal, COVID-19, tuberculosis, bacterial pneumonia, and viral pneumonia were 1840, 433, 394, 2780, and 1345 respectively. ResNet50-based MANet had the best accuracy at 97.06%.

Some state-of-the-art architectures contain numerous parameters, therefore, training such architectures requires large labeled datasets. To address that, modification of the deep learning architectures is required to get a smaller architecture. For instance, a method to detect COVID-19 in X-ray images was proposed in [75]. This method modified the original DarkNet architecture [111], where the changes mainly were fewer layers and changes in the filter sizes. This approach was trained and tested for two classes classification (COVID-19 positive, normal) and three classes classification (COVID-19, normal, Pneumonia). The dataset used for this experiment was based on 1) 127 COVID-19 X-ray images from [33], and 2) randomly selecting 1000 images for pneumonia and normal classes (500 for each class) from [42]. The training and testing were done using a 5-fold cross-validation approach. The results for two-class classification were 98.08% (accuracy) and 87.02% (accuracy) for three-class classification. Although the accuracy was high, the total number of COVID-19 images used in this study was very few and the dataset was imbalanced.

An approach for chest X-ray content-based image retrieval (CBIR) was proposed in [112]. This approach was based on ResNet50 pre-trained on ImageNet. The ResNet50 was fine-tuned using chest X-ray images of COVID-19, pneumonia, and normal. Then feature extraction from two different layers within the ResNet50 was done. Moreover, an attention

module was plugged in to learn the spatial attention mask. The extracted features and the attention mask were combined using element-wise multiplication [113]. Followed by convolution layers to output the generated embedding features. This approach used multi-similarity loss [114]. The dataset used for this approach is based on a COVID-19 publicly available dataset [39] and new chest X-ray images collected from hospitals in Massachusetts, the United States of America, and South Korea. The total number of images used in this approach was 18,055 chest X-ray images. Since chest X-ray images often have texts overlaid on the image, a pre-processing step was done. This pre-processing step included image cropping, resizing, windowing, and lung segmentation. The lung segmentation was done using an ensemble of five neural networks of the same backbone structure EfficientNet [115]. The results showed high accuracy for CBIR compared to the baseline (original ResNet50), where the accuracy was 83.9% and COVID-19 sensitivity was 85%.

An ensemble deep learning model approach was proposed in [116], where the deep learning networks were MobileNet and ShuffleNet. These two neural networks were trained in parallel and the learned features were merged together and used for a fully connected network for classification. The data to evaluate this method was from four datasets: BIMCV COVID-19 dataset [35], Shenzhen dataset [117], Montgomery dataset [117], and CoronHack dataset [33]. The accuracy and F1-score of this method were 95.83% and 95.94% respectively.

Training neural networks to detect COVID-19 using chest X-ray images, followed by extracting the features from the trained deep learning models for features selection and classification using SVM was done in [118]. This approach classified X-ray images into pneumonia, COVID-19, or normal class. This approach used a dataset collected by the author, where each class had 364 X-ray images. The author created an enhanced version of the dataset using the approach proposed in [119]. The training of deep learning models was done on both raw and enhanced datasets. Four neural networks were trained from scratch, where the neural networks were: AlexNet [19], VGG16 [48], GoogleNet [57], and ResNet [18]. After training the neural networks, feature extraction was done from each neural network where the number of features extracted from each network was 1000 features. Then feature selection was done using two meta-heuristic algorithms: Binary Gray Wolf Optimization [120] and Binary particle swarm optimization [121]. The selected features were used to train a Support Vector Machine (SVM) algorithm for classification into one of three classes: pneumonia, COVID-19, or normal. The best result was got using a combination of AlexNet and VGG16 features, where the accuracy was 99.08%. Another approach that used two neural networks, MobileNet and SqueezeNet was proposed in [122]. This approach comprised of pre-processing the chest X-ray images using a fuzzy color technique to remove noise from the images, followed by stacking to combine the enhanced images with original images for better qual-

ity [123], [124]. After the pre-processing, training of two neural networks, MobileNet and SqueezeNet was performed. Then using social mimic optimization (SMO) [125], features selection was applied. Then the SVM was used to classify chest X-ray images as normal, pneumonia, or COVID-19. The dataset used in this approach was based on [33] and [40]. The accuracy of this approach was 98.25%.

Lung segmentation and COVID-19 localization approach was proposed in [126]. This approach uses chest X-ray images with segmentation ground truth for training U-Net, U-Net++, and Pyramid Networks (FPN). The authors constructed the largest benchmark dataset with chest X-ray images of 33920 image. An approach to address the shortcomings of imbalanced dataset was proposed in [127], where MixMatch algorithm is applied [128]. Data imbalance was corrected using the loss function weights for under-represented class. A prognostication approach of patients with COVID-19 using Artificial Intelligence (AI) and chest X-ray images was proposed in [129]. This approach uses chest X-ray images and clinical data to train EfficientNet deep neural network. Then deep features extracted from the trained model and clinical data were used to build time-to-event models for the purpose of predicting the risk of disease progression.

A semi-supervised approach for classifying COVID-19 X-ray images was proposed in [130]. This approach is called Semi-supervised Open set Domain Adversarial network (SODA). Features extraction was done from pre-trained network, then feeding the latent features to different classification networks. An approach to validate deep learning results for COVID-19 classification using X-ray images was proposed in [131]. This approach is called Hide-and-Seek, which uses modified versions of the training, validation, and testing sets. The modifications include cropping the lungs out of X-ray images to study the effect on deep learning features learning and results. A deep anomaly detection approach called confidence aware anomaly detection (CAAD) was proposed in [132]. This approach consists of feature learning and extraction, anomaly detection and confidence prediction modules. This approach was used to detect viral pneumonia cases using in-house built X-ray images dataset.

2) MULTIPLE MODEL-BASED APPROACHES

Multiple model-based approaches presented in this subsection are for methods that integrate and train more than one deep learning model to detect COVID-19 using chest X-ray images, as detailed in the following paragraph.

In [133] a convolutional neural network and an autoencoder were trained, where the latent space of the autoencoder was the input to a convolutional neural network. This method shows 2% better accuracy compared to a VGG16 model on a chest X-ray images dataset comprising 400 COVID-19 positive cases and 400 normal cases [33]. Another approach that combines Xception and ResNet50v2 networks for learning features and detecting COVID-19 using chest X-ray images datasets was proposed in [134]. This approach trains

Xception and ResNet50-v2 in parallel where the network heads of both networks were removed. The learned features of size $10 \times 10 \times 2048$ from each of the two networks were concatenated. Then a convolution layer was applied with a filter size of 1×1 and no activation, followed by a dropout and classification layer. This approach classified chest X-ray images into normal, COVID-19, and pneumonia. The authors also proposed a training mechanism for imbalanced data, where the majority class is divided into many subsets of size equal to the minority class. Then, each subset is combined with the minority class to form a training set. A deep learning model was trained for each combined training set, which resulted in n trained models. The dataset used for training and validating this approach was based on [33] the RSNA pneumonia detection challenge dataset. This approach showed an accuracy of 91.40% whereas the accuracy of detecting COVID-19 was 99.56%

V. DISCUSSION

Evidence-based medicine (EBM) is defined as integrating current best evidence with clinical expertise [137]. EBM is a continuously improved process that requires a non-stop addition of new tools and research. The current vision of radiology medical practice is to shift from volume-based practice using standardized methods toward more affordable yet high-quality services that are patient-centered with a focus on the patient's outcome [138], [139]. Integrating quantitative radiological-imaging-based evidence based on literature and continuous modeling improvement to serve as a robust decision support tool along with histopathological findings is an overdue step [140]. Within the context of the presented work, EBM is planned by integrating X-ray image analysis and machine learning techniques as a decision support tool in medicine. Specifically, the potential value of artificial neural network (ANN) algorithms in incorporating evidence-based practice in the clinic is promising in much literature. [141], [142]. In this study, we reviewed studies that aim at recruiting artificial neural networks as an analytical metric to support the decision taken in the detection of COVID-19 infection, which is a considered relatively new approach in medicine. According to these studies, neural network algorithms showed a valid promise to aid in adopting the evidence-based medicine treatment tailored for each patient by enhancing the evidence collection approach [143], [144], and [145]. These studies warrant further investigation of the validity of such approaches by investigating artificial neural network algorithms on phase two and three randomized clinical bases. Deep learning classifiers might provide the physician with tools that go beyond the traditional anatomical descriptors of the image. Such a method will increase the mass of scans, expedite scanning time, and reduce the efforts of radiologists in analyzing imaging scans based on infection status.

During our literature review process, we came across articles that used different radiological images, computed tomography or X-ray images, as the source data to train

the classification algorithms. A comparison between CT and X-ray has to be drawn. The feasibility of recruiting CT image data for a decision-supportive tool for COVID-19 patients has many considerations. Because computerized imaging tomography is helical three-dimensional imaging that requires intensive work to analyze based on two aspects. First, is the physical aspect, where the three-dimensional image processing of CT requires training the models on larger images of raw data [146]. This will require heavy analysis, coding, and labor-intensive assessment of both the code and the reliability of the images. X-ray images, on the other hand, are two-dimensional images that are easier to analyze. The size of the CT raw data generated also requires heavy storage and computing capacity. From a broader perspective, it is a challenging task to generate intra- and inter-institutional reproducibility in terms of imaging parameters and protocol because of the difficulty of enforcing the same parameters and protocols throughout institutions and training all technicians to rapidly reproduce the same method repeatedly. For example, in the case of implementing the same discretization level and reconstruction algorithms [140]. These parameters, among others, may affect image quality. From an operational perspective, the scheduled maintenance routine, quality assurance program, and common malfunctions of the CT machine will add to a higher downtime compared to the X-ray, which will translate into hardship in conducting research and generating data. On a similar note, the cost of acquiring a CT image is expensive and may not be cost-effective. Second, as for the clinical aspect, the time needed to generate a CT image requires a longer time as opposed to a chest X-ray. Such a factor means that having a COVID-19 suspected patient waiting for his or her turn in addition to the scan time, generation of the images, analysis by a radiologist, and transferring the reports to the attending physician will add to a higher waiting time for the patient. In addition, COVID-19 patients may complain of severe lung symptoms, which will make them unable to hold their breath properly during CT scanning, especially elderly and pediatrics [147], [148]. Movement during the CT scan will cause motion artifacts that will affect the imaging quality by reducing the lung volume or deforming the voxel intensity value. Consequently, it will become more challenging for a machine-learning algorithm to extract features from the images due to noise and poor quality.

From the aforementioned discussion, conducting a quantitative imaging analysis, by the virtue of machine learning algorithms, is feasible using X-ray chest radiographs. On a similar note, the efficacy of building machine learning models roots in the validity and robustness of the mined data such as X-ray images. Furthermore, the total number of images in each class is important, because they will have a pronounced statistical effect on the data, such as skewness, kurtosis, and deviated means, which will perturb the validity of the model.

In some articles reviewed, there is a limited number of COVID-19 chest X-ray images versus pneumonia chest X-rays. Therefore, ensuring a balanced dataset is of utmost

TABLE 2. Summary of reviewed papers which include the datasets, performance, and the category of the COVID-19 detection methods.

Ref	Datasets and Source code	Performance	Category
[47]	<p>Datasets: Chest X-ray images:1) 146 images of COVID-19 from [33]. 2) 420 images of pneumonia from [36].</p> <p>Source code: Not available.</p>	<p>Accuracy (%): 1) VGG16: 96.88 2) VGG19: 95.31 3) Inception ResNetV2: 89.06 4) Xception: 95.31 5) InceptionV3: 92.66 6) MobileNet: 89.06 7) DenseNet121: 92 8) Ensemble: 98</p>	Transfer Learning - Fine-tuning
[101]	<p>Datasets: Chest X-ray:1) a publicly available dataset of pneumonia cases [36]. 2) Kaggle dataset [102]. 3) a publicly available dataset of pneumonia and COVID-19 cases. [33].</p> <p>Source code: Not available.</p>	Accuracy: 97%	Transfer Learning - Feature extraction
[106]	<p>Datasets: Chest X-ray:1) a publicly available dataset of COVID-19 cases [33]. 2) a dataset by Qatar University [40] [54] 3) Pneumonia cases dataset [100].</p> <p>Source code: Not available.</p>	Accuracy: 99.2%	Training from scratch - Single Model.
[77]	<p>Datasets: Chest X-ray:1) a publicly available dataset of X-ray images [42]. 2) a publicly available dataset of pneumonia cases [41]. 3) a publicly available dataset of COVID-19 cases [33]</p> <p>Source code: Available HERE</p>	Accuracy : 88%	Transfer Learning - Fine-tuning.
[108]	<p>Datasets: Chest X-ray images for segmentation of lung from [109], [32], [70] and for classification from [109], [32], [70], [33]</p> <p>Source code: Available HERE</p>	Accuracy = 88.9% F1-score = 84.4%	Training from scratch - Multiple models
[75]	<p>Datasets: Chest X-ray images from two publicly available datasets: 1) Total of 127 COVID-19 positive cases from [33]. 2) Total of 1000 images selected randomly from [42] for two classes: no findings and pneumonia.</p> <p>Source code: Available HERE</p>	<p>Results for two classes classification: Accuracy: 98.08% F1-score: 96.51%</p> <p>Results for three classes classification: Accuracy: 87.02% F1-score: 87.37%</p>	Training from scratch - Multiple models.
[66]	<p>Datasets: X-ray images from two publicly available datasets. 1) 290 COVID-19 X-ray images from [33]. 2) 967 X-ray images contains three classes: pneumonia-bacterial, pneumonia-viral, and normal</p> <p>Source code: Available HERE</p>	<p>Four classes classification: F1-score = 89.8% Accuracy = 89.6%</p> <p>Three classes classification: F1-score = 95.6% Accuracy = 95%</p> <p>Two classes classification: F1-score = 98.5% Accuracy = 99%</p>	Transfer learning - Fine-tuning

importance for the robustness and reliability of any learning model. In the articles by [62], [64], [68], and [99] the data was imbalanced, which cause bias in the trained model. Other researchers have applied several approaches to overcome the

data imbalance problem, such as under-sampling the majority class, until the number of images becomes balanced, which was applied in [66]. In [67] and [89], random sampling of images from each class in the combined dataset was done to

TABLE 3. Summary of reviewed papers which include the datasets, performance, and the category of the COVID-19 detection methods.

Ref	Datasets and Source code	Performance	Category
[55]	<p>Datasets: Chest X-ray images from two publicly available datasets: 1) CoronaHack-Chest dataset, where randomly selecting 50 images from COVID-19 cases and the other classes was done. [33]. 2) Vancouver General Hospital (VGH), British Columbia, Canada dataset which contains 85 COVID-19 X-ray images. The dataset used for this approach was available HERE Source code: Not available.</p>	<p>DarkNet-19 accuracy = 94.28% ResNet-50 accuracy = 93.69%</p>	Transfer Learning - Fine-tuning
[76]	<p>Datasets: chest X-ray Images from two datasets: 1) RSNA Pneumonia detection challenge [41]. 2) COVID-19 X-rays [33] Source code: Not available.</p>	<p>Accuracy = 95.5%</p>	Transfer Learning - Fine-tuning
[53]	<p>Datasets: Chest X-ray from a publicly available COVID-19 dataset [40] [54]. Source code: Not available.</p>	<p>Accuracy: 98% Precision: 100%</p>	Transfer Learning - Fine-tuning
[116]	<p>Datasets: Chest X-ray images from four publicly available datasets. 1) 3296 COVID-19 chest X-ray images from [35]. 2) 662 non-COVID-19 chest X-ray images from [117]. 3) 138 non-COVID-19 chest X-ray images from [117]. 4) 3343 non-COVID-19 chest X-ray images from [33]. Source code: Not available.</p>	<p>Accuracy = 95.83% F1-score = 95.94%</p>	Training from scratch - Multiple Models
[86]	<p>Datasets: Chest X-ray images from a publicly available dataset [31], which contains chest X-ray images for three classes: COVID-19, normal, and viral pneumonia. The total number of images is 219, 1345, and 1341 images for COVID-19, pneumonia, and normal cases respectively. Source code: Available HERE</p>	<p>Average F1-score = 0.97</p>	Transfer Learning - Features extraction.
[89]	<p>Datasets: chest X-ray images from two publicly available datasets: 1) COVID-19 datasets [33]. 2) chest X-ray datasets of pneumonia, normal, and COVID-19 [40]. Total images of COVID-19 is 295, 65 of normal images, pneumonia X-ray images of 98. where the total chest X-ray images is 458 . Source code: Not available.</p>	<p>Accuracy of 98.25%</p>	Transfer Learning - Features extraction
[118]	<p>Datasets: The authors used their own chest X-ray dataset, where the images are categorized into three classes: Pneumonia, COVID-19, and normal. Total images per category was 364 chest X-ray image. Source code: Available HERE</p>	<p>Accuracy = 99.08%</p>	Training from scratch - Multiple Models

TABLE 4. Summary of reviewed papers which include the datasets, performance, and the category of the COVID-19 detection methods.

Ref	Datasets and Source code	Performance	Category
[39]	Datasets: COVIDx dataset: a benchmark dataset consisting of 13975 chest X-ray images including COVID-19, pneumonia, and normal chest X-ray images. This dataset was collected from multiple publicly available datasets Source code: Not available.	Accuracy = 93.3%	Training from scratch - Single Model
[88]	Datasets: chest X-ray images from two publicly available datasets [33] [33], and [37]. Source code: Not available.	Accuracy = 92.18% F1 score = 92.07%	Transfer Learning - Features extraction
[62]	Datasets: chest X-ray images from publicly available datasets [33] and [37] Source code: Not available.	Accuracy = 98.75% (for two classes classification) Accuracy = 93.48% (for three classes classification task)	Transfer Learning - Fine-tuning
[74]	Datasets: chest X-ray images from publicly available datasets [33] [75] [42]. Source code: Not available.	Accuracy = 98%	Transfer Learning - Fine-tuning
[68]	Datasets: chest X-ray images from three publicly available dataset: [33] [36] [70]. Source code: Not available.	Accuracy = 100% AUC = 1.0	Transfer Learning - Fine-tuning
[90]	Datasets: chest X-ray images collect from multiple publicly available datasets [33] [42]. The collected data was named RYDLS-20. Source code: Not available.	F1-score of 0.889	Transfer Learning - Features Extraction
[79]	Datasets: chest X-ray images from [37], and COVID-19 dataset from Sylhet Medical College, Bangladesh. Source code: Available HERE	Accuracy = 96.9%	Training from scratch - Single Model
[134]	Datasets: chest X-ray images from two publicly available datasets [33] and RSNA pneumonia detection challenge. Source code: Available HERE	Accuracy for all classes = 91.40% Accuracy for COVID-19 = 99.56%	Training from scratch - Multiple Models
[72]	Datasets: chest X-ray images from two publicly available datasets: 1) COVID-19 dataset [33] 2) Kaggle chest X-ray pneumonia dataset [37]. Source code: Not available.	Accuracy = 98.26%	Transfer Learning - Fine-tuning

TABLE 5. Summary of reviewed papers which include the datasets, performance, and the category of the COVID-19 detection methods.

Ref	Datasets and Source code	Performance	Category
[89]	<p>Datasets: chest X-ray Images from three publicly available datasets 1) NIH chest X-ray dataset COVID19 chest X-ray dataset [42]. 3) COVID-19 chest X-ray images dataset [33]. Source code: Not available.</p>	Accuracy = 98.64% F1-score = 98.63%	Transfer Learning - Features Extraction
[99]	<p>Datasets: chest X-ray images from two publicly available datasets: 1) chest X-ray images from [100], of normal and pneumonia cases. 2) chest X-ray images of COVID-19 patients [33]. Source code: Not available.</p>	Accuracy of feed-forward neural network using raw images = 98.82% Accuracy using CNNs =95.48%. Accuracy using texture features as input to feed-forward neural network was 98.56%.	Transfer Learning - Features Extraction
[110]	<p>Datasets: chest X-ray images from three publicly available datasets: 1) Montgomery County and Shenzheno. 3 People's Hospital [70]. 2) Chest X-ray images dataset [37]. 3) COVID-19 chest X-ray images [33]. Source code: Not available.</p>	Accuracy using MANet is 97.06%	Training from scratch - Single Model
[67]	<p>Datasets: chest X-ray images from two publicly available datasets: 1) chest X-ray images of 219, 1345,1341 for COVID-19, pneumonia, and normal respectively [40] 2) Total of 142 COVID-19 chest X-ray images [45]. Source code: Not available.</p>	Accuracy for three classes classification was 99.08% Accuracy for two classes classification was 99.52%	Transfer Learning - Fine-tuning
[107]	<p>Datasets: Chest X-ray images from two publicly available datasets 1) 219 COVID-19 cases chest X-ray images [40]. 2) total of 1107 chest X-ray images of effusion disease and normal cases [34]. Source code: Not available.</p>	Accuracy:for two classes classification 98.62%. For three classes classification 95.77%	Training from scratch - Single Model
[133]	<p>Datasets: Chest X-ray images from publicly available dataset [33], which consists of 400 COVID-19 cases and 400 normal cases. Source code: Not available.</p>	Accuracy: 98%	Training from scratch - Multiple Models
[64]	<p>Datasets: Chest X-ray from two publicly available datasets: 1) 193 images were selected randomly for each of normal and pneumonia classes. [36] 2) 163 COVID-19 images from [33]. Source code: Not available.</p>	Accuracy: 98%	Transfer Learning - Fine-tuning
[65]	<p>Datasets: Chest X-ray images from two publicly available datasets: 1) Total of 141 COVID-19 positive cases from [33]. 2) Total of 1341 images of normal X-ray images from [36] . Source code: Not available.</p>	Accuracy on validation set = 99.45% F1-score on validation set = 97.13%	Transfer Learning - Fine-tuning
[112]	<p>Datasets: chest X-ray images from public COVID-19 dataset [39] and chest X-ray images collected from affiliated hospitals in MA USA, and South Korea . Source code: Not available.</p>	Accuracy = 83.9%	Training from scratch - Single Model

TABLE 6. Summary of reviewed papers which include the datasets, performance, and the category of the COVID-19 detection methods.

Ref	Datasets and Source code	Performance	Category
[80]	Datasets: Chest X-ray images of posterior-anterior view (666 images), anterior-posterior view (582 images) [135] [33] [41] Source code: Available HERE	F1-score = 62%	Transfer Learning - Fine-tuning
[126]	Datasets: chest X-ray images with segmentation ground truth [126]. Source code: Not available.	Dice coefficient 97.99%	Train from Scratch- Single Model
[127]	Datasets: Chest X-ray images from publicly available datasets [33] [33], and Costa Rica dataset. Source code: Not available.	Accuracy of 95.5%	Training from scratch - Single Model
[129]	Datasets: Dataset collected from multiple hospitals including: University of Pennsylvania Health System PA, USA and Brown University hospitals, RI, USA Source code: Not available.	ROC-AUC 0.846	Training from Scratch - Single Model
[81]	Datasets: Chest X-ray images from CheXpert [136]. Source code: Available HERE.	ROC-AUC 0.97	Transfer Learning - Fine-tuning
[82]	Datasets: Chest X-ray images from publicly available dataset [100] [33] Source code: Not available.	F1-score 0.8953	Transfer Learning - Fine-tuning
[43] [83]	Datasets: Chest X-ray images of pneumonia and COVID-19 cases (Early-QaTa-COV19 and QaTa-COV19 datasets) [43] Source code: Not available.	Accuracy: 95.6%	Transfer Learning - Fine-tuning
[105]	Datasets: Chest X-ray images [33] [40]. Source code: Not available.	Accuracy 98.36%	Transfer Learning - Features Extraction
[84]	Datasets: Chest X-ray images [41] [33]. Source code: Not available.	AUC 0.857	Transfer Learning - Fine-tuning
[44]	Datasets: Chest X-ray images [44]. Source code: Not available.	AUC 0.89	Transfer Learning - Fine-tuning
[130]	Datasets: Chest X-ray images [42] [33]. Source code: Not available.	AUC 0.90	Transfer Learning - Features extraction
[85]	Datasets: Chest X-ray images [41] and Private dataset from Emory University Hospital, GA, USA. Source code: Not available.	AUC 0.93	Transfer Learning - Fine-tuning
[131]	Datasets: Chest X-ray images [42] [33]. Source code: Not available.	Accuracy 91.3%	Training from scratch - Multiple Models
[132]	Datasets: Chest X-ray images [132]. Source code: Not available.	Accuracy 72.77%	Training from scratch - Single Model
[13]	Datasets: Chest X-ray images [33]. Source code: Not available.	AUC 0.988	Transfer Learning - Features Extraction
[61]	Datasets: Chest X-ray images [33]. Source code: Not available.	AUC 0.988	Transfer Learning - Features Extraction

generate a smaller balanced dataset for training deep learning models. In [90], the author used a Synthetic Minority Over-sampling Technique (SMOTE) to synthesize new examples (features) for the minority class [98] to balance the dataset. Another approach to balance the instances in the dataset was applied in [72], where data augmentation was extensively applied for the minority class. Another novel approach for training neural networks with an imbalanced dataset was proposed in [133]. This training mechanism for an imbalanced dataset started by dividing the majority class into many subsets of size equal to the minority class. Then, each subset

is combined with the minority class to form a smaller training set. The training was done for all the subsets of the majority class n in combination with the minority class, which gives f_n trained models. The drawback of this approach is the time consumption of training multiple n neural networks.

The published datasets consist of chest X-ray scans that were designed for chest X-ray research. However, the datasets need careful filtration and assessment to be used for COVID-19 detection. As we previously mentioned, a machine learning model relies on the information extracted from the training images. We noticed that the chest X-ray

images include scanned areas such as the shoulders and neck, which provide insignificant information for determining the presence of COVID-19 infection. In fact, such parts of the image may influence the accuracy of the model classification.

In the case of COVID-19 infection, along with other pulmonary diseases, visual inspection and investigation of the lung by an expert is done using chest X-ray images. Therefore, using the entire chest X-ray images for deep learning or extracting features without segmentation of the lung may make the model learn and extract features from other irrelevant parts of the X-ray images which are not related to the lung infection. Moreover, the chest X-ray images often have texts around the images which could lead the model to learn these artifacts and distinguish between images based on features unrelated to the whole lung region. An approach to segment the lung lobes from chest X-ray images for learning effective and discriminative features using a deep learning model was proposed in [108]. This approach first segment the lung lobes using a FC-DeeSenseNet103 deep learning network. The segmentation was used to mask out parts of chest X-ray images outside the lung lobes. Then the resulted images are used for training a classification deep learning model. Another approach to segment the whole lung region using the ensemble of EfficientNet [115] was done in [112]. In [110], segmentation of the whole lung region was done using U-Net with ResNet backbone. The goal of these approaches is to allow the deep learning model to learn to focus on lung related features by masking out the irrelevant area of the chest X-ray images. Therefore, applying a segmentation stage to chest X-ray images in order to discard irrelevant content of the images is very crucial for training deep learning neural networks.

We observed that transfer learning of pre-trained deep learning models on images of different domains (such as natural images from ImageNet) was mostly facilitated in the reviewed papers. Although transfer learning then fine-tuning of pre-trained models is a promising approach, it is important that the data sample be large enough to avoid over-fitting, especially in large neural networks such as ResNet [18]. Additionally, some freezing of the pre-trained neural network layers should be applied or truncation of the neural network layers could help avoid over-fitting issues. For instance, in [68], the authors truncated the Inception neural network pre-trained on ImageNet for fine-tuning a relatively smaller dataset of chest X-ray images [51].

The COVID-19 datasets used in the reviewed articles were based on publicly available datasets provided by researchers. Most COVID-19 datasets have a limited number of chest X-ray images, therefore, some researchers have aggregated multiple datasets for training and testing deep learning models. However, we noticed that most of the articles used multiple datasets of COVID-19 including the combined datasets, which could cause a duplicate of images in the dataset, particularly if the duplicate images are in both the training and testing sets. Therefore, combining multiple datasets into a joint dataset could cause the unreliable performance of deep

learning models. We suggest that the authors use a single dataset for training, and the other for testing while applying approaches to compute quantitative similarity metrics between images in the datasets, for the purpose of avoiding image duplication [149]. Moreover, recent studies show that a model trained on images from one institute does not generalize well to other images collected from a different institute [150]. The lack of such a universal model could place a barrier in front of the use of deep learning models in clinics. Therefore, we recommend the generalization of deep learning models by ensuring that the model is trained on a wide variety of different chest X-ray images, which are collected under different conditions and protocols.

There are other factors observed in the chest X-ray images datasets that could cause perturbation to the classification model. These factors are a combination of pediatrics with adult X-rays, different field of view (FOV) scans, machine energy output (KVp), and electrical current (mAS). Therefore, unifying the acquisition protocols of chest X-ray images is important for efficient and general machine learning models.

Some researchers have proposed deep learning approaches for multi-class classification such as COVID-19, bacterial pneumonia, viral pneumonia, and normal using X-ray images [73], [75], [80], [86], [89], [107], [108], and [110]. However, these approaches uses entire X-ray images scans which could learn non-lung related features. Therefore, an optimal solution would be to segment the lung lobes from each X-ray image to learn multi-class effectively.

Data and label quality are vital for effective machine learning models. The consistency of data labels poses a challenge to learning algorithms. These inconsistent labeling could result from human errors because of the similarity between COVID-19 infection and pneumonia in terms of visual chest X-ray appearance. Therefore, the data should be assigned with detailed and clear labels. However, the labels of the data are required to be annotated by more than one user or verified by experts in the field to ensure the correctness of the labels. Therefore, we suggest building a dataset of images for COVID-19 and other pneumonia cases, where the labels for each image are verified by more than one expert in the field. An approach that may help in assisting labeling of data was proposed in [112], where images similar to a given image are retrieved for the expert to facilitate image labeling.

VI. CONCLUSION AND FUTURE RESEARCH DIRECTION

Deep learning has revealed many promising avenues for detecting COVID-19 cases using the data from chest X-ray images.

A promising future direction for research is combining multiple information sources for deep learning models. For instance, adding information from clinical records of patients such as disease history and vital signs in combination with model-driven information derived from chest X-ray images should improve deep learning models to further support radiologists' decisions. For deep learning models, combining

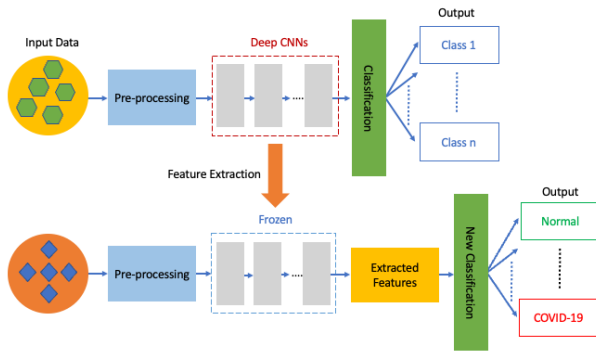


FIGURE 6. Feature extraction in transfer learning.

information when training the model should help in detecting COVID-19 more effectively improving accuracy and generalisability.

This review paper summarizes deep learning approaches for COVID-19 detection using chest X-ray images. Furthermore, this article summarizes the available datasets of chest X-ray images used in the reviewed approaches. Also, the article presents a discussion of the reviewed approaches and highlights directions for improvement. We also discuss the future direction of COVID-19 detection using chest X-ray images.

APPENDIX A

A. A SUMMARY OF DEEP LEARNING MODEL ARCHITECTURES AND VISUALIZATIONS

Image classification is a highly time-consuming process. To train a model with high accuracy, one needs a large dataset of labeled data. However, collecting these datasets is complex and expensive. Transfer learning approaches using a pre-trained deep learning model may be one solution to overcome these difficulties. In transfer learning, a deep neural network model is pre-trained on a general and large dataset, such as ImageNet. Then transfer the model to a different dataset of the same domain. In transfer learning, two approaches can be applied: feature extraction and fine-tuning.

Figure 6 describes a general concept of feature extraction in transfer learning. The pre-trained network is used to extract features from a new dataset. Then input images can be classified with the resulting features by simply substituting a new classifier layer or applying classification algorithms such as Support Vector Machine (SVM). Now the new model can classify images in the new dataset with the extracted features and a new classifier layer that is repurposed to classify images in the new dataset.

Figure 7 depicts a scenario of a fine-tuning model. A deep convolutional neural network (DCNN) model is first trained on a different and larger dataset. The early layers of the DCNN are more generic, but the later layers are more dependent on the dataset. Fine-tuning refers to freezing early layers, then unfreezing some top layers of the pre-trained model, and finally substituting a new classifier layer. The unfrozen layers

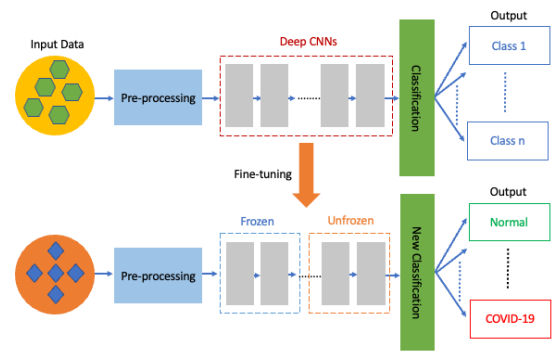


FIGURE 7. Fine-tune in transfer learning.

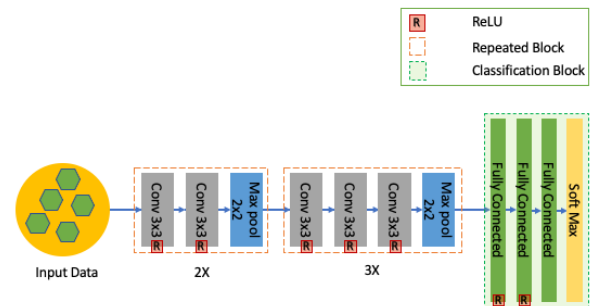


FIGURE 8. VGG-16.

and new classifiers will be trained from scratch over the new dataset. Learned features are fine-tuned to be more relevant to the new data set.

Figure 8 presents a VGG-16 (or OxfordNet) model architecture. VGG-16 was proposed by Karen Simonyan and Andrew Zisserman in 2014 [48]. It comprises 16 convolutional layers and achieved 92.7% top-5 test accuracy on the ImageNet dataset that contains over 14 million images with 1000 classes. VGG-16 is still widely used for image classification and localization problems. The stacks of convolutional layers are followed by fully connected layers and a Softmax activation layer.

Fully connected layers generally refer to the last few layers in a convolutional neural network. The output from the last pooling or convolutional layers becomes the input for the fully connected layer, which is a flattened vector. Activation functions (often called transfer functions) are operated through activation layers. Typically, for hidden layers (which receive the input from the previous layer and send the output to the next layer), three types of activation functions are used; Rectified Linear Activation (ReLU), Logistic (Sigmoid), and Hyperbolic Tangent (Tanh). For output layers (which produce predictions), linear, logistic (sigmoid), and Softmax activation functions are considered. Pooling layers usually come after the convolutional layers. Pooling layers are used to reduce the dimensions of the features and the number of parameters. This process increases so computational

- [8] S. Hu, G. Yang, H. Ye, J. Xia, W. Menpes-Smith, Y. Gao, Z. Niu, Y. Jiang, L. Li, X. Xiao, M. Wang, and E. F. Fang, "Weakly supervised deep learning for COVID-19 infection detection and classification from CT images," *IEEE Access*, vol. 8, pp. 118869–118883, 2020.
- [9] A. Tahamtan and A. Ardebili, "Real-time RT-PCR in COVID-19 detection: Issues affecting the results," *Exp. Rev. Mol. Diag.*, vol. 20, no. 5, pp. 453–454, May 2020.
- [10] D. Chen, S. Ji, F. Liu, Z. Li, and X. Zhou, "A review of automated diagnosis of COVID-19 based on scanning images," in *Proc. 6th Int. Conf. Robot. Artif. Intell.*, Nov. 2020, pp. 97–104.
- [11] H. S. Alghamdi, G. Amoudi, S. Elhag, K. Saeedi, and J. Nasser, "Deep learning approaches for detecting COVID-19 from chest X-ray images: A survey," *IEEE Access*, vol. 9, pp. 20235–20254, 2021.
- [12] W. Hariri and A. Narin, "Deep neural networks for COVID-19 detection and diagnosis using images and acoustic-based techniques: A recent review," *Soft Comput.*, vol. 25, no. 24, pp. 15345–15362, Dec. 2021.
- [13] M. A. Mohammed, K. H. Abdulkareem, A. S. Al-Waisy, S. A. Mostafa, S. Al-Fahdawi, A. M. Dinar, W. Alhakami, A. Baz, M. N. Al-Mhiqani, H. Alhakami, N. Arbai, M. S. Maashi, A. A. Mutlag, B. García-Zapirain, and I. de la Torre Díez, "Benchmarking methodology for selection of optimal COVID-19 diagnostic model based on entropy and toposis methods," *IEEE Access*, vol. 8, pp. 99115–99131, 2020.
- [14] W. C. S. Low, J. H. Chuah, C. A. T. H. Tee, S. Anis, M. A. Shoaib, A. Faisal, A. Khalil, and K. W. Lai, "An overview of deep learning techniques on chest X-ray and CT scan identification of COVID-19," *Comput. Math. Methods Med.*, vol. 2021, pp. 1–17, Jun. 2021.
- [15] S. V. Ligi, S. S. Kundu, R. Kumar, R. Narayanamoorthi, K. W. Lai, and S. Dhanalakshmi, "Radiological analysis of COVID-19 using computational intelligence: A broad gauge study," *J. Healthcare Eng.*, vol. 2022, pp. 1–25, Feb. 2022.
- [16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [17] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2006, pp. 2169–2178.
- [18] 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.
- [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25, Dec. 2012, pp. 1097–1105.
- [20] D. Massiceti, L. Zintgraf, J. Bronskill, L. Theodorou, M. Tobias Harris, E. Cutrell, C. Morrison, K. Hofmann, and S. Stumpf, "ORBIT: A real-world few-shot dataset for teachable object recognition," 2021, *arXiv:2104.03841*.
- [21] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1452–1464, Jun. 2018.
- [22] H. Jiang, G. Larsson, M. M. G. Shakhnarovich, and E. Learned-Miller, "Self-supervised relative depth learning for urban scene understanding," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 19–35.
- [23] T. Salem, S. Workman, and N. Jacobs, "Learning a dynamic map of visual appearance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12435–12444.
- [24] C. Zhang, Z. Cui, Y. Zhang, B. Zeng, M. Pollefeys, and S. Liu, "Holistic 3D scene understanding from a single image with implicit representation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8833–8842.
- [25] S. Deng, X. Xu, C. Wu, K. Chen, and K. Jia, "3D AffordanceNet: A benchmark for visual object affordance understanding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1778–1787.
- [26] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, May 2019, doi: 10.1016/j.zemedi.2018.11.002.
- [27] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018.
- [28] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
- [29] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. Hershey, PA, USA: IGI Global, 2010, pp. 242–264.
- [30] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, pp. 1345–1359, Nov. 2010.
- [31] A. M. V. Dadario. (2020). *COVID-19 X Rays*. Accessed: Dec. 8, 2021. [Online]. Available: <https://www.kaggle.com/dsv/1019469>
- [32] B. van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database," *Med. Image Anal.*, vol. 10, pp. 19–40, Feb. 2007.
- [33] 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*.
- [34] K. Murali. (2019). *CXR-Data*. Accessed: Dec. 8, 2021. [Online]. Available: <https://www.kaggle.com/murali0861/cxrdata>
- [35] M. D. L. I. Vayá, J. M. Saborit, J. A. Montell, A. Pertusa, A. Bustos, M. Cazorla, J. Galant, X. Barber, D. Orozco-Beltrán, F. García-García, M. Caparrós, G. González, and J. M. Salinas, "BIMCV COVID-19+: A large annotated dataset of RX and CT images from COVID-19 patients," 2020, *arXiv:2006.01174*.
- [36] P. Mooney. (2018). *Chest X-Ray Images (Pneumonia)*. Accessed: Dec. 8, 2021. [Online]. Available: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia/>
- [37] D. Kermany, K. Zhang, M. Goldbaum, "Labeled optical coherence tomography (OCT) and chest X-ray images for classification," *Mendeley Data*, vol. 2, no. 2, 2018.
- [38] Praveen. (2020). *Coronahack-Chest X-Ray Datasets*. [Online]. Available: <https://www.kaggle.com/praveengovi/coronahack-chest-xraydataset>
- [39] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, Dec. 2020.
- [40] M. E. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, and N. Al Emadi, "Can AI help in screening viral and COVID-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132665–132676, 2020.
- [41] P. Rui and K. Kang. (2018). *National Ambulatory Medical Care Survey: 2015 Emergency Department Summary Tables. Table 27*. Accessed: Dec. 8, 2021. [Online]. Available: <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>
- [42] 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.
- [43] M. Ahishali, A. Degerli, M. Yamac, S. Kiranyaz, M. E. Chowdhury, K. Hameed, T. Hamid, R. Mazhar, and M. Gabbouj, "Advance warning methodologies for COVID-19 using chest X-ray images," *IEEE Access*, vol. 9, pp. 41052–41065, 2021.
- [44] I. Castiglioni, D. Ippolito, M. Interlenghi, C. B. Monti, C. Salvatore, S. Schiaffino, A. Polidori, D. Gandola, C. Messa, and F. Sardanelli, "Machine learning applied on chest X-ray can aid in the diagnosis of COVID-19: A first experience from lombardy, Italy," *Eur. Radiol. Experim.*, vol. 5, no. 1, pp. 1–10, Dec. 2021.
- [45] 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.
- [46] 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.
- [47] M. Qjidaa, Y. Mechbal, A. Ben-fares, H. Amakdouf, M. Maaroufi, B. Alami, and H. Qjidaa, "Early detection of COVID19 by deep learning transfer model for populations in isolated rural areas," in *Proc. Int. Conf. Intell. Syst. Comput. Vis. (ISCV)*, Jun. 2020, pp. 1–5.
- [48] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [49] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-V4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4285.
- [50] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.

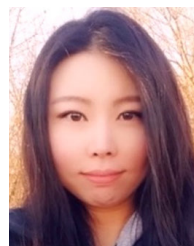
- [51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
- [52] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [53] O. El Gannour, S. Hamida, B. Cherradi, A. Raihani, and H. Moujahid, "Performance evaluation of transfer learning technique for automatic detection of patients with COVID-19 on X-ray images," in *Proc. IEEE 2nd Int. Conf. Electron., Control, Optim. Comput. Sci. (ICECOCS)*, Dec. 2020, pp. 1–6.
- [54] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. A. Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughair, M. S. Khan, and M. E. H. Chowdhury, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput. Biol. Med.*, vol. 132, May 2021, Art. no. 104319.
- [55] M. Elgendi, M. U. Nasir, Q. Tang, R. R. Fletcher, N. Howard, C. Menon, R. Ward, W. Parker, and S. Nicolaou, "The performance of deep neural networks in differentiating chest X-rays of COVID-19 patients from other bacterial and viral pneumonias," *Frontiers Med.*, vol. 7, p. 550, Aug. 2020.
- [56] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*.
- [57] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [58] J. Redmon. (2016). *DarkNet: Open Source Neural Netw. C*. Accessed: Dec. 8, 2021. [Online]. Available: <http://pjreddie.com/darknet/>
- [59] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.
- [60] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.
- [61] A. Shazia, T. Z. Xuan, J. H. Chuah, J. Usman, P. Qian, and K. W. Lai, "A comparative study of multiple neural network for detection of COVID-19 on chest X-ray," *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, pp. 1–16, Dec. 2021.
- [62] 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.
- [63] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [64] H. Amin, A. Darwish, and A. E. Hassanien, "Classification of COVID19 X-ray images based on transfer learning inceptionV3 deep learning model," in *Digital Transformation and Emerging Technologies for Fighting COVID-19 Pandemic: Innovative Approaches*. Cham, Switzerland: Springer, 2021, pp. 111–119.
- [65] S. Ghosh and M. Bandyopadhyay, "Detection of coronavirus (COVID-19) using deep convolutional neural networks with transfer learning using chest X-ray images," in *Machine Learning Approaches for Urban Computing*. Singapore: Springer, 2021, pp. 63–77.
- [66] 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.
- [67] A. Gupta, Anjum, S. Gupta, and R. Katarya, "InstaCovNet-19: A deep learning classification model for the detection of COVID-19 patients using chest X-ray," *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106859.
- [68] D. Das, K. C. Santosh, and U. Pal, "Truncated inception net: COVID-19 outbreak screening using chest X-rays," *Phys. Eng. Sci. Med.*, vol. 43, no. 3, pp. 915–925, Sep. 2020.
- [69] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [70] S. Jaeger, S. Candemir, S. Antani, Y.-X. Wang, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imag. Med. Surg.*, vol. 4, no. 6, p. 475, 2014.
- [71] S. Jaeger and S. Candemir. (2020). *Tuberculosis Chest X-Ray Image Data Sets*. Accessed: Dec. 8, 2021. [Online]. Available: <https://lhncbc.nlm.nih.gov/LHC-publications/pubs/TuberculosisChestXrayImageDataSets.html>
- [72] F. Ucar and D. Korkmaz, "COVIDiagnosis-Net: Deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images," *Med. Hypotheses*, vol. 140, Jul. 2020, Art. no. 109761.
- [73] W. Jia, C. Xiu-Yun, Z. Hao, X. Li-Dong, L. Hang, and D. Si-Hao, "Hyperparameter optimization for machine learning models based on Bayesian optimization," *J. Electron. Sci. Technol.*, vol. 17, no. 1, pp. 26–40, 2019.
- [74] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105608.
- [75] 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.
- [76] S. Misra, S. Jeon, S. Lee, R. Managuli, I.-S. Jang, and C. Kim, "Multi-channel transfer learning of chest X-ray images for screening of COVID-19," *Electronics*, vol. 9, no. 9, p. 1388, Aug. 2020.
- [77] X. Li, C. Li, and D. Zhu, "COVID-mobileXpert: On-device COVID-19 patient triage and follow-up using chest X-rays," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2020, pp. 1063–1067.
- [78] A. Wong, M. Javad Shafiee, B. Chwyl, and F. Li, "FermiNets: Learning generative machines to generate efficient neural networks via generative synthesis," 2018, *arXiv:1809.05989*.
- [79] T. Mahmud, M. A. Rahman, and S. A. Fattah, "CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization," *Comput. Biol. Med.*, vol. 122, Jul. 2020, Art. no. 103869.
- [80] M. Momeny, A. A. Neshat, M. A. Hussain, S. Kia, M. Marhamati, A. Jahanbakhshi, and G. Hamarneh, "Learning-to-augment strategy using noisy and denoised data: Improving generalizability of deep CNN for the detection of COVID-19 in X-ray images," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104704.
- [81] G. Wang et al., "A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images," *Nature Biomed. Eng.*, vol. 5, no. 6, pp. 509–521, Apr. 2021.
- [82] N. E. M. Khalifa, F. Smarandache, G. Manogaran, and M. Loey, "A study of the neutrosophic set significance on deep transfer learning models: An experimental case on a limited COVID-19 chest X-ray dataset," *Cognit. Comput.*, vol. 13, pp. 1–10, Jan. 2021.
- [83] M. Yamac, M. Ahishali, A. Degerli, S. Kiranyaz, M. E. H. Chowdhury, and M. Gabbouj, "Convolutional sparse support estimator-based COVID-19 recognition from X-ray images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 1810–1820, May 2021.
- [84] R. B. Fricks, F. Ria, H. Chalian, P. Khoshpouri, E. Abadi, L. Bianchi, W. P. Segars, and E. Samei, "Deep learning classification of COVID-19 in chest radiographs: Performance and influence of supplemental training," *J. Med. Imag.*, vol. 8, no. 6, Dec. 2021, Art. no. 064501.
- [85] M. Zandehshahvar, M. van Assen, H. Maleki, Y. Kiarashi, C. N. D. Cecco, and A. Adibi, "Toward understanding COVID-19 pneumonia: A deep-learning-based approach for severity analysis and monitoring the disease," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, Dec. 2021.
- [86] S. Tiwari and A. Jain, "Convolutional capsule network for COVID-19 detection using radiography images," *Int. J. Imag. Syst. Technol.*, vol. 31, no. 2, pp. 525–539, Jun. 2021.
- [87] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," 2017, *arXiv:1710.09829*.
- [88] K. El Asnaoui and Y. Chawki, "Using X-ray images and deep learning for automated detection of coronavirus disease," *J. Biomolecular Struct. Dyn.*, vol. 39, no. 10, pp. 1–12, 2020.
- [89] W. Jin, S. Dong, C. Dong, and X. Ye, "Hybrid ensemble model for differential diagnosis between COVID-19 and common viral pneumonia by chest X-ray radiograph," *Comput. Biol. Med.*, vol. 131, Apr. 2021, Art. no. 104252.
- [90] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla, and Y. M. G. Costa, "COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios," *Comput. Methods Programs Biomed.*, vol. 194, Oct. 2020, Art. no. 105532.

- [91] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.
- [92] L. Nanni, A. Lumini, and S. Brahnham, "Local binary patterns variants as texture descriptors for medical image analysis," *Artif. Intell. Med.*, vol. 49, no. 2, pp. 117–125, 2010.
- [93] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 1363–1366.
- [94] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Proc. Int. Conf. Image Signal Process.* Berlin, Germany: Springer, 2008, pp. 236–243.
- [95] A. Ramirez Rivera, R. Castillo, and O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1740–1752, May 2013.
- [96] M. Crosier and L. D. Griffin, "Using basic image features for texture classification," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 447–460, 2010.
- [97] T. Song, H. Li, F. Meng, Q. Wu, and J. Cai, "LETRIST: Locally encoded transform feature histogram for rotation-invariant texture classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 7, pp. 1565–1579, Jul. 2018.
- [98] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
- [99] 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.
- [100] 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.
- [101] J. Rabbah, M. Ridouani, and L. Hassouni, "A new classification model based on StackNet and deep learning for fast detection of COVID-19 through X rays images," in *Proc. 4th Int. Conf. Intell. Comput. Data Sci. (ICDS)*, Oct. 2020, pp. 1–8.
- [102] B. Bachir. (2020). *COVID-19 Chest Xray*. [Online]. Available: <https://www.kaggle.com/bachrr/covid-chest-xray/>
- [103] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, Jan. 1992.
- [104] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzivasvas, "A comparison of machine learning techniques for customer churn prediction," *Simul. Model. Pract. Theory*, vol. 55, pp. 1–9, Jun. 2015.
- [105] N.-A.-A. Alam, M. Ahsan, M. A. Based, J. Haider, and M. Kowalski, "COVID-19 detection from chest X-ray images using feature fusion and deep learning," *Sensors*, vol. 21, no. 4, p. 1480, Feb. 2021.
- [106] E. Irmak, "A novel deep convolutional neural network model for COVID-19 disease detection," in *Proc. Med. Technol. Congr. (TIPTe-KNO)*, Nov. 2020, pp. 1–4.
- [107] S. Lafraxo and M. E. Ansari, "CoviNet: Automated COVID-19 detection from X-rays using deep learning techniques," in *Proc. 6th IEEE Congr. Inf. Sci. Technol. (CiSt)*, Jun. 2020, pp. 489–494.
- [108] 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.
- [109] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-I. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *Amer. J. Roentgenol.*, vol. 174, no. 1, pp. 71–74, 2000.
- [110] Y. Xu, H.-K. Lam, and G. Jia, "MANet: A two-stage deep learning method for classification of COVID-19 from chest X-ray images," *Neurocomputing*, vol. 443, pp. 96–105, Jul. 2021.
- [111] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.
- [112] A. Zhong, X. Li, D. Wu, H. Ren, K. Kim, Y. Kim, V. Buch, N. Neumark, B. Bizzo, W. Y. Tak, S. Y. Park, Y. R. Lee, M. K. Kang, J. G. Park, B. S. Kim, W. J. Chung, N. Guo, I. Dayan, M. K. Kalra, and Q. Li, "Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101993.
- [113] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [114] X. Wang, X. Han, W. Huang, D. Dong, and M. R. Scott, "Multi-similarity loss with general pair weighting for deep metric learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5022–5030.
- [115] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
- [116] M. Owais, H. S. Yoon, T. Mahmood, A. Haider, H. Sultan, and K. R. Park, "Light-weighted ensemble network with multilevel activation visualization for robust diagnosis of COVID19 pneumonia from large-scale chest radiographic database," *Appl. Soft Comput.*, vol. 108, Sep. 2021, Art. no. 107490.
- [117] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 577–590, Feb. 2014.
- [118] M. Canayaz, "MH-COVIDNet: Diagnosis of COVID-19 using deep neural networks and meta-heuristic-based feature selection on X-ray images," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102257.
- [119] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new image contrast enhancement algorithm using exposure fusion framework," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, Springer, 2017, pp. 36–46.
- [120] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.
- [121] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE ICNN*, vol. 4, Nov./Dec. 1995, pp. 1942–1948.
- [122] M. Toğaçar, B. Ergen, and Z. Cömert, "COVID-19 detection using deep learning models to exploit social mimic optimization and structured chest X-ray images using fuzzy color and stacking approaches," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103805.
- [123] (2019). *Fuzzy Color Image Enhancement Algorithm*. [Online]. Available: <https://github>
- [124] (2019). *Image Stacking: Simple Code to Load and Process Images*. [Online]. Available: <https://github>
- [125] S. Balochian and H. Balochian, "Social mimic optimization algorithm and engineering applications," *Exp. Syst. Appl.*, vol. 134, pp. 178–191, Nov. 2019.
- [126] A. M. Tahir, M. E. H. Chowdhury, A. Khandakar, T. Rahman, Y. Qiblawey, U. Khurshid, S. Kiranyaz, N. Ibtihaz, M. S. Rahman, S. Al-Maadeed, S. Mahmud, M. Ezeddin, K. Hameed, and T. Hamid, "COVID-19 infection localization and severity grading from chest X-ray images," *Comput. Biol. Med.*, vol. 139, Dec. 2021, Art. no. 105002.
- [127] S. Calderon-Ramirez, S. Yang, A. Moemeni, D. Elizondo, S. Colreavy-Donnelly, L. F. Chavarria-Estrada, and M. A. Molina-Cabello, "Correcting data imbalance for semi-supervised COVID-19 detection using X-ray chest images," *Appl. Soft Comput.*, vol. 111, Nov. 2021, Art. no. 107692.
- [128] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
- [129] Z. Jiao et al., "Prognostication of patients with COVID-19 using artificial intelligence based on chest X-rays and clinical data: A retrospective study," *Lancet Digit. Health*, vol. 3, no. 5, pp. e286–e294, May 2021.
- [130] J. Zhou, B. Jing, Z. Wang, H. Xin, and H. Tong, "SODA: Detecting COVID-19 in chest X-rays with semi-supervised open set domain adaptation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Mar. 21, 2021, doi: [10.1109/TCBB.2021.3066331](https://doi.org/10.1109/TCBB.2021.3066331).
- [131] R. Sadre, B. Sundaram, S. Majumdar, and D. Ushizima, "Validating deep learning inference during chest X-ray classification for COVID-19 screening," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, Dec. 2021.
- [132] 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.
- [133] H. Hanafi, A. Pranolo, and Y. Mao, "CAE-COVIDX: Automatic COVID-19 disease detection based on X-ray images using enhanced deep convolutional and autoencoder," *Int. J. Adv. Intell. Inform.*, vol. 7, no. 1, pp. 49–62, 2021.
- [134] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of xception and ResNet50 V2," *Informat. Med. Unlocked*, vol. 19, Jan. 2020, Art. no. 100360.

- [135] M. Nishio, S. Noguchi, H. Matsuo, and T. Murakami, "Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: Combination of data augmentation methods," *Sci. Rep.*, vol. 10, no. 1, pp. 1–6, Dec. 2020.
- [136] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghighi, 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.
- [137] L. P. Lavelle, R. M. Dunne, A. G. Carroll, and D. E. Malone, "Evidence-based practice of radiology," *Radiographics*, vol. 35, no. 6, pp. 1802–1813, 2015.
- [138] D. E. Malone, "Evidence-based practice in radiology: An introduction to the series," *Radiology*, vol. 242, no. 1, pp. 12–14, Jan. 2007.
- [139] Evidence-Based Radiology Working Group, "Evidence-based radiology: A new approach to the practice of radiology," *Radiology*, vol. 220, no. 3, pp. 566–575, Sep. 2001.
- [140] B. A. Altazi, G. G. Zhang, D. C. Fernandez, M. E. Montejo, D. Hunt, J. Werner, M. C. Biagioli, and E. G. Moros, "Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms," *J. Appl. Clin. Med. Phys.*, vol. 18, no. 6, pp. 32–48, Nov. 2017.
- [141] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England J. Med.*, vol. 380, no. 14, pp. 1347–1358, 2019.
- [142] S. Wang and R. M. Summers, "Machine learning and radiology," *Med. Image Anal.*, vol. 16, no. 5, pp. 933–951, Jul. 2012.
- [143] Y. Liu, A. Jain, C. Eng, D. H. Way, K. Lee, P. Bui, K. Kanada, G. D. O. Marinho, J. Gallegos, S. Gabriele, and V. Gupta, "A deep learning system for differential diagnosis of skin diseases," *Nature Med.*, vol. 26, no. 6, pp. 900–908, 2020.
- [144] L. Dai, L. Wu, H. Li, C. Cai, Q. Wu, H. Kong, R. Liu, X. Wang, X. Hou, Y. Liu, X. Long, Y. Wen, L. Lu, Y. Shen, Y. Chen, D. Shen, X. Yang, H. Zou, B. Sheng, and W. Jia, "A deep learning system for detecting diabetic retinopathy across the disease spectrum," *Nature Commun.*, vol. 12, no. 1, pp. 1–11, Dec. 2021.
- [145] P. Wang, X. Xiao, J. R. G. Brown, T. M. Berzin, M. Tu, F. Xiong, X. Hu, P. Liu, Y. Song, and D. Zhang, "Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy," *Nature Biomed. Eng.*, vol. 2, no. 10, pp. 741–748, 2018.
- [146] O. Ratib, "PET/CT image navigation and communication," *J. Nucl. Med.*, vol. 45, no. 1, pp. 46S–55S, 2004.
- [147] W. Li, H. Cui, K. Li, Y. Fang, and S. Li, "Chest computed tomography in children with COVID-19 respiratory infection," *Pediatric Radiol.*, vol. 50, no. 6, pp. 796–799, May 2020.
- [148] C.-C. Lai, W.-C. Ko, P.-I. Lee, S.-S. Jean, and P.-R. Hsueh, "Extra-respiratory manifestations of COVID-19," *Int. J. Antimicrobial Agents*, vol. 56, no. 2, Aug. 2020, Art. no. 106024.
- [149] M. van Heel, "Similarity measures between images," *Ultramicroscopy*, vol. 21, no. 1, pp. 95–100, 1987.
- [150] K. B. Ahmed, G. M. Goldgof, R. Paul, D. B. Goldgof, and L. O. Hall, "Discovery of a generalization gap of convolutional neural networks on COVID-19 X-rays classification," *IEEE Access*, vol. 9, pp. 72970–72979, 2021.
- [151] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.



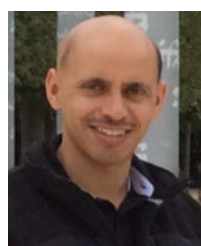
BADERALDEEN ALTAZI received the Bachelor of Science degree in nuclear engineering with a concentration in medical physics from King Abdulaziz University, Jeddah, Saudi Arabia, in July 2009, the Masters of Science and the Doctor of Philosophy degrees in applied physics with a concentration in medical physics from the University of South Florida (USF), Tampa, FL, USA, in May 2014 and October 2017, respectively, and the degree from the Radiation Oncology Physics Residency Program, McGill University, in March 2021. He is a Consultant, Radiation Oncology Physicist with the King Abdullah Medical City at the Holy Capital and affiliated with Batterjee Medical College, College of Medicine. He joined the USF-Moffitt Cancer Center CAMPEP Ph.D. Program. He also published in several peer-reviewed scientific journals. During the Ph.D. program, he had the privilege of joining several fields related to societies, such as the American Association of Physicists in Medicine (AAPM), the American Society for Radiation Oncology (ASTRO), and the Radiological Society of North America (RSNA). He was awarded the ASTRO Basic/Translational Science Abstract Award in the Radiation Physics category for abstract titled, "Comparing Radiomic Features and SUV as Predictors for Cervical Cancer Treatment Outcomes" ASTRO 57th Annual Meeting, San Antonio, TX, USA, in October 2015.



JISOO HWANG is currently pursuing the Ph.D. degree in technology with Purdue University, USA. Her research interests include machine learning, computer vision, and affective computing.



SAMUEL HAWKINS (Member, IEEE) received the Ph.D. degree in computer science from the University of South Florida, in December of 2017. He is currently an Assistant Professor at Bradley University, where he teaches and researches medical imaging and machine learning. Previously, he was a Postdoctoral Research Fellow at the Moffitt Cancer Center. His paper "Predicting Malignant Nodules from Screening CTs" won the *Journal of Thoracic Oncology* December 2016 Editor's Choice Article. His research interests include machine learning, image processing, data mining, and radiomics.



SAEED S. ALAHMARI (Member, IEEE) received the Ph.D. degree in computer science from the University of South Florida, in 2020. He is currently an Assistant Professor of computer science at Najran University, Najran, Saudi Arabia. He has authored or coauthored many journals and conference papers. His research interests include learning from noisy and limited labeled data, machine learning, deep learning, medical image understanding, computer vision, and deep learning repeatability and explainability.



TAWFIQ SALEM (Member, IEEE) received the Ph.D. degree in computer science from the University of Kentucky, in 2019. He is currently a Visiting Assistant Professor with the Department of Computer and Information Technology, Purdue University, USA. His research interests include machine learning, computer vision, and remote sensing.

• • •