

Received January 11, 2021, accepted January 21, 2021, date of publication January 25, 2021, date of current version February 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3054484

Deep Learning Approaches for Detecting COVID-19 From Chest X-Ray Images: A Survey

HANAN S. ALGHAMDI¹, GHADA AMOUDI¹, SALMA ELHAG¹, KAWTHER SAEEDI¹, AND JOMANAH NASSER²

¹Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

²Faculty of Medicine, King Abdulaziz University, Jeddah 80215, Saudi Arabia

Corresponding author: Hanan S. Alghamdi (hsaalghamdi@kau.edu.sa)

This work was supported by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under Grant GCV19-34-1441.

ABSTRACT Chest X-ray (CXR) imaging is a standard and crucial examination method used for suspected cases of coronavirus disease (COVID-19). In profoundly affected or limited resource areas, CXR imaging is preferable owing to its availability, low cost, and rapid results. However, given the rapidly spreading nature of COVID-19, such tests could limit the efficiency of pandemic control and prevention. In response to this issue, artificial intelligence methods such as deep learning are promising options for automatic diagnosis because they have achieved state-of-the-art performance in the analysis of visual information and a wide range of medical images. This paper reviews and critically assesses the preprint and published reports between March and May 2020 for the diagnosis of COVID-19 via CXR images using convolutional neural networks and other deep learning architectures. Despite the encouraging results, there is an urgent need for public, comprehensive, and diverse datasets. Further investigations in terms of explainable and justifiable decisions are also required for more robust, transparent, and accurate predictions.

INDEX TERMS Chest x-ray, coronavirus, COVID-19, deep learning, radiological imaging.

I. INTRODUCTION

Early diagnosis of the coronavirus disease (COVID-19) is essential to reduce the spread of the virus and provide care for preventing complications. The daily increments in COVID-19 cases worldwide and the limitations of the current diagnostic tools impose challenges in identifying and managing the pandemic. Researchers worldwide are actively participating to find effective diagnostic procedures and accelerate the development of a vaccine and treatments. As of the writing of this paper, three diagnostic procedures are commonly used: blood tests, viral tests, and medical imaging [1]. Blood tests detect the presence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) antibodies in the blood. However, the reliability of this test in diagnosing COVID-19 is as low as 2% or 3% [2]. Viral tests detect the antigens of SARS-CoV-2 using samples from the respiratory tract. The rapid diagnostic test (RDT) is a type of antibody detection test that is fast and can produce results in 30 min. However, the availability of RDT test kits is limited, and its

effectiveness depends on the sample quality and the time of onset of illness. Furthermore, the test can yield false positive results because it does not distinguish COVID-19 from other viral infections; therefore, it is not recommended for diagnosing COVID-19 [3]. Another commonly used viral test is reverse transcription polymerase chain reaction (RT-PCR). RT-PCR is the gold-standard tool used as the first-line screening choice [4]. However, large-scale studies have found that the test result sensitivity ranges between 50–62% [4]. This implies that an initial negative RT-PCR result can be obtained. Therefore, to ensure the correctness of the test result for diagnosis, multiple RT-PCR tests are performed over a 14-day observation period. In other words, an RT-PCR negative result for a suspected case of COVID-19 is only considered as a true negative when there are no positive RT-PCR results after multiple tests have been taken over the 14-day observation period [5]. This can be frustrating for the patient and costly for the healthcare authorities owing to the shortage of RT-PCR test kits in several countries [6].

Because COVID-19 targets the respiratory system, chest radiology scans are an important tool for diagnosis and early management. Chest X-rays (CXR) have been used

The associate editor coordinating the review of this manuscript and approving it for publication was Carmen C. Y. Poon ¹.

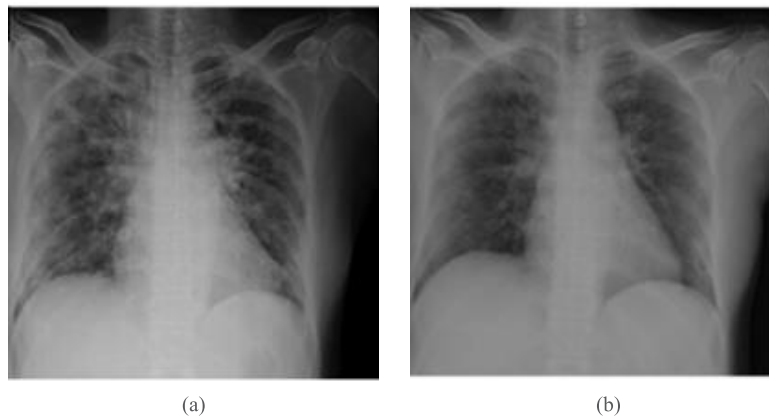


FIGURE 1. CXR findings: (a) GGO and (b) GGO with consolidation [8].

as a first-line diagnostic tool in Italy and various other countries [7]. The condition of the lungs can be effectively detected using radiology scans along with the different stages of illness or recovery [8]. Radiologists have recorded a range of abnormalities found in the radiology scans of COVID-19 patients. Fig. 1 shows two examples of COVID-19 features in CXR images, namely, bilateral GGO and bilateral and multifocal GGO with consolidation.

CXR is a widely available tool in most clinical settings; it is less time-consuming in terms of patient preparation and immediate diagnosis. Consequently, CXR can be used for patient triage, deciding the priority of patient treatments, and utilizing medical resources.

In the medical imaging domain, deep learning (DL) techniques have been used to improve the performance of image analysis significantly [9], [10]. For example, DL has been successfully applied to microscopy images [11], brain tumor classification [12], MRI images [13], and retinal photographs [14].

Convolutional neural networks (CNNs) are commonly used for medical imaging [15], [16]; they have various architectures and applications. Therefore, since the first few months of the pandemic, DL approaches have been extensively explored for diagnosing COVID-19 from radiology photographs. In this paper, we review the latest research contributions of the application of DL for the detection of COVID-19 from CXR images by comparing the existing DL technologies, highlighting the challenges, and identifying the required future investigations.

To understand how CNNs and other DL architectures could facilitate the diagnosis of COVID-19 via CXR images, this paper reviews and critically assesses the preprint and published reports made available between March and May 2020 on this topic. The articles were found in several common research databases, such as PubMed, ScienceDirect, Springer, IEEE, ACM, Scopus, ArXiv, and MedRxiv. The keywords used in the search included “transfer learning,” “convolutional,” “deep learning,” “radiograph,” “chest x-ray,” “CXR,” “COVID,” and “Coronavirus;” this list

was regularly updated since the beginning of this study. We reviewed the paper abstracts and excluded those studies that considered DL for computed tomography images and those that used traditional machine learning algorithms. When articles from multiple resources overlapped, only the most recent articles were considered. Fig. 2 shows a histogram of the distribution dates of the papers included in this review.

Section II provides a detailed review of DL applications for COVID-19 CXR image analysis, including the architectures used and CXR datasets employed. Section III presents the methodology analysis and performance comparisons of various application of deep learning modeling. Section IV highlights the current challenges and discusses the findings of this survey, including public dataset establishment, model optimization, model uncertainty, and the unexplained black-box decisions made by the DL models. Finally, the paper is concluded with some suggestions for further investigations in Section V.

II. LITERATURE SURVEY

We reviewed 34 articles investigating the use of DL models to examine CXR images with SARS-CoV-2 viral infections. The vast majority of the studies (71%) implemented transfer learning using publicly available CNN architectures trained on the ImageNet dataset. These architectures with their parameters and hyperparameter settings are publicly available [17]. However, 29% of the studies went beyond using off-the-shelf tools and implemented novel architectures. In the following subsections, we provide a general overview of the main approaches and datasets that were used in the research works reviewed in this survey.

A. CLASSIFICATION TASK FORMULATION

The COVID-19 detection results are given by classifying the CXR images into 2–4 classes, i.e., binary or multi-class classification. Each class represents one or more labels: “healthy,” “no finding,” “bacterial pneumonia,” “viral pneumonia,” or “COVID-19.” Two-class classification

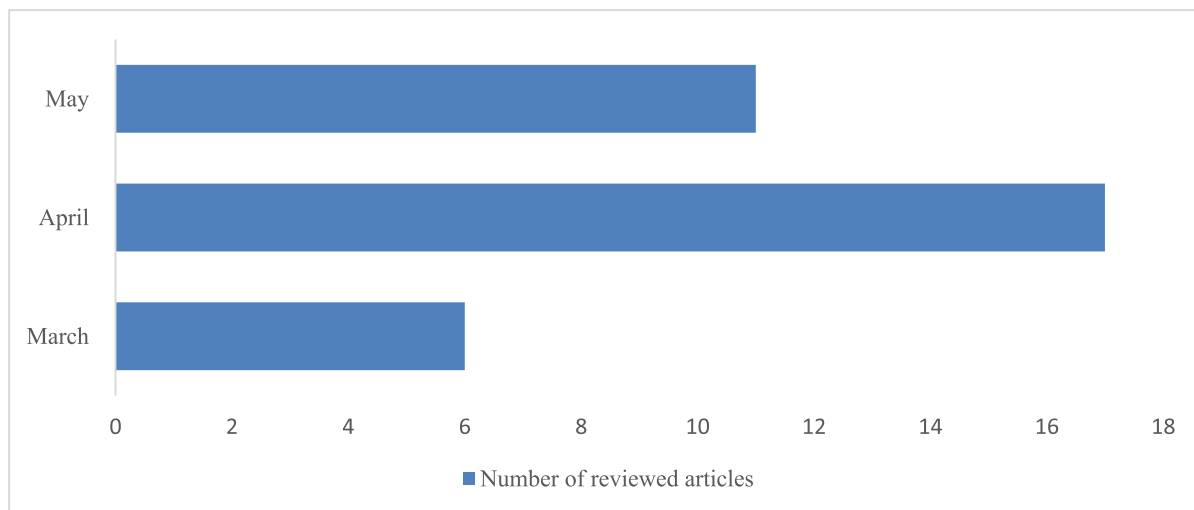


FIGURE 2. Distribution of the reviewed articles in the months of 2020.

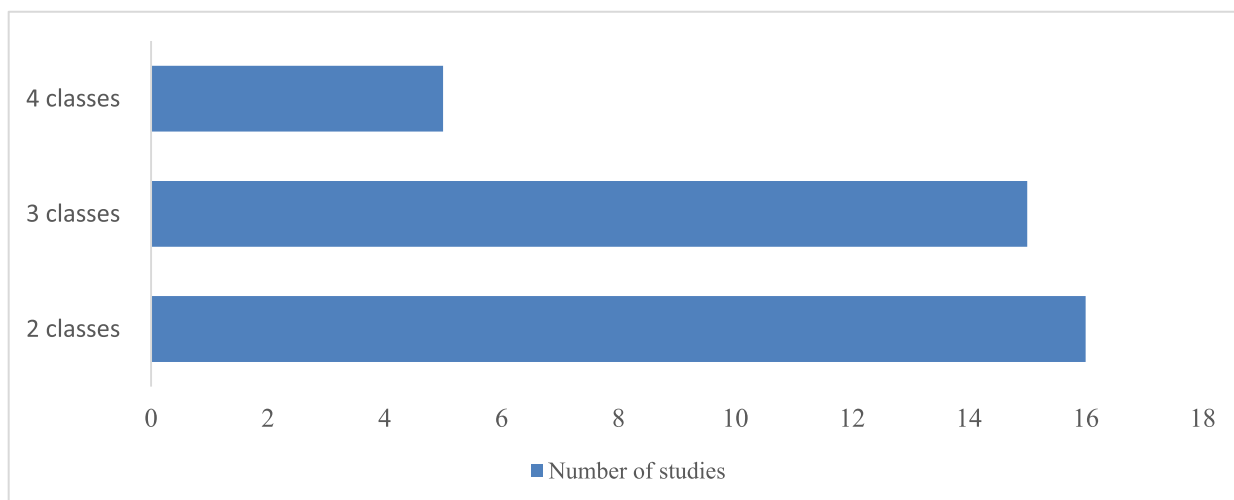


FIGURE 3. Distribution of studies according to classification task formulation.

is called binary classification, and its results include the COVID-19 label and either of the following labels: “healthy,” “no finding,” “bacterial pneumonia,” or “viral pneumonia.” The three-class results include “COVID-19,” “healthy or no finding,” and “pneumonia.” The four classes results include “COVID-19,” “healthy or no finding,” “bacterial pneumonia,” and “viral pneumonia.” Most of the reviewed research used two or three classes. Fig. 3 shows the number of reviewed studies grouped by the number of classes used in the classification task.

B. DATASETS

In the reviewed articles, 14 different datasets were used. A summary of these datasets is presented in Table 1. Each row specifies the reference, dataset name, a brief description, and whether the dataset contains COVID-19 samples. According to our survey, COVID-19 Image Data Collection [18] is the

most cited dataset. It contains images extracted from various online publications and websites in an attempt to provide AI researchers with COVID-19 photographs to develop deep learning-based models. Each image in this dataset is accompanied by a set of attributes such as sex, age, date, survival, and clinical notes.

Table 2 reveals the quantitative values related to the datasets including number of records, image resolution, types, and the URL. The URL is the address to access the dataset, which offers opportunity for researchers to reuse the dataset, compare the results, and enhance their knowledge. Table 2 shows that Github.com and Kaggle.com are the mostly used platforms to store and make datasets available online.

It is worth noting that, some studies referred to the same dataset by a different name; for instance, the COVID-19 Image Data Collection [18] has been referred to as the

TABLE 1. Description of the COVID-19 datasets used in the reviewed research.

No.	Reference	Name(s)	Description	COVID-19 Samples
1	[18]	COVID-19 Image Data Collection	Public CXR images of COVID-19 collected from various online sources with ongoing updates.	✓
2	[20]	Figure 1 COVID-19 CXR Dataset Initiative (COVIDx)	48 CXR images of COVID-19	✓
3	[61]	ActualMed COVID-19 CXR Dataset Initiative	N/A	✓
4	[62]	COVID-19 Radiography Database	2,905 CXR images with 219 COVID-19 positive cases	✓
5	[63]	Japanese Society of Radiological Technology	105 CXR images of COVID-19 cases	✓
6	[22] sometimes referenced as [64]	ChestX-ray8 ChestX-ray14	Contains 112,120 frontal-view CXR images of 30,805 patients labeled with 14 thoracic diseases. This dataset is also called the NIH CXR dataset and RSNA Pneumonia Detection Challenge Dataset.	×
7	[65]	SIRM COVID-19 Database	68 COVID-19 CXR images	✓
8	[19]	Radiopaedia.org	An open-source website that shares radiographs. As of the publication of this work, it contained 20 COVID-19 CXR images.	✓
9	[66]	ChexPert Dataset	224,316 chest CXR images of 65,240 patients labeled for 14 thoracic diseases.	×
10	[67]	Twitter COVID-19 CXR Dataset/ COVID-19 Chest Imaging at thread reader	135 CXR images showing SARS-CoV-2 viral infection. This dataset is provided by a cardiothoracic radiologist from Spain through his Twitter account.	✓
11	[68] taken from [69]	Pediatric Pneumonia Chest X-ray	5,856 pediatric CXR pneumonia images of children aged 1–5 years, labeled as “pneumonia” or “normal.”	×
12	[69] or [70] sometimes referenced as [23]	Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification/Kaggle Chest X-Ray Images (Pneumonia)	A dataset composed of OCT and 5,863 CXR images of normal, viral pneumonia, and bacterial pneumonia cases.	×
13	[71]	Open-I Repository	Open access biomedical image search engine provided by the US National Library of Medicine—CXR images of COVID-19 can be found with the relevant publication.	✓

TABLE 2. COVID-19 dataset sizes, formats, and download links.

No.	Reference	# Images	Resolution	Format	URL
1	[18]	315	4248 × 3480	JPG	https://github.com/ieee8023/covid-chestxray-dataset
2	[20]	48	various	JPG, PNG	https://github.com/agchung/Figure1-COVID-chestxray-dataset
3	[61]	N/A	various	PNG	https://github.com/agchung/Actualmed-COVID-chestxray-dataset
4	[62]	2,905			https://github.com/tawsifur/COVID-19-Chest-X-ray-Detection
5	[63]	105	4020 × 4892	N/A	http://db.jsrt.or.jp/eng.php
6	[22] sometimes referenced as [64]	112,120	1024 × 1024	DICOM PNG	https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data
7	[65]	68	various	JPEG	https://www.kaggle.com/nih-chest-xrays/data https://www.sirm.org/category/senza-categoria/COVID-19/
8	[19]	N/A	various	JPEG and others	Radiopaedia.org
9	[66]	224,316	320 × 320	N/A	https://stanfordmlgroup.github.io/competitions/chexpert/
10	[67]	135	2012 × 2012	JPG	https://twitter.com/ChestImaging
11	[68] taken from [69]	5,856	various	JPEG	https://www.kaggle.com/andrewmvd/pediatric-pneumonia-chest-xray
12	[69] or [70] sometimes referenced as [23]	5,863	various	JPEG	https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
13	[71]	N/A	various	various	https://openi.nlm.nih.gov/

“Montreal database collection” in some studies, although the original name is used in most of the reviewed research. This

dataset is composed of images from different sources, including Radiopaedia.org [19], the Italian Society of Medical and

Interventional Radiology (SIRM) COVID-19 Database [65], and Figure1.com [20]. Note that these resources were used in some studies with the COVID-19 Image Data Collection [18], creating the possibility of data duplication in the combined datasets.

The COVID-19 Radiography Database [21], which is the winner of the COVID-19 Dataset Award, is a dataset composed from six different sources, including the SIRM COVID-19 Database, COVID-19 Image Data Collection, Twitter COVID-19 CXR Dataset, RSNA Pneumonia Detection Challenge dataset [22], Kaggle CXR Images (Pneumonia) [23] and other CXR images from multiple published studies. Rahman *et al.* [21] stated that they only gathered images from published work and addressed redundancy by comparing the CXR images from different studies with those in the COVID-19 Image Data Collection to eliminate duplications. Other datasets, such as the datasets obtained from Peshmerga Hospital, Erbil, Kurdistan [24], were obtained from local hospitals and are not publicly available.

C. TRANSFER LEARNING

Transfer learning has been widely adopted in medical imaging applications [15], [25]. Transfer learning is beneficial in situations where the training examples are insufficient for training a model from scratch. Tajbakhsh *et al.* [25] demonstrated that a pre-trained CNN with adequate fine tuning might outperform or perform as well as a CNN trained from scratch. Consequently, and because of the limited training datasets, transfer learning has been actively explored for the detection of COVID-19 from CXR images.

In this survey, the reviewed works that utilized transfer learning can be categorized into three groups. In the first group, a pre-trained CNN on a large-scale natural image dataset was used to initialize the weights of a new network that will be trained on the target CXR data. For instance, models trained on ImageNet were used in [26], [27], and [28].

The second group consists of studies wherein some of the early layers of the pre-trained model on large-scale natural image dataset were frozen and their weights kept unchanged while the final layers were finetuned [29]. This practice is based on the fact that the early layer features are more generic (e.g., edges), whereas the later-layer features are more specific to a particular task or dataset [17]. Examples of works that implemented the finetuning approach for the radiological photographs can be found in [26], [27], [30], and [31].

In the third group of studies, transfer learning was implemented using a model pre-trained on a similar target domain; for example, Afshar *et al.* [31] trained a model on a radiography dataset of patients with and without pneumonia. They then trained the model further on COVID-19 CXR images. The studies in this group claimed that the use of models trained on ImageNet is not the best option for medical applications because the source (natural images) and target domains (e.g., CXR images) are different [30], [31]. However, the results of a comparative study by Cheplygina [29]

did not fully support this assumption; the study examined 12 articles that compared the use of medical images to natural images in transfer learning in medical imaging research. The goal of the study was to determine which source images are better in medical transfer learning tasks: natural images such as ImageNet or medical images. Among the 12 articles examined, the study found that six articles supported each claim, i.e., each claim is supported equally; therefore, the study concluded that the selection of the model and source data depends on the task at hand among other factors.

D. CNN ARCHITECTURES

In recent years, CNN architectures [9] have managed to achieve human expert-level performance in a wide range of complex visual tasks, including medical image assessment and pathology detection. Numerous CNN architectures have been proposed in the literature since the very first successful CNN in 1998. Known as LeNet and developed by Yann LeCun, it was widely used for handwritten digit recognition [32]. Compared to current models, LeNet is considered to be a shallow architecture; it contains three convolutional, two average pooling, and two fully connected layers. In the following subsections, we briefly describe the CNN architectures used in the reviewed studies along with their usage and results for COVID-19 detection from CXR images.

1) AlexNet

AlexNet [33] is similar to LeNet, but it is deeper and contains three stacked convolutional layers. AlexNet won the 2012 ILSVRC challenge and achieved a top-five error rate of 17%. To overcome the overfitting problem, the authors used a dropout regularization technique and data augmentation in AlexNet.

Razzak *et al.* [34] used AlexNet for the binary and multiclassification of COVID-19 cases. They achieved a test set accuracy of 97.04% for COVID-19/healthy binary classification and 63.27% for COVID-19/healthy/bacterial pneumonia/viral pneumonia multiclassification tasks.

Kumar and Kumari [35] used AlexNet as a feature extractor to feed a support vector machine (SVM) classifier and achieved an accuracy of 93.0%. Abbas *et al.* [36] also used AlexNet for the feature extraction of three classes, namely, normal, COVID-19, and SARs. However, in their work, they proposed that a class decomposition layer should be added to partition each class into multiple sub-classes. These subsets were reassembled to produce the final predictions. They used AlexNet to find features for the proposed decomposition layer and achieved an accuracy of 95.12% for their proposed model DeTraC.

2) GoogleNet

GoogleNet [37] won the 2014 ILSVRC challenge and achieved a top-five error rate of 6.67%. This network is significantly deeper than the previous CNNs; in addition to the pooling and convolutional layers, GoogleNet contains an inception module (IM). This module acts as a small network

and can learn cross-channel correlations (depth-wise) along with spatial correlations. It consists of six convolutional layers: four 1×1 convolutional layers, one max pooling layer, and one concatenation layer [37]. The IM serves as a bottleneck layer and induces several advantages. First, it enables the training of significantly deeper models while reducing the number of learnable parameters by nearly ten times. Second, the output of an IM is configured to be smaller than its input in terms of the number of feature maps. Thus, the IM reduces the dimensionality. Third, an IM can capture complex patterns at multiple scales along with the spatial and depth dimensions. Other variants of GoogleNet have been proposed using slightly different inception components, and they have achieved better performances. Examples include Inception-V2, Inception-V3 [38], Inception-ResNet, and Inception-v4 [39].

Razzak *et al.* [34] used GoogleNet in a similar manner as AlexNet, i.e., for the binary and multiclassification of COVID-19 cases, and the test set accuracy improved to 98.15% for COVID-19/healthy binary classification and 75.51% for COVID-19/healthy/bacterial-pneumonia/viral-pneumonia multiclassification tasks.

Similarly, Kumar and Kumari [35] used GoogleNet as AlexNet for feature extraction and achieved an accuracy of 93% using an SVM classifier.

3) VGGNet

VGGNet [40] was proposed by the Visual Geometry Group (VGG) at Oxford University and was the runner up of the 2014 ILSVRC challenge; it achieved a top-five error rate of 7.3%. With a total of 16 or 19 convolutional layers, VGGNet has the advantage of architectural simplicity [40]. However, it used three times more parameters than AlexNet [38].

Moutounet-Cartan () [41] evaluated five different CNN architectures followed by a flat multi-layer perceptron. They found that VGG16 yielded the best test accuracy of 70.6% for detecting COVID-19 cases over three classes of COVID-19/no findings/other pneumonia, followed by VGG19 with an accuracy of 70.2%; meanwhile, the InceptionResNetV2 and InceptionV3 architectures yielded significantly lower accuracies of 45.7% and 47.7%, respectively. Rahaman [42] also found that VGG19 achieved the highest testing accuracy of 89.3% compared to the 14 other deep CNN architectures.

Kumar and Kumari [35] also used VGG16 and VGG19 to extract the features of COVID-19 to feed the SVM with the final accuracies of 92.7% and 92.9%, respectively.

4) ResNet

ResNet [43] introduced a residual learning component to the CNN architecture. The residual unit (RU) consists of a regular layer with a skip connection. The skip connection allows the input signal of a layer to traverse the network by connecting it to the output of that layer. Thus, the RUs enabled the training of an extremely deep model of 152 layers, which won the 2015 ILSVRC challenge and achieved a top-five error rate

of under 3.6%. Other variants of ResNet have 34, 50, and 101 layers.

As shown in Fig. 4, ResNet was the most widely utilized CNN architecture in the reviewed papers. Minaee *et al.* [28] applied ResNet18 and ResNet50 on an imbalanced dataset of 100 COVID-19 images and 3000 non-COVID images; they achieved a sensitivity of 98% for both architectures. ResNet50 yielded an accuracy of 89.2% for detecting COVID-19 from CXR in [44]; Kumar and Kumari [35] used three variants of ResNet with an SVM classifier including ResNet18, which provided an accuracy of 91%. ResNet50 provided an accuracy of 95% and ResNet101 provided an accuracy of 89.2%.

5) XCEPTION

Xception [45] was proposed by Francois Chollet, and it outperformed Inception-V3 on a huge image classification task comprising 350 million images and 17,000 classes. In contrast to the IM of GoogleNet, the main concept of Xception involves learning the cross-channel and spatial patterns separately. Thus, in Xception, the IM is replaced with a special layer called depth-wise separable convolution. Compared with a traditional convolutional layer, the separable convolution layer has fewer parameters, lower computational cost, and an overall better accuracy of 79% in comparison to Inception-V3 (78.2% accuracy) on the ImageNet dataset [45].

El-Din Hemdan [46] used Xception to diagnose COVID-19 in normal patients. Xception obtained the highest precision among the deep learning classifiers for detecting only positive COVID-19 cases; however, it exhibited significantly worse performance in classifying the normal cases. In [35], Xception yielded a precision and sensitivity of 0.830 and 0.894 in identifying COVID-19 patients, respectively.

6) SENet

The Squeeze and Excitation Network (SENet), proposed by Hu *et al.* [47], was the winner of the 2017 ILSVRC Challenge with a top-five error rate of 2.251%. SENet extends the GoogleNet IM and ResNet architectures and boosts their performances. SENet introduced a new component, namely, the SE block. This block has been added to every IM or RU in ResNet. The extended versions are called SE-Inception and SE-ResNet, respectively. The SE block consists of three simple layers, a layer for global average pooling across the feature maps, a squeeze dense layer, and a dense layer. The output of the SE block recalibrates the feature maps generated by the IM or RU in a way that downscales the irrelevant feature maps and retains the relevant features. Note that the SE block focuses on the cross-channel patterns, instead of the spatial patterns, and learns the features that are active together. In other terms, it learns the objects in the images that usually appear together.

Razzak *et al.* [34] used SENet for COVID-19 detection and achieved accuracies of 98.89% and 94.39% for binary and multiclassification tasks, respectively.

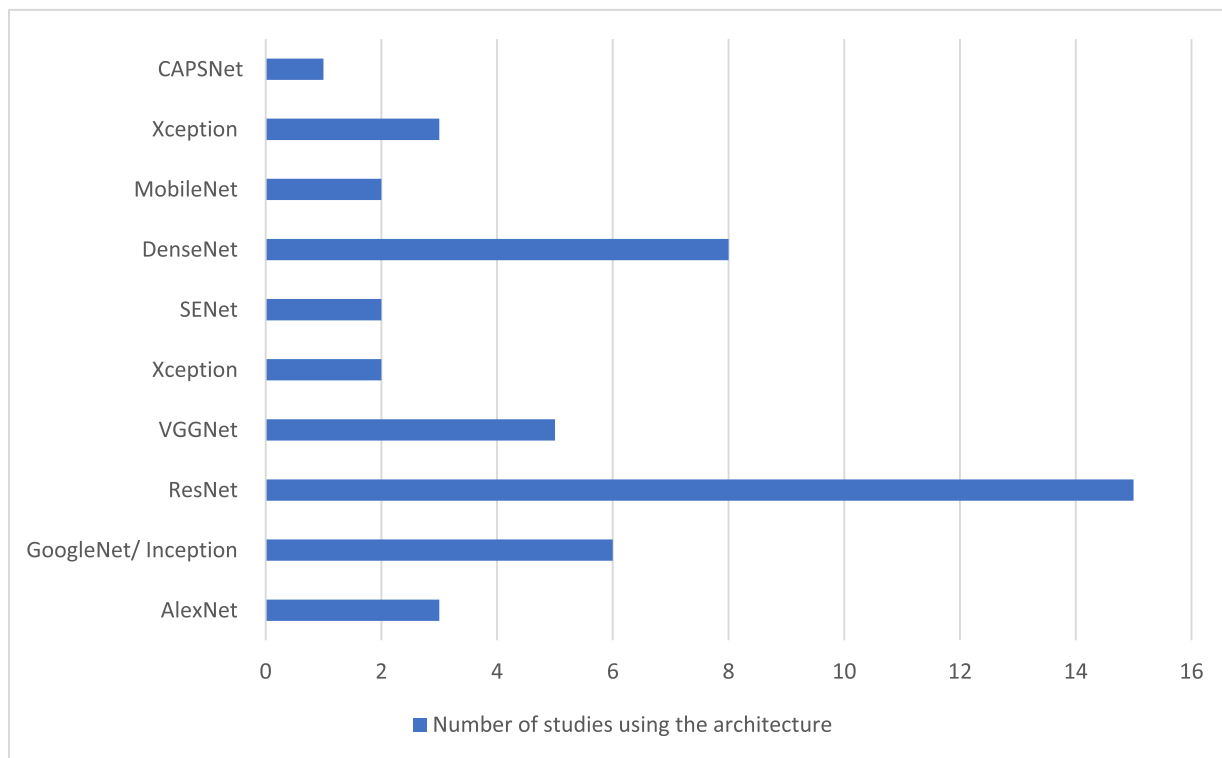


FIGURE 4. CNN architectures used in the reviewed work.

Minaee *et al.* [28] used SENet as a ResNet variant on a highly imbalanced dataset and achieved similar sensitivity of 98%, but improved specificity of 92.9% in comparison to 90.7% and 89.6% of the ResNet models, respectively.

7) DenseNet

DenseNet [48] connects each layer to every other forward layer in the network. Therefore, instead of L connections between L layers in the regular CNN architecture, DenseNet contains $L(L+1)/2$ layer connections. All subsequent layers use the feature maps generated by any layer in the network, which enables all layers to reuse and propagate features, including the final ones. On ImageNet, DenseNet achieved a top-five error rate of 6.12%; however, it requires fewer parameters and less computational cost than other state-of-the-art CNN architectures, such as ResNet.

As shown in Fig. 4, DenseNet with its variants is the second most used architecture in the reviewed studies. Razzak *et al.* [34] used DenseNet and achieved accuracies of 98.75% and 93.46% for binary and multiclassification, respectively. In [28], the application of DenseNet resulted in similar sensitivity as the aforementioned architectures, but a lower specificity of 75.1%. Kumar and Kumari [35] and El-Din Hemdan *et al.* [46] used DenseNet201 for COVID-19 detection and achieved accuracies of 93.8% and 90%, respectively.

8) MobileNet

MobileNet [49] is a lightweight CNN designed for mobile and embedded vision systems. MobileNet utilizes depth-wise

separable convolution to generate a lighter architecture and introduces two trade-off hyperparameters to balance the latency and accuracy. MobileNet has been shown to perform well across a wide range of applications [49].

MobileNet was used for COVID-19 detection from the CXR images in [34] and [46], achieving accuracies of 96.30% and 60%, respectively.

9) ShuffleNet

Other advanced CNN architectures include ShuffleNet [50], which outperformed MobileNet on the ImageNet classification task. Compared with AlexNet, ShuffleNet achieved a speedup of 13 times with comparable accuracy. To reduce the computational cost, ShuffleNet introduces channel shuffle and pointwise group convolution operations.

From the reviewed studies, Alqudah *et al.* [51] used ShuffleNet for the automatic extraction of features, which were then fed to four different classifiers: Random Forest, Softmax, SVM, and KNN. The accuracies achieved by these classifiers with the ShuffleNet features were 80%, 99.35%, 95.81, and 99.35%, respectively.

E. OTHER DL ARCHITECTURES

In this survey, we found two additional DL architectures that are not based on the basic CNN components but were also suggested for COVID-19 detection, namely, autoencoder and Capsule Network (CapsNet).

An autoencoder [52] is a type of unsupervised neural network. The features learned by autoencoders, also called

codings, are a compressed representation of the input image. This makes the autoencoders useful for feature extraction, dimensionality reduction, and pre-training supervised models.

Autoencoders were used by Khobahi *et al.* [53] for the detection of COVID-19 and achieved an accuracy of 93.50%.

CapsNet was first introduced by Hinton *et al.* [54]. It contains a special arrangement of neurons, called a capsule; a traditional neuron outputs a scalar and a capsule output a vector. The length of the vector represents the existence of an object in the image and the vector orientation represents the object properties [55]. CapsNet has been demonstrated to be a powerful and promising tool for managing the variations in the orientation, position, and lighting.

CapsNet was used by [31] for identifying the COVID-19 cases. By pre-training with a dataset of X-ray images, CapsNet in [31] provided an accuracy of 98.3%.

III. METHODOLOGY ANALYSIS

As discussed in previous sections, various DL architectures (CNNs in particular) have been proposed for the detection of COVID-19 from CXR images in a relatively short time. In this section, we provide in-depth insights about the reviewed studies.

A. DATASETS

Table 3 details some well-known CNN architectures applied for COVID-19 classification along with the datasets used. All reviewed studies used publicly available datasets, except the studies of Iqbal Khan *et al.* [90] and Gomes *et al.* [24].

Most of the studies combined datasets to enlarge the training set. However, due to the limited number of available COVID-19 samples, an imbalance problem occurs, which is one of the major challenges. The table also illustrates that most of the studies employed a number of different architectures to compare the classification result or to build an ensemble model to achieve better performance.

The dataset collected by Cohen *et al.* [18] was used in more than 85% of the articles surveyed here. The collection of CXR images in the dataset of Cohen *et al.* was obtained from online publications rather than from original medical sources, which may have reduced the image quality and led to undesirable learning models. This dataset also does not provide lesion or infected area annotation. Visual annotation would help to get more insights regarding the reasons of the prediction decisions made by human experts and would help in comparisons with deep learning architectures.

The disagreement between human annotators should be also provided to allow better model evaluation.

B. DEEP LEARNING MODELS CONSTRUCTING

The review indicates that transfer learning was preferred by most researchers, and broad interest in this approach continues. DL, particularly transfer learning, enables rapid model development while outperforming other approaches. ResNet [43], DenseNet [48], Inception [38], and VGG [40]

are among the most utilized pre-trained architectures. A properly trained transfer learning model will usually outperform a model trained from scratch. The smaller the dataset, the better the performance. However, as the number of labeled COVID-19 CXR images is currently limited, common pre-trained models such VGG or ResNet, with millions of parameters, can easily overfit the training data. Thus, particular attention should be paid towards choosing the appropriate metrics for evaluation and for selecting appropriate and representative testing data.

Some studies investigated more than one version of the same base CNN, VGG16, and VGG19. As shown in Fig. 4, most authors employed a ResNet followed by DenseNet. In the implementation of transfer learning, a new model requires a pre-trained network chosen from among the widely adopted networks that are trained on the ImageNet dataset as a starting point. Although most studies exploited architectures trained on ImageNet, Duchesne *et al.* [58] and Bassi and Attux [59] applied transfer learning using ChexNet [60]. ChexNet is a 121-layer dense CNN model (DenseNet) trained on the ChestX-ray14 dataset [22], which contains 112,120 frontal-view CXR images labeled with 14 different thoracic diseases, including pneumonia. In ChexNet, the final DenseNet fully connected layer is replaced with a fully connected layer whose output is produced using a nonlinear sigmoid function. The weights of the network were initialized with weights from a model pre-trained on ImageNet.

Table 4 summarizes the contributions and novelty of some articles reviewed in this survey along with the datasets they used for evaluation. The approaches applied varied from building frameworks and models from scratch to exploiting transfer learning along with some advanced feature extraction methods. For example, in [53], autoencoders were used for feature extraction. In [26], the authors combined anomaly detection scores with classification scores in the last layer of the network. The anomaly detection component was used to generate a large anomaly score of CXR images with COVID-19. The authors demonstrated that this hybrid approach outperforms other individual task learning models. The authors of [61] proposed to automatically generate a new deep architecture named COVID-Net, which has been tailored particularly for COVID-19 CXR image classification. COVID-Net is open source along with its COVIDx dataset, comprising 13,975 CXR images. The main advantage of this approach is that the architectural design choices made by generative synthesis can achieve a balance between multiple objectives such as performance and computational cost. This approach can be further investigated and applied to other medical image classification tasks by specifying the requirements such as the desired sensitivity and specificity.

C. PERFORMANCE COMPARISONS

It was difficult to compare the studies included in this survey owing to the variations in the size of the testing sets and the lack of standard performance measures, which further complicated the identification of the most efficient DL models

TABLE 3. Well-known CNN architectures used for the detection of COVID-19.

Research study	Datasets	CNN architecture(s) employed
[72]	[18], [65], [22]	InceptionV3
Deep-COVID [28]	[18], [66]	ResNet18, ResNet50, SqueezeNet, DenseNet121
[73]	[18], [23]	Bayesian ResNet50V2
[74]	[18], [22], [66], [68], [66], [64]	VGG 16, InceptionV3, Xception, DenseNet121, NASNet-mobile
[31]	[18], [23]	CapsNet
[26]	[18], [22]	ResNet18
DeTraC [36]	[18], [63]	AlexNet, ResNet18
[75]	[18], [64]	ResNet50V2, Xception
[51]	[18], [64], [65], [19], [23]	AOCT-Net, MobileNet, ShuffleNet
[34]	[18] [23]	AlexNet, SqueezeNet, GoogLeNet, VGG, MobileNet, ResNet18, ResNet50, ResNet101, DenseNet
[53]	[18], [64]	Autoencoders, ResNet18
[76]	[18], [65], [23]	DenseNet161
[57]	[23]	ResNet152
[41]	[18], [20]	VGG16, VGG19, InceptionResNetV2, InceptionV3, Xception
[6]	[18], [69]	RestNet50
[77]	Locally collected from Hospital San Gerardo, Monza, Italy, and IRCCS Policlinico San Donato, Italy	RestNet10
[46]	[18]	VGG19, DenseNet121, InceptionV3, ResNetV2, Inception, ResNet-V2, Xception, MobileNetV2
[35]	[18], (Winner of the COVID-19 Dataset Award) [21]	AlexNet, DenseNet201, GoogleNet, InceptionV3, ResNet18, ResNet50, ResNet101, VGG16, VGG19, XceptionNet, InceptionResNetV2
[44]	[18], [65], [69]	ResNet50, VGG16
[27]	[18], [23]	ResNet50, InceptionV3, InceptionResNetV2
[79]	[20]	SqueezeNet
[80]	[66]	DenseNet121
[42]	[20]	DenseNet121
[42]	[18][23]	VGG16, VGG19, ResNet50, ResNet50V, ResNet101, ResNet101V2, ResNet152, ResNet152V2, DenseNet121, DenseNet169, DenseNet201, MobileNetV1, XceptionNet, InceptionV3, InceptionResNetV2
[82]	[18]	CNN

for detecting COVID-19 from CXR images. Most authors evaluated the DL models in terms of accuracy, sensitivity, and specificity metrics. However, the difficulty of comparing different approaches increases when non-standard metrics and datasets from multiple sources are used. Thus, it is essential to develop a public COVID-19 dataset that is comprehensive and accessible by the AI research community. In addition, standards for evaluating the performance of prediction models must be established.

Table 5 shows the results of the reviewed articles in terms of the classification metrics such as accuracy, precisions, recall, AUC, and F1-score. The reader can refer to the glossary appendix for the definition of these terms.

It can be observed from the table that most of the models achieved high accuracies; however, as demonstrated in Table 4, most of these results were obtained over a limited number of COVID-19 samples. Thus, the results are not representative unless the study considered the class

TABLE 4. Articles proposing novel methods for COVID-19 detection via CXR images.

Research study	Dataset(s)	Contribution(s)
[30]	[18], [65], [19]	The authors built a model from scratch trained on more than 100,000 CXR images and used transfer learning to overcome the problem of limited number of COVID-19 samples.
[26]	[18], [22]	Anomaly detection was used in the last layer to classify COVID-19 CXR images. This layer generates a scalar anomaly score that assigns statistically significantly large classification scores and anomaly scores to CXR images with COVID-19.
COVID-Net [61]	[18], [22], (Winner of the COVID-19 Dataset Award) [21]	The authors used GenSynth to generate a deep CNN tailored for COVID-19 by specifying the design requirements, thereby obtaining a sensitivity of 80% and PPV of 80%.
DeTraC [36]	[18], [63]	Original classes were decomposed into several sub-classes using K-nearest neighbors; transfer learning (ImageNet pre-trained ResNet) was applied, and the final classification was refined using the error-correction criteria.
CheXNet [59]	[18], [65], [23], (Winner of the COVID-19 Dataset Award) [21]	CheXNet was trained with transfer learning twice.
COVID-DA [83]	[18], [22], [21]	A domain adaptation method was proposed to transfer the domain knowledge from the well-labeled source domain (pneumonia) to the partially labeled target domain (COVID-19). Domain discrepancy was minimized via domain adversarial learning and the task difference between domains was eliminated by proposing a novel classifier separation scheme.
CoroNet [53]	[18], [22]	Autoencoders were used for feature extraction.
[57]	[23]	SMOTE was used to handle the class imbalance problem.
[9]	[18]	DarkNet is an end-to-end architecture proposed as a classifier for YOLO real-time object detection. The authors introduced different filters on each layer with an implementation of 17 convolutional layers.
[11]	[18]	Multi-level segmentation thresholding was applied to split grayscale images into areas of intensities to reduce the number of objects in a lung image. An SVM was then used to classify the COVID-19 cases. The MATLAB 2019a DL toolbox was used.
[44]	[18], [65], [69]	The proposed model tuned both the VGG16 and ResNet50 models with a balanced set of CXR images by replacing the last three layers with the trainable part, followed by a 64-unit connected layer with dropout, 10-fold validation, and a classification layer with a sigmoid output.
[84]	[18], [69]	The authors built a framework called discriminative cost-sensitive learning. They stated that this technique should be adopted in clinical settings, such as COVID-19 diagnosis, because it offers two features, i.e., fine-grained classification and cost-sensitive learning.
[80]	[61]	Two types of COVID-19 infections were distinguished using CXR images because the treatment for each type differs significantly. Additionally, this study considered the progression of the disease.
[81]	Locally collected	Proposed a model called Cascade-SENet, which is based on the SENet architecture [47]. Additionally, the study implemented U-Net [98] and introduced adaptive histogram equalization with limited contrast to preprocess and enhance the quality of the images.

imbalance problem in the test set. Fig. 5 illustrates the results of the studies that presented the most common metrics, i.e., accuracy, specificity, and sensitivity. As evident, Alqudah *et al.* [56] achieved the highest performance with respect to all measures. The methods proposed by Razzak *et al.* [34] and Kumar *et al.* [57] also achieved comparable results for all metrics; as demonstrated in Table 5, these two studies dealt with class imbalance by including

an equal number of samples for each class and employing the synthetic minority oversampling (SMOTE) technique, respectively.

IV. DISCUSSIONS

Despite the encouraging results of the DL architectures, there exist several issues that must be addressed to ensure that the diagnostic process is more accurate, transparent,

TABLE 5. Performance metrics of the methods used in the reviewed research.

Research study	Accuracy (%)	Precision / PPV (%)	Specificity (%)	Sensitivity/ Recall/TPR (%)	AUC	F1-Score
CovidAID [90]		98.60		1	0.99	
[72]	96.90					
[30]	95.30					
Deep-COVID [28]			90.00	97.00		
[73]	89.82					
[74]	93.00	93.15	88.52	97.53	0.95	94.57
COVID-CAPS [31]	95.70		95.80	90.00	0.97	
[26]			87.84	90.00	0.95	
COVID-Net [60]	93.30	98.90		91.00		
DeTraC [36]	95.12		91.87	97.91		
[75]	91.40	35.27	99.56	80.53		
[56]	99.46	99.46	99.73	99.46		99.46
CheXNet [59]	97.80	97.80		97.80		97.80
COVID-DA [62]		98.15		88.33	0.98	92.98
[34]	97.20	97.67	98.51	97.50		97.50
CoroNet [53]	93.50	93.63		93.50		93.51
[76]		99.33		99.33		
[57]	97.70		98.80	97.70	0.98	97.70
[41]	93.90			88.00	0.97	
[8]						
[5]	99.00	99.00		99.80		99.80
[92]		82.00	82.00	78.00	0.89	
[46]	90.00	83.00		1		91.00
[35]	95.38					95.52
[44]	91.24					
[79]	98.26		99.13			
[91]	95.00	97.00		100.0		
[58]	89.78	89.85	99.63	89.79		
[80]	82.70	92.70	62.50			
[81]				80.00		
[82]	97.00					
[42]	89.3	90.00		89.00		90.00

and trustworthy. In this section, we emphasize the current research challenges associated with the detection of COVID-19 from CXR images.

A. CLASS IMBALANCE PROBLEM

COVID-19 datasets encounter the problem of class imbalance. The uneven distribution of classes raises concerns related to the robustness of the machine learning algorithm. Some studies, such as the study of Kumar *et al.* [57], proposed the use of SMOTE to mitigate this problem. Another approach suggested by Ucar and Korkmaz [78] and Rajaraman and Antani [74] involved the implementation

of data augmentation to enlarge the number of images obtained from the limited number of COVID-19 cases. Multiple datasets consist of images containing unrelated visual features and misleading artifacts, which are not usually addressed by the studies covered in this review.

Table 6 lists the number of samples used for the training and testing sets by the studies included in this survey. As shown, the size of the datasets varies widely. COVID-19 cases range from 11 to 1,536 cases, whereas the total sample sizes range from 50 to 224,316 cases.

As shown, the size of the datasets varies widely. Moreover, the COVID-19 cases are limited and significantly small in

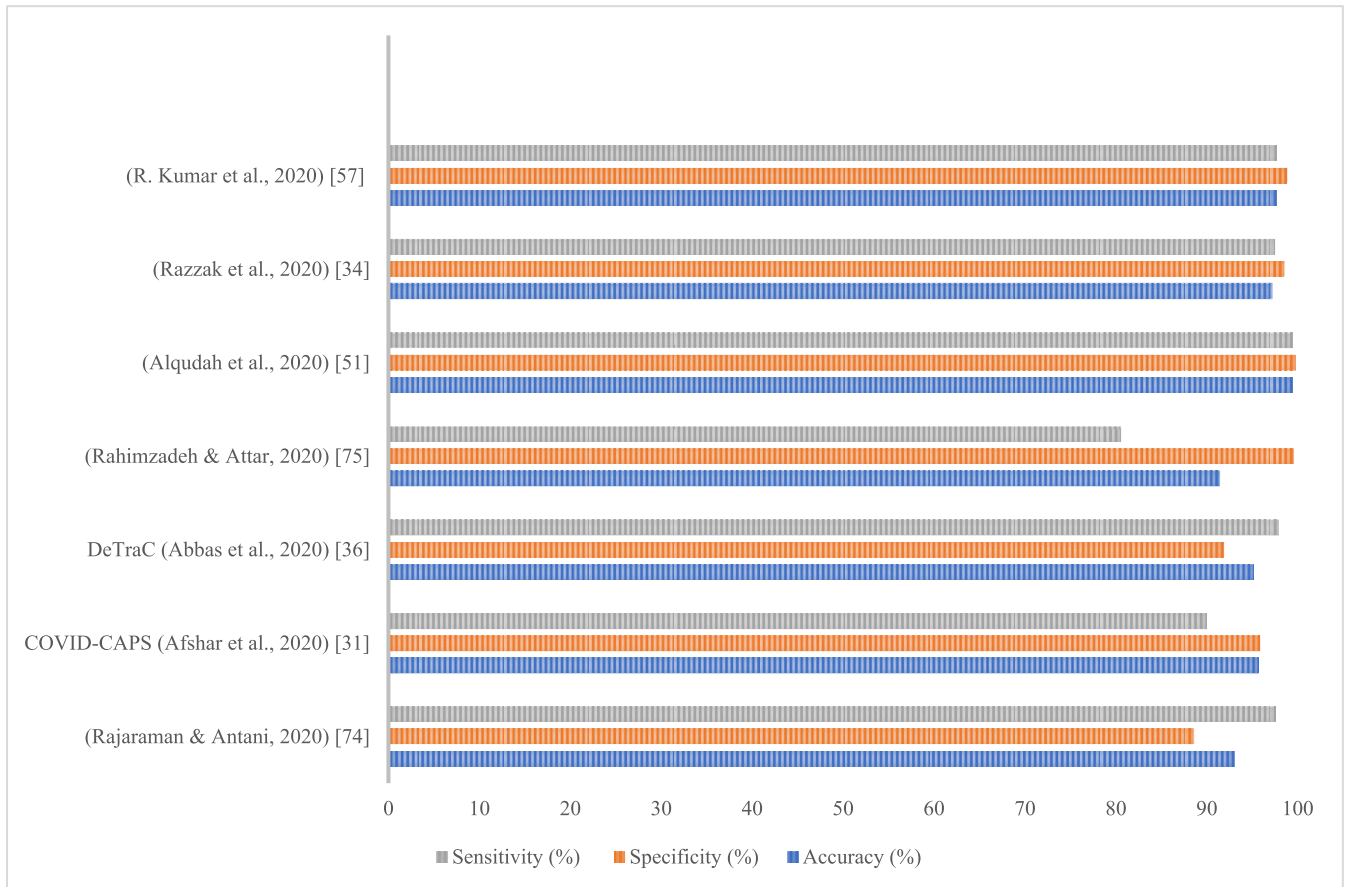


FIGURE 5. Performance comparison of the results of the reviewed research.

number compared to the total number of samples. Thus, several studies considered the class imbalance problem by employing different techniques. Alqudah *et al.* [56] handled this problem by keeping the number of samples in each class equal to 310. Other studies such as [59] and [61] utilized data augmentation. Razzak *et al.* [34], Castiglioni *et al.* [77], El-Din Hemdan *et al.* [46], Hall *et al.* [44], and De Moura *et al.* [76] also considered using a fixed number of samples for each class. Khobahi *et al.* [53] and Duchesne *et al.* [58] used a class-weighted entropy loss function. Kumar *et al.* [57] used the SMOTE technique; Medhi and Hussain [81] and Han *et al.* [83] employed cost-sensitive learning.

B. EXPLAINING DEEP MODEL PREDICTIONS

In most studies covered in this review, DL architectures are used as black-box classifiers, and an explanation of the model decisions is lacking. Explainable AI is an emerging AI subfield that refers to the techniques and methods used to understand the paths taken by machine learning models for decision making. The GSInquire approach [92] was used by Wang and Wong [61] to highlight the areas used by the DL classifier to drive predictions. None of the studies considered defining a region of interest for detecting the symptoms or infections related to COVID-19. However, deep

neural network architectures contain numerous optimization parameters; therefore, they heavily rely on large annotated datasets to avoid overfitting.

With the increased adoption of the DL models, the demand for explaining how these DL models make decisions has also increased [92]. The lack of transparency and interpretability in DL models hinders their adoption, especially in situations where transparency is crucial, such as in CXR diagnostic scenarios [92]. In the reviewed literature, 11 out of 26 studies (42%) applied DL visualization techniques; Fig. 6 illustrates the number of studies that utilized explanatory techniques. Table 7 summarizes the techniques used and studies that implemented them. As seen in Table 7, the most used method is GRAD-CAM, and next most used is the CAM method. Table 8 shows examples extracted from the reviewed studies showing the explanatory methods used in each work. As illustrated, the GRAD-CAM and CAM methods work in a similar manner, using heat maps, while the other methods each highlight the affected area differently. In the next subsections, we briefly explain each of these techniques.

1) GRAD-CAM

Grad-CAM is the most widely used technique by the reviewed studies, as illustrated in Fig. 6; it was implemented in [30], [73], [74], [26], [82], and [80]. Grad-CAM is a visual

TABLE 6. Sizes of the datasets used in the reviewed research.

Research study	Total samples		Training samples		Testing samples		Method for handling class imbalance
	ALL	COVID-19	ALL	COVID-19	ALL	COVID-19	
CovidAID [90]	3,969	106	3,516	80	424	19	Fixed ratio of classes in each batch
[72]	3,550	864					
[30]	109,895	225					
Deep-COVID [28]	5,071	71	2,031	31	3,040	40	
[73]	5,941	68	4,753		1,188		
[74]	37,220	314	3,883		11,706	314	
COVID-CAPS [31]	94,638	315		284		31	Modification of the loss function
[26]	1,531	100	764	50	767	50	
COVID-Net [60]	13,975	358					Batch re-balancing
DeTraC [36]	196	11	138		58		
[75]	15,085	180	3,783	149	11,302	31	Fixed ratio of classes in each training batch: # samples = 633 Equal # samples for each class = 310
[56]	930	310					
CheXNet [59]	2,339	187	2,159	127	180	60	Data augmentation
COVID-DA [62]	11,663	318	10,718	258	945	60	Focal loss
[34]	800	240	800	200	160	40	Equal # samples for each class = 200
CoroNet [53]	18,529	99	16,576	89	1,953	10	Class-weighted entropy loss function
[76]	621	207	366	125	255	82	Equal # samples for each class = 2017
[57]	8,588	62					SMOTE
[41]	327	129					
[8]		127		125			
[5]	9,672	161	7,254		2,418		Data augmentation
[92]	610	324	500	250	110	74	Equal # samples for each class
[46]	50	25	50	20	50	5	Equal # samples for each class
[35]	316	158	0.6		0.2		
[44]	455	135	204	102	251	33	Equal # samples for each class = 102
[79]	5,949	1,536		153		153	Data augmentation
[91]	1,493	284		228		56	
[58]	6,320	464					Class-weighted entropy loss function
[80]	224,316						
[81]	16,130			10		10	
[82]	1,141						New conditional loss, learned with joint balance optimization and cost-sensitive learning
[42]	860	260	580	180	140	40	

explanation technique that distinguishes between classes in the visualization task [93], and it offers several features that might have encouraged the reviewed studies to use it. First, Grad-CAM does not require any changes in the architecture, unlike other techniques such as CAM, which alter the architecture of the CNN [93]. Second, Grad-CAM is highly class-discriminative, i.e., it not only highlights the regions

attended to by the classifier during classification but also differentiates between the classes.

2) CAM

CAM is a visualization technique that replaces the fully connected layers with the global average pooling and convolutional layers to create class-specific feature maps

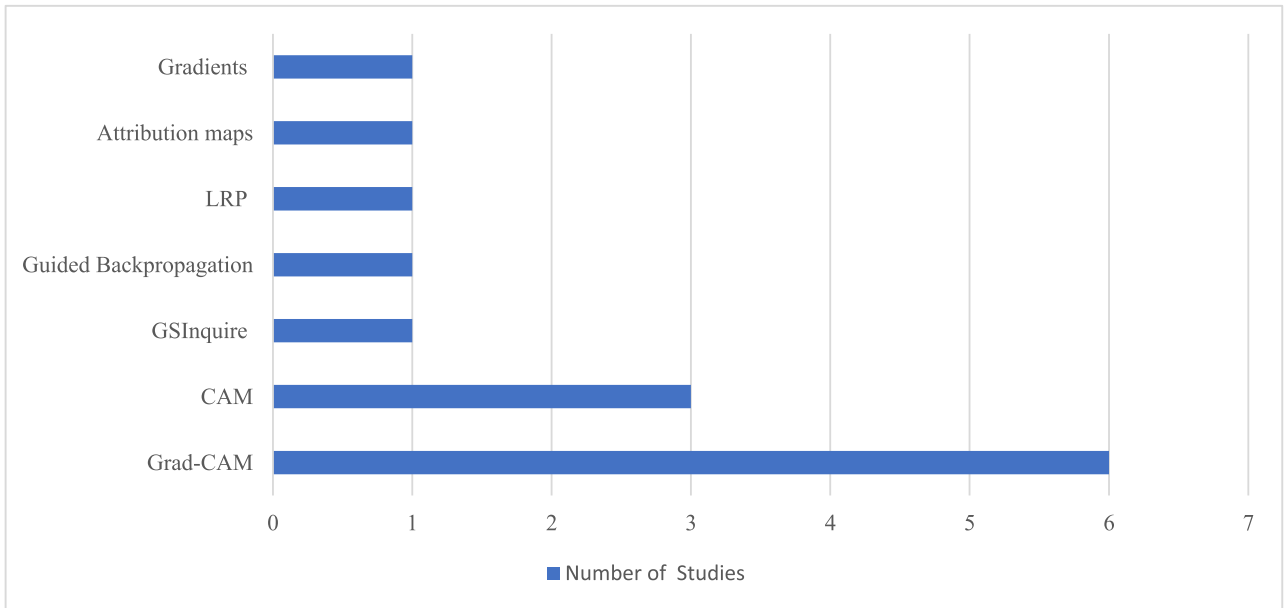


FIGURE 6. Number of studies utilizing explanatory techniques.

TABLE 7. Methods used for explaining classification decisions.

Visualization technique	Studies utilizing the technique
Grad-CAM	[30], [73], [74], [83], [81]
CAM	[73], [51], [79]
GSInquire	[61]
Guided backpropagation	[73]
LRP	[59]
Attribution maps	[53]
Gradients	[73]

because the global average pooling layer enables localization in CNNs [70]. CAM was employed in three studies— [73], [51], and [78]—as shown in Fig. 6.

3) GSInquire

This technique is based on generative synthesis, which is an algorithm that can produce generators that create deep neural networks automatically [92]. This method was adopted by Wang and Wong [61].

4) GUIDED BACKPROPAGATION

In guided backpropagation, the process of finding the discriminative part in an image starts with a high-level feature map. Next, the algorithm reverses the data flow of the network, starting from neuron activation in a given layer back to the image. Consequently, the created image highlights the part of the input image that is responsible for activating the neuron [96]. Guided backpropagation was employed by Ghoshal and Tucker [73].

5) LRP

LRP employs the network weights and the activations created by forward propagation to propagate the output back

to the input layer. Thus, pixels that contribute to the output can be visualized [97]. This method was used by Bassi and Attux [59].

6) ATTRIBUTION MAPS

Attribution maps, used by Khobahi *et al.* [53], are heatmaps that show areas in the input image that are responsible for the classification output. To construct an attribution map, a generative model removes parts of the image, where the attribution value of an input area is approximated by the changes in the classification probability [98].


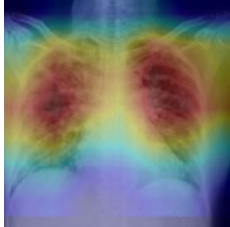



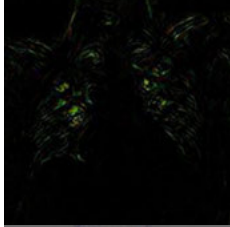

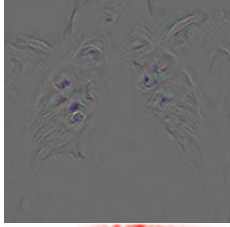

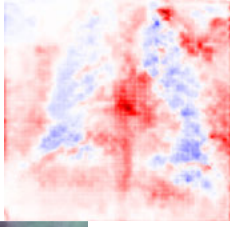
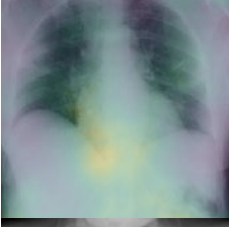

7) GRADIENTS

Gradient is a technique used to visualize a deep convolutional network trained using ImageNet [99]. This method was implemented by Ghoshal and Tucker [73]. A gradient finds the gradient of the class score for the input image and uses backpropagation to represent the gradient as a heatmap.

C. MANAGING CLASSIFICATION UNCERTAINTY

Uncertainty in DL represents the level of confidence in the result obtained by the classifier [73]. Obtaining a high softmax output does not imply high certainty, even though the softmax output can be mistakenly confused with model confidence [100] [101]. A DL model that considers uncertainty enables uncertain cases to be treated with caution. Thus, when a model generates a result with high uncertainty, this suggests that human intervention is recommended to examine the result further [78]. Ghoshal and Tucker [73] stated that the estimation of uncertainty in AI solutions can increase their adoption in clinical settings because it provides a more transparent solution that can be trusted by physicians. In the reviewed articles, we found that the study in [73] dealt with model uncertainty. In this research, drop weights and

TABLE 8. Examples of explanatory techniques used in the reviewed studies.

Visualization Technique	COVID-19 Case Before Highlighting	COVID-19 Case After Highlighting	Source
Grad-CAM			[30]
CAM			[79]
Gradients			[73]
Guided Backpropagation			[73]
LRP			[59]
Attribution maps			[53]
GSInquire			[61]

Bayesian CNNs were implemented to compute the uncertainty. The study also implemented transfer learning using a pre-trained ResNet50V2 model finetuned on COVID-19 data. The model generated the output certainty as low or high confidence based on the input CXR image, and the authors illustrated how the confidence level could affect the decision process when using real COVID-19 radiograph examples. Their accuracy ranged between 86.02% and 89.82%. The model showed a relatively high correlation between model uncertainty and prediction accuracy.

According to this review, more studies are required to investigate the uncertainty in the predictions made by a model, which define the level of confidence in the results produced by the model. In classification problems, the data samples can be close to a threshold or decision boundaries, which reduces the confidence accuracy of the classifier regarding the final decision. However, this is not usually discussed in the DL medical imaging literature and was only discussed by Ghoshal and Tucker [73] in the studies included in this review.

D. COVID-19 SEVERITY ASSESSMENT

Other problems that remain to be tackled in the COVID-19 CXR imaging literature include disease progression assessment and prognosis analysis. CXR imaging analysis could also help in identifying high-risk patients and the areas that urgently require attention and support. These issues and problems require more involvement of medical personnel at all stages of DL model development, evaluation, and validation.

Triage is an important stage during the COVID-19 pandemic owing to the growing number of patients who require rapid and accurate intensive care and resources. DL studies that aim to predict, track, and assess the progress and severity of COVID-19 patients help in efficiently triaging patients. Duchesne *et al.* [58] and Islam and Fleischer [79] considered tracking the progress of COVID-19 patients in their studies. Duchesne *et al.* [58] monitored and predicted patient progress using the extracted DL features, which can predict whether a patient's case would "worsen" or "improve" with an accuracy of 82.7%. Islam and Fleischer [79] used feature-embedded machine learning to distinguish L-type and H-type patients using their CXR images. Moreover, to detect and monitor disease progression and recovery, they categorized multiple images from the same patient.

E. TRAINING DATASET QUALITY

Our findings indicate that it is highly likely that some data samples overlap. The same images could be used multiple times in training, particularly when the authors have collected their data from several online resources that include data originally from the same source. One solution to this problem is to run an image similarity assessment process. It is very crucial to detect the amount of duplication in the training and test sets to avoid overusing data samples and overfitting. This problem has not been discussed in any of the works reviewed in this survey.

Annotation is time-consuming and requires radiologists to grade images at pixel level to specify the COVID-19 biomarkers and complications. Currently, the CXR images found in public datasets are only labeled as normal/healthy, pneumonia, and COVID-19. Establishing a dataset annotated with the main characteristics employed by radiologists to derive their decisions would considerably assist in embedding more useful features to a DL model; thus, the transparency of the decisions made by the model can be enhanced and clinically acceptable automatic detection systems can be obtained. The pixel-level annotation of COVID-19 signs would also be beneficial in determining the disease severity for effectively using resources or prioritizing treatment in heavily affected regions.

The lack of uniformity in the CXR images is another problem that must be addressed. Scalable deep neural network classifiers should be built using samples from several diverse resources. However, image diversity introduces the problem of preprocessing efficiency, and more investigations are required in this direction. Moreover, it is essential to ensure that the CXR images are appropriate for automatic analysis. Images should be of sufficient quality and free from misleading features such as descriptive texts or numbers. Therefore, including a quality assessment component in COVID-19 automatic classification systems is highly recommended. This would also enhance the clinical trust toward the computer-aided diagnosis system.

F. TRANSFER LEARNING FROM GENERAL OBJECT RECOGNITION TASK

Despite the advantages of employing transfer learning, most of the works used deep models pre-trained on general object recognition tasks such as the ImageNet dataset. Only a few studies, such as Mangal *et al.* [89] and Bassi and Attux [59], used architectures pre-trained on a large dataset of CXR images. However, even these architectures are based on networks originally designed for the ImageNet dataset, which raises several questions about their robustness and effectiveness when used in practical clinical scenarios. The automatic selection and optimization of deep neural network architectures and their hyperparameters is another important research domain that could contribute positively to the COVID-19 classifiers. For instance, a generative synthesis approach was used in [61] by specifying the human design requirements.

V. CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH

This study presents a comprehensive review of the diverse DL methods used to detect COVID-19 from CXR images. The current status of this research is discussed here. Besides, the most common pretrained CNN architectures were explained. The datasets that were utilized by different studies are presented and discussed, and the current challenges associated with the current approaches are highlighted. This survey indicated the significant potential of DL methods

in the automatic diagnosis of COVID-19 from the currently available datasets; however, medical personnel and computer scientists should work together closely and utilize their complementary expertise to validate the usefulness of DL techniques.

It was found that CNN-based transfer learning was used in most studies using the same dataset collected by Cohen *et al.* [18]. Despite the encouraging performance achieved, there is still significant room for improvements. First, public, comprehensive, and diverse datasets need to be established. The datasets should be validated by experts and annotated with the corresponding lesions of lung diseases. Incorporating the detection of signs with the classification output would increase both the prediction accuracy and the models' transparency. Second, as the medical research to determine the main characteristics of COVID-19 is still ongoing, it is essential to utilize more features extracted based on recommendations of medical personnel. Given the small size of available CXR COVID-19 datasets, integrating domain knowledge would help create models that mimic human expert diagnostic patterns and focus on the signs or regions they pay particular attention to. However, appropriate domain knowledge should first be determined. The trade-off between the automatically learned deep features and the extracted domain knowledge features should be managed to achieve the desired performance. Third, it is important to measure the amount of disagreement between radiologists to develop a benchmark for use in the prediction evaluation of the deep learning models. Fourth, considering that clinicians often refer to previous analogous cases to make reliable decisions regarding diagnoses, we believe that semi-supervised learning has great potential yet to be unlocked. Semi-supervised algorithms employ few labeled samples and many unlabeled data as part of the training set. Semi-supervised modelling can not only reduce the cost of data annotation, but also help in discovering hidden patterns and relations in the data. Fifth, as seen, most studies in this survey utilized traditional data augmentation operations to deal with the scarcity of COVID-19 CXR images. The promising results achieved by generative adversarial networks (GAN) are worth further investigation. Finally, the promising results achieved by automatically generating a deep CNN architecture tailored for COVID-19 classification task using GenSynth [60] can also be utilized when researchers make a larger and comprehensive COVID-19 CXR dataset available.

APPENDIX

Overfitting: This refers to the phenomenon in which a machine learning model learns a function that fits the training samples perfectly with low error but high variance. Such a model would poorly generalize previously unseen samples.

Regularization: This technique is used for controlling the overfitting phenomenon by adding an additional penalty term in the cost function to avoid extreme values of the model parameters.

Data augmentation: This is a regularization approach that generates an enormous amount of artificial data samples through employing multiple transformations such as flipping, rotating, shifting, resizing, and changing the lighting conditions.

Transfer learning: This is a concept based on representation learning with the underlying assumption that some features are common to many different tasks. In this process, a model that has been trained in a certain setting is used to improve generalization in another setting [83].

Accuracy: This parameter measures the overall performance of a model. It is calculated as the percentage of the correctly classified data samples by the model.

Sensitivity, recall, or true positive rate (TPR): These parameters measure the number of positive cases correctly predicted by the model.

Specificity: This measures the negative cases covered by the model.

Precision: This measures the accuracy of the model in predicting positive samples.

F1-score: This is the harmonic mean of precision and recall.

Receiver Operating Characteristic (ROC): This curve is a plot that displays the trade-off between precision and recall over a series of cut-off points. The closer the curve to the top left corner, the better the classifier.

Area Under the Curve (AUC): AUC of the ROC is used to evaluate a classifier. For a perfect classifier, the AUC would be equal to one.

Explainable Artificial Intelligence: This includes techniques used to visualize and explain DL models, e.g., by highlighting important features in the images used by the model to reach a prediction.

ImageNet: This is a benchmark image dataset organized according to the WordNet hierarchy and consists of a collection of 14,197,122 annotated images of numerous everyday objects such as animals, food, devices, and flowers.

ILSVRC: The ImageNet Large-Scale Visual Recognition Challenge, which was started in 2010 and ran annually until 2017.

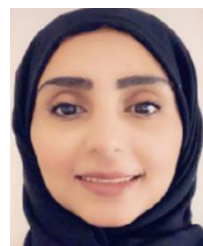
REFERENCES

- [1] World Health Organization. (2020). *Guidance and Standard Operating Procedure: COVID-19 Virus Testing in NHS Laboratories*. Accessed: Jun. 5, 2020. [Online]. Available: <https://www.england.nhs.uk/coronavirus/publication/guidance-and-standard-operating-procedure-covid-19-virus-testing-in-nhs-laboratories/>
- [2] World Health Organization. (2020). *WHO Warns That Few Have Developed Antibodies to Covid-19*. Accessed: Jun. 5, 2020. [Online]. Available: <https://www.theguardian.com/society/2020/apr/20/studies-suggest-very-few-have-had-covid-19-without-symptoms>
- [3] World Health Organization. (2020). *Advice on the Use of Point-of-Care Immunodiagnostic Tests for COVID-19*. Accessed: Jun. 5, 2020. [Online]. Available: <https://www.who.int/news-room/commentaries/detail/advice-on-the-use-of-point-of-care-immunodiagnostic-tests-for-covid-19>
- [4] J.-L. He, L. Luo, Z.-D. Luo, J.-X. Lyu, M.-Y. Ng, X.-P. Shen, and Z. Wen, "Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside Wuhan, China," *Respiratory Med.*, vol. 168, Jul. 2020, Art. no. 105980.

- [5] E. B. G. Kana, M. G. Z. Kana, A. F. D. Kana, and R. H. A. Kenfack, "A Web-based diagnostic tool for COVID-19 using machine learning on chest radiographs (CXR)," *medRxiv*, Jan. 2020, Art. no. 20063263.
- [6] H. Y. F. Wong, H. Y. S. Lam, A. H. T. Fong, S. T. Leung, T. W. Y. Chin, C. S. Y. Lo, M. M.-S. Lui, J. C. Y. Lee, K. W. H. Chiu, T. W. H. Chung, and E. Y. P. Lee, "Frequency and distribution of chest radiographic findings in patients positive for COVID-19," *Radiology*, vol. 296, Mar. 2020, Art. no. 201160.
- [7] A. E. Hassani, L. N. Mahdy, K. A. Ezzat, H. H. Elmousalimi, and H. A. Ella, "Automatic X-ray COVID-19 lung image classification system based on multi-level thresholding and support vector machine," *medRxiv*, Mar. 2020, Art. no. 20047787.
- [8] T. Ozturk, M. Talu, 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.
- [9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciampi, M. Ghafoorian, J. A. W. M. 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, Dec. 2017.
- [10] F. Altuf, S. M. S. Islam, N. Akhtar, and N. K. Janjua, "Going deep in medical image analysis: Concepts, methods, challenges, and future directions," *IEEE Access*, vol. 7, pp. 99540–99572, 2019.
- [11] F. Xing, Y. Xie, H. Su, F. Liu, and L. Yang, "Deep learning in microscopy image analysis: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4550–4568, Oct. 2018.
- [12] K. Muhammad, S. Khan, J. D. Ser, and V. H. C. de Albuquerque, "Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–16, 2020, doi: 10.1109/TNNLS.2020.2995800.
- [13] J. Liu, Y. Pan, M. Li, Z. Chen, L. Tang, C. Lu, and J. Wang, "Applications of deep learning to MRI images: A survey," *Big Data Mining Anal.*, vol. 1, no. 1, pp. 1–18, Mar. 2018.
- [14] P. Seebock, J. I. Orlando, T. Schlegl, S. M. Waldstein, H. Bogunovic, S. Klimesch, G. Langs, and U. Schmidt-Erfurth, "Exploiting epistemic uncertainty of anatomy segmentation for anomaly detection in retinal OCT," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 87–98, Jan. 2020.
- [15] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
- [16] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: A review," *J. Med. Syst.*, vol. 42, no. 11, pp. 1–13, Nov. 2018.
- [17] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," *Insights Imag.*, vol. 9, no. 4, pp. 611–629, Aug. 2018.
- [18] J. Paul Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "COVID-19 image data collection: Prospective predictions are the future," 2020, *arXiv:2006.11988*. [Online]. Available: <http://arxiv.org/abs/2006.11988>
- [19] (2021). *Radiopaedia.org*. Accessed: Jan. 26, 2021. [Online]. Available: <https://radiopaedia.org/>
- [20] *Figure 1—Covid-19 Clinical Library*. Accessed: 2020. [Online]. Available: <https://www.figure1.com/covid-19>
- [21] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al-Emadi, M. B. I. Reaz, and T. I. Islam, "Can AI help in screening viral and COVID-19 pneumonia?" 2020, *arXiv:2003.13145*. [Online]. Available: <http://arxiv.org/abs/2003.13145>
- [22] 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.
- [23] P. Mooney. *Chest X-Ray Images (Pneumonia) | Kaggle*. Kaggle.com. Accessed: 2018. [Online]. Available: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>
- [24] J. C. Gomes et al., "IKONOS: An intelligent tool to support diagnosis of COVID-19 by texture analysis of X-ray images," *Res. Biomed. Eng. Aug.*, 2020, doi: 10.1007/s42600-020-00091-7.
- [25] 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.
- [26] 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-ray images using confidence-aware anomaly detection," 2020, *arXiv:2003.12338*. [Online]. Available: <http://arxiv.org/abs/2003.12338>
- [27] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," 2020, *arXiv:2003.10849*. [Online]. Available: <http://arxiv.org/abs/2003.10849>
- [28] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-covid: Predicting covid-19 from chest X-ray images using deep transfer learning," *arXiv:2004.09363*, vol. 1, Apr. 2020. [Online]. Available: <https://arxiv.org/abs/2004.09363>
- [29] V. Cheplygina, "Cats or CAT scans: Transfer learning from natural or medical image source data sets?" *Current Opinion Biomed. Eng.*, vol. 9, pp. 21–27, Mar. 2019.
- [30] S. Basu, S. Mitra, and N. Saha, "Deep learning for screening COVID-19 using chest X-ray images," 2020, *arXiv:2004.10507*. [Online]. Available: <http://arxiv.org/abs/2004.10507>
- [31] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," 2020, *arXiv:2004.02696*. [Online]. Available: <http://arxiv.org/abs/2004.02696>
- [32] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks Alex," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, vol. 1, 2012, pp. 1097–1105.
- [34] I. Razzak, S. Naz, A. Rehman, A. Khan, and A. Zaib, "Improving Coronavirus (COVID-19) diagnosis using deep transfer learning," *medRxiv*, no. 1, Jan. 2020, Art. no. 20054643.
- [35] P. Kumar and S. Kumari. *Detection of Coronavirus Disease (COVID-19) Based on Deep Features*. Accessed: Mar. 9, 2020. [Online]. Available: <https://www.Preprints.Org/Manuscript/202003.0300/V1>
- [36] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," 2020, *arXiv:2003.13815*. [Online]. Available: <http://arxiv.org/abs/2003.13815>
- [37] 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.
- [38] 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. (CVPR)*, Jun. 2016, pp. 2818–2826.
- [39] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.
- [40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. Conf. Track (ICLR)*, 2015, pp. 1–14.
- [41] P. G. B. Moutounet-Cartan, "Deep convolutional neural networks to diagnose COVID-19 and other pneumonia diseases from posteroanterior chest X-Rays," 2020, *arXiv:2005.00845*. [Online]. Available: <http://arxiv.org/abs/2005.00845>
- [42] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, M. A. Rahman, Q. Wang, S. Qi, F. Kong, X. Zhu, and X. Zhao, "Identification of COVID-19 samples from chest X-ray images using deep learning: A comparison of transfer learning approaches," *J. X-Ray Sci. Technol.*, vol. 28, no. 5, pp. 821–839, Sep. 2020, doi: 10.3233/XST-200715.
- [43] 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.
- [44] L. O. Hall, R. Paul, D. B. Goldgof, and G. M. Goldgof, "Finding covid-19 from chest X-rays using deep learning on a small dataset," 2020, *arXiv:2004.02060*. [Online]. Available: <http://arxiv.org/abs/2004.02060>
- [45] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.
- [46] E. El-Din Hemdan, M. A. Shouman, and M. E. Karar, "COVIDX-net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images," 2020, *arXiv:2003.11055*. [Online]. Available: <http://arxiv.org/abs/2003.11055>
- [47] 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.

- [48] 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. 2261–2269.
- [49] 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*. [Online]. Available: <http://arxiv.org/abs/1704.04861>
- [50] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," 2017, *arXiv:1707.01083*. [Online]. Available: <http://arxiv.org/abs/1707.01083>
- [51] A. Alqudah, S. Qazan, and A. Alqudah, "Automated systems for detection of COVID-19 using chest X-ray images and lightweight convolutional neural networks," *Res. Square*, Apr. 2020.
- [52] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proc. Unsupervised Transf. Learn. (ICML)*, 2012, pp. 37–50.
- [53] S. Khobahi, C. Agarwal, and M. Soltanalian, "CoroNet: A deep network architecture for semi-supervised task-based identification of COVID-19 from chest X-ray images," *medRxiv*, Jan. 2020, Art. no. 20065722.
- [54] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming autoencoders," in *Artificial Neural Networks and Machine Learning (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence, Lecture Notes in Bioinformatics)*, vol. 6791 LNCS, no. PART 1. Berlin, Germany: Springer, 2011, pp. 44–51.
- [55] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 3857–3867.
- [56] A. M. Alqudah, S. Qazan, and A. Alqudah, "Automated systems for detection of COVID-19 using chest X-ray images and lightweight convolutional neural networks," *Res. Square*, Apr. 2020.
- [57] R. Kumar, R. Arora, V. Bansal, V. J. Sahayashela, H. Buckchash, J. Imran, N. Narayanan, G. N. Pandian, and B. Raman, "Accurate prediction of COVID-19 using chest X-ray images through deep feature learning model with smote and machine learning classifiers," *medRxiv*, Apr. 2020, Art. no. 20063461.
- [58] S. Duchesne, D. Gourdeau, P. Archambault, C. Chartrand-Lefebvre, L. Dieumegarde, R. Forghani, and C. Gagne, "Tracking and predicting COVID-19 radiological trajectory using deep learning on chest X-rays: Initial accuracy testing," *medRxiv*, vol. 1, no. 418, May 2020, Art. no. 20086207.
- [59] P. R. A. S. Bassi and R. Attux, "A deep convolutional neural network for COVID-19 detection using chest X-Rays," 2020, *arXiv:2005.01578*. [Online]. Available: <http://arxiv.org/abs/2005.01578>
- [60] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, *arXiv:1711.05225*. [Online]. Available: <http://arxiv.org/abs/1711.05225>
- [61] 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.
- [62] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. Reajul Islam, M. S. Khan, A. Iqbal, N. Al-Emadi, M. B. I. Reaz, and T. I. Islam, "Can AI help in screening viral and COVID-19 pneumonia?" 2020, *arXiv:2003.13145*. [Online]. Available: <http://arxiv.org/abs/2003.13145>
- [63] 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, Jan. 2000.
- [64] *RSNA Pneumonia Detection Challenge Dataset*. Accessed: 2018. [Online]. Available: <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data%0D%0D%0A>
- [65] *Italian Society of Medical and Interventional Radiology (SIRM) COVID-19 DATABASE*. Accessed: 2020. [Online]. Available: <https://www.sirm.org/en/category/articles/covid-19-database/>
- [66] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, and J. Seekins, "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 590–597.
- [67] ChestImaging. *Twitter COVID-19 CXR Dataset, @ChestImaging (Twitter Account)*, *Cardiothoracic Radiologist*. Accessed: 2020. [Online]. Available: <https://twitter.com/ChestImaging>
- [68] Larxel. *Pediatric Pneumonia Chest X-ray*. Kaggle.com. Accessed: 2020. [Online]. Available: <https://www.kaggle.com/andrewmvd/pediatric-pneumonia-chest-xray>
- [69] D. Wang, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (OCT) and chest X-Ray images for classification," *Mendeley Data*, vol. 2, Jun. 2018, doi: [10.17632/rscbjbr9sj.2](https://doi.org/10.17632/rscbjbr9sj.2).
- [70] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, and J. Dong, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [71] *Open Access Biomedical Image Search Engine*, Nat. Library Med., Bethesda, MD, USA, 2020. [Online]. Available: <https://open.nlm.nih.gov/>
- [72] S. Asif and Y. Wenhui, "Automatic detection of COVID-19 using X-ray images with deep convolutional neural networks and machine learning," *medRxiv*, May 2020, Art. no. 20088211.
- [73] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection," 2020, *arXiv:2003.10769*. [Online]. Available: <http://arxiv.org/abs/2003.10769>
- [74] S. Rajaraman and S. Antani, "Training deep learning algorithms with weakly labeled pneumonia chest X-ray data for COVID-19 detection," *medRxiv*, pp. 1–13, Jan. 2020.
- [75] M. Rahimzadeh and A. Attar, "A new modified deep convolutional neural network for detecting COVID-19 from X-ray images," 2020, *arXiv:2004.08052*. [Online]. Available: <http://arxiv.org/abs/2004.08052>
- [76] J. De Moura, L. R. García, P. F. L. Vidal, M. Cruz, L. A. López, E. C. Lopez, J. Novo, and M. Ortega, "Deep convolutional approaches for the analysis of COVID-19 using chest X-ray images from portable devices," *IEEE Access*, vol. 8, pp. 195594–195607, 2020, doi: [10.1109/ACCESS.2020.3033762](https://doi.org/10.1109/ACCESS.2020.3033762).
- [77] I. Castiglioni, D. Ippolito, M. Interlenghi, C. B. Monti, C. Salvatore, S. Schiaffino, A. Polidori, D. Gandola, C. Messa, and F. Sardanelli, "Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: A first experience from Lombardy, Italy," *medRxiv*, Jan. 2020.
- [78] F. Ucar and D. Korkmaz, "COVIDDiagnosis-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.
- [79] M. T. Islam and J. W. Fleischer, "Distinguishing I and h phenotypes of Covid-19 using a single X-ray image," *medRxiv*, Jan. 2020.
- [80] D. Lv, W. Qi, Y. Li, L. Sun, and Y. Wang, "A cascade network for detecting COVID-19 using chest X-rays," May 2020, *arXiv:2005.01468*. [Online]. Available: <http://arxiv.org/abs/2005.01468>
- [81] K. Medhi and I. Hussain, "Automatic detection of COVID-19 infection from chest X-ray using deep learning," *medRxiv*, pp. 1–6, Jan. 2020.
- [82] Y. Zhang, S. Niu, Z. Qiu, Y. Wei, P. Zhao, J. Yao, J. Huang, Q. Wu, and M. Tan, "COVID-DA: Deep domain adaptation from typical pneumonia to COVID-19," 2020, *arXiv:2005.01577*. [Online]. Available: <http://arxiv.org/abs/2005.01577>
- [83] T. Li, Z. Han, B. Wei, Y. Zheng, Y. Hong, and J. Cong, "Robust screening of COVID-19 from chest X-ray via discriminative cost-sensitive learning," 2020, *arXiv:2004.12592*. [Online]. Available: <http://arxiv.org/abs/2004.12592>
- [84] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI (Lecture Notes in Computer Science: Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [85] 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.
- [86] 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, no. 1, pp. 19–40, Feb. 2006.
- [87] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. 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, pp. 475–477, 2014.
- [88] Praveen, "Corona hack: Chest X-ray-dataset," Tech. Rep.

- [89] A. Mangal, S. Kalia, H. Rajgopal, K. Rangarajan, V. Namboodiri, S. Banerjee, and C. Arora, "CovidAID: COVID-19 detection using chest X-ray," Apr. 2020, *arXiv:2004.09803*. [Online]. Available: <http://arxiv.org/abs/2004.09803>
- [90] 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.
- [91] S. Schiaffino, A. Polidori, D. Gandola, and C. Messa, "Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: A first experience from Lombardy, Italy," *medRxiv*, Jan. 2020.
- [92] Z. Q. Lin, M. J. Shafiee, S. Bochkarev, M. S. Jules, X. Y. Wang, and A. Wong, "Do explanations reflect decisions? A machine-centric strategy to quantify the performance of explainability algorithms," vol. 1910, 2019, *arXiv:1910.07387*. [Online]. Available: <http://arxiv.org/abs/1910.07387>
- [93] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020.
- [94] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.
- [95] A. Wong, M. J. Shafiee, B. Chwyl, and F. Li, "FermiNets: Learning generative machines to generate efficient neural networks via generative synthesis," 2018, *arXiv:1809.05989*. [Online]. Available: <http://arxiv.org/abs/1809.05989>
- [96] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," in *Proc. 3rd Int. Conf. Learn. Represent. Work. Track (ICLR)*, 2015, pp. 1–14.
- [97] E. Lindwurm, "InDepth: Layer-wise relevance propagation," *Medium*, 2019.
- [98] C. Agarwal, D. Schonfeld, and A. Nguyen, "Removing input features via a generative model to explain their attributions to an image classifier's decisions," 2019, *arXiv:1910.04256*. [Online]. Available: <http://arxiv.org/abs/1910.04256>
- [99] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," 2013, *arXiv:1312.6034*. [Online]. Available: <http://arxiv.org/abs/1312.6034>
- [100] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 1050–1059.
- [101] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.



HANAN S. ALGHAMDI received the Ph.D. degree in computer science from the University of Surrey, Guildford, U.K., in 2018. She is currently working as an Assistant Professor with King Abdulaziz University, Jeddah, Saudi Arabia. Regarding professional experience, she worked as a Strategic Communication Manager with the Savola Group, Saudi Arabia, and as a Demonstrator with the University of Surrey, U.K. She has published several articles in the field of retinal images analysis, examples include "Measurement of Optical Cup to Disk Ratio in Fundus Images for Glaucoma screening," "Ensemble Learning Optimization for Diabetic Retinopathy Image Analysis," and "Automatic Optic Disc Abnormality Detection in Fundus Images: A Deep Learning Approach." She is involved in several research activities. Her research interests include artificial intelligence, machine learning, deep learning, evolutionary algorithms, optimization, and medical image analysis. She has participated as a Reviewer for multiple journals and conferences locally and internationally, such as IEEE Access Journal, International Conference on Computational Intelligence and Intelligent Systems, CIIS, International Conference on Machine Learning and Human-Computer Interaction, MLHMI, and International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI.



GHADA AMOUDI received the bachelor's degree from King Saud University, Riyadh, Saudi Arabia, in 1996, the master's and Ph.D. degrees in computer science from Dalhousie University, Canada, in 2002 and 2016, respectively, and the Diploma degree in professional development (teaching and learning) from King Abdulaziz University, in 2019. Since 2005, she has been working as an Educator and as a Lecturer with Effat University, from 2005 to 2006. From 2006 to 2011, she was with Arab Open University. Since 2011, she has been an Assistant Professor with King Abdulaziz University. She is currently working on applying AI methods on kidney stone disease research. Her research interests include machine learning, deep learning data science, and social network analysis. She is also a member of the Saudi Artificial Intelligence Society.



SALMA ELHAG received the B.S., M.S., and Ph.D. degrees in computer science from the University of Khartoum, Sudan, in 2003, 2006, and 2014, respectively. Since 2014, she has worked as a Teaching Assistant and then as a Lecturer with the University of Khartoum. Since 2011, she worked as a Lecturer with the Faculty of Computing and Information Technology, King Abdul-Aziz University, as an Assistant Professor, since 2014, and as an Associate Professor, since 2019. Her research interests include artificial intelligence, computation intelligence, network security, data mining, artificial neural networks, fuzzy logic, and evolutionary algorithms.



KAWTHER SAEEDI received the Ph.D. degree from the School of Computing, Manchester University, U.K. She is currently an Assistant Professor with the Information System Department, King Abdulaziz University. With regards to professional experience, she worked as a Technology Specialist with ING Financial Market, Amsterdam, and as a Sales Executive Engineer with IT Integrated Solutions, Aljuffali, Jeddah. She has managed projects of different scale and technologies. She is a certified Professional Scrum Master, a Senior Software Engineer, and a System Architect. She has lead the IS Department for four years. She is interested in applied research involving new technological trends. Her core research interests include software engineering, blockchain, and artificial intelligence.



JOMANAH NASSER was born in Riyadh, Saudi Arabia. She graduated from the Medical College, King Abdulaziz University, in 2019. She completed her internship, in 2020. She published a research on The impact of Stone Composition on Renal Function. She is in the process of publishing a research on Unusual Neurological Complications Nasal Desmopressin for Nocturnal Enuresis.