IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 1, 2021, accepted March 15, 2021, date of publication March 30, 2021, date of current version April 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3069937

Intelligent Pneumonia Identification From Chest X-Rays: A Systematic Literature Review

WASIF KHAN, NAZAR ZAKI, AND LUQMAN ALI
Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University (UAEU), Al Ain, United Arab Emirates

Corresponding authors: Nazar Zaki (nzaki@uaeu.ac.ae) and Wasif Khan (201990025@uaeu.ac.ae)

ABSTRACT Chest radiography is a significant diagnostic tool used to detect diseases afflicting the chest. The automatic detection techniques associated with computer vision are being adopted in medical imaging research. Over the last decade, several remarkable advancements have been made in the field of medical diagnostics with the application of deep learning techniques. Various automated systems have been proposed for the rapid detection of pneumonia from chest X-rays. Although several algorithms are currently available for pneumonia detection, a detailed review summarizing the literature and offering guidelines for medical practitioners is lacking. This study will help practitioners to select the most effective and efficient methods from a real-time perspective, review the available datasets, and understand the results obtained in this domain. It will also present an overview of the literature on intelligent pneumonia identification from chest X-rays. The usability, goodness factors, and computational complexities of the algorithms employed for intelligent pneumonia identification are analyzed. Additionally, this study discusses the quality, usability, and size of the available chest X-ray datasets and techniques for coping with unbalanced datasets. A detailed comparison of the available studies reveals that the majority of the applied datasets are highly unbalanced and limited, providing unreliable results and rendering methods that are unsuitable for large-scale use. Large-scale balanced datasets can be obtained via smart techniques, such as generative adversarial networks. Current literature has indicated that deep learning-based algorithms achieve the best results for pneumonia classification with an accuracy of 98.7%, a sensitivity of 0.99, and a specificity of 0.98. The higher accuracy offered by deep-learning algorithms in addition to their appropriate class balancing techniques serves as a good reference for further research.

INDEX TERMS Chest radiography, computer vision, deep learning, generative adversarial networks, medical imaging, pneumonia detection, unbalanced datasets.

I. INTRODUCTION

A chest X-ray (CXR) is a simple, economical, and commonly adopted mechanism for diagnosing lung infections [1]. The CXR can be perceived as either normal or presenting a disease, such as lung cancer, tuberculosis, or pneumonia by an experienced radiologist. One of the most common lung diseases is pneumonia, a lung infection caused by viruses, bacteria, or fungi [2]. Pneumonia is life-threatening to infants, older adults, patients placed on a ventilator in hospitals, and patients with asthma. Moreover, pneumonia is a high-risk illness, especially in developing countries where millions of people are impoverished and lack access to medical facilities. The World Health Organization (WHO) [3] estimates that each year, more than four million deaths are caused by pneumonia and other diseases associated with air pollution [3]. More than 150 million people, mainly children under five years old, are infected with pneumonia annually [4]. Viral pneumonia tends to be mild, whereas bacterial pneumonia is severe, especially in children [5]. Fungal pneumonia can occur in patients with weak immune systems. Because of the low cost of CXRs, they are more commonly requested than other medical modalities such as magnetic resonance imaging (MRI) and computed tomography (CT) [6]. The demand for CXRs translates to thousands of readings per radiologist annually. However, there is a shortage of radiologists in both developing and developed countries [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Easter Selvan Suviseshamuthu.

The mortality from lung diseases can be reduced via accurate and timely diagnosis. In developing countries, where diagnoses and treatment are delayed because of the shortage of experienced radiologists, pneumonia in children is associated with alarming death rates. The massive imbalance between the number of doctors and the population of a specific area also hinders timely diagnosis. Furthermore, the resolution of CXRs is lower than those of MRI and CT; and therefore, they cannot be easily interpreted even by experienced radiologists. The decision-making of medical staff can be supplemented by computer-aided diagnostic (CAD) tools, which combine aspects of computer vision and machine learning (ML) with radiological image analysis to recognize and extract patterns [8]. A typical CAD system sequentially processes the input data (i.e., CXRs), extracts and classifies the features. In the first step, the CXR data are preprocessed. In the second step, the features are extracted from input images using various techniques such as Gaussian filters [9], morphological operation [10], and edge detection [11]. In the third step, the extracted features are distinguished using a suitable classifier such as a support vector machine (SVM) [12], the random forest (RF) algorithm [13], or the neural network [14].

The publicly available CXR datasets for pneumonia are highly class-imbalanced, indicating that more images are available in one class than in the other. The accuracy of a classification system is considerably affected by class imbalance. For instance, the dataset compiled by Kermany *et al.* [2] (used as an example in this study) includes 5,856 images, among which 4,273 images belong to the disease (pneumonia) class and only 1,583 images belong to the normal class. Several preprocessing techniques can be used to resolve the class imbalance problems. Traditionally, the minority class is oversampled by duplicating randomly selected samples, whereas the majority class is undersampled. However, random oversampling introduces an overfitting problem, and valuable information can potentially be lost because of undersampling. Besides, the sampling technique exhibits limited generalization and variance [15]. The training dataset is often expanded via data augmentation, resulting in false representations of the original images to avoid overfitting [16]. More recently, artificial datasets have been generated by generative adversarial networks (GANs) [17], which include two neural networks: a generator and a discriminator. The generator synthesizes artificial samples with the required variations from the input data distribution, and the discriminator differentiates between the samples generated by the generator and those in the input data. Thoroughly engineered architectures, such as deep convolutional GAN (DCGAN) [18], styleGAN [19], and Cycle GAN [20], have been introduced for more robust synthesized data generation.

Traditional CAD-based systems [21] have successfully classified lung diseases from CXRs, but they require extensive handcrafted techniques for feature-extraction from images, followed by applying ML classifiers [12], [13]. To overcome this limitation, several artificial intelligence (AI) based solutions were developed. Data-driven deep learning (DL) methods for instance achieved automatic end-to-end feature extraction and classification. Over the past decade, the convolutional neural network (CNN), a type of deep neural network, has achieved groundbreaking results in different tasks related to pattern recognition. Inspired by the visual cortex of humans [22], CNNs differentiate several classes in image recognition problems [16]. However, CNN requires a large volume of training data. Medical image classification using CNN is usually performed either by training the CNN from scratch, using an existing pre-trained network without retraining, or fine-tuning a pre-trained network on a target dataset [23].

Recently, DL-based algorithms, such as SegNet [24], U-Net [25], Chexnet [26], and CardiacNet [27], have become the default choice for medical imaging applications. DL-based techniques are applied for pneumonia detection from CXRs [28].

In this study, the current literature on pneumonia identification from CXR images is presented and the contributions are summarized as follow:

- Review and analysis of the usability, goodness factors, and computational complexity of the algorithms that were used for pneumonia identification.
- Review and analysis of the quality, usability, size, and class balance extent of the available CXR images datasets.
- Discuss and provide a comparative analysis of the algorithms used for pneumonia identification.
- Recommend future directions based on the current literature for the medical practitioner and research community which could assist them to select the best methods from a real-time perspective, perceive the available datasets, and understand the results obtained in this domain.

The remainder of this paper is organized as follows. Section II describes the research method adopted, Section III elaborates on the datasets and presents their details. Section IV focuses on data preprocessing and augmentation techniques that solve the unbalanced class problem, and Section V reviews the techniques applied to lung disease detection. Section VI describes the evaluation metrics followed by a comparative analysis and discussion in section VII. Section IX finalizes the conclusion of the paper.

II. RESEARCH METHOD

An unbiased research method [29] that can ensure complete coverage of the relevant research is required for conducting a systematic literature review.

A. DATA SOURCES

We used four electronic databases (ED) as primary data sources to search for the relevant studies. The electronic databases used during the search process are presented in Table 1.

TABLE 1. Data sources.

Identifiers	Databases	URL	Results
ED1	IEEE Xplore	https://ieeexplore.ieee.org/	74
ED2	Science Direct	https://www.sciencedirect.com/	80
ED3	Springer Link	https://link.springer.com/	74
ED4	ACM	https://dl.acm.org/	79

B. SEARCH TERMS

The following search terms were used based on the research question and the literature:

- Pneumonia detection OR health safety infectious diseases of chest OR X-ray dataset OR deep learning for pneumonia
- Chest radiography OR chest X-rays OR chest disease detection

The search string for the automated search is as follows: (Chest OR pneumonia OR diseases OR X-ray OR X-ray OR CXR OR lung diseases); data balancing AND chest X-rays OR pneumonia, Generative Adversarial Networks (GANs) AND chest X-rays

C. STUDY SELECTION PROCEDURE

The study selection procedure included identification, screening, eligibility, and inclusion of the research papers.

D. INCLUSION AND EXCLUSION CRITERIA

We extracted relevant studies from different data sources under the following inclusion and exclusion criteria.

1) INCLUSION

- Studies related to pneumonia identification using CXRs
- Studies not focused on pneumonia detection but contributing to solving the problem
- Studies published in a peer-reviewed conference or journal
- Studies published between 2010 and 2020 inclusively.
- Studies published in English
- Peer-reviewed publications

2) EXCLUSION

- Studies other than pneumonia identification using CXRs
- Studies published in languages other than English
- Studies with no validation of the proposed technique
- Editorials, short papers, posters, technical reports, patents, reviews, Wikipedia, surveys, and extended papers

E. QUALITY ASSESSMENT CRITERIA

The following quality criteria were defined for our systematic review:

- Are the study objectives clearly defined?
- Are the estimation methods well defined?
- Is the estimation accuracy measured and validated?
- Are the limitations of the study explicitly stated?
- Is the result published in a reputable venue?

III. DATASETS

A. THE CHEST X-RAY14 DATASET

Wang *et al.* [1] presented a chest X-ray database named Chest X-ray14 [1], which contains 108,948 images of eight (currently 14) diseases obtained from 32,717 patients (collected from 1992 to 2015). This dataset was acquired from Picture Archiving and Communication Systems (PACS). The eight common thoracic diseases, i.e., atelectasis, effusion, infiltration, cardiomegaly, mass, nodule, pneumonia, and pneumothorax, were shortlisted as keywords in a search of the PACS system, and the associated images were extracted. Some sample images of multiple chest pathologies are presented in Fig. 1. The typical size of the X-ray images (3000 × 2000 pixels) was resized to 1024 × 1024 pixels without losing any significant details or contents. Among the 108,948 images, 24,636 images contained one or more pathologies, whereas the remaining 84,312 images were normal. The mean age of the patients was 46.9 years, with 63,340 (56.5%) CXRs for male and 48,780 (43.5%) CXRs for female patients. Less than 1,500 pneumonia images were found in this publicly available dataset.

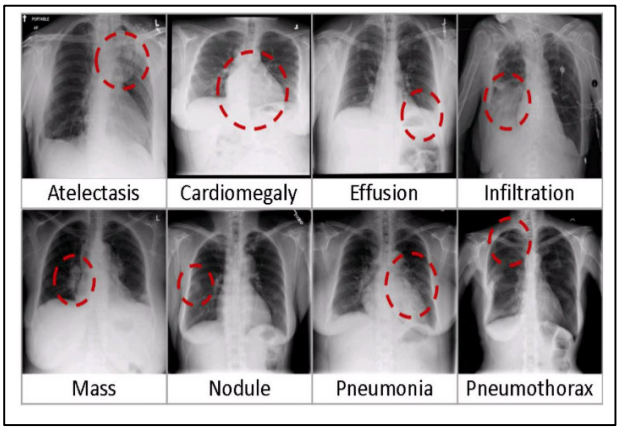

FIGURE 1. Examples of lung-disease images extracted from the Chest X-rays14 dataset.

B. PEDIATRIC CXRS FOR PNEUMONIA DETECTION

Kermany *et al.* [2] collected and labeled pediatric CXRs from the Guangzhou Women and Children's Medical Centre (Guangzhou, China). CXR imaging was performed as part of the routine clinical care of the patients. In this study, the authors selected 5,856 CXR images from pediatric patients aged 1–5 years: 4,273 pneumonia images (2,780 bacterial and 1,493 viral) and 1,583 normal images. The X-ray images are available in various dimensions such as 1040 × 664, 1224 × 1000, and 1848 × 1632 pixels. Fig. 2 shows a sample image obtained from the dataset of Kermany *et al.* [2].

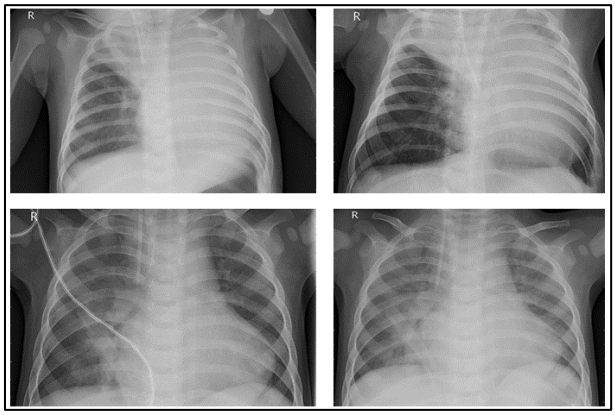

FIGURE 2. Sample images from the Pediatric CXR dataset.

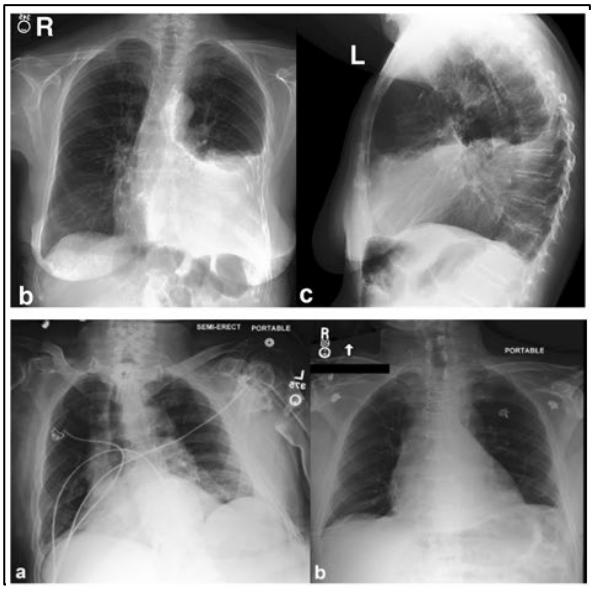

FIGURE 3. Sample images are taken from the MIMIC dataset.

C. MIMIC CXR

Currently, MIMIC–CXR–JPG [30] is the largest publicly available CXR dataset worldwide. It contains more than 377,110 CXRs obtained via 227,827 studies. It contains the data of 14 chest diseases sourced from the Beth Israel Deaconess Medical Centre (Massachusetts, USA) from 2011 to 2016. The CXRs are publicly available in full resolution (typically 2544×3056). Sample images from this dataset are presented in Fig. 3.

D. OPEN-I INDIANA DATASET

Demner-Fushman *et al.* [31] collected the Indiana (USA) dataset from multiple hospitals associated with the Indiana University School of Medicine. This dataset comprises 7, 470 chest radiographs (resolution: 512×512 pixels) and 3,955 associated reports. The images are annotated with disease view (frontal or lateral) and include pulmonary edema, cardiac hypertrophy, pleural effusion, and opacity.

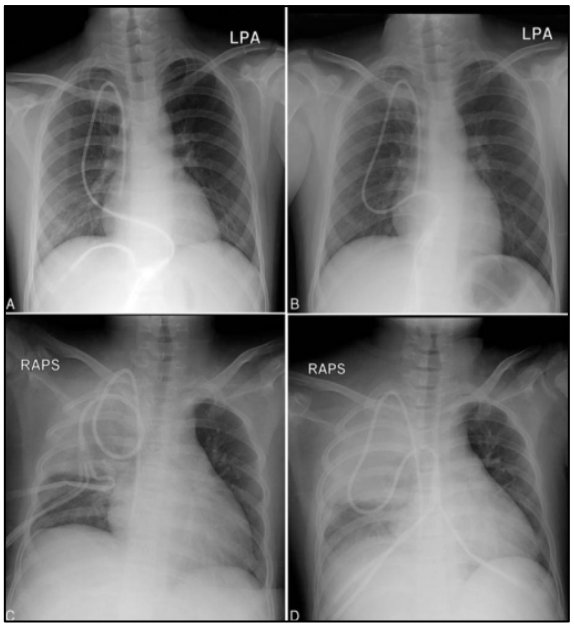

FIGURE 4. Sample images from the Indiana dataset.

This dataset is publicly available for users but contains only 40 pneumonia samples. Sample images from this dataset are presented in Fig. 4.

E. MC DATASET

Collaborating with Montgomery County's (MC) screening program for tuberculosis (USA), Jaeger *et al.* [32] obtained 138 chest images, including 80 from healthy subjects and 58 from tuberculosis patients. Among these images, 63 were obtained from males, 74 were obtained from females, and one was obtained from a patient of unknown/other gender. The resolution of the images varied from 4020×4892 to 4892×4020 pixels. This dataset is publicly available but contains no pneumonia samples. Therefore, it cannot be used as a pneumonia detection dataset. Sample images from the MC dataset are presented in Fig. 5.

F. SHENZHEN DATASET

In addition to the MC database, Jaeger *et al.* [32] obtained the Shenzhen dataset from Shenzhen's hospital (Guangdong Province, China), named the Guangdong Medical College. This dataset, which was collected as part of the hospital's routine within 1 month, contains 662 chest radiographs, including 326 healthy and 336 tuberculosis cases. Among the 662 images, 442 were obtained from males, 213 were obtained from females, and seven were obtained from patients of unknown/other gender. The dataset is publicly available in PNG format, and the image size is approximately 3000×3000 pixels. Some examples are displayed in Fig. 6.

G. KIT DATASET

Ryoo and Kim [33] collected 10,848 images from the Korea Institute of Tuberculosis. This dataset includes 7,020 images

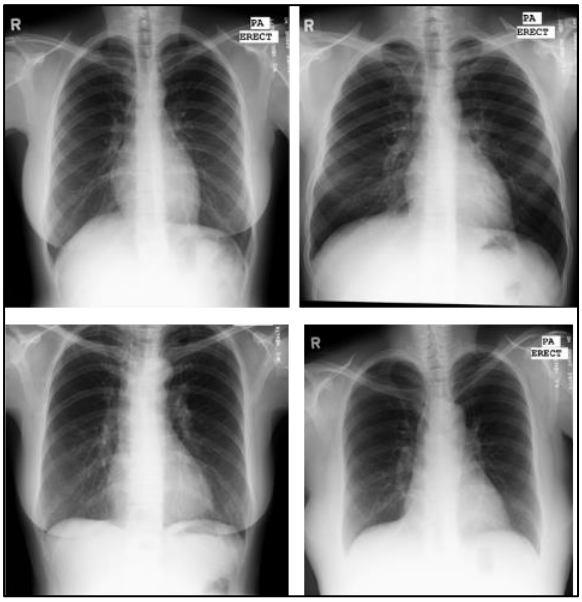

FIGURE 5. Sample images from the MC dataset.

FIGURE 6. Sample images from the Shenzhen dataset.

from healthy subjects and 3,828 images from tuberculosis cases.

H. JSRT DATASET [34]

The Japanese Society of Radiological Technology [34] collected 247 chest images, among which 154 presented with pulmonary nodules (100 nodules were malignant and 54 were benign), and 93 were nodule-free. Among these 154 images, 119 were from men (68 with nodules and 51 without nodules) and 128 were from women (86 with nodules and 42 without nodule). The average age of the patients with nodules was 60 years. All X-ray images are sized 2048 × 2048 pixels, and the depth of the grayscale is 12 bits. Examples are shown in Fig. 7.

IV. DATA PREPROCESSING

The majority of the CXRs are obtained in Digital Imaging and Communications in Medicine (DICOM) format with a large amount of metadata. However, this format cannot be easily understood by experts outside the radiology domain [30].

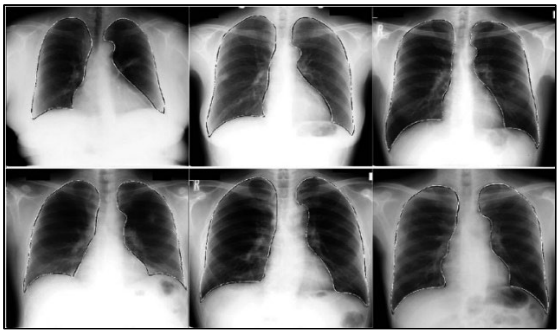

FIGURE 7. Sample images from the JSRT dataset.

In other domains, images are usually stored in formats such as PNG and JPEG, which are processed using compression algorithms to conserve the information in the image without losing any desired information. First, the information of the concerned patient is deidentified to satisfy the mandatory privacy standards. This step requires the removal of the patient's details, identifiers, and dates using a customized algorithm. In this algorithm, image processing and an optical character recognition system are combined to detect and remove text from the X-ray image. After de-identification, the DICOM images are converted into JPEG or bitmap format using previously described approaches [1], [30].

However, the dimensions of normal X-ray images (3000 × 2000 pixels) are difficult to process due to the high computational cost. Therefore, the X-ray images must be dimensionally reduced to an optimal size while preserving the vital information in them. For example, Wang *et al.* [1] reduced the size of X-ray images to 1024 × 1024 pixels and 512 × 512 pixels. The sizes of the MIMIC [30] and pediatric CXR datasets were set to 2048 × 2048 and 1024 × 1024 pixels, respectively.

A. DATA BALANCING, AUGMENTATION, AND ENHANCEMENT USING TRADITIONAL TECHNIQUES

The majority of the pneumonia image datasets suffer from the class-imbalanced issue. The imbalance problem has been traditionally using the following techniques. The mean squared error (MSE) is effective to evaluate classification models on balanced datasets [35] but often fails when the data are imbalanced because it equally captures the errors in the minority and majority classes [35], [36]. The loss function is biased toward the majority class because the error associated with the majority class overwhelms that associated with the minority class. Wang *et al.* [35] proposed the mean false error (MFE) and its improved version, i.e., the mean squared false error (MSFE), for training the neural networks on class-imbalanced data [37]. MFE is inspired by the concepts of false-positive rate and false-negative rate and can be used to estimate the average error in each class. Further, the mean false positive error (FPE) associated with the majority class and the mean false-negative error (FNE) associated with the

minority class is added, as shown in (1).

$$l = FPE + FNE. \tag{1}$$

Although MFE performs better on imbalanced data than the MSE [38], however, the MFE loss does not ensure that FNE will also be reduced because the FPE error contributes more to the sum (MFE) due to the presence of considerably more samples in the majority class. Therefore, the MFE loss function is insensitive to the error associated with the minority class and cannot guarantee higher performance when using this class [35]. This problem is effectively solved by the MSFE, in which the loss function (l') is calculated as

$$\begin{aligned} l' &= (FPE)^2 + (FNE)^2 \\ &= \frac{1}{2}((FPE + FNE)^2 + (FPE - FNE)^2 \end{aligned} \tag{2}$$

To minimize MSFE, the errors associated with the majority and minority classes are simultaneously minimized, resulting in improved unbiased classification accuracy [39], [40]. A better classification performance can be obtained based on the MSFE when compared with that obtained based on the MSE [41]; however, the improvement is only minor [35], [40]. Data sampling can also help in data balancing problems. Samples can be randomly duplicated from the minority class and randomly removed from the majority class. However, as mentioned above, random duplication and undersampling result in overfitting (owing to duplicate samples) and loss of important information, respectively. Undersampling and oversampling methodologies for solving data-unbalanced problems have been substantially developed since the late 1990s [42]–[48]. Numerous researchers have suggested and employed different sampling techniques.

Several reviews have focused on the effectiveness and usability of these methodologies [43], [44]. Random sampling, as proposed by some researchers, has several shortcomings. First, it can eliminate important relevant samples from the data, risking the overfitting problem. Kubat [45] selected samples from the original population via one-sided selection, forming an undersampled dataset. The Tomek links algorithm [48] and the condensed nearest neighbor method [42] were employed to identify the bad samples from the data. The latter algorithm filters out the noisy and bad examples from the majority class that needs to be undersampled.

Laurikkala [46] suggested the neighborhood cleaning rule for filtering excess examples from the majority class data. The authors calculated three nearest neighbors for every example (Di) from the training set. If Di fits within the common class but has been misclassified by the chosen nearest neighbors, it is detached from the dataset. If Di is located within the less common class and has been misclassified by the chosen nearest neighbors, it should also be detached from the dataset. This approach can hit a computational bottleneck when processing large and highly imbalanced datasets.

The synthetic minority oversampling technique (SMOTE) [47] creates simulated data based on the similarities between pairs of the existing minority samples. New instances in the minority class are created by obtaining samples from the target class and combining their features with the features of their neighbors. The classification performance of the minority class can be improved using the newly synthesized instances by balancing its data with those of the majority class [49]. However, SMOTE has limited generalizability and fails when data cannot be linearly separated [50]. Further, it cannot handle high-dimensional data [50]–[52]. Therefore, Wang [53] incorporated SMOTE into a locally linear embedding algorithm (LLE) that maps high-dimensional data into a low-dimensional space, where data are more separable. The LLE–SMOTE algorithm reverse-maps the instance synthesized in the new feature space to its original input space [53], [54]. The authors collected three datasets of CXR images and verified the pulmonary detection performances of their method using multiple classifiers (k-neural network, SVM, and naïve Bayes). The minority-to-majority class ratio was approximately 1:25. The classification accuracy of the LLE–SMOTE algorithm was greater than that of the conventional SMOTE by 2–4% [53]. However, the LLE–SMOTE algorithm cannot be generalized to new data because it is based on the LLE algorithm [55].

Krawczyk *et al.* [56] presented several data sampling techniques for class-imbalanced problems (SMOTEBoost [57], OverBagging [58], RAMO [59], ADASYN [60], and conventional SMOTE [47]) and tested them on a dataset of 340 images for breast cancer detection. The majority class contained 144 images of intermediate malignancy, whereas the minority class contained 26 images of high malignancy (giving an approximate imbalance ratio of 6:1). As the base classifier, they employed an SVM with a Gaussian kernel and a minimal optimization training procedure. SMOTEBoost achieved the highest sensitivity (88.46%) and specificity (exact negative rate = 88.8%) among the tested oversampling methods. The data sampling techniques proposed in a previous study [56] are given below.

The adaptive synthetic sampling approach for imbalanced learning (ADASYN) algorithm [60] uses the weighted distributions of different instances from the minority class based on the difficulty level. Here, more synthetic instances are generated for a minority class that is harder to learn when compared with the instances that are easy to learn. It improves learning concerning the data distribution by reducing the bias caused by class imbalance and shifts the classification boundary toward difficult instances [61]. Because ADASYN is adaptive, its precision can be affected [62].

Ranked minority oversampling (RAMO) [59] is an extension of ADASYN that ranks the samples from the minority class during each iteration according to a sampling probability distribution. Unlike ADASYN, in which the number of synthetic instances must be specified, RAMO determines the probability of generating synthetic instances for each minority example [56]. However, RAMOboost is sensitive to noise [62].

Instead of randomly sampling the whole dataset, OverBagging [58] separately and randomly oversamples the minority class in each bagging iteration. The main advantage associated with this method is that it will increase the minority samples and the majority of samples can be used as a part of each bootstrap iteration. However, the complexity of the algorithm increases with the increasing number of samples during the training process.

The SMOTEBoost algorithm [57] is the combination of the SMOTE and boosting procedure. Here, the prediction accuracy of the minority class improves when using SMOTE; further, boosting improves the accuracy by focusing on the difficult examples in both the minority and majority classes. Although SMOTEBoost outperforms SMOTE, it is difficult to optimize and the boosting technique is sensitive to noise.

B. DATA BALANCING USING GENERATIVE ADVERSARIAL NETWORKS (GANS)

Traditional oversampling techniques are usually employed after the feature extraction step in training data [63]. However, as DL algorithms utilize end-to-end feature extraction and classification, incorporating oversampling techniques with DL models would require costly parameter tuning [63], [64]. Furthermore, these techniques can work for low dimension tabular data but not for high-dimensional image data [65], [66]. Therefore, GAN is a suitable alternative when applied to image datasets. A GAN generates new samples based on the learned input (training) data distribution p. GANs comprise two multilayer perceptrons: a discriminator (D) and a generator (G). The generator creates samples from the simple distribution p(g) such that $p(data) = p(g)$. During the training process, G maximizes the error-generating probability of D and creates fake samples following the input data distribution. After several training steps, the goal $p(data) = p(g)$ is achieved and D cannot discriminate between the real samples and samples from G. Thus, synthetic samples can be generated using GANs.

GANs have attracted attention because of their high performance; however, they are unstable during training and often produce nonsensical outputs from G. To overcome these problems, researchers have recently developed modified GANs such as DCGAN [18], styleGAN [19], and Cycle GAN [20]. DCGAN is more stable during training than the traditional GAN, and both its generator and discriminators are deep CNNs [67]. The GAN-based network StyleGAN can generate high-resolution images (e.g. 1024×1024 pixels) in which the generator network is modified via stacking in fully connected layers. Although StyleGANs can generate high-dimensional realistic images [68], the images are spoiled by artifacts such as water splotches and phase artifacts [69], [70]. CycleGAN [20] is used to translate an unpaired image from a source domain to a target domain using a technique such as style transfer, season transformation, and object transfiguration. CycleGAN uses the cycle- consistency loss to return the translated image to the original image [71]. In this type of network, only one style can be transferred per network; moreover, the algorithm with the cycle-consistent loss function becomes complicated with the increasing number of transferred styles [72].

Wei *et al.* [73] proposed a structure-correcting adversarial network (SCAN) for organ segmentation (left lung, right lung, and heart) from CXRs. SCAN is suitable for small training datasets because of its critical network that can differentiate between the ground truth annotations and the mask synthesized by the segmentation network [73]. The critic network and segmentation network can be trained end-to-end to achieve realistic segmentation outcomes [73], [74]. The algorithm was evaluated using the publicly available JSRT [34] and Montgomery [32] datasets. The authors of a previous study [73] trained a fully convolutional network (FCN) on a small dataset of grayscale CXR images. The segmentation network could be optimized using FCN, whereas SCAN differentiated the segmentation network predictions from the ground truth annotations. Data augmentation was not applied because it did not improve the results. The algorithm was tested by the intersection-over-union (IoU) and dice coefficient metrics. In the JSRT evaluation, the adversarial training improved the performance from 92.9% in FCN to 94.7% in the proposed algorithm.

Adar *et al.* [75] isolated the region of interest (ROI) in a synthetic liver lesion using a GAN architecture and classified its features using CNN. They used a limited dataset of 182 CT images of liver lesions (53 cysts, 64 metastases, and 65 hemangiomas). The sensitivity and specificity were improved from 78.6% and 88.4%, in classic data augmentation to 85.7% and 92.4%, respectively, in the GAN-generated synthetic data augmentation. However, the ROI segmented in their paper was excised only from a 2D CT image. Gupta *et al.* [76] proposed a GAN for bone-lesion detection. To reduce the computational resources required, they extracted a lesion patch from the X-ray rather than from the full image and thereby obtained many patches. For training, the lesions were annotated manually under expert advice. A non-lesion patch was translated into a lesion patch.

Salehinejad *et al.* [74] proposed DCGAN for generating artificial CXR images from real X-ray images. In the present study, the dataset was obtained from the Radiology Information System, which stores the data of four diseases, and the classification was performed by using AlexNet [16]. The accuracy improved from 88.4% [75] to 92.1%, but the generated images had a relatively low resolution. Chuquicusma *et al.* [77] artificially generated CT images of lung nodules by using DCGAN and checked their quality in a visual Turing test conducted by two radiologists. The DCGAN was judged to have produced highly realistic samples when the radiologists could not distinguish the original images from the DCGAN-generated samples.

Chuquicusma *et al.* [77] extracted 1145 lung nodules (635 benign and 510 malignant) from the Lung Image Database Consortium image collection. They generated lung nodules (benign and malignant) using DCGAN and compared the qualities of the generated and original nodules.

As a visual Turing test, they presented the generated and real images to radiologists in a radiology room to evaluate their perception and classify between real and generated lung nodules (malignant, benign, and both). The experienced radiologists, one with 13 years of experience (Radiologist 1) and the other with four years of experience (Radiologist 2), were not shown the sample before the experiments and could not observe each other's evaluation. Both the radiologists were presented with 36 high-quality samples on a 6×6 squared grid. Radiologist 1 accurately assessed 67% and 58% of the fake and real generated nodules, respectively. Meanwhile, Radiologist 2 accurately evaluated 100% and 92% of the fake and real generated nodules, respectively. The interobserver agreements of the benign and malignant real cases were 44.91% and 58.56%, respectively. The experimental results revealed that realistic samples can be generated using DCGAN. However, some samples contained signs of both benign and malignant nodules. These ambiguous, low-quality instances should be improved.

Baur *et al.* [78] proposed the deeply discriminated GAN (DDGAN) algorithm, which can be trained end-to-end and has multiple discriminators attached to different levels of the generator. Realistic high-resolution samples of skin lesions can be obtained from a set of 2,000 training samples using the modified Laplacian pyramid GAN (LAPGAN) algorithm. This algorithm was evaluated on the 2017 International Skin Imaging Collaboration (ISIC) [79] dataset, containing 2000 samples (374 melanomas, 1372 benign lesions, and 254 seborrheic keratosis samples). The performance of LAPGAN was compared with those of DCGAN and DDGAN. LAPGAN slightly outperformed DDGAN because it involved multiple noise sources, which are beneficial for sample diversity and realism. However, the proposed LAPGAN algorithm proved to be challenging to train, and its hyperparameters were difficult to adjust. Meanwhile, when trained on 374 images, DCGAN failed but DDGAN could be trained easily with fast convergence and no degradation from severe high-frequency artifacts; accordingly, it realistically generated high-quality images of 256×256 pixels.

Malygina *et al.* [80] attempted to solve the unbalanced data problem using CycleGAN for sample creation and DenseNet for classification. The authors used the binary classification of pneumonia versus normal samples on the Chest X-rays14 dataset [2]. To balance the dataset, they generated new samples using CycleGAN and achieved an area under the curve value of 0.9939. However, no other performance metrics were evaluated, and no comparisons with related works or nonstandard data divisions were reported. Unlike GANs, which learn mappings only from random noise vectors, conditional GANs (cGANs) learn mappings from both observed images and random noise vectors. A cGAN conditions the input and generates a corresponding output image, synthesizing a novel sample with specific attributes. However, it is computationally expensive and restricted to attribute labeling.

Mahapatra *et al.* [81] generated CXR images from the Segmentation in Chest Radiographs (SCR) database, which contained 247 lung images (153 nodules and 93 healthy), using a cGAN and an active learning technique. They adopted a three-step methodology: sample generation using a conditional GAN, classification/segmentation of a lung nodule, and sample informativeness calculation. They labeled the dataset of 16 samples and manually segmented the masks in each class, thereby generating up to 200 synthetic images from each test image using cGAN. After their informativeness was verified using a Bayesian neural network (BNN), the generated images were added to the training data to fine-tune the previously trained classifier. These steps were repeated until the performance of the classifier did not change. The authors varied the number of samples in the initial training set and showed that the method achieved almost the same results as fully supervised learning (training and testing on the whole dataset) even when training and testing on 35% of the full dataset. After their informativeness was verified using a BNN, the generated images were added to the training data to fine-tune the classifier. The test set includes 400 images (200 normal and 200 nodular). When training on 35% of the labeled data, the accuracy was 91.9% versus 92.4% in fully supervised learning.

V. DATA MINING TECHNIQUES

A. TRADITIONAL MACHINE LEARNING

Oliveria *et al.* [21] proposed PneumoCAD, which is an ML-based network that can be used to identify pneumonia and normal pediatric CRX images. The authors obtained a subset of 40 images (20 pneumonia and 20 non-pneumonia images) from a knowledge database for training [21]. The system was evaluated using another 20 random test images extracted from the same database. The texture features were extracted using eight wavelet transform coefficients: Haar, three members of the Daubechies family (Db 2, Db 4, and Db 8), two members of the Coiflets family (Coif 2 and Coif 4), and two members of the biorthogonal family (Bior 2.2 and Bior 4.4). The features of pneumonia and normal images were differentiated by the weighted nearest neighbor method based on the Euclidean distance. In the evaluation experiments, the Haar wavelet transforms provided the highest accuracy of feature extraction from X-rays (97%) but a mediocre specificity (80%). Even though the PneumonCAD prototype can classify pneumonia images, its specificity is not promising. Furthermore, this method lacks a preprocessing technique that can be used to eliminate noise from the input images. A similar ML-based pneumonia detection system called PneumoCAD was developed by Macedo *et al.* [60]. PneumoCAD accepts CXRs as the input and extracts their useful features using the Haar wavelet transform. The pneumonia images are then classified using a k-nearest neighbor (KNN) classifier. The model was trained and evaluated on a dataset of 156 verified child chest radiographs (78 pneumonia and 78 normal). The average accuracy of the KNN ($k = 9$) was 91.75%. Although its accuracy should be improved, this study will assist the future development of CAD systems.

Sousa *et al.* [82] extended PneumoCAD [21] by comparing the performances of various ML algorithms concerning childhood pneumonia detection from CXRs. The authors used a self-created dataset known as the PneumoCAD dataset, which contained 156 grayscale chest radiographs. These images were first annotated by expert radiologists following the WHO guidelines. The authors extracted several texture-based features for the classification task: the coefficient of variation, energy, contrast, average energy, correlation, entropy, difference variance, average deviation, difference entropy, inverse, difference moment, sum average, residual mean, sum entropy, variance, sum variance, standard deviation, and suavity [83]–[86]. The sequential forward elimination algorithm was used to determine the useful features. For instance, the best features in the case of SVM were the correlation, average deviation, difference variance, and standard deviation. Finally, the classification was performed using three classifiers, i.e., SVM, KNN, and naïve Bayes, with classification accuracies of 77%, 70%, and 68%, respectively. Sousa *et al.* [87] further extended PneumoCAD by proposing five ML classifiers such as KNN, naïve Bayes, multilayer perceptron, decision tree, and SVM and three-dimensionality reduction techniques (sequential forward selection, principal component analysis (PCA), and kernel PCA (KPCA). The authors used the dataset and features given in a previous study [82] and performed a 10-fold cross-validation test on each classifier. For each combination, the authors evaluated the pneumonia detection performance of their algorithm on CXRs. The highest accuracy (96%) was achieved when using the 13-dimensional features produced by KPCA and classified by naïve Bayes.

Depeursinge *et al.* [88] compared the performances of five ML classifiers (naïve Bayes, KNN, J48 decision trees, multilayer perceptrons, and SVM) for tissue segmentation from high-resolution CT images. Their self-collected dataset contained 843 radiographs covering six pathologies (emphysema, fibrosis, micronodules, ground glass, macro-nodules, and healthy) and their annotated regions of interest (ROIs). From the ROIs, the authors extracted gray-level histograms, air components, and quincunx-wavelet frame coefficients with B-spline wavelets. They extracted 39 texture-based attributes and optimized the parameters of each classifier by grid searching. The performance was evaluated based on McNemar's statistical tests and the accuracy measure. The obtained features were then classified by various ML classifiers (NB, KNN, SVM, MLP, and J48). The SVM achieved the best values of each metric, with a mean testing accuracy of 88.3% and an accuracy of 96.4% for binary classification (pathology vs. normal). However, the detection accuracy of macro-nodules was only 42.5%.

Noor *et al.* [89] proposed a wavelet transform-based method that can be used to extract features from the CXR images obtained from the Institute of Respiratory Medicine, Malaysia. They applied the modified principal component (ModPC) method for dimensional reduction and distinguishing normal images from pneumonia images. Although the proposed method can successfully classify normal and pneumonia samples, the wavelet transform cannot preserve the phase information. Ambita *et al.* [90] proposed locally adaptive regression kernel descriptors for feature extraction, PCA for dimensional reduction, and an SVM classifier for pneumonia detection from CXR images [2]. The quality of the input images was improved using image preprocessing techniques such as Gaussian filters, power law transformation, and Otsu thresholding. The precision and recall were 0.95 and 0.96, respectively, on a sample size of 1000 and 0.98 (both measures) on a sample size of 400. The accuracy decreased with the increasing sample size; furthermore, the authors did not follow the standard data division [2].

Khatri *et al.* [91] proposed EarthMovers Distance (EMD) for pneumonia detection from CXR images. The authors preprocessed the images by cropping the lung regions, normalized the intensity, and calculated the EMD difference to distinguish the pneumonia samples from non-pneumonia samples. The classification accuracy was 83.3%, but EMD is sensitive to rotation, scaling, and intensity variations, and cannot feasibly handle large amounts of data.

Varela-Santos and Patricia Melin [92] proposed a ML-based approach for pneumonia classification from Chest X-rays [1]. The authors first used image pre-processing by extracting the lung regions (by resizing the image) followed by the image enhancement technique using histogram equalization. For feature extraction and classification, the authors employed Gray Co-Occurrence Matrix (GLCM) and ANN, respectively. Experimental results show that the mean accuracy of 92% using the neural network with the Gradient Descent, Adaptive Learning, and Momentum Algorithm. However, it uses only binary classification on the chest X-rays 14 dataset, further, the accuracy still needs improvement. Similar work is done by Pavithra et.al. [93], who employed power law transformation and median filter for image preprocessing followed by Gabor filter for feature extraction. Finally, ANN is used for the classification of Chest X-rays and which achieved an accuracy of 94.82%. However, this lack of details in the work presented to further evaluate their performance.

Yao *et al.* [94] proposed a ML-based automated system based on which five diseases, including pneumonia, can be identified. They collocated 40 CTs and employed SVM classifiers with texture-analysis ability. The authors initially segmented the lung regions using region growing and dynamic programming. Subsequently, they divided them into 16×16 texture blocks and extracted a feature vector containing 25 texture features, including mean, variance, energy, and correlation. The co-occurrence matrix was also obtained. Finally, the CT images were classified using an SVM. The authors achieved an accuracy of 80% during the pneumonia detection task. However, the proposed system struggled to distinguish bacterial pneumonia from H1N1 influenza.

Self *et al.* [95] compared the diagnostic performance of CXRs with CT scans for lung opacities (infiltrate, consolidation, pneumonia, or bronchopneumonia) in emergency

department patients. The authors used a cohort of 4237 subjects who had chest CT scans and 3423 similar subjects that underwent CXRs. CT scans and CXRs interpreted by radiologists show that sensitivity and positive predictive rate of CXR in the detection of pulmonary opacity was only 43.5% and 26.9%, respectively. Thus, the authors show that CXR could lead to significant misdiagnosis of pneumonia.

Naydenova *et al.* [96] employed ML methods in a novel diagnostic process. These methods consider numerous clinical measurements that can be performed using cheap and easily operable devices as input. To evaluate their findings, the authors collected a dataset of 1093 Gambian children aged 2–59 months, including 777 pneumonia samples and 316 healthy samples. They applied preprocessing techniques to data with missing values and extracted 47 clinical characteristic features. The useful features were extracted using nine feature selection techniques (correlation coefficients, minimum redundancy, Gram–Schmidt, relief, orthogonalization, selection operator, least angle shrinkage, elastic net, and sparse linear discriminant analysis). The authors divided the data into ten equals sections. Each feature selection technique was iterated 50 times in each section to achieve unbiased results. The ten most relevant features were then obtained by majority voting. Finally, the extracted features were learned by the SVM, logistic regression, and RF ML procedures. The authors also employed a predictive algorithm with four predictive attributes (temperature, heart rate, oxygen saturation, and respiratory rate). The authors reported a sensitivity of 96.6%, a specificity of 96.4%, and an AUC of 97.8%. However, these individual features must be combined to generalize the pneumonia detection system beyond the control cases (which are inapplicable to clinical scenarios). Another study done by Antin *et al.* [97] proposed a supervised learning technique for pneumonia detection in CXR images and used the NIH database [1] of 112120 radiographs collected from 30805 patients. The images are pre-annotated with ''one'', ''more than one'' or ''no'' diseases by expert radiologists. The authors selected 500 random samples and employed t-SNE (TDistributed Stochastic Neighbor Embedding) [98] for visualization. The k-means clustering was performed to label each point from t-SNE. Further, the authors used classification on Chest X-ray performance by using logistic regression classifier, DenseNet but the accuracy of the proposed method was poor i.e. AUC = 0.60.

B. DEEP LEARNING

A CNN comprises three main parts: the input images, an in-depth feature extractor, and a classifier. Through its multiple layers, the feature extractor automatically learns the essential features from the raw input (or preprocessed) images. The learned features are passed to a classifier such as SoftMax, where they are classified based on the learned features. A CNN contains several layers, i.e., a convolutional layer, a pooling layer, an activation layer, a dropout layer, and a classifier (e.g., SoftMax) [99]. A CNN can be built from scratch [5] by employing an existing pre-trained network without retraining or by fine-tuning a pre-trained network on a target dataset.

1) TRANSFER LEARNING (OF PRETRAINED NETWORKS)

Wang *et al.* [1] collected the largest CXR dataset, which contained 108, 948 images of eight (currently 14) diseases obtained from 32, 717 patients. The authors classified multiple pathologies in the CXRs using various pre-trained models (AlexNet [16], GoogleNet, VGG16, and ResNet). The ResNet model with AUC = 0.63 provided higher accuracy when compared with those provided by the remaining pre-trained networks. Although the accuracy of their proposed system is not promising, the authors provided a baseline dataset for the research community. To improve the classification performance on this dataset, Yao *et al.* [8] proposed a two-stage end-to-end model with a DCNN encoder and an RNN (long short-term memory) decoder that can be used to predict the labels of the pathologies in the Chest X-ray14 data. The DCNN was a modified DenseNet trained from scratch on the Chest X-ray14 dataset. The classification performance of the proposed method improved from that of a previous study [1], achieving an AUC of 0.71. Rajpurkar *et al.* [26] proposed Chexnet, which is a pre-trained 121-layer DenseNet trained on the Chest X-ray14 dataset, to further improve the pneumonia detection performance of the previous methods [1], [8]. The proposed algorithm improved the accuracy to 76% and enhanced the F1 metric. For pneumonia detection, the dataset was divided into training (98, 637 images), validation (6351 images), and testing (420 images) sets. The dataset was split into training, validation, and test sets in proportions of 70%, 10%, and 20%, respectively, to ensure that the algorithm was applied to all 14 diseases in the Chest X-ray14 dataset. The authors compared the performance of the proposed Chexnet model with the assessments of the experienced radiologists. Chexnet outperformed the manual assessments, with an F1 score of 0.435 (versus 0.387 in the manual assessments). The authors also visualized the lung nodules in CXRs using a heat map. However, the AUC of pneumonia detection was only 0.76, indicating the need for improvement.

Kermany *et al.* [2] proposed a transfer learning algorithm for retinal diagnosis via optical coherence tomography (OCT). Their method, which was based on the transfer learning of retinal OCT images, achieved state-of-the-art performance. The same transfer learning system was applied to a pediatric CXR dataset for pneumonia classification [2]. When comparing normal and pneumonia images, the accuracy, specificity, and sensitivity of this system were 92.8%, 0.90%, and 0.93%, respectively. Meanwhile, the accuracy, specificity, and sensitivity associated with the detection of bacterial versus viral pneumonia were 90.7%, 0.88, and 0.90, respectively. Although the authors created a useful baseline by creating the pediatric CXR dataset and employed the transfer learning model, the accuracy of the proposed system was not promising. In another work done by Ayan *et al.* [100] compared the performances of two pre-trained networks

(VGG16 and Xception) on the same dataset [2]. For VGG-16 network, the accuracy, sensitivity, and specificity of distinguishing normal from pneumonia images were 0.87, 0.82, and 0.91 while 0.82, 0.85, and 0.76, for Xception. The experimental results show that the VGG16 network is better in detecting normal cases while the Xception network is better in detecting pneumonia cases. Thakur *et al.* [101] proposed a pre-trained VGG16 network for pneumonia detection on the CXRs dataset [2]. Experimental results show an accuracy of 90.54%, precision, recall, and F-1 score of 87.9%, 98.7%, and 92.9%, respectively. However, the performance of the proposed models [100], [101] was not promising.

Li *et al.* [102] proposed the attention-guided squeeze-and-excitation network (SENet) architecture [103] based on the transfer learning-based model for pneumonia detection. This method erases the pneumonia region in the CXR and replaces it with non-pneumonia pixels to increase the attention of the CNN. The lung regions are segmented by U-Net and classified using the proposed SENet. The experimental results were obtained at different threshold levels (T). When T (IoU) = 0.3, the precision and recall were 0.611 and 0.835, respectively, and the accuracy (model score) and false-positive rate were 0.262 and 0.194, respectively. Jain *et al.* [104] proposed several DL-based models for pneumonia detection from CXRs: a customized CNN, another customized CNN, VGG16, VGG19, ResNet50, and Inception-v3, with validation accuracies of 85.26%, 92.31%, 87.28%, 88.46%, 77.56%, and 70.99%, respectively. The experimental results show that the customized CNN outperformed the pre-trained model, with an accuracy of 92.3% and a recall of 98%. However, the precision of both the studies [102], [104] was low. Similar work done by Ureta *et al.* proposed the combination of three different pre-trained networks (Resnet-18, VGG16, and DenseNet) for pneumonia classification on chest X-rays dataset [2]. The fully connected layer was replaced with a linear classifier with a single output [ureta] thus obtaining an accuracy of 96.6% with specificity and sensitivity of 97.8% and 96% respectively. The authors performed experiments on Chest X-rays 14 dataset for binary classification (pneumonia vs non-pneumonia) to evaluate the performance on an external dataset. However, the accuracy was only 51%.

Chhikara *et al.* [105] applied a preprocessing technique followed by a pre-trained Inceptionv3 model for pneumonia detection from pediatric CXRs [2]. Pre-processing involved median filter for noise reduction, gamma correction: to display the image correctly, Contrast Limited Adaptive Histogram Equalization (CLAHE) for image enhancement, and JPEG compression to compress the image size without compromising the image quality. Experimental results performed using the pre-trained incpetionV3 model show an accuracy of 90.01%, precision, recall, and F-1 score of 0.907, 0.957, and 0.932, respectively. Schwyzer [106] proposed a DL-based ResNet-34 model to classify pulmonary infections in CXR dose-equivalent CT. The experimental results show that the specificities for standard dose and reduced dose were 93.8% and 93.3%, respectively. However, the sensitivity was lower for the reduced dose (71.0%) than that for the standard dose (82.9%).

Togacar *et al.* [107] applied three DL methods (Alexnet, VGG16, and VGG19) in a deep feature learning model that extracts CXR images at the eighth fully connected layer. The 1000 features obtained from each DL model were reduced to 100 by considering minimum redundancy and maximum relevance. Finally, the authors combined the features obtained from Alextnet, VGG16, and VGG19, obtaining a feature vector size of 300, which achieved a classification accuracy of 99% when LDA classifier was used on a CXR dataset [2]. Although promising results were obtained using the proposed method, the authors did not apply the standard data division presented previously [2]; instead, they balanced only the normal images by data augmentation. Similarly, Liang *et al.* [108] proposed 49 layers of the convolutional layer residual structure that can overcome the overfitting for pneumonia detection on pediatric CXR datasets [2]. The network was pre-trained on Chest X-rays14 [1] instead of ImageNet, which is a major advantage. The experimental results show that the accuracy, precision, and recall of the proposed method were 0.905, 0.891, and 0.967, respectively.

Zech *et al.* [109] proposed pre-trained Densnet-1212 and ResNet-50 to evaluate the generalizability of DL to external site data. Three datasets, i.e., Chest X-ray-14 (NIH), Mount Sinai Hospital (MSH; 48,915 radiographs), and Indiana University Network for Patient Care (IU; 7,470 radiographs), were used to evaluate the performance of the proposed model. The authors used only frontal radiographs for experimentation and used 100% of IU data for the external test because the data did not contain patient identifiers. NIH and MSH contain patient identifiers; therefore, they were separated into training (70%), validation (20%), and test (10%) sets. The training was performed on three dataset combinations: MSH only, NIH only, and combined MSH–NIH. The extensive experiments conducted indicated that the accuracy, sensitivity, specificity, and AUC were 0.732, 0.950, 0.706, and 0.931, respectively, when performing the internal test using the MSH–NIH model. However, during the external test (trained on MSH–NIH, and test on IU), the accuracy, sensitivity, specificity, and AUC were 0.238, 0.974, 0.230, and 0.815, respectively. Thus, the experimental results revealed that the model trained on an internal dataset could not provide promising results when considering an external set.

Putha *et al.* [110] trained a DL system on the largest available dataset, which contained 2.3 million CXR images with different pathologies. The DL algorithm (ResNet) was validated using 100, 000 of these images (CQ100k) and was further evaluated using 2000 images (CQ2000) collected from three hospitals in India. The pathologies were obtained from the images and their associated reports using a natural language programming (NLP) algorithm. 1930 X-ray images were independently verified by experienced radiologists to evaluate the correctness of the NLP algorithm. The accuracy of pathology classification on the CQ2000 dataset ranged

from 89.0% to 99.0%. The same results were obtained on the CQ100k dataset, but the accuracy of detecting normal X-ray images from this dataset reduced to 86.0%. This study did not validate the detection of other pathologies, such as rib fractures and pneumothorax, and did not localize the lung legions. Furthermore, its accuracy must be improved. This study is a useful baseline for conducting further research.

Yee *et al.* [111] employed pre-trained Inception V3 as feature extractor and three other ML classifiers: KNN, SVM, and neural network for pneumonia classification on CXR dataset [2]. Experimental results were performed to show that among the three classifiers, a maximum ACU of 93% was achieved when SVM was used as a classifier, however, the sensitivity was only 83.3%. Asnaoui *et al.* [112] proposed comparison of different pre-trained DL models: VGG16, VGG19, InceptionV3, Xception, DenseNet201, MobileNet_v2, ResNet-50, Inception_Resnet_V2 and customized CNN for pneumonia classification using chest X-rays. Experimental results show that Inception_Resnet_v2, mobileNet_v2 gave an accuracy of around 96% while the other algorithms did not perform well with accuracy, not over 85%.

2) DL AND NONIMAGE FEATURES

Baltruschat *et al.* [28] supplemented the CXR dataset with nonimage features (patient age, gender, and acquisition type). The authors applied the transferred learning technique (ResNet) with and without fine-tuning and a CNN trained from scratch. The data were split in two ways. In the first data split, patients with multiple follow-up records were assigned to a single subset. This splitting enabled a vast range of patient numbers, for example, 22420 images can be split into 5817 patients (split 2) or 6245 patients (split 5). The second dataset split was described by Wang *et al.* [1]. The former split improved the accuracy of classifying pneumonia by transferred learning from 75.3% to 76.7% (for image dimensions of 448×448 pixels). The latter split reduced the AUC result of detecting pneumonia with ResNet-38 to 0.71 but achieved superior results with respect to the five remaining diseases in the CXR dataset.

Bar *et al.* [113] concatenated the Decaf [114] features with the low-level features (the GIST and bag-of-words features) and classified lung pathologies using a feature selection method. For feature selection, they selected the 5000 most significant features from among the $18,920$ features obtained by concatenating the Decaf and low-level features. The authors demonstrated the effectiveness of using the low-level features with the concatenation of DL features for pathology classification. They also proposed that selecting only the intelligent and useful features from the whole feature set would improve the classification. Although the effectiveness of their method was clarified [113], the classification accuracy should be improved.

Er *et al.* [115] performed a comparative study of different neural networks in chest disease diagnosis. The dataset was obtained from Diyarbakir Chest Diseases Hospital located in the southeast of Turkey, which included the epicrisis reports of patients with tuberculosis (50 samples), chronic obstructive pulmonary disease (71), pneumonia (60), asthma (44), and lung cancer (32) as well as healthy samples (100). The authors collected 357 samples and analyzed 38 nonimage features, including complaints of coughing, weakness, chest aches, and high body temperature. The experimental results showed that the probabilistic neural network (PNN) [116] achieved the highest average accuracy (92.16%) but low accuracy for pneumonia detection (88.33%). The most successful network in the pneumonia analysis (60 samples) was the multilayer neural network with the Levenberg–Marquart architecture [117] (one hidden layer), which achieved an accuracy of 91.67%, but the average accuracy for all the pathologies was 90.48%. Furthermore, no other metrics were presented in their study.

3) CUSTOMIZED CNN

Stephan *et al.* [4] proposed a CNN with four convolution layers, which were trained from scratch using the pediatric CXR dataset proposed by Kermany *et al.* [2]. The data split was performed differently from that in the study of [2] by assigning 2134 images to the validation set and increasing the training dataset via augmentation techniques. The authors reported an accuracy of 93.7% by building a less computationally complex DL model on the validation set; however, this performance could not be consolidated using other metrics. This lack of further verification and performance in case of a nonstandard division of the dataset is considered a significant drawback of their study. Raheel [5] proposed an 18-layer customized DCNN and trained it on the pediatric CXR dataset. They followed the dataset division of the original authors [16]. The experimental result indicated that the accuracy of the proposed system (94.3%)was1.6% greater than that of Kermany *et al.* [2] and that the sensitivity was 99%. However, the specificity was only 86%.

Verma *et al.* [118] proposed a customized CNN to classify lung nodule in CXR images into three classes: pulmonary TB, viral pneumonia, and bacterial pneumonia. The authors [118] preprocessed the images via data augmentation to avoid overfitting. The experimental results indicated an overall accuracy of 99.01%, but the experimental details were missing from the study. Omar [119] similarly proposed a customized CNN model for pneumonia detection from CXRs [2] and achieved an overall accuracy of 87.65%. However, Omar's study omitted other performance metrics, including precision and recall. Bhatt *et al.* [120] proposed a computationally cheap nine-layer customized CNN for pneumonia classification from CXRs [2]. After randomly dividing the data into training and test sets, the authors achieved an accuracy of 96.18%, a sensitivity of 98.16%, and a specificity of 91.29% [98]. These authors did not follow the standard data division proposed in a previous study [2]; in addition, the specificity is not promising.

Wu *et al.* [121] proposed an adaptive median filter CNN and RF (ACNN–RF) for pneumonia detection using

CXRs [2]. Initially, they removed noise from the images via median filtering. The features in the cleaned images were extracted by a CNN and were classified by RF based on GridSearchCV. The experimental results indicated that the proposed method achieved an accuracy of 96.9% as well as a precision and recall of 0.90 and 0.95, respectively. Furthermore, the testing time of the proposed system was 625.4 s, whereas that of the CNN model was 921.2 s. However, the proposed model exhibited low precision. In another study conducted by Sarkar *et al.* [122] proposed a preprocessing method followed by a DL model that achieved an accuracy of 98.82% and AUC of 0.99726 on CXRs [2]. The authors used bilateral filtering, the CLAHE algorithm for image enhancement, and a modified DL model using the GAP layer rather than a flattened layer. Although the authors considered promising accuracy, the authors did not follow the standard data division proposed in a previous study [2].

A CNN is perceived as a black box that outputs a performance without relaying the understanding of the process. Such a lack of transparency can adversely affect decision-making. To improve this situation, Rajarman *et al.* [123] proposed a visual explanation of the prediction and activation of CNNs. Two models (VGG16 and a CNN built from scratch) were evaluated, and their results were presented visually. The lung boundaries were detected using an atlas-based detection algorithm [84], and the classification was performed by a DCNN. The VGG16 achieved a better learning and optimization outcome when compared with that obtained using the customized CNN. The VGG16 improved the accuracy of distinguishing between pneumonia and normal images from 92.8% in Kermany *et al.* [2] to 96.2%. Meanwhile, the accuracy of distinguishing between viral and bacterial pneumonia improved from 90.7% to 93.6%. Although promising results were achieved, the specificity of classifying viral and bacterial pneumonia by the VGG16 network was only 85.9%.

Abiyev *et al.* [124] proposed customized CNN, competitive neural networks (CpNNs) with unsupervised learning, and backpropagation neural networks (BPNNs) with supervised learning for pathology detection from chest X-rays. Multiple experiments performed show that CpNN converges faster than CNN; however, if the accuracy is the primary concern then the accuracy of the proposed CNN is higher (92.4%) than CpNN (80.04%), BPNN, (89.57%) VGG16 (86%), VGG19 (92%), and CNN with GIST (92%). However, the accuracy still needs improvements, further, other performance metrics such as sensitivity and specificity were not given.

4) SEGMENTATION BY FCN AND RCNN

Gu *et al.* [125] proposed a two-step method that can classify bacterial and viral pneumonia from chest radiographs. In the first step, the lung regions were extracted from the X-ray images using an FCN. In the second step, the lung regions were classified into pediatric viral or bacterial pneumonia using a DL network with handcrafted features. The authors used the publicly available MC [32] and JSRT [34] datasets for segmentation and the data from the Guangzhou Women and Children's Medical Center [2] for binary classification. Feature extraction by transfer learning achieved higher accuracy ($80.0\% \pm 2.02\%$) when compared with those achieved by other feature extraction techniques. The performance was further improved using an ensemble of features (gray-level co-occurrence matrix, wavelets, a histogram of oriented gradients, and DCNN features), although the improvement was slight (AUC $= 82.3\% \pm 0.14\%$). The authors presented that segmentation and utilization of the handcrafted features will increase the accuracy; however, the method must be further improved because the sensitivity and specificity of using ensemble features were 0.556 and 0.926, respectively.

Mask–RCNN [126] is a deep neural network designed for tasks such as segmentation. Jaiswal *et al.* [127] proposed mask–RCNN with ResNet101 (and ResNet50) as a backbone detector for pneumonia detection and trained it on the dataset of the Radiological Society of North America [128], which contained almost $30,000$ annotated X-ray images. Mask–RCNN outperformed several object detection techniques, such as YOLO3 and U-Net when evaluated on the test set (the Society of Thoracic Radiology dataset). The proposed model predicted the bounding box of each CXR, its label, and its masks with the respective class. The performance of the proposed model can be further improved by improving the architecture of the proposed model; however, this would necessitate the adjustment of the hyperparameters. Wang *et al.* [129] presented cooperative CNN (co-CNN) for pneumonia detection and localization on CXRs [128]. The authors used U-Net for lung segmentation, co-CNN for pneumonia classification, and localization. Various experiments performed show a prediction score of 0.2409, sensitivity and specificity of 0.89 and 0.78, respectively which needs improvement.

The method of Li Zhe *et al.* [130] identifies and localizes disease lesions in CXRs [1] under a limited amount of supervision. The proposed method accurately visualizes the location of the disease in the x-ray image, improving the disease interpretation. First, the image input is processed by Preact–Resnet (Resnet-v2) [91], which extracts the feature tensors of size $h = h/32, w = w/32$ and $c = 2048$. Here, h, w, and c are the height, width, and number of channels of the input image, respectively. The image is then divided into a $P \times P$ grid for predicting K possible disease types. As the recognition network, they applied FCN [92]. The image model was Resnet-v2-50, and the patch-slice size was selected from $\{12, 16, 20\}$. The CXR dataset contains 984 labeled bounding boxes for 880 images; the reaming $111,240$ images are unannotated. The first evaluation was performed by five-fold cross-validation on 70% of the unannotated images and 70% of the annotated images. The second evaluation checked the effectiveness of the supervision provided by the bounding boxes. To this end, the proportion of unannotated bounding boxes was reduced from 80% to 0%. The authors showed that in some cases, the result of 40% (44496) unannotated images with 80% (704) annotated images outperformed the result

of 80% (88; 892) unannotated images. The authors successfully show that the proposed model significantly outperform [1] in identification and localization of multiple pathologies with limited number of annotated data, however, in terms of pneumonia the performance did not improve from [1] i.e. the AUC was 0.60 in the proposed work however, it was 0.63 in [1].

5) ENSEMBLE METHODS

Chouhan *et al.* [131] proposed an ensemble of different state-of-the-art DL algorithms (Alexnet, Inception v3, ResNet, GoogleNet, and DenseNet-121) for pneumonia detection from pediatric CXRs. Multiple experiments confirmed that the ensemble method outperformed the remaining methods [2], achieving an accuracy of 96.4% and a sensitivity of 99.0% (versus an average accuracy and sensitivity of 92.8% and 93.2%, respectively, in previous methods). The ensemble method can be applied to improve model accuracy. However, the ensemble method is computationally expensive because it required multiple architectures to be combined. Further, the precision (93.2%) of the proposed model was not appreciably different from those of the previous methods, which is another disadvantage of the study [131].

Vijendran *et al.* [132] proposed a multilayered method called online sequential extreme learning machines (OSELM) for pneumonia detection from the Chest X-ray and MNIST datasets. The multilayered OSELM achieved an accuracy of 91.7% on the Chest X-rays dataset, which exceeded those obtained using the SVM (86.5%) and conventional extreme learning machines (89.8%). Although the traditional CNN outperformed OSLEM in terms of accuracy (92%), the testing time was 2.434 s in CNN but only 1.2 s in the multilayered OSELM. Thus, the accuracy of multilayered OSELM was similar to that of traditional CNN but within a much shorter runtime. However, the authors did not provide any additional performance metrics to evaluate the performance of the compared methods.

Islam *et al.* [133] proposed ensemble DCCN models (AlexNet, VGG16, VGG19, ResNet-50, ResNet-101, and ResNet-152) that can be used to accurately detect abnormalities (cardiomegaly and tuberculosis) in CXRs. In the Indiana, JSRT, and Shenzhen datasets, the accuracies of cardiomegaly and tuberculosis detection using the ensemble DCCNs were 93.0% and 90%, respectively, which was 17% greater than that associated with the rule-based methods. Although the proposed ensemble model achieved promising accuracy, its sensitivity for tuberculosis detection was only 88% and its specificity for cardiomegaly detection was 92%. These performances should be improved.

Sirazitdinov *et al.* [134] proposed an ensemble of RatinaNet and mask–R-CNN for pneumonia detection and localization in the dataset of 26,684 images from Kaggle Pneumonia [128]. The precision, recall, and F1 score of the ensemble were 0.75, 0.79, and 0.77, respectively. The experimental results reveal that the performance of the proposed ensemble model improved when compared with those of the state-of-the-art models (the F1 score for ResNet-50 was 0.68 and that for DenseNet-121 was 0.731). However, further, improvement is required to minimize false positives and false negatives.

Ko *et al.* [135] proposed a weighted ensemble of object detection models (mask–R-CNN and RetinaNet) for feature extraction from CXRs. The weighted ensemble of two mask R-CNNs and three RetinaNet models delivered the best performance, with a mean average precision of 0.21746 (versus 0.19984 associated with the individual model). However, no comparisons with the existing works were reported.

Hashmi *et al.* [136] proposed a weighted classifier comprising different DL models (ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3), which can be used to identify pneumonia from CXRs [2]. The experimental results confirmed a remarkable accuracy of 98.43% and precision and recall values of 98.26% and 99.0%, respectively. Although this method outperformed all the existing methods, the testing images used slightly differed from those of Chouhan *et al.* [131]. Specifically, Hashmi *et al.* [115] used 700 test images (400 pneumonia and 300 normal), whereas Chouhan *et al.* [131] used 624 test images (390 pneumonia and 324 normal). The data division of Chouhan *et al.* [131] was similar to that of the baseline study [2]. Yan *et al.* proposed multi-scale CNN (MSCNN) for pneumonia classification on chest CT scans. The proposed MSCNN contains an ensemble of a multiscale spatial pyramid (MSSP) for image decomposition and then fed to MSCNN which is pretrained EfficientNetB0 for classification. The authors collected 416 CT scans samples from 206 COVID-19 patients, 412 CT scans of pneumonia collected from 412 pneumonia patients. Experimental results were performed to show that the proposed AI system achieved promising results with AUC of 93.4% on scan level while AUC of 96.2% was achieved on slice level.

Yu *et al.* 2021 [137] achieved a promising accuracy of 98.7% for pneumonia classification from CXR [2] when a DL-based CGNet model was employed. The authors first used a pre-trained DL-based model for feature extraction and then the extracted feature were reconstructed based on features correlation using Euclidean distance. Finally, the reconstructed features were used for classification using ANN which achieved an accuracy of 98.7%, sensitivity, and specificity of 0.99 and 0.98, respectively.

6) ANALYSES OF PARTIAL DATASETS WITH NONSTANDARD DIVISIONS

Some researchers have performed nonstandard divisions of partial datasets. Rahmat *et al.* [138] employed Faster RCNN, based on which pathological and normal CXRs can be classified. The authors selected a subset of 200 pathological and normal images from the Chest X-ray14 dataset. The accuracy of Faster RCNN was 62.0%, which was greater than the judgment accuracies of a general practitioner and a medical student but still very low because the data were highly imbalanced. The specificity (54.39%) was also low,

but Faster RCNN is useful for localizing the pathologies (including pneumonia) in CXRs.

Chandra *et al.* [139] employed ML methods on a subset of the Chest X-ray 14 [1] dataset. The authors initially segmented the lung regions and then extracted eight statistical features (mean, variance, standard deviation, skewness, kurtosis, smoothness, uniformity, and entropy) from the lung ROI. The extracted features were allocated to multiple classifiers: RF, MLP, logistics regression, sequential minimal optimization, and classification via regression. In the experimental evaluation, logistic regression outperformed the remaining classifiers with accuracy, sensitivity, and specificity values of 96.63%, 93.68%, and 97.57%, respectively. However, the method was evaluated using only 412 images (206 normal and 206 pneumonia).

Varshni *et al.* [140] applied six DL algorithms (Xception, VGG16, VGG-19, ResNet-50, DenseNet-121, and DenseNet-169) as the feature extractors and combined them with multiple ML classifiers (RF, KNN, Naive Bayes, and SVM) for pneumonia detection in CXRs. Because the Chest X-ray14 dataset [1] contains 1431 pneumonia images, they selected 1431 normal images to balance the dataset for binary classification. The AUC on the 573-image test dataset was 0.8002 in a series of experiments with the appropriate parameters of the SVM classifier in DenseNet169. However, there were several limitations: only binary classification was used in this study. The performance of pneumonia detection was not promising, and the applied pre-trained networks were computationally expensive.

7) PNEUMONIA AND COVID-19

The recent outbreak of the COVID-19 disease and its rapid spread worldwide poses a global health challenge [141]. This spread to the lungs, causing pneumonia. As a result, many researchers have focused on developing algorithms that can identify pneumonia.

Mahmud *et al.* [142] proposed deep CNN-based CovXNet for pneumonia and COVID-19 detection. The proposed CovXNet extracts features from CXRs via depth-wise convolution with varying dilution rates. Multiple forms of CovXNets were used for different resolutions of CXR images followed by a stacking algorithm for optimization. Gradient-based discriminative localization employed in the localization of abnormal lung regions. Experiments were performed to verify the result, and an accuracy of 96.9% was achieved in classifying viral pneumonia and COVID-19, 94.7 for bacterial pneumonia and COVID-19, and 97.4% for COVID and normal CXRs. However, in the classification among COVID/normal/viral/bacterial pneumonia and COVID-19/Viral Pneumonia/Bacterial pneumonia, the accuracy was only 90.2% and 89.6%, respectively. Rahimzada *et al.* [143] proposed concatenation of pretrained Xception and ResNet50V2 for pneumonia and covid-19 detection using CXRs [144], [128]. Experimental results show that concatenation of Xception and ResNet50V2 resulted in an overall accuracy of 91%, with promising accuracy of 99% for COVID-19. However, the sensitivity for COVID-19 and pneumonia was only 8.53 and 8.35%, respectively. Furthermore, poor precision (35.27%) and low specificity (88.09%) were achieved for COVID-19 and normal samples, respectively.

Jin *et al.* 2021 [145] proposed a three-step DL-based model for pneumonia and COVID-19 classification. They first used AlexNet to extract features from 1743 CXRs [1], [144], followed by ReliefF algorithm [146], [147] which ranked the features based on their importance. Finally, the top n features selected were used for classification in conjunction with the SVM classifier. The experimental results showed that the proposed algorithm can achieve an accuracy of 98.6% using only 40 features. The proposed model was faster as compared to other state-of-the-art models.

Dey *et al.* 2021 [148] employed concatenated features obtained from the state-of-the-art DL-based models and traditional ML algorithms. The authors extracted Gray-Level Co-Occurrence Matrix, discrete and complex wavelet transform (DWT, and CWT, respectively) from CXRs [2] and employed PCA for feature selection. They also extracted features using a VGG19 based DL algorithm which was concatenated with handcrafted features as an input to ML classifiers (KNN, SVM, RF, and DT). Experiments showed that the use of combined features and the RF classifier can result in an accuracy of 97.8%, with precision and recall of 0.94 and 0.96, respectively. Although the proposed approach showed promising results compared to those obtained using only the DL model, it is computationally complex [145].

Das *et al.* 2021 [149] evaluated the performance of three DL-based CNN, VGG16, and ResNet50 models for pneumonia and COVID-19 classification from CXRs. They obtained 2,686 CXRs (1,345 viral pneumonia and 1,341 normal) from [2] and 219 CXRs from [144]. The experimental results showed that VGG16 achieved a better classification accuracy of 97.6% with precision and recall of 0.965 and 0.965, respectively as compared to Resnet50 and CNN that achieved accuracy results of 94.5% and 93.6%, respectively.

Rajpal *et al.* 2021 [150] proposed a transfer learning-based model for pneumonia classification. They obtained a dataset from different sources such as [144], [151], [152] and [153]. The dataset contained 1,560 CXRs from pneumonia, COVID-19, and normal classes with 520 CXRs in each class. A feature vector of size 252 was obtained using the Resnet50 model, which was reduced to 64 features using PCA, followed by dense and dropout layers. A final feature vector of size 16 was obtained using the Resnet50 model. The authors also used handcrafted feature vectors of size 252 (14 FFT, 112 DWT, 1126 texture features). The experimental work on the combined feature vector resulted in a classification accuracy of 97% with an average sensitivity of 0.978.

Tuncer *et al.* 2021 [154] proposed a fuzzy-transform model (F-transform) [155] to obtain a novel fuzzy tree for pneumonia classification. They used a dataset obtained from online sources which contained 435 CXRs (150 normal, 135 COVID-19, and 150 pneumonia). They first obtained

three levels of fuzzy trees and a multi-kernel local binary pattern for features extraction. Iterative neighborhood component (INCA) was employed to select only appropriate features based on the correlation among the features and a feature vector of size 616 was achieved. Finally, classifiers such as DT, LD, SVM, and KNN were used for classification. SVM achieved better performance with an accuracy of 97.01%. However, the accuracy in detecting the pneumonia class was 94.8%.

Danilov *et al.* 2021 [156] evaluated the performance of four DL models for pneumonia classification from 2631 CXRs: EfficientNet B1, B3, VGG16 and MobileNetV2 [2], [144], [151]–[153]. After fine-tuning, the proposed models achieved an accuracy of only 78% for pneumonia classification. To further improve the classification performance, the authors employed Grad-CAM [157], which provided guided attention to the lung area based on U-Net segmentation. The Grad-CAM technique improved the classification performance by 1%.

TABLE 2. Calculation of the performance metrics.

Actual Label	Prediction (Disease)	Prediction (Normal)
Positive (Disease)	TP	FN
Negative (Normal)	FP	TN

VI. PERFORMANCE METRICS

The performance of a pneumonia detection system was evaluated based on various performance metrics. The accuracy metric determines the correctness of the identified instances in both classes of binary classification. This measure must be supplemented by other metrics such as precision, recall, F1 score, and AUC. The precision, recall, and F1 score are given by Eqs. (3), (4), and (5), respectively. Table 2 shows the structure of a confusion matrix.

$$Precision = TP/(TP + FP), \tag{3}$$

$$Recall = \frac{TP}{(TP + FN)}, \tag{4}$$

$$F1\ score = \frac{2\,(Precision \times Recall)}{(Precsion + Recall)}. \tag{5}$$

VII. COMPARATIVE ANALYSIS AND DISCUSSION

This section compares and briefly discusses the state-of-the-art approaches and datasets used for pneumonia identification from CXRs. The evaluation is mainly focused on ensuring robustness and usability.

It can be seen from Table 3 that there are multiple datasets in the literature for pneumonia detection. The most famous and significant datasets are the MIMIC CXR [30] and Chest X-rays14 [1] datasets, which contain multiple pathologies. However, both these datasets contained fewer examples of pneumonia when compared with other classes, making them highly imbalanced. The pediatric Chest X-ray dataset [2] can be considered as the most relevant dataset because it contained only pneumonia and normal samples. Various techniques [56], [73]–[75], [78], [80], [81] have been proposed to tackle the class imbalance problem, but a small number of researchers have followed these techniques. The authors of the majority of the studies have undersampled the datasets to obtain a fewer number of balanced images. Therefore, it is difficult to improve the acceptance of the above methods in clinical applications.

TABLE 3. Comparison of datasets comprised of chest X-ray W.R.T detection of pneumonia.

Dataset	Statistics	Relevance
Chest X-rays14 dataset [1]	1,08,948 images	Relevant but fewer than 1500 pneumonia cases (1.3%)
Pediatric Chest X-rays [2]	5856 images	All pneumonia cases
MIMIC CXR [30]	3,77,110 images	Relevant but only a few cases of pneumonia: 15,769 in the training set (6.9%)
Indiana dataset [31]	7470 images	40 pneumonia cases
MC dataset [32]	138 images	No pneumonia cases
Shenzhen dataset [32]	662 images	No pneumonia cases
KIT dataset [33]	10,848 images	No pneumonia cases
JSRT dataset [34]	247 images	No pneumonia cases
PneumoCAD dataset [21]	156 images	Relevant but only 78 cases
Depeursinge dataset [65]	843 images	Relevant but only 400 cases
Naydenova [73]	1093 images	Relevant but only 777 cases

Furthermore, the majority of the CXR datasets contained no pneumonia samples [32]–[34]. The composite samples in the present dataset can be exploited as a temporary solution for pneumonia classification, but the combined number of images is only approximately 5,000 which is insufficient for accurate and reliable detection.

Table 4 presents a brief overview of the pneumonia detection efforts. The majority of the efforts achieved comparable accuracies, but the results were evaluated on small scales with unbalanced datasets. An accuracy of 96% has been achieved in the domain of traditional ML algorithms [64] by exploiting NB with dimensionality reduction techniques. The traditional ML algorithms are computationally cheap but require extensive handcrafted techniques to extract useful features from the input data and are also affected by noise. To address these problems, researchers have developed automatic feature-learning techniques such as DL algorithms, which are effective in many medical and image recognition tasks. Various DL-based methods have achieved promising accuracy [2], [4], [5], [26], [84], [85], [103].

The accuracy of DL can be further enhanced using an ensemble modeling approach, wherein the predictions of multiple models are combined to improve the overall classification performance. Although DL has achieved state-of-the-art performance, several limitations remained; in particular, these techniques require computational resources, extensive experimentation, and hyperparameter tuning. Furthermore, in most studies, DL is considered as a black box without visualization results, making it inefficient for real-time use.

TABLE 4. Comparison of the pneumonia detection techniques.

Authors	Method used	Dataset	Advantages	Challenges
Abiyev et al. [124]	CNN, CpNN, and BpNN	Chest X-rays14 [1]	Accuracy = 92.4%, CpNN (80.04%), BPNN, (89.57%) VGG16 (86%), VGG19 (92%), and CNN with GIST (92%), CPNN converges fastest	Accuracy needs improvement, CNN converges slower than CPNN
Ambita et al. [90]	Gaussian filter, power law transformation and Otsu thresholding, locally adaptive regression kernel (LARK) descriptors, PCA, SVM, RF, and DT	Pediatric chest X-rays [2]	Precision and recall = 0.95 and 0.96, respectively, for 1000 data; precision and recall = 0.98 each for 400 data.	Performance decreases with increasing sample size, nonstandard data division.
Antin et al. [97]	K-means clustering	Chest X-rays14 [1]	k-mean for clustering, AUC = 0.60	Nonstandard division of data, AUC is low
Asnaoui et al. [112]	Pre-trained models (VGG16, VGG19, InceptionV3, Xception, DenseNet201, MobileNet_v2, ResNet-50, Inception_Resnet_V2 and customized CNN)	Pediatric chest X-rays [2] and CT	Comparison of multi models preprocessing: intensity normalization and Contrast Limited Adaptive Histogram Equalization (CLAHE). Data augmentation Inception_Resnet_v2, mobileNet_v2 gave an accuracy of around 96%	No comparison with related works. Data division for training and testing not according to related work [2], only pneumonia samples,
Ayan et al. 2019 [100]	Transfer learning of VGG16 and Xception	Pediatric chest X-rays [2]	Accuracy, sensitivity, and specificity were 0.87, 0.82, and 0.91, respectively for VGG16 and 0.82, 0.85, and 0.76, respectively, for Xception	Performance is not promising; needs improvements.
Baltruschat et al. [28]	Transfer learning and CNN from scratch	Chest X-ray dataset [1]	On nonimage features, AUC = 71%	Low AUC for pneumonia detection
Bar et al. [113]	PiCodes features and Decaf	93 X-rays from Sheba Medical center	Feature selection, low-level features, AUC = 0.79	Low AUC for pneumonia detection, evaluated on a small dataset
Bhatt et al. [120]	Nine-layer customized CNN	Pediatric chest X-rays [2]	Accuracy = 96.18%, sensitivity = 98.16%, specificity = 91.29%, Computationally cheap	Poor specificity, accuracy not promising, nonstandard data division [2], limited to binary classification
Chandra et al. [139]	Lung segmentation, RF, MLP, logistics regression, SMO, and classification via regression	Chest X-ray14 [1] subset; 206 pneumonia and 206 normal	Accuracy = 96.63%, sensitivity and specificity = 93.68% and 97.57%, respectively.	Evaluated only on 412 images (206 normal, 206 pneumonia).
Chhikara et al. [105]	Modified InceptionV3, preprocessing (median filtering, gamma correction, CLAHE, JPEG compression)	Pediatric chest X-rays [2]	Accuracy = 90.01%, precision, recall, and F-1 score = 0.907, 0.957, and 0.932, respectively.	Accuracy needs improvement, precision is also not promising.
Chouhan et al. [131]	Ensemble (Alexnet, Inception v3, ResNet, GoogleNet, and DenseNet-121)	Chest X-ray14 [1]	Proposed ensemble model, accuracy = 96.4%, sensitivity = 99.0%	Low precision (93.2%), computationally complex
Danilov et al. [156]	EfficientNet B1, B3, VGG16 and MobileNetV2, Grad-CAM	2631 CXRs from[113] [151] [152] [150] [2]	Accuracy of 78% using guided attention-based TL	Accuracy needs improvement.
Das et al. [149]	CNN, VGG16 and ResNet50	Pediatric chest X-rays [2], COVID [144]	VGG16 achieved an accuracy of 97.6% with precision and recall of 0.965 and 0.965.	Non-standard data division.
Depeursinge et al. [88]	NB, KNN, J48 decision trees, MLP, and SVM	Depeursinge dataset [88]	88.3% accuracy with SVM, 96.4% accuracy for binary classification (pathology vs. normal).	The detection accuracy of macro nodules was only 42.5%.
Dey et al. [148]	GLCM, DWT, CWT, VGG-19, RF , SVM, KNN	Pediatric chest X-rays [2]	VGG19 and handcrafted features with RF achieved 97.8 with precision and recall of 0.94 and 0.96	Computationally complex, non-standard data division.
Er et al. [115]	MLNN, PNN, GRNN, and RBF-NN	Self-generated dataset [44]	Average accuracy = 92.16%	The accuracy of pneumonia detection was only 88.33%, no other performance metrics were evaluated.
Gu et al. [125]	FCN [92] and DCNN	MC [40], JSRT [42], and Chest X-rays 14	Segments the lungs using FCN, accuracy = 82.3%	Low sensitivity and specificity of ensemble features (0.556 and 0.926, respectively)

TABLE 4. *(Continued.)* **Comparison of the pneumonia detection techniques.**

Hashmi et al. [136]	Weighted classifier (ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3), data augmentation	Pediatric chest X-rays [2]	Accuracy = 98.43%, AUC = 99.76	
Islam et al. [133]	Ensemble DCCN (AlexNet, VGG16, VGG19, ResNet-50, ResNet-101, ResNet-152)	MC [40] and JSRT [42]	Accuracy = 93.0% for cardiomegaly detection, 90% for tuberculosis detection, 17% improvement over the rule-based method	Sensitivity for tuberculosis detection was only 88%; specificity for cardiomegaly detection was 92%
Jain et al. [104]	Six DL models (two customized CNNs, VGG16, VGG19, ResNet50, and Inception-v3)	Pediatric chest X-rays [2]	Validation accuracies of 1st and 2nd customized CNNs, VGG16, VGG19, ResNet50, and Inception-v3 for pneumonia detection from CXRs were 85.26%, 92.31%, 87.28%, 88.46%, 77.56%, and 70.99%, respectively	Accuracy not promising, precision was also low.
Jaiswal et al. [127]	Mask–RCNN [81] with ResNet 101	RSNA [128]	Image segmentation combined with image augmentation along with dropout and L2 regularization, Mean score = 0.21	Performance can be further improved
Jin et al. [145]	Alexnet for feature extraction, ReliefF for feature ranking, SVM	Chest X-ray14 [1], Cohen et al. [144]	Accuracy of 98.6% with using only 40 features, total classification time was 5.91 seconds.	only limited amount of images with non-standard data division
Kermany et al. [2]	Transfer learning	Pediatric chest X-rays [2]	Normal versus pneumonia images, the accuracy, specificity, and sensitivity of this system were 92.8%, 0.90, and 0.93, respectively	Accuracy not promising, low specificity
Khatri et al. [91]	Preprocessing and EMD	36 Chest X-rays	Accuracy = 83.33%.	EMD is sensitive to rotation, scaling, and intensity variations. Not feasible for large amounts of data
Ko et al. [135]	Ensemble model (mask R-CNN and RetinaNet	Kaggle pneumonia (RSNA) [128]	mAP improves from 0.19984 to 0.21746	A subset of the dataset, computationally complex
Li et al. [102]	SE-Resent, U-Net for lung segmentation, data augmentation, post-processing	RSNA pneumonia dataset [128]	T (IoU) precision and recall = 0.611 and 0.835, respectively. Accuracy and FPR = 0.262 and 0.194, respectively	Requires manual post-processing, precision needs improvement.
Li et al. [130]	FCN and ResNet-v2-50	Chest X-ray14 [1]	Disease localization with a limited number of annotated data	AUC = 0.60 for pneumonia classification
Liang et al. [108]	49 layers CNN with residual DL architecture	Pediatric chest X-rays [2]	Pretrained on Chest X-rays14, accuracy = 90.05 %, recall = 0.967, precision = 0.891	Low precision
Macedo [158]	Haar wavelet transform, KNN	PneumoCAD dataset [21]	Developing PneumoCAD, Accuracy = 91.75%	Accuracy needs improvement
Mahmud et al. [142]	Deep CNN with different dilution rates	Pediatric chest X-rays [2] and COVID	CovXNet, Localization, multiple experiments	Multi-class classification needs improvement
Malygina et al. [80]	DenseNet and CycleGAN	Chest X-ray14 [1]	CycleGAN for data balancing, AUC = 0.9994	No performance metrics were other than AUC, no comparison with related works, nonstandard data division.
Naydenova et al. [96]	Feature detection techniques; SVM, and RF	Clinical data [96]	Clinical data, AUC = 97.8%, sensitivity = 96.6%, specificity = 96.4%	Not applicable to realistic clinical scenarios.
Noor et al. [89]	Wavelet transform and modified principal component (ModPC)	100 samples from the Institute of Respiratory Medicine (IPR), Malaysia,	Misclassification probability is at most 0.15.	Small sample size, wavelet transform cannot preserve the phase information
Oliveria et al. [21]	(Haar) wavelet transform as a feature extractor and KNN as a classifier	PneumoCAD dataset [21]	Proposed PneumoCAD, AUC = 97%, sensitivity = 100	Specificity = 90%, no image processing technique
Omar et al. [119]	Customized CNN	Pediatric chest X-rays [2]	Customized CNN, accuracy = 87.65%	Low accuracy, no additional performance metrics were evaluated
Pavitra et al. [93]	Image preprocessing, Gabor filtering, ANN	116 Chest X-rays	Accuracy = 94%	Lack of details in the paper
Putha et al. [110]	ResNet	CQ dataset [110]	89.0% accuracy, baseline for future research	Accuracy needs improvement, validation needed for detecting other pathologies such as pneumothorax and localization of lung nodules

TABLE 4. *(Continued.)* **Comparison of the pneumonia detection techniques.**

Raheel. [5]	18-layer DCNN	Pediatric chest X-rays [2]	Customized CNN, accuracy = 94.3%, sensitivity = 99.0%	Poor specificity (88%)
Rahimzada et al. [143]	proposed concatenation of pre-trained Xception and ResNet50V2	Cohen et al. [144] and RSNA [128]	proposed concatenation of pre-trained Xception and ResNet50V2, 91% average accuracy	Low precision of COVID-19, poor specificity for normal samples, the sensitivity of 87% for pneumonia
Rahmat et al. [138]	Faster-RCNN	Chest X-ray14 [1]	Employs Faster RCNN for pathology detection in CXR, comparison with a general practitioner, accuracy = 62.0%, sensitivity = 72%, specificity =54.39%	Performance needs improvement, nonstandard data division.
Rajarman et al. [123]	Atlas-based detection algorithm [96] and DCNN	Pediatric chest X-rays [2]	Visualizes the nodules, accuracy = 96.2%	Specificity was only 0.859 when classifying viral vs. bacterial pneumonia
Rajpal et al. [150]	PCA, ResNet50, FFT, DWT, texture feature	Pediatric chest X-rays [2] and cohen et al. [144]	Accuracy of 97%, sensitivity of 0.97.	Classification performance for pneumonia needs improvement
Sara et al. [111]	Inception V3 as FE and SVM KNN and NN as classier	Pediatric chest X-rays [2]	Multiple CNN classifiers	Sensitivity was only 83.5%.
Sarkar et al. [122]	Image enhancement (bilateral filtering, CLAHE algorithm), DL, model visualization via Grad-CAM, and saliency map	Pediatric chest X-rays [2]	Accuracy = 98.82%, AUC = 0.99726, sensitivity = 0.9927, specificity = 0.9555	Nonstandard data division and comparison
Schwyzer et al. [106]	Pre-trained ResNet-34	Reduced dose CT from 100 patients	Standard dose: AUC = 0.923, sensitivity = 82.9%, specificity = 93.8, Reduced dose: AUC = 0.881, sensitivity = 71.0%, specificity = 93.3%	Poor sensitivity
Self et al. [95]	Diagnostic comparison of CT and CXR	4237 CT scans and 3423 CXRs	CXR shows poor sensitivity (43.5%) and poor positive predictive value (26.9%)	Classification not performed with a CAD tool.
Sirazitdinov et al. [134]	Ensemble of mask R-CNN and RatinaNet	RSNA [128]	Precision, recall, and F1 scores = 0.75, 0.79, and 0.77, respectively	Needs improvement to minimize false positive and false negative instances
Sousa et al. [82]	Haar wavelet transform, sequential forward elimination (SFE) algorithm, SVM, KNN, and NB	PneumoCAD dataset [21]	Extensions of pneumoCAD, accuracies of 77%, 70%, and 68%	Accuracy needs improvement; no other performance metrics were evaluated
Sousa et al. [87]	KNN, naive Bayes, MLP, decision tree and SVM combined with PCA, sequential forward elimination, and kernel PCA	PneumoCAD dataset [21]	Extension of pneumoCAD, 96% accuracy with NB	Naïve Bayes assumes that all features are independent.
Stephan et al. [4]	CNN with four convolution layers	Pediatric chest X-rays [2]	Less computationally complex model, accuracy = 93.7%	Nonstandard division of dataset, no performance metrics other than accuracy were evaluated
Thakur et al. [101]	Pretrained VGG16	Pediatric chest X-rays [2]	Accuracy = 90.54%, recall = 98.7%, precision = 87.7%, F1 score = 92.9%	No comparison with related works, accuracy, and precision not promising
Togacar et al. [107]	Deep feature learning model, Alexnet, VGG16, VGG19, Classification via SVM, LDA, KNN, LR, DT	Chest X-ray dataset [2]	A combination of features achieves 99.4% accuracy	Nonstandard data division
Tuncer et al. [154]	MKLBP, INCA, DT, LD, SVM, KNN	Kaggle, Github	Cubic SVM achieved 97.01% classification results.	Classification performance for pneumonia needs improvement
Ureta et al. [159]	(Resnet-18, VGG16 and DenseNet	Pediatric chest X-rays [2] and Chest X-rays14	Model combination	Low accuracy on CXR-14 dataset
Varela-Santos et al. [92]	Histogram equalization, GLCM, ANN	Chest X-ray14 [1]	Accuracy = 92%	Resizing of the image to extract ROI is subjective, limited to binary classification, accuracy needs improvement
Varshni et al. [140]	Xception, VGG16, VGG-19, ResNet-50, DenseNet-121 and DenseNet-169 with SVM, KNN, RF, NB	Chest X-ray14 [1]	Pretrained networks for feature extraction and ML classifiers for classification, AUC = 80.02%	Limited to binary classification, performance needs improvement
Verma et al. [118]	Customized CNN, image processing (data augmentation, annealing) method, and learning rate variation)	Shenzen and Pediatric chest X-rays [2]	Overall accuracy = 99.01 for classifying TB, viral pneumonia, bacterial pneumonia, and normal	Low precision for bacterial pneumonia, lack of experimental details

TABLE 4. *(Continued.)* **Comparison of the pneumonia detection techniques.**

Vijendran et al. [132]	Multilayer online sequential extreme learning machine (OSLEM)	5860 Chest X-rays	Accuracy = 91.7% for multilayered OSELM and 92% for CNN; testing time = 2.434 s for CNN and 1.23 s for multilayered OSELM	Accuracy lower than that of CNN, no other performance metrics were evaluated
Wang [129]	U-Net, Co-CNN (ResNet50 and ResNet101) , detection and localization	Kaggle [128]	Prediction score = 0.2409, sensitivity and specificity = 0.89 and 0.78, respectively.	Results need improvement
Wang et al. [1]	Transfer learning of pre-trained AlexNet, GoogleNet, VGG16, ResNet.	Chest X-rays14 [1]	On the largest dataset, AUC = 63%	Accuracy not promising, the dataset contains only eight diseases; however, it contains 14.
Wu et al. [121]	Image enhancement, ACNN–RF, GridSearchCV-based RF	Pediatric chest X-rays [2]	Accuracy = 97%, precision and recall = 0.90 and 0.95, respectively, runtime = 625.2 s	Precision needs improvement, nonstandard data division.
Yao et al. [94]	Texture features, SVM	CT images	Accuracy = 80%	Accuracy needs improvement, difficult to classify between bacterial pneumonia and H1N1 influenza
Yao et al. [8]	LSTM and CNN	Chest X-rays14 [1]	AUC = 0.71	Accuracy not promising
Yu et al [137]	CGNet	Pediatric chest X-rays [2]	Accuracy of 0.98, sensitivity of 1, and specificity of 0.97.	-
Zech et al. [109]	ResNet-50 and DenseNet-121, detailed experiments	NIH [1], IU, MSH	MSH–NIH model: accuracy, sensitivity, specificity, and AUC = 0.732, 0.950, 0.706, and 0.931 and respectively in the internal test, and 0.238, 0.974, 0.230, and 0.815, respectively in the external test .	The study was limited to frontal radiographs, the performance of the external test was not promising.

VIII. LIMITATIONS OF THE STUDY

There are several future directions to improve the review presented in this paper. Firstly, the review only considers articles that were published in English. The selected articles are only focused on X-rays images, however, CT scans, MRI, and fMRI data can also be considered. Due to the recent outbreak of the COVID-19, many researchers have shifted their attention to developing advanced ML models to assist in pneumonia identification. As a result, many papers were published in the past year. This review covers many of the most influential papers on pneumonia and COVID-19, however; we limited our review to methods developed to extract valuable features from X-rays. Furthermore, due to resemblance to COVID-19, papers related to the string ''pneumonia AND machine learning'' and published between 2010 and 2019 were only 31 on IEEE explorer, compared to 94 in 2020 only. This shows that there is a need for a separate review study to focus only on pneumonia and COVID-19. In addition, this study did not consider performing experimental works to analyze and compare the performance of the existing methods.

IX. CONCLUSION

Chest radiography has played a vital role in the examination and diagnosis of chest diseases. Thus, automatic detection has become a popular topic in computer vision research for medical imaging. Many algorithms using various techniques have been developed, but a literature review summarizing the currently available practices has been lacking. Based on this review, practitioners can select an appropriate method from a real-time perspective, understand the currently available datasets, and appreciate the currently achieved results in this domain.

This study presented an overview of the current literature on pneumonia identification from the chest X-ray data. It summarized the topic and analyzed the usability, goodness factors, and computational complexities of the present algorithms. We observed that multiple datasets are available for a prescribed task.

Further, traditional ML methods can be employed when high computational resources are scarce but cannot be employed for industrial-scale intelligent detection of pneumonia. Therefore, DL can be used.

Most of the CXR datasets [1], [2], [30] are highly imbalanced, and only some researchers have followed the proposed balancing techniques. In most cases, the authors have utilized undersampling and obtained their results with few images. These results are unreliable and cannot be exploited on an industrial scale. We suggest collecting large-scale data that should be balanced and contain approximately one million samples of pneumonia cases. Such datasets can be difficult to create because they rely on humans as the gold standard, are subject to copyright and privacy laws, and may require verification by multiple sources, which is expensive, time-consuming, and subjective. Therefore, based on the literature, we suggest generating high-quality synthesized images using GANs along with data augmentation techniques. Unlike traditional oversampling techniques such as

SMOTE, which generates samples based on the line between instances of the minority class [160], [161], the GANs generative model learns highly informative features from majority classes and uses them to generate images for minority classes [161], [162]. Furthermore, GANs work well with images thus incorporating GANs based synthesized images with the training data would significantly improve the classification performance [67], [163]–[167].

Finally, a significant volume of pneumonia detection works have focused on DL algorithms; few studies are based on traditional ML [21], [82], [87]–[89], [91], [92], [94]–[97], [158]. The current best performance with a low computational requirement was provided by Sousa *et al.* [87]. However, the works of Hashmi *et al.* [136] and Chouhan *et al.* [131] can also be considered remarkable if computational power is not constrained. Therefore, the main challenges to be addressed in the future are listed below:

- Researchers have mainly focused on DL algorithms for pneumonia detection, but preprocessing techniques, such as image enhancement, data augmentation techniques (classic and synthetic data augmentation) need to be considered to improve the classification performance [89], [91], [105], [118]. Pre-processing of CXRs helps in improving the quality of input image by eliminating noise, adjust low or high frequencies, adjusting image contrast, etc. Using data augmentations helps in avoiding overfitting, increase the training set by producing realistic samples as well as helps to improve the model robustness.
- The majority of the existing works have only used frontal radiographs. To improve the detection accuracy, lateral view radiographs should also be considered [26], [168]. It is shown that in frontal views, up to 15% of the lung can be hidden by cardiovascular structures and the diaphragm thus lateral views may provide information on 15% of the lung that is hidden from frontal views [168].
- In the current literature, the majority of work has been done on normal human lung anatomy (without any structural deformities) however there is a need for an intelligent pneumonia identification system that can consider pathological deformities such as deformed lungs due to accidents, disease, or post-surgical alterations, e.g., pneumonectomy or lobectomy [169] into consideration. Moreover, the algorithms developed for the segmentation of lung regions are based on adult chest x-rays datasets [1] and may not perform well on pediatric CXRs [2] due to the differences in lung appearance and sizes between adults and pediatrics [169], [170].
- Few authors have incorporated nonimage features, such as patient age, sex information, to the CXR data. Including a patient's records is expected to improve the classification performance and reduce the false positives rate. For example, pneumonia would be an appropriate term for a patient of pulmonary abnormality with a history of fever and cough rather than less specific terms such as infiltration or consolidation [26], [171].
- The majority of the ML-based pneumonia identification systems lack transparency on how the data split was performed. Most of the datasets are divided into training and testing sets randomly. The studies must include descriptions of how the training and testing split was being selected. Furthermore, distinct datasets should be used for training and testing to evaluate the diagnostic accuracy and generalizability of the proposed algorithms [109].
- Several models for pneumonia detection can perform up to 99% accuracy, however, often the sensitivity or specificity of the models are not promising. Thus, to evaluate the model performance all matrices such as sensitivity, specificity, AUC, and F1 measure should be taken into consideration. Moreover, the models were able to perform well for binary classes, but not for multi-class classification (viral/bacterial pneumonia/normal/COVID-19) [142]. Therefore, researchers should focus on building new methods for intelligent disease detection from CXRs using lightweight algorithms that can efficiently identify pneumonia and other chest-related diseases across different datasets.

ACKNOWLEDGMENT

This work was supported by the College of Information Technology and the office of the Associate Provost for Research, United Arab Emirates University (UAEU).

REFERENCES

[1] 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. 3462–3471.

[2] D. S. Kermany, G. Michael, C. Wenjia, C. V. Carolina, L. Huiying, L. B. Sally, and M. E. A. Alex, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[3] *WHO | World Health Organization*. Accessed: Mar. 31, 2020. [Online]. Available: https://www.who.int/

[4] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *J. Healthcare Eng.*, vol. 2019, pp. 1–7, Mar. 2019.

[5] R. Siddiqi, "Automated pneumonia diagnosis using a customized sequential convolutional neural network," in *Proc. 3rd Int. Conf. Deep Learn. Technol. (ICDLT)*, 2019, pp. 64–70.

[6] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 167–179, Jan. 2019.

[7] A. Oates, K. Halliday, A. C. Offiah, C. Landes, N. Stoodley, A. Jeanes, K. Johnson, S. Chapman, S. M. Stivaros, J. Fairhurst, A. Watt, M. Paddock, K. Giles, K. McHugh, and O. J. Arthurs, "Shortage of paediatric radiologists acting as an expert witness: Position statement from the british society of paediatric radiology (BSPR) national working group on imaging in suspected physical abuse (SPA)," *Clin. Radiol.*, vol. 74, no. 7, pp. 496–502, Jul. 2019.

[8] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," 2017, *arXiv:1710.10501*. [Online]. Available: http://arxiv.org/abs/1710.10501

[9] G. Deng and L. W. Cahill, "An adaptive Gaussian filter for noise reduction and edge detection," in *Proc. IEEE Conf. Rec. Nucl. Sci. Symp. Med. Imag. Conf.*, Oct. 1993, pp. 1615–1619.

[10] X. Zhuang and R. M. Haralick, "Morphological structuring element decomposition," *Comput. Vis., Graph., Image Process.*, vol. 35, no. 3, pp. 370–382, Sep. 1986.
[11] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
[12] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.
[13] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[14] B. Yegnanarayana, *Artificial Neural Networks*. New Delhi, Delhi: PHI, 2009.
[15] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets: A review," *GESTS Int. Trans. Comput. Sci. Eng.*, vol. 30, no. 1, pp. 25–36, 2006.
[16] A. Krizhevsky, S. Ilya, and E. H. Geoffrey, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[17] I. Goodfellow, P.-A. Jean, M. Mehdi, X. Bing, W.-F. David, O. Sherjil, C. Aaron, and B. Yoshua, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014.
[18] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434
[19] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.
[20] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
[21] L. L. G. Oliveira, S. A. E. Silva, L. H. V. Ribeiro, R. M. de Oliveira, C. J. Coelho, and A. L. S. S. Andrade, "Computer-aided diagnosis in chest radiography for detection of childhood pneumonia," *Int. J. Med. Informat.*, vol. 77, no. 8, pp. 555–564, Aug. 2008.
[22] J. Kim, O. Sangjun, Y. Kim, and M. Lee, "Convolutional neural network with biologically inspired retinal structure," *Procedia Comput. Sci.*, vol. 88, pp. 145–154, Jan. 2016.
[23] R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?" *ICT Exp.*, vol. 4, no. 4, pp. 247–254, Dec. 2018.
[24] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[25] O. Ronneberger, F. Philipp, and B. Thomas, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
[26] 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
[27] A. Mortazi, K. Rashed, R. Kawal, B. Jeremy, and B. Ulas, "CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2017, pp. 377–385.
[28] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, "Comparison of deep learning approaches for multi-label chest X-ray classification," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, Dec. 2019.
[29] D. Moher, P.-P. Group, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. Stewart, "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement," *Systematic Rev.*, vol. 4, no. 1, pp. 1–9, Dec. 2015.
[30] A. E. W. Johnson, T. J. Pollard, N. R. Greenbaum, M. P. Lungren, C.-Y. Deng, Y. Peng, Z. Lu, R. G. Mark, S. J. Berkowitz, and S. Horng, "MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs," 2019, *arXiv:1901.07042*. [Online]. Available: http://arxiv.org/abs/1901.07042
[31] D. Demner-Fushman, M. D. Kohli, M. B. Rosenman, S. E. Shooshan, L. Rodriguez, S. Antani, G. R. Thoma, and C. J. McDonald, "Preparing a collection of radiology examinations for distribution and retrieval," *J. Amer. Med. Inform. Assoc.*, vol. 23, no. 2, pp. 304–310, Mar. 2016.
[32] S. Jaeger, S. Candemir, S. Antani, Y.-X. Wáng, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imag. Med. Surg.*, vol. 4, no. 6, p. 475, 2014.
[33] S. Ryoo and H. J. Kim, "Activities of the Korean institute of tuberculosis," *Osong Public Health Res. Perspect.*, vol. 5, pp. S43–S49, Dec. 2014.
[34] 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.
[35] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, "Training deep neural networks on imbalanced data sets," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 4368–4374.
[36] S. Bermejo and J. Cabestany, "Oriented principal component analysis for large margin classifiers," *Neural Netw.*, vol. 14, no. 10, pp. 1447–1461, Dec. 2001.
[37] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, Dec. 2019.
[38] S. Mousavi, A. Fotoohinasab, and F. Afghah, "Single-modal and multimodal false arrhythmia alarm reduction using attention-based convolutional and recurrent neural networks," *PLoS ONE*, vol. 15, no. 1, Jan. 2020, Art. no. e0226990.
[39] S. Mousavi, F. Afghah, and U. R. Acharya, "SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0216456.
[40] C. Vo and P. N. Hua, "An enhanced CNN model on temporal educational data for program-level student classification," in *Proc. Asian Conf. Intell. Inf. Database Syst.*, 2020, pp. 442–454.
[41] P. H. G. Nguyen and T. N. V. Chau, "A CNN model with data imbalance handling for course-level student prediction based on forum texts," in *Proc. Int. Conf. Comput. Collective Intell.*, 2018, pp. 479–490.
[42] P. Hart, "The condensed nearest neighbor rule (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 3, pp. 515–516, May 1968.
[43] M. A. Maloof, "Learning when data sets are imbalanced and when costs are unequal and unknown," in *Proc. Workshop Learn. Imbalanced Data Sets (ICML)*, 2003, pp. 515–516.
[44] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newslett.*, vol. 6, no. 1, pp. 20–29, Jun. 2004.
[45] M. Kubat and M. Stan, "Addressing the curse of imbalanced training sets: One-sided selection," in *Proc. ICML*, 1997, pp. 179–186.
[46] J. Laurikkala, "Improving identification of difficult small classes by balancing class distribution," in *Proc. Conf. Artif. Intell. Med. Eur.*, Berlin, Germany, 2001, pp. 63–66.
[47] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
[48] I. Tomek, "Two modifications of CNN," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-6, no. 11, pp. 769–772, 1976.
[49] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates," *Inf. Sci.*, vol. 425, pp. 76–91, Jan. 2018.
[50] G. Douzas, F. Bacao, and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE," *Inf. Sci.*, vol. 465, pp. 1–20, Oct. 2018.
[51] E. S. M. Abd and A. Ajith, "A review of class imbalance problem," *J. Netw. Innov. Comput.*, vol. 1, no. 2013, pp. 332–340, 2013.
[52] C. Bunkhumpornpat, S. Krung, and L. Chidchanok, "Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*, 2009, pp. 475–482.
[53] J. Wang, M. Xu, H. Wang, and J. Zhang, "Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding," in *Proc. 8th Int. Conf. Signal Process.*, 2006.
[54] V. García, J. S. Sánchez, R. Martín-Félez, and R. A. Mollineda, "Surrounding neighborhood-based SMOTE for learning from imbalanced data sets," *Prog. Artif. Intell.*, vol. 1, no. 4, pp. 347–362, Dec. 2012.
[55] Y. Xie, D. Jiang, X. Wang, and R. Xu, "Robust transfer integrated locally kernel embedding for click-through rate prediction," *Inf. Sci.*, vol. 491, pp. 190–203, Jul. 2019.

[56] B. Krawczyk, Ł. Jeleń, A. Krzyżak, and F. Thomas, "Oversampling methods for classification of imbalanced breast cancer malignancy data," in *Proc. Int. Conf. Comput. Vis. Graph.*, 2012, pp. 483–490.

[57] N. V. Chawla, L. Aleksandar, O. H. Lawrence, and W. B. Kevin, "SMOTEBoost: Improving prediction of the minority class in boosting," in *Proc. Eur. Conf. Princ. Data Mining Knowledge Discovery*, 2003, pp. 107–119.

[58] S. Wang and X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models," in *Proc. IEEE Symp. Comput. Intell. Data Mining*, Mar. 2009, pp. 324–331.

[59] S. Chen, H. He, and E. A. Garcia, "RAMOBoost: Ranked minority oversampling in boosting," *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1624–1642, Oct. 2010.

[60] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw., IEEE World Congr. Comput. Intell.*, Jun. 2008, pp. 1322–1328.

[61] B. Tan, J. Yang, Y. Tang, S. Jiang, P. Xie, and W. Yuan, "A deep imbalanced learning framework for transient stability assessment of power system," *IEEE Access*, vol. 7, pp. 81759–81769, 2019.

[62] S. Barua, M. M. Islam, X. Yao, and K. Murase, "MWMOTE-majority weighted minority oversampling technique for imbalanced data set learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 2, pp. 405–425, Feb. 2014.

[63] S. S. Mullick, S. Datta, and S. Das, "Generative adversarial minority oversampling," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1695–1704.

[64] S. Ando and Y. H. Chun, "Deep over-sampling framework for classifying imbalanced data," in *Proc. Joint Eur. Conf. Mach. Learning Knowledge Discovery Databases*, 2017, pp. 770–785.

[65] V. Sampath, I. Maurtua, J. J. A. Martín, and A. Gutierrez, "A survey on generative adversarial networks for imbalance problems in computer vision tasks," *J. Big Data*, vol. 8, no. 1, pp. 1–59, Dec. 2021.

[66] J. Engelmann and S. Lessmann, "Conditional wasserstein GAN-based oversampling of tabular data for imbalanced learning," 2020, *arXiv:2008.09202*. [Online]. Available: http://arxiv.org/abs/2008.09202

[67] Q. Wu, Y. Chen, and J. Meng, "DCGAN-based data augmentation for tomato leaf disease identification," *IEEE Access*, vol. 8, pp. 98716–98728, 2020.

[68] G. Yildirim, N. Jetchev, R. Vollgraf, and U. Bergmann, "Generating high-resolution fashion model images wearing custom outfits," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3161–3164.

[69] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8110–8119.

[70] B. Carl and W. Jevin. *Which Face Is Real?* Accessed: 2021. [Online]. Available: https://www.whichfaceisreal.com/learn.html

[71] Y. Lei, J. Harms, T. Wang, Y. Liu, H. Shu, A. B. Jani, W. J. Curran, H. Mao, T. Liu, and X. Yang, "MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks," *Med. Phys.*, vol. 46, no. 8, pp. 3565–3581, Aug. 2019.

[72] T. Kaneko, H. Kameoka, K. Tanaka, and N. Hojo, "CycleGAN-VC2: Improved cyclegan-based non-parallel voice conversion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 6820–6824.

[73] W. Dai, D. Nanqing, W. Zeya, L. Xiaodan, Z. Hao, and P. X. Eric, "Scan: Structure correcting adversarial network for organ segmentation in chest X-rays," in *Proc. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support*, 2018, pp. 263–273.

[74] H. Salehinejad, E. Colak, T. Dowdell, J. Barfett, and S. Valaee, "Synthesizing chest X-ray pathology for training deep convolutional neural networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 5, pp. 1197–1206, May 2019.

[75] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[76] A. Gupta, S. Venkatesh, S. Chopra, and C. Ledig, "Generative image translation for data augmentation of bone lesion pathology," 2019, *arXiv:1902.02248*. [Online]. Available: http://arxiv.org/abs/1902.02248

[77] M. J. M. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, "How to fool radiologists with generative adversarial networks? A visual turing test for lung cancer diagnosis," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 240–244.

[78] C. Baur, S. Albarqouni, and N. Navab, "MelanoGANs: High resolution skin lesion synthesis with GANs," 2018, *arXiv:1804.04338*. [Online]. Available: http://arxiv.org/abs/1804.04338

[79] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.

[80] T. Malygina, E. Elena, and D. Ivan, "GANs' N Lungs: Improving pneumonia prediction," 2019, *arXiv:1908.00433*. [Online]. Available: https://arxiv.org/abs/1908.00433

[81] D. Mahapatra, B. Antony, S. Sedai, and R. Garnavi, "Deformable medical image registration using generative adversarial networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1449–1453.

[82] R. T. Sousa, O. Marques, F. A. A. M. N. Soares, I. I. G. Sene, L. L. G. de Oliveira, and E. S. Spoto, "Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs," *Procedia Comput. Sci.*, vol. 18, pp. 2579–2582, Jan. 2013.

[83] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," 2015, *arXiv:1510.00149*. [Online]. Available: http://arxiv.org/abs/1510.00149

[84] P. W. Huang and S. K. Dai, "Design of a two-stage content-based image retrieval system using texture similarity," *Inf. Process. Manage.*, vol. 40, no. 1, pp. 81–96, Jan. 2004.

[85] M. Kokare, P. K. Biswas, and B. N. Chatterji, "Texture image retrieval using new rotated complex wavelet filters," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 35, no. 6, pp. 1168–1178, Dec. 2005.

[86] M. K. Bashar, T. Matsumoto, and N. Ohnishi, "Wavelet transform-based locally orderless images for texture segmentation," *Pattern Recognit. Lett.*, vol. 24, no. 15, pp. 2633–2650, Nov. 2003.

[87] R. T. Sousa, O. Marques, G. T. F. Curado, R. M. da Costa, A. S. Soares, F. A. A. M. N. Soares, and L. L. G. de Oliveira, "Evaluation of classifiers to a childhood pneumonia computer-aided diagnosis system," in *Proc. IEEE 27th Int. Symp. Computer-Based Med. Syst.*, May 2014, pp. 477–478.

[88] A. Depeursinge, J. Iavindrasana, A. Hidki, G. Cohen, A. Geissbuhler, A. Platon, P.-A. Poletti, and H. Müller, "Comparative performance analysis of state-of-the-art classification algorithms applied to lung tissue categorization," *J. Digit. Imag.*, vol. 23, no. 1, pp. 18–30, Feb. 2010.

[89] N. M. Noor, O. M. Rijal, A. Yunus, and S. A. R. Abu-Bakar, "A discrimination method for the detection of pneumonia using chest radiograph," *Comput. Med. Imag. Graph.*, vol. 34, no. 2, pp. 160–166, Mar. 2010.

[90] A. A. E. Ambita, N. V. B. Eujene, and C. N. Prospero, "Locally adaptive regression kernels and support vector machines for the detection of pneumonia in chest X-ray images," in *Proc. Asian Conf. Intell. Inf. Database Syst.*, 2020, pp. 129–140.

[91] A. Khatri, J. Rishabh, V. Hariom, M. Nityam, R. Priya, and J. Rajiv, "Pneumonia identification in chest X-ray images using EMD," in *Trends in Communication, Cloud, and Big Data*. Singapore: Springer, 2020.

[92] S. Varela-Santos and M. Patricia, "Classification of X-ray images for pneumonia detection using texture features and neural networks," in *Intuitionistic and Type-2 Fuzzy Logic Enhancements in Neural and Optimization Algorithms: Theory and Applications*. Cham, Switzerland: Springer, 2020.

[93] R. Pavithra and S. Y. Pattar, "Detection and classification of lung disease-pneumonia and lung cancer in chest radiology using artificial neural network," *Int. J. Sci. Res. Publications*, vol. 5, no. 10, 2015.

[94] J. Yao, A. Dwyer, R. M. Summers, and D. J. Mollura, "Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification," *Acad. Radiol.*, vol. 18, no. 3, pp. 306–314, Mar. 2011.

[95] W. H. Self, D. M. Courtney, C. D. McNaughton, R. G. Wunderink, and J. A. Kline, "High discordance of chest X-ray and computed tomography for detection of pulmonary opacities in ED patients: Implications for diagnosing pneumonia," *Amer. J. Emergency Med.*, vol. 31, no. 2, pp. 401–405, Feb. 2013.

[96] E. Naydenova, A. Tsanas, C. Casals-Pascual, and M. De Vos, "Smart diagnostic algorithms for automated detection of childhood pneumonia in resource-constrained settings," in *Proc. IEEE Global Humanitarian Technol. Conf. (GHTC)*, Oct. 2015, pp. 377–384.

[97] B. Antin, K. Joshua, and M. Emil, "Detecting pneumonia in chest X-rays with supervised learning," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2017.

[98] L. V. D. Maaten and H. Geoffrey, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
[99] F. Q. Lauzon, "An introduction to deep learning," in *Proc. 11th Int. Conf. Inf. Sci., Signal Process. Appl. (ISSPA)*, 2012, pp. 1438–1439.
[100] E. Ayan and H. M. Unver, "Diagnosis of pneumonia from chest X-ray images using deep learning," in *Proc. Sci. Meeting Elect.-Electron. Biomed. Eng. Comput. Sci. (EBBT)*, Apr. 2019, pp. 1–5.
[101] S. Thakur, G. Yajash, A. Sahil, U. Rohit, and S. Geetanjali, "Chest X-ray images based automated detection of pneumonia using transfer learning and CNN," in *Proc. Int. Conf. Artif. Intell. Appl.*, 2020, pp. 329–335.
[102] B. Li, G. Kang, K. Cheng, and N. Zhang, "Attention-guided convolutional neural network for detecting pneumonia on chest X-rays," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 4851–4854.
[103] 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.
[104] R. Jain, P. Nagrath, G. Kataria, V. S. Kaushik, and D. J. Hemanth, "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning," *Measurement*, vol. 165, Dec. 2020, Art. no. 108046.
[105] P. Chhikara, S. Prabhjot, G. Prakhar, and B. Tarunpreet, "Deep convolutional neural network with transfer learning for detecting pneumonia on chest X-rays," in *Proc. Adv. Bioinf., Multimedia, Electron. Circuits Signals*, 2020, pp. 155–168.
[106] M. Schwyzer, K. Martini, S. Skawran, M. Messerli, and T. Frauenfelder, "Pneumonia detection in chest X-ray dose-equivalent CT: Impact of dose reduction on detectability by artificial intelligence," *Academic Radiol.*, Jul. 2020.
[107] M. Toğaçar, B. Ergen, Z. Cömert, and F. Özyurt, "A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models," *IRBM*, vol. 41, no. 4, pp. 212–222, 2019.
[108] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 104964.
[109] J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study," *PLOS Med.*, vol. 15, no. 11, Nov. 2018, Art. no. e1002683.
[110] P. Putha, T. Manoj, R. B. R. Tarun, A. C. Justy, G. Shalini, S. Namita, R. M. K. Sundeep, R. Pooja, and W. Prashant, "Can artificial intelligence reliably report chest X-rays?: Radiologist validation of an algorithm trained on 2.3 million X-rays," 2018, *arXiv:1807.07455*. [Online]. Available: https://arxiv.org/abs/1807.07455
[111] S. L. K. Yee and W. J. K. Raymond, "Pneumonia diagnosis using chest X-ray images and machine learning," in *Proc. 10th Int. Conf. Biomed. Eng. Technol.*, Sep. 2020, pp. 101–105.
[112] K. El Asnaoui, Y. Chawki, and A. Idri, "Automated methods for detection and classification pneumonia based on X-ray images using deep learning," 2020, *arXiv:2003.14363*. [Online]. Available: http://arxiv.org/abs/2003.14363
[113] Y. Bar, D. Idit, W. Lior, L. Sivan, K. Eli, and G. Hayit, "Chest pathology identification using deep feature selection with non-medical training," *Comput. Methods Biomech. Biomed. Eng., Imag. Visualizat.*, vol. 6, no. 3, p. 2018, 259-263.
[114] J. Donahue, Y. J. V. Oriol, H. Judy, Z. Ning, T. Eric, and D. Trevor, "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.
[115] O. Er, N. Yumusak, and F. Temurtas, "Chest diseases diagnosis using artificial neural networks," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7648–7655, Dec. 2010.
[116] D. F. Specht, "Probabilistic neural networks," *Neural Netw.*, vol. 3, no. 1, pp. 109–118, 1990.
[117] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, Nov. 1994.
[118] D. Verma, C. Bose, N. Tufchi, K. Pant, V. Tripathi, and A. Thapliyal, "An efficient framework for identification of tuberculosis and pneumonia in chest X-ray images using neural network," *Procedia Comput. Sci.*, vol. 171, pp. 217–224, Jan. 2020.
[119] H. S. Omar and A. Babalik, "Detection of pneumonia from X-ray images using convolutional neural network," in *Proc. Book*, 2019, p. 183.
[120] R. Bhatt, Y. Sudhanshusinh, and N. S. Jignesh, "Convolutional neural network based chest X-ray image classification for pneumonia diagnosis," in *Proc. Int. Conf. Emerg. Technol. Trends Electron. Commun. Netw.*, 2020, pp. 254–266.
[121] H. Wu, X. Pengjie, Z. Huiyi, L. Daiyi, and C. Ming, "Predict pneumonia with chest X-ray images based on convolutional deep neural learning networks," *J. Intell. Fuzzy Syst. Preprint*, vol. 39, no. 3, pp. 2893–2907, 2020.
[122] R. Sarkar, H. Animesh, S. Koulick, and G. Preetam, "A novel method for pneumonia diagnosis from chest X-ray images using deep residual learning with separable convolutional networks," in *Computer Vision and Machine Intelligence in Medical Image Analysis*. Singapore: Springer, 2020.
[123] S. Rajaraman, C. Sema, T. George, and A. Sameer, "Visualizing and explaining deep learning predictions for pneumonia detection in pediatric chest radiographs," *Med. Imag., Comput.-Aided Diagnosis*, vol. 10950, Mar. 2019, Art. no. 109500S.
[124] R. H. Abiyev and M. K. S. Ma'aitah, "Deep convolutional neural networks for chest diseases detection," *J. Healthcare Eng.*, vol. 2018, pp. 1–11, Aug. 2018.
[125] X. Gu, L. Pan, H. Liang, and R. Yang, "Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography," in *Proc. 3rd Int. Conf. Multimedia Image Process. (ICMIP)*, 2018, pp. 88–93.
[126] K. He, G. Georgia, D. Piotr, and G. Ross, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.
[127] A. K. Jaiswal, T. Prayag, K. Sachin, G. Deepak, K. Ashish, and J. R. Joel, "Identifying pneumonia in chest X-rays: A deep learning approach," *Measurement*, vol. 145, pp. 511–518, Oct. 2019.
[128] *RSNA Pneumonia Detection Challenge*. Radiological Society of North America. Accessed: 2018. [Online]. Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge
[129] K. Wang, Z. Xiaohong, H. Sheng, and C. Feiyu, "Automatic detection of pneumonia in chest X-ray images using cooperative convolutional neural networks," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, 2019, pp. 328–340.
[130] Z. Li, W. Chong, H. Mei, X. Yuan, W. Wei, L. Li-Jia, and F.-F. Li, "Thoracic disease identification and localization with limited supervision," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8290–8299.
[131] V. Chouhan, S. K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damaševičius, and V. H. C. de Albuquerque, "A novel transfer learning based approach for pneumonia detection in chest X-ray images," *Appl. Sci.*, vol. 10, no. 2, p. 559, Jan. 2020.
[132] S. Vijendran and R. Dubey, "Deep online sequential extreme learning machines and its application in pneumonia detection," in *Proc. 8th Int. Conf. Ind. Technol. Manage. (ICITM)*, Mar. 2019, pp. 311–316.
[133] M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, "Abnormality detection and localization in chest X-rays using deep convolutional neural networks," 2017, *arXiv:1705.09850*. [Online]. Available: http://arxiv.org/abs/1705.09850
[134] I. Sirazitdinov, M. Kholiavchenko, T. Mustafaev, Y. Yixuan, R. Kuleev, and B. Ibragimov, "Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database," *Comput. Electr. Eng.*, vol. 78, pp. 388–399, Sep. 2019.
[135] H. Ko, H. Ha, H. Cho, K. Seo, and J. Lee, "Pneumonia detection with weighted voting ensemble of CNN models," in *Proc. 2nd Int. Conf. Artif. Intell. Big Data (ICAIBD)*, May 2019, pp. 306–310.
[136] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, "Efficient pneumonia detection in chest Xray images using deep transfer learning," *Diagnostics*, vol. 10, no. 6, p. 417, Jun. 2020.
[137] X. Yu, S.-H. Wang, and Y.-D. Zhang, "CGNet: A graph-knowledge embedded convolutional neural network for detection of pneumonia," *Inf. Process. Manage.*, vol. 58, no. 1, Jan. 2021, Art. no. 102411.
[138] A. Ismail, R. Taufik, and A. Sharifah, "Chest X-ray image classification using faster R-CNN," *Malaysian J. Comput.*, vol. 4, no. 1, pp. 225–236, 2019.
[139] T. B. Chandra and V. Kesari, "Pneumonia detection on chest X-ray using machine learning paradigm," in *Proc. 3rd Int. Conf. Comput. Vis. Image Process.*, 2020, pp. 21–33.
[140] D. Varshni, K. Thakral, L. Agarwal, R. Nijhawan, and A. Mittal, "Pneumonia detection using CNN based feature extraction," in *Proc. IEEE Int. Conf. Electr., Comput. Commun. Technol. (ICECCT)*, Feb. 2019, pp. 1–7.
[141] N. Zaki, H. Alashwal, and S. Ibrahim, "Association of hypertension, diabetes, stroke, cancer, kidney disease, and high-cholesterol with COVID-19 disease severity and fatality: A systematic review," *Diabetes Metabolic Syndrome, Clin. Res. Rev.*, vol. 14, no. 5, pp. 1133–1142, Sep. 2020.
[142] T. Mahmud, M. A. Rahman, and S. A. Fattah, "CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization," *Comput. Biol. Med.*, vol. 122, Jul. 2020, Art. no. 103869.

[143] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of xception and ResNet50V2," *Informat. Med. Unlocked*, vol. 19, Jan. 2020, Art. no. 100360.
[144] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "COVID-19 image data collection: Prospective predictions are the future," 2020, *arXiv:2006.11988*. [Online]. Available: http://arxiv.org/abs/2006.11988
[145] W. Jin, S. Dong, C. Dong, and X. Ye, "Hybrid ensemble model for differential diagnosis between COVID-19 and common viral pneumonia by chest X-ray radiograph," *Comput. Biol. Med.*, vol. 131, Apr. 2021, Art. no. 104252.
[146] S. Kilicarslan, K. Adem, and M. Celik, "Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network," *Med. Hypotheses*, vol. 137, Apr. 2020, Art. no. 109577.
[147] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," in *Proc. Eur. Conf. Mach. Learn.*, 1994, pp. 171–182.
[148] N. Dey, Y.-D. Zhang, V. Rajinikanth, R. Pugalenthi, and N. S. M. Raja, "Customized VGG19 architecture for pneumonia detection in chest X-rays," *Pattern Recognit. Lett.*, vol. 143, pp. 67–74, Mar. 2021.
[149] A. K. Das, K. Sidra, K. Chiranjeev, and S. Ditipriya, "TLCoV—An automated Covid-19 screening model using transfer learning from chest X-ray images," *Chaos, Solitons Fractals*, vol. 144, Mar. 2021, Art. no. 110713.
[150] S. Rajpal, N. Lakhyani, A. K. Singh, R. Kohli, and N. Kumar, "Using handpicked features in conjunction with ResNet-50 for improved detection of COVID-19 from chest X-ray images," *Chaos, Solitons Fractals*, vol. 145, Apr. 2021, Art. no. 110749.
[151] T. Rahman and A. A. K. M. Chowdhury. *COVID-19 Radiography Database—Kaggle*. Accessed: 2020. [Online]. Available: https://www.kaggle.com/tawsifurrahman/
[152] A. G. Chung. *GitHub—Agchung/Figure1-COVID-Chestxray-Dataset*. Accessed: 2020. [Online]. Available: https://github.com/agchung/Figure1-COVID-chestxray-dataset
[153] A. G. Chung. *GitHub—Agchung/Actualmed-COVID-Chestxray-Dataset*. Accessed: 2020. [Online]. Available: https://github.com/agchung/Actualmed-COVID-chestxray-dataset
[154] T. Tuncer, F. Ozyurt, S. Dogan, and A. Subasi, "A novel covid-19 and pneumonia classification method based on F-transform," *Chemometric Intell. Lab. Syst.*, vol. 210, Mar. 2021, Art. no. 104256.
[155] I. Perfilieva, N. Vilém, and D. Antonín, "Fuzzy transform in the analysis of data," *Int. J. Approx. Reasoning*, vol. 48, no. 1, pp. 36–46, 2008.
[156] V. Danilov, A. Karpovsky, A. Kirpich, D. Litmanovich, D. Nefaridze, O. Talalov, S. Semyonov, A. Proutski, V. Koniukhovskii, V. Shvartc, and Y. Gankin, "COVID-19/pneumonia classication based on guided attention," Res. Square, Tech. Rep., 2021.
[157] K. Li, W. Ziyan, P. Kuan-Chuan, E. Jan, and F. Yun, "Guided attention inference network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 12, pp. 2996–3010, Dec. 2019.
[158] S. Macedo and D. Oliveira, "Desenvolvimento de um sistema de auxílio ao diagnóstico de pneumonia na infância utilizando visão computacional," Instituto De Informatica, Universidade Federal De Goias, Goiânia, Brazil, Tech. Rep., 2012.
[159] J. Ureta, O. Aran, and J. P. Rivera, "Detecting pneumonia in chest radiographs using convolutional neural networks," in *Proc. 12th Int. Conf. Mach. Vis. (ICMV)*, Jan. 2020, Art. no. 114331.
[160] J.-H. Oh, J. Y. Hong, and J.-G. Baek, "Oversampling method using outlier detectable generative adversarial network," *Expert Syst. Appl.*, vol. 133, pp. 1–8, Nov. 2019.
[161] Z. Zhang, J. Han, K. Qian, C. Janott, Y. Guo, and B. Schuller, "Snore-GANs: Improving automatic snore sound classification with synthesized data," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 300–310, Jan. 2020.
[162] G. Douzas and F. Bacao, "Effective data generation for imbalanced learning using conditional generative adversarial networks," *Expert Syst. Appl.*, vol. 91, pp. 464–471, Jan. 2018.
[163] S. Guan, "Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks," *J. Med. Imag.*, vol. 6, no. 3, p. 1, Mar. 2019.
[164] E. Wu, W. Kevin, C. David, and L. William, "Conditional infilling GANs for data augmentation in mammogram classification," in *Image Analysis for Moving Organ, Breast, and Thoracic Images*. Cham, Switzerland: Springer, 2018.
[165] Y. Dan, Z. Yong, L. Xiang, L. Shaobo, H. Ming, and H. Jianjun, "Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials," *NPJ Comput. Mater.*, vol. 6, no. 1, pp. 1–7, 2020.
[166] T. Koga, N. Nonaka, J. Sakuma, and J. Seita, "General-to-detailed GAN for infrequent class medical images," 2018, *arXiv:1812.01690*. [Online]. Available: http://arxiv.org/abs/1812.01690
[167] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1316–1324.
[168] S. Raoof, D. Feigin, A. Sung, S. Raoof, L. Irugulpati, and E. C. Rosenow, "Interpretation of plain chest roentgenogram," *Chest*, vol. 141, no. 2, pp. 545–558, Feb. 2012.
[169] S. Candemir and S. Antani, "A review on lung boundary detection in chest X-rays," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 14, no. 4, pp. 563–576, Apr. 2019.
[170] R. Arthur, "Interpretation of the paediatric chest X-ray," *Paediatric Respiratory Rev.*, vol. 1, no. 1, pp. 41–50, Mar. 2000.
[171] N. Zaki and E. A. Mohamed, "The estimations of the COVID-19 incubation period: A scoping reviews of the literature," *J. Infection Public Health*, Feb. 2021.

WASIF KHAN received the M.Sc. degree from Kasetsart University, Bangkok, Thailand, under the TAIST-Tokyo Tech. program in collaboration with the Tokyo Institute of Technology, Japan, in 2017. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University (UAEU). He mainly focuses on developing intelligent machine learning algorithms for various medical and civil applications. His research interests include computer vision and deep learning.

NAZAR ZAKI is currently a Professor of Computer Science and the Director of the Big Data Analytics Center, United Arab Emirates University (UAEU). He mainly focuses on developing intelligent data mining algorithms to solve problems in domains, such as biology, health, education, business, and social networks. He published more than 120 scientific results in reputable journals and conferences. His research interests include artificial intelligence and machine learning. He received several scholarship awards, such as the College Recognition Award for Excellence in Scholarship, in 2007, 2012, and 2016, and the Best Paper Award in leading conferences, such as the ACM Genetic and Evolutionary Computation Conference and the Chancellor's Annotation Award in technology. He is also a frequent recipient of certificates of achievement for publishing in top journals.

LUQMAN ALI received the B.Sc. degree in electrical communication engineering from the University of Engineering and Technology at Peshawar, Pakistan, in 2014, and the M.Sc. degree from Kasetsart University, Thailand, under the TAIST-Tokyo Tech. program in collaboration with the Tokyo Institute of Technology, in 2017. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University. He mainly focuses on developing intelligent machine learning algorithms for various medical and civil applications. His research interests include computer vision and deep learning. His awards and honors include the TAIST Tokyo Scholarship, Kasetsart University, the Soyama Scholarship, Tokyo Institute of Technology, Sakura Exchange, Tokyo Institute of Technology, and a Full Fellowship, United Arab Emirates University.