Automatic Diagnosis of Pneumothorax From Chest Radiographs: A Systematic Literature Review

TAHIRA IQBAL\textsuperscript{1}, ARSLAN SHAUKAT\textsuperscript{2,1}, MUHAMMAD USMAN AKRAM\textsuperscript{1}, ZARTASHA MUSTANSAR\textsuperscript{2}, AND AIMAL KHAN\textsuperscript{1}

\textsuperscript{1}Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan
\textsuperscript{2}Research Center for Modelling and Simulation (RCMS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

Corresponding author: Arslan Shaukat (arslanshaukat@ceme.nust.edu.pk)

ABSTRACT Among various medical imaging tools, chest radiographs are the most important and widely used diagnostic tool for the detection of thoracic pathologies. Remarkable research is being carried out to propose robust automatic diagnostic tools for the detection of pathologies from the chest radiographs. Artificial Intelligence techniques have been found to give promising results in automating the field of medicine. Lot of research has been done for automatic detection of pneumothorax from chest radiographs while proposing several frameworks based on artificial intelligence techniques. Undoubtedly, several models are available for automatic diagnosis of pneumothorax, however a summarized review of the existing literature is still missing. This study summarizes the existing literature for pneumothorax detection from chest x-rays along with describing the available chest radiographs datasets. It will help the researchers to select the optimal and most effective model with respect to the real-time scenarios. The comparative analysis of the literature is provided in terms of goodness, usability, and quality along with highlighting the research gaps for further investigation. From the literature, it is evident that pneumothorax is more common in men as compared to women. Additionally, the proposed models have achieved incredible results for pneumothorax detection on selected datasets, however, the effectiveness of proposed models in real-time cases cannot be claimed, as none of the models have been implemented clinically yet. This issue can be solved by external validation of the models. Furthermore, the class-imbalance problem in most of the medical image dataset has been solved by algorithm-level-techniques. Moreover, there is need to put more effort in combined detection and localization of pneumothorax as mostly research is limited to either classification or localization of pathology. So far, best results have been achieved by deep-learning based models with Area-under-receiver-operating-characteristic-curve (AUC) of 88.87\% for classification, and Dice-similarity-coefficient (DSC) of 88.21\% for localization of pneumothorax. Thus, the outstanding abilities of deep learning techniques can be deployed for developing robust models for pneumothorax detection.

INDEX TERMS Chest radiographs, chest x-rays datasets, class-imbalance, deep learning, ensemble, machine learning, pre-processing, pneumothorax, detection.

I. INTRODUCTION

Pneumothorax, also known as Collapsed Lung, refers to a condition in which the air present inside the lung is leaked into the pleural cavity (empty space between lungs and the chest walls). This air exerts pressure on the lungs which results in collapse of lungs. The most commonly observed symptoms include sudden sharp pain in chest and shortness of breath.

It occurs because of several reasons such as injury in chest area or some other lung diseases, however sometimes there is no apparent reason for the lung collapse [1]. Almost 30-39\% patients suffering from chest trauma have chances to develop pneumothorax [2]. The size of affected area of the lung by pneumothorax varies from small to large. Small or medium sized pneumothoraxes (i.e., up to 40\% affected lung) are not life-threatening if treated in time. However, delayed diagnosis and treatment may lead to death especially in the patients with large-sized pneumothorax or those whose are already on
ventilators [3]. Thus, the early diagnosis of pneumothorax is a crucial thing and require expert medical personnel. Several medical imaging techniques are available for diagnosis of chest diseases, including Chest radiography, Magnetic Resonance Image (MRI), computed tomography (CT) scan and Ultrasound. However, because of the cheap cost and availability of X-ray machines almost everywhere, doctors prefer X-rays instead of other medical imaging techniques [4].

Analyzing chest radiographs requires a lot of expertise because of the complex overlapping structure of thoracic cavity and the fact that the position and size of pneumothorax vary from patient to patient. But the problem is that there is a smaller number of radiologists throughout the world and a single radiologist has to analyze thousands of X-rays per year [5]. Additionally, CXRs have lower resolution as compared to CT scans or MRIs, which makes the diagnosis of pathology difficult and time consuming. Hence in order to lessen the burden on the radiologist and to provide him a second opinion while analyzing a radiograph, a Computer Aided Design (CAD) can be a game changer, and it has become a topic of interest in the field of chest radiography. The very first CAD system was proposed in 1960 and the results proved that the diagnostic performance was increased by a CAD system [6]. In the medical field, CAD systems have been used for several purposes including arrhythmia detection [7], classification of skin cancer [8] and diabetic retinopathy identification [9]. Any CAD system comprises of four main steps [10]. (1) Preprocessing of the data (2) extraction of region of interest. (3) Feature extraction step, which can be done using different ways like Local binary pattern, textural feature extraction or fixed feature extraction using Convolutional Neural Networks (CNN). (4) Classification of the input into respective class based on the extracted features. The last step needs a classifier to be trained on some data, (which are the features extracted from the input data). Different machine learning classifiers include Linear Regression, Support Vector Machine, Naive Bayes Classifier, Random Forest or Neural Networks [11]. Before the advent of deep learning techniques, different thoracic pathologies have been successfully detected using traditional machine learning approaches [12]. However, it’s main drawback is that the features are handcrafted so the choice of optimal features is an important decision to be made. On contrary, the process of feature extraction and classification have been automated by deep learning techniques. Convolutional Neural Networks have achieved astonishing results in automating several processes in different fields of life including agricultural domain [14], surveillance [15] and medical field [16].

The only downside of deep learning methods is the dire need of enormous data for training purpose, which is not easy to acquire, and it is time consuming to label and annotate unlimited number of samples. However, this problem is solved to some extent by the availability of large number of annotated datasets on different platforms like Kaggle [17] and Google-Dataset-Search [18]. These free datasets allow the researchers to contribute to the automation of relevant field. Although huge number of medical datasets are now publicly available, they face the problem of class-imbalance [19]. This issue is resolved in different research by using various techniques including data level and algorithm level techniques [20].

In the field of automatic diagnosis of thoracic pathologies, different CAD systems have been proposed which found their roots from deep learning. Examples include lung nodule detection [21], detection of pneumonia and tuberculosis [22], [23] and diagnosis of other lung diseases including pneumothorax [24].

A systematic literature review is important as it summarizes and analyzes the existing literature on relevant topic along with highlighting the limitations and shortcomings of existing research. Such literature reviews are available for pneumonia, TB and multiple thoracic diseases [4], [10], [13], however, regardless of the fact that lot of research has been done for automatic diagnosis of pneumothorax, a brief overview of the existing work for pneumothorax is not available so far. Thus, there is a need to explore and summarize the available literature. This paper presents an overview of the existing work for pneumothorax detection using Chest radiographs. The main contributions of this paper are summarized below:

- Analysis of the usability, pros and cons of the techniques utilized for pneumothorax detection.
- Briefly summarizing the openly available CXRs datasets for automatic diagnosis of pneumothorax.
- Providing a comparative analysis of existing work on pneumothorax detection.
- Suggesting research ideas on the basis of existing literature which will help the researchers in selection of dataset and optimal technique on the basis of available resources and real time scenarios.

The remaining paper is organized as follows. Section II briefly talks about the research methodology used for the systematic literature review. The different openly available Chest X-rays datasets are mentioned in detail in Section III. Commonly adapted data pre-processing techniques are described in Section IV. In Section V the techniques proposed by researchers for pneumothorax detection are discussed. Section VI gives the overview of the evaluation metrics used in most of the literature and Section VII provides comparative analysis and discussion for the existing work. Limitations for this study are provided in Section VIII. In Section IX conclusion is drawn based on the entire review.

II. RESEARCH METHOD

An unbiased research methodology is adopted for a systematic literature review (SLR) in order to ensure the evaluation of all the existing research relevant to the said field.

A. DATA SOURCES

The literature is obtained from different electronic databases including IEEE Xplore, Springer, ScienceDirect, arXiv,
The total number of articles retrieved was 7,940. Additional 2,740 research articles were obtained from Google Scholar database. After removing duplicate records from the retrieved articles, 6,552 articles were left which were screened on the basis of titles. Finally, 142 full-text articles, out of which 33 articles met the selection criteria mentioned in Section II-C. These articles were analyzed in terms of goodness and drawbacks of the technique deployed for pathology detection. In order to provide relevant information (like data preprocessing) for pathology detection, additional articles were also accessed. The whole process of article selection was carried out by first two authors. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [26] diagram for the literature selection is shown in Fig. 1.

**B. SEARCH TERMS**
Certain search terms have been utilized in order to search the relevant literature and the valuable work done in the field of automatic diagnosis of pneumothorax.

- Pneumothorax OR pneumothorax detection using artificial intelligence.
- Automatic diagnosis of chest pathologies OR pneumothorax from chest radiographs.
- Chest X-rays datasets OR datasets for pneumothorax identification.

Mentioned below is the search string for the said purpose:
Pneumothorax OR chest x-rays OR artificial intelligence OR automatic detection.

**C. STUDY SELECTION PROCEDURE**
Some criteria are adapted in order to select the relevant studies out of all the literature retrieved from the data sources. These inclusion and exclusion criteria are briefly explained below:

1) **INCLUSION CRITERIA**
   - Studies relevant to pneumothorax detection by using chest radiographs.
   - Studies that were not just specified to pneumothorax, instead covered multiple chest diseases detection including pneumothorax.
   - Studies utilizing public and private datasets.
   - Studies published in peer reviewed publications along with the pre-prints (arXiv).
   - Studies that were published from 2010 to 2020.
   - Studies using English language as mode of sharing the research.

2) **EXCLUSION CRITERIA**
   - Studies which did not add value in the field of automatic diagnosis of pneumothorax.
   - Studies relevant to pneumothorax using chest radiography reports in textual format.
   - Studies relevant to pneumothorax using medical imaging techniques other than X-rays such as CT scans and MRIs.
   - Studies published in languages other than English.
   - Web articles, Wikipedia, posters, and short papers.

**D. QUALITY ASSESSMENT**
The quality of all the selected 33 articles was accessed based on following five questions. Each question was given a weight of 1. The passing score for quality assessment was set as 3 (inclusive) by the leading authors [25]. It can be seen in Table 1 that all our included papers have achieved score greater than or equal to 3. The questions for quality assessment of our systematic literature review are defined below:

**Q1.** Are the research objectives clearly defined?
**Q2.** Are the selected approaches and techniques clearly explained?
**Q3.** Is the class-imbalance issue explicitly discussed?
**Q4.** Are the result evaluation metrics clearly defined?
**Q5.** Are the limitations well mentioned?

**III. DATASETS**
The openly available dataset which can be utilized for automatically detecting the presence of pneumothorax are discussed in this section.

**A. NIH CHEST X-RAY 14 DATASET**
A very large chest x-rays dataset was presented by Wang in 2017 [27] containing 112,120 CXRs from 32,717 patients and named as Chest X-ray 14 dataset [28]. Picture Archiving and Communication Systems (PACS) of the hospitals affiliated
TABLE 1. Quality assessment for each paper.

| Author | Q1 | Q2 | Q3 | Q4 | Q5 | Marks (Total=5) |
|--------|----|----|----|----|----|----------------|
| Geva [102] | Yes | Yes | N/A (balance dataset) | Yes | No | 4 |
| Chan [58] | Yes | Yes | N/A (balance dataset) | Yes | No | 4 |
| Jun [74] | Yes | Yes | Yes | Yes | Yes | 5 |
| Taylor [40] | Yes | Yes | Yes | Yes | Yes | 5 |
| Sze [75] | Yes | Yes | Yes | Yes | No | 4 |
| Park [108] | Yes | No | Yes | No | 3 |
| Luo [41] | Yes | Yes | N/A (balance dataset) | Yes | No | 4 |
| Ouyang [112] | Yes | Yes | No | Yes | No | 3 |
| Mostayed [92] | Yes | Yes | No | Yes | No | 3 |
| Gooden [114] | Yes | Yes | No | Yes | No | 3 |
| Vintasvkyi [76] | Yes | Yes | Yes | Yes | No | 4 |
| Abedalla [61] | Yes | Yes | No | Yes | No | 3 |
| Groza [73] | Yes | Yes | Yes | Yes | No | 4 |
| Jakhar [60] | Yes | Yes | No | Yes | No | 3 |
| Tolkachev [110] | Yes | Yes | No | Yes | No | 3 |
| Wang Yaqi [107] | Yes | Yes | Yes | Yes | Yes | 5 |
| Wang [27] | Yes | Yes | Yes | Yes | No | 4 |
| Rajpurkar [22] | Yes | Yes | Yes | Yes | Yes | 5 |
| Yao Li [26] | Yes | Yes | No | Yes | No | 3 |
| Li [45] | Yes | Yes | Yes | Yes | No | 4 |
| Kumar [116] | Yes | Yes | Yes | Yes | No | 4 |
| Yao [42] | Yes | Yes | Yes | Yes | Yes | 5 |
| Guan [117] | Yes | Yes | No | Yes | No | 3 |
| Rubin [59] | Yes | Yes | No | Yes | Yes | 4 |
| Salehinejad [125] | Yes | Yes | Yes | Yes | No | 4 |
| Baltruschat [128] | Yes | Yes | Yes | Yes | No | 4 |
| Tang [119] | Yes | Yes | N/A (balance dataset) | Yes | No | 4 |
| Allouzi [43] | Yes | Yes | Yes | Yes | No | 4 |
| Guendel [120] | Yes | Yes | Yes | Yes | No | 4 |
| Mojkowska [122] | Yes | Yes | No | Yes | Yes | 4 |
| Mondal [123] | Yes | Yes | No | Yes | No | 3 |
| Bharati [124] | Yes | Yes | No | Yes | Yes | 4 |
| Prakash [44] | Yes | Yes | No | Yes | No | 3 |

Q1. Are the research objectives clearly defined? Q2. Are the selected approaches and techniques clearly explained? Q3. Is the class-imbalance issue explicitly discussed? Q4. Are the result analysis techniques clearly defined? Q5. Are the limitations well mentioned?

with National Institutes of Health Clinical Center (NIH) were used for acquiring the chest radiographs. This dataset initially covered 8 thoracic pathologies but later it was extended to 14 chest diseases including atelectasis, edema, consolidation, emphysema, cardiomegaly, effusion, mass, fibrosis, infiltration, hernia, nodule, pneumothorax, pleural thickening and pneumonia. Out of 112,120 CXRs, 60,412 samples are those with no-pathology while remaining 51,708 CXRs have one or more pathologies. In the whole dataset 5298 CXRs are labelled to have pneumothorax. Moreover, the dataset also provides bounding boxes information which can be used for disease localization, however, this information is available for 980 CXR images only. The resolution of each CXR in this dataset is $1024 \times 1024$. In addition to CXR images, the metadata, e.g., patient gender, age, and view position of the CXR are also available. Training and testing lists are also provided in this dataset in which the samples are split in such a manner that CXRs from the same patient belong to either training or testing list.

B. RANDOM SAMPLE OF NIH CHEST X-RAY DATASET (RS-NIH)

A small version of NIH Chest X-ray 14 (NIH CXR14) dataset is available on Kaggle [29] and contains 5% of the whole dataset which were randomly selected. The samples were selected in such a way that the percentage of occurrence of each pathology is the same as the NIH Chest Xray14 dataset. Total of 5606 CXRs are present covering 14 thoracic pathologies and 15th class being the samples with no-pathology. Just like in the NIH CXR14 dataset, the resolution of each image in this dataset is $1024 \times 1024$. Among 5606 CXRs, 3044 images are of healthy persons while remaining CXRs correspond to one or more thoracic diseases. Metadata like patient age, patient gender, patient’s follow up, view position etc. are also supplied for each CXR image in the database.

C. CHEXPERT

Another large dataset was made available in 2019 by the Stanford group [30], [31]. It contains 224,316 Chest x-rays of 65,420 persons obtained from Stanford Hospital, collected during 2002 to 2017. It covers 14 common chest diseases including atelectasis, enlarged cardamom, lung lesion, cardiomegaly, edema, fracture, lung opacity, pneumonia, consolidation, pleural effusion, pneumothorax, pleural other, no finding and support devices. The typical size of each CXR is $320 \times 320$ in this dataset. A unique feature of this dataset is that it provides uncertainty labels as well for every pathology, i.e., those CXRs in which the labeler couldn’t properly interpret regarding presence or absence of certain pathology is stated to be uncertain for a particular pathology. These uncertainty labels can be dealt with different techniques [30], [32]. The dataset contains 16627 CXRs with no pathology while remaining CXRs have either one or more chest pathologies. 17313 samples belong to pneumothorax class while there is uncertainty regarding the declaration of 2663 CXRs to have pneumothorax.

D. MIMIC-CXR-JPG

To date, it is the largest openly available Chest X-rays dataset, which was made public in 2019 [33], [34]. It contains 377,110 CXRs which are related to 22,782 radiologic studies and covers 14 chest diseases including No Finding, atelectasis, cardiomegaly, edema, consolidation, enlarged
TABLE 2. Summary of databases with respect to number of pneumothorax samples.

| Publication Year | Dataset                          | Total No of samples | No. of Pnu samples | No. of class in dataset | Official Data split | Data Collection Year | Age Group (Years) | Gender Distribution (%) |
|------------------|---------------------------------|---------------------|--------------------|-------------------------|---------------------|---------------------|---------------------|------------------------|
| 2017             | NIH Chest X-ray14 dataset [28]  | 112,120             | 5298               | 14                      | ✓                   | 1992-2015           | 01-94               | 50-58                  | 56                     | 44                     |
| 2017             | Random Sample of NIH Chest X-ray dataset [29] | 5606             | 271                | 14                      | ×                   | 1992-2015           | 01-94               | 47-66                  | 56                     | 44                     |
| 2019             | CheXpert [30]                   | 224,316             | 17,313             | 14                      | ✓                   | 2002-2017           | 18-90               | 59-66                  | 59                     | 41                     |
| 2019             | MIMIC-CXR-JPG [34]              | 377,110             | 9317               | 14                      | ✓                   | 2011-2016           | N/A                 | N/A                    | N/A                    | N/A                    |
| 2019             | SIIM ACR Pneumothorax Segmentation [37] | 10,675 (stage1)    | 12,047 (stage2)    | 2                      | ✓                   | 1992-2015           | 07-85               | 50-60                  | 55                     | 45                     |

Pnu: Pneumothorax

Cardio mediastinum, lung lesion, fracture, lung opacity, pleural Effusion, pneumothorax, pneumonia, pleural other and support Devices. The CXRs were acquired from Beth Israel Deaconess Medical Centre (Massachusetts, USA) in the duration of 5 years (2011 to 2016). 75163 images are of the people with no pathology while remaining CXRs belong to the patients who have one or multiple thoracic pathologies. There are 9317 CXR images which are labelled as pneumothorax in the whole dataset while for 868 images, there is uncertainty regarding presence of pneumothorax.

E. SIIM ACR PNEUMOTHORAX SEGMENTATION

A very recent dataset provided by Society for Informatics in Medicine (SIIM) in 2019 is a very major contribution in the field of automatic detection and localization of pneumothorax [35]. The images were acquired from NIH Chest Xray14 dataset which were later annotated on pixel level using artificial intelligence techniques and reviewed by expert radiologists. The dataset is available on Cloud Healthcare [36], [37] in which the CXR images are provided in Digital Imaging and Communications in Medicine (DICOM) format while corresponding masks for each image are in Run-Length Encoding (RLE format) format, however, many individuals have shared the same dataset in PNG format on Kaggle [38]. Since this dataset was released as challenge/competition and invited people to participate in developing robust segmentation model, thus it has two stages, the stage1 has 10,675 samples for training purpose and 1372 samples belong to the test set, while in stage2, the number of samples in both training and testing set was increased to 12,047 and 3205 samples respectively. Most of the researchers utilizing this dataset have evaluated their proposed models on these held-out test sets (i.e., stage1 test set and stage2 test set).

F. SUMMARIZED DETAILS OF EACH DATASET

The total number of CXRs labelled as pneumothorax in each of the above-mentioned dataset is stated in tabular form in Table 2. The Table 2 also highlights if the dataset provides data-splits for training and testing purposes or not, which is important for the purpose of uniformity. Additionally, the subject demographics like gender distribution and age groups are also mentioned which have been obtained by analyzing the metadata provided with these datasets. It is to be noted that subject demographics are not available for MIMIC-CXR-JPG as it does not provide the supplementary metafiles [39].

IV. DATA PRE-PROCESSING

In almost all CAD systems, the input data has to go through several pre-processing steps. These steps may include:

- Formatting & Resizing
- Enhancement
- Region of Interest (ROI) Extraction
- Data balancing

If not all, involvement of one of the above-mentioned pre-processing steps is a must in CAD system for chest pathology detection. These preprocessing steps are briefly explained in this section.

A. FORMATTING & RESIZING

Mostly the chest radiographs are acquired in DICOM format, which not only contains the X-ray image of the chest but also contains meta-data like patient’s name, age, gender etc. [10]. This DICOM format is not only incomprehensible (except for radiologists), but also many machine learning algorithms are programmed to take the input image in other formats like JPEG or PNG. So, in order to change the format of the CXR from DICOM to one of the above-mentioned image formats, first of all character recognition and image processing techniques are used to remove the patient metadata from the CXR image. After the removal of metadata, the image in DICOM format is converted to one of the previously stated formats (JPEG, PNG or bitmap) using compression algorithms [33], [40].

A fixed resolution of input image is required in any CAD system. Thus, resizing of the CXR image is a very common step in most of the proposed models for automatic diagnosis of chest pathologies including pneumothorax [41]–[45]. Some of the common techniques for resizing of the CXR images include downsampling,
upsampling by nearest-neighbor-interpolation and bilinear interpolation [46], [47].

B. ENHANCEMENT
Quality of the chest X-ray image is one of the key factors for robust classification and accurate segmentation of region of pathology. Research prove that the performance of any CAD system improves greatly when pre-processed and enhanced images are used as compared to the unprocessed data. So, several image processing techniques are adopted in order to improve the chest radiographs quality while highlighting the structural information and noise removal. Some of the commonly adapted image enhancement approaches for chest radiographs include edge sharpening [48], [49], noise elimination [43], [50] and contrast enhancement [51]–[54].

Edge sharpening is a process of enhancing the edges and contours of the image, thus highlighting and revealing the fine details. This can be done by using different types of filters including Sobel Edge and Canny Edge detection methods [55]. Noise elimination refers to denoising the image with the aim to preserve the image details as much as possible using filters such as Mean, Median and Gaussian filters [56]. Contrast enhancement is a way of changing the range of an image’s brightness values which enhances the overall contrast of the image, thus making the details more visible. One of the commonly used techniques for contrast stretching is histogram equalization [57]. It can be safely said that pre-processing of the image before feeding it to any machine learning algorithm increases the accuracy by removing the noise and distortions while preserving the important details in the image.

C. REGION OF INTEREST (ROI) EXTRACTION
In the medical field, diagnosis of a disease is not sufficient without the exact information of the region of pathology. Thus, in CAD systems, extraction of region of interest (ROI) is highly appreciated. In automatic diagnosis of chest pathology, segmentation of ROI can either be a preprocessing step [43], [58], [59] or it might be the sole purpose of the research [27], [60], [61]. Segmentation can be done using image-processing and deep learning-based methods. In image-processing based methods, ROI extraction is carried out using certain rules such as pixel location, texture and intensity values. Methods include thresholding, region-based methods and edge detection methods [62]–[65]. In Deep learning-based methods, each pixel of the image is assigned a class label, either background or foreground. The aim is to extract the whole lung or region of pathology from the whole CXR image. From these extracted ROI, features such as gray scale value, textural information or location information are extracted which can then be fed to any classifier like Support Vector machine (SVM), Random Forest (RF) or neural network [4]. Commonly used segmentation methods include Fully Convolutional networks (FCN) [66] such as U-Net and SegNet [141], [67].

D. DATA BALANCING
It is a known fact that most of the medical image datasets including chest radiographs face the problem of Class-Imbalance. A dataset is said to be class-imbalanced if it has an unequal number of samples for each class [68]. The problem with class-imbalance data is that when trained, as it is, on a machine learning classifier, it may yield high accuracy, but this performance is biased towards the majority class. Such a model is not useful, especially in the medical domain, where the main concern is accurate prediction of minority class [69]. Thus, it is important to balance the dataset before training on a classifier or segmentation model.

The solutions mostly adapted for solving class imbalance can be categorized as data-level, algorithm-level and hybrid approaches [70]. Data-level approaches solve the class-imbalance problem by modifying the class distribution in a dataset [40], [71]–[73]. In algorithm level approaches, the class-imbalance problem is solved by altering the classifier algorithm [20], [74]–[76]. Hybrid approaches utilizes the techniques from both data-level and algorithm level approaches.

1) DATA-LEVEL APPROACHES
a: OVERSAMPLING
As the name suggests, oversampling solves the class-imbalance problem by increasing the samples of minority class. The most commonly used method for over sampling is Random Oversampling (ROS) which randomly duplicates the minority class samples [77], however this may lead to overfitting. A more efficient way of oversampling is Synthetic minority over-sampling technique (SMOTE) in which synthetic samples are created by interpolating the neighboring data samples [78]. However SMOTE fails to perform properly when the data is not linearly separable. This issue was solved by LLE-SMOTE [79] in which locally linear embedding algorithm (LLE) was incorporated with traditional SMOTE algorithm. Cluster-based oversampling is another way to solve class-imbalance problem in which the clusters of datasets are created, which are then over-sampled separately [80], this method is efficient for solving the intra-class and inter-class imbalance problem. Another oversampling technique is DataBoost-IM which utilizes the concept of boosting to identify difficult samples from minority class in order to generate synthetic data [81]. Recently, Generative Adversarial Network (GAN) based on neural networks have gained attention because of high performance [82], in which synthetic samples for minority class are generated based on the learned data. Several outperforming variants of GAN have been proposed including DCGAN [83], CycleGAN [84] and styleGAN [85].

b: UNDERSAMPLING
One of the popular ways of dealing with class-imbalance problem, in which samples from the majority class are randomly removed so that all classes in the dataset have same sample size. Although undersampling has been proven to be
better as compared to oversampling in some cases, however, the fact that it might remove some important data samples has directed the researchers to optimize the concept of undersampling [86]. E.g., in [87], a modification of undersampling named as One-Sided Selection was proposed which discards the redundant samples instead of random samples.

2) ALGORITHMS-LEVEL APPROACHES

a: COST-SENSITIVE LEARNING

Minority class samples are often misclassified due to the fact that traditional classifiers minimize the least square error. In cost-sensitive learning, different costs are assigned to majority and minority class. Weight-balancing [88] is one of the cost-sensitive learning approaches in which different weights are assigned to the majority and minority class, calculated by (1).

\[
C_i = \frac{(1 + B) C_{\text{majority}} + (1 - B) C_{\text{minority}}}{2m} - \frac{\text{if } y_i = +1}{2m} - \frac{\text{if } y_i = -1}{2m} \quad (1)
\]

for \( i = 1 \ldots m \) where \( C > 0 \) and \(-1 < B < 1\). B is called the balancing factor. In (1), the denominators cater the unequal class distribution in the dataset and an additional compensation is provided by parameter B.

Weighted extreme learning machine (WELM) and weighted support vector machine (WSVM) are also similar approaches to assign different costs to majority and minority class samples [89], [90], however the limitation of this strategy is to decide the weights to be assigned to majority and minority class. Another cost-sensitive technique is to manipulate the learning rate so that samples with more cost have greater contribution in updating the weights and then network training is carried out with the aim to lessen the misclassification cost [91]. Utilization of loss function such as Cross-Entropy loss function (CEL), and Weighted CEL (W-CEL) are relatively new approaches to solve the class imbalance problem [27], [41], [75], [92].

b: THRESHOLDING

Threshold-moving or post scaling is one of the algorithm-level approaches which involves adjustment of the classifier’s decision threshold in the test phase. The class-probabilities of the output are also changed based on the threshold [93]. Different ways are available to adjust the outputs of the network. The most commonly used method is prior class probabilities. Prior class probabilities are calculated based on the frequency of each class in the dataset. In neural networks, posterior probabilities are calculated by Bayesian theorem [94].

3) HYBRID APPROACHES

In this approach, different techniques are adapted from data-level and algorithm level methods. One of the most commonly used methods is ensemble models. BalanceCascade and EasyEnsemble are kind of ensemble models in which the dataset is divided into several balanced subsets and different algorithm level techniques are then used to train classifiers on balanced subsets [95]. Another ensemble-based model is SMOTEBest, which utilized the concept of boosting and SMOTE oversampling [96]. Several researches have proved the effectiveness of previously mentioned ensemble models [97]–[99].

V. PNEUMOTHORAX DETECTION METHODOLOGIES

This section is divided into two main categories. Firstly, the research done exclusively for automatic detection of pneumothorax is discussed, followed by the existing valuable literature covering multiple pathologies detection including pneumothorax. These two categories are sub-categorized as Machine learning (ML) and Deep learning (DL) based approaches [100], [101]. These sub-categories are split into either Classification (i.e., publications in which main focus was to label a CXR as either normal or pneumothorax), Localization (papers in which the main goal was to achieve the area of pneumothorax) or Classification and Localization (research which covers both classification of CXR as either normal or pneumothorax, and identification of region of pathology).

A. PATHOLOGY SPECIFIC RESEARCH

1) MACHINE LEARNING BASED APPROACHES

a: CLASSIFICATION

Geva [102] proposed an exclusive framework for automated diagnosis of pneumothorax from the chest x-ray images by combining the texture analysis method with supervised learning techniques. Two main steps were involved in the proposed framework. (1) Detection of abnormal texture of the lung caused by pneumothorax, using the texture analysis process. Two types of textural feature extraction approaches including Local Binary Pattern (LBP) and Maximum Response Filters (MR) were experimented in this research. This way labeled patches of images were generated after which the local analysis values were integrated with global image representation. (2) The global images were indulged into training purpose for identifying the presence of pathology. The global images used for training were designed on the basis of the shape of the lung. Both local and global data was incorporated for a supervised learning process. Gentle AdaBoost and KNN were used as classifiers for the training purpose. Local dataset was used in this research which was obtained from Sheba Medical Center, and it comprised a total of 108 cases out of which 48 CXRs were diagnosed with pneumothorax. When evaluated on the test set the proposed model achieved 81% sensitivity and 87% specificity value.

b: CLASSIFICATION AND LOCALIZATION

Machine learning techniques were employed for automatic detection and localization of pneumothorax in [58]. For the classification part, firstly the lung regions were extracted from the CXR images after which features were obtained using the Uniform Local Binary pattern (ULBP) method. Support Vector Machine (SVM) with RBF kernel was used.
as a classifier. The second method was the segmentation of the abnormal areas of the lung from the CXR images. The extraction of the lung region was carried out after removing the background and noise in the CXR images, later the textural analysis was performed using Local Binary Pattern in order to define the smooth and complex regions. Moreover, the rib boundaries were highlighted using the Sobel Edge detection method. The final image was the segmented abnormal region with pneumothorax in form of a black-and-white mask. The dataset was obtained from a Chinese Hospital containing 84 CXRs. For training purpose 36 normal CXRs and 22 CXRs with pneumothorax were used while the evaluation was done on 16 normal and 10 pneumothorax CXRs. For classification purpose 82.2% accuracy was obtained while for the second proposed method 85.8% accuracy was achieved on the test set.

2) DEEP LEARNING BASED APPROACHES

a: CLASSIFICATION

Jun [74] proposed an ensemble-based framework in order to distinguish between normal and pneumothorax CXRs. The ensemble model comprised of three identical CNN architectures trained on three different input sizes. Pretrained ResNet50 [103] architecture was used which was trained on the same training set with different resolutions, i.e., 512 × 512, 384 × 384 and 256 × 256. ImageNet weights were used for model initialization and RMSprop optimizer was selected for model training. The ensemble was created by averaging the probabilities generated by each of the trained models. NIH Chest X-ray 14 dataset (NIH) was used in this experiment which contains 112,120 images however for this study, 59,156 samples of Normal CXRs and 5225 samples of pneumothorax were used. These 64,381 CXRs were randomly divided into 80% and 20% for training and testing purposes respectively. To avoid the problem of class imbalance, instead of directly using the probability from the softmax layer, a cut-off value was selected so as to maximize the sum of specificity and sensitivity in the ROC curve. When evaluated on the test set, the ensemble model achieved AUC of 91.1% for pneumothorax detection.

Taylor [40] collected Chest X-ray images from Center for Digital Health Innovation at UCSF in order to design a model for automatic detection of pneumothorax. The dataset contained 13,292 CXRs belonging to two classes, pneumothorax and No-pathology. The images were in DICOM format which were converted to JPEG and resized to 512 × 512. The dataset was split into 70% training, 15% validation and 15% testing set. The training set had the issue of class imbalance as it contained 2214 positive and 7095 negative samples, so this issue was resolved using random under-sampling technique. Image Augmentation was also used in order to avoid overfitting. Training with several CNN architectures including VGG-16, VGG19 [104], ResNet, Inception [105] and Xception [106] was carried out in this study. Performance was assessed on the internal and external test set which was obtained from NIH Chest Xray14 dataset. The results showed that there was a performance decline in case of external test set. For the internal test set the proposed framework attained AUC of 94% with sensitivity of 55% and specificity of 90% while on the external test set AUC of 75% with 49% sensitivity and 85% specificity was achieved.

Sze [75] proposed a pneumothorax detection model in which the power of transfer learning was explored. A 122-layered deep neural network was proposed which was named as tCheXNet based on the fact that it utilized the trained model CheXNet [22] in order to initialize the model weights. CheXpert dataset was selected for this research. The model was trained using 94,948 CXRs in which only 13911 CXRs belonged to pneumothorax class, this class- imbalance problem was solved by using weighted binary cross entropy loss function. The proposed framework was tested on a test set containing 7 pneumothorax images and 195 normal CXR and achieved AUC of 70.8%.

A study was carried out by Park et al. [108] in order to estimate the performance of CNN for the automatic diagnosis of pneumothorax from the CXRs after Percutaneous transthoracic needle biopsy (PTNB). PTNB is a famous and widely used technique for detection of pulmonary lesions. The data was obtained from PACs of two local hospitals, and it comprised of 1596 CXRs with pneumothorax and 11137 normal CXRs. During training, 1343 pneumothorax CXRs were used, out of which 90% samples were randomly assigned to training set and remaining 10% were treated as validation set. A held-out test set containing 250 normal cases and 253 images with pneumothorax were used for internal testing purpose. Pretrained YOLO Darknet-19 [109] was fine-tuned on the CXRs dataset with SGD optimizer and initial learning rate was set to 0.001. The input resolution was kept as 1024 × 1024. Contrast Limited Adaptive Histogram Equalization was used for preprocessing. The trained model achieved sensitivity of 89.7% and specificity of 96.4% with AUC of 98.4% on the internal test set. The model was also tested on an external validation set containing 309 pneumothorax and 1020 normal CXRs and achieved sensitivity, specificity and AUC of 63.4%, 93.5% and 90.5% respectively.

In [107] a framework evolved from deep learning techniques was proposed for diagnosis of pneumothorax. The classification model was named as ChestNet in which data cleaning was done using two Networks in Network (NIN) and a new data augmentation method based on random histogram equalization was introduced for data imbalance problem. The training and evaluation of the proposed model was done on two datasets including ZJU-2 and NIH Chest Xray14 dataset. The ZJU-2 dataset was obtained from a hospital affiliated with Zhejiang University School of Medicine China. The CXRs were randomly divided into 80% training and 20% test set. For training purpose, the samples of ZJU-2 and NIH were combined together. The training set contained 26891 non-pneumothorax samples, 1507 pneumothorax samples from ZJU-2 and 5298 pneumothorax samples from NIH. The proposed model achieved AUC of 98.44%...
of 256 were experimented while keeping input resolution including U-Net, LinkNet and Tiramisu (FCDenseNet67) segmentation purpose, three different types of architectures utilizing Guided Attention Inference Network (GAIN). For model and the weakly annotated masks were obtained by level classification was performed using the ResNet101 annotated on pixel level along with the weakly annotated label smoothing regularization) technique. (2) Segmentation had some errors which were corrected using SLSR (spatial a rough estimation of the area of pathology. These masks generated the attention masks, these attention masks gave two main stages. (1) Training an image-level classifier of 92.01% with Mean pixel accuracy (MPA) of 93.01%. Moreover, data augmentation by means of horizontal flip was done for the whole training set. Upon evaluation on the test set, the trained model achieved a Dice coefficient score (DSC) of 92.01% with Mean pixel accuracy (MPA) of 93.01%.

In [112] a deep learning-based framework was proposed to automatically segment out the region of pathology. The main idea behind this framework was to lessen the dependency on the pixel-level annotations for segmentation purpose. There were two main stages. (1) Training an image-level classifier which not only predicted the class of the CXR images but also generated the attention masks, these attention masks gave a rough estimation of the area of pathology. These masks had some errors which were corrected using SLSR (spatial label smoothing regularization) technique. (2) Segmentation model was trained using some of the images which were well annotated on pixel level along with the weakly annotated masks generated by the image level classifier. The Image level classification was performed using the ResNet101 model and the weakly annotated masks were obtained by utilizing Guided Attention Inference Network (GAIN). For segmentation purpose, three different types of architectures including U-Net, LinkNet and Tiramisu (FCDenseNet67) [113] were experimented while keeping input resolution of 256 × 256. The experimental data was obtained from Chinese Hospital, and it contained a total of 5400 CXRS with 3400 pneumothorax cases and 2000 normal CXRs. The dataset was evenly and randomly divided into training and testing sets, thus the training set comprised of 4000 CXRS while 1400 CXRs were present in the test set. When evaluated on a test set, the model performance was reported as 66.69% in terms of Intersection over Union (IoU) score.

A novel segmentation framework was proposed for identifying the location of pneumothorax in [41]. Three modules were combined in the proposed framework including Fully Convolutional (FC) DenseNet, scSE which is a spatial and channel squeeze and excitation module, and finally a multi-scale module. Single channel images which were resized to 256 × 256 resolution were used in this research. The dataset was obtained from the institutions’ PACs. The dataset contained 11051 frontal CXR images out of which 5566 belonged to Pneumothorax and 5485 were the x-rays of healthy persons. The data was split into training, validation and test set. For solving the pixel-level-class-imbalance, weights were added in the cross-entropy function and named as weighted cross entropy loss (W-CEL). Adam optimizer was adopted with initial learning rate set to $1e^{-4}$. Moreover, data augmentation by means of horizontal flip was done for the whole training set. Upon evaluation on the test set, the trained model achieved a Dice coefficient score (DSC) of 92.01% with Mean pixel accuracy (MPA) of 93.01%.

Mostayed [92] suggested a modification for the traditional U-Net architecture [141] which is famous for segmentation purpose. The modification was proposed with the aim to lessen the network’s trainable parameters by replacing the concatenation operation in the skip connection of traditional U-Net architecture with content-adaptive convolution. The proposed method was experimented with two different datasets, one of which was SIIM Pneumothorax dataset. 12,047 CXRS were utilized in this research which were split into training and validation sets. The input resolution of images was kept equal to 128 × 128. A customized loss function was defined by combining binary cross entropy and soft dice loss. The U-Net model achieved dice coefficient score of 75.36% with 7.76M parameters while the modified U-Net achieved 76.04% dice coefficient score with 7.09M parameters. Hence the modified U-Net not only increased the dice coefficient score but also reduced the resources utilization.

In [76], a weakly supervised segmentation model was proposed with the aim to reduce the dependency on pixel level annotation and utilizing only image-level labels. The proposed method consisted of three consecutive steps. (1) Class Activation Maps generation was done by training ResNet50 architecture using the image-level annotations of the training data and then GradCAM++ method was utilized in order to generate the activation maps from the trained model. (2) These generated activation maps were trained on IRNet, which is known for its ability to improve the inter-object boundaries. This step was performed with the aim of improving the quality of these maps prior to the segmentation task. (3) Finally, the output from the second step was trained on the U-Net model with ResNet50 backbone. Experiments were performed on two different datasets, one of which was SIIM Pneumothorax dataset. Here the training set of stage 1 containing 2379 samples of pneumothorax and 8296 normal cases was used. The stage1 test set was further split into validation and test set, thus validation and test had 145 positive and 541 negative cases each. Weighted binary cross entropy (BCE) was used for avoiding the class imbalance issue, while the overfitting problem was reduced by data augmentation method. SGD was used as an optimizer with initial learning rate of 6e-5. The image resolution was kept 512 × 512 with a batch size of 48. The model evaluation was done on the test set which achieved dice coefficient score of 76.69%.

Abedalla [61] proposed a two-stage segmentation framework (2ST-Unet) in order to extract the region of pathology. The main architectural block was U-Net with ResNet34 encoder, while the weights of the model trained on ImageNet dataset were used to initialize the encoder part. The decoder block comprised of five blocks containing several convolutional and batch normalization layers and RELU activation function layer. The first phase was trained with image resolution of 256 × 256 and then the weights from this phase were used to initialize the training of the second phase for which image resolution was kept 512 × 512. For model training,
Adam optimizer was used, and initial learning rate was set to 0.001 which was decayed by using cosine annealing scheduling method in every epoch. SIIM Pneumothorax dataset was used for this research with the officially provided data-split for training and testing set. 2669 pneumothorax and 9378 normal CXRs were present in the training set. The evaluation was performed on the test set of stage1 and stage 2 separately containing 1372 and 3205 samples respectively. Test Time Augmentation was also applied, and the trained model achieved a Dice coefficient score of 85.02% for stage1 and 83.56% for stage 2 test set.

A major contribution in the field of extraction of the area of pneumothorax was made in [73]. Here an ensemble of three LinkNet networks [111] was used with three different backbones including SENet154, seresnext50, seresnext101. The original encoder part of the LinkNet was replaced by these backbone architectures. The decoder block consisted of a convolutional layer followed by an up-sampling and another convolution layer. Sigmoid activation layer was used for acquiring the probability values. The training for each of the LinkNet models consisted of three phases. (1) Training for 40 epochs with Adam optimizer and Cosine Annealing scheduling. (2) Training for 15 epochs with CyclicLR scheduling. (3) Training using the initial setup for 15 epochs. Data augmentation during training and Test Time Augmentation were incorporated in this research. “Non-empty sapling” was performed in order to ensure the presence of positive samples in every batch. The experiment was done using SIIM Pneumothorax dataset. 10675 samples were present in the training set while the evaluation was done using stage1 and stage2 test set separately. The proposed ensemble model achieved dice coefficient score of 88.21% and 86.14% for stage1 and stage2 test sets respectively.

In [60], a segmentation model was proposed for the precise prediction of the portion of lung affected by pneumothorax. The model generated a binary mask to assist the radiologist in identifying the location and size of pneumothorax. U-Net architecture was used in this study in which the traditional encoder part was replaced with the ResNet model. Apart from resizing the images from 1024 × 1024 to 256 × 256, contrast correction was also done in order to obtain a uniform color range in the image. SIIM Pneumothorax dataset was used for experiment purpose containing 12,047 CXRs. 80% of the dataset was used for training purpose while the remaining 20% was used for validation and testing purpose. During training of the U-Net architecture, initial learning rate of 0.0001 with Adam optimizer was used. For better training, binary cross-entropy loss function and early stopping call-back were employed. The model performance was evaluated in terms of Dice Similarity Coefficient (DSC) and IoU, which was reported to be 84.3% and 82.6% respectively.

Tolkachev [110] proposed a segmentation model for identifying the area of lung affected by pneumothorax using SIIM Pneumothorax dataset containing 12,047 CXRs. A U-Net model was used in which various CNNs including SE-ResNext50, ResNet34, DenseNet121 and SE-ResNet101 were used to replace the traditional encoder. In addition to resizing the images from 1024 × 1024 to 768 × 768, some other preprocessing was also applied before training the segmentation model. In order to enhance the model performance, different data augmentation techniques were also exploited. The encoder part of the U-Net architecture was initialized with ImageNet weights while binary cross entropy was used as a loss function. Stratified five-fold-cross-validation method was adopted during the training process. The performance of the proposed model was evaluated on a test set containing 1372 CXRs for which it achieved a dice coefficient score of 85.74%.

c: CLASSIFICATION & LOCALIZATION

In [114], detection and localization of pneumothorax was performed using a local dataset containing 1003 CXRs obtained from University of Washington Medical Centre. The dataset had 437 pneumothorax samples and 566 non-pathology samples. Five-fold-cross-validation was used to split the data into training and testing sets. Three different deep learning techniques including CNN, Multiple Instance Learning (MIL) and Fully Convolutional Networks (FCN) were experimented. For CNN, a pre-trained ResNet-50 architecture was fine-tuned using single channel input with resolution of 448 × 448. MIL was used for both classification and localization using the previously trained ResNet50 model as patch classifier. For FCN, a U-Net model was deployed in order to detect the location of pneumothorax. The three mentioned models achieved AUC of 96%, 93% and 92% each. The authors concluded that CNN achieved the best result for classification purpose while MIL and FCN performed well for identifying the location of pneumothorax.

B. MULTIPLE PATHOLOGIES RELATED RESEARCH

Many authors have contributed to the field of automatic detection of multiple thoracic pathologies including pneumothorax. The problem of multiple pathologies classification is solved either as multi-class or multi-label classification problem. Multi-class classification problems are those in which the labels are dealt independently, and the classifier assigns any one of the N labels to each sample, where N refers to the number of labels/classes in a dataset, and \( N \in \{1, \ldots, \infty\} \). Multi-label classification problems are those in which a sample can be annotated with more than one label [115]. This section summarizes the valuable work done for the automatic detection of thoracic pathologies.

1) DEEP LEARNING BASED APPROACHES

a: CLASSIFICATION

A 121-layered CNN architecture was proposed in [22] for the detection of 14 pathologies from the chest radiographs which was named as CheXNet. It was basically a 121 layered Dense Convolutional Network, the weights of the model trained on ImageNet dataset were used for initializing the proposed
network. CheXNet was initially proposed for pneumonia detection but later it was extended to the detection of other thoracic diseases. The trained model generated a vector with binary labels and the length of the vector corresponded to the total number of classes in the classification problem, i.e., 14. NIH Chest X-ray-14 (NIH) dataset was used for this study which was divided in such a manner that 70% was used for training, 10% for validation and remaining 20% was declared as test set. While splitting the data, the “no-patient-wise-overlap” was ensured, i.e., same patient’s radiograph was not present in more than one split of the data. To cater the class-imbalance problem, binary cross-entropy was used as a loss function with Adam optimizer and input resolution of 224 × 224. The model performance was evaluated in terms of AUC with an achieved value of 88.87% for the detection of pneumothorax. Although heat maps were generated in order to visualize the area of pathology, however, the accurateness of localization was not reported.

Automatic detection of 14 thoracic diseases was carried out while solving the multi-label classification problem in [116]. Two approaches were used to tackle the multi-label classification problem, first one was training a CNN model with two loss functions including binary cross entropy (after transforming the multi-label problem into binary classification problem using Binary Relevance approach) and PairWise Error Loss. Second one was designing a six level cascaded architecture, especially designed to tackle the multi-label classification problem that took advantage from the training strategies of boosting methods. The base CNN architecture utilized in this research was DenseNet161 which was initialized with “He norm” for training purpose. During training Stochastic Gradient Descent (SGD) optimizer was used with a learning rate of 0.1 The class-imbalance problem was solved using weighted binary cross-entropy function. NIH Chest X-ray-14 dataset was used for experimentation which was randomly split into 80% training and 20% testing set. In case of pneumothorax detection, best results were obtained from the cascaded model with reported AUC of 85.94% on the test set.

In [42], multi-label classification problem was studied using openly available NIH Chest X-ray 14 dataset. The interdependency among labels was leveraged using LSTMs. An encoder and decoder based RNN was proposed for automated detection of thoracic diseases form the CXR images. The encoder part of the proposed classifier was based on the DenseNet. The images were resized to 256 × 256 and instead of transfer learning method, the model was trained from scratch. Maximum Log-Likelihood Estimation (MLE) was optimized during model training. The dataset was divided such that 70% was used for training, 10% for validation and 20% as test set while considering all the 14 classes. Weighted cross-entropy was used to tackle the class-imbalance. Upon testing, the model attained overall AUC of 79.8% while on pneumothorax 84.1% AUC was achieved.

In [117], a three-branch attention guided convolutional neural network (AGCNN) was proposed for the automatic diagnosis of 14 thoracic diseases including pneumothorax. The model was trained using global and local images on two baseline models, ResNet50 and DenseNet121. The global images were trained using transfer learning method and heat maps were generated which were used to crop the identified area of pathology in a CXR image. The cropped area was treated as local image, this way local images from all the training samples were obtained. These local images were then trained on CNN architecture. Finally, the last pooling layer for both the models (for training global and local images) were concatenated. Since the research dealt with multi-label classification problem, so final output was a vector with binary values, indicating the presence or absence of respective classes. NIH Chest X-ray-14 dataset was used which was randomly split in such a way that 70% was used for training, 10% for validation and 20% as test set. For training and testing purpose, the images were resized to 224 × 224, and model was trained using SGD optimizer. The proposed AGCNN achieved best results with DenseNet121 as baseline CNN architecture, with reported AUC of 87.1% on whole test set and 92.1% on pneumothorax detection.

For the automatic diagnosis of 14 different chest diseases, MIMIC CXR dataset was used, consisting of 473,064 CXR images [59]. CNN was trained with different experimental settings. The main task was to find the effectiveness of using different view position of the CXRs. DenseNet121 was used as base architecture with little modification. 70% of the data was used for training purpose, validation was done using 10% of data while the remaining 20% was used for testing purpose. Each subset was further divided based on the view positions i.e., lateral view position, anteroposterior (AP) and posteroanterior (PA) view positions. The CNN model was trained separately on these three viewpoints using ImageNet weights and input size was set to 512 × 512. A DualNet was also proposed in which both frontal view positions (AP and PA) along with Lateral view CXRs were used for training purpose. When evaluated, the results showed that DualNet performed better in terms of AUC with an achieved value of 72.1% on the whole test set and 62.5% for pneumothorax detection.

An automatic detection model was proposed with the aim to detect five different thoracic pathologies including pneumothorax in [125]. A deep CNN was proposed which was similar to AlexNet. The data was acquired from the local hospital’s Radiology Information System (RIS). One of the motives of this research was to explore the effect of oversampling by DCGAN on classification performance. The original dataset contained 4013 samples of pneumothorax, 15781 Normal, and remaining 36,626 samples belonged to four other pathologies. The resolution of input images was kept as 256 × 256 and the batch size of 128 was used during model training. The CNN model was trained for 100 epochs and Adam was chosen as optimizer. The model output was a vector with five probability values corresponding to each pathology. The evaluation was done on 1000 CXRs with equal contribution from each class.
It was found that overall performance was enhanced when DCGAN generated CXRs were added in the training set. For pneumothorax diagnosis, the model trained with the original dataset achieved accuracy of 57.99% while the accuracy value was increased to 88.84% when the model was trained with original and DCGAN-generated samples.

A multi-label classifier was designed with the aim to automatically diagnose 14 thoracic pathologies in [118]. NIH Chest Xray-14 (NIH) dataset was used for experimentation purpose. Two different protocols for data split were considered. Firstly, the data was randomly divided into 70% training, 10% validation and 20% testing set, and secondly the official data split provided by the NIH was used. ResNet50 was used as base architecture. Three main domains were covered in this research: (1) the effect of transfer learning with and without fine tuning and training a model from scratch. (2) Effect of high resolution of input image. (3) Effect of using non-image features like gender, age and viewpoint along with the image features in order to train the classifier. During training, the image resolution was kept as 448 × 448. To tackle the class imbalance problem, class-averaged Binary Cross Entropy (BCE) was used as a loss function. Best overall results were achieved when the model was exclusively trained on the selected dataset with non-image features incorporated during training. Although GradCAM was used to generate heat maps in order to visualize the area of pathology, however the precision of localization methodology was not reported. For the random split the best achieved results for pneumothorax detection were obtained when pre-trained ResNet50 was fine-tuned on the selected dataset with an achieved AUC of 87.0%. In case of official split, 84.0% AUC was achieved.

In [119], the automatic differentiation between normal and abnormal CXRs was the main motive. The main idea behind this study was to differentiate between normal and abnormal CXRs based on the synthetic image produced by the proposed model. If the trained model performed poorly while reconstruction of the input CXR, then the CXR was declared as abnormal one, i.e., the CXR belongs to one or more thoracic pathologies. The proposed framework was similar to Generative adversarial networks (GAN), and it comprised of three main modules. (1) Auto encoder. (2) A decoder. (3) Another Encoder which was programmed to increase the consistency between the latent spaces of the two encoders. For model training and evaluation, the resolution of images was set to 64 × 64. The experiments were performed using the NIH Chest Xray14 dataset and training was done using only the normal CXRs (with no pathology). The training data consisted of 4479 normal CXRs and 0 abnormal samples, while for validation purpose 849 normal and 857 CXRs with any pathology were used. In the test set, there were 667 abnormal and 677 normal CXRs present. The CXRs with at least one of the 14 thoracic diseases including pneumothorax were declared as abnormal. Four types of loss functions including image reconstruction loss, adversarial learning loss, encoding consistency loss and feature map consistency loss were utilized in this study. Upon testing, the model achieved AUC of 84.1% for discriminating between normal and abnormal CXRs.

In [43], automatic detection of 14 thoracic diseases including pneumothorax was performed using transfer learning method. The CNN architecture chosen for experiment purpose was DenseNet121, which was trained on the selected datasets using ImageNet weights as the network’s initializer. The last fully connected layer with 14 channels was removed and a 1024D feature vector was obtained from the trained model. As a preprocessing step, the CXR images were resized to 224 × 224. Adam optimizer was used, and initial value of learning rate was set to 0.001. 1024D features obtained from the DenseNet121 architecture were fed to the Logistic Regression classifier for training purpose. Since the research aim was to study multi-label classification, three different problem transformation techniques were experimented including Binary Relevance (BR), Classifier Chain (CC) and Label Power set (LP). Two openly available CXRs dataset were used in this study, i.e., 112,120 CXRs belong to the NIH chest X-ray-14 dataset (NIH) and 134,327 CXRs from CheXpert dataset. Each dataset was randomly divided such that 80% was used for training and 20% was kept aside for testing purposes. The results proved that BR performed best of all the three multi-label approaches experimented in this research. For the detection of pneumothorax, the model achieved AUC of 92.9% and 86% on the test set of NIH and CheXpert respectively. For the NIH dataset, another set of experimentation was performed with different data-split, i.e., 70% of data used for training and 20% data was used for testing purpose, while remaining 10% data was ignored. The trained model achieved AUC of 88.2% for pneumothorax detection.

A novel framework for detection of multiple thoracic diseases was proposed in which DenseNet121 was used as baseline CNN architecture in [120]. The main aspect of this framework was that it incorporated spatial information of the disease along with exploiting the high resolution of CXR image. It was found that training a model with higher resolution of input and providing location information of the pathology yield better classification results. The model training was carried out using two openly available datasets including NIH Chest Xray14 (NIH) dataset and PLCO dataset [121]. In this research, the classification problem was dealt as multi-class classification problem. 70% of data from each dataset was used as training set, 10% as validation set and the remaining 20% was used as test set. Patient-wise split was considered while allocating training, validation and test set. The training was done by combining the training sets of both the datasets, however, the labels were dealt independently, i.e., the images with the same label in both the datasets were treated as different classes. However, the testing was performed separately on the test set of NIH and PLCO datasets. In case of the NIH dataset both the official train-test split and random split of the dataset were considered while evaluating the proposed model. Weighted cross entropy loss was used in order to cater the class imbalance problem. The PLCO data didn’t contain...
pneumothorax samples, however, for the NIH dataset, AUC values of 84.6% and 87.2% were obtained for pneumothorax detection on the test sets with official split and random split respectively.

In [122], deep learning model was trained in order to detect four pathologies including pneumothorax, nodule, airspace opacity and fracture. Two different datasets were deployed in this study including a local dataset (DS1) obtained from different hospitals of India and second being NIH Chest Xray14 dataset. The DS1 contained total of 759,611 CXRs while 112,120 CXRs from NIH dataset were used in this study. The x-ray images were randomly split into training, validation and testing sets while considering that same patient’s CXRs were present in only one set, either training, validation or test set. Certain selection criteria were followed prior to training the model. The test set of DS1 contained 1818 while 1962 CXRs were present in NIH-test set belonging to the four pathologies. Here 94 samples from DS1 and 195 samples from NIH belonged to pneumothorax. Training was done using the Pre-Trained Xception model by combining the training sets of DS1 and NIH, while performance was evaluated on the test sets of two datasets separately. The model performance was evaluated in terms of AUC with obtained values of 95% and 94% for DS1 and NIH respectively.

A deep learning-based framework was proposed for the automatic diagnosis of multiple thoracic diseases from the CXR images in [123]. Xception model was used in this multi-label classification problem. The research was carried out using Random Sample of NIH Chest X-ray (RS-NIH) dataset which was randomly divided into 75% training and 25% testing set. The images were resized to 128 × 128. Adam optimizer with a learning rate of 0.001 was chosen in this study. The model was trained for 28 epochs with binary cross entropy loss function. All the 14 classes present in the dataset including pneumothorax were used in this experiment, however the No-finding samples were excluded during model training. The accuracy achieved by whole test set was reported to be equal to 88.76% while the AUC achieved for pneumothorax detection on the test set was 54%.

Bharati [124] proposed a new framework (VDSNet) for automatically identifying the presence or absence of thoracic pathology. The proposed framework was the combination of VGG16 architecture, data augmentation and spatial Transformer network STN. The CXR images were converted from RGB to gray prior to model training and the pixel values of images were normalized so that the pixel values lie within the range [0-255]. In addition to images, the Meta data like age, gender and view position of the CXR were also used for training purpose. Along with accuracy, which was the main performance measure in this research, $F_{0.5}$ was also calculated. Two datasets including Random Sample of NIH Chest X-ray (RS-NIH) dataset and NIH Chest X-ray-14 (NIH) dataset were used. Although both datasets have 14 classes each, however this research dealt it as a binary classification problem, i.e., the model was trained in order to detect if the CXR is normal or abnormal (any thoracic pathology). The proposed model achieved 70.8% accuracy and 64% $F_{0.5}$ score on RS-NIH while 73% accuracy and 68% $F_{0.5}$ score was achieved on NIH dataset.

Deep learning techniques were studied in [44] with the aim to detect the presence of thoracic diseases as a multi-label classification problem. The presented work comprised of two parts, firstly a five layered CNN architecture was proposed which was trained from scratch with random initialization of weights and secondly pre-trained VGG-16 architecture was fine-tuned on the selected dataset. Random Sample of NIH Chest X-ray (RS-NIH) dataset was used in this research, which was randomly split into 80% training 20% for testing purpose. Out of 80% data, 20% was declared as validation set. The images were resized to 128 × 128, and Adam optimizer was used with the initial value of learning rate set assigned to be 0.001. Since this study dealt with multi-label classification problems, categorical cross-entropy function loss function was used. GradCAM method was used to generate a heat map in order to visualize the area of pathology, however the pathology localization was not the main goal so only classification model performance was reported in terms of accuracy. Upon testing, the model achieved overall accuracy of 83.67% and 97.8% for the two different trained models.

b: CLASSIFICATION & LOCALIZATION

In [27], along with proposing a very large chest radiographs database (i.e., NIH Chest X-ray 14 dataset), experiments were performed in order to classify and locate multiple pathologies in CXR images. The dataset contained 112,120 frontal view CXRs which was divided into 70% training, 10% validation and 20% testing set. Multi-label classification was performed using different pretrained CNN architectures including AlexNet, GoogleNet, ResNet50 and VGG16. It was found that ResNet50 performed best of all. Various types of loss functions including Hinge Loss, Euclidean Loss and Cross Entropy Loss (CEL) were experimented to tackle the multi-label classification task. Since CEL performed best of all, it was modified a little bit to solve the class imbalance problem and the new loss function was named as weighted CEL (W-CEL). 79.93% AUC was achieved for the automatic detection of pneumothorax in the test set. Moreover, localization of the pathologies was also performed by generating bounding boxes which were obtained by applying an ad-hoc thresholding-based method on the heat maps obtained from the trained DCNN models. The performance of pathology localization was evaluated for only 983 images for which 1600 annotated bounding boxes were present and the accuracy for pneumothorax localization was 17.35% with Average False Positive Rate of 52.43%. Although the performance needs further improvement, however there is no doubt that collection of the CXR dataset on such a large scale was a big step.

A framework was proposed in [126] for the detection and localization of multiple thoracic pathologies by combining the concepts of multi-resolution and MIL (multi-instance learning). The LogSum-Exp pooling function
was parameterized with a learnable lower bound adaptation (LSE-LBA). This resulted in more robust diagnosis along with high resolution saliency maps for localization of pathology. The proposed model consisted of ResNet architecture which was used to down-sample the input image and then the image resolution was preserved using DenseNet. Training and evaluation were performed using NIH Chest X-ray 14 dataset with the official data split. The proposed model achieved AUC of 76.1% on whole test set, while for pneumothorax detection and localization, the model achieved AUC of 80.5% and DSC of 3.9% with lower-bound adaptation ($r_0 = 0$).

A novel approach for identification and localization of chest diseases from the radiographs was proposed by Li et al. [45]. The main goal of this study was to lessen the dependency of the location information for localization tasks, since the acquisition of pixel-level annotations for each pathology is an expensive job. The images were first processed with ResNetV2 after which patch slicing layer was used to resize the features extracted from the convolutional neural network using max pooling or bilinear interpolation. These resized images were then trained on a fully convolutional recognition network which generated two outputs, the label prediction and an image with location information. NIH Chest Xray14 dataset was used with all the 14 labels including pneumothorax. For experimentation the images with bounding boxes (880 images) were separated from the images with no-bounding boxes (111,240 images) and these two sets of images were named as annotated and unannotated. The dataset was split such that 70% of data was declared as a training set, 10% was used as validation and the remaining 20% was kept aside for testing purpose. When evaluated, the model achieved AUC value of 87% and IoU score of 63% for pneumothorax detection and localization respectively.

VI. PERFORMANCE METRICS

There are many different performance metrics which can be used to evaluate the classification models. In case of class-imbalance datasets, AUC is preferred over other performance measures [127]. Moreover, in literature mostly authors have reported the superiority of their proposed methods in terms of AUC. However few researchers have calculated other performance metrics including accuracy, sensitivity, precision and specificity. AUC is defined in terms of true positive (tpr) and false positive rate (fpr) [128] and the formulae for these two are given below:

\[
\text{tpr} = \frac{\text{Correctly classified positive samples}}{\text{Total No of Positive samples}} \tag{2}
\]

\[
\text{fpr} = \frac{\text{Incorrectly classified negative samples}}{\text{Total No. of negative samples}} \tag{3}
\]

The other performance measures can be defined with the help of a confusion matrix as shown in Table 3.

| Actual Class | Predicted Class |
|--------------|-----------------|
| Negative (Normal) | Negative (Normal) | Positive (Pathology) |
| Positive (Pathology) | Positive (Pathology) |

TN: True Negative, FP: False Positive, FN: False Negative, TP: True Positive

The expressions for calculating accuracy, recall/ sensitivity, specificity and precision [129] are given below.

\[
\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{FN} + \text{TP}} \tag{4}
\]

\[
\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}
\]

\[
\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}
\]

\[
\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{7}
\]

In order to prove the authenticity of a localization model, the measure of overlap between the original location information and predicted location is tried to be maximized and is calculated by means of Dice coefficient score (DSC) and Intersection over Union (IoU) as follows [130].

\[
\text{Dice coefficient} = \frac{2|X \cap Y|}{|X| + |Y|} \tag{8}
\]

\[
\text{IoU} = \frac{|X \cap Y|}{|X \cup Y|} \tag{9}
\]

VII. ANALYSIS AND DISCUSSION

This section compares the existing techniques for pneumothorax detection on the basis of usability in real life and reliability along with highlighting the research gaps which can be studied in future. Table 4 presents the summarized literature review in tabular form along with highlighting the limitations in each research.

Note that, for the papers in which binary classification was considered (i.e., presence of pathology or NO pathology, without particularly specifying any pathology) [44], [119], [124], we have reported the accuracy/AUC of the binary classifiers. While for all the papers in which more than one pathology was considered in either multi-label or multi-class classification problem, we have reported the performance for the detection of pneumothorax instead of model performance on the whole test set containing multiple pathologies.

Moreover, the quality assessment for each paper is provided in Table 1, in which the five questions mentioned in Section II-D are answered. In Table 1, it can be observed that most of the authors have not clearly mentioned the limitations in their research articles, however, the research objective, the selected approaches, and the evaluation metrics are clearly explained in all the included papers. The last column shows...
the marks obtained by each paper on the basis of quality assessment.

A. COMPARATIVE ANALYSIS

- From Table 2, it is quite evident that almost all the chest X-ray datasets have very few samples of pneumothorax as compared to other class’s samples, thus making the datasets highly class imbalance. Several approaches have been proposed in literature for solving class imbalance problem [27], [77], [86], [88], [94], [95], however, in the literature mostly researchers have used the classifier-level-approaches to tackle this issue [22], [27], [75], [113], [118], [120] and few have used data-level approaches [40], [125]. Thus, it can be expected that classifier-level approaches perform better, yet none of the approaches can be declared as the best one to be used in CAD systems for medical diagnosis.

- The gender distribution in Table 2 shows that men are more likely to be affected by pneumothorax as compared to women.

- From literature, it is clear that mostly researchers have preferred DL techniques instead of ML approaches. This might be due to the reason that ML requires handcrafted features which must be acquired very carefully and, considerably, as unnecessary features affect the accuracy of a classifier [58], [102]. On the other hand, DL algorithms do the job of feature extraction by itself. Despite the several limitations of DL technique such as requirement of huge amount of training data, computational complexity, extensive resource consumption and optimizing the hyperparameters, mostly researchers have adapted DL techniques because of their outstanding performance [22], [43], [61], [73], [110], [116].

- The performance of deep learning models can further be improved by ensemble modelling in which different models are trained on the training set and final results are combined via voting or averaging methods. The same is evident from [73], [74]. However, ensemble models face the problem of time and computational complexity.

- Although NIH Chest X-ray 14 dataset provided an official split of data, however many researchers have used their own split instead of official data split, this creates lack of uniformity in the training and testing data of different researchers. Hence for the datasets which provide official data split, shall be used as it is, instead of random split of dataset of author’s own choice in order to make fair comparison of the achieved results with the benchmark results. Table 4 describes the type of data split used in each paper as either random or official data split.

- In some of the research, huge amount of data was used for training purpose while very small test sets were used for evaluation of models [75], [76], [119] [125]. However, for more reliability and authenticity, the evaluation shall be performed on maximum possible samples.

- Assigning samples to training, validation and testing sets must be done with great care while ensuring that there is no-patient-wise-overlap in these sets, i.e., samples from the same patient must not be present in more than one set. However, it has been observed that in literature, although few authors have explicitly declared the no-patient-wise-overlap in paper [22], [120], [122], but most of the authors missed this important point and did not follow the patient-wise-split. (i.e., dataset was split randomly without ensuring that the chest x-rays from the same patient should be in the same split, either training or testing split) [42], [43], [74], [117]. Thus, extra attention must be paid to ensure the no-patient-wise-overlap, which otherwise may lead to biasness of results and poor performance of models in real time scenarios.

B. RESEARCH GAPS

There are few research gaps which need to be covered in future:

- The existing literature for pneumothorax detection proved the effectiveness of DL methods as they can make intelligent decisions without human intervention, thus DL approaches are preferred over ML. Since the performance of a deep learning model is dependent on the size of the training data [131], while all the CXRs dataset available till date have very few samples of pneumothorax compared to whole dataset size, so contribution shall be made by the researchers in collecting larger dataset with greater number of pneumothorax samples. The data collection by traditional method of obtaining samples of thousands of patients might face the problem of privacy and copyright issue, hence assistance can be taken from latest technology of synthetic image generation such as GAN, as research have proved the robustness of GAN generated samples for classification problems in different fields including medical domain [125], [132], [133], [134].

- It is a known fact that most of the medical image datasets are imbalanced in nature, i.e., contain fewer samples of pathology as compared to normal samples [135]. This might generate biased results toward majority class samples. Most of the researchers have tried to solve this issue using algorithm level approaches, e.g., adding weights in the loss function in order to give more priority to fewer occurring samples as in [27], [41], and [75]. However, a comparison of existing class imbalance approaches like the one made in [20] has never been made particularly for chest radiographs dataset. As the class imbalance approaches are domain dependent [136] so conducting such a comparison and later proposing a new technique might give surprising results and can be an interesting research topic.

- Most of the researchers used the NIH Chest Xray14 dataset while very few have used CheXpert and
### Table 4. Comparison of existing approaches for pneumothorax detection.

| Author          | Year | Technique | Methodology Description                                                                 | Class Imbalance Solution          | Dataset          | Result (%) | Domain | Data split | External testing | Limitation                                                                 |
|-----------------|------|-----------|-----------------------------------------------------------------------------------------|-----------------------------------|------------------|-------------|---------|-------------|------------------|---------------------------------------------------------------------------|
| Geva [102]      | 2015 | ML        | Feature extraction from Local Binary Pattern and KNN as classifier.                     | Dataset was already class-balanced | Local            | Sen=81     | C       | R          | ×                | Small and private dataset, sensitivity needs improvement                |
| Chan [38]       | 2018 | ML        | Classification ULBP with SVM, (Localization) LBP with Sobel Edge Detector              | Dataset was already class-balanced | Local            | Classification: Accuracy=82.2 Localization: Accuracy=85.8 | C & L   | R          | ×                | Private dataset, Other performance metrics such as AUC, precision and recall are not evaluated. Evaluation on a very small dataset. |
| Jun [74]        | 2018 | DL        | Ensemble of three ResNet 50 architecture with different input resolution               | Thresholding, i.e., cut-off applied to the probability value | NIH              | AUC=91.1   | C       | R          | ×                | Non-standard data split, Patient-wise split was not considered.          |
| Taylor [40]     | 2018 | DL        | Multiple CNNs including VGG16, ResNet, Inception and Xception.                        | Random undersampling              | Local & NIH      | Local: AUC=94, Sen=55 NIH AUC=75, Sen=49 | C       | R          | ✓                | Performance need improvement as sensitivity is only 55%, Low performance on external test set. |
| Sze [75]        | 2019 | DL        | Transfer learning using CheXNet                                                        | Weighted binary cross entropy function | CheXpert         | AUC=70.8   | C       | R          | ×                | Extremely small test set, Performance needs improvement                  |
| Park [108]      | 2019 | DL        | Transfer Learning using YOLO Darknet19                                                | No proper explanation for class imbalance | Local            | Internal: AUC=98.4, Sen=89 External: AUC=90.5, Sen=63 | C       | R          | ✓                | Private Dataset. No proper explanation for solving class imbalance. Very low performance on external test set. |
| Luo [41]        | 2019 | DL        | Fully convolutional multiscale seSE DenseNet                                          | Dataset was already class-balanced | Local            | DSC=92.0   | L       | R          | ×                | Details for train-test split are missing, private dataset was used       |
| Ouyang [112]    | 2019 | DL        | Multiple segmentation models                                                           | No proper explanation for class imbalance | Local            | IoU=66.69  | L       | R          | ×                | IoU score is not promising                                                |
| Mostayed [92]   | 2019 | DL        | Modified traditional U-Net                                                            | No proper explanation for class imbalance | SIIM             | DSC=76.0   | L       | R          | ×                | Lack of details regarding data split.                                    |
| Gooben [114]    | 2019 | DL        | (Classification) Pre-trained ResNet50 (Localization) MIL and U-Net                     | No proper explanation for class imbalance | Local            | Classification: AUC=96 Localization: AUC=93 | C & L   | R          | ×                | Extremely small and private dataset. Performance measures other than AUC were not evaluated. |
| Viniaovskyi [76] | 2020 | DL        | CAM generation, IRNet, U-Net model                                                     | Weighted binary cross entropy loss function | SIIM             | DSC=76.69  | L       | 0ff        | ×                | Evaluation on a very small test, no comparison with existing literature. Performance needs improvement |
| Abedalla [91]   | 2020 | DL        | Two stage U-Net based framework                                                        | No proper explanation for class imbalance | SIIM             | Stage1: DSC=85.02 Stage2: DSC=83.56 | L       | 0ff        | ×                | No proper explanation for solving imbalance issue (in terms of images number and pixel level imbalance. |
| Groza [73]      | 2020 | DL        | Ensemble of LinkNet(s)                                                                | Non-empty sampling to ensure positive samples | SIIM             | Stage1: DSC=88.21 Stage2: DSC=86.14 | L       | 0ff        | ×                | Computationally complex                                                   |
| Jakhar [60]     | 2021 | DL        | U-Net with ResNet backbone                                                            | No proper explanation for class imbalance | SIIM             | DSC=83.4   | L       | R          | ×                | Random data split was used.                                              |
| Tolkachev [110] | 2021 | DL        | U-Net with various CNNs                                                               | No proper explanation for class imbalance | SIIM             | Stage1: DSC=85.47 | L       | 0ff        | ×                | Comparison with related work is needed.                                   |
TABLE 4. (continued.) Comparison of existing approaches for pneumothorax detection.

| Wang [107] | 2021 | DL | Network In Network and Histogram Equalization | Data augmentation | Local & NIH | Classification: AUC=98.4 | Localization: NIH: AUC=99.06 | C | R | × | Only AUC and accuracy were used as a performance measure, other performance metrics such as recall, and precision needs to be evaluated. Non-standard random split of dataset. |
|--------------|------|----|-----------------------------------------------|-------------------|------------|----------------------|------------------------------|----|----|   |                                |
| Rajpurkar [22] | 2017 | DL | CheXNet using DenseNet121 | Binary cross entropy | NIH | AUC= 88.87 | C | R | × | Non-standard random data split |
| Yao Li [126] | 2018 | DL | Multi resolution and Multi instance learning | No proper explanation for class imbalance | NIH | Classification: AUC=80.5 | Localization: DSC=3.9 | C | L | 0ff | Very low dice coefficient score. Comparison with related work is missing |
| Li [45] | 2018 | DL | ResNet50, patch slicing, FCN | Added weights to images with bounding boxes | NIH | Classification: AUC=87 | Localization: IoU=63 | C | L | 0ff | Non-standard Random data split was used. |
| Kumar [116] | 2018 | DL | DenseNet161, Binary Relevance, Pair Wise Approach | Weighted binary cross entropy | NIH | AUC=85.94 | C | R | × | Random data split, performance needs improvement |
| Yao [42] | 2018 | DL | Encoder and Decoder based RNN, Interdependency among labels was leveraged using LSTMs | Weighted cross entropy | NIH | AUC=84.1 | C | R | × | Random split of dataset, patient-wise overlap should be considered in data split |
| Guan [117] | 2018 | DL | AGCNN using ResNet50 and DenseNet121 | No proper explanation for class imbalance | NIH | AUC=92.1 | C | R | × | Non-standard random data split, patient wise overlap was not considered. Class imbalances needs to be explained |
| Rubin [39] | 2018 | DL | DenseNet121 trained on different view positions of CXRs | No proper explanation for class imbalance | MIMIC | AUC= 62.5 | C | R | × | Very low performance in terms of AUC |
| Salchinea [125] | 2018 | DL | Proposed network similar to AlexNet, Oversampling by DCGAN | Oversampling by DCGAN | Local | Accuracy=88.8 | C | R | × | Other performance metrics like AUC, recall etc., were missing. Test set was completely class-balanced which does not support the claimed research. |
| Baltrusch at [118] | 2019 | DL | Transfer Learning with and without fine tuning was explored using ResNet50 | Class-averaged binary cross entropy | NIH | Random split: AUC=87.0 | Official split: AUC=84.0 | C | R | 0ff | Performance needs improvement |
| Tung [119] | 2019 | DL | Generative Adversarial Network to distinguish between normal and abnormal CXRs | Balanced dataset | NIH | AUC=84.1 | C | R | × | Particular pathology shall be the main focus. Data selection criteria was not mentioned clearly. Evaluation on a very small test set. |
| Allaouzi [43] | 2019 | DL | Transfer Learning using DenseNet121 | Problem transformation techniques including LP, BR and CC | NIH & CheXpert | AUC=92.9 | CheXpert: AUC=86 | C | R | × | Random split of the dataset was used. Patient-wise-split was not explicitly mentioned |
**TABLE 4. (continued.)** Comparison of existing approaches for pneumothorax detection.

| Author            | Year | Methodology                        | Problem                                                                 | Loss Function          | Evaluation | Split      | Performance Details                                                                 |
|-------------------|------|------------------------------------|------------------------------------------------------------------------|------------------------|------------|-----------|------------------------------------------------------------------------------------|
| Guendel [120]     | 2019 | DL       | DenseNet121 was trained using spatial and location information of pathology. | Weighted cross entropy loss function                  | Local & NIH | C, R & Off | Performance needs improvement especially for official data split                   |
| Mojkowska [122]   | 2019 | DL       | Pre-trained Xception model                                                  | No proper explanation for class imbalance              | Local & NIH | C, R      | Evaluation on a very small test. Non-standard random data split. External validation required. |
| Mondal [123]      | 2019 | DL       | Multi-label classification problem using Xception                          | No proper explanation for class imbalance              | RS-NIH     | C, R      | Very low AUC. Patient-wise-split was not explicitly mentioned.                     |
| Bharati [124]     | 2020 | DL       | VGG16, data augmentation, STN                                                | No proper explanation for class imbalance              | RS-NIH & NIH| C, R      | Lack experimental detail regarding data split. Unfair comparison with related work as it is binary classification problem. Other performance measures like AUC etc. are missing. |
| Prakash [44]      | 2021 | DL       | Own CNN architecture and transfer learning using VGG16                       | No proper explanation for class imbalance              | RS-NIH     | C, R      | Performance in terms of other metrics is required. Unfair comparison with related work. Patient-wise-split was not explicitly mentioned. |

MIMIC-CXR-JPG dataset, so these datasets need to be explored further for pneumothorax detection purpose.

- It is very clear that from the last decade the trend has been shifted form Machine Learning (ML) to Deep learning techniques (DL). Mostly, the proposed frameworks used pre-trained Convolutional Neural Networks (e.g., VGG16, ResNet, and DenseNet) for classification, and for localization purpose (e.g., U-Net). So, there is great scope to engage one-self in this field. However, it can be observed that for classification purpose mostly researchers have used DenseNet or ResNet, so it can be an interesting experiment to try some other and new CNN architectures like EfficientNet [137] in automating the field of automatic detection of thoracic pathologies especially pneumothorax. Similarly for the purpose of localization, there are few other and relatively new segmentation models like Feature Pyramid Network (FPN) [138], LinkNet [111] and Pyramid Scene Parsing Network (PSPNet) [139] which need to be explored.

- It has been observed that most of the existing literature had considered only the frontal chest X-rays, however almost 15% of the information is added by the lateral view chest radiograph [140]. Thus, the performance of pneumothorax detection models can be improved by using both frontal and lateral view chest radiographs, as is evident from [59] in which different view positions of chest x-rays were used.

- Mostly, images alone are used for training a deep learning model while ignoring the non-image-features. In future, non-image features like patient’s age, gender, CXR view-position etc. shall be considered along with the image for model training, as its effectiveness is obvious from [118].

- It can be easily observed from the literature that for pneumothorax detection (either disease specific or multiple pathologies detection system), very less work has been done for combined diagnosis and localization of affected area. The reason might be that, before the availability of SIIM ACR pneumothorax dataset (the sole purpose of which was to contribute to the field of segmentation of the area of lung affected by pneumothorax) mostly authors have used was NIH Chest Xray14 dataset which provide bounding box information for only 980 images out of which only 98 samples belong to pneumothorax class. Moreover, the other dataset like CheXpert and MIMIC-CXR-JPG mainly focused on providing image-level labels instead of pixel-level annotation. It is understandable that obtaining pixel level annotation is not an easy job, however localization of a pathology especially pneumothorax is important as the treatment depends entirely on the size of pneumothorax. This problem can be solved by proposing such localization techniques which lessen the dependency on the requirement of pixel level annotation for training...
and later predicting the affected area of the lung. Efforts have been made in order to solve this problem (as in [45], [112]), however there is still great scope of improvement in this domain.

- Generalization is a very important matter to be considered while proposing a framework for automatic detection of pathology. Generalizability means that a model trained and validated on a dataset obtained from same experimental setting shall perform well on the data obtained from a different experimental setting. In other words, a model trained on dataset A shall be able to give vigorous performance for dataset B as well. However, it can be observed that in most of the research only internal test data (i.e., test data obtained from same experimental setting as the training data) was used for evaluating the proposed model, while very few have reported the performance of the proposed framework on external test data [40], [108]. Hence validation of the results on external dataset shall be made in order to safely use the proposed framework in real life.

VIII. LIMITATIONS OF THE STUDY

Just like any other publication, this paper has few limitations which need to be catered in future. Firstly, it covers the literature utilizing chest radiographs only, while the research work done using CT scan or MRIs need to be reviewed yet. Secondly this review paper does not consider the research work done for pneumothorax detection using chest radiograph’s textural report. Thirdly only the papers published during 2010 to 2020 in English language are considered, so the related work in other languages need to be focused-on in a separate review paper. Additionally, the results reported in the paper were relied on and no experiment was performed in order to reproduce and analyze the results of the existing techniques.

IX. CONCLUSION

Chest radiography is one of the cheapest and easily available diagnostic tools used for the diagnosis of various chest pathologies including pneumothorax. The successful employment of deep learning techniques in various fields of medicine has encouraged the researchers to contribute towards automating the diagnostic process from chest radiographs. Hence many different frameworks have been proposed for automatic diagnosis of pneumothorax utilizing various machine learning and especially deep learning techniques. However, a summarized overview of the existing literature for pneumothorax detection is not available so far. This paper presented a systematic literature review of the research carried out in the last decade using chest radiographs for the automatic detection of pneumothorax. It provided a summarized overview of existing work along with highlighting usability and cons of present models. Additionally, the computational complexities and research gaps were also discussed. This would guide the researchers to contribute towards filling those gaps and selecting optimal techniques for further research.

From literature, it is evident that pneumothorax is more common in men as compared to women. Additionally, the external testing of proposed models must be done in order to utilize them for real-life diagnosis. Finally, the current best performing model for the NIH Chest Xray14 dataset was provided by Baltruschat [118], as the authors have used the official data-split for training and testing purposes while ensuring the no-patient-wise-overlap and achieved AUC of 84.0%. For localization, the work done by Groza [73] can be declared as best one with highest Dice coefficient score (DSC) achieved till date for localization of pneumothorax, i.e., 88.21% DSC, however the computational complexity can’t be ignored. Overall best results for pneumothorax detection were obtained by Rajpurkar [22], if utilization of official split is not a constraint, as the author catered the patient-wise split and also generated heat map to visualize the location of pathology with AUC of 88.87% for pneumothorax detection.

CONFLICT OF INTEREST

The authors have no conflict of interest to disclose.

REFERENCES

[1] P. Zarogoulidis, I. Kioumis, G. Pitsioul, K. Porrpods, S. Lampaki, A. Papaiwannou, N. Katsikogiannis, B. Zaric, P. Bransilav, N. Secen, G. Dryllis, N. Machariiotis, A. Rapti, and K. Zarogoulidis, “Pneumothorax: From definition to diagnosis and treatment,” J. Thoracic Disease, vol. 6, no. 4, pp. S372–S376, Oct. 2014, doi: 10.3978/j.issn.2072-1439.2014.09.24.

[2] R. M. Shorr, M. Crittenden, M. Indeck, S. L. Hartunian, and A. Rodriguez, “Blunt thoracic trauma analysis of 515 patients,” Ann. Surg., vol. 206, no. 2, pp. 200–205, Aug. 1987, doi: 10.1097/00000658-198708000-00013.

[3] L. Husain, D. Wayman, K. Carmody, L. Hagopian, and W. Baker, “Sonographic diagnosis of pneumothorax.” J. Emergencies, Trauma, Shock, vol. 5, no. 1, p. 76, 2012, doi: 10.4103/0974-2700.93116.

[4] C. Qin, D. Yao, Y. Shi, and Z. Song, “Computer-aided detection in chest radiography based on artificial intelligence: A survey,” Biomed. Eng. OnLine, vol. 17, no. 1, pp. 1–23, Aug. 2018, doi: 10.1186/s12938-018-0544-y.

[5] A. Oates, K. Halliday, A. C. Offiah, C. Landes, N. Stoodley, A. Jeannes, 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, doi: 10.1016/j.crad.2019.04.016.

[6] G. S. Lodwick, T. E. Keats, and J. P. Dorst, “The coding of roentgen images for computer analysis as applied to lung cancer,” Radiology, vol. 81, no. 2, pp. 185–200, Aug. 1963, doi: 10.1148/81.2.185.

[7] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, “Cardiologist-level arrhythmia detection with convolutional neural networks,” Jul. 2017, arXiv:1707.01886.

[8] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, no. 7639, pp. 115–118, Jan. 2017, doi: 10.1038/nature21056.

[9] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” J. Amer. Med. Assoc., vol. 316, no. 22, pp. 2402–2410, Dec. 2016, doi: 10.1001/jama.2016.17216.

[10] W. Khan, N. Zaki, and L. Ali, “Intelligent pneumonia identification from chest X-rays: A systematic literature review,” IEEE Access, vol. 9, pp. 51747–51771, 2021, doi: 10.1109/access.2021.3069957.
A. S. Lundervold and A. Lundervold, "An overview of deep datasets search in chest radiography: A survey," Quant. Imag. Med. Surg., vol. 3, no. 2, pp. 89–99, Apr. 2013, doi: 10.2976/2223-4292.2013.04.03.

A. Mucherino, P. Papajorgji, and P. M. Pardalos, "A survey of data mining techniques applied to agriculture," Oper. Res., vol. 9, no. 2, pp. 121–140, Jun. 2009, doi: 10.1287/ijor.1101.0045.6.

K. Muhammad, T. Hassan, and S. L. Baik, "Efficient CNN based summarization of surveillance videos for resource-constrained devices," Pattern Recognit. Lett., vol. 130, pp. 370–375, Feb. 2020, doi: 10.1016/j.patrec.2018.08.059.

J. W. Zuo, F. Zhou, Z. Li, and L. Wang, "Multi-resolution CNN networks for chest diseases detection," J. Digit. Imag., vol. 33, no. 2, pp. 490–496, Apr. 2020, doi: 10.1007/s12531-019-09316-2.

L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Comparison," Jan. 2019, arXiv:1901.07042.

P. R. Smith, "Bilinear interpolation of digital images," IEEE Trans. Med. Imag., vol. 106, pp. 249–259, Oct. 2018, doi: 10.1109/BIBM.2018.8983009.

A. T. Iqbal, C. J. Coelho, and A. L. S. S. Andrade, "Computer-aided diagnosis," IEEE Trans. Med. Informat., vol. 77, no. 8, pp. 555–564, Aug. 2008, doi: 10.1109/TMI.2008.2000120.

A. Mucherino, P. Papajorgji, and P. M. Pardalos, "A survey of data mining techniques applied to agriculture," Oper. Res., vol. 9, no. 2, pp. 121–140, Jun. 2009, doi: 10.1287/ijor.1101.0045.6.

K. Muhammad, T. Hassan, and S. L. Baik, "Efficient CNN based summarization of surveillance videos for resource-constrained devices," Pattern Recognit. Lett., vol. 130, pp. 370–375, Feb. 2020, doi: 10.1016/j.patrec.2018.08.059.

J. W. Zuo, F. Zhou, Z. Li, and L. Wang, "Multi-resolution CNN networks for chest diseases detection," J. Digit. Imag., vol. 33, no. 2, pp. 490–496, Apr. 2020, doi: 10.1007/s12531-019-09316-2.

L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," Feb. 2018, arXiv:1710.10501.

I. Allaouzi and M. B. Ahmed, "A novel approach for multi-label chest X-ray classification of common thorax diseases," IEEE Access, vol. 7, pp. 64279–64288, 2019, doi: 10.1109/ACCESS.2019.2916849.

P. Chowdhury and A. Hazra, "Chest disease radiography in a cloud world: Using convolutional neural networks and transfer learning," Evolving Syst., vol. 12, no. 2, pp. 567–579, Jun. 2021, doi: 10.1109/IS2015.2018.8096865.

D. Bozij, M. Grigic, and K. Delac, "Comparative analysis of interpolation methods for bilateral asymmetry," in Proc. ELMAR, Sep. 2010, pp. 1–7.

R. W. P. S. Silveira, C. W. V. A. Araujo, V. C. Silva, and G. Zanetti, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3462–3471, doi: 10.1109/CVPR.2017.369.

Academic Torrents. (2017). NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories. Accessed: Nov. 11, 2020. [Online]. Available: https://academictorrents.com/details/55748f1aad8d824c83b757676938b3325a

Kaggle. (2017). Random Sample of NIH Chest X-ray Dataset. Accessed: Nov. 11, 2020. [Online]. Available: https://www.kaggle.com/nih-chest-xrays/sample/activity

J. Irvin, P. Rajurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, J. Seekins, D. A. Mong, S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungen, and A. Y. Ng, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert confidence annotations," Jun. 2019, arXiv:1901.07042.

Chexpert: A Large Dataset of Chest X-Rays and Competition for Automated Chest X-Ray Interpretation. Accessed: Nov. 11, 2020. [Online]. Available: https://www.stanfordmlgroup.github.io/competitions/chexpert/

H. H. Pham, T. T. Le, D. Q. Tran, D. T. Ngo, and H. Q. Nguyen, “Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels,” Neurocomputing, vol. 437, pp. 186–194, Jun. 2021, doi: 10.1016/j.neucom.2020.02.030.

A. E. W. Johnson, T. J. Pollard, N. R. Greenbaum, M. P. Lungen, C.-Y. Deng, Y. Peng, Z. Lu, R. G. Mark, S. J. Berkowitz, and S. Horng, “MIMIC-CXR-JPG, a publicly available database of labeled chest radiographs,” Nov. 2019, arXiv:1901.10742.

Physiobank. (2019). MIMIC-CXR-JPG—Chest Radiographs With Structured Labels V2.0.0. Accessed: Nov. 11, 2020. [Online]. Available: https://www.physionet.org/content/mimic-cxr-jpg/2.0.0/

A. J. Larrazabal, N. Nieto, V. Peterson, D. H. Milone, and E. Ferrante, “Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis,” Proc. Nat. Acad. Sci. USA, vol. 117, no. 23, pp. 12592–12594, Jun. 2020, doi: 10.1073/pnas.1901021117.

A. G. Taylor, C. Mielke, and J. Mongan, “Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study,” Plod Med., vol. 15, no. 11, Nov. 2018, Art. no. e1002697, doi: 10.1371/journal.pmed.1002697.

G. Luo, Z. Liu, Q. Wang, Q. Liu, Y. Zhou, W. Xu, J. Huang, J. Fu, and J.-Z. Cheng, “Fully convolutional multi-scale ScSE-DenseNet for automatic pneumothorax segmentation in chest radiographs,” in Proc. IEEE Int. Conf. Biomed. Image (IBIBM), San Diego, CA, USA, Nov. 2019, pp. 1551–1555, doi: 10.1109/IBIBM.2019.8983009.

L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, “Learning to diagnose from scratch by exploiting dependencies among labels,” Feb. 2018, arXiv:1710.10501.

I. Allaouzi and M. B. Ahmed, “A novel approach for multi-label chest X-ray classification of common thorax diseases,” IEEE Access, vol. 7, pp. 64279–64288, 2019, doi: 10.1109/ACCESS.2019.2916849.

P. Chowdhury and A. Hazra, “Chest disease radiography in a cloud world: Using convolutional neural networks and transfer learning,” Evolving Syst., vol. 12, no. 2, pp. 567–579, Jun. 2021, doi: 10.1109/IS2015.2018.8096865.

D. Bozij, M. Grigic, and K. Delac, "Comparative analysis of interpolation methods for bilateral asymmetry," in Proc. ELMAR, Sep. 2010, pp. 1–7.

R. P. Smith, "Bilinear interpolation of digital images," Ultramicroscopy, vol. 6, no. 2, pp. 201–204, Jan. 1981, doi: 10.1016/0304-3991(81)90061-9.

X. Xu, Y. Wang, G. Yang, and Y. Hu, "Image enhancement method based on fractional wavelet transform," in Proc. IEEE Int. Conf. Signal Image Process. (ICISP), Beijing, China, Aug. 2016, pp. 194–197, 10.1109/SIPROCESS.2016.7888251.
[9] S. Khorbagrade, A. Tiwari, C. Y. Patil, and V. Narke, “Automatic detection of major lung diseases using chest radiographs and classification by feed-forward artificial neural network,” in Proc. IEEE 1st Int. Conf. Power Electron., Intell. Control Energy Syst. (ICPEICES), Delft, India, Jul. 2016, pp. 1–5, doi: 10.1109/ICPEICES.2016.7853685.

[50] Y. H.-M. Benjamin Kwan and H. K. W. Kwan, “Improved lung nodule visualization on chest radiographs using digital filtering and contrast enhancement,” World Acad. Sci., Eng. Technol., Int. J. Med., Health, Biomed., Bioeng. Pharmaceutical Eng., vol. 5, no. 112, pp. 700–703, Dec. 2011, doi: 10.5281/zenodo.1063290.

[51] I. C. Mehta, Z. J. Khan, and R. R. Khotpal, “Analysis and review of chest radiograph enhancement techniques,” Inf. Technol. J., vol. 5, no. 3, pp. 577–582, Apr. 2006, doi: 10.3923/itj.2006.577.582.

[52] A. Sharma, D. Raju, and S. Ranjan, “Detection of pneumonia clouds in chest X-ray using image processing approach,” in Proc. Nirma Univ. Int. Conf. Eng. (NUICON), Ahmedabad, India, Nov. 2017, pp. 1–4, doi: 10.1109/NUICON.2017.8352667.

[53] R. H. Sherrier and G. A. Johnson, “Regionally adaptive histogram equalization of the chest,” IEEE Trans. Med. Imag., vol. MI-6, no. 1, pp. 1–7, Mar. 1987, doi: 10.1109/TMI.1987.4307791.

[54] S. S. K and N. C. Naveen, “Algorithm for pre-processing chest-X-ray images using multi-level enhancement operation,” in Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET), Chennai, India, Mar. 2016, pp. 2182–2186, doi: 10.1109/WiSPNET.2016.7566629.

[55] ImageFiltering. Accessed May 19, 2021. [Online]. Available: https://ai.stanford.edu/~syyeung/cvweb/tutorial1.html.

[56] H. He and E. A. Garcia, “Learning from imbalanced data,” J. Artif. Intell. Res., vol. 16, no. 1, pp. 371–396, Jan. 2002, doi: 10.1613/jair.654.

[57] Y. M. Benjamin Kwan and H. K. W. Kwan, “Improved lung nodule visualization on chest radiographs,” in Proc. IEEE 17th Int. Symp. Biomed. Imag. Workshops (ISBI Workshops), Iowa City, IA, USA, Apr. 2020, pp. 1–4, doi: 10.1109/ISBIWorkshops2020.2020.9153444.

[58] J. Rubin, D. Sanghavi, C. Zhao, K. Lee, A. Qadir, and M. Xu-Wilson, “Detection of pneumonia clouds in chest X-ray using image processing approach,” in Proc. Nirma Univ. Int. Conf. Eng. (NUICON), Ahmedabad, India, Nov. 2017, pp. 1–4, doi: 10.1109/TKDE.2008.239.

[59] V. Groza and A. Kuzin, “Pneumothorax segmentation with effective conditional post-processing in chest X-ray,” in IEEE Proc. 17th Int. Symp. Biomed. Imag. Workshops (ISBI Workshops), Iowa City, IA, USA, Apr. 2020, p. 1–4, doi: 10.1109/ISBIWorkshops2020.2020.9153444.

[60] T. I. Jun, D. Kim, and D. Kim, “Automated diagnosis of pneumothorax using an ensemble of convolutional neural networks with multi-sized chest radiography images,” Apr. 2018, arXiv:1804.06821.

[61] S. S. K and N. C. Naveen, “Regionally adaptive histogram equalization of the chest,” IEEE Trans. Med. Imag., vol. MI-6, no. 1, pp. 1–7, Mar. 1987, doi: 10.1109/TMI.1987.4307791.

[62] Y.-H. Chan, Y.-Z. Zeng, H.-C. Wu, M.-C. Wu, and H.-M. Sun, “Effective pneumothorax detection for chest X-ray images using local binary pattern and support vector machine,” J. Healthcare Eng., vol. 2018, pp. 1–11, Aug. 2018, doi: 10.1155/2018/2908517.

[63] A. Awain. (Jul. 16, 2020). Noise filtering in digital image processing. Medium. Accessed: May 19, 2021. [Online]. Available: https://medium.com/image-visions/noise-filtering-in-digital-image-processing-d12b5266847c.

[64] T. K. Jemis and S. Shukla, “Class imbalance learning using UnderBagging based kernalized extreme learning machine,” Neurocomputing, vol. 329, pp. 172–187, Feb. 2019, doi: 10.1016/j.neucom.2018.10.056.

[65] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” J. Artif. Intell. Res., vol. 16, no. 1, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[66] J. Wang, M. Xu, H. Wang, and J. Zhang, “Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding,” in Proc. 9th Int. Conf. Signal Process., vol. 3, 2006, pp. 1–4, doi: 10.1109/ICOSP.2006.345752.

[67] T. Io and N. Japkowicz, “Class imbalances versus small disjuncts,” ACM SIGKDD Explor. Newsletter., vol. 6, no. 1, pp. 40–49, Jun. 2004, doi: 10.1145/1007730.1007737.

[68] J. Guo and H. Guo, “Harmonizing learning from imbalanced datasets by leveraging and data generation: The DataBoost-I(M) approach,” ACM SIGKDD Explor. Newsletter., vol. 6, no. 1, pp. 30–39, Jun. 2004, doi: 10.1145/1007730.1007736.

[69] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” Jun. 2014, arXiv:1406.2661.

[70] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” Jan. 2016, arXiv:1511.06434.

[71] 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), Venice, Italy, Oct. 2017, pp. 2242–2251, doi: 10.1109/ICCV.2017.244.

[72] 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. 4396–4405, doi: 10.1109/CVPR.2019.00453.

[73] C. Drummond and R. C. Holte, “C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling,” in Proc. Workshop Learn. From Imbalanced Datasets II, Washington, DC, USA, vol. 11, Aug. 2003, pp. 1–8. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.68.6858&rep=rep1&type=pdf.

[74] M. Kubit and S. Matwin, “Addressing the curse of imbalanced training sets: One-sided selection,” presented at the ICML, Jan. 1997. Accessed: May 7, 2021. [Online]. Available: https://openreview.net/forum?id=BJWDhmzZaW.

[75] B. Raskutti and A. Kowalczyk, “Extreme re-balancing for SVMs: A case study,” SIGKDD Explor. Newsletter., vol. 6, no. 1, pp. 60–69, Jun. 2004, doi: 10.1145/1007730.1007739.
[126] L. Yao, J. Prosky, E. Poblenz, B. Covington, and K. Lyman, “Weakly supervised medical diagnosis and localization from multiple resolutions,” Mar. 2018, arXiv:1803.07703.

[127] U. R. Salunkhe and S. N. Mali, “Classifier ensemble design for imbalanced data classification: A hybrid approach,” Proc. Comput. Sci., vol. 85, pp. 725–732, Jun. 2016, doi: 10.1016/j.procs.2016.05.259.

[128] T. Fawcett, “An introduction to ROC analysis,” Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

[129] M. da Silva Santos, M. Ladeira, G. C. G. Van Erven, and G. L. da Silva, “Machine learning models to identify the risk of modern slavery in Brazilian cities,” in Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2019, pp. 740–746, doi: 10.1109/ICMLA.2019.00132.

[130] N. Tistison and J. Gee, “Introducing Dice, Jaccard, and other label overlap measures to ITK.” Insight J., vol. 2, p. 707, Dec. 2009. [Online]. Available: http://hdl.handle.net/10380/3141

[131] J. G. A. Barbedo, “Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification.” Comput. Electron. Agrovol., vol. 153, pp. 46–53, Oct. 2018, doi: 10.1016/j.compag.2018.08.013.

[132] Q. Wu, Y. Chen, and J. Meng, “DCGAN-based data augmentation for tomato leaf disease identification,” IEEE Access, vol. 8, pp. 98716–98728, 2020, doi: 10.1109/ACCESS.2020.2997001.

[133] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, M. Bar-Or, Z. Gross, G. Friedman, S. Rotshtein, Z. Kupiec, Y. Yilma, R. Assaf, L. Gepner, S. Avidan, Z. Assia, S. Aizenstein, S. Amiel, I. Ben-Haroush, and J. G. A. Barbedo, “Impact dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification.” Comput. Electron. Agrovol., vol. 153, pp. 46–53, Oct. 2018, doi: 10.1016/j.compag.2018.08.013.

[134] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn., May 2019, pp. 6105–6114. Accessed: Dec. 11, 2020. [Online]. Available: http://proceedings.mlr.press/v97/tan19a.html

[135] S. Seferbekov, V. Iglovikov, A. Buslaev, and A. Shvets, “Feature pyramid network for multi-class land segmentation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Salt Lake City, UT, USA, Jun. 2018, pp. 272–273, doi: 10.1109/CVPRW.2018.00051.

[136] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Medical Image Computing and Computer-Assisted Intervention. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

TAHIRA IQBAL received the B.S. degree in computer engineering from the Department of Computer and Software Engineering, College of Electrical & Mechanical Engineering, National University of Sciences and Technology (CEME NUST), Pakistan, in 2018, and the M.S. degree in computer engineering from the National University of Sciences and Technology. Her research interests include deep learning, machine learning, and medical image analysis.

ARSLAN SHAUKAT received the B.S. and M.S. degrees in computer engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2003 and 2005, respectively, and the Ph.D. degree in computer science from The University of Manchester, U.K., in 2010. He is currently an Associate Professor with the Department of Computer and Software Engineering, College of Electrical & Mechanical Engineering, National University of Sciences and Technology (CEME NUST). He has published various research papers in refereed journals and conference proceedings. His research interests include machine learning, pattern recognition, digital image, and speech processing. He has been a member of technical program committees of numerous international conferences and a reviewer of international journals. He was a recipient of academic awards, including the Best Teacher Award, in 2018, and the Best Research Paper Award, in 2019.

MUHAMMAD USMAN AKRAM received the Ph.D. degree in computer engineering (specializing in medical imaging analysis) from the National University of Sciences and Technology, Pakistan. He is currently an Associate Professor with the Department of Computer and Software Engineering, College of Electrical & Mechanical Engineering, National University of Sciences and Technology (CEME NUST). He is one of the youngest Ph.D. in Pakistan and has published more than 115 research papers in peer-reviewed journals and conferences. In addition, he has also worked as a reviewer for many journals, including IEEE, Elsevier, and Springer. He has received academic awards, including the Best University Teacher Award, in 2016, and the Best Researcher Award, in 2019.

ZARTASHA MUSTANSAR received the Ph.D. degree from The University of Manchester, U.K. She was selected by Microsoft Research Cambridge (MSR) to pursue research in physical sciences and engineering in Manchester. She is currently working as an Assistant Professor with the Research Center for Modeling & Simulation (RCMS), NUST. She has published 32 research papers in various peer-reviewed journals and conferences. Her research interest includes biomechanical engineering, specially associated with health care.

AIMAL KHAN received the B.Eng. degree from the College of Electrical & Mechanical Engineering, National University of Sciences and Technology (CEME NUST), Pakistan, in 2016, and the M.S. degree in digital communication engineering from the University of Kiel, Germany, and the Ph.D. degree from the University of Rostock, Germany, in 2012. His Ph.D. research work was sponsored by DFG Germany. He had worked as a Research and Development Engineer at Seleon GmbH (pvt Ltd.), for three years, in which he was engaged in the development of efficient communication protocols and algorithms for robust performance of medical electronic devices. He is currently an Assistant Professor in the Department of Computer & Software Engineering, (CEME NUST). His awards and honors include NUST scholarship for the master’s degree in Germany and gold medal from Government of Pakistan for excellent performance in higher secondary school examination.