

Lung Disease Detection and Classification Based on AI, Deep Learning and Machine Learning: A Comprehensive Survey

Harith S. Hassan ^{a, }, Sinan A. Naji ^{b, }, and Abdullah G. Jaber ^{b, }

^aPostgraduate studies department, Research and Development Directorate, Ministry of Higher Education and Scientific Research, Baghdad, Iraq

^bDepartment of Business Information Technology, College of Business Informatics, University of Information Technology and Communications, Baghdad, Iraq

CORRESPONDENCE

Harith S. Hassan
harith.s.hassan@gmail.com

ARTICLE INFO

Received: Feb. 13, 2026

Revised: Jul. 04, 2026

Accepted: Jul. 14, 2026

Published: Sep. 30, 2026



© 2026 by the author(s).
Published by Mustansiriyah University. This article is an Open Access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

ABSTRACT: Background: Lung diseases significantly contribute to substantial global morbidity and mortality, which pose considerable diagnostic challenges. Manual analysis of medical images like chest X-rays and CT scans is often subject to human error, requires specialized expertise, and is time-consuming. To build fully automated systems that analyze the information contained in medical images, robust and efficient algorithms are required. In response, artificial intelligence (AI), especially deep learning (DL), has become an innovative approach that shows excellent results for the automation of lung disease detection and classification. **Objective:** Following a strict systematic literature review methodology, this study aims to provide an extensive survey of artificial intelligence, machine learning, and deep learning approaches for the detection and classification of lung diseases. **Methods:** The survey discusses over 30 reported approaches in the field. These approaches have been categorized and evaluated based on the methodology and reported results. We also discuss relevant datasets and benchmarking that have been used in state-of-the-art models and have undergone peer review. **Results:** The survey reveals a strong trend towards powerful ensemble and hybrid architectures like ResNet, VGG, RVCNet, Vision Transformers (ViTs), etc., and their variations. These models have achieved high accuracy for detecting and classifying various diseases. **Conclusions:** This work highlights the advancement and recent progress in using AI and deep learning to detect and classify lung diseases while addressing identified gaps and limitations related to dataset availability, standardization, and imbalance, in addition to model transparency and the computational cost, providing a prospective roadmap for further research or implementation.

KEYWORDS: Deep learning; Lung disease detection; Medical image classification; Transfer learning; Explainable AI

INTRODUCTION

Lung disease denotes various types of diseases or illnesses that prevent the proper functionality of the lungs [1]. Currently, lung diseases can lead to fatal conditions that pose significant diagnostic challenges [2]. The manual interpretation of medical images such as chest X-rays and CT scans is prone to human diagnostic errors, is time-consuming, and is dependent on physicians' expertise [3]. The rapidly expanding research in automatic lung disease detection is based on the premise that information about lung condition, illness, class of diseases, and underlying pathology can be extracted from medical data. In recent decades, lung disease detection and classification have attracted significant attention, and they have been studied by radiologists, medical physicians, and engineers [4]. In response, artificial intelligence (AI), particularly deep learning (DL), has emerged as a transformative approach, offering powerful tools to detect subtle patterns that humans might miss, enabling automated procedures, gaining new insights, and improving accuracy [5], [6]. Furthermore, big data has made a variety of large medical datasets accessible for researchers, enabling the implementation of DL models to detect and classify lung diseases with enhanced accuracy and efficiency [7].

The most widespread lung diseases include pneumonia, tuberculosis, asthma, fibrosis, cancer, interstitial lung disease, and COVID-19 [8], [9]. Typically, lung diseases are attributed to many factors, including bacterial, viral, and fungal factors, as well as genetic factors, lifestyle choices, and environmental causes [10], [11]. While numerous studies have demonstrated promising results in automatic lung disease detection, the field lacks a comprehensive thematic synthesis of these methodologies. The current literature exhibits significant variability in model architecture, preprocessing techniques, dataset selection, and evaluation metrics, which makes comparative analysis difficult. Distinct from other recent reviews in the field, this paper attempts to provide a comprehensive recent survey by adopting a wide-ranging, broad-scope methodology, such as knowledge-based systems, traditional machine learning, deep learning, explainable AI, and multimodal methodologies. It also highlights research gaps, methodological comparisons, evaluation metrics, publicly available datasets, and summary tables that not only focus on models' accuracy but also on generalizability, interpretability, and transparency, in addition to dataset limitations. With over 30 reported studies, this paper aims to give a comprehensive and critical survey of the current methodologies for automatic diagnosis and classification of lung diseases. Furthermore, we provide an overview of multiple models that utilize diverse datasets that have been used in these studies. When available, we have reported the corresponding performance metrics of each method.

MATERIALS AND METHODS

To maintain methodological integrity, transparency and consistent reproducibility, this survey follows a structured systematic review methodology to explore, evaluate, and synthesize the recent state of knowledge on the utilization of artificial intelligence and deep learning for lung disease identification and classification.

Information Sources and Search Strategy

A systematic, comprehensive literature search was implemented across five indexed databases, namely: Web of Science, PubMed, Scopus, IEEE Xplore, and Google Scholar. The search was defined as recent peer-reviewed articles from the last six years. The search strings strategy was formed using Boolean operators (AND/OR) combining multiple controlled keywords, limits were applied for publication period and document type where supported, Example search strategy for Google Scholar: ("machine learning" OR "artificial intelligence" OR "deep learning" OR "transfer learning" OR "explainable AI" OR "decision support system" OR "expert system") AND ("lung disease" OR "respiratory disease" OR pneumonia OR tuberculosis OR "lung cancer" OR "COVID-19") AND (2020/01/2025/12/31).

Inclusion and Exclusion Criteria

To guarantee structural integrity and transparency in this survey, a careful selection framework was established. The literature was systematically weighed against a set of inclusion and exclusion criteria designed to keep only high-quality, relevant research studies. Figure 1 demonstrates a complete PRISMA diagram showing the full selection process.

1 Inclusion Criteria

(1) Peer-reviewed journal articles stating primary quantitative findings. (2) Studies clearly focusing on AI, ML, or DL-based classification, detection, or segmentation of lung diseases. (3) Approaches employing medical diagnostic modalities, including imaging data (X-ray, CT scans, ultrasound). (4) Articles that have open access and are published completely in the English language.

2 Exclusion Criteria

(1) Non-peer-reviewed papers, extended abstracts, preprints, and short book chapters. (2) Studies missing precise validation metrics (e.g., accuracy, sensitivity, specificity). (3) Dissertations and theses, editorials, and commentary papers.

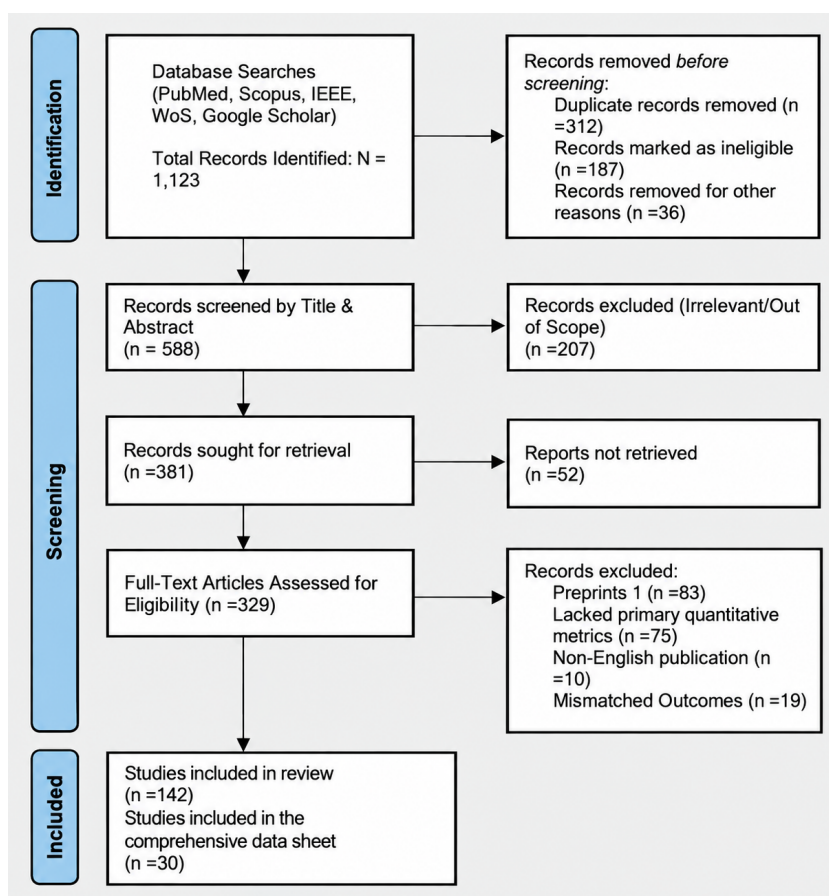


Figure 1. PRISMA flow diagram charting the systematic study selection process

Data Collection and Analysis

Both qualitative and quantitative comparative methodologies were systematically used across the final selected studies. Applicable key data elements were collected in an organized and uniform manner using a uniform extraction form. Key variables recorded covered a comprehensive set of fields, encompassing the study author, the specific lung pathology, data modality, model architecture, preprocessing methods, and the reported performance metrics.

LUNG DISEASE DIAGNOSIS BASED ON AI AND DEEP LEARNING

In this section, we review existing techniques for diagnosing lung diseases using various AI methods while focusing on deep learning works. Usually, these diagnostic methods consist of various phases, such as data preparation, preprocessing, segmentation, model design and training, fine tuning, and finally the evaluation and testing against prominent benchmark datasets [3], [12]. We classify AI detection methods into three categories; some methods clearly overlap category boundaries and are discussed at the end of this section. Table 1 compares the strengths and limitations of lung disease diagnostic methodologies.

Knowledge-Based Methods: These methods are designed mainly as expert systems for medical applications. Generally, most expert systems are rule-based methods that encode human knowledge of what constitutes a typical infection/disease. Usually, the rules capture the relationship between symptoms of the disease and the extent of the infection.

Machine Learning-Based Methods (ML): These methods aim to build models that are able to learn from data without explicit programming. Generally, ML models are trained by feeding them a large amount of data along with ground truth output. It repeatedly makes predictions, compares them to the ground truth, and adjusts its internal parameters to reduce prediction errors (i.e., through a process called optimization). These methods are designed mainly to detect fractures, lung cancer, breast cancer, brain bleeding, etc.

Table 1. A comparative taxonomy of lung disease diagnostic paradigms

Approach	Algorithmic Strengths	Critical Limitations	Medical Feasibility Transparency
Knowledge-Based Expert Systems	Runs with no training data	Highly dependent on human experts for rule creation	High Transparency: doctors can trace logical rules
Conventional Machine Learning (ML)	Very effective on low-dimensional layouts.	Depending on manual feature engineering (e.g., LBP, GLCM descriptors)	Moderate Transparency: Models like Decision Trees offer interpretability.
	Lower computational costs.	Affected by feature selection bias	Moderate Viability: Sensitive to human error during annotation.
Deep Learning (DL) Architectures	Automatic end-to-end multi-level feature extraction.	Serious computational constraints.	Low Transparency: Deep "black-box" models cause clinical suspicion.
	Outstanding scaling on massive datasets.	Susceptible to overfitting small/imbalanced datasets.	High Viability: Produces state-of-the-art diagnostic accuracy.

Deep Learning-Based Methods (DL): These methods aim to build models that are able to automatically discover the features needed for classification directly from raw data without human-engineered features. For example, DL models take the raw pixels and, through their deep layers, learn to identify edges, shapes, patterns, and objects by themselves. The correlations between an input image and the stored patterns are computed for detection.

These AI models are then used for detection, diagnosis, and classification. Table 2 summarizes the most significant and foundational studies with important representative methodologies that are reflective of the research in lung disease detection and classification within these three categories. To deliver a robust, uniform synthesis of these methodologies without disrupting the structural chronicle of this survey, a comprehensive evaluation matrix maps 30 primary experimental studies against their exact data modalities, datasets, core architecture, and variable metrics.

KNOWLEDGE-BASED METHODS

Knowledge-based methods, particularly expert systems, for medical diagnosis act as a powerful 'second opinion' tool. These systems use structured human knowledge that is encoded as rules to simulate expert reasoning that support clinical decision-making. Unlike purely machine learning models, these systems embed domain expertise to provide explainable recommendations. Recently, expert systems have shown significant advantage over machine learning, as these systems function effectively without relying on large training datasets [13]. Furthermore, computerized decision support systems (CDSS) are employed to improve healthcare by supporting the decision-making process. The CDSS assists providers in evidence-based decision making [10]. A critical aspect of expert systems and CDSS was the creation of a knowledge base, typically developed through extensive interviews with expert physicians to translate human expertise into if/then logical rules. In these systems, indirect methods like Dempster-Shafer theory and Forward Chaining were used to diagnose lung diseases based on symptoms [2], [14] or logical criteria for pediatric patients [15]. Recent advancements in expert systems display flexibility of use, from fuzzy inference systems designed for early identification to mobile applications that predict COVID-19 risk based on the semantic rules [16], [17]. Subramaniam et al, developed an expert system for classifying COVID-19 and pneumonia [18]. The input lung X-ray images were enhanced utilizing different preprocessing algorithms; then optimal features were extracted using Haar transform filters. The authors customized DNN for the multi-classification achieving 91% accuracy. Zulkifli *et al.* built an expert system by using several architectures to classify X-ray images for diagnosing COVID-19, pneumonia, tuberculosis, and normal lungs [19]. Personal interviews with specialists and patients were used to create a knowledge-based model. Classification with Inception3 model employing a dataset of 409 images showed the best performance with 86% accuracy, 91% sensitivity, precision and F1 score. Al-Issa *et al.* created a decision support system by applying and comparing knowledge-driven AI models (i.e., DenseNet201, DarkNet, Xception, and VGG16) all trained on the ImageNet dataset for multi-class lung diseases classification [11]. Xception model achieved the highest detection accuracy of 94.78%, yet when applying majority voting and feature ensemble techniques for feature extraction, the accuracy enhanced to 97.79%, 98.5% sensitivity, 99.07% specificity, and 98% precision.

Table 2. Some representative studies that include significant innovations and show the best performance

Ref.	Disease	Method	Model Architecture	Data Used		Metrics		
				Dataset	Samples	Acc%	Sen%	Spe%
[20]	Multi Class (13 diseases + normal)	Deep Learning – Segmentation (U-Net)	LCTfound, Vision Foundation Model	LungCT-28M	105,184 CT scans	–	–	–
[21]	Multi Class (5 diseases + normal)	Deep Learning – CNN / Transfer Learning	Lightweight two-dimensional convolutional neural network (LW-2DCNN)	Publicly available AIHub dataset	10,000 CT scans	99.80	99.68	
[4]	Atelectasis Pneumothorax	Deep Learning – Attention-Based	Attention-Based Modified DenseNet201	NIH chest X-ray	112,120 X-ray scans	95.34	96.00	–
[22]	Lung cancer	Machine Learning	Chameleon swarm-based support vector machine (CS-SVM)	Public datasets available on Kaggle	1,097 images	95.64	95.99	94.82
[5]	Multi Class	Deep Learning – CNN / Transfer Learning	Hybrid framework VGG19 + CNN	Consolidated multi-class collection of chest X-ray	80,000	96.48	93.75	–
[3]	Tuberculosis	Deep Learning – CNN / Transfer Learning	NFNet (Normalization-Free Network)	TBX11K	11,200 X-ray	96.91	91.81	98.42
[23]	Respiratory diseases	Deep Learning – CNN / Transfer Learning	CNN + Parallel Pooling, Linear Discriminant Analysis (LDA) + Random Subspace Ensembles (RSE)	ICBHI 2017	920 lung sounds. 6898 pulmonary cycles	71.15	61.00	86.00
[24]	Pulmonary nodules	Image Processing / Mathematical	TSASM (Three-Stage Automated Segmentation Method)	LIDC Image Dataset + RTH Image Dataset	100 CT scans	–	–	–

Kvak *et al.* introduced (Carebot) a CDSS along with DL to automatically identify and localize abnormal lung lesion anomalies in X-ray images [25]. The model can recognize and localize the small-scale lesions using YOLO-5 architecture that a human radiologist could overlook. The accuracy of the model end results was compared with the results of five individual radiologists achieving higher detection rates and lower false-negative rates. Cheng *et al.* developed a CDSS to support the identification of lung cancer in CT images [24]. The authors proposed a three-stage automated segmentation method mathematical framework (TSASM), which involved automated lung segmentation and feature extraction with discrete wavelet packet transform (DWPT), and rough set theory (LEM2 algorithm) for classification which outperformed other compared classifiers. On the other hand, Miyachi *et al.*

utilized Learning to Rank (LTR) approaches to develop a CDSS that generates a ranked list of differential diagnoses to enhance diagnostic accuracy by mitigating confirmation bias [26]. This system is especially helpful to aid the detection of difficult to diagnose or rare cases while reducing possible diagnostic errors.

In general, current CDSS research increasingly includes Explainable AI (XAI) techniques such as SHAP, LIME, etc. to improve trust and reliability of expert systems and CDSS [27], [28]. More details and representative work concerning Explainable AI will be discussed in next sections. A close assessment of knowledge-based systems uncovers a core limitation; they are heavily dependent on manual expert input for collecting knowledge. Although some semantic rule-based systems can perform well without the need for large training datasets, they lack acceptable generalisability when challenged with discrepancies in clinical presentations or patients with complex or rare conditions. Furthermore, some recent hybrid approaches that combine deep learning models with expert systems, could enhance total classification accuracy but they often lack the explainability feature that initially made traditional rule-based systems mainly valuable in clinical environment. Practically, building medical expert systems, CDSS, and Explainable AI models implies important preprocessing steps. These may include image segmentation and feature selection which are discussed below.

1 Image Segmentation

Image segmentation is considered a crucial preprocessing step in medical images. The automatic segmentation step aims to divide the source input image into several meaningful regions based on shared features or predefined rules, enabling more precise analysis and understanding of image contents. The final goal of the segmentation step is to detect and localize what is called Regions of Interest (ROIs) in medical images that identify some kind of infection or disease. Practically, accurate segmentation of tumors, organs, or lesions is essential for diagnosis and treatment plan. Liu *et al.* developed a lung segmentation method based on multi-scale superpixel analysis for interstitial lung diseases detection from CT scans [29]. The proposed method employed multiple RF classifiers to segment lung areas within the CT scans. Figure 2 shows the general architecture of the proposed system. The experimental results showed high segmentation accuracy averaging 96.45% using Dice Similarity Coefficient metric. Nair *et al.* proposed an enhanced lung cancer detection system integrating Improved Random Walker segmentation with Random Forest RF [30]. The improved Random Walker algorithm was used first to initially segment and highlight lung nodules, then ANN + RF classifiers were employed to categorize lung nodules into benign and malignant categories achieving classification accuracy of 94.76% for ANN, 99.65 % for RF, Sensitivity of 99.78% for ANN, 99.98% for RF, and Specificity of 28.97% for ANN, and 94.74% for RF. Abdikerimova *et al.* used an autocorrelation function method as initial segmentation stage to outline the regions with abnormalities in lung X-rays. After segmentation, the authors classify the images with k-means achieving an accuracy of 98% [31].

2 Feature Selection

Feature selection is a highly important step that enhances decision-making models. The main idea of feature selection in expert systems and ROI detection is to choose the most relevant features/variables based on domain knowledge while ignoring the irrelevant ones. In other words, finding features that accurately predict an outcome to build more efficient, accurate, and understandable models. Furthermore, feature selection also plays a critical role in managing the “curse of dimensionality” problem (i.e., too many features). This implies improving interpretability and robustness of rule-based systems or CDSS models. The significance of effective feature selection to achieve accurate diagnoses was acknowledged by many researchers. For instance, Medeiros *et al.* developed a classifier to discriminate between COVID-19 positive and negative patients by using and evaluating six distinct types of customized texture descriptors [32]. The authors observed that extracting features from X-ray or CT images using individual texture descriptors caused weak prediction of COVID-19; however, combining features from multiple descriptors such as LBP, GLCM, SCM, GABOR had enhanced performance, achieving 95.06% accuracy, 95.91% sensitivity, and 95.98% specificity. In related study, Chen *et al.* proposed a predictive method based on multi-layer artificial neural network (NNET) to analyze the extracted radiomic features and classify patient’s CT scans into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) based on radiomic features [33]. Their model achieved an area under the curve (AUC) metric of 0.90 with sensitivity and specificity of 0.85.

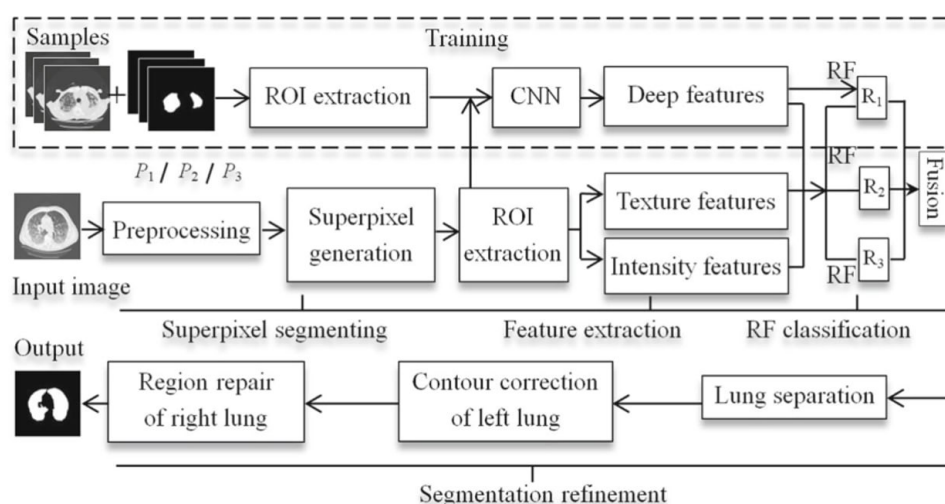


Figure 2. Image segmentation for ROI detection was adopted by C. Liu *et al.* The ROI are detected from CT scans based on multi-scale superpixel to group pixels with similar intensities and spatial characteristics [29]

Machine Learning-Based Methods

Machine learning ML is a core branch of AI that has been successfully applied as a detector/classifier of abnormal ROI and infection details in medical images [34]. Compared to other detection and classification models, ML techniques showed excellent and promising performance. These models have become an important partner for doctors and healthcare experts, in order to enhance their expertise, achieve faster and more accurate diagnosis, and reduce human error due to fatigue. Practically, machine learning-based methods aim at building systems that are able to learn by example or from experience without the need for definite programming logic. Typically, a lot of data and ground truth output are fed into ML models to train them. The ML models continuously generate predictions, compare them to the ground truth, and modify its internal parameters to boost the model performance. It stops training phase at the point where the model has learned the generalizable patterns from the training data. In other words, the training phase should be stopped when it reaches the best results with minimum prediction errors.

The core process of ML models generally follows these steps:

Data Acquisition: Collecting medical images/sound files for training the system (i.e., X-rays, MRIs, CT scans, dermatology photos, pathology slides, sound files, etc.). Practically, training phase requires a huge amount of training datasets.

Data Preprocessing: The aim of this step is to standardize the dataset to ensure consistency for the model (e.g., images resizing, data normalizing, noise removing, etc.).

Annotation: For supervised learning, the ground truth is needed to be prepared by specialists to label the data (e.g., outlining the infected region, drawing bounding box around tumors, generating binary images, etc.).

Model Training: The model, normally a Neural Network (NNs), is trained with a huge amount of data accompanied by its ground truth (i.e., thousands, sometimes millions, of these labeled data).

Validation & Testing: The performance of the ML model is tested and evaluated using new dataset that has not been seen before. Many evaluation metrics like accuracy, sensitivity, specificity, etc., will be discussed in the next sections.

Generally, ML models had been used in three main medical tasks: for categorizing an entire image (i.e., classification), locates and identifies diseases or abnormalities within an image (i.e., object detection), and creating a detailed map of diseased tissue (i.e., semantic segmentation). Practically, many representative works used neural networks for lung disease detection and classification. Wu *et al.* designed a multilayer machine learning classifier based on fractional-order convolution, combined with Gray Relational Analysis (GRA) and a Radial Bayesian Network (RBN) to identify ROI and predict seven lung diseases from X-ray images, reporting 83.57% accuracy, 98.68% sensitivity, and 82.42% precision on the NIH dataset [35]. In a relevant study, Wang *et al.* developed an algorithm to classify adenocarcinoma (AC) and squamous cell carcinoma (SCC) subtypes of lung cancer in patients [36]. The authors combined a KNN classifier with feature reduction methods PCA, SVM-RFE and

the borderline2-SMOTE resampling method, to reduce dataset imbalance, resulting in area under the receiver operating characteristic curve AUC-ROC of 0.63.

Alternatively, Kocks *et al.* designed a classification model named AC/DC Asthma/COPD Differentiation Classification, which is an ensemble of 100 classifiers that employ electronic health records to distinguish between three classes of lung diseases with similar symptoms [37]. This model produced a 73% classification accuracy outperforming respiratory specialists or primary physicians. Oyediran *et al.* employed the Chameleon Swarm Algorithm (CSA) to enhance Support Vector Machine SVM performance to classify lung cancer images by using the best parameters [22], integrating CSA into the SVM contributed to enhanced performance while reducing time for malignant nodule recognition, achieving an average 95.64 accuracy, 95.99 sensitivity, and 94.82% specificity.

Siordia-Millán *et al.* employed Artificial Neural Network (ANN) among few other algorithms in order to examine health signs and laboratory results to discriminate between Pulmonary Thromboembolism (PTE), Pneumonia, normal cases [38]. The authors created the study dataset based on 173 laboratory reports extracted from the EHRs. The highest AUC-ROC results for structured data were achieved by the naïve Bayes algorithm, while BiLSTM model scored only 57.6% for unstructured data. Li *et al.* designed a Bayesian Network (BN) based on Probabilistic graphical model (DAG) to analyse electronic medical records (EMRs) of 33,571 children under 5 years old to diagnose community-acquired pneumonia (CAP) [39]. The network achieved 99.47% recall on external validation dataset. Sherafatian *et al.* used Recursive Partitioning and Regression Trees (RPART) to classify lung cancer and identified microRNA biomarkers [40]. The authors analysed data of 1068 samples from The Cancer Genome Atlas (TCGA) dataset and employed focusing on binary classification of lung cancer and healthy tissue with AUC of 91.2%. Random Forests networks were also efficient, Atzeni *et al.* designed a machine learning framework integrating data from electronic health records, in addition to personal air quality monitor data to predict the probability of Chronic Obstructive Pulmonary Disease (COPD) flare ups [41].

While Yang *et al.* combined Random Forest with Bagging and SHAP (Shapley Additive Explanations) to explore the correlation of a number of factors including weather, pollution and medical visits on respiratory health [42]. The authors analysed the relationship between these factors and their effect on respiratory diseases in Tianjin, China, from 2013 to 2019, to predict daily patient volume with strong generalization and stability. Finally, Saputra *et al.* focused in their study on lung texture analysis to improve the accuracy of pneumonia and COVID-19 detection in X-ray images by using a Modified K-Nearest Neighbor (MKNN) and Gray Level Co-occurrence Matrix (GLCM) statistical method [43] which led to classify images with 93.3% accuracy. The primary limitation of classical machine learning methods in remains in their strong dependence on manually (expert-defined) feature extraction processes. Although combining multiple automatic descriptors (e.g., LBP, GLCM, and Gabor filters) partly improves feature representation, it also escalates the computational and statistical challenges associated with high-dimensional representations. This, as response, require demanding, multi-stage feature reduction techniques such as PCA or SVM-RFE. In conclusion, traditional, ML models serve predominantly as supportive diagnostic tools for carefully selected and pre-processed datasets, rather than as strong, automated and complete systems that are capable of directly managing high-resolution raw medical images.

Deep Learning Methodologies

Deep Learning (DL) is an interesting, advanced subfield of ML that employs neural networks and unstructured data to discover complicated patterns used in many objects' detection and recognition tasks. The key principle behind DL lies in the "depth" of neural networks architecture and the capacity of automatic feature extraction [44]. The main structure of most DL models implies an encoder and decoder in which the encoder takes the input data, such as images, and transforms it into a compact, compressed and discriminative set of meaningful features. Subsequently, the decoder takes the encoded representation from the encoder and reconstructs the original data or generating a new output. The next sections discuss the most common architecture used in DL.

1 Convolutional Neural Network (CNN) Architectures

A Convolutional Neural Network (CNN) is a type of DL model that handle two-dimensional array data such as medical images by applying convolutional operations to automatically extract multi-level representations of features. Furthermore, CNNs are typically used for image recognition and classification tasks. The most commonly used CNNs models are LeNet, AlexNet, VGG, GoogLeNet (Inception), and ResNet. Several studies proposed adapting CNNs models to overcome the challenge

of multiclass lung disease classification. VGG-Net is considered one of the best CNN architectures that is used for object detection and image classification [45]. Figure 3 shows the structure of standard VGG-16 which is an interesting CNNs architecture. Later many versions of VGG with many variations had been proposed. Ibrahim *et al.* developed a model that combined VGG-19 Net and custom CNN to classify lung cancer, pneumonia, COVID-19, plus normal lungs achieving 98.05% accuracy and sensitivity, 99.5% specificity [46]. Likewise, Lyu *et al.* addressed the challenges of binary and multi-classification by proposing an innovative multi-level cross residual network (ML-xResNet) achieving 85.88% accuracy for multiclass and 92.19% for binary lung cancer classification [47]. An extensive dataset was collected and prepared by Abdulahi *et al.* then the authors used Wide Residual Networks to create their model PulmoNet that achieved 91% accuracy, 95.01% sensitivity, 88.65% AUC score for classifying lung pneumonia types [48]. Other studies aimed for higher accuracy by employing various modern architecture. Moussaid *et al.* proposed the Post-Pandemic Classification Algorithm (PPCA) algorithm that is based on EfficientNetB7 to classify lung pneumonia types achieving an accuracy of 99.81% for X-ray images and 99.88% for CT scans [6].

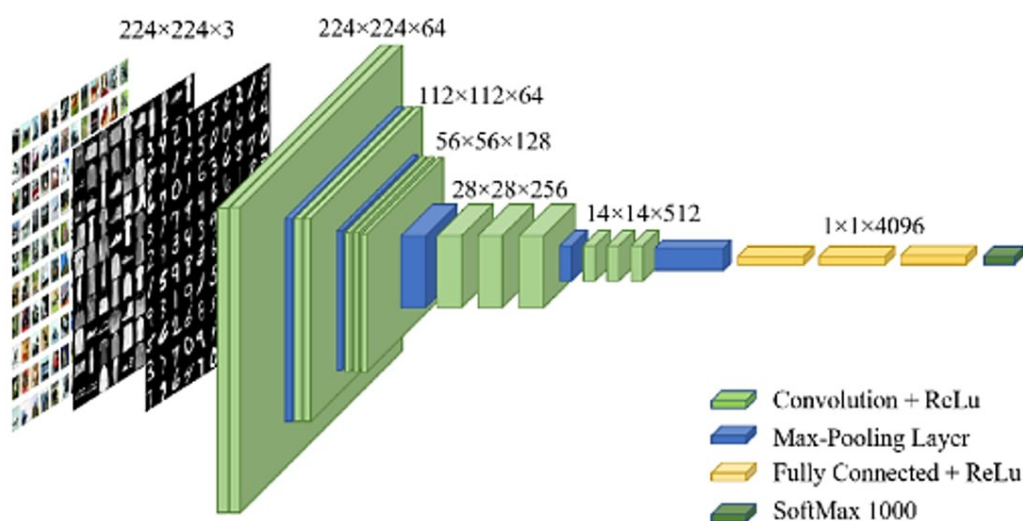


Figure 3. VGG16 Model, which consists of 16 convolutional layers, adapted by De *et al.* [49]

In parallel with adopting baseline architecture, creating customized deep learning models has also shown good performance. Jawahar *et al.* proposed CovMnet, a lightweight custom CNN model that achieved 97.4% accuracy, 100% sensitivity, 97.73% specificity in the COVID-19 classification task of chest X-rays [50]. In the same way, Ullah *et al.* created a custom-built CNN framework named DeepLungNet, for the classification of five severe lung diseases employing Leaky ReLU and fire modules, in addition to augmentation to improve generalizability and robustness. The model achieved 97.47% average accuracy, 93.6% sensitivity, 93.2% precision, and 93.4% F1 score [51]. Apart from fundamental architectural design, some studies are also devoted to performance improvement and clinical translation. Sriporn *et al.* evaluated the performance of multiple models to predict lung disease by applying the Mish activation function in combination with seven optimizers using a dataset of 5810 images for training and validation [52]. The authors showed that fine tuning parts like Mish activation function and Nadam optimizer in combination with a DenseNet-121 can improve accuracy of lung disease prediction to 98.88%. Approaching practical applicability, Kasinathan & Jayakumar proposed a hybrid cloud-based lung tumor detector and stage classifier system Cloud-LTDSC that analyzes CT and PET scans to classify lung tumors [53]. The system combines active contour segmentation and multilayer CNN with the Cloud environment for real-time detection and diagnosis in remote destinations, resulting in 98.6% accuracy, 95.9% sensitivity, 93.9% specificity, and 94% and demonstrating a way for integrated clinical systems.

2 Transfer Learning Based Models

Transfer learning is a subfield of machine learning that repurposes knowledge and fine-tunes pre-trained models for a new domain, it has shown significant promise in the application of diagnosing pulmonary diseases from medical images, especially with new data [54]. This approach has been used

to obtain remarkable accuracy in the field of medical image analysis. For example, Al-Sheikh *et al.* designed a classification framework to automatically categorize six diverse lung diseases by employing 1714 CT scans and X-ray images from two publicly available datasets [55]. This classification framework joined a custom CNN with VGG-16Net and AlexNet for feature extraction and classification. The authors proposed k-symbol Lerch transcendent functions, a novel image enhancement preprocessing step, achieving 98.80% accuracy on CT scans and 98.60% on X-ray images. Similarly, Badawi and Elgazzar created Chest-Xray15, a balanced dataset that comprise of 15,000 normal, pneumonia, and COVID-19 images from 11 publicly available datasets in addition to augmented images [56]. The authors customized a VGG-16 Net model to achieve the highest accuracy of 99.62% in binary classification, and 95.48% in multi-class classification.

Kor *et al.* developed a fine-tuned model for the automatic classification of COVID-19, viral pneumonia, and normal lungs, focusing on processing raw images, applying minimal preprocessing on the Cohen and Kermany datasets [57]. The study employed pre-trained NASNet-Mobile model and added a dropout layer to mitigate overfitting, achieving an average classification accuracy of 97.78%. There have been attempts in the line of streamlining diagnostic pipelines, such as an end-to-end system that directly processes raw images. Kim *et al.* proposed a robust and scalable deep learning framework employing pretrained EfficientNet V2M model, accomplishing 82.15% accuracy classifying multi lung diseases from X-rays images evaluated on both NIH public dataset, and SCH in house datasets [58]. This approach input raw images directly and trains the entire network at once eliminating the intermediate steps. Phudowski and Mulawka designed a novel integrated clinical decision-support system with a graphical user interface (GUI) to classify four classes of diseases [59]. The authors used U-Net for segmentation then ResNet-50 model for classification using the CheXpert dataset that contains 224,316 labeled X-ray images. Prior segmentation resulted in a 21% improvement over no segmentation when detecting pneumothorax. Similarly, Vidal *et al.* constructed a novel multi-stage transfer learning methodology to segment the lung regions from chest X-ray images obtained with portable devices to support clinicians [60]. The authors repurposed brain MRI U-NET by adapting it to general chest X-rays using the Kermany *et al.* dataset, then fine-tuning it on the limited low-quality images in the CHUAC dataset created by radiology staff at a Spanish hospital using portable devices. The segmentation performance achieved a Dice coefficient of 0.976 for COVID-19 patients, 0.980 for health, and 0.9769 for non-COVID pulmonary diseases. Zak and Krzyżak proposed a transfer learning framework using VGG-16 Net, ResNet-50, and InceptionV3 Net to classify lung diseases addressing data scarcity by utilizing small common datasets [54]. Furthermore, the authors used U-Net for segmentation then Inception V3 for classification on the Shenzhen and Montgomery chest X-ray datasets images to achieve 89% accuracy for identifying pneumonia and tuberculosis.

Comparative studies further note that the choice of model depends on the task at hand. Tang *et al.* proposed a CNN solution to classify X-ray images as normal or abnormal cases using several leading CNN architectures, including AlexNet, VGG-16, GoogLeNet, ResNet18, ResNet50, and DenseNet121 [61]. The models were trained on 8,500 images extracted from the NIH Chest X-ray dataset, then a comprehensive evaluation of these architectures was conducted to find out designs that balance accuracy and speed such as in ResNet, is generally highly effective on the mid-sized data, and models that are more complex do not necessarily produce better results. Data fusion techniques can also improve the classification results, Yang *et al.* proposed a transfer learning technique in utilizing CT scans and X-rays data fusion to classify COVID-19 and pneumonia [45]. The authors created and employed a unified dataset that comprising 8,461 X-ray images, they also used the SARS-CoV-2 Dataset which contain 2,481 CT scans. The authors employed Fast.AI framework and ResNet-50 to achieve 96% classification accuracy from CT scans, while a finetuned VGG-16 model achieved 99% accuracy from X-ray images. Brunese *et al.* developed a DL framework for pulmonary diseases and COVID-19 detection in chest X-rays, substantially reducing detection time [62]. The framework depends on a sequential classification approach based on VGG-16 which detects whether an image is affected by a disease with 96% accuracy, then a second VGG-16 classifies that image as pneumonia or COVID-19 with an accuracy of 98%. The authors combined dataset from three public data sources with 6,523 X-ray images, and Gradient-weighted Class Activation Mapping (Grad-CAM) was also employed to highlight COVID-19 infected areas. Priyadarsini *et al.* developed a deep learning framework for classifying lung opacity, pneumonia, tuberculosis, COVID-19, and lung cancer using 7,385 chest X-ray and CT images from three Kaggle datasets, supplemented by extensive augmentation techniques [63]. The authors compared three CNN architectures: sequential CNN, functional CNN, and a pre-trained VGG-16 model. The sequential model demonstrated superior performance achieving accuracy of 98.43% for pneumonia and 99.4% for tuberculosis detection, while the functional model showed outstanding accuracy of 99.9% for cancer detection. Shamrat *et al.* developed MobileLungNetV2, a customized deep learning model based on MobileNetV2 for multiclass classification of fifteen different

pulmonary diseases [64]. The authors employed the ChestX-ray14 dataset, applying preprocessing to produce a balanced dataset of 51,987 X-ray images. Performance of the model was evaluated against other transfer learning models on the same dataset, with MobileNetV2 achieving the highest accuracy of 91.6%. Furthermore, the study fine-tuned the model by adding several layers achieving 96.97% accuracy, 96.83 sensitivity, 99.78% specificity, and 96.71% precision.

More works proposed to use hybrid models that combine deep learning with other algorithms to improve the performance of the proposed systems. More details about hybrid models can be found in [5], [9], [65], [66], [67], [68]. While transfer learning models achieve high classification accuracies (often exceeding 95%), an operational analysis uncovers that these architectures regularly capture shortcuts rather than authentic pathological features. Many models reporting almost-perfect results on public chest X-ray datasets demonstrate extreme performance drops when employed in an external environment. This performance drop strikes since deep networks heavily utilize dataset-specific artifacts, medical tools variations, and image background characteristics instead of separating identified pulmonary lesions. Therefore, high accuracy in transfer learning often led to a considerable lack of generalizability, declining the clinical strength in the absence of a preceding segmentation module responsible for isolating relevant anatomical regions of interest preceding to classification stage.

3 Explainable Artificial Intelligence (XAI)

Although deep learning models showed excellent results in medical diagnoses, detection, and classification, these models have a “black-box” nature that raises challenges for clinical adoption and trust. In other words, if users do not understand why an AI made a decision, they may not trust it, emphasizing the necessity for developing an explainable AI approaches [69]. Generally, explainable artificial intelligence (XAI) approaches aim to make the decisions and outputs of prediction and classification models understandable and explainable for physicians, enabling them to interpret and trust the predictions with greater transparency [70]. These methods identify the ROI that mostly influenced the decision making, and support the physicians to interpret the individual predictions and understand the logic behind them [27], [71], [72]. Furthermore, XAI helps in detecting biases in AI decisions and identifying model errors that helps developers to refine the system.

The explainability was integrated into deep learning methodologies through two main strategies, most studies apply post-hoc explanation techniques such as LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations), LRP (Layer-wise Relevance Propagation), and Grad-CAM to visualize how models reach specific decisions after the predictions have been generated by the model [62], [73], [74]. On the other hand, some methods aim to design architectures that are fundamentally transparent, considering interpretability from the beginning, employing approaches such as attention mechanisms, rule-based extraction, and decision trees [75]. In this direction, Naz *et al.* developed an explainable AI framework that classify and interpret several lung diseases using a transfer learning ResNet50 model, utilizing X-ray images from two public datasets achieving a classification accuracy of 97% on the COVIDNet and 93% on the COVID-CT dataset [69]. The authors applied Local Interpretable Model-agnostic Explanations (LIME) to clarify ResNet50 model predictions, by highlighting the specific regions of X-ray image that have a significant influence on classification results.

Chadaga *et al.* designed a decision support system to diagnose COVID-19 from blood test reports, employing stacked ML, random forest, deep neural networks and one-dimensional convolutional networks in addition to explainable AI techniques such as SHAP, LIME, QLattice, and ELI5 obtaining 96% accuracy, 95% sensitivity, 94% precision and F1 score, and 98% AUC [27]. The authors used these explainable AI methods to enhance model transparency and explainability and emphasize the importance of critical features conforming COVID-19.

Patel *et al.* integrated the Grad-CAM explainable AI approach with EfficientNet to classify five lung diseases from a diverse dataset of 7,389 X-ray images achieving 96% accuracy [72]. Grad-Cam highlights the areas in images using heatmaps that draw attention to the critical ROI and boost model interpretability and transparency for both radiologists and physicians. The study indicates that explainable methods assist to interpret false negatives and positives, enabling future refinements on the highlighted relevant features [72], Figure 4 demonstrates how heat maps generated by Grad-CAM indicate visually the precise location of pulmonary abnormalities, allowing radiotherapists to validate if the model's focus in line with the standard medical diagnosis. Brunese *et al.* proposed a similar technique to diagnose COVID-19, employing VGG-16 and explainable AI with Grad-CAM achieving 98% detection accuracy on a dataset of 6,523 chest X-ray images [62]. The authors utilized

the Grad-CAM to generate heatmap visualizations outlining the areas of the X-ray that influenced the model's prediction based on clinically relevant features.

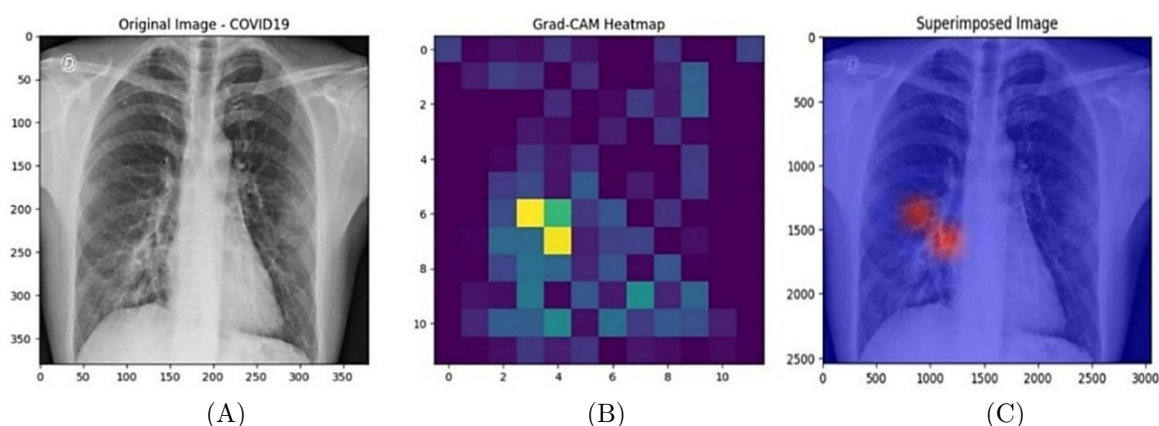


Figure 4. The GradCAM visualizations for COVID-19: (A) original image; (B) heatmap highlighting salient regions; (C) superimposed heatmap on the original image; adopted by Patel *et al.* [72]

In a subsequent study, Khan *et al.* proposed a deep transfer learning and explainable AI framework using customized VGG16 and EfficientNet models to diagnose and classify COVID-19, pneumonia and lung opacity achieving 99.1% accuracy [73]. The study integrated the Whale-Elephant Herding algorithm for feature optimization, then the Grad-CAM technique for explainability verifying that the model supports its predictions on meaningful pulmonary features, improving interpretability and adoption trust. Ho and Gwak adapted DenseNet-121 within a self-training knowledge distillation (KD) framework to identify thorax diseases in X-ray images on the ChestX-ray14 dataset and identified 14 lung classes with an average 0.826 AUC [76]. The study tested several saliency mapping techniques such as Guided Backpropagation, SmoothGrad, and SmoothGrad integrated Gradients to generate heatmaps that highlight the important (abnormal) regions in chest X-ray images, and they found SmoothGrad IG to be the most effective. Rajpoot *et al.* integrated multiple transfer learning models including VGG-16, DenseNet-169 and ResNet50 to build an ensemble framework to classify COVID-19 from chest X-rays and CT scans achieving 96.18% for CT scans and 99% for X-rays, focusing on explainable AI to improve interpretability [77]. The authors applied Grad-CAM, LIME, SHAP and Grad-CAM++ techniques to highlight the key regions that influenced the model decision making process through visual explanations, enhancing interpretability and trust. Amann *et al.* argued that explainability is not purely technical, but also has different dimensions as legal, ethical and medical. The authors implemented a comprehensive ethical evaluation and assessment about explainability responsibility [28]. The authors utilized AI based clinical decision support system (as a case study; they analyzed the system through multiple perspectives, and before conducted the “Principles of Biomedical Ethics” assessment framework beforehand. The authors stress it is essential to address explainability challenges which affect policymakers, physicians, and system developers. E. Ihongbe *et al.* conducted a human-centered user study involving 26 medical experts to evaluate LIME and Grad-CAM visual explainability algorithms [78]. The authors first used AI models to diagnose pneumonia and COVID-19 using publicly available datasets. They then applied LIME and Grad-CAM algorithms to generate visual explanations of the models’ predictions. The qualitative user study evaluated the visual explanation of both LIME and Grad-CAM from the end user perspective concerning comprehensibility, confidence and clinical relevance. Although most participants favored Grad-CAM heatmaps compared to LIME, participants were indeterminate about whether explainability has a significant role in building trust toward AI systems adoption.

4 Recurrent Neural Networks (RNNs)

RNNs are a special form of neural networks that are particularly effective with time series and sequential data [79]. These networks can keep information (i.e., has a memory with loops) from previous inputs and use it to create the next output in the sequence. According to Galić, *et al.*, RNNs could be integrated with CNNs to extract spatial features from sequential data to improve the results [80]. The RNNs had been incorporated in studies utilizing sound recordings, as it is a powerful tool particularly

effective for identifying both sequential and temporal patterns. Among RNN variants, Long Short-Term Memory networks (LSTM) are considered the most employed when building hybrid models that usually combine them with CNNs. The RNN-LSTM was utilized by multiple studies, including [81], [82], to achieve high accuracy classifying lung diseases from lung sound recordings. Although many studies utilized sound recording data, a system designed by Islam *et al.* achieved 99.86% accuracy in detecting COVID-19, confirming that RNNs are capable of sequentially and effectively processing spatial features produced by a CNN from chest X-rays [79]. Similarly, Goyal and Singh study demonstrated great flexibility in hybrid/ensemble classification approaches using medical images and not sound recordings [66]. Furthermore, the flexibility of RNNs was demonstrated by Lam, *et al.* when they used EHR data of 40,703 patients to predict acute respiratory distress syndrome (ARDS) and 12 other associated diseases, achieving an area under the curve (AUC) of 0.842 [71].

5 Segmentation-Integrated Models

Recent studies effectively use deep learning segmentation-integrated models as a foundational step to improve the detection and classification of various lung diseases. This segmentation methodology, where the region of interest is first outlined before classification, has continuously returned high accuracy. So, instead of using separate models for detection and outlining, these models perform both tasks together for better accuracy and efficiency. One example, Crasta *et al.* presented a lung cancer segmentation and classification model that is applied on the 3D representation of CT scan [83]. The system employed a volumetric 3D-VNET, an extension of U-Net for effective segmentation of ROI, then the authors employed 3D-ResNet to classify the segmented ROI as cancerous or noncancerous with 99.2% accuracy, 98.8% sensitivity, and 99.6% specificity. Similarly, Sharma *et al.* applied a customized and improved U-Net model as a preprocessing step for tuberculosis detection on 704 chest X-ray images, achieving 96.35% accuracy with U-net [84]. This precise identification of the ROIs allowed the Xception classification model to achieve up to 99.29% accuracy with an AUC of 0.999. Vinta *et al.* developed a hybrid framework and applied a foundational improved U-Net model for HRCT image segmentation that was enhanced through dense skip connections, custom loss function, multi-scale feature extraction, and deep supervision [85]. The proposed hybrid framework was able to classify six types of ILD from the MedGIFT dataset using MobileUNetV3 with 99.10% accuracy, 98.05% sensitivity, 98.65% precision and F1 score of 98.13%. Following a comparable approach, Shafi & Chinnappan proposed a Modified SegNet (M-SegNet) architecture that is based on the authors' developed tanh-Softplus activation function for robust segmentation of lung image ROI, achieving 96.23% segmentation accuracy [86]. Following segmentation, a hybrid classification model based on Scale Dot Product Attention-based SqueezeNet (SDPA-SqueezeNet) and a DCNN were employed to determine the level of disease severity, achieving mean 94.4% accuracy on the HRCTCov19 dataset. To improve the performance of lung cancer segmentation, Shetty *et al.* employed The Lung Image Database Consortium (LIDC-IDRI) dataset for lung area segmentation [87]. The Lung Lobe Segmentation using Bayesian fuzzy clustering method was used to segment the entire lung boundaries in the CT images, then cancer regions are segmented using the novel optimized deformable model that incorporates Water Cycle Algorithm and Sea Lion Optimization algorithms (WSLnO), achieving 93.03% segmentation accuracy and 91.23% classification accuracy using Shepard Convolutional Neural Network ShCNN model.

6 Multimodal and NLP-Enhanced Models

Recent deep learning advances for lung disease diagnosis integrate multiple data sources such as, electronic health records, genomics, natural language processing (NLP), clinical notes besides medical images and sound recordings, to achieve a more inclusive representation by combining different data types. For instance, Sangeetha *et al.* proposed a Multimodal Fusion Deep Neural Network (MFDNN) which was trained on clinical data, genomics, and medical imaging that achieved 92.5% accuracy, 86.4% sensitivity, 87.4% precision and 86.2% F1 score for cancer detection, outperforming traditional models like CNNs [88]. Some other studies have utilized NLP to analyze unstructured text; for instance, [89] Liang *et al.* developed an AI system to diagnose lung diseases using clinical features from 31,267 electronic health records EHR data. These records include multimodal data such as clinical data, laboratory results, medical reports, and patient demographics. They employed Bidirectional Encoder Representations from Transformers (BERT) for text understanding, Bi-LSTM and CRF for feature extraction, and LightGBM for classification. This proposed system outperformed both human experts and ChatGPT 4.0 with 0.899 F1 score. In a different mixed-image-text strategy, Wang L. proposed a model named DeepMRD which is based on a modified 3D ResNet-18, that incorporates

both radiological images and NLP-extracted information from medical reports [90]. This model achieved 95% accuracy and proved to be time-efficient for decision-making. Together, these studies indicate that models augmented with multimodal data and NLP are driving progress toward more accurate and clinically relevant diagnostic system development

7 Vision-Transformer-Based Methods

Vision Transformers (ViTs) are a state-of-the-art Deep Learning (DL) architecture motivated by Natural Language Processing (NLP) philosophies, which integrates transformer architecture with self-attention mechanisms and capable of capturing long-range, universal and complex relationships across the input image [91]. The main idea of ViTs is to divide the input image into fixed-size patches. Then, ViTs treats an image as a sequence of patches and applies multiple layers of multi-head self-attention for feature extraction. Singh *et al.* proposed a self-attention ViT for pneumonia detection from CXRs reaching 97.61% accuracy, 95% sensitivity, 98% specificity and 95% F1 score, outperforming traditional approaches [91]. The authors stated that ViT required longer training times and computational resources compared to CNNs. Hybrid systems that incorporated ViT together with CNNs have demonstrated high performance. In one such approach, Ukwuoma *et al.* proposed a framework which utilized an ensemble of DenseNet201, VGG16, and GoogleNet for deep features extraction, then a Global Second-Order Pooling technique to enhanced ensemble models capabilities [8]. Finally, a ViT multi-head self-attention is applied using a scaled dot-product attention mechanism achieved 98% for COVID-19 and pneumonia detection. In a subsequent study, Wang *et al.* integrated ViT and ResNet18 to create PneuNet, a model to diagnose COVID-19 and differentiate between different pneumonia types that has very similar symptoms in chest X-ray images [92]. The authors applied multi-head attention to channel-based patches extracted by the ResNet to identify lung disease abnormalities, achieving 94.96% accuracy in the three-categories classification.

8 Generative Models and Synthetic Data Generation

Generative models such as GANs, VAEs, and Diffusion models are powerful deep learning paradigms. They address data shortage and imbalance by generating realistic synthetic samples from prior examples, which in turn improves the performance, robustness, and generalizability of diagnostic systems. Many innovative representative frameworks show this potential. Rezaei and Ahmadi proposed an innovative hierarchical Generative Adversarial Network (GAN) with an ensemble CNNs detection system [93]. This system is based on two steps: lung segmentation that utilizes data from Luna dataset using U-net model generator, and nodule localization with R-CNN mask generator. The model generates high-quality artificial lung images at several scales to augment the training dataset, afterwards the ensemble CNNs combines the predictions of multiple CNNs to increase the model sensitivity detecting true small nodules reaching 90.1%. Li *et al.* proposed CAGAN, Classifier-Augmented GAN framework designed to detect and localize lung lesions caused by COVID-19 in chest CT scans [94]. The framework combines abnormality map generator to indicate ROI, a discriminator to ensure generated images are realistic, and classification network (3DResNet) to focus on lesion-related regions. The method outperformed both traditional CAM-based and previous GAN-based augmentation approaches in accuracy recording 0.884 while requiring less labeled images. Many research reports the advantages of synthesizing data. Another research has applied GANs to synthesize chest X-ray and CT images to perform balancing on COVID-19 and lung cancer classifier tasks with accuracy over 98% [95], [96]. Many studies demonstrated the advantages of producing synthesized data using GANs for lung disease diagnosis. Gulakala *et al.* constructed GAN architecture based on Residual U-Net with instance and batch normalization to generate 2000 synthetic COVID-19 chest X-rays [95], tackling lung disease data imbalance as the original dataset contain 417 images only; the authors proposed a compact CNN classifier achieving 99.2% accuracy. Similarly, Salama *et al.* developed a generative framework based on convolutional generative auto-encoder for lung cancer detection and classification that produced high-quality synthetic medical images [97]. These produced images are used for class balancing, expanding a small lung cancer dataset to 1000 balanced synthetic images varying in size and location representing different lung tumor types. ResNet50 model was utilized to train the classifier network on both real and synthetic image and detect benign and malignant cases with accuracy of 98.91%. Synthetic generation is not limited to images, for instance Convolutional VAEs have been employed to generate fake lung sounds to augment sound datasets and increase diagnostic performance for lung diseases [98]. This shows the flexibility of generative models for creating strong and well-balanced datasets for medical AI.

9 Attention Mechanisms

Advanced attention mechanisms are DL techniques that enable the NNs to focus on the most important parts of the data input. This will improve their ability to handle complex and long sequences of information. It also provides a visual representation of image regions that are most influential in the model prediction [99]. These include dual attention (channel and spatial), self-attention modules outside of transformer architectures, and related innovations. Attention mechanisms have growing importance in deep learning for medical imaging, and demonstrated substantial success across several diagnostic tasks [94]. For instance, in lung cancer diagnosis, Chen *et al.* utilized a deep attention-based multiple instance learning (MIL) methodology for classifying lung cancer employing radiomic characteristics extracted from CT nodules [100]. The method achieved mean accuracy of 0.807 on the LIDC-IDRI dataset, outperforming other MIL methods. A novel oversampling technique was also proposed to address class imbalance and improve the model performance. Attention visualization highlighted the diagnosed nodules for interpretability and to provide justification for the model's predictions. Similarly, for cancer nodule detection, UrRehman *et al.* presented a DL framework for lung cancer nodule detection and classification in CT images, that integrated a dual attention mechanism into CNN [101]. To highlight the most important features and regions within lung images, the channel and spatial attention modules were both employed to improve the custom (CNN) architecture. Evaluation of this framework using data from LUNA16 dataset demonstrated that integrating the attention modules with the custom CNN exhibited high performance delivering 95.40% accuracy. In pneumonia screening, Li proposed a ResNet architecture that implements attention mechanisms to improve pneumonia detection in chest X-ray images [102]. The author proposed a unique methodology for attention mechanism by generating side branches incorporating spatial and channel attention modules during the convolution process. These auxiliary branches generate additional feature maps, to enrich the representational capacity achieving 96% accuracy. Enhanced Focal Loss function also proposed to overcome the issue of class imbalance and optimize the model performance, achieving high accuracy of 98.6% compared to relevant models.

SOUND AND SIGNAL-BASED APPROACHES

Sound and signal-based approaches use audio signals for lung disease detection. These methods require special hardware/sensors to capture sound signal data of lungs and/or respiratory. Then, AI models are trained to listen to lung recordings data and recognize sound patterns linked to diseases. For instance, crackles are often associated with pneumonia, while wheezes are a hallmark of asthma and COPD. Generally, the incorporation of lung sounds data for the detection and classification of breathing diseases is rather uncommon compared to the use of medical imaging data. However, it is an emerging field of research, particularly with the enhanced capabilities provided by AI. Recent research demonstrates that the use of these models results in strong diagnostic performance.

Single-Modality Sound-Based

In this approach, the specialists capture and analyze lung sound data, including respiratory sounds such as abnormal breathing patterns, coughing, and gasping. Some studies convert recordings to spectrogram images for preprocessing and training the models [23], [103], or employing the numerical feature sequences of the audio file without converting them to images [81]. For example, Zhang *et al.* presented an approach to diagnose lung diseases from audio recordings collected by recording breathing cycles using a digital stethoscope, then four different models were implemented and evaluated on the extracted features [81]. The authors stated that the Long short-term memory (LSTM) model achieved the highest accuracy of 98.82%; this outperformed comparable models. Fraiwan *et al.* developed a two-stage detection framework that starts with CNN then adds Bidirectional Long Short-Term Memory units (BDLSTM) to detect pulmonary diseases from lung sounds [82]. The integration of the two models boosted efficiency to 99.62% accuracy when classifying the raw lung sound signals (waveforms) into multiple respiratory diseases. Borwankar *et al.* introduced a unique preprocessing method that uses augmentation with a delay function for sound recordings [103]. Their feature transformation approach merged Mel spectrogram features with other preprocessing techniques and fed these into a CNN, reaching 99.40% precision, with 99.1% sensitivity and 99.3% F1 score. In another study, Taloba and Matoog combined SVM and KNN classifiers using Mel-Frequency Cepstral Coefficients MFCC features extraction method to improve the detection of COPD with an accuracy up to 94% [104], demonstrating that there are different computational approaches capable to be used.

Generative augmentation approaches have been utilized to mitigate the issue of limited and imbalanced datasets. Saldanha *et al.* explored using Convolutional VAE, Multilayer Perceptron VAE, and Conditional VAE Variational Autoencoders generative models to synthesis lung sounds in a various class [98]. These generative models were employed to create realistic synthetic medical images to augment the training dataset and address class imbalance in ICBHI sound dataset especially in the underrepresented classes. The authors evaluated these models and found Convolutional VAE as the most effective. The models trained with augmented data showed improved diagnostic performance. Using spectrogram images of stethoscope sounds recordings can aid the early diagnosis of lung diseases. Demir *et al.* created a hybrid model that consist of a CNN and Parallel Pooling approach to extract deep features from spectrogram images of the ICBHI 2017 [23]. The authors employed a linear discriminant analysis (LDA) classifier and random subspace ensemble (RSE) to classify the images achieving 71% accuracy which was higher when compared to relevant methods on the same dataset, this model also scored 61% sensitivity, 86% specificity, 69% precision and 65% F1 score. Together, these studies display a strong trend toward using advanced feature extraction, data augmentation, and diverse model designs to improve the accuracy of lung disease detection from sound recordings.

Multimodal Sound-Based Approaches

Multimodal approaches combine sound data with other data sources such as medical images and clinical information. These methods often employ data balancing techniques and use advanced deep learning frameworks to fuses different data modalities to improve diagnostic performance of lung diseases. For instance, Malik *et al.* applied an important preprocessing method which is continuous wavelet transform to convert cough sound files to 2D scalogram images [7]. The authors designed DCDD_NET, trained and evaluated on 20 public datasets to classify nine types of lung diseases using cough spectrograms, X-rays, and CT scans. To balance the discrepancies and imbalances among these data types, the Borderline- Synthetic Minority Oversampling Technique SMOTE algorithm was applied leading to a significant model performance. The proposed model achieved 96.67%, high classification accuracy, 95.76% sensitivity, 96.82% precision, 95.61% F1 score, and 99.43% AUC. Digital lung auscultation has been the subject of investigations beside cough-based sounds studies. As an example, Siebert *et al.* created DeepBreath model, integrating data from lung recorded sounds of 120 patients, with LUS images and clinical information for the automatic diagnosis and categorization of chronic obstructive disease COPD and interstitial lung disease (ILD) [105]. The model trains on these various data employing long short-term memory LSTM, CNN, and transforms to model could separate disease subtypes and predict severity achieving greater than 80 area under the ROC curve. Figure 5 demonstrates the utilization of multiple data modalities including lung sounds recording, X-rays and CT scans in the classification of lung diseases. These findings suggest that multimodal approaches incorporating sound-based data are promising for accurately diagnosing pulmonary disease.

DATA COLLECTION AND PREPARATION

Datasets are the fundamental resource that powers AI and deep learning models. The quality, volume, diversity, and characteristics of the dataset are essential for building, training and refining AI models to address feasibility, accuracy and reliability. These models rely on the data for training first, then to be capable of identifying patterns, segmenting area of interests, and finally generating predictions. The majority of data used in lung diseases detection and classification are represented by medical images which is fundamental for AI models to function [58]. Labeling datasets and creating ground truth images are a highly important step for enhancing datasets in supervised-learning based methods. Practically, wider variety of data sources may increase the probability of finding hidden patterns and correlations. The lack of coherent and well labeled datasets is one of the main challenges of AI-based medical diagnoses [106]. Studies usually include methods of collection, preparation, and preprocessing as the main data centric procedures [72]. Data aggregation using unified methods and formats in addition to create a global centralized data repositories is significant to develop faster and more accurate models [107].

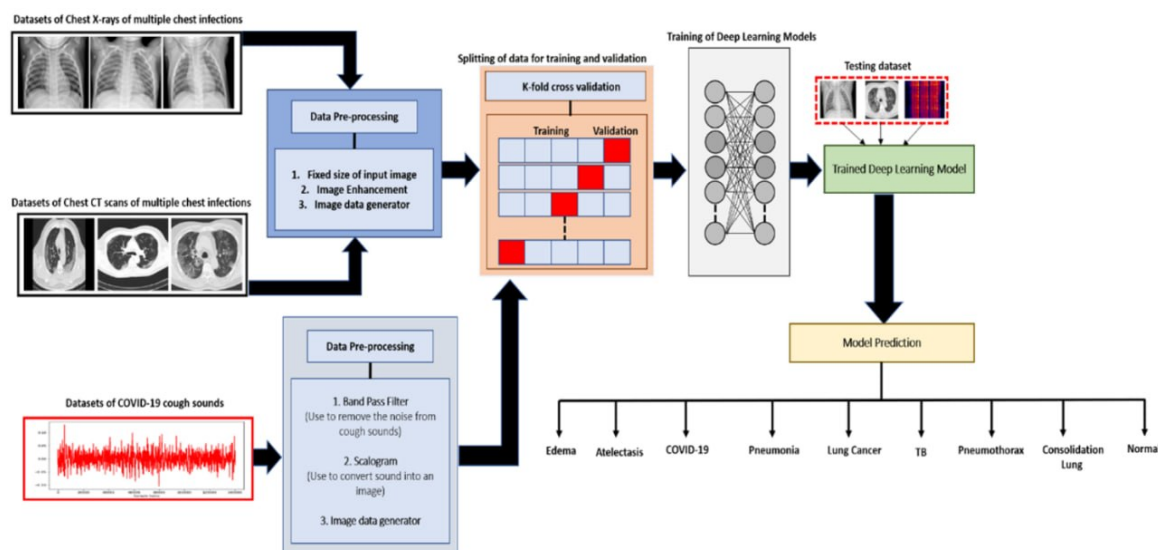


Figure 5. Lung cough sounds, X-rays and CT scans as an input for the deep learning classification model, adopted from Malik *et al.* [7]

Data Collection

Medical imaging devices are used to generate visual representations of internal organs, making use of technologies such as radiographic imaging (X-ray), computed tomography (CT), lung ultrasound (LUS), whole-slide imaging (WSI), positron emission tomography (PET), and chest MRI [85], [108]. An alternative approach that has emerged recently entails converting lung sounds recordings by a digital stethoscope into images for further image-based diagnostic analysis [7], [81]. These various imaging data feed computer-aided diagnosis CAD systems, which function as key tools for radiologists. By supporting image interpretation, AI models in CAD systems can improve diagnostic accuracy and reduce the time required for a diagnosis [107]. AI diagnostic models usually perform better when data from different imaging sources are used simultaneously. For instance, combining both chest X-rays with CT scans has proven multi-classification performance improvements, due to the capability of CT scans to pinpoint abnormalities at earlier stages than X-rays and more identify features in more precise manner [46]. Such use of data improves a model's ability to differentiate irregular patterns [5]. The development of the current AI diagnostic models relies on the access to large datasets, thus creating critical ethical considerations regarding data usage, privacy, and the importance of participants informed consent. Maintaining participants' trust is significant and can be maintained by informing the patients with their rights, including the choice of removing their data from research datasets [109].

Data Preprocessing

Real-world data usually contains errors, missing values, duplicates, noise, etc. Data preprocessing is a critical step that aims to deal with common raw data issues and transforms them into a clean, structured format suitable for model training [88]. Furthermore, well-structured clean data trains the AI model faster and more efficiently on computational resources [110], [111]. Key preprocessing techniques involve standardization, normalization, resizing, feature enhancement, noise removal, removing duplicates, handling missing values, correcting errors, smoothing, etc. [55], [58]. Applying these procedures boosts AI model performance and reliability in patterns identification and disease diagnosis and classification [108]. Studies confirm that preprocessing techniques like histogram equalization and adaptive masking improves accuracy among other metrics [112]. Segmentation alone can increase diagnostic accuracy by as much as 21% for COVID-19/pneumonia classification [59]. These principles extend beyond images to other modalities, such as sound-based datasets requiring wavelet smoothing and normalization [82]. On the other hand, handling medical data requires addressing significant privacy concerns. To prevent patient re-identification from metadata, solutions like differential privacy and federated learning are crucial for protecting patient identity [109].

Cropping and Resizing

Cropping and resizing are essential image preprocessing techniques that are employed to limit image dimensionality and to standardize medical images specifications before model training. Moreover, it's important to recognize the specific lung regions precisely to improve accuracy, as they represent the region of interest [31]. Different datasets have varying image properties and sizes; resizing techniques ensure uniform input dimension, while cropping reduces the volume by focusing only on the region of interest within the image [58], [83]. Removing irrelevant information from an image will consequently ensure accurate identification, decrease computational requirements, simplify training, and enhance performance [113].

Image Enhancement

Image enhancement techniques are a key preprocessing step in AI and DL models, specifically in image classification. These methods improve the performance, feature representation, accuracy, and efficiency of AI models by enhancing image visual quality. They underline prominent features and perform adjustments to make the images more suitable for analysis. Image enhancement includes improving sharpness, contrast, brightness, and removing noise [55]. Many techniques have been used to prepare the data for the model [73] such as noise reduction in using median filtering and Wiener filtering or non-local means filters, histogram equalization, and intensity adjustment [66], Contrast Limited Adaptive Histogram Equalization (CLAHE) [6], and denoising with Gaussian or Wiener filters [86].

Normalization

Normalization is an important preprocessing step that modifies the range of varying values to a standard scale. It is crucial for ensuring data standardization and consistency, which decreases variability, prevents features with large value ranges from overpowering the learning process, and ultimately enhances model performance and generalizability [114]. Some typical procedures applied to normalize image data include histogram equalization, min-max scaling, z-score normalization, and intensity clipping. Despite strengthened model outcomes with normalization, it can lead to a compromise between the benefits and the conservation of original data characteristics [58]. In contrast to normalization-based approaches, Acharya *et al.* presented and evaluated a normalization-free network NFNet to mitigate the issue of performance degradation associated with small, highly variable batch sizes, achieving 96.91% accuracy on a multi-class classification using the TBX11K dataset [3].

Augmentation

The generalizability of classification tasks has been identified as a significant gap [98], as the AI models focus on specific diseases, they typically require a large amount of data. Hence, medical data augmentation methods enhance the performance of AI models primarily by increasing the diversity of the dataset without annotating or collecting new data [113]. These methods are especially integral in cases of lack of data or inadequate labeling [115]. The principal objective of augmentation methods is to expand the dataset with synthetic data, which is mostly a transformed version of the existing data in every training epoch, this helps overcome class imbalance and reduce overfitting [46], [63]. Researchers often use several augmentation methods, these include contrast adjustment, rotation, mix up, mirroring, cropping, zooming, flipping, etc. Additionally, generative models such as VAEs Variational Autoencoders and Generative Adversarial Networks (GAN) are utilized to enhance the diversity and the quantity of dataset images as discussed in Section 3.3.8 [106]. Nevertheless, it has been observed that these augmentation methods may cause a loss of information in CT images. For instance, Salama and Aly reported that model performance was declined when three of these techniques were used on the same images [96].

Dataset Split and Class Balancing

For valid training and testing of AI models, data partitioning is essential especially in case when some classes are under-represented. It's a common approach to divide the dataset into three classic subsets: training, validation and test. The predominant training set is used to train the model in the first place. Secondly, the validation set typically comprises 10-15% of data which is used to tune the model parameters and prevent overfitting to the training dataset. Finally, the testing dataset

(also representing approximately 10–15% of the data) is employed to assess how well the model can perform. But in some cases, the researchers ignore the validation set and try to allocate 80% of the data for training and the remaining 20% for testing [51].

POPULAR DATASETS

This section presents the most popular lung diseases datasets that had been utilized by other representative works to train and evaluate AI models.

NIH Chest X-ray Dataset

The NIH dataset is a U.S. database containing chest x-ray images captured and prepared by a team of researchers at the U.S. National Institutes of Health NIH [116]. The NIH dataset is publicly available, comprising 112,120 images with 1024×1024 resolution for 32,717 patients with various lung diseases. The dataset was automatically annotated based on physicians' medical reports with an accuracy of over 90%. Currently, it is considered the largest in the field of lung diseases, covering 14 disease classes [116].

Tuberculosis Chest X-rays Dataset (Shenzhen set)

This dataset is compiled by the American National Library of Medicine (NLM) in association with the Third People's Hospital of Shenzhen City and the Guangdong Medical College in China [117]. The dataset is a collection of 662 annotated frontal X-ray images with 3000×3000 resolution that comprise two cases, normal and abnormal (tuberculosis). The dataset population comprises 69% male and 31% female, and it is available publicly on the NLM/NIH website and Kaggle [118].

MIMIC-CXR Dataset

Medical Information Mart for Intensive Care (MIMIC-CXR) dataset is a large collection of chest x-ray images and their associated radiology reports. It has 371,920 images from 65,079 patients. The images and reports are de-identified for privacy [119]. A group of scientific institutions led by the Institute of Medical Engineering & Science, Massachusetts, designed this dataset to aid advances in natural language processing, clinical data mining in addition to medical imaging. This dataset was made available through PhysioNet [119], though a request to access the data should be submitted after completing a course in "human research and data privacy". The images were made available in their original DICOM format. Later, Version 2 of this dataset, which was released in 2024, has created a parallel collection in JPEG format with metadata containing labels from the radiology reports.

ICBHI Dataset

The ICBHI dataset implies a lung sound recordings collection that was prepared to support the scientific ICBHI 2017 challenge and submitted to the International Conference on Biomedical Health Informatics, which was held in Thessaloniki, Greece. The database was published at the conference proceedings of IFMBE in 2018 [120]. This dataset is primarily used for research in the field of respiratory sound analysis, aiming to improve the detection and classification of lung diseases, which are crucial for diagnosing respiratory conditions. The dataset consists of 6,898 recordings from 126 patients of all age groups in 920 annotated audio samples. This dataset was made publicly available on the official challenge webpage [120].

Tuberculosis Chest X-rays Dataset (Montgomery set)

Referred to as the Montgomery County chest X-ray set (MC), it is another dataset that was created by the American National Library of Medicine in collaboration with the Department of Health and Human Services, Montgomery County, Maryland, USA [117]. This dataset contains 138 labeled chest X-rays of which 58 exhibited signs of tuberculosis, and the other 80 are normal. The dataset demographics were 46% male and 54% female and available publicly on the NLM/NIH website and Kaggle [118].

LIDC-IDRI Dataset

The Lung Image Database Consortium image collection and Image Database Resource Initiative (LIDC-IDRI) initiated by the American National Cancer Institute (NCI) resulted from collaboration among eight medical imaging corporations and seven academic centers. It comprises 1,018 cases with 244,527 images in DX, CT, and CR data types. It is the most extensive and widely used collection in the Cancer Imaging Archive TCIA project for cancer research, and it is accessible to the public with annotations through the Cancer Imaging Archive (TCIA) website. The TCIA project was launched in 2011, and it is administered by the University of Arkansas for Medical Sciences and funded by the American National Cancer Institute (NCI) [121]. The archive hosts over 50 million images from 50,000 subjects, organized into 175 cancer collections [122]. The LUNA16 challenge is a subset of the LIDC-IDRI dataset. This subset includes 888 CT scans with 1,186 annotated nodules, which is a subset of LIDC that has been used for the automatic detection of pulmonary nodules in lung CT images [123]. It is available publicly on the challenge website and Kaggle.

CheXpert Dataset

The CheXpert Dataset is considered one of the largest datasets. It contains 224,316 chest images from 65,240 patients that were prepared and collected from Stanford Hospital by a team from Stanford University/USA. The images were labeled to include the 14 lung pathology classes, which indicate the positive, negative, or uncertain presence of a disease. These labels were annotated by an algorithm (automated labeler) that extracts observations from radiology reports and then voted on by three board-certified radiologists to set the ground truth [124]. Technically, the image resolution is 4000×4000 pixels, with another smaller version of 320×320 . This dataset is available at the Stanford AIMI website.

Chest X-Ray Images Dataset

This dataset comprises 5,232 chest X-ray images and includes both viral and bacterial pneumonia. These images were collected and labelled for patients aged one to five years old from Guangzhou Women and Children's Medical Centre, Guangzhou, China [125]. The authors made it available on the Mendeley Data website. In subsequent updates, it has expanded to include more samples; currently it has 5,863 images (1,583 normal, 4,280 pneumonia).

Mendeley Chest X-Ray Image Dataset

The Chest X-ray Image is a relatively small dataset that offers 4,350 high-quality chest X-ray images. These images were collected from various hospitals in Bangladesh and credited to Jagannath University/Bangladesh. It is categorized into four distinct classes: normal, lung opacity, COVID-19, and viral pneumonia, for this reason, it is more suitable for multi-classification purposes. It's available publicly on the Mendeley data website [126].

COVID-19 Radiography Dataset

A database that is collected from various public datasets and then prepared by a team of researchers [127]. Different sub databases were utilized to produce one unified database. It contains 1579 normal, 423 COVID-19, and 1485 viral pneumonia, a total of 3,487 chest X-ray images. This database was updated over time to include another class (lung opacity image and lung masks) and increased the number of images to 33,920. The update was published on the dataset page on Kaggle, where it's available publicly online [128].

Tuberculosis X-ray Dataset

Tuberculosis X-ray Dataset (TBX) is a large-scale dataset consisting of 11,200 x-ray images with bounding box annotations. This dataset is considered balanced, as it contains 5,600 normal and 5,600 TB-positive samples. Images were 512×512 pixels in size, and covered five classes: normal, unhealthy but non-TB, active TB, latent TB, and uncertain TB. The dataset is publicly accessible on Kaggle [129].

COUGHVID Dataset

The Embedded Systems Laboratory (ESL), Switzerland, and the European Commission sponsored the creation of COUGHVID dataset, the largest crowdsourced cough dataset. It comprises more than 30,000 cough recordings in the third update of this dataset, with more than 2,000 recordings labeled by experienced physicians. This dataset includes diverse demographics and has considered age, gender, location, COVID-19 status and respiratory condition factors. The dataset and its related labels are publicly accessible on Zenodo [130].

RSNA Dataset

RSNA Dataset is an enormous dataset of pulmonary embolism (PE) CT scans. It also includes annotations by a large group of thoracic radiologist experts. This dataset was created by the Radiological Society of North America (RSNA) and consists of 2,995,147 images from 12,195 patients, covering Acute PE, Acute and chronic PE, and Chronic PE classes. RSNA de-identified images and annotations are publicly accessible on Kaggle [131].

NLST Dataset

The National Lung Screening Trial (NLST) is a comprehensive, publicly available medical imaging dataset acquired from the National Lung Screening study, which was a randomized supervised trial carried out by the National Cancer Institute (NCI) in the United States. It includes more than 200,000 CT scans from 53,500 participants. The dataset can be reached with restrictions through the NCI's Cancer Data Access System (CDAS) [132]. Table 3 shows a summary of the above-mentioned datasets.

Table 3. Summary of the most common datasets that had been utilized by other representative works to train and evaluate AI models

Dataset	No. of samples	Type	Specialty	No. of Classes
NIH Chest X-ray dataset (ChestX-ray)	112,120	X-ray	Various	14
Tuberculosis Chest X-rays (Shenzhen No.3)	662	X-ray	Tuberculosis	2
MIMIC-CXR	371,920	X-ray	Various	14
ICBHI2017 (sound dataset)	6,898	Sound recording	Various	8
Tuberculosis (Montgomery)	138	X-ray	Tuberculosis	2
LIDC-IDRI	244,527	CT scans	Lung Cancer	4
CheXpert	224,316	X-rays	Various	14
Chest X-Ray Images (Pneumonia)	5,232	X-ray	Pneumonia	3
Mendeley Chest X-Ray Image	4,350	X-ray	Various	4
Covid-19 Radiography Database	3,487	X-ray	COVID-19	4
Tuberculosis X-ray (TBX11 K)	11,200	X-ray	Tuberculosis	6
The COUGHVID dataset (sound dataset)	30,000	Sound recording	COVID-19	3
RSNA Pulmonary Embolism Detection	2,995,147	CT scans	pulmonary embolism	10
National Lung Screening Trial (NLST)	203,099	LDCT	Lung Cancer	2

CHALLENGES IN AI FOR LUNG DISEASE DETECTION

AI systems for detecting and classifying lung disease must cope with several challenges. Addressing these challenges is crucial for advancing the development and deployment of these systems. The identified challenges must be resolved before AI can be effectively implemented in real-world clinical settings.

Data-Related Challenges

The development of robust and generalizable models demands large datasets of, high-quality images, however several obstacles persist. One main obstacle is the dependency on large volumes of high-quality, annotated medical imaging data. The data related challenges can be attributed to: Data availability and quality, lack of standardization, data imbalance, generalizability and dataset shift, manual preprocessing.

Model Interpretability and Complexity

Deep learning models are computationally demanding and have a complex design, while offering limited transparency. This combination makes medical staff uncertain to adopt DL based medical systems for routine diagnostics.

Lack of Transparency

Deep learning models remain 'black boxes' to users because of their complex architecture and vast parameter sets, making their results hard to interpret. Without transparency, health professionals hesitate to adopt these models in real-world settings, as healthcare professionals cannot trace the reasoning behind predictions [36], [106]. Even though the Explainable AI (XAI) techniques like Grad-CAM and LIME highlight important reasoning, these techniques still do not explain why certain features are significant. Therefore, research continues to improve interpretability, hoping that this issue will boost trust and accountability in clinical settings.

Computational Demands

Training and deploying effective models, particularly 3D CNNs, need significant computational power and energy, which limits their potential use in real world clinical settings [34], [133]. Training these models raises its own challenges, issues like vanishing gradients can disrupt the training process and diminish model stability. Fine tuning CNN parameters to acquire optimal weights and biases also present a key optimization challenge. To overcome these challenges, several studies are integrating metaheuristic algorithms like Ebola optimization search algorithm (EOSA) with CNNs to improve parameter optimization and boosts classification accuracy [65].

Ethical, Accountability, and Regulatory Challenges Beyond Accuracy

Accuracy is a widely adopted evaluation metric, but it doesn't capture the full picture. Evaluations should include other key factors, such as computational efficiency. A model might achieve high accuracy but need a substantial amount of time to provide feedback, which may be impractical in a clinical environment. This becomes obvious in imbalanced datasets. This is typical in sound-based diagnosis, where achieving a high accuracy score can be misleading if the model is simply prioritizing the majority class [104]. So, evaluating models requires a comprehensive evaluation of a model to include a variety of metrics which will be discussed in the next sections [66], [134].

Robust Validation Techniques

Validating models properly helps establish their generalizability and avoiding overfitting to new data. Some typical ways are hold-out validation, which divides the Datasets for training, validation, and testing, or by applying cross-validation techniques. While a crucial aspect of internal validation involves balancing the need to retain critical information and preventing the model from merely memorizing the training examples [58]. Testing a model on an entirely separate dataset from a different organization is regarded as a crucial and definitive measure for assessing its generalizability [66].

Explainable AI (XAI)

Explainable AI (XAI) is very important for in the evaluation of medical models, it not only shows whether the predictions are accurate but it also explains the logic behind a model gives a particular prediction. This issue has been already discussed.

Algorithmic Bias

AI models built upon imbalanced or biased datasets might reinforce or worsen disparities in health care, considering their performance could vary among demographic groups [68]. To address bias, it is essential to employ a wide range of training data (i.e., from different sources), along with continuous evaluation for various effects in clinical applications.

Accountability and Consent

Utilizing patient data for training AI models presents a number of challenges associated with data privacy, data usage consent, and determining responsibilities in the event of diagnostic mistakes. These issues highlight the need for clear and transparent guidelines and legal regulations [28].

Patient Trust and Clinical Adoption

Generally, the success of AI systems when it comes to detecting lung diseases comes down to trust from both patients and medical staff. This is needed to encourage adoption and make sure AI tools are safe, effective, and unbiased. To achieve this, it's essential to ensure transparency, build awareness, and ensure active involvement from relevant parties. Table 4 shows the main challenges associated with lung diseases' identification and classification.

Table 4. The main challenges are associated with lung diseases identification and classification

Challenge Area	Core Issues	References
Data Availability/Quality	Scarcity of diverse, annotated data	[67], [72]
Standardization	Protocol/annotation inconsistency	[83]
Generalizability	Dataset shift, poor external validation	[65], [135]
Data Imbalance	Bias toward majority classes	[53], [136]
Interpretability	Black-box nature, clinician distrust	[36], [106]
Computational Demands	High resource requirements	[34]
Algorithmic Bias/Equity	Disparate impact, ethical concerns	[68]
Validation/Evaluation	Overfitting, misleading metrics	[86], [110]

Performance Evaluation Metrics and Challenges

A full evaluation is needed to assess the overall performance, effectiveness, and clinical applicability. To achieve this, multiple performance evaluation metrics are needed [11].

1 Performance Metrics Used for Classification Tasks

For classification tasks, various metrics derived from the fundamental concepts of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are employed [108]. This section presents the main metrics used in classification tasks. Accuracy is performance metric that indicates the ratio of correct predictions both true positive and true negative to the total number of predictions. The accuracy metric is calculated as in (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision as the measure of how many of the positive predictions were correct. This metric can measure superior precision by predicting “positive”, although it ignores false negatives (FN) which

may cause the model to miss many actual positives. Therefore, Precision is often combined with Recall for a balanced interpretation of a model performance. The Precision is calculated as in (2):

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall is also called Sensitivity or true positive rate, measures what portion of the actual positive cases were found correctly. In lung disease diagnosis, it is very important to achieve a high recall so that unhealthy cases are not overlooked. This metric is helpful in cases where false positives lead to high costs, for instance, performing unnecessary biopsies. The recall metric is calculated as in (3):

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Specificity or true negative rate, refers to how well the model can identify negative cases (e.g., healthy patients). Its use is critical in scenarios where preventing false alarms is the top priority, such as when screening populations with a low risk for a particular disease. The specificity metric is calculated as in (4):

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

The F1 score, which is a balanced ground between precision and recall, it helps to evaluate classification models, with its value ranging from 0 to 1. It's particularly useful when dealing with imbalanced classes because it takes into account both false positives and false negatives. The F1-Score metric is calculated as in (5):

$$F1 \text{ score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

The area under the receiver operating characteristic curve (ROC) shows how well the model separate classes by measuring the true positive rate changes compared to the false positive rate. Meanwhile, the AUC (Area Under the Curve) or AUROC gives us single number that captures this curve characteristic. The AUROC is the probability (P) that a model or classifier will give a randomly chosen positive example a higher score than a randomly chosen negative example.

CLINICAL SAFETY

Although many of the studies report certain model classification or segmentation performance using raw accuracy measure as the main or the only metric, evaluating medical artificial intelligence models solely on high accuracy is misleading [27], introduce serious limitations and critical safety risks. In medical settings, it's vital to employ other clinical safety metrics [25], especially sensitivity and specificity. Achieving high sensitivity is fundamental to ensure a low false-negative rate and mitigate missing a true positive case that demands urgent care. At the same time, high specificity measures are necessary to reduce false-positive rates, in that way guarding the patients from psychological concerns, unnecessary diagnostic obligations, and financial strain of unnecessary visitation. The issue of relying mainly on accuracy could be more severe when dealing with imbalanced datasets, a model can demonstrate high accuracy by simply prioritizing the majority of healthy samples and overlooking the few critical abnormalities [133].

Performance Metrics Used for Segmentation Tasks

This section presents the performance metrics used to evaluate the segmentation results [134]. The most used metrics are:

Dice Similarity Coefficient (DSC): A metric that evaluates the similarity between two sets, typically a predicted segmentation (A) and a ground-truth segmentation (B). It's calculated as twice the area of their intersection divided by the sum of their individual areas, emphasizing the degree of overlap.

Intersection over Union (IoU): This metric is a way to measure the overlap between a predicted segmentation or bounding box (A) and the actual ground-truth area (B).

Pixel Accuracy: This metric calculates the percentage of positively classified pixels (TP) in the output segmented images to the total number of pixels.

Model Validation Techniques

Researchers rely on key validation methods to check how well AI models for lung disease detection and segmentation perform in a variety of situations [102]. The main goal is to make sure these models are dependable and can also handle new, unfamiliar/unseen data.

Hold-Out Validation (Data Splitting): A very common approach is to split the dataset into smaller groups. The first part is the training set, which the model uses for learning patterns from the data. Then, there's a validation set to help adjust and fine-tune the model's settings. Finally, a separate test set is used to check how well the model performs on completely new data it has not seen before. Using an independent test set is crucial to avoid bias, particularly when working with limited data.

Cross-validation splits the dataset multiple times; in this way researchers can observe how the model performs on fresh data. By averaging these, the model's real predictive power is demonstrated, especially when dealing with limited data. K-fold cross-validation is a common technique, where the data is divided into 'K' equal folds, then train and test K different times, using a different fold for testing each round while the rest trains the model. Since every data point serves both roles, the performance estimate will be more accurate and fairer [11], [13]. Other Validation Methods include LOOCV which may be used with small datasets, though the computational demands can get intense, potentially making it impractical for real world applications [13]. While MCCV deal better with large datasets, providing solid performance indicators [11]. On the other hand, the Stratified K-Fold Cross-Validation is more suitable for imbalanced datasets (i.e., when classes are not equally represented), although it is not suitable for regression problems [57].

Cohen's kappa is a statistical measure that assesses the agreement between two classifiers, or what researchers call raters after removing random agreement. This metric indicates how well they correspond between predicted classifications and the manually marked ground truth [82]. While researchers mainly use this metric for classification, it can also evaluate the effectiveness of a segmentation model by comparing its pixel-level predictions with ground truth data.

External validation is the best way to test if a model can be generalized, by testing how it performs on a separate unused dataset. This step is important to reduce bias which is a frequent issue if performance is measured on data that is not fully independent [137]. So, the model doesn't learn specific issues from one dataset, like artifacts that come from imaging mistakes, researchers use external datasets to validate it [54].

This validation can be done in several ways, including temporal validation and geographic validation. By offering it gives a rigid method for testing AI model's stability and reliability. External validation helps make sure the metrics accurately reflect model's potential performance in actual real world clinical use.

IDENTIFICATION OF CRITICAL RESEARCH GAPS

Even with the widespread implementation of AI-driven tools in lung disease detection, analyzing the current literature reveals several critical research gaps that obstruct effective clinical translation:

Over-reliance on Static Datasets and Data Leakage

Most reviewed models are greatly dependent on the same most used public datasets (e.g., NIH ChestX-ray and MIMIC-CXR) evaluate performance. This establishes a key limitation which is the lack of independent external validation on diverse demographic groups. As a result, the models trained on these datasets continue to be at risk from dataset shift and undetected data leakage.

Superficial Explainability (XAI)

While techniques such as Grad-CAM are broadly adopted, they only produce visual heatmaps without offering any understanding into the underlying pathophysiological reasoning. A considerable gap persists in the development of inherently explainable models that translate AI-generated representations into standardized clinical classification systems

The Multimodal Fusion Deficit

Although multimodal methods offer evident theoretical benefits, genuine multimodal fusion is still notably underexplored. The effective incorporation of high-resolution medical images with unstructured electronic health records (EHRs) and temporal audio signals, and other types of data remains in

its early stages, with the majority of current systems continuing to operate in strict singular modality framework.

Computational Constraints in Resource-Limited Settings

Despite the considerable promise demonstrated by emerging foundation models and transformer-based architectures, their practical deployment is strictly constrained by high-performance computing requirements. A major gap persists in the development of optimized lightweight architectures, such as efficient U-Net variants for accurate segmentation, that can achieve state-of-the-art accuracy while functioning effectively within the limited hardware structure of local medical environments.

FUTURE PERSPECTIVES

Foundation Models & large vision-language models

In the last decade, foundation models have emerged as a novel class of artificial intelligence. These models are trained via self-supervised learning (SSL) on large scale and diverse labeled and unlabeled datasets. The extensive vast scale and the diverse training of these models enable them to be generalized and adopted in various domains and applications, including those involving the use on new or unseen data. These factors empower foundation models to excel in no-shot, or few shot learning capabilities. This allows the model to capture generalized representations of unseen target domains, without being trained on that type of data specifically, with the minimal reliance on fine-tuning labeled data. To effectively mitigate the lack of expert annotations in medical workflows, the adoption of large-scale medical vision backbones is becoming vital for bridging the scarcity of expert annotations gap. A prime example of this paradigm was distinctly illustrated by Gao *et al.* foundation model (LCTfound). This model was trained on a large 105,000 CT scans using denoising diffusion advanced self-training method. This model exhibits remarkable data efficiency by learning vast, generalized structural representations during pre-training. This model achieved outstanding diagnostic and segmentation accuracy when evaluated on rare clinical anomalies like pulmonary alveolar proteinosis (PAP) [20]. Another major example is the MedSam, a foundational architectural model introduced by Ma *et al.* and trained using over 1.5 million (image/mask), enabling efficient and accurate image segmentation performance across wide range of images including the medical [138].

In the same context, the generative capabilities of medical AI are increasingly expanding to the recent progress in large vision-language models (LVLMs). One impactful example is RoentGen, a foundational vision-language model created by Bluethgen *et al.* and developed to produce highly accurate, realistic and disease specific X-ray images directly from unstructured medical text reports [139]. To improve the model capabilities of generating high quality images based on complex medical reports, RoentGen was finetuned using Latent Diffusion Model (LDM) trained on the vast MIMIC-CXR dataset to reduce the gap between general domain natural language and complex medical jargon. The generation of these synthetic images play an important role especially when dealing with rare or underrepresented diseases. Therefor LVLMs contribute to enhanced performance, generalizability and explainability of AI systems.

Federated Learning

Many of the recent deep learning methodologies achieve optimal accuracy in lung disease diagnoses, the real-world application of this system in clinical settings faces some challenges such as patients' privacy [140] regulations (e.g., HIPAA, GDPR) , bias and data quality and scarcity, addressing these issues is essential for successful implementation [141]. To tackle these challenges, Federated Learning (FL) directly focuses on distributing the training process across various hospitals. These hospitals keep the patient's sensitive information locally on site, and the way they contribute to the model training is by sharing anonymous mathematical weights, not the images to a central server. In this way, federated learning architecture preserves strict data confidentiality, resolve data imbalance and improve global generalization of data. This approach was well presented through the work of Makkar and Santosh (SecureFed), which is a framework that uses advanced federated learning techniques to enable secure, privacy focused architecture that is optimized for detecting chest x-ray lung abnormalities using the collaborative multi-site approach [142]. SecureFed was designed to overcome common federated learning issues such as varying protocols and imaging equipment around different hospitals, by employing an improved secure aggregation protocol that amplifies consistent

performance across all involved institutions, and overall diagnostic unbiased accuracy, while keeping the local data of patients secure.

Lightweight and Efficient Architectures

The adoption of AI models that are inherently efficient and increasingly lighter and less complex is rapidly growing in the field of medical diagnosis. These models can be fine-tuned specifically for the implementation environment, while balancing effective variables such as model size and complexity, mobility, budget and computational demands, leading examples in this space are U-net, SqueezeNet, and EfficientNetV2. These compact, high efficiency models are ideal for edge devices, point of care applications, and mobile friendly computer aided diagnoses (CAD) systems that must preserve high diagnostic performance even in limited resources settings. This offers significant benefits especially for the patient's populations who are geographically remote or who have mobility constraints, providing them with reliable and timely medical care. For instance, Sanida & Dasygenis created a high performance, lightweight CNN system that proves it isn't necessary for large scale model to be more effective in point of care, and more compact portable, in constrained environment could be as robust, their proposed system achieved 98.56% diagnosing lung diseases, surpassing existing state of the art architecture [143]. A recent study proposed by Choi & Yang created a lightweight two-dimensional CNN (LW-2DCNN) adapted for high-performance lung disease classification using chest CT scans. Through architectural optimization of the model complexity – efficiency, LW-2DCNN demonstrated a classification accuracy of 99.80% across five lung diseases [21].

CONCLUSION

Building AI and DL models for lung disease detection and classification is an active area of research with many key benefits such as reducing human errors, maintaining consistent medical performance, early identifying of subtle patterns that might be missed by the human eye, leading to earlier intervention, provide precise measurements (e.g., tumor volume), much faster than manual analysis, alert experts to focus on complex cases, etc. Furthermore, these models can be deployed via mobile apps in remote areas that suffer from limited access to experts. Recent studies indicate that AI and DL models have excellent diagnoses and classification results. One major trajectory of research advancement in this field involves integrating hybrid and ensemble models. Researchers have tested combining different architectures using stacking or weighted ensemble methods, these integrated models - like RVCNet, ResNet, VGG, and a custom CNN demonstrated better functionality compared to a single model. This progress resulted in robust performance, with many models achieving 90-99% accuracy in multi disease classification of lung diseases. Additionally, research efforts are expanding to areas beyond image-based diagnostic models. Examining lung sounds through signal and multimodal sound-based frameworks, exhibit a potential for higher diagnostic and classification accuracy. Multimodal sound-based models achieve strong classification accuracy from audio recordings. Some studied proposed the integration of imaging modalities and sound bases data to boost disease diagnosing abilities. This literature survey highlighted more than 60 studies that cover recent progress in applying AI and deep learning for detecting and classifying lung diseases through advanced preprocessing like segmentation and normalization, combined with image enhancement techniques and robust ensemble architecture. However, wide medical applications still face major challenges such as limited databases, imbalanced data, wrong labels, interpretability issues, and generalizability problems. Although various innovative frameworks achieved diagnostic accuracy above 90%, the actual test is in obtaining stable outcomes in real clinical environments. Explainable AI becomes more and more important in building trusted models while making sure that these systems are accurate and make sense, with decisions that experts can interpret and visualize.

To shift from experimental research to vigorous actual clinical deployment, the focus of research must move from pursuing slight accuracy increases on static, overused datasets to addressing deep methodological gaps. Future progress must highlight the importance of robust external validation, the integration of multimodal data natively (combining radiomics, electronic health records, audio signals, with images), and the optimization of lite, computational - resource efficient architectures. By concentrating on these specific bottlenecks, AI can achieve its capability as a reliable, transparent, and globally accessible diagnostic aid partner in respiratory health care.

SUPPLEMENTARY MATERIAL

None.

AUTHOR CONTRIBUTIONS

Harith S. Hassan: Conceptualization, Investigation, and Writing - original draft, review & editing. Sinan A. Naji: Supervision, Formal analysis, and Validation. Abdullah G. Jaber: Investigation, visualization.

FUNDING

This research received no external funding.

DATA AVAILABILITY STATEMENT

Authors declare that no new data were created.

ACKNOWLEDGMENTS

The authors would like to thank University of Information Technology and Communications for its support in the present work.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DECLARATION OF GENERATIVE AI USE

During the preparation of this manuscript, the authors used Quillbot online language editing (quillbot.com) for grammar checking and language polishing. After using this tool, the authors thoroughly reviewed and revised the generated content, taking full responsibility for the accuracy and integrity of the final manuscript.

REFERENCES

- [1] L. Choridah, R. Rulaningtyas, L. Muqmiroh, S. Suprayitno, and K. Ain, "Detection of lung disease using relative reconstruction method in electrical impedance tomography system," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 4, pp. 2136–2145, 2023, doi: [10.11591/beej.v12i4.4940](https://doi.org/10.11591/beej.v12i4.4940).
- [2] H. Y. Riskiawan, T. Rizaldi, D. P. S. Setyohadi, and M. M. D. Utami, "Modelling expert system for lung disease," *Journal of Physics: Conference Series*, vol. 1569, no. 2, Art no. 022011, Jul. 2020, doi: [10.1088/1742-6596/1569/2/022011](https://doi.org/10.1088/1742-6596/1569/2/022011).
- [3] V. Acharya, G. Dhiman, K. Prakasha, P. Bahadur, A. Choraria, S. M, S. J, S. Prabhu, K. Chadaga, W. Viriyasitavat, et al., "AI-Assisted tuberculosis detection and classification from chest X-rays using a deep learning normalization-free network model," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, Art no. 2399428, 2022, doi: [10.1155/2022/2399428](https://doi.org/10.1155/2022/2399428).
- [4] U. Chutia, A. S. Tewari, J. P. Singh, and V. K. Raj, "Classification of lung diseases using an attention-based modified Densenet model," *Journal of Imaging Informatics in Medicine*, vol. 37, no. 4, pp. 1625–1641, 2024, doi: [10.1007/s10278-024-01005-0](https://doi.org/10.1007/s10278-024-01005-0).
- [5] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," *Alexandria Engineering Journal*, vol. 64, pp. 923–935, Feb. 2023, doi: [10.1016/j.aej.2022.10.053](https://doi.org/10.1016/j.aej.2022.10.053).
- [6] A. Moussaid, N. Zrira, I. Benmiloud, Z. Farahat, Y. Karmoun, Y. Benzidia, S. Mouline, B. El Abdi, J. E. Bourkadi, and N. Ngote, "On the implementation of a post-pandemic deep learning algorithm based on a hybrid CT-scan/X-ray images classification applied to pneumonia categories," *Healthcare*, vol. 11, no. 5, Art no. 662, 2023, doi: [10.3390/healthcare11050662](https://doi.org/10.3390/healthcare11050662).
- [7] H. Malik, T. Anees, A. S. Al-Shamayleh, S. Z. Alharthi, W. Khalil, and A. Akhunzada, "Deep learning-based classification of chest diseases using X-rays, CT scans, and cough sound images," *Diagnostics*, vol. 13, no. 17, Art no. 2772, 2023, doi: [10.3390/diagnostics13172772](https://doi.org/10.3390/diagnostics13172772).

- [8] C. C. Ukwuoma, Z. Qin, M. B. B. Heyat, F. Akhtar, A. Smahi, J. K. Jackson, S. Furqan Qadri, A. Y. Muaad, H. N. Monday, and G. U. Nneji, "Automated lung-related pneumonia and COVID-19 detection based on novel feature extraction framework and vision transformer approaches using chest X-ray images," *Bioengineering*, vol. 9, no. 11, Art no. 709, 2022, doi: [10.3390/bioengineering9110709](https://doi.org/10.3390/bioengineering9110709).
- [9] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Informatics in Medicine Unlocked*, vol. 20, Art no. 100391, 2020, doi: [10.1016/j.imu.2020.100391](https://doi.org/10.1016/j.imu.2020.100391).
- [10] I. Chouvarda, E. Perantoni, and P. Steiropoulos, "Respiratory decision support systems," in *Wearable Sensing and Intelligent Data Analysis for Respiratory Management*. Elsevier, 2022, pp. 299–322, doi: [10.1016/b978-0-12-823447-1.00008-7](https://doi.org/10.1016/b978-0-12-823447-1.00008-7).
- [11] Y. Al-Issa, A. M. Alqudah, H. Alquran, and A. A. Issa, "Pulmonary diseases decision support system using deep learning approach," *Computers, Materials and Continua*, vol. 73, no. 1, pp. 311–326, 2022, doi: [10.32604/cmc.2022.025750](https://doi.org/10.32604/cmc.2022.025750).
- [12] A. Novak, S. Ather, A. Gill, P. Aylward, G. Maskell, G. W. Cowell, A. T. Espinosa Morgado, T. Duggan, M. Keevill, O. Gamble, et al., "Evaluation of the impact of artificial intelligence-assisted image interpretation on the diagnostic performance of clinicians in identifying pneumothoraces on plain chest X-ray: A multi-case multi-reader study," *Emergency Medicine Journal*, vol. 41, no. 10, pp. 602–609, 2024, doi: [10.1136/emmermed-2023-213620](https://doi.org/10.1136/emmermed-2023-213620).
- [13] J. Antão, J. de Mast, A. Marques, F. M. Franssen, M. A. Spruit, and Q. Deng, "Demystification of artificial intelligence for respiratory clinicians managing patients with obstructive lung diseases," *Expert Review of Respiratory Medicine*, vol. 17, no. 12, pp. 1207–1219, 2023, doi: [10.1080/17476348.2024.2302940](https://doi.org/10.1080/17476348.2024.2302940).
- [14] A. S. Sitanggang, A. D. Damarullah, and Wartika, "Analysis of expert system lung disease diagnosis system of web-based disease in cihaur puskesmas," *IOP Conference Series: Materials Science and Engineering*, vol. 879, no. 1, Art no. 012065, 2020, doi: [10.1088/1757-899x/879/1/012065](https://doi.org/10.1088/1757-899x/879/1/012065).
- [15] S. Yazdani, C. Lerner, D. Kulkarni, A. Kamzan, and R. C. Henry, "A new expert system with diagnostic accuracy for pediatric upper respiratory conditions," *Healthcare Analytics*, vol. 2, Art no. 100042, Nov. 2022, doi: [10.1016/j.health.2022.100042](https://doi.org/10.1016/j.health.2022.100042).
- [16] M. Y. Santoso, A. M. Disrinama, and H. N. Amrullah, "An application of ANFIS for lung diseases early detection system," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 5, no. 1, pp. 29–36, 2020, doi: [10.22219/kinetik.v5i1.996](https://doi.org/10.22219/kinetik.v5i1.996).
- [17] D. Çelik Ertuğrul and D. Celik Ulusoy, "A knowledge-based self-pre-diagnosis system to predict COVID-19 in smartphone users using personal data and observed symptoms," *Expert Systems*, vol. 39, no. 3, Art no. e12716, 2021, doi: [10.1111/exsy.12716](https://doi.org/10.1111/exsy.12716).
- [18] U. Subramaniam, M. M. Subashini, D. Almakhlles, A. Karthick, and S. Manoharan, "An expert system for COVID-19 infection tracking in lungs using image processing and deep learning techniques," *BioMed Research International*, vol. 2021, no. 1, Art no. 1896762, 2021, doi: [10.1155/2021/1896762](https://doi.org/10.1155/2021/1896762).
- [19] Z. Zulkifli, R. A. Soeprihatini, S. Sfenrianto, Z. Wiyanti, P. Bintoro, F. Fitriana, S. Sukarni, N. A. Putri, and D. Y. A. Andini, "Expert system for diagnosis of lung disease from X-ray using CNN and SVM," *International Journal of Artificial Intelligence Research*, vol. 7, no. 2, Art no. 98, 2023, doi: [10.29099/ijair.v7i1.870](https://doi.org/10.29099/ijair.v7i1.870).
- [20] Z. Gao, G. Zhang, H. Liang, J. Liu, L. Ma, T. Wang, Y. Guo, Y. Chen, Z. Yan, X. Chen, et al., "A lung CT vision foundation model facilitating disease diagnosis and medical imaging," *Nature Communications*, vol. 17, no. 1, Art no. 35, 2025, doi: [10.1038/s41467-025-66620-z](https://doi.org/10.1038/s41467-025-66620-z).
- [21] H.-S. Choi and J. Yang, "High-Performance lung disease identification and explanation using a reciprocally-enhanced lightweight convolutional neural network," *IEEE Access*, vol. 13, pp. 84954–84965, May 2025, doi: [10.1109/access.2025.3568463](https://doi.org/10.1109/access.2025.3568463).
- [22] M. O. Oyediran, O. S. Ojo, I. A. Raji, A. E. Adeniyi, and O. J. Aroba, "An optimized support vector machine for lung cancer classification system," *Frontiers in Oncology*, vol. 14, Art no. 1408199, Dec. 2024, doi: [10.3389/fonc.2024.1408199](https://doi.org/10.3389/fonc.2024.1408199).
- [23] F. Demir, A. M. Ismael, and A. Sengur, "Classification of lung sounds with CNN model using parallel pooling structure," *IEEE Access*, vol. 8, pp. 105376–105383, Jun. 2020, doi: [10.1109/access.2020.3000111](https://doi.org/10.1109/access.2020.3000111).
- [24] C.-H. Cheng, H.-H. Chen, and T.-L. Chen, "A clinical decision-support system based on three-stage integrated image analysis for diagnosing lung disease," *Symmetry*, vol. 12, no. 3, Art no. 386, 2020, doi: [10.3390/sym12030386](https://doi.org/10.3390/sym12030386).
- [25] D. Kvak, A. Chromcová, R. Hrubý, E. Janů, M. Bíroš, M. Pajdaković, K. Kvaková, M. A. Al-antari, P. Polášková, and S. Strukov, "Leveraging deep learning decision-support system in specialized oncology center: A multi-reader retrospective study on detection of pulmonary lesions in chest x-ray images," *Diagnostics*, vol. 13, no. 6, Art no. 1043, 2023, doi: [10.3390/diagnostics13061043](https://doi.org/10.3390/diagnostics13061043).

- [26] Y. Miyachi, O. Ishii, and K. Torigoe, "Design, implementation, and evaluation of the computer-aided clinical decision support system based on learning-to-rank: Collaboration between physicians and machine learning in the differential diagnosis process," *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, Art no. 26, 2023, doi: [10.1186/s12911-023-02123-5](https://doi.org/10.1186/s12911-023-02123-5).
- [27] K. Chadaga, S. Prabhu, V. Bhat, N. Sampathila, S. Umakanth, and R. Chadaga, "A decision support system for diagnosis of COVID-19 from non-covid-19 influenza-like illness using explainable artificial intelligence," *Bioengineering*, vol. 10, no. 4, Art no. 439, 2023, doi: [10.3390/bioengineering10040439](https://doi.org/10.3390/bioengineering10040439).
- [28] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: A multidisciplinary perspective," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, Art no. 310, 2020, doi: [10.1186/s12911-020-01332-6](https://doi.org/10.1186/s12911-020-01332-6).
- [29] C. Liu, R. Zhao, W. Xie, and M. Pang, "Pathological lung segmentation based on random forest combined with deep model and multi-scale superpixels," *Neural Processing Letters*, vol. 52, no. 2, pp. 1631–1649, 2020, doi: [10.1007/s11063-020-10330-8](https://doi.org/10.1007/s11063-020-10330-8).
- [30] S. S. Nair, V. M. Devi, and S. Bhasi, "Enhanced lung cancer detection: Integrating improved random walker segmentation with artificial neural network and random forest classifier," *Heliyon*, vol. 10, no. 7, Art no. e29032, 2024, doi: [10.1016/j.heliyon.2024.e29032](https://doi.org/10.1016/j.heliyon.2024.e29032).
- [31] G. Abdikerimova, A. Shekerbek, M. Tulenbayev, S. Beglerova, E. Zakharevich, G. Bekmagambetova, Z. Manbetova, and M. Baibulova, "Detection of chest pathologies using autocorrelation functions," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, pp. 4526–4534, 2023, doi: [10.11591/ijece.v13i4.pp4526-4534](https://doi.org/10.11591/ijece.v13i4.pp4526-4534).
- [32] E. P. Medeiros, M. R. Machado, E. D. G. de Freitas, D. S. da Silva, and R. W. R. de Souza, "Applications of machine learning algorithms to support COVID-19 diagnosis using X-rays data information," *Expert Systems with Applications*, vol. 238, Art no. 122029, Mar. 2024, doi: [10.1016/j.eswa.2023.122029](https://doi.org/10.1016/j.eswa.2023.122029).
- [33] B. T. Chen, Z. Chen, N. Ye, I. Mambetsariev, J. Fricke, E. Daniel, G. Wang, C. W. Wong, R. C. Rockne, R. R. Colen, et al., "Differentiating peripherally-located small cell lung cancer from non-small cell lung cancer using a CT radiomic approach," *Frontiers in Oncology*, vol. 10, Art no. 593, Apr. 2020, doi: [10.3389/fonc.2020.00593](https://doi.org/10.3389/fonc.2020.00593).
- [34] D.-E.-M. Nisar, R. Amin, N.-U.-H. Shah, M. A. A. Ghamdi, S. H. Almotiri, and M. Alruily, "Healthcare techniques through deep learning: Issues, challenges and opportunities," *IEEE Access*, vol. 9, pp. 98 523–98 541, Jul. 2021, doi: [10.1109/access.2021.3095312](https://doi.org/10.1109/access.2021.3095312).
- [35] J.-X. Wu, P.-Y. Chen, C.-M. Li, Y.-C. Kuo, N.-S. Pai, and C.-H. Lin, "Multilayer fractional-order machine vision classifier for rapid typical lung diseases screening on digital chest X-ray images," *IEEE Access*, vol. 8, pp. 105 886–105 902, Jun. 2020, doi: [10.1109/access.2020.3000186](https://doi.org/10.1109/access.2020.3000186).
- [36] C. Wang, Y. Long, W. Li, W. Dai, S. Xie, Y. Liu, Y. Zhang, M. Liu, Y. Tian, Q. Li, et al., "Exploratory study on classification of lung cancer subtypes through a combined k-nearest neighbor classifier in breathomics," *Scientific Reports*, vol. 10, no. 1, Art no. 5880, 2020, doi: [10.1038/s41598-020-62803-4](https://doi.org/10.1038/s41598-020-62803-4).
- [37] J. W. Kocks, H. Cao, B. Holzhauser, A. Kaplan, J. M. FitzGerald, K. Kostikas, D. Price, H. K. Reddel, I. Tsiligianni, C. F. Vogelmeier, et al., "Diagnostic performance of a machine learning algorithm (asthma/chronic obstructive pulmonary disease [COPD] differentiation classification) tool versus primary care physicians and pulmonologists in asthma, COPD, and asthma/COPD overlap," *The Journal of Allergy and Clinical Immunology: In Practice*, vol. 11, no. 5, pp. 1463–1474.e3, 2023, doi: [10.1016/j.jaip.2023.01.017](https://doi.org/10.1016/j.jaip.2023.01.017).
- [38] S. Siordia-Millán, S. Torres-Ramos, R. A. Salido-Ruiz, D. Hernández-Gordillo, T. Pérez-Gutiérrez, and I. Román-Godínez, "Pneumonia and pulmonary thromboembolism classification using electronic health records," *Diagnostics*, vol. 12, no. 10, Art no. 2536, 2022, doi: [10.3390/diagnostics12102536](https://doi.org/10.3390/diagnostics12102536).
- [39] J. Li, Y. Wang, Q. Sheng, X. Liu, Z. Xing, F. Sun, Y. Wang, S. Li, Y. Li, Y. Yu, et al., "Interpretable modeling and discovery of key predictors for pneumonia diagnosis in children based on electronic medical records," *DIGITAL HEALTH*, vol. 8, Art no. 205520762211311, Jan. 2022, doi: [10.1177/20552076221131185](https://doi.org/10.1177/20552076221131185).
- [40] M. Sherafatian and F. Arjmand, "Decision tree-based classifiers for lung cancer diagnosis and subtyping using TCGA miRNA expression data," *Oncology Letters*, vol. 18, no. 2, pp. 2125–2131, 2019, doi: [10.3892/ol.2019.10462](https://doi.org/10.3892/ol.2019.10462).
- [41] M. Atzeni, G. Cappon, J. K. Quint, F. Kelly, B. Barratt, and M. Vettoretti, "A machine learning framework for short-term prediction of chronic obstructive pulmonary disease exacerbations using personal air quality monitors and lifestyle data," *Scientific Reports*, vol. 15, no. 1, Art no. 2385, 2025, doi: [10.1038/s41598-024-85089-2](https://doi.org/10.1038/s41598-024-85089-2).
- [42] X. Yang, Y. Li, L. Liu, and Z. Zang, "Prediction of respiratory diseases based on random forest model," *Frontiers in Public Health*, vol. 13, Art no. 1537238, Feb. 2025, doi: [10.3389/fpubh.2025.1537238](https://doi.org/10.3389/fpubh.2025.1537238).

- [43] S. Kana Saputra, I. Taufik, M. Hidayat, and D. F. Dharma, "Pneumonia identification based on lung texture analysis using modified k-nearest neighbour," *Journal of Physics: Conference Series*, vol. 2193, no. 1, Art no. 012070, 2022, doi: [10.1088/1742-6596/2193/1/012070](https://doi.org/10.1088/1742-6596/2193/1/012070).
- [44] N. Ghaffar Nia, E. Kaplanoglu, and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," *Discover Artificial Intelligence*, vol. 3, no. 1, Art no. 5, 2023, doi: [10.1007/s44163-023-00049-5](https://doi.org/10.1007/s44163-023-00049-5).
- [45] D. Yang, C. Martinez, L. Visuña, H. Khandhar, C. Bhatt, and J. Carretero, "Detection and analysis of COVID-19 in medical images using deep learning techniques," *Scientific Reports*, vol. 11, no. 1, Art no. 19638, 2021, doi: [10.1038/s41598-021-99015-3](https://doi.org/10.1038/s41598-021-99015-3).
- [46] D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, "Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases," *Computers in Biology and Medicine*, vol. 132, Art no. 104348, May 2021, doi: [10.1016/j.compbiomed.2021.104348](https://doi.org/10.1016/j.compbiomed.2021.104348).
- [47] J. Lyu, X. Bi, and S. H. Ling, "Multi-Level cross residual network for lung nodule classification," *Sensors*, vol. 20, no. 10, Art no. 2837, 2020, doi: [10.3390/s20102837](https://doi.org/10.3390/s20102837).
- [48] A. T. Abdulahi, R. O. Ogundokun, A. R. Adenike, M. A. Shah, and Y. K. Ahmed, "PulmoNet: A novel deep learning based pulmonary diseases detection model," *BMC Medical Imaging*, vol. 24, no. 1, Art no. 51, 2024, doi: [10.1186/s12880-024-01227-2](https://doi.org/10.1186/s12880-024-01227-2).
- [49] S. De, M. Lederer, Y. Raffel, D. Lehninger, S. Thunder, M. P. Jank, T. Ali, and T. Kaempfe, "Monolithic-3D inference engine with IGZO based ferroelectric thin film transistor synapses," *TechRxiv*, Aug. 2022, doi: [10.36227/techrxiv.2051813.1.v1](https://doi.org/10.36227/techrxiv.2051813.1.v1).
- [50] M. Jawahar, J. A. L. V. Ravi, J. Prassanna, S. G. Jasmine, R. Manikandan, R. Sekaran, and S. Kannan, "CovMnet-Deep learning model for classifying coronavirus (COVID-19)," *Health and Technology*, vol. 12, no. 5, pp. 1009–1024, 2022, doi: [10.1007/s12553-022-00688-1](https://doi.org/10.1007/s12553-022-00688-1).
- [51] N. Ullah, M. Marzougui, I. Ahmad, and S. A. Chelloug, "DeepLungNet: An effective DL-based approach for lung disease classification using cris," *Electronics*, vol. 12, no. 8, Art no. 1860, 2023, doi: [10.3390/electronics12081860](https://doi.org/10.3390/electronics12081860).
- [52] K. Sriporn, C.-F. Tsai, C.-E. Tsai, and P. Wang, "Analyzing lung disease using highly effective deep learning techniques," *Healthcare*, vol. 8, no. 2, Art no. 107, 2020, doi: [10.3390/healthcare8020107](https://doi.org/10.3390/healthcare8020107).
- [53] G. Kasinathan and S. Jayakumar, "Cloud-based lung tumor detection and stage classification using deep learning techniques," *BioMed Research International*, vol. 2022, no. 1, Art no. 4185835, 2022, doi: [10.1155/2022/4185835](https://doi.org/10.1155/2022/4185835).
- [54] M. Zak and A. Krzyżak, "Classification of lung diseases using deep learning models," in *Computational Science – ICCS 2020*. Springer International Publishing, 2020, pp. 621–634, doi: [10.1007/978-3-030-50420-5_47](https://doi.org/10.1007/978-3-030-50420-5_47).
- [55] M. H. Al-Sheikh, O. Al Dandan, A. S. Al-Shamayleh, H. A. Jalab, and R. W. Ibrahim, "Multi-class deep learning architecture for classifying lung diseases from chest X-ray and CT images," *Scientific Reports*, vol. 13, no. 1, Art no. 19373, 2023, doi: [10.1038/s41598-023-46147-3](https://doi.org/10.1038/s41598-023-46147-3).
- [56] A. Badawi and K. Elgazzar, "Detecting coronavirus from chest X-rays using transfer learning," *COVID*, vol. 1, no. 1, pp. 403–415, 2021, doi: [10.3390/covid1010034](https://doi.org/10.3390/covid1010034).
- [57] H. Kör, H. Erbay, and A. H. Yurttakal, "Diagnosing and differentiating viral pneumonia and COVID-19 using X-ray images," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 39 041–39 057, 2022, doi: [10.1007/s11042-022-13071-z](https://doi.org/10.1007/s11042-022-13071-z).
- [58] S. Kim, B. Rim, S. Choi, A. Lee, S. Min, and M. Hong, "Deep learning in multi-class lung diseases' classification on chest X-ray images," *Diagnostics*, vol. 12, no. 4, Art no. 915, 2022, doi: [10.3390/diagnostics12040915](https://doi.org/10.3390/diagnostics12040915).
- [59] J. Płodowski and J. Mulawka, "Machine learning in recognition of basic pulmonary pathologies," *Applied Sciences*, vol. 12, no. 16, Art no. 8086, 2022, doi: [10.3390/app12168086](https://doi.org/10.3390/app12168086).
- [60] P. L. Vidal, J. de Moura, J. Novo, and M. Ortega, "Multi-stage transfer learning for lung segmentation using portable X-ray devices for patients with COVID-19," *Expert Systems with Applications*, vol. 173, Art no. 114677, Jul. 2021, doi: [10.1016/j.eswa.2021.114677](https://doi.org/10.1016/j.eswa.2021.114677).
- [61] Y.-X. Tang, Y.-B. Tang, Y. Peng, K. Yan, M. Bagheri, B. A. Redd, C. J. Brandon, Z. Lu, M. Han, J. Xiao, et al., "Automated abnormality classification of chest radiographs using deep convolutional neural networks," *npj Digital Medicine*, vol. 3, no. 1, Art no. 70, 2020, doi: [10.1038/s41746-020-0273-z](https://doi.org/10.1038/s41746-020-0273-z).
- [62] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays," *Computer Methods and Programs in Biomedicine*, vol. 196, Art no. 105608, Nov. 2020, doi: [10.1016/j.cmpb.2020.105608](https://doi.org/10.1016/j.cmpb.2020.105608).

- [63] M. Jasmine Pemeena Priyadarsini, K. Kotecha, G. K. Rajini, K. Hariharan, K. Utkarsh Raj, K. Bhargav Ram, V. Indragandhi, V. Subramaniaswamy, and S. Pandya, "Lung diseases detection using various deep learning algorithms," *Journal of Healthcare Engineering*, vol. 2023, no. 1, Art no. 3563696, 2023, doi: [10.1155/2023/3563696](https://doi.org/10.1155/2023/3563696).
- [64] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, "High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images," *Computers in Biology and Medicine*, vol. 155, Art no. 106646, Mar. 2023, doi: [10.1016/j.combiomed.2023.106646](https://doi.org/10.1016/j.combiomed.2023.106646).
- [65] T. I. A. Mohamed, O. N. Oyelade, and A. E. Ezugwu, "Automatic detection and classification of lung cancer CT scans based on deep learning and Ebola optimization search algorithm," *PLOS ONE*, vol. 18, no. 8, Art no. e0285796, 2023, doi: [10.1371/journal.pone.0285796](https://doi.org/10.1371/journal.pone.0285796).
- [66] S. Goyal and R. Singh, "Detection and classification of lung diseases for pneumonia and COVID-19 using machine and deep learning techniques," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 3239–3259, 2021, doi: [10.1007/s12652-021-03464-7](https://doi.org/10.1007/s12652-021-03464-7).
- [67] M. Nahiduzzaman, M. O. Faruq Goni, M. Robiul Islam, A. Sayeed, M. Shamim Anower, M. Ahsan, J. Haider, and M. Kowalski, "Detection of various lung diseases including COVID-19 using extreme learning machine algorithm based on the features extracted from a lightweight CNN architecture," *Biocybernetics and Biomedical Engineering*, vol. 43, no. 3, pp. 528–550, 2023, doi: [10.1016/j.bbe.2023.06.003](https://doi.org/10.1016/j.bbe.2023.06.003).
- [68] F. B. Alam, P. Podder, and M. R. H. Mondal, "RVCNet: A hybrid deep neural network framework for the diagnosis of lung diseases," *PLOS ONE*, vol. 18, no. 12, Art no. e0293125, 2023, doi: [10.1371/journal.pone.0293125](https://doi.org/10.1371/journal.pone.0293125).
- [69] Z. Naz, M. U. G. Khan, T. Saba, A. Rehman, H. Nobanee, and S. A. Bahaj, "An explainable AI-enabled framework for interpreting pulmonary diseases from chest radiographs," *Cancers*, vol. 15, no. 1, Art no. 314, 2023, doi: [10.3390/cancers15010314](https://doi.org/10.3390/cancers15010314).
- [70] R. K. Singh, R. Pandey, and R. N. Babu, "COVIDScreen: Explainable deep learning framework for differential diagnosis of COVID-19 using chest X-rays," *Neural Computing and Applications*, vol. 33, no. 14, pp. 8871–8892, 2021, doi: [10.1007/s00521-020-05636-6](https://doi.org/10.1007/s00521-020-05636-6).
- [71] C. Lam, R. Thapa, J. Maharjan, K. Rahmani, C. F. Tso, N. P. Singh, S. Casie Chetty, and Q. Mao, "Multitask learning with recurrent neural networks for acute respiratory distress syndrome prediction using only electronic health record data: Model development and validation study," *JMIR Medical Informatics*, vol. 10, no. 6, Art no. e36202, 2022, doi: [10.2196/36202](https://doi.org/10.2196/36202).
- [72] A. N. Patel, R. Murugan, G. Srivastava, P. K. R. Maddikunta, G. Yenduri, T. R. Gadekallu, and R. Chengoden, "An explainable transfer learning framework for multi-classification of lung diseases in chest X-rays," *Alexandria Engineering Journal*, vol. 98, pp. 328–343, Jul. 2024, doi: [10.1016/j.aej.2024.04.072](https://doi.org/10.1016/j.aej.2024.04.072).
- [73] M. A. Khan, M. Azhar, K. Ibrar, A. Alqahtani, S. Alsubai, A. Binbusayyis, Y. J. Kim, and B. Chang, "COVID-19 classification from chest X-ray images: A framework of deep explainable artificial intelligence," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–14, Jul. 2022, doi: [10.1155/2022/4254631](https://doi.org/10.1155/2022/4254631).
- [74] E. Mahamud, N. Fahad, M. Assaduzzaman, S. Zain, K. O. M. Goh, and M. K. Morol, "An explainable artificial intelligence model for multiple lung diseases classification from chest X-ray images using fine-tuned transfer learning," *Decision Analytics Journal*, vol. 12, Art no. 100499, Sep. 2024, doi: [10.1016/j.dajour.2024.100499](https://doi.org/10.1016/j.dajour.2024.100499).
- [75] Q. Teng, Z. Liu, Y. Song, K. Han, and Y. Lu, "A survey on the interpretability of deep learning in medical diagnosis," *Multimedia Systems*, vol. 28, no. 6, pp. 2335–2355, 2022, doi: [10.1007/s00530-022-00960-4](https://doi.org/10.1007/s00530-022-00960-4).
- [76] T. K. K. Ho and J. Gwak, "Utilizing knowledge distillation in deep learning for classification of chest X-ray abnormalities," *IEEE Access*, vol. 8, pp. 160 749–160 761, Sep. 2020, doi: [10.1109/access.2020.3020802](https://doi.org/10.1109/access.2020.3020802).
- [77] R. Rajpoot, M. Gour, S. Jain, and V. B. Semwal, "Integrated ensemble CNN and explainable AI for COVID-19 diagnosis from CT scan and X-ray images," *Scientific Reports*, vol. 14, no. 1, Art no. 24985, 2024, doi: [10.1038/s41598-024-75915-y](https://doi.org/10.1038/s41598-024-75915-y).
- [78] I. E. Ihongbe, S. Fouad, T. F. Mahmoud, A. Rajasekaran, and B. Bhatia, "Evaluating Explainable Artificial Intelligence (XAI) techniques in chest radiology imaging through a human-centered lens," *PLOS ONE*, vol. 19, no. 10, pp. 1–27, 2024, doi: [10.1371/journal.pone.0308758](https://doi.org/10.1371/journal.pone.0308758).
- [79] M. M. Islam, M. Z. Islam, A. Asraf, M. S. Al-Rakhami, W. Ding, and A. H. Sodhro, "Diagnosis of COVID-19 from X-rays using combined CNN-RNN architecture with transfer learning," *Benchmark Transactions on Benchmarks, Standards and Evaluations*, vol. 2, no. 4, Art no. 100088, 2022, doi: [10.1016/j.tbench.2023.100088](https://doi.org/10.1016/j.tbench.2023.100088).
- [80] I. Galić, M. Habijan, H. Leventić, and K. Romić, "Machine learning empowering personalized medicine: A comprehensive review of medical image analysis methods," *Electronics*, vol. 12, no. 21, Art no. 4411, 2023, doi: [10.3390/electronics12214411](https://doi.org/10.3390/electronics12214411).

- [81] P. Zhang, A. Swaminathan, and A. A. Uddin, "Pulmonary disease detection and classification in patient respiratory audio files using long short-term memory neural networks," *Frontiers in Medicine*, vol. 10, Art no. 1269784, Nov. 2023, doi: [10.3389/fmed.2023.1269784](https://doi.org/10.3389/fmed.2023.1269784).
- [82] M. Fraiwan, L. Fraiwan, M. Alkhodari, and O. Hassanin, "Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 10, pp. 4759–4771, 2021, doi: [10.1007/s12652-021-03184-y](https://doi.org/10.1007/s12652-021-03184-y).
- [83] L. J. Crasta, R. Neema, and A. R. Pais, "A novel deep learning architecture for lung cancer detection and diagnosis from computed tomography image analysis," *Healthcare Analytics*, vol. 5, Art no. 100316, Jun. 2024, doi: [10.1016/j.health.2024.100316](https://doi.org/10.1016/j.health.2024.100316).
- [84] V. Sharma, Nillmani, S. K. Gupta, and K. K. Shukla, "Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images," *Intelligent Medicine*, vol. 4, no. 2, pp. 104–113, 2024, doi: [10.1016/j.imed.2023.06.001](https://doi.org/10.1016/j.imed.2023.06.001).
- [85] S. R. Vinta, B. Lakshmi, M. A. Safali, and G. S. C. Kumar, "Segmentation and classification of interstitial lung diseases based on hybrid deep learning network model," *IEEE Access*, vol. 12, pp. 50 444–50 458, Mar. 2024, doi: [10.1109/access.2024.3383144](https://doi.org/10.1109/access.2024.3383144).
- [86] S. M. Shafi and S. K. Chinnappan, "Segmenting and classifying lung diseases with m-segnet and hybrid squeezenet-cnn architecture on CT images," *PLOS ONE*, vol. 19, no. 5, Art no. e0302507, 2024, doi: [10.1371/journal.pone.0302507](https://doi.org/10.1371/journal.pone.0302507).
- [87] M. V. Shetty, J. D., and S. Tunga, "Optimized deformable model-based segmentation and deep learning for lung cancer classification," *The Journal of Medical Investigation*, vol. 69, no. 3.4, pp. 244–255, 2022, doi: [10.2152/jmi.69.244](https://doi.org/10.2152/jmi.69.244).
- [88] S. Sangeetha, S. K. Mathivanan, P. Karthikeyan, H. Rajadurai, B. D. Shivahare, S. Mallik, and H. Qin, "An enhanced multimodal fusion deep learning neural network for lung cancer classification," *Systems and Soft Computing*, vol. 6, Art no. 200068, Dec. 2024, doi: [10.1016/j.sasc.2023.200068](https://doi.org/10.1016/j.sasc.2023.200068).
- [89] H. Liang, T. Yang, Z. Liu, W. Jian, Y. Chen, B. Li, Z. Yan, W. Xu, L. Chen, Y. Qi, et al., "LungDiag: Empowering artificial intelligence for respiratory diseases diagnosis based on electronic health records, a multicenter study," *MedComm*, vol. 6, no. 1, Art no. e70043, 2025, doi: [10.1002/mco2.70043](https://doi.org/10.1002/mco2.70043).
- [90] L. Wang, "Deep learning techniques to diagnose lung cancer," *Cancers*, vol. 14, no. 22, Art no. 5569, 2022, doi: [10.3390/cancers14225569](https://doi.org/10.3390/cancers14225569).
- [91] S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek, and S. Selvarajan, "Efficient pneumonia detection using vision transformers on chest X-rays," *Scientific Reports*, vol. 14, no. 1, Art no. 2487, 2024, doi: [10.1038/s41598-024-52703-2](https://doi.org/10.1038/s41598-024-52703-2).
- [92] T. Wang, Z. Nie, R. Wang, Q. Xu, H. Huang, H. Xu, F. Xie, and X.-J. Liu, "PneuNet: deep learning for COVID-19 pneumonia diagnosis on chest X-ray image analysis using vision transformer," *Medical & Biological Engineering & Computing*, vol. 61, no. 6, pp. 1395–1408, 2023, doi: [10.1007/s11517-022-02746-2](https://doi.org/10.1007/s11517-022-02746-2).
- [93] S. R. Rezaei and A. Ahmadi, "A hierarchical GAN method with ensemble CNN for accurate nodule detection," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, no. 4, pp. 695–705, 2022, doi: [10.1007/s11548-022-02807-9](https://doi.org/10.1007/s11548-022-02807-9).
- [94] X. Li, X. Fei, Z. Yan, H. Ren, C. Shi, X. Zhang, I. Mumtaz, Y. Luo, and X. Wu, "CAGAN: Classifier-augmented generative adversarial networks for weakly-supervised COVID-19 lung lesion localisation," *IET Computer Vision*, vol. 18, no. 1, pp. 1–14, 2023, doi: [10.1049/cvi2.12216](https://doi.org/10.1049/cvi2.12216).
- [95] R. Gulakala, B. Markert, and M. Stoffel, "Generative adversarial network based data augmentation for CNN based detection of COVID-19," *Scientific Reports*, vol. 12, no. 1, Art no. 19186, 2022, doi: [10.1038/s41598-022-23692-x](https://doi.org/10.1038/s41598-022-23692-x).
- [96] W. M. Salama and M. H. Aly, "Framework for COVID-19 segmentation and classification based on deep learning of computed tomography lung images," *Journal of Electronic Science and Technology*, vol. 20, no. 3, Art no. 100161, 2022, doi: [10.1016/j.jnlest.2022.100161](https://doi.org/10.1016/j.jnlest.2022.100161).
- [97] W. M. Salama, A. Shokry, and M. H. Aly, "A generalized framework for lung cancer classification based on deep generative models," *Multimedia Tools and Applications*, vol. 81, no. 23, pp. 32 705–32 722, 2022, doi: [10.1007/s11042-022-13005-9](https://doi.org/10.1007/s11042-022-13005-9).
- [98] J. Saldanha, S. Chakraborty, S. Patil, K. Kotecha, S. Kumar, and A. Nayyar, "Data augmentation using variational autoencoders for improvement of respiratory disease classification," *PLOS ONE*, vol. 17, no. 8, Art no. e0266467, 2022, doi: [10.1371/journal.pone.0266467](https://doi.org/10.1371/journal.pone.0266467).

- [99] Y. Choi, W. Yu, M. B. Nagarajan, P. Teng, J. G. Goldin, S. S. Raman, D. R. Enzmann, G. H. J. Kim, and M. S. Brown, "Translating AI to clinical practice: Overcoming data shift with explainability," *RadioGraphics*, vol. 43, no. 5, Art no. e220105, 2023, doi: [10.1148/rg.220105](https://doi.org/10.1148/rg.220105).
- [100] J. Chen, H. Zeng, C. Zhang, Z. Shi, A. Dekker, L. Wee, and I. Bermejo, "Lung cancer diagnosis using deep attention-based multiple instance learning and radiomics," *Medical Physics*, vol. 49, no. 5, pp. 3134–3143, 2022, doi: [10.1002/mp.15539](https://doi.org/10.1002/mp.15539).
- [101] Z. UrRehman, Y. Qiang, L. Wang, Y. Shi, Q. Yang, S. U. Khattak, R. Aftab, and J. Zhao, "Effective lung nodule detection using deep CNN with dual attention mechanisms," *Scientific Reports*, vol. 14, no. 1, Art no. 3934, 2024, doi: [10.1038/s41598-024-51833-x](https://doi.org/10.1038/s41598-024-51833-x).
- [102] D. Li, "Attention-enhanced architecture for improved pneumonia detection in chest X-ray images," *BMC Medical Imaging*, vol. 24, no. 1, Art no. 6, 2024, doi: [10.1186/s12880-023-01177-1](https://doi.org/10.1186/s12880-023-01177-1).
- [103] S. Borwankar, J. P. Verma, R. Jain, and A. Nayyar, "Improvise approach for respiratory pathologies classification with multilayer convolutional neural networks," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 39 185–39 205, 2022, doi: [10.1007/s11042-022-12958-1](https://doi.org/10.1007/s11042-022-12958-1).
- [104] A. I. Taloba and R. Matoog, "Detecting respiratory diseases using machine learning-based pattern recognition on spirometry data," *Alexandria Engineering Journal*, vol. 113, pp. 44–59, Feb. 2025, doi: [10.1016/j.aej.2024.11.009](https://doi.org/10.1016/j.aej.2024.11.009).
- [105] J. N. Siebert, M.-A. Hartley, D. S. Courvoisier, M. Salamin, L. Robotham, J. Doenz, C. Barazzzone-Argiroffo, A. Gervais, and P.-O. Bridevaux, "Deep learning diagnostic and severity-stratification for interstitial lung diseases and chronic obstructive pulmonary disease in digital lung auscultations and ultrasonography: Clinical protocol for an observational case-control study," *BMC Pulmonary Medicine*, vol. 23, no. 1, Art no. 191, 2023, doi: [10.1186/s12890-022-02255-w](https://doi.org/10.1186/s12890-022-02255-w).
- [106] W. Quanyang, H. Yao, W. Sicong, Q. Linlin, Z. Zewei, H. Donghui, L. Hongjia, and Z. Shijun, "Artificial intelligence in lung cancer screening: Detection, classification, prediction, and prognosis," *Cancer Medicine*, vol. 13, no. 7, Art no. e7140, 2024, doi: [10.1002/cam4.7140](https://doi.org/10.1002/cam4.7140).
- [107] A. A. Trusculescu, D. Manolescu, E. Tudorache, and C. Oancea, "Deep learning in interstitial lung disease—how long until daily practice," *European Radiology*, vol. 30, no. 11, pp. 6285–6292, 2020, doi: [10.1007/s00330-020-06986-4](https://doi.org/10.1007/s00330-020-06986-4).
- [108] M. Shimja and K. Kartheeban, "Empowering diagnosis: An astonishing deep transfer learning approach with fine tuning for precise lung disease classification from CXR images," *Automatika*, vol. 65, no. 1, pp. 192–205, 2023, doi: [10.1080/00051144.2023.2290737](https://doi.org/10.1080/00051144.2023.2290737).
- [109] J. Herington, M. D. McCradden, K. Creel, R. Boellaard, E. C. Jones, A. K. Jha, A. Rahmim, P. J. Scott, J. J. Sunderland, R. L. Wahl, et al., "Ethical considerations for artificial intelligence in medical imaging: Data collection, development, and evaluation," *Journal of Nuclear Medicine*, vol. 64, no. 12, pp. 1848–1854, 2023, doi: [10.2967/jnumed.123.266080](https://doi.org/10.2967/jnumed.123.266080).
- [110] G. Mohan, M. M. Subashini, S. Balan, and S. Singh, "A multiclass deep learning algorithm for healthy lung, COVID-19 and pneumonia disease detection from chest X-ray images," *Discover Artificial Intelligence*, vol. 4, no. 1, Art no. 20, 2024, doi: [10.1007/s44163-024-00110-x](https://doi.org/10.1007/s44163-024-00110-x).
- [111] R. Alaufi, M. Kalkatawi, and F. Abukhodair, "Challenges of deep learning diagnosis for COVID-19 from chest imaging," *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 14 337–14 361, 2023, doi: [10.1007/s11042-023-16017-1](https://doi.org/10.1007/s11042-023-16017-1).
- [112] A. Gielczyk, A. Marciniak, M. Tarczewska, and Z. Lutowski, "Pre-processing methods in chest X-ray image classification," *PLOS ONE*, vol. 17, no. 4, Art no. e0265949, 2022, doi: [10.1371/journal.pone.0265949](https://doi.org/10.1371/journal.pone.0265949).
- [113] S. Arvind, J. V. Tembhurne, T. Diwan, and P. Sahare, "Improvised light weight deep CNN based U-Net for the semantic segmentation of lungs from chest X-rays," *Results in Engineering*, vol. 17, Art no. 100929, Mar. 2023, doi: [10.1016/j.rineng.2023.100929](https://doi.org/10.1016/j.rineng.2023.100929).
- [114] P. Yadav, V. Rastogi, A. Yadav, and P. Parashar, "Artificial intelligence: A promising tool in diagnosis of respiratory diseases," *Intelligent Pharmacy*, vol. 2, no. 6, pp. 784–791, 2024, doi: [10.1016/j.ipha.2024.05.002](https://doi.org/10.1016/j.ipha.2024.05.002).
- [115] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, Art no. 60, 2019, doi: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0).
- [116] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 3462–3471, doi: [10.1109/cvpr.2017.369](https://doi.org/10.1109/cvpr.2017.369).

- [117] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, pp. 475–477, 2014, doi: [10.3978/j.issn.2223-4292.2014.11.20](https://doi.org/10.3978/j.issn.2223-4292.2014.11.20).
- [118] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 577–590, 2014, doi: [10.1109/TMI.2013.2290491](https://doi.org/10.1109/TMI.2013.2290491).
- [119] A. Johnson, T. Pollard, R. Mark, S. Berkowitz, and S. Horng, "MIMIC-CXR Database," *PhysioNet*, Jul. 2024, doi: [10.13026/4jqj-jw95](https://doi.org/10.13026/4jqj-jw95).
- [120] B. M. Rocha, D. Filos, L. Mendes, G. Serbes, S. Ulukaya, Y. P. Kahya, N. Jakovljevic, T. L. Turukalo, I. M. Vogiatzis, E. Perantoni, et al., "An open access database for the evaluation of respiratory sound classification algorithms," *Physiological Measurement*, vol. 40, no. 3, Art no. 035001, 2019, doi: [10.1088/1361-6579/ab03ea](https://doi.org/10.1088/1361-6579/ab03ea).
- [121] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, et al., "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013, doi: [10.1007/s10278-013-9622-7](https://doi.org/10.1007/s10278-013-9622-7).
- [122] J. Eary and L. Shankar, "COVID-19 update from the NCI cancer imaging program," *Radiology: Imaging Cancer*, vol. 2, no. 3, Art no. e204017, 2020, doi: [10.1148/rycan.2020204017](https://doi.org/10.1148/rycan.2020204017).
- [123] A. A. A. Setio, A. Traverso, T. de Bel, M. S. Berens, C. v. d. Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts, et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, Dec. 2017, doi: [10.1016/j.media.2017.06.015](https://doi.org/10.1016/j.media.2017.06.015).
- [124] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghighi, R. Ball, K. Shpankaya, et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 590–597, 2019, doi: [10.1609/aaai.v33i01.3301590](https://doi.org/10.1609/aaai.v33i01.3301590).
- [125] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018, doi: [10.1016/j.cell.2018.02.010](https://doi.org/10.1016/j.cell.2018.02.010).
- [126] M. A. Talukder, "Chest X-Ray image," version 1, *Mendeley Data*, 2023, doi: [10.17632/m4s2jn3csb.1](https://doi.org/10.17632/m4s2jn3csb.1).
- [127] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, et al., "Can AI help in screening viral and COVID-19 Pneumonia?" *IEEE Access*, vol. 8, pp. 132 665–132 676, Jul. 2020, doi: [10.1109/access.2020.3010287](https://doi.org/10.1109/access.2020.3010287).
- [128] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. Abul Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan, et al., "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Computers in Biology and Medicine*, vol. 132, Art no. 104319, May 2021, doi: [10.1016/j.combiomed.2021.104319](https://doi.org/10.1016/j.combiomed.2021.104319).
- [129] Y. Liu, Y.-H. Wu, Y. Ban, H. Wang, and M.-M. Cheng, "Rethinking computer-aided tuberculosis diagnosis," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2020, pp. 2643–2652, doi: [10.1109/cvpr42600.2020.00272](https://doi.org/10.1109/cvpr42600.2020.00272).
- [130] L. Orlandic, T. Teijeiro, and D. Atienza, "The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Scientific Data*, vol. 8, no. 1, Art no. 156, 2021, doi: [10.1038/s41597-021-00937-4](https://doi.org/10.1038/s41597-021-00937-4).
- [131] E. Colak, F. C. Kitamura, S. B. Hobbs, C. C. Wu, M. P. Lungren, L. M. Prevedello, J. Kalpathy-Cramer, R. L. Ball, G. Shih, A. Stein, et al., "The RSNA pulmonary embolism CT dataset," *Radiology: Artificial intelligence*, vol. 3, no. 2, Art no. e200254, 2021, doi: [10.1148/ryai.2021200254](https://doi.org/10.1148/ryai.2021200254).
- [132] National Lung Screening Trial Research Team, "The national lung screening trial: Overview and study design," *Radiology*, vol. 258, no. 1, pp. 243–253, 2011, doi: [10.1148/radiol.10091808](https://doi.org/10.1148/radiol.10091808).
- [133] D. Yang, Y. Miao, C. Liu, N. Zhang, D. Zhang, Q. Guo, S. Gao, L. Li, J. Wang, S. Liang, et al., "Advances in artificial intelligence applications in the field of lung cancer," *Frontiers in Oncology*, vol. 14, Art no. 1449068, Sep. 2024, doi: [10.3389/fonc.2024.1449068](https://doi.org/10.3389/fonc.2024.1449068).
- [134] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Research Notes*, vol. 15, no. 1, Art no. 210, 2022, doi: [10.1186/s13104-022-06096-y](https://doi.org/10.1186/s13104-022-06096-y).
- [135] R. Hertel and R. Benlamri, "A deep learning segmentation-classification pipeline for X-ray-based COVID-19 diagnosis," *Biomedical Engineering Advances*, vol. 3, Art no. 100041, Jun. 2022, doi: [10.1016/j.bea.2022.100041](https://doi.org/10.1016/j.bea.2022.100041).

- [136] E. Chamseddine, N. Mansouri, M. Soui, and M. Abed, "Handling class imbalance in COVID-19 chest X-ray images classification: Using SMOTE and weighted loss," *Applied Soft Computing*, vol. 129, Art no. 109588, Nov. 2022, doi: [10.1016/j.asoc.2022.109588](https://doi.org/10.1016/j.asoc.2022.109588).
- [137] M. K. Hasan, M. A. Alam, L. Dahal, S. Roy, S. R. Wahid, M. T. E. Elahi, R. MartÃ, and B. Khanal, "Challenges of deep learning methods for COVID-19 detection using public datasets," *Informatics in Medicine Unlocked*, vol. 30, Art no. 100945, 2022, doi: [10.1016/j.imu.2022.100945](https://doi.org/10.1016/j.imu.2022.100945).
- [138] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, Art no. 654, 2024, doi: [10.1038/s41467-024-44824-z](https://doi.org/10.1038/s41467-024-44824-z).
- [139] C. Bluethgen, P. Chambon, J.-B. Delbrouck, R. van der Sluijs, M. Połacin, J. M. Zambrano Chaves, T. M. Abraham, S. Purohit, C. P. Langlotz, and A. S. Chaudhari, "A vision-language foundation model for the generation of realistic chest X-ray images," *Nature Biomedical Engineering*, vol. 9, no. 4, pp. 494–506, 2024, doi: [10.1038/s41551-024-01246-y](https://doi.org/10.1038/s41551-024-01246-y).
- [140] F. N. Felder and S. L. Walsh, "Exploring computer-based imaging analysis in interstitial lung disease: Opportunities and challenges," *ERJ Open Research*, vol. 9, no. 4, Art no. 00145–2023, 2023, doi: [10.1183/23120541.00145-2023](https://doi.org/10.1183/23120541.00145-2023).
- [141] S. A. Alowais, S. S. Alghamdi, N. Alsuehany, T. Alqahtani, A. I. Alshaya, S. N. Almohareb, A. Aldairem, M. Alrashed, K. Bin Saleh, H. A. Badreldin, et al., "Revolutionizing healthcare: The role of artificial intelligence in clinical practice," *BMC Medical Education*, vol. 23, no. 1, Art no. 689, 2023, doi: [10.1186/s12909-023-04698-z](https://doi.org/10.1186/s12909-023-04698-z).
- [142] A. Makkar and K. Santosh, "SecureFed: Federated learning empowered medical imaging technique to analyze lung abnormalities in chest X-rays," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 8, pp. 2659–2670, 2023, doi: [10.1007/s13042-023-01789-7](https://doi.org/10.1007/s13042-023-01789-7).
- [143] T. Sanida and M. Dasygenis, "A novel lightweight CNN for chest X-ray-based lung disease identification on heterogeneous embedded system," *Applied Intelligence*, vol. 54, no. 6, pp. 4756–4780, 2024, doi: [10.1007/s10489-024-05420-2](https://doi.org/10.1007/s10489-024-05420-2).