Article

A Survey on Multimodal Approaches for Lung Disease Diagnosis using Deep Learning

Zhaniya Medeuova [1]

[1]Department of Computer Science, SDU University, Almaty, Kazakhstan

DOI: 10.47344/tx89w092

Abstract

Lung disorders are a major global health issue. A quick and accurate diagnosis is essential for proper treatment. In order to increase diagnostic accuracy, recent multimodal techniques are gaining popularity. This study carried out a comprehensive analysis of research articles on multimodal approaches that were published between 2020 and 2024 in Scopus and Google Scholar. The results show that there is limited study on the multimodal approach and on a variety of lung disorders such as asthma, TB, pneumonia, and chronic obstructive pulmonary disease. Several studies concentrated mainly on the detection and binary classification of COVID-19. The field has several challenges, including limited datasets, high computing costs, difficulties in integrating multiple modalities, and lack of accessibility of the models. Future studies should look at a wider range of lung diseases, increase the accessibility of datasets, improve fusion methods for merging data from many sources, and create models that are easier to understand and use fewer resources. Resolving these issues will improve patient outcomes by advancing the real-world use of deep learning in medical diagnosis.

Keywords: deep learning, multimodal approach, lung diseases, medical imaging, lung sounds, regression, classification, diagnostics.

I. Introduction

The respiratory system plays a crucial role in the human body, facilitating the exchange of oxygen and carbon dioxide [1]. Despite its flexibility, it remains at risk for numerous diseases that can significantly affect lung function and overall human health. Lung diseases cover a broad category of disorders such as pneumonia, tuberculosis, chronic obstructive pulmonary disease, and lung cancer, among others. These diseases are a major cause of morbidity and mortality on a global scale [2].

The World Health Organization informs that in 2019 around 3.23 million victims were COPD. In the same year, it was reported that chronic respiratory diseases were responsible for 4 million deaths overall. In the United States, asthma affects more than 23.3 million adults and 6.6% children, resulting in significant treatment costs and reduced quality of life [3]–[5]. Furthermore, in 2024

Email: zhaniya.medeuova@sdu.edu.kz ORCID: 0009-0004-7409-9792

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

Kazakhstan had one of the highest rates of lung disorders globally [6]. These statistics highlight the impact of lung diseases on global health and the need for better diagnostic methods that can quickly and accurately identify diseases.

Traditional methods for diagnosing lung disease are medical history reviews, blood tests, lung sound, chest X-rays, and CT scans, etc. [7]. However, these methods have their own drawbacks, such as the dependence on expert analysis and limited accessibility in the environment. Sometimes, these methods can be the cause of human error. That is why manual checking and image-based analysis emphasize the need for more automated and standardized diagnostic processes [8], [9].

Nowadays deep learning has become a solution for these issues, providing precise and automatic diagnostic skills. Due to the increasing availability of medical imaging and acoustic data, researchers have created deep learning algorithms that can accurately identify lung problems [10]. In order to identify diseases, these models have shown remarkable success in evaluating lung sound recordings, CT scans, and chest X-rays. Notable developments include the application of Recurrent Neural Network for lung sound analysis and Convolutional Neural Network for image based classification. For instance, Çallı et al. emphasized the efficacy of deep learning models like VGGNet and ResNet in chest X-ray processing, Ahmed et al. investigated CNN based architectures for lung disease identification using chest imaging [11], [12]. Likewise, Sfayyih et al. examined the function of acoustic signal analysis in identifying lung diseases, stressing the significance of CNN models based on spectrograms [13].

Kieu et al. examined 98 research from 2016 to 2020. They presented a taxonomy that included ensemble techniques, algorithms, transfer learning, augmentation, and features. Large image sizes, a lack of publicly available datasets, data imbalance, and significant error correlation in ensemble models are some of the main issues noted. In order to overcome these problems, the authors proposed using cloud computing, different feature extraction, dataset sharing, and enhanced ensemble approaches. This survey article offers insightful information, more research is necessary given recent developments in datasets and model designs [14]. AI based lung sound categorization for the diagnosis of respiratory diseases was reviewed by Wanasinghe et al., who highlighted developments in deep learning models, data augmentation, feature extraction, and explainability. With fusion models reaching up to 98% accuracy, CNN performed incredibly. However, several obstacles persist, such as the scarcity of datasets, the dependence on individual feature representations, and the absence of explainable AI methodologies. Developing clinical support tools for real-world applications, increasing model interpretability, and diversifying datasets should be the main goals of future research [9]. In their assessment of deep learning-based acoustic analysis for lung disease diagnosis, Sfayyih et al. emphasized the expanding use of Deep Learning Convolutional Neural Networks (DLCNNs) in the detection of obstructive lung diseases. There are no as many reviews on signal-based lung disease detection as there once was. Although they show potential, DLCNNs need to be further validated through extensive research. Data standardization, clinical acceptance, and enhancing diagnostic reliability should be the main areas of future study to assist industry applications and medical practitioners [13].

Despite these developments, most of the other research has focused on single-modal strategies that leverage acoustic analysis, medical imaging, or other discrete data sources. Deep learning techniques for lung illness diagnosis have been evaluated in a variety of survey publications, these researchers mainly focus on single-method approaches such as respiratory sound categorization or CNN-based medical imaging analysis or other types of data [15]. On the other hand, diagnosing lung disease usually requires a variety of clinical data sources, such as the patient's medical history, symptoms, and other relevant information. The multimodal approach can improve diagnostic accuracy, reduce biases, and increase predictability by integrating multiple data sources [15]. And this survey aims to close this gap by providing an overview of multimodal deep learning methods for diagnosing lung diseases. The objectives include assessing the effectiveness of multimodal models, identifying challenges that retard the progression in this field, and exploring solutions that can be implemented to improve model accessibility and performance in a variety of lung diseases.

The following sections present a detailed review of multimodal deep learning techniques. The second section describes the strategy used to collect and examine the relevant literature, including research published in Russian, Kazakh, and English. The third section outlines the fundamental steps needed for deep learning applications, including feature extraction, data preprocessing, model training, and evaluation. The fourth section classifies current techniques and examines breakthroughs in this area. In conclusion, the importance of deep learning in improving the diagnosis of lung diseases and the potential impact of multimodal approaches will be addressed.

II. Methodology

This research uses a systematic process to identify and analyze recent work on the multimodal approach. The methodology is divided into major steps that include the process of selecting the articles, the filtering process, and the analysis of the selected articles. The research was carried out in the Scopus and Google Scholar databases, with an emphasis on Q1-ranked papers published between 2020 and 2024. The research terms used were a combination of phrases such as "deep learning", "detection", "lung disease"

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

(including asthma, chronic obstructive pulmonary disease, COPD, lung cancer, tuberculosis, pneumonia, COVID-19) and with terms like ”image”, ”audio”, and ”sound” to ensure that suitable research is obtained.

The selection process is summarized in Figure 1b. The initial search yielded 535 papers from Scopus and 550 from Google Scholar. A filtering process was then applied to exclude duplicate records and retain only studies that explicitly utilized both image and audio or sound data in a multimodal approach. This step reduced the selection to 47 studies. Further eligibility screening was performed on the basis of predefined inclusion and exclusion criteria. The inclusion criteria required studies to focus on multimodal deep learning models for lung disease detection, provide clear experimental results and evaluation metrics, be published in English, Russian, or Kazakh and appear in peer-reviewed journals or conferences. Studies were excluded if they used only a single data modality (either image or audio), covered diseases beyond the scope of this research, or lacked clear methodological details or experimental validation.

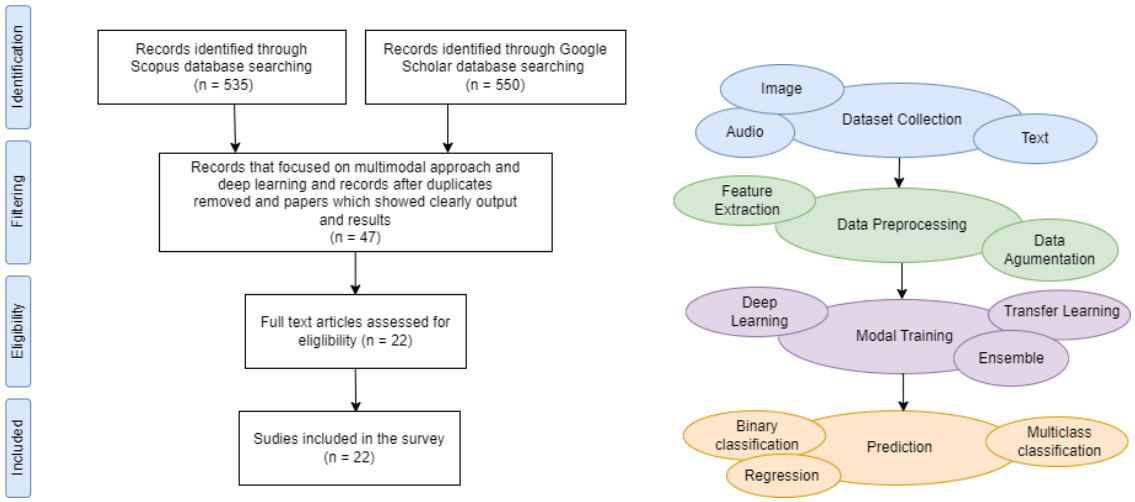
Following this process, 22 articles were considered eligible for inclusion in the final survey. These selected studies provided meaningful information on current trends and challenges of multimodal deep learning in lung disease detection. And the results of recent studies are summarized in Table I to provide a better understanding of the different modalities and their uses in the diagnosis of lung diseases.

Table I summarizes the various research studies that were analyzed in this survey, emphasizing the variety of modalities, datasets, and neural network architectures that were used. This indicates the diversity of approaches currently being explored in the field of lung disease diagnosis using multimodal deep learning techniques.

This methodology section included the selection of relevant studies, a filtering process was used to ensure that only multimodal approaches were included, and the final set of studies was assessed using predefined criteria. The selected articles provide information on current trends, challenges and advances in the integration of multiple data modalities for improved diagnostic accuracy.

III. Fundamental Steps in Applying Deep Learning for Lung Disease Detection

Deep learning plays an essential role in the identification of lung diseases by analyzing medical images and patient data. The process consists of four key steps, they are data collection, data preprocessing, training model, and prediction making [14]. The overview of the process is illustrated at Figure 1b.



(a) (b)

Fig. 1: (a) The survey methodology, (b) Overview of using DL for lung disease detection

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

TABLE I: Summary of multimodal deep learning approaches for lung disease diagnosis

Study	Modality	Datasets Used	Neural Network Architecture	Key Results
Kumar et al., 2023 [18]	img + text	Manually collected (289 patients, future 65k records)	DenseNet121, ResNet50, LSTM, SVM fusion	Intermediate fusion improved accuracy by 2.9%
Malik et al., 2024 [19]	img + audio	24 public datasets (CXR, Cough sound, RSNA, etc.)	CNN + BANL, RBAP, MWDG	Achieved SOTA performance across diseases
Kumar et al., 2024 [20]	img + text	3,256 patient records (India)	CNN, Denoising Autoencoder, Cross-Modal Transformer	Addressed data imbalance, high accuracy for TB classification
Abhishek et al., 2024 [21]	img + audio	1,979 respiratory sound recordings	Hybrid CNN-GRU model	High accuracy in common respiratory diseases, overfitting risk
Sangeetha et al., 2024 [22]	img + text	TCIA, TCGA	MFDNN, CNN, DNN, Intermediate Fusion	92.5% accuracy in lung cancer classification
Varunkumar et al., 2024 [23]	img + img	RIDER Lung CT, Kaggle X-ray	CNN with dilated convolutions, multimodal fusion	Limited dataset diversity, generalizability issues
Hamdi et al., 2021 [24]	img + text	Public IPF dataset (33,026 CT + 1,549 records)	EfficientNet, DenseNet, LSTM, Attention Fusion	Multimodal integration improved prediction accuracy
Kumar et al., 2024 [25]	img + audio + text	AIIMS, Raipur (CT, X-ray, cough, lung sounds)	EfficientNet, RNN, U-Net, OpenL3, RVFL neuro-fuzzy model	COPD prediction using multimodal fusion
Deng et al., 2024 [35]	img + text	East China hospitals, Kaggle COVID-19 CT	CNN + Contrastive Learning + Early Fusion	Contrastive learning improved performance, Grad-CAM interpretation
Adeshina et al., 2022 [26]	img + audio	COVIDx, SARS-CoV-2 CT-scan dataset	CNN, ResNet, DenseNet, XResNet, Self-Attention	91.07% accuracy, effective multimodal cascaded approach
Thandu et al., 2024 [27]	img + audio	Chest X-ray (COVID-19 Radiography) + COUGHVID	DSPANN + Blockchain-based Privacy (ECHFA)	Data quality challenges, complex attention mechanisms
Liu et al., 2024 [28]	img + text	4 hospitals (China), Chest CT, Clinical Features	DenseNet-201 + DNNs + Early Fusion	Outperformed junior radiologists, 11 key clinical features identified
Farhan et al., 2023 [29]	img + img	CXRTD, PCXRA, CCSC, NIH Chest X-ray	CNN, LSTM, SVM, Decision Tree	Improved severity grading performance
Lay et al., 2022 [30]	img + text	Shenzhen, Montgomery X-ray Dataset	EfficientNet, XGBoost, U-Net	AUC improved by 0.0213 over unimodal models
Mayya et al., 2021 [36]	img + text	COVID-19 Chest X-ray, RSNA Pneumonia Dataset	ResNet18, NLP, Grad-CAM, Deep NN Ensemble	X-ray + diagnosis reports enhanced accuracy
Wu et al., 2021 [31]	img + text	TCIA (422 NSCLC patients)	3D-ResNet, Clinical Embedding Layer, Fusion	Improved survival prediction using multimodal fusion

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

A. Dataset Collection and Data Preprocessing

When collecting data, data can be in the form of chest X-rays, CT scans, medical records of patients, coughing, and breathing sounds [10], [11]. Researchers choose between public medical databases or manually acquire data from hospitals and clinics. To ensure that the model can identify a wide variety of lung disorders, balanced data are crucial. Once data is collected, they are processed to make them clean and ready for use. This includes eliminating noise, improving image quality, and being standardized in terms of size and format. In medical imaging, pre-processing can be in altering contrast, segmentation of lung regions, and removal of extraneous detail. In non-image data, such as patient symptoms or audio, pre-processing can be used to structure information in a well-defined format. The purpose of this step is to clean the data so that the model learns only meaningful patterns [33].

B. Training the Model and Prediction

Before the training step, the model gets a large number of labeled samples to be able to understand its features and patterns of lung diseases. Researchers can use neural network architectures that are appropriate for medical image and sound analysis. During training time, the model continuously changes its internal parameters so that it can better identify diseases. A well-trained model predicts the results of the new data. After being trained, the model is tested with new images or patient data to verify its performance. When given a new X-ray or CT scan, the model makes a decision about whether a patient is healthy or has a specific lung disease [14]. Certain models also give us a probability score that informs us about how certain or confident the model is in its decision. This method can help physicians diagnose patients more quickly and accurately when it is integrated into a clinical workflow.

IV. Taxonomy and Trends in Multimodal Approaches for Lung Disease Diagnosis

This section shows the taxonomy and trends in multimodal approaches to the diagnosis of lung diseases. Modalities, feature engineering, data augmentation, fusion techniques, illness categories, and output types are the six key qualities into which the taxonomy groups the important methodologies used in recent studies. These attributes describe the methods of data acquisition, feature extraction, model enhancement, and prediction. These attributes are discussed in detail in subsections A to B, along with a study of the corresponding research.

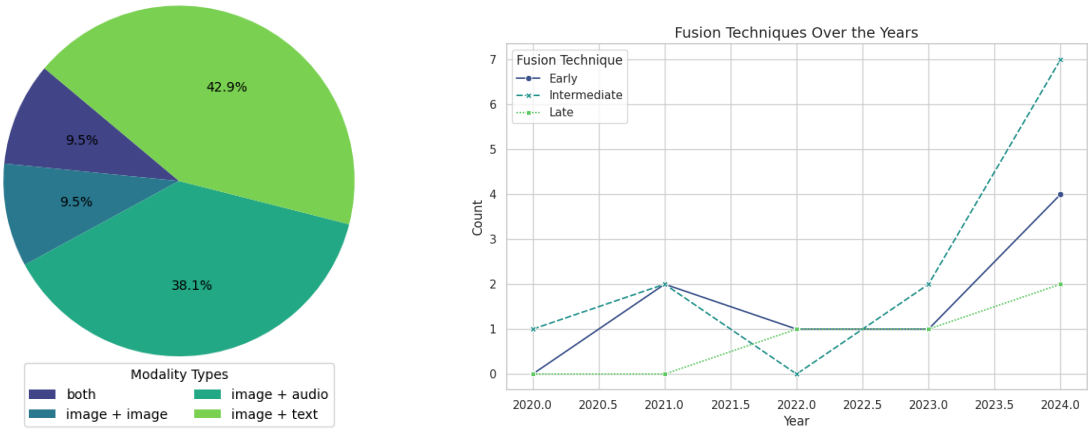
A. Modalities type

Lung disease detection using deep learning is based on various data modalities, often combining multiple sources for better accuracy. Figure 2a shows that some studies use only medical images, such as CT, X-rays, and PET scans, to identify lung abnormalities [29], [33]. Others improve detection by integrating images with respiratory or cough sounds, capturing both structural and acoustic patterns [15], [19], [21], [26], [27], [33], [37], [38]. Another approach combines images with clinical records, including patient demographics, diagnostic reports, and lab results, providing additional diagnostic context [18], [20], [22], [24], [28], [30], [31], [35], [36]. Studies using image and audio data focus primarily on COVID-19, pneumonia, tuberculosis, lung cancer, asthma, and COPD, while image and text combinations are commonly applied to lung cancer, tuberculosis, chronic bronchitis, and pulmonary fibrosis. Some research incorporates the three modalities: images, audio, and text, to improve disease prediction, particularly for COVID-19, COPD, and other complex respiratory conditions [17], [25], [32]. The choice of modality depends on the characteristics of the disease and the available diagnostic data, with multimodal approaches enhancing the accuracy of classification.

B. Feature engineering

Feature engineering is essential for the diagnosis of multimodal lung disease because it has a direct impact on the way deep learning models extract relevant representations from medical data. Handcrafted features and learned features are two main categories into which feature engineering methodologies can be divided. Medical pictures and audio data are manually processed to extract hand-crafted features based on domain-specific knowledge. Texture descriptors, shape characteristics, and statistical qualities are frequently used in imaging modalities, whereas Mel frequency cepstral coefficients (MFCC) and spectrum features are frequently used in audio-based diagnostics. On the other hand, deep learning models, in particular, Convolutional Neural Networks, which are suited to recognizing complex patterns in unstructured information without the need for explicit feature selection automatically extract learned features. Using pre-trained architectures like VGG19, Inception-v3, ResNet, DenseNet, and EfficientNet to increase feature extraction and classification performance, transfer learning has been widely used in recent research. These models are refined on lung disease datasets to extract high-level features relevant to illness detection after being pre-trained on vast datasets. In order

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.



(a) (b)

Fig. 2: (a) Distribution of Modalities, (b) Fusion techniques over time

to minimize dimensionality and maintain the most discriminative features, some research incorporates feature selection methods such as principal component analysis (PCA) and recursive feature elimination (RFE) in addition to feature extraction based on deep learning [18], [22]. This improves the performance of the model. Furthermore, hybrid techniques that integrate learned and handcrafted features have attracted a lot of interest since they allow for a more thorough representation of multimodal data, which eventually improves diagnostic adaptability and accuracy. Multimodal approaches can improve lung disease detection by using these feature engineering techniques to capture high- and low-level data representations, which will improve prediction performance.

C. Data augmentation

Deep learning-based lung disease identification often employs data augmentation to improve model generalization and address data limitation. Rotation, scaling, translation, flipping, contrast alterations, and noise injection are popular augmentation procedures in medical imaging. For specialization on lung regions, some investigations use segmentation-based augmentations such as cropping and scaling. Furthermore, image quality is enhanced by preprocessing techniques such contrast limited adaptive histogram equalization (CLAHE) and histogram matching [36]. Using pitch shifting, temporal stretching, noise injection, and speed perturbation, augmentation techniques alter respiratory sounds for audio-based classification [33]. These techniques help models adjust to changes in recording conditions and sound quality. Furthermore, by increasing the representation of imbalanced classes, data balancing techniques such as MWDG (Multiple-Way Data Generation) and SMOTE (Synthetic Minority Oversampling Technique) reduce model bias [19]. Horizontal flipping, rotation, and width/height shifts are used in public datasets such as POCOVID-Net and NIH Chest X-Ray, in addition to preprocessing techniques such as CLAHE and scaling. Principal component analysis (PCA), image embedding, clustering for defect detection, and Fourier transform are the complex augmentation methods. They are frequently used in manually collected datasets. Preprocessing techniques such as wavelet transformations, noise reduction, and Mel frequency cepstral coefficients (MFCC) improve the accuracy in audio samples [37]. Augmentation has drawbacks despite its benefits. Unrealistic data produced by excessive changes can result in poor model generalization [37]. Complex procedures raise computing costs, and improper augmentation strategies could result in biases. Additionally, broad, high-quality real-world data is still necessary for developing a strong and reliable deep learning model, and augmentation cannot completely replace it.

D. Fusion techniques

In order to improve the quality and strength of computational models, fusion techniques are essential for combining various data sources. Figure 2b shows that several fusion strategies have been used, such as early fusion (E), intermediate fusion (I),

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

and late fusion (L), according to the reviewed publications. The method by which and when the data is joined during processing differ in these methods, which affects model performance and computing efficiency. With 10 experiments, intermediate fusion was the most commonly utilized strategy among the 22 papers surveyed [17], [20]–[24], [27], [29], [33]. Before making a final judgment, features that have been retrieved from several modalities or sources are combined using feature-level integration, which is a common component of intermediate fusion approaches. The Progressive Split Deformable Field Fusion Module (PSDFM), which uses intermediate fusion to improve representation learning, is a notable example [27]. Seven studies used early fusion (E), suggesting a preference for input-level direct data integration [15], [19], [28], [31], [32], [35], [36]. This method is frequently used in situations where it is possible to efficiently mix raw data from many sources prior to feature extraction. Four articles reported the use of late fusion (L), which combines predictions from different models and is frequently used in ensemble-based techniques to increase the accuracy of regression or classification [25], [30], [37]. The flexibility of fusion techniques in complicated problem domains was demonstrated by certain papers that used a combination of fusion procedures, such as L, I and E, I [18], [26].

However, a study specifically mentioned the lack of fusion techniques, implying that independent processing of data sources would be better in some circumstances. The performance of the model is significantly affected by the fusion technique method. Intermediate fusion often outperforms early and late fusion because it allows feature representations from multiple modalities to be refined before final decision making, leading to more discriminative patterns. However, it can be challenging to compute [26]. On the other hand, early fusion ensures that raw data is combined before feature extraction, which can be valuable when different modalities share a common feature space but may struggle with heterogeneous data [18]. Late fusion provides flexibility by allowing independent model predictions to be combined, but may not fully leverage interactions between different data sources. The effectiveness of each method depends on factors such as data heterogeneity, model complexity, and available computational resources. Studies have shown that hybrid approaches, such as the combination of early and intermediate fusion, can further improve performance utilizing data-level and feature-level integration [26].

In general, fusion methods are still being developed, and hybrid fusion models which use several levels of integration to optimize the advantages of various data sources are becoming progressively more popular. Future studies might concentrate on refining fusion techniques to strike a balance between prediction performance and computational economy across a range of application domains.

E. Disease types

The reviewed studies cover a broad spectrum of lung diseases, demonstrating the extensive application of computational models in clinical diagnosis. As shown in Figure 3b, COVID-19 was the most frequently occurring disease to be examined, occurring in nine studies, reaffirming its persistent relevance in clinical imaging [15], [17], [26], [27], [32], [36], [38]. Pneumonia was also a significant area of research, studies of its various forms, including bacterial, viral, lobar, lobular, and Staphylococcus aureus pneumonia (SAP) demonstrating the need for precise diagnostic models [15], [17]–[19]. Tuberculosis (TB) has also been explored frequently, with particular studies differentiating pulmonary TB [15], [19], [20], [37]. Other respiratory infections including bronchitis, lower and upper respiratory tract infections (LRTI, URTI), and bronchiolitis were also explored [37]. Chronic lung diseases sush as Chronic Obstructive Pulmonary Disease (COPD), asthma, and chronic bronchitis were also extensively explored, with the need for early diagnosis and long-term monitoring [25]. Lung cancer, particularly non-small cell lung cancer (NSCLC), was also a significant area of research in various studies [15], [19], [22]. Some studies also explored relatively uncommon but clinically important conditions including Idiopathic Pulmonary Fibrosis (IPF), pleural effusion, and pulmonary edema [24].

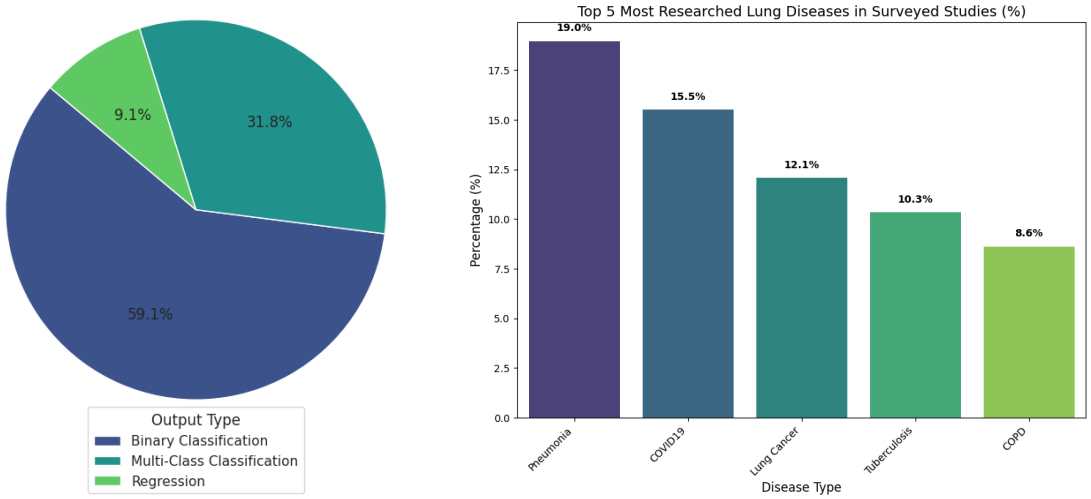
The studies used publicly available datasets or manually collected data. Most of the research used publicly available datasets, ensuring standardized imaging data for training and evaluation. However, some studies included manually collected datasets from hospitals and medical institutions, especially for diseases that are underrepresented in publicly available data [17], [18], [20], [21], [32], [35]–[37]. According to Table 1, ChestX-ray14, COVIDx, Tuberculosis Chest X-ray, RSNA Pneumonia Detection Challenge Dataset, and LIDC-IDRI are the public datasets most commonly used. Large-scale model training was made possible by these datasets, which offered categorized medical imaging data, eliminating the need for manual collection. A smaller number of studies applied datasets that were manually collected, mostly from imaging facilities and hospital records. For rare disorders where public datasets were not enough, such as pleural effusion, pulmonary fibrosis, or mixed-disease classification tasks, these datasets were especially valuable. In comparison to publicly available datasets, personally gathered datasets frequently have smaller sample numbers, but provide greater control over patient demographics and imaging conditions.

Large-scale model training is made easier by publicly accessible datasets, but these datasets frequently contain biases that may hinder the generalizability and performance of the models. Ethnic representation is an important issue. There is a lack of diversity in many large scale datasets, like ChestX-ray14 and COVIDx, because most of the images are taken from particular populations [36]. Because of this, models developed using these datasets might not work consistently across ethnic groups, which could reduce

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

the diagnostic accuracy for underrepresented groups. The distribution of ages is also a significant factor. Adult and elderly patients make up a larger percentage of many datasets, while young children are still underrepresented. For diseases like pneumonia and bronchiolitis, which occur differently in children than in adults, this can present difficulties. Models may perform less well in predicting outcomes for younger patients if they are not trained in a balanced age distribution. In addition, a common limitation is an imbalance in the severity of the disease. Since severe cases are more commonly diagnosed and documented in medical settings, they often make up a larger percentage of public datasets. This makes early stage diagnosis more challenging by biasing model training toward identifying diseases at a later stage. Early detection, which is essential for prompt medical intervention, may be difficult for models trained on unbalanced datasets.

The range of diseases covered in these studies highlights the need for strong deep learning models capable of addressing a variety of lung conditions. To improve predictability, future research may focus on improving classification performance in a range of diseases and ensuring that datasets incorporate world differences. Curating datasets that more accurately reflect a range of age groups, disease severity levels, and populations should be the main goal.



(a) (b)

Fig. 3: (a) Distribution of output types in lung disease diagnosis studies, (b) Distribution of most studied lung diseases in multimodal research

F. Output types

Various types of output were used in the investigated research. Figure 3a shows that the three main types of these outputs were probabilistic estimation, regression based prediction, and classification. Classification tasks, especially binary classification, were a popular type among reviewed articles [18], [20], [22], [23], [25], [26], [28], [30], [32], [35], [36], [38]. A unique case was when a model was categorized according to severity levels rather than type of disease, including mild, moderate, severe, and deadly [29]. Also in regression models used to estimate patient disease severity. Using metrics like the MAE and Concordance Index to estimate survival time for patients with non-small cell lung cancer. Regression based methods were also employed to monitor the severity of COPD and the development of idiopathic pulmonary fibrosis. Probabilistic outputs, which provide confidence scores for the existence or severity levels of the diseases. In multiclass classification tasks, where probability distributions aided in improving decision making in unclear situations, such methods were frequently used. These probabilistic outputs were frequently evaluated using metrics like the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC). The metrics used for the evaluation were chosen based on the selection of the output type. F1-score, recall, specificity, accuracy, and precision were

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

frequently used in binary classification models. Log Loss, Fowlkes-Mallows Index (FMI), and Matthews Correlation Coefficient (MCC) were used in multiclass classification studies. Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R2 score were commonly used to evaluate regression models. It is crucial to use these metrics depending on their strengths and limitations [14]. AUC-ROC and other performance metrics based on probability were commonly used to assess probabilistic models. However, it does not capture data imbalances as well as the F1 score, MCC [34]. Both RMSE and MAE give distinct viewpoints on prediction error in regression models, with RMSE penalizing larger errors progressively. The best evaluation method for a task can be chosen with the help of a structured comparison of these metrics. The comparative analysis presented in Table II underscores the importance of understanding the strengths and limitations inherent in different models, highlighting areas that require further exploration.

TABLE II: Comparison of multimodal models: strengths, weaknesses, and metrics

Study	Strengths	Weaknesses	Evaluation Metrics
Kumar et al., 2023 [18]	Adaptive batch sizes, effective multimodal fusion	Small dataset, data quality issues	Accuracy, Precision, F1 Score
Malik et al., 2024 [19]	Early fusion, data augmentation	Data imbalance, high computational cost	Accuracy (99.01%), MCC, FMI
Kumar et al., 2024 [20]	Cross-modal attention, effective fusion	Small dataset, high computational cost	Accuracy (95%), AUC-ROC, MCC
Abhishek et al., 2024 [21]	Robust feature extraction, real-time processing	Limited class diversity, noisy data	Accuracy (98%), AUC, F1 Score
Sangeetha et al., 2024 [22]	Improved accuracy, effective feature extraction	Privacy concerns, AI interpretability	Accuracy (92.5%), Precision, Recall
Varunkumar et al., 2024 [23]	CNN for feature extraction, hierarchical fusion	Lack of diverse datasets, model interpretability	Accuracy (94%), F1 Score
Hamdi et al., 2021 [24]	CNN+LSTM fusion, attention mechanism	Lung segmentation noise, training complexity	Accuracy (97%), R^2 Score (91%)
Deng et al., 2024 [35]	Hierarchical fine-tuning, contrastive learning	Small dataset, overfitting risk	Accuracy (90.14%), F1 Score
Adeshina et al., 2022 [26]	End-to-end training, self-attention. Discriminative fine-tuning.	Complexity in training models. Sensitivity to hyperparameter tuning.	Accuracy (91.26%), XResNet
Thandu et al., 2024 [27]	Uses multimodal data fusion, achieves high diagnostic accuracy, blockchain for privacy	Scalability, interpretability	Accuracy (98%), AUC (97%)
Liu et al., 2024 [28]	Early fusion, transfer learning	Small sample size, imbalance	AUC (0.92), Accuracy (78%)
Farhan et al., 2023 [29]	CNN+handcrafted features, optimized CNN	Class imbalance, long training times	Accuracy (98.78%), F1 Score
Lay et al., 2022 [30]	Demographic data fusion, late fusion	Small dataset, generalization issues	AUC (0.9574)
Mayya et al., 2021 [36]	Feedback mechanism, Grad-CAM interpretability	Limited dataset, X-ray variability	Accuracy (97%)
Wu et al., 2021 [31]	3D-ResNet, batch normalization	Data variety issues, complex survival model	MAE (0.162), C-index (0.6580)

The reviewed studies highlight a growing trend toward the integration of multiple modalities, advanced feature engineering, and data fusion techniques to improve diagnostic accuracy. The taxonomy reveals that the majority of approaches rely on deep learning, leveraging handcrafted and learned features to optimize performance. Intermediate fusion emerges as the most effective method, striking a balance between enhanced representation learning and computational efficiency. Additionally, publicly available datasets remain the primary source for training models, despite concerns about data diversity. Upcoming advancements should focus on improving fusion techniques, guaranteeing dataset inclusivity, and resolving feature selection issues to increase the diagnostic

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

accuracy for a wider variety of lung conditions.

V. Conclusion

This study investigated the application of deep learning to identify lung diseases by merging various data sets, including lung sounds and medical imaging. Studies show that, in contrast to the use of a single data type, multimodal techniques can increase diagnostic accuracy. But there are still a number of difficulties. The lack of studies that examine a broad spectrum of lung disorders is a major problem. Instead of classifying several lung diseases such as asthma, TB, pneumonia, and chronic obstructive pulmonary disease (COPD), the majority of current research concentrates on the detection or binary classification of COVID-19. This restricts how these models can be used in the real world. The difficulty of combining several data types in a way that improves model performance is another significant obstacle. Large, high-quality datasets are also necessary for deep learning models. However, there are not enough publicly accessible multimodal datasets that cover a range of lung disorders.

Furthermore, doctors find it difficult to believe the predictions made by AI models because they are sometimes complex and difficult to understand. The adoption of these techniques in hospitals with limited resources is further hampered by their high computing costs. It is essential to expand the focus of future studies to include lung conditions other than COVID-19. Improving techniques to efficiently integrate clinical, audio, and visual information can improve diagnosis. Creating larger and more balanced databases with a variety of disease categories should be another priority for researchers. Creating models that can operate with smaller datasets and reduce dependence on enormous amounts of labeled data is another crucial avenue. Enhancing transparency and explainability will contribute to a rise in medical professionals' trust. Lastly, to ensure that these complex algorithms can be applied successfully in actual medical situations, cooperation between AI researchers and healthcare professionals is essential. Deep learning can significantly improve early diagnosis and treatment for a variety of lung diseases by addressing these issues, ultimately improving patient outcomes.

References

[1] Y. Sugandi, I. Soesanti, and H. A. Nugroho, "A Systematic Literature Review of Convolutional Neural Network Architecture for Lung Disease Detection," in Proc. 2023 International Conference on Information and Communications Technology (ICOIACT), 2023, pp. 230-235. DOI: 10.1109/ICOIACT59844.2023.10455864.

[2] P. P. Jasmine, K. Kotecha, G. Rajini, K. Hariharan, K. Raj, K. Ram, V. Indragandhi, V. Subramaniyaswamy, and S. Pandya, "Lung Diseases Detection Using Various Deep Learning Algorithms," Journal of Healthcare Engineering, vol. 2023, pp. 1-13, 2023. DOI: 10.1155/2023/3563696.

[3] GBD 2019 Chronic Respiratory Diseases Collaborators, "Global Burden of Chronic Respiratory Diseases and Risk Factors, 1990-2019: An Update from the Global Burden of Disease Study 2019," EClinicalMedicine, vol. 59, p. 101936, 2023. DOI: 10.1016/j.eclinm.2023.101936.

[4] World Health Organization, "Chronic Obstructive Pulmonary Disease (COPD)," WHO, Nov. 6, 2024. Available: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd).

[5] J. Li, Y. Meng, L. Ma, S. Du, H. Zhu, Q. Pei, and X. Shen, "A Federated Learning Based Privacy-Preserving Smart Healthcare System," IEEE Transactions on Industrial Informatics, vol. 18, pp. 2021-2031, 2022. DOI: 10.1109/TII.2021.3098010.

[6] A. Gafizkyzy, "Qazaqstan ökpe auruynan älem boiynşa üşınşı orynğa şyqqan," Qazaqstan TV, Dec. 5, 2024. Available: https://qazaqstan.tv/news/203518/.

[7] K. Bartziokas, A. Papaporfyriou, G. Hillas, A. Papaioannou, and S. Loukides, "Global Initiative for Chronic Obstructive Lung Disease (GOLD) Recommendations: Strengths and Concerns for Future Needs," Postgraduate Medicine, vol. 135, 2022. DOI: 10.1080/00325481.2022.2135893.

[8] J. P. Allinson, N. Chaturvedi, A. Wong, I. Shah, G. C. Donaldson, J. A. Wedzicha, and R. Hardy, "Early Childhood Lower Respiratory Tract Infection and Premature Adult Death from Respiratory Disease in Great Britain: A National Birth Cohort Study," Lancet (London, England), vol. 401, no. 10383, pp. 1183–1193, 2023. DOI: 10.1016/S0140-6736(23)00131-9.

[9] T. Wanasinghe, S. Bandara, S. Madusanka, D. Meedeniya, M. Bandara, and I. De la Torre Díez, "Lung Sound Classification for Respiratory Disease Identification Using Deep Learning: A Survey," International Journal of Online and Biomedical Engineering (iJOE), vol. 20, pp. 1-15, 2024. DOI: 10.3991/ijoe.v20i10.49585.

[10] A. Ijaz, M. Nabeel, U. Masood, T. Mahmood, M. S. Hashmi, I. Posokhova, A. Rizwan, and A. Imran, "Towards Using Cough for Respiratory Disease Diagnosis by Leveraging Artificial Intelligence: A Survey," Informatics in Medicine Unlocked, vol. 29, p. 100832, 2022. DOI: 10.1016/j.imu.2021.100832.

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

[11] S. Ahmed and S. Kadhem, "Using Machine Learning via Deep Learning Algorithms to Diagnose the Lung Disease Based on Chest Imaging: A Survey," International Journal of Interactive Mobile Technologies (iJIM), vol. 15, p. 95, 2021. DOI: 10.3991/ijim.v15i16.24191.

[12] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, "Deep Learning for Chest X-ray Analysis: A Survey," Medical Image Analysis, vol. 72, p. 102125, 2021. DOI: 10.1016/j.media.2021.102125.

[13] A. H. Sfayyih, N. Sulaiman, and A. H. Sabry, "A Review on Lung Disease Recognition by Acoustic Signal Analysis with Deep Learning Networks," Journal of Big Data, vol. 10, no. 1, p. 101, 2023. DOI: 10.1186/s40537-023-00762-z.

[14] S. T. H. Kieu, A. Bade, M. H. A. Hijazi, and H. Kolivand, "A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy, Issues and Future Directions," Journal of Imaging, vol. 6, no. 12, p. 131, 2020. DOI: 10.3390/jimaging6120131.

[15] H. Malik, T. Anees, A. S. Al-Shamaylehs, 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 (Basel, Switzerland), vol. 13, no. 17, p. 2772, 2023. DOI: 10.3390/diagnostics13172772.

[16] R. Hertel and R. Benlamri, "Deep Learning Techniques for COVID-19 Diagnosis and Prognosis Based on Radiological Imaging," ACM Computing Surveys, vol. 55, 2022. DOI: 10.1145/3576898.

[17] U. Sait, G. L. K. V, S. Shivakumar, T. Kumar, R. Bhaumik, S. Prajapati, K. Bhalla, and A. Chakrapani, "A deep-learning based multimodal system for Covid-19 diagnosis using breathing sounds and chest X-ray images," Applied Soft Computing, vol. 109, p. 107522, 2021. DOI: 10.1016/j.asoc.2021.107522.

[18] S. Kumar, O. Ivanova, A. Melyokhin, and P. Tiwari, "Deep-learning-enabled multimodal data fusion for lung disease classification," Informatics in Medicine Unlocked, vol. 42, p. 101367, 2023. DOI: 10.1016/j.imu.2023.101367.

[19] H. Malik and T. Anees, "Multi-modal deep learning methods for classification of chest diseases using different medical imaging and cough sounds," PLoS One, vol. 19, no. 3, p. e0296352, 2024. DOI: 10.1371/journal.pone.0296352.

[20] S. Kumar and S. Sharma, "An Improved Deep Learning Framework for Multimodal Medical Data Analysis," Big Data and Cognitive Computing, vol. 8, no. 10, p. 125, 2024. DOI: 10.3390/bdcc8100125.

[21] S. Abhishek, A. Ananthapadmanabhan, T. Anjali, S. Remya, A. Perathur, and R. Bentov, "Multimodal Integration of Enhanced Novel Pulmonary Auscultation Real-Time Diagnostic System," IEEE MultiMedia, vol. PP, pp. 1–26, 2024. DOI: 10.1109/MMUL.2024.3422022.

[22] S. Skb, M. S. Kumar, P. Karthikeyan, H. Rajadurai, B. Shivahare, S. Mallik, and H. Qin, "An Enhanced Multimodal Fusion Deep Learning Neural Network for Lung Cancer Classification," Systems and Soft Computing, vol. 6, p. 200068, 2023. DOI: 10.1016/j.sasc.2023.200068.

[23] K. Varunkumar, M. Zymbler, and S. Kumar, "Multimodal Deep Dilated Convolutional Learning for Lung Disease Diagnosis," Brazilian Archives of Biology and Technology, vol. 67, 2024. DOI: 10.1590/1678-4324-2024231088.

[24] A. Hamdi, A. Aboeleneen, and K. Shaban, "MARL: Multimodal Attentional Representation Learning for Disease Prediction," in Proc. 3rd Int. Conf. Artif. Intell. Comput. Vis. (AICV 2021), Springer, 2021, pp. 14–27. DOI: 10.1007/978-3-030-87156-7_2.

[25] S. Kumar, A. V. Shvetsov, and S. H. Alsamhi, "FuzzyGuard: A Novel Multimodal Neuro-Fuzzy Framework for COPD Early Diagnosis," IEEE Internet of Things Journal, 2024. DOI: 10.1109/JIOT.2024.3467176.

[26] S. A. Adeshina and A. P. Adedigba, "Bag of Tricks for Improving Deep Learning Performance on Multimodal Image Classification," Bioengineering, vol. 9, no. 7, p. 312, 2022. DOI: 10.3390/bioengineering9070312.

[27] A. L. Thandu and P. Gera, "Privacy-centric multi-class detection of COVID-19 through breathing sounds and chest X-ray images: Blockchain and optimized neural networks," IEEE Access, vol. 12, pp. 89968-89985, 2024. DOI: 10.1109/ACCESS.2024.3418202.

[28] T. Liu, Z. Zhang, Q. Zhou, et al., "MI-DenseCFNet: Deep learning–based multimodal diagnosis models for Aureus and Aspergillus pneumonia," European Radiology*, vol. 34, pp. 5066–5076, 2024. DOI: 10.1007/s00330-023-10578-3.

[29] A. M. Q. Farhan, S. Yang, A. Q. S. Al-Malahi, and M. A. Al-antari, "MCLSG: Multi-modal classification of lung disease and severity grading framework using consolidated feature engineering mechanisms," Biomedical Signal Processing and Control, vol. 85, p. 104916, 2023. DOI: 10.1016/j.bspc.2023.104916.

[30] J. Lay and B. Pardamean, "Detection of pulmonary tuberculosis on chest X-ray images using multimodal ensemble," ResearchGate, 2022. DOI: 10.13140/RG.2.2.11678.61763.

[31] Y. Wu, J. Ma, X. Huang, S. Ling, and S. Su, "DeepMMSA: A Novel Multimodal Deep Learning Method for Non-small Cell Lung Cancer Survival Analysis," in Proc. IEEE SMC Conf., 2021, pp. 1468–1472. DOI: 10.1109/SMC52423.2021.9658891.

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.

[32] S. Kumar, R. Nagar, S. Bhatnagar, R. Vaddi, S. K. Gupta, M. Rashid, A. K. Bashir, and T. Alkhalifah, "Chest X-ray and cough sample based deep learning framework for accurate diagnosis of COVID-19," Computers & Electrical Engineering, vol. 103, p. 108391, 2022. DOI: 10.1016/j.compeleceng.2022.108391.

[33] Z. Tariq, S. K. Shah, and Y. Lee, "Multimodal lung disease classification using deep convolutional neural network," in Proc. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 2530–2537. DOI: 10.1109/BIBM49941.2020.9313208.

[34] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, p. 6, 2020. DOI: 10.1186/s12864-019-6413-7.

[35] S. Deng, X. Zhang, and S. Jiang, "A diagnostic report supervised deep learning model training strategy for diagnosis of COVID-19," Pattern Recognition, vol. 149, p. 110232, 2024. DOI: 10.1016/j.patcog.2023.110232.

[36] V. Mayya, K. Karthik, S. S. Kamath, K. Karadka, and J. Jeganathan, "COVIDDX: AI-based clinical decision support system for learning COVID-19 disease representations from multimodal patient data," in Proc. International Conference on Health Informatics, 2021.

[37] S. Kumar, V. Bhagat, P. Sahu, M. K. Chaube, A. K. Behera, M. Guizani, R. Gravina, M. Di Dio, G. Fortino, E. Curry, and S. H. Alsamhi, "A novel multimodal framework for early diagnosis and classification of COPD based on CT scan images and multivariate pulmonary respiratory diseases," *Computer Methods and Programs in Biomedicine*, vol. 243, p. 107911, 2024. DOI: 10.1016/j.cmpb.2023.107911.

[38] M. J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, and N. Shukla, "COVID-19 detection through transfer learning using multimodal imaging data," *IEEE Access*, vol. 8, pp. 149808–149824, 2020. DOI: 10.1109/ACCESS.2020.3016780.

Received: March 10, 2025. Reviewed: March 18, 2025. Accepted: March 20, 2025.