# Machine Learning for Respiratory Sound Analysis: A Survey

www.surveyx.cn


## Abstract

Machine learning's integration into respiratory diagnostics has revolutionized the field, significantly enhancing the accuracy and efficiency of diagnosing respiratory conditions. This survey paper explores the transformative impact of machine learning, particularly deep learning models such as convolutional neural networks (CNNs), in classifying and detecting respiratory sounds. These advancements address the limitations of traditional auscultation methods by offering automated solutions that reduce inter-listener variability and improve diagnostic consistency. The paper also highlights the role of hybrid and multimodal approaches, which integrate diverse data sources and computational techniques to further enhance diagnostic accuracy. Despite these advancements, challenges such as data limitations, model complexity, and clinical applicability persist. The need for large, well-labeled datasets and the development of interpretable models capable of real-time operation remain critical areas for future research. Addressing these challenges will require innovative solutions and interdisciplinary collaboration to ensure effective generalization across diverse clinical settings. The potential of self-supervised pre-trained models to improve classification tasks underscores the importance of continued research and development in this domain. By leveraging these models, future research can enhance machine learning capabilities in respiratory diagnostics, ultimately leading to more accurate, efficient, and accessible healthcare solutions.


## 1 Introduction

### 1.1 Significance of Respiratory Sound Analysis

The analysis of respiratory sounds is crucial for diagnosing respiratory conditions, providing insights that enable timely and accurate medical interventions. Respiratory diseases, including asthma, chronic obstructive pulmonary disease (COPD), and sleep apnea, are leading causes of morbidity and mortality globally, highlighting the necessity for early diagnosis to improve treatment outcomes [1]. Beyond individual patient care, respiratory sound analysis enhances the detection accuracy of various conditions, significantly influencing healthcare outcomes [2].

Traditional auscultation methods, while valuable, suffer from inter-listener variability and reliance on in-person examinations, which may lead to inaccuracies [3]. The integration of advanced computational techniques, particularly machine learning, offers a promising avenue to mitigate these limitations by improving diagnostic accuracy and efficiency [4]. This is especially vital in low-resource settings where access to skilled physicians is limited, positioning automated respiratory sound analysis as a critical diagnostic tool.

Accurate detection and classification of adventitious respiratory sounds, such as crackles and wheezes, are essential for developing automated systems that enhance diagnostic precision [5]. Standardizing methodologies and comparisons in the literature is necessary for effective sound detection and classification, ultimately improving healthcare outcomes. The analysis of respiratory sounds parallels the need for accurately modeling interrelations among multiple organs in computational anatomy,

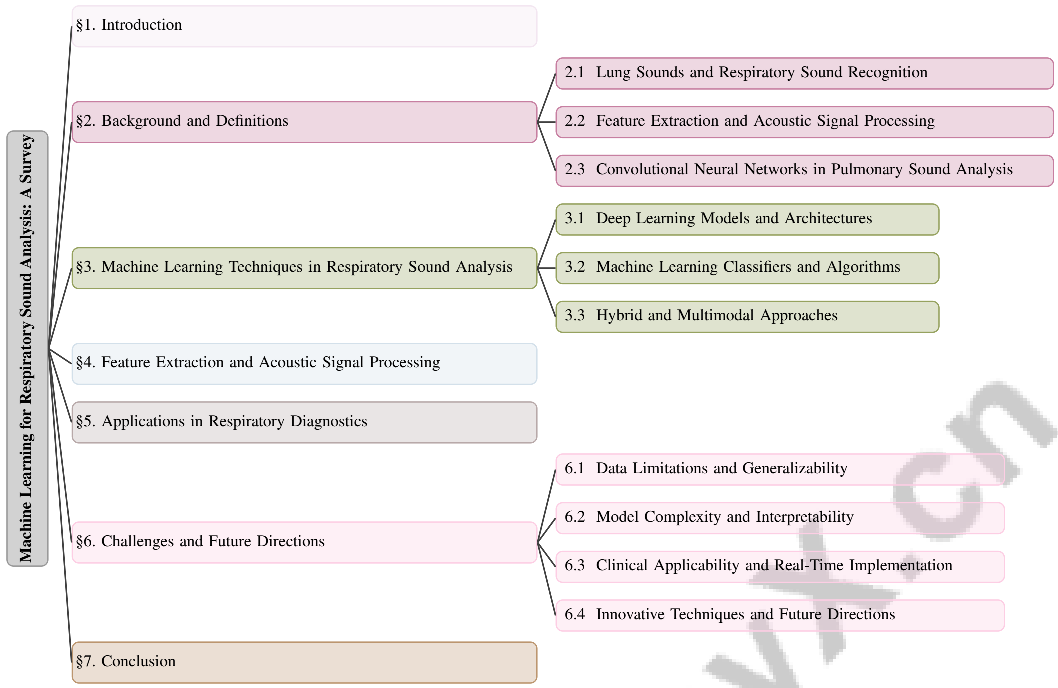


Figure 1: chapter structure

underscoring its significance in comprehensive health assessments. Moreover, effective data mining algorithms are crucial due to the challenges posed by limited labeled data in respiratory sound analysis, necessitating accurate labeling and analysis for advancing automated diagnostic systems.

## 1.2 Role of Machine Learning

Machine learning significantly enhances the accuracy and efficiency of respiratory sound analysis by automating the detection, classification, and diagnosis of lung sounds, thereby addressing the limitations of traditional auscultation methods. Techniques such as convolutional neural networks (CNNs) have been instrumental in this transformation, providing improved accuracy and efficiency in respiratory sound analysis [6, 3].

Innovative models like the audio-spectrogram vision transformer (AS-ViT) convert audio waveforms into visual spectrogram representations, facilitating the detection and categorization of abnormal lung sounds [7]. This method exemplifies how machine learning transforms raw audio data into actionable diagnostic insights. Additionally, multi-task learning frameworks enhance diagnostic efficiency by concurrently classifying lung sounds and related diseases, streamlining the diagnostic process [8].

Machine learning also improves the reliability and speed of diagnostic processes. For example, tensor-network machine learning (TN-ML) methods have enhanced the reliability of lung cancer screening, paralleling improvements in respiratory sound analysis [9]. Adaptive transfer learning algorithms, such as HMM-FLDA, refine the classification and segmentation of physiological event states from non-stationary data, showcasing machine learning's versatility in handling complex medical datasets [10].

Pre-trained models serve as efficient feature extractors, significantly reducing training time and computational demands, as demonstrated in pneumonia severity classification [1]. This efficiency is crucial in scenarios requiring rapid diagnosis. Furthermore, machine learning techniques have enhanced the quality of noisy neonatal chest sound recordings, illustrating their capability to improve data quality across various clinical contexts [11].

Machine learning frameworks, through advanced signal processing and feature extraction techniques, substantially enhance the detection and classification of respiratory sounds, contributing to more accurate diagnoses and efficient healthcare delivery [12].



AI-generated, for reference only.

### 1.3 Structure of the Survey

This survey is systematically organized to provide a comprehensive overview of the intersection between machine learning and respiratory sound analysis. Following the introduction, which establishes the significance of respiratory sound analysis and the transformative role of machine learning, the survey progresses into the **Background and Definitions** section. This section elucidates key concepts and terminologies, offering foundational knowledge essential for understanding subsequent discussions, including subsections on lung sounds and respiratory sound recognition, feature extraction and acoustic signal processing, and the application of convolutional neural networks in pulmonary sound analysis.

The next section, **Machine Learning Techniques in Respiratory Sound Analysis**, explores various methodologies employed in the field, examining deep learning models and architectures, machine learning classifiers and algorithms, as well as hybrid and multimodal approaches, highlighting their applications and effectiveness in respiratory diagnostics.

In the **Feature Extraction and Acoustic Signal Processing** section, the survey discusses critical methods that enhance the accuracy of machine learning models, including preprocessing and noise reduction, diverse feature extraction techniques, and the concept of end-to-end learning in respiratory sound analysis.

The survey then transitions to **Applications in Respiratory Diagnostics**, showcasing practical implementations of machine learning through case studies and examples, covering the classification and detection of respiratory sounds, the diagnosis of respiratory diseases, and the integration of respiratory sound analysis with machine learning models to improve diagnostic accuracy.

The penultimate section, **Challenges and Future Directions**, addresses current limitations and potential advancements in the field, discussing data limitations and generalizability, model complexity and interpretability, clinical applicability, and real-time implementation, concluding with an exploration of innovative techniques and future directions for research.

Finally, the **Conclusion** section synthesizes the survey's key points, emphasizing the impact of machine learning on respiratory diagnostics and the importance of continued research and development in this multidisciplinary domain.The following sections are organized as shown in Figure 1.

## 2 Background and Definitions

### 2.1 Lung Sounds and Respiratory Sound Recognition

Lung sounds, encompassing both normal and abnormal signals like crackles and wheezes, are crucial for diagnosing respiratory conditions such as asthma and COPD [13]. The traditional use of stethoscopes is often limited by noise interference and overlapping heart sounds, complicating diagnostics [11]. The subjective nature of auscultation, prone to inter-listener variability, necessitates automated systems for consistent and objective diagnostics [6, 14]. Deep learning models are increasingly employed to classify lung sounds and related diseases [15], with recurrent neural networks enhancing classification by identifying critical breathing phases [16]. Integrating lung sounds with other modalities, such as ultrasound images, enriches diagnostics and offers non-invasive assessments of conditions like pneumonia [1].

Datasets like ICBHI, containing diverse recordings, are pivotal for recognizing respiratory sounds [17]. The challenges in lung sound recognition echo those in speaker recognition, emphasizing the need for algorithms that handle variability and limited data [18]. The scarcity of comprehensive respiratory disease databases complicates analysis, highlighting the need for robust datasets for AI training and evaluation [19]. Addressing the benchmark for classifying respiratory diseases underscores the importance of differentiating between normal and abnormal sounds for accurate diagnostics [3]. These advancements are vital for timely and precise diagnostic outcomes, enhancing healthcare delivery [20].

### 2.2 Feature Extraction and Acoustic Signal Processing

Transforming raw respiratory audio into meaningful representations is crucial for accurate pulmonary condition classification. The non-linear, non-stationary nature of lung sounds necessitates advanced feature extraction techniques. Traditional methods like mel-frequency cepstral coefficients (MFCCs) are common but often insufficient for respiratory sounds [21]. Recent advancements include empirical mode decomposition (EMD) and discrete wavelet transform (DWT) for denoising, enhancing the clarity and quality of audio signals [22]. Lightweight CNN architectures combined with hybrid scalogram-based features exemplify innovative approaches [23], while unsupervised learning techniques highlight the importance of feature extraction from raw audio data [24, 25].

Combining various feature sets with classifiers such as SVM and Logistic Regression in a parallel structure enhances diagnostic system performance [26]. The challenges of time alignment across devices further underscore the complexities in multi-device speech processing [4]. These sophisticated feature extraction methods are essential for improving classification accuracy and system reliability in respiratory diagnostics.

### 2.3 Convolutional Neural Networks in Pulmonary Sound Analysis

Convolutional Neural Networks (CNNs) significantly enhance pulmonary sound analysis by automating and refining respiratory diagnostics. CNNs excel in processing and extracting features from complex acoustic signals, thereby improving diagnostic accuracy [27]. The multi-input CNN (MICNN) exemplifies the integration of respiratory cycle features for improved classification [13]. Current research advocates for tailored convolutional feature extraction approaches using Maximum Entropy (ME) and Signal-to-Noise Ratio (SNR) to enhance performance [28].

Integrating CNNs with clinical measurements, as demonstrated by KAMP-Net, showcases their potential in predictive analytics by combining CNN-extracted features with clinical data for mortality risk prediction [29]. CNNs streamline diagnostics by classifying lung sounds from electronic stethoscopes, reducing manual interpretation needs [6]. Hybrid models combining CNNs with LSTMs capture both spatial and temporal dynamics, enhancing respiratory sound analysis [3].

CNNs offer robust frameworks for feature extraction and classification, significantly improving the accuracy and efficiency of respiratory diagnostics. As deep learning research progresses, integrating CNNs with advanced feature extraction techniques and complementary models promises further improvements in diagnostic accuracy and patient care. This evolution is supported by CNNs' ability to extract complex features from medical images, as seen in studies fusing imaging data with clinical information for enhanced Computer-Aided Diagnosis (CAD). The application of diverse pre-trained CNN architectures and sophisticated feature selection methods has shown potential in detecting conditions like Acute Lymphoblastic Leukemia (ALL), highlighting the superiority of automated systems over traditional diagnostics. Optimizing CNN models based on input data complexity and quality aims to refine these systems, ultimately leading to better health outcomes [28, 14, 30, 31].

In recent years, the application of machine learning techniques in respiratory sound analysis has gained significant attention due to its potential to enhance diagnostic accuracy and patient outcomes. As outlined in Figure 2, this figure illustrates the hierarchical structure of these techniques, detailing the primary categories of deep learning models, machine learning classifiers, and hybrid approaches, along with their respective subcategories and advancements. This comprehensive overview not only highlights the diversity of methodologies employed in the field but also underscores the continuous evolution of these technologies, paving the way for future innovations in respiratory diagnostics.

## 3 Machine Learning Techniques in Respiratory Sound Analysis

### 3.1 Deep Learning Models and Architectures

Deep learning has revolutionized respiratory sound analysis through advanced models that enhance the classification and diagnosis of respiratory conditions. Central to this is the application of Convolutional Neural Networks (CNNs), which leverage hierarchical feature extraction to process complex pulmonary acoustics [6]. Methods like CNN-LSC effectively categorize lung sound recordings, improving diagnostic capabilities. The integration of CNNs with Long Short-Term Memory (LSTM) networks further enhances classification by capturing temporal dependencies, bypassing preprocessing through sound augmentation techniques [3].

Hybrid models such as CNN-LSTM, which combine convolutional layers for feature extraction with LSTM layers, exemplify CNNs' adaptability in managing temporal dependencies critical for accurate

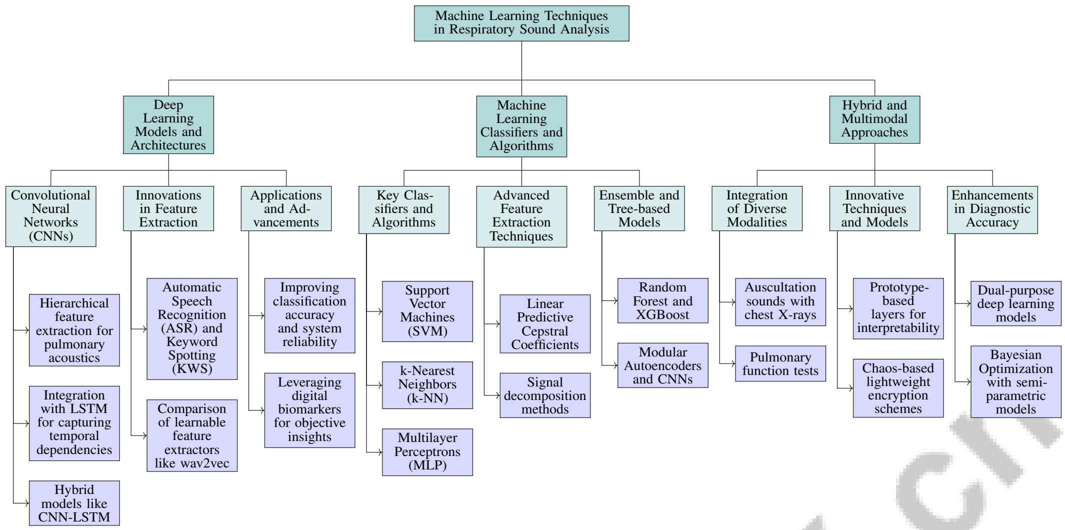


Figure 2: This figure illustrates the hierarchical structure of machine learning techniques in respiratory sound analysis, detailing the primary categories of deep learning models, machine learning classifiers, and hybrid approaches, along with their respective subcategories and advancements.

classification [25]. Additionally, KAMP-Net illustrates the integration of deep learning features with clinical measurements, showcasing CNNs' potential in predictive analytics [29].

Recent advancements in automatic speech recognition (ASR) and keyword spotting (KWS) highlight deep learning's transformative potential in feature extraction [4]. Vienting et al. emphasize the importance of novel techniques, comparing learnable feature extractors like wav2vec against traditional methods to improve model performance [24].

Exploring diverse deep learning models, including Random Forest and Decision Trees, within a unified framework enhances reproducibility and applicability, as demonstrated in systematic evaluations for coronary heart disease prediction [2]. These advancements underscore deep learning's critical role in improving classification accuracy and system reliability.

The integration of innovative methodologies and hybrid models—such as artificial intelligence and machine learning techniques for audio analysis—promises significant improvements in diagnostic accuracy and efficiency. By leveraging digital biomarkers from respiratory sounds and voice analysis, these advancements aim to provide objective insights that surpass traditional subjective assessments, facilitating remote and non-invasive evaluations and ultimately improving patient outcomes [32, 33, 34].

As illustrated in Figure 3, this figure encapsulates the hierarchical categorization of deep learning models and architectures utilized in respiratory sound analysis. It highlights key models such as CNN-LSC, CNN-LSTM, and KAMP-Net, while also emphasizing hybrid models and advanced feature extraction methods. The first example showcases a CNN architecture with multiple layers, emphasizing an ensemble hidden layer that integrates outputs from various convolutional layers to enhance learning. The second example presents a comparative analysis of model performances, where varying the number of layers impacts results, providing insights into optimization processes for effective respiratory sound analysis [35, 36].

### 3.2 Machine Learning Classifiers and Algorithms

Machine learning classifiers and algorithms are integral to respiratory sound analysis, employing advanced computational techniques to enhance classification and diagnosis. Table 1 presents a comprehensive overview of machine learning methods employed in respiratory sound analysis, illustrating the diversity of algorithm types, feature extraction techniques, and their application contexts. Classifiers such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Multilayer Perceptrons (MLP) effectively process complex acoustic signals in respiratory diagnostics

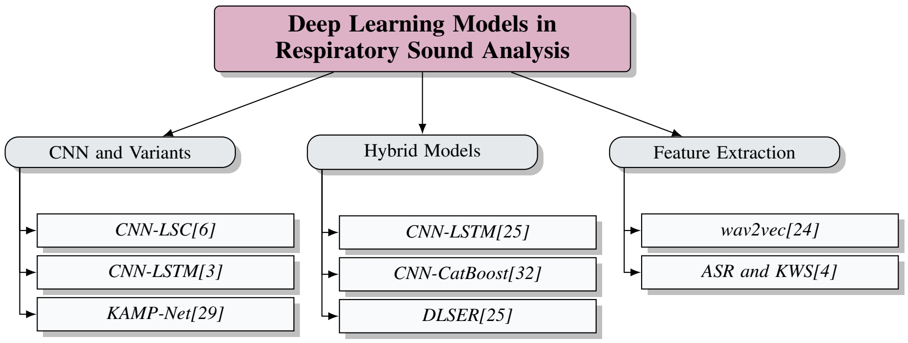


Figure 3: This figure illustrates the hierarchical categorization of deep learning models and architectures used in respiratory sound analysis, highlighting key models like CNN-LSC, CNN-LSTM, and KAMP-Net. It also emphasizes hybrid models and advanced feature extraction methods.

| Method Name | Algorithm Types | Feature Extraction | Application Context |
|---|---|---|---|
| MTL[8] | Deep Learning Architectures | Mel Frequency Cepstral | Respiratory Diagnostics |
| HMM-FLDA[10] | Hmm-FLDA | Fisher's Lda | Health Monitoring |
| RLUS-COVID[1] | Xception, Resnet50, Vgg16 | Deep Features Extraction | Clinical Settings |

Table 1: Overview of Machine Learning Methods for Respiratory Sound Analysis, detailing the algorithm types, feature extraction techniques, and application contexts. The table highlights the integration of deep learning architectures and advanced feature extraction methods to enhance diagnostic accuracy and efficiency in various clinical and health monitoring settings.

[8]. These classifiers support systems like the Lung Sound Classification Method (LSCM), which utilizes signal decomposition and feature extraction for robust classification outcomes [10].

Incorporating deep learning architectures such as Xception, ResNet50, and VGG16 as feature extractors enables rapid classification with minimal additional training, proving efficient in clinical settings [1]. This approach is particularly advantageous for swift diagnostic decisions, allowing rapid processing of respiratory sound data.

The HMM-FLDA method employs Fisher's linear discriminant analysis to adapt to changes in physiological data distribution, enhancing robustness in respiratory sound classification [10]. MobileNet achieves notable accuracy, with 74

Tree-based models like Random Forest and XGBoost enhance performance in small and imbalanced datasets typical in medical applications. The integration of ensemble methods with advanced models, such as Modular Autoencoders and CNNs, significantly improves classification performance by leveraging the strengths of multiple classifiers. Modular Autoencoders, for example, learn diverse representations from unlabelled data, improving feature extraction and enabling effective supervised learning. Additionally, semi-supervised learning techniques optimize classification outcomes across various applications, yielding substantial performance gains in benchmark datasets [37, 35, 38, 39, 31]. Prototype-based methods further enhance interpretability in neural networks, aiding respiratory sound classification by establishing a metric space based on distance to class prototypes.

The deployment of machine learning classifiers and algorithms in respiratory sound analysis substantially enhances diagnostic accuracy and efficiency. By integrating advanced feature extraction techniques, such as Linear Predictive Cepstral Coefficients and various signal decomposition methods, alongside robust classifiers like SVMs and MLPs, automated diagnostic systems for respiratory diseases are poised for significant performance improvements. These innovations facilitate accurate analysis of audio-based biomarkers, such as cough sounds and lung auscultation recordings, enabling timely identification of respiratory conditions and promoting the development of non-invasive screening tools, ultimately advancing respiratory diagnostics and patient care [40, 32, 41, 42, 34].

### 3.3 Hybrid and Multimodal Approaches

Hybrid and multimodal approaches are critical in enhancing respiratory sound analysis, integrating diverse data sources and computational techniques to improve diagnostic accuracy and efficiency.

Combining auscultation sounds with modalities like chest X-rays and pulmonary function tests exemplifies a hybrid approach that enriches diagnostics by providing a comprehensive view of respiratory health [43]. This multimedia database facilitates holistic assessments, improving detection and classification of respiratory conditions.

The Rene architecture demonstrates the potential of multimodal approaches, achieving up to 23

Prototype-based layers in neural networks introduce a geometrically interpretable classification approach, allowing for intuitive analysis of respiratory sounds [44]. Such innovations highlight the significance of developing models that achieve high accuracy while providing insights into decision-making processes.

Innovative techniques, such as chaos-based lightweight encryption schemes, could inspire hybrid approaches in respiratory sound analysis by ensuring data integrity and privacy [45]. This is crucial for developing robust systems capable of handling sensitive medical data.

The dual-purpose deep learning model exemplifies hybrid training benefits, utilizing mixed training sets of lung and tracheal sounds to enhance model robustness and adaptability [46]. This illustrates the potential of combining different respiratory sound types to improve classification outcomes.

Furthermore, integrating Bayesian Optimization with a semi-parametric residual Gaussian Process model enhances auscultation quality, showcasing the innovative use of hybrid models to tackle challenges in traditional auscultation methods [5].

Hybrid and multimodal approaches are vital for enhancing respiratory sound analysis, integrating advanced machine learning algorithms with diverse audio-based biomarkers to enable accurate identification and diagnosis of respiratory diseases. Leveraging audio data from various sources—cough sounds, lung and tracheal sounds, and voice abnormalities—improves diagnostic accuracy and facilitates timely interventions in respiratory health. Studies demonstrate that combining different audio analysis methods significantly enhances detection of respiratory symptoms, addressing limitations of traditional diagnostic tools and promoting a comprehensive understanding of respiratory conditions [7, 32, 42, 46, 34]. By integrating diverse data sources and computational techniques, these approaches enhance diagnostic system accuracy and reliability, ultimately improving patient outcomes and healthcare delivery.

# 4 Feature Extraction and Acoustic Signal Processing

## 4.1 Preprocessing and Noise Reduction

Preprocessing and noise reduction are critical in refining respiratory sound analysis, directly influencing the precision of machine learning applications. These processes ensure that features extracted from audio signals accurately reflect physiological events, thereby enhancing predictive capabilities. Techniques such as data normalization and quantum state mapping, as demonstrated by the TN-ML method, effectively minimize data variability before analysis [9]. The HMM-FLDA method underscores the importance of initializing models with baseline data, enhancing adaptability to data variations and improving classification accuracy [10]. Additionally, preprocessing methods like normalization, clipping, and padding prepare audio signals for deep learning frameworks, as seen in raw waveform processing evaluations [25].

Advanced techniques, including 3D point cloud scans, enhance sound data quality by filtering noise, thus providing comprehensive respiratory profiles [5]. Addressing data imbalance through oversampling and undersampling ensures models are trained on representative datasets, enhancing their generalizability and performance [2]. Effective preprocessing and noise reduction are vital for distinguishing normal from abnormal lung sounds, thereby improving machine learning classifier performance and enabling timely respiratory disease diagnoses. Techniques such as Empirical Mode Decomposition, Mel-frequency cepstral coefficients, and feature selection algorithms significantly enhance classification accuracy, proving indispensable in healthcare applications [40, 47, 41, 42, 34]. By employing normalization, filtering, and data augmentation, these processes enable accurate classification and diagnosis of respiratory conditions, ultimately improving healthcare outcomes.

## 4.2 Feature Extraction Techniques

Feature extraction is pivotal in respiratory sound analysis, transforming raw audio signals into structured representations for effective classification and diagnosis. Mel Frequency Cepstral Coefficients (MFCCs) are a key technique in this domain, particularly in multi-task learning approaches, due to their ability to capture essential spectral characteristics, addressing the non-linear and non-stationary nature of respiratory sounds [8, 21]. Advanced methodologies, such as spectrogram image generation, enhance classification accuracy by training CNNs to recognize intricate patterns associated with respiratory conditions [6]. This visual representation underscores the importance of robust feature extraction.

As illustrated in Figure 4, the categorization of feature extraction techniques in respiratory sound analysis emphasizes audio signal processing, graph neural networks, and multimodal approaches. Each category highlights specific methods such as MFCCs, spectrograms, GCNs, and multimodal networks, showcasing their application in various contexts including automatic speech recognition (ASR), graph classification, and medical imaging. The integration of graph convolutional networks (GCNs) for high-level feature extraction offers a novel approach, improving classification accuracy through logical graph predictions [38]. KAMP-Net exemplifies multimodal approaches by combining CNN-extracted imaging features with clinical knowledge to enhance mortality risk predictions in lung cancer patients, showcasing CNNs' versatility in extracting meaningful features from diverse data sources [29]. Hybrid neural network/hidden Markov model systems further emphasize optimizing feature extraction and classification performance, illustrating the potential of hybrid approaches in advancing respiratory sound analysis [24].

Exploring diverse feature extraction techniques is crucial for advancing respiratory sound analysis. Integrating advanced methodologies such as data augmentation and employing deep learning techniques that fuse medical imaging with clinical data enables the development of more accurate and efficient machine learning models. This progress leads to improved diagnostic capabilities and healthcare outcomes, especially in areas like medical imaging for conditions such as Alzheimer's and respiratory diseases, where timely intervention is critical [30, 34, 48].

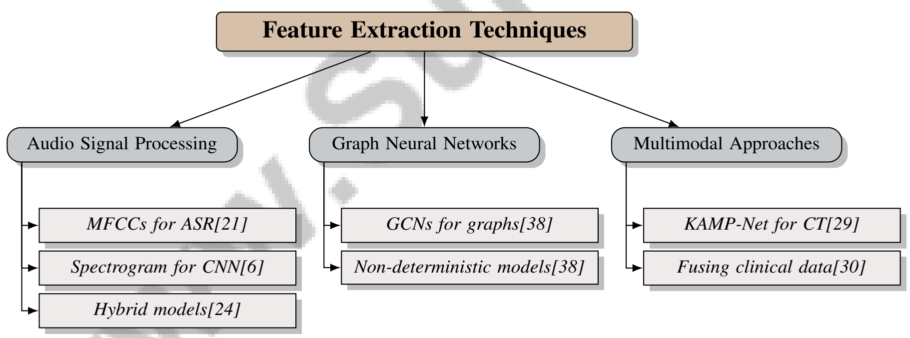


Figure 4: This figure illustrates the categorization of feature extraction techniques in respiratory sound analysis, emphasizing audio signal processing, graph neural networks, and multimodal approaches. Each category highlights specific methods such as MFCCs, spectrograms, GCNs, and multimodal networks, showcasing their application in various contexts including ASR, graph classification, and medical imaging.

## 4.3 End-to-End Learning and Feature Extraction

End-to-end learning has transformed respiratory sound analysis by enabling models to learn directly from raw audio inputs, eliminating the need for manual feature engineering. This approach employs deep learning architectures to automatically extract and optimize features, enhancing diagnostic accuracy and efficiency. WaDeNet, for example, uses wavelet decomposition for spectral feature extraction from speech signals, facilitating direct learning for classification tasks [49]. This method highlights the potential of wavelet transforms in lossless compression and effective feature extraction, improving model performance [50].

CNNs are integral to end-to-end learning frameworks, adept at identifying complex patterns for high accuracy in classifying breathing phases from spectrograms [51]. Their ability to compute



AI-generated, for reference only.

piecewise linear functions underscores their effectiveness in capturing intricate acoustic patterns [31]. Additionally, convolutional recurrent neural networks (CRNNs) process sequential data frame-by-frame through recurrent layers, computing features based on hidden states or outputs [52].

Innovative methods, such as trainable basis functions of spectrograms, exemplify advancements in end-to-end learning, enabling models to learn optimal representations directly from data [53]. This aligns with end-to-end learning principles, where feature extraction blocks process raw audio to generate time-frequency representations, subsequently classified by advanced neural networks [54].

End-to-end learning and automated feature extraction present a powerful framework for analyzing respiratory sounds, significantly reducing manual preprocessing needs. This advancement enhances machine learning models' performance in accurately identifying and diagnosing respiratory conditions, such as wheezes and crackles, based on audio biomarkers. Recent studies demonstrate high precision and recall rates, making these models reliable tools for clinical applications, including mobile diagnostics and remote monitoring, addressing the growing demand for efficient respiratory disease diagnosis amid global health challenges [40, 55, 34]. As research progresses, integrating novel techniques and architectures promises to further advance the field, ultimately improving patient care and healthcare delivery.

## 5 Applications in Respiratory Diagnostics

### 5.1 Classification and Detection of Respiratory Sounds

The classification and detection of respiratory sounds are pivotal in automated diagnostics, with machine learning significantly enhancing both accuracy and efficiency. Advanced models like the Multi-path CNN-BiGRU have demonstrated marked improvements in identifying common adventitious sounds, underscoring the efficacy of these architectures in respiratory sound analysis [16]. Comprehensive datasets, such as those comprising 680 lung sound clips for training and validation, ensure rigorous testing and reliability in clinical applications [56].

Innovative approaches, exemplified by BeamLearning, highlight the advantages of end-to-end deep learning frameworks, which are robust against measurement noise and capable of processing raw audio data without preprocessing—a critical feature in noisy clinical environments [20]. Hybrid deep neural network models, integrating CNNs with other architectures, have outperformed traditional methods in various applications, indicating significant potential for respiratory sound classification [21].

Recent advancements illustrate the transformative potential of machine learning methodologies in respiratory diagnostics, enabling accurate assessments of pulmonary conditions through audio-based biomarkers, such as cough sounds and lung acoustics. This shift addresses the limitations of traditional subjective interpretations, particularly highlighted during the COVID-19 pandemic, as machine learning continues to revolutionize respiratory health monitoring [32, 34]. By integrating advanced feature extraction techniques with cutting-edge models, these methodologies enhance the accuracy and reliability of automated diagnostic systems, ultimately improving patient outcomes and healthcare delivery.

### 5.2 Diagnosis of Respiratory Diseases

Machine learning has significantly advanced the diagnosis of respiratory diseases, enhancing both accuracy and efficiency in clinical settings. Techniques such as the TN-ML method have achieved over 98.5% concordance with pathological results, demonstrating their effectiveness in diagnosing lung cancer and other respiratory conditions [9]. Multi-task learning frameworks, which simultaneously classify lung sounds and identify related diseases, exemplify the potential of machine learning to streamline diagnostic processes [8].

CNN-based methods have proven effective in classifying lung sounds, with hybrid CNN-LSTM models notably outperforming traditional approaches, thereby improving diagnostic accuracy in clinical practice [6, 3]. The HMM-FLDA method further illustrates machine learning's potential in health monitoring by effectively classifying physiological event states [10]. Automated auscultation systems have also demonstrated the ability to capture diagnostic-quality heart and lung sounds comparable to traditional methods, reinforcing the role of machine learning in respiratory diagnostics [5].

Additionally, Random Forest models achieving 84% accuracy in predicting coronary heart disease highlight the versatility of machine learning algorithms in managing complex medical datasets [2]. The implementation of these sophisticated models enhances diagnostic capabilities and improves patient outcomes. As research progresses, the incorporation of diverse data sources, particularly audio-based biomarkers, alongside advanced machine learning techniques, is poised to further enhance disease identification and monitoring. This evolution is underscored by a growing body of studies utilizing audio analysis to recognize respiratory symptoms while addressing challenges such as environmental noise interference, ultimately aiming to reduce reliance on subjective clinical interpretation [32, 42, 27, 34, 48].

### 5.3 Integration with Machine Learning Models

Integrating respiratory sound analysis with machine learning models is crucial for advancing automated diagnostic capabilities. This integration employs sophisticated computational techniques to enhance the accuracy and efficiency of respiratory diagnostics. Hybrid models, which combine various machine learning frameworks, have demonstrated superior performance over traditional meta-learning methods. Effective loss functions and attention mechanisms further enhance these models, enabling them to excel in audio and speech processing tasks [37].

In respiratory diagnostics, merging machine learning models with respiratory sound analysis facilitates automatic classification and detection of respiratory conditions. Techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are utilized to process complex acoustic signals, extracting meaningful features essential for accurate diagnostics. These advanced models adeptly handle the non-linear and non-stationary characteristics of respiratory sounds, significantly improving diagnostic accuracy for various respiratory conditions. By leveraging audio-based biomarkers, these models effectively identify and classify critical respiratory symptoms, such as wheezes and crackles, even amidst background noise, providing reliable diagnostic outcomes for healthcare professionals [33, 57, 58, 42, 34].

The integration process often utilizes pre-trained models, such as CNNs and frameworks like wav2vec, serving as effective feature extractors. This approach enhances performance in tasks like image classification and automatic speech recognition (ASR) while alleviating the computational burden of training models from scratch. By leveraging pre-trained models, researchers achieve superior results with reduced computational resources and untranscribed data [24, 31]. This is particularly advantageous in scenarios demanding rapid diagnosis, facilitating swift processing of respiratory sound data.

Multi-task learning frameworks further enhance diagnostic efficiency by enabling simultaneous classification of lung sounds and associated diseases, streamlining the diagnostic process. Advanced deep learning models, such as MobileNet and ResNet50, extract relevant features from lung sound recordings, achieving high accuracy rates—74% for lung sound analysis and 91% for lung disease classification—using the ICBHI 2017 Respiratory Sound Database. By automating audio biomarker analysis, healthcare professionals can rely on objective, data-driven insights to improve diagnostic accuracy and reduce the burden of manual assessments [32, 59, 8, 34]. This integration not only enhances diagnostic accuracy but also contributes to more efficient healthcare delivery by minimizing the time and resources required for manual interpretation.

Integrating respiratory sound analysis with machine learning models represents a significant advancement in respiratory diagnostics. By leveraging advanced computational techniques and hybrid models, this integration improves the accuracy and reliability of diagnostic systems, ultimately enhancing patient outcomes and healthcare delivery. As research advances, the integration of innovative machine learning methodologies with diverse audio-based data sources, such as cough and lung sounds, is set to significantly enhance the accuracy and efficiency of respiratory disease diagnostics. This evolution is driven by the development of automated systems capable of analyzing complex audio signals to provide objective insights into pulmonary conditions, reducing reliance on subjective clinical assessments and improving early diagnosis and treatment outcomes. The ongoing exploration of digital biomarkers and publicly available datasets further expands the horizons of respiratory diagnostics, particularly in light of recent global health challenges like the COVID-19 pandemic [32, 27, 34, 58].

# 6 Challenges and Future Directions

The rapid evolution of respiratory sound analysis technologies presents substantial challenges, primarily in standardizing methodologies for audio-based diagnostics, integrating machine learning algorithms for sound classification, and mitigating environmental noise when detecting respiratory symptoms like wheezes and crackles [60, 32, 34]. These challenges include data limitations, model complexity, and clinical applicability. Understanding data limitations and their impact on model generalizability is crucial for developing robust models suitable for diverse clinical environments.

## 6.1 Data Limitations and Generalizability

Data limitations significantly impede the generalization of machine learning models in respiratory sound analysis. Small, imbalanced datasets can lead to overfitting and poor generalization [3]. The impracticality of acquiring large, well-labeled datasets, especially in urgent healthcare settings, further complicates this issue [1]. Uneven sample distribution across classes can degrade model performance, highlighting the necessity for balanced datasets [3].

Variability in audio quality from different recording devices poses additional challenges, potentially compromising classification accuracy [6]. This variability necessitates robust algorithms capable of operating in diverse acoustic environments. High noise levels in neonatal intensive care units, for example, complicate the reliability of chest sound recordings [11]. Conventional filtering methods often struggle to remove noise, such as White Gaussian Noise (WGN), complicating accurate signal recognition [61].

Methods reliant on specific datasets, like the PhysioNet Apnea-ECG dataset, limit their applicability across different scenarios. Domain mismatch and the need for validation across diverse populations and recording devices further challenge model generalizability. For instance, Vienting et al.'s benchmark focuses on English speech data, which may not translate effectively to other languages or dialects [24].

Addressing these challenges requires larger, more diverse datasets that reflect real-world conditions to enable effective model generalization [38]. Data augmentation and advanced feature extraction techniques can enhance the robustness and reliability of respiratory sound analysis systems. Resolving time synchronization issues and creating algorithms that accommodate varied acoustic environments are crucial for improving model generalization [4].

## 6.2 Model Complexity and Interpretability

Model complexity and interpretability present significant challenges in clinical applications of respiratory sound analysis. Complex architectures, such as convolutional neural networks (CNNs), are necessary to capture intricate respiratory sound patterns but can be computationally intensive, complicating their use in real-time applications and resource-limited environments [11]. While CNNs excel in feature extraction, their theoretical frameworks may not fully account for the complexities of real-world datasets, necessitating deeper architectures for more nuanced tasks [31].

The use of pre-trained models can introduce challenges in complexity and interpretability, particularly when the pre-trained model's quality does not align with specific tasks [13]. This reliance may hinder the adaptation of models to new, unseen data, impacting their generalizability in diverse clinical scenarios.

Hybrid models, such as CNN-TT, aim to enhance computational efficiency but may involve performance trade-offs compared to more complex models [62]. The interpretability of complex models remains a critical concern; prototype-based neural network layers can enhance interpretability by elucidating model decisions but may also increase execution times, limiting their applicability in real-time contexts [11]. Balancing model complexity and interpretability is essential for effective clinical application.

Advancements in computational techniques that improve respiratory sound analysis efficiency while maintaining diagnostic accuracy are crucial for enhancing healthcare outcomes, particularly in the timely identification and management of respiratory diseases. Recent studies emphasize integrating machine learning algorithms to analyze audio biomarkers from respiratory sounds, facilitating the detection of conditions such as coughs and wheezes, thereby reducing reliance on subjective clinician interpretation and promoting non-invasive diagnostic methods [40, 32, 42, 27, 34].

### 6.3 Clinical Applicability and Real-Time Implementation

The clinical applicability and real-time implementation of machine learning models in respiratory sound analysis face significant challenges, particularly in integrating advanced computational techniques within existing healthcare frameworks. Developing models that operate efficiently in real-time, providing immediate diagnostic insights without sacrificing accuracy, is a primary challenge. This requires lightweight models capable of rapid audio data processing, even in resource-constrained environments [11].

Integrating machine learning models into clinical workflows demands high reliability and robustness, particularly in managing the variability of respiratory sounds recorded in diverse clinical settings. Variations in acoustic environments, recording devices, and patient populations can significantly affect model performance, underscoring the need for robust algorithms capable of generalizing across these conditions [6]. This challenge is compounded by the necessity for models to handle noisy data effectively, as clinical environments often introduce substantial background noise that can degrade recorded sound quality [11].

Real-time implementation also requires model interpretability, providing clinicians with clear insights into decision-making processes. The complexity of deep learning models, such as CNNs, often limits interpretability, posing barriers to clinical adoption [31]. While prototype-based neural network layers enhance interpretability, they can also increase execution times, limiting real-time applicability [11]. Thus, developing models that balance interpretability and computational efficiency is vital for successful clinical implementation.

Potential solutions include hybrid models that leverage the strengths of various machine learning frameworks, enhancing both performance and interpretability [62]. Additionally, employing pre-trained models as feature extractors can reduce computational demands, allowing for faster processing without compromising accuracy [13]. Implementing robust preprocessing techniques can also enhance data quality, ensuring models receive clear input signals even in noisy clinical environments [11].

To effectively address the challenges of clinical applicability and real-time implementation, a multi-faceted approach is essential, one that accounts for healthcare's unique demands while incorporating advanced techniques such as data augmentation in medical imaging and innovative solutions for multi-device speech processing. These strategies can bolster the robustness and interoperability of machine learning models by utilizing diverse datasets, thereby enhancing performance in real-world clinical settings [4, 48]. By developing efficient and interpretable models and leveraging advanced preprocessing techniques, machine learning can significantly enhance the diagnostic capabilities of respiratory sound analysis, ultimately improving patient care and healthcare delivery.

### 6.4 Innovative Techniques and Future Directions

The advancement of respiratory sound analysis is poised to progress significantly through innovative techniques and strategic future research directions. Key areas of focus should include enhancing data annotation methods, exploring unsupervised and semi-supervised learning approaches, and developing models capable of generalizing across diverse datasets [63]. Expanding datasets to incorporate a broader range of conditions and patient populations is critical for improving model generalizability and performance, essential for validating methods across different languages and populations [21].

Future research should prioritize improving performance on low-certainty samples, vital for advancing machine learning applications in respiratory sound analysis [9]. Integrating real-time monitoring capabilities into multi-task learning frameworks alongside larger datasets can significantly enhance diagnostic processes [8]. Additionally, refining the use of pre-trained models for various respiratory



AI-generated, for reference only.

conditions and optimizing performance across different datasets will further improve diagnostic accuracy and efficiency [1].

Innovative feature extraction techniques, such as Maximum Entropy (ME) and Signal-to-Noise Ratio (SNR) metrics, should be examined across diverse datasets and neural network architectures. Investigating the temporal dynamics of data for dynamic applications can also contribute to developing more robust models [28]. Furthermore, exploring transformer compression techniques and developing lightweight models will enhance the applicability of machine learning in varied clinical settings [7].

The integration of hybrid approaches that combine multiple deep learning techniques presents promising avenues for future exploration. Such methods could improve model interpretability and performance, leading to more effective respiratory diagnostics [12]. Future research should also focus on validating hybrid models in diverse clinical contexts to ensure their applicability and effectiveness across various environments [62].

Additionally, enhancing sound separation in noisy environments through deep learning approaches is critical for advancing innovative techniques in respiratory sound analysis [11]. By addressing these areas, researchers can significantly propel the field forward, leading to more accurate and accessible diagnostic tools that improve patient outcomes and healthcare delivery. As research continues to evolve, the exploration of innovative techniques and strategic research directions promises to enhance the capabilities of machine learning models in respiratory diagnostics, ultimately contributing to superior healthcare solutions.

## 7 Conclusion

The application of machine learning in respiratory diagnostics has revolutionized the field, significantly enhancing the precision and efficiency of diagnosing respiratory ailments. With the advent of deep learning techniques, notably convolutional neural networks (CNNs), the classification and detection of respiratory sounds have become more accurate and reliable, addressing the inherent limitations of traditional auscultation by reducing inter-listener variability and ensuring consistent diagnostic results.

Furthermore, the integration of hybrid and multimodal strategies has enriched the analytical process by combining various data inputs and computational methodologies, thereby refining diagnostic precision. Advanced feature extraction techniques have been pivotal in crafting models capable of navigating the intricate nature of respiratory sounds, thus improving healthcare delivery.

Despite these advancements, challenges such as data scarcity, model complexity, and clinical applicability remain. The need for extensive, well-annotated datasets and the development of interpretable models capable of real-time operation are crucial areas for future exploration. Addressing these issues will require sustained innovation and cross-disciplinary collaboration to ensure that machine learning models can be effectively generalized across diverse clinical environments and patient demographics.

The potential of self-supervised pre-trained models to enhance classification tasks underscores the necessity for ongoing research and development, as demonstrated in studies focusing on voice quality classification. Harnessing the capabilities of these models could further bolster machine learning's role in respiratory diagnostics, paving the way for more accurate, efficient, and accessible healthcare solutions.

## References


[1] Pierre Raillard, Lorenzo Cristoni, Andrew Walden, Roberto Lazzari, Thomas Pulimood, Louis Grandjean, Claudia AM Gandini Wheeler-Kingshott, Yipeng Hu, and Zachary MC Baum. Rapid lung ultrasound covid-19 severity scoring with resource-efficient deep feature extraction, 2022.

[2] Jamal Al-Karaki, Philip Ilono, Sanchit Baweja, Jalal Naghiyev, Raja Singh Yadav, and Muhammad Al-Zafar Khan. Predicting coronary heart disease using a suite of machine learning models, 2024.

[3] Ali Mohammad Alqudah, Shoroq Qazan, and Yusra M Obeidat. Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds. *Soft Computing*, 26(24):13405–13429, 2022.

[4] Gregory Ciccarelli, Jarred Barber, Arun Nair, Israel Cohen, and Tao Zhang. Challenges and opportunities in multi-device speech processing, 2022.

[5] Yifan Zhu, Alexander Smith, and Kris Hauser. Automated heart and lung auscultation in robotic physical examinations, 2022.

[6] Murat Aykanat, Özkan Kılıç, Bahar Kurt, and Sevgi Saryal. Classification of lung sounds using convolutional neural networks. *EURASIP Journal on Image and Video Processing*, 2017:1–9, 2017.

[7] Whenty Ariyanti, Kai-Chun Liu, Kuan-Yu Chen, and Yu Tsao. Abnormal respiratory sound identification using audio-spectrogram vision transformer, 2024.

[8] Suma K V au2, Deepali Koppad, Preethi Kumar, Neha A Kantikar, and Surabhi Ramesh. Multi-task learning for lung sound  lung disease classification, 2024.

[9] Yu-Jia An, Sheng-Chen Bai, Lin Cheng, Xiao-Guang Li, Cheng en Wang, Xiao-Dong Han, Gang Su, Shi-Ju Ran, and Cong Wang. Intelligent diagnostic scheme for lung cancer screening with raman spectra data by tensor network machine learning, 2023.

[10] Xichen She, Yaya Zhai, Ricardo Henao, Christopher W. Woods, Christopher Chiu, Geoffrey S. Ginsburg, Peter X. K. Song, and Alfred O. Hero. Adaptive multi-channel event segmentation and feature extraction for monitoring health outcomes, 2020.

[11] Ethan Grooby, Chiranjibi Sitaula, Davood Fattahi, Reza Sameni, Kenneth Tan, Lindsay Zhou, Arrabella King, Ashwin Ramanathan, Atul Malhotra, Guy A. Dumont, and Faezeh Marzbanrad. Noisy neonatal chest sound separation for high-quality heart and lung sounds, 2022.

[12] Alyaa Hamel Sfayyih, Nasri Sulaiman, and Ahmad H Sabry. A review on lung disease recognition by acoustic signal analysis with deep learning networks. *Journal of big Data*, 10(1):101, 2023.

[13] Truc Nguyen and Franz Pernkopf. Crackle detection in lung sounds using transfer learning and multi-input convolitional neural networks, 2021.

[14] Sabit Ahamed Preanto, Md. Taimur Ahad, Yousuf Rayhan Emon, Sumaya Mustofa, and Md Alamin. A study on deep feature extraction to detect and classify acute lymphoblastic leukemia (all), 2024.

[15] Dat Ngo, Lam Pham, Anh Nguyen, Ben Phan, Khoa Tran, and Truong Nguyen. Deep learning framework applied for predicting anomaly of respiratory sounds, 2020.

[16] Fu-Shun Hsu, Shang-Ran Huang, Chien-Wen Huang, Chao-Jung Huang, Yuan-Ren Cheng, Chun-Chieh Chen, Jack Hsiao, Chung-Wei Chen, Li-Chin Chen, Yen-Chun Lai, Bi-Fang Hsu, Nian-Jhen Lin, Wan-Lin Tsai, Yi-Lin Wu, Tzu-Ling Tseng, Ching-Ting Tseng, Yi-Tsun Chen, and Feipei Lai. Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database-hf$_lung_v1$, 2022.

[17] Siddhartha Gairola, Francis Tom, Nipun Kwatra, and Mohit Jain. Respirenet: A deep neural network for accurately detecting abnormal lung sounds in limited data setting, 2021.

[18] Zhongxin Bai and Xiao-Lei Zhang. Speaker recognition based on deep learning: An overview, 2021.

[19] Rizwana Zulfiqar, Fiaz Majeed, Rizwana Irfan, Hafiz Tayyab Rauf, Elhadj Benkhelifa, and Abdelkader Nasreddine Belkacem. Abnormal respiratory sounds classification using deep cnn through artificial noise addition. *Frontiers in medicine*, 8:714811, 2021.

[20] Hadrien Pujol, Éric Bavu, and Alexandre Garcia. Beamlearning: an end-to-end deep learning approach for the angular localization of sound sources using raw multichannel acoustic pressure data, 2021.

[21] Jawid Ahmad Baktash and Mursal Dawodi. Employing hybrid deep neural networks on dari speech, 2023.

[22] Syed Zohaib Hassan Naqvi and Mohammad Ahmad Choudhry. An automated system for classification of chronic obstructive pulmonary disease and pneumonia patients using lung sound analysis. *Sensors*, 20(22):6512, 2020.

[23] Samiul Based Shuvo, Shams Nafisa Ali, Soham Irtiza Swapnil, Taufiq Hasan, and Mohammed Imamul Hassan Bhuiyan. A lightweight cnn model for detecting respiratory diseases from lung auscultation sounds using emd-cwt-based hybrid scalogram, 2020.

[24] Peter Vieting, Christoph Lüscher, Wilfried Michel, Ralf Schlüter, and Hermann Ney. On architectures and training for raw waveform feature extraction in asr, 2021.

[25] Zeynep Hilal Kilimci, Ulku Bayraktar, and Ayhan Kucukmanisa. Evaluating raw waveforms with deep learning frameworks for speech emotion recognition, 2023.

[26] Kerlos Atia Abdalmalak and Ascensión Gallardo-Antol'in. Enhancement of a text-independent speaker verification system by using feature combination and parallel-structure classifiers, 2024.

[27] Alyaa Hamel Sfayyih, Ahmad H Sabry, Shymaa Mohammed Jameel, Nasri Sulaiman, Safanah Mudheher Raafat, Amjad J Humaidi, and Yasir Mahmood Al Kubaiaisi. Acoustic-based deep learning architectures for lung disease diagnosis: a comprehensive overview. *Diagnostics*, 13(10):1748, 2023.

[28] Nidhi Gowdra, Roopak Sinha, and Stephen MacDonell. Examining convolutional feature extraction using maximum entropy (me) and signal-to-noise ratio (snr) for image classification, 2021.

[29] Hengtao Guo, Uwe Kruger, Ge Wang, Mannudeep K. Kalra, and Pingkun Yan. Knowledge-based analysis for mortality prediction from ct images, 2019.

[30] Songxiao Yang, Xiabi Liu, Zhongshu Zheng, Wei Wang, and Xiaohong Ma. Fusing medical image features and clinical features with deep learning for computer-aided diagnosis, 2021.

[31] Vinoth Nandakumar, Arush Tagade, and Tongliang Liu. Why do cnns excel at feature extraction? a mathematical explanation, 2023.

[32] Wei Zhang and Xiaoyan Liu. Decoding breath: Machine learning advancements in diagnosing pulmonary diseases via lung sound analysis. *Scientific and Engineering Letters*, 12(1):1–11, 2024.

[33] Yoonjoo Kim, YunKyong Hyon, Sunju Lee, Seong-Dae Woo, Taeyoung Ha, and Chaeuk Chung. The coming era of a new auscultation system for analyzing respiratory sounds. *BMC Pulmonary Medicine*, 22(1):119, 2022.

[34] Panagiotis Kapetanidis, Fotios Kalioras, Constantinos Tsakonas, Pantelis Tzamalis, George Kontogiannis, Theodora Karamanidou, Thanos G Stavropoulos, and Sotiris Nikoletseas. Respiratory diseases diagnosis using audio analysis and artificial intelligence: a systematic review. *Sensors*, 24(4):1173, 2024.

[35] Henry W J Reeve and Gavin Brown. Modular autoencoders for ensemble feature extraction, 2015.

[36] Kanru Hua. Do wavenets dream of acoustic waves?, 2018.

[37] Athul Raimon, Shubha Masti, Shyam K Sateesh, Siyani Vengatagiri, and Bhaskarjyoti Das. Meta-learning in audio and speech processing: An end to end comprehensive review, 2024.

[38] Masoud Kargar, Nasim Jelodari, and Alireza Assadzadeh. Multistage non-deterministic classification using secondary concept graphs and graph convolutional networks for high-level feature extraction, 2024.

[39] Aoran Shen, Minghao Dai, Jiacheng Hu, Yingbin Liang, Shiru Wang, and Junliang Du. Leveraging semi-supervised learning to enhance data mining for image classification under limited labeled data, 2024.

[40] Morten Grønnesby, Juan Carlos Aviles Solis, Einar Holsbø, Hasse Melbye, and Lars Ailo Bongo. Feature extraction for machine learning based crackle detection in lung sounds from a health survey, 2017.

[41] Andrine Elsetrønning, Adil Rasheed, Jon Bekker, and Omer San. On the effectiveness of signal decomposition, feature extraction and selection on lung sound classification, 2020.

[42] Himadri Mukherjee, Priyanka Sreerama, Ankita Dhar, Sk Md Obaidullah, Kaushik Roy, Mufti Mahmud, and KC Santosh. Automatic lung health screening using respiratory sounds. *Journal of Medical Systems*, 45:1–9, 2021.

[43] Gokhan Altan, Yakup Kutlu, Yusuf Garbi, Adnan Ozhan Pekmezci, and Serkan Nural. Multimedia respiratory database (respiratorydatabase@tr): Auscultation sounds and chest x-rays, 2021.

[44] Sascha Saralajew, Lars Holdijk, Maike Rees, and Thomas Villmann. Prototype-based neural network layers: Incorporating vector quantization, 2019.

[45] Awais Aziz Shah, Ahsan Adeel, Jawad Ahmad, Ahmed Al-Dubai, Mandar Gogate, Abhijeet Bishnu, Muhammad Diyan, Tassadaq Hussain, Kia Dashtipour, Tharm Ratnarajah, and Amir Hussain. A novel chaos-based light-weight image encryption scheme for multi-modal hearing aids, 2022.

[46] Fu-Shun Hsu, Shang-Ran Huang, Chang-Fu Su, Chien-Wen Huang, Yuan-Ren Cheng, Chun-Chieh Chen, Chun-Yu Wu, Chung-Wei Chen, Yen-Chun Lai, Tang-Wei Cheng, Nian-Jhen Lin, Wan-Ling Tsai, Ching-Shiang Lu, Chuan Chen, and Feipei Lai. A dual-purpose deep learning model for auscultated lung and tracheal sound analysis based on mixed set training, 2023.

[47] Burcu Acar Demirci, Yücel Koçyiğit, Deniz Kızılırmak, and Yavuz Havlucu. Adventitious and normal respiratory sound analysis with machine learning methods. *Celal Bayar University Journal of Science*, 18(2):169–180, 2022.

[48] Manuel Cossio. Augmenting medical imaging: A comprehensive catalogue of 65 techniques for enhanced data analysis, 2023.

[49] Prithvi Suresh and Abhijith Ragav. Wadenet: Wavelet decomposition based cnn for speech processing, 2020.

[50] Philippe Reiter. Cloud detection through wavelet transforms in machine learning and deep learning, 2020.

[51] Cristina Jácome, Johan Ravn, Einar Holsbø, Juan Carlos Aviles-Solis, Hasse Melbye, and Lars Ailo Bongo. Convolutional neural network for breathing phase detection in lung sounds, 2019.

[52] Gil Keren and Björn Schuller. Convolutional rnn: an enhanced model for extracting features from sequential data, 2017.

[53] Kwan Yee Heung, Kin Wai Cheuk, and Dorien Herremans. Understanding audio features via trainable basis functions, 2022.

[54] Emre Çakır and Tuomas Virtanen. End-to-end polyphonic sound event detection using convolutional recurrent neural networks with learned time-frequency representation input, 2018.

[55] Morten Grønnesby, Juan Carlos Aviles Solis, Einar Holsbø, Hasse Melbye, and Lars Ailo Bongo. Feature extraction for machine learning based crackle detection in lung sounds from a health survey. *arXiv preprint arXiv:1706.00005*, 2017.

[56] Ji Soo Park, Kyungdo Kim, Ji Hye Kim, Yun Jung Choi, Kwangsoo Kim, and Dong In Suh. A machine learning approach to the development and prospective evaluation of a pediatric lung sound classification model. *Scientific Reports*, 13(1):1289, 2023.

[57] Pengfei Zhang, Zhihang Zheng, Shichen Zhang, Minghao Yang, and Shaojun Tang. Rene: A pre-trained multi-modal architecture for auscultation of respiratory diseases, 2024.

[58] Negin Melek. Responding to challenge call of machine learning model development in diagnosing respiratory disease sounds, 2021.

[59] Fu-Shun Hsu, Shang-Ran Huang, Chien-Wen Huang, Chun-Chieh Chen, Yuan-Ren Cheng, and Feipei Lai. Multi-path convolutional neural networks efficiently improve feature extraction in continuous adventitious lung sound detection, 2021.

[60] Renard Xaviero Adhi Pramono, Stuart Bowyer, and Esther Rodriguez-Villegas. Automatic adventitious respiratory sound analysis: A systematic review. *PloS one*, 12(5):e0177926, 2017.

[61] Angkoon Phinyomark, Chusak Limsakul, and Pornchai Phukpattaranont. A novel feature extraction for robust emg pattern recognition, 2009.

[62] Jun Qi, Hu Hu, Yannan Wang, Chao-Han Huck Yang, Sabato Marco Siniscalchi, and Chin-Hui Lee. Exploring deep hybrid tensor-to-vector network architectures for regression based speech enhancement, 2020.

[63] Yan Tian, Zhaocheng Xu, Yujun Ma, Weiping Ding, Ruili Wang, Zhihong Gao, Guohua Cheng, Linyang He, and Xuran Zhao. Survey on deep learning in multimodal medical imaging for cancer detection, 2023.





AI-generated, for reference only.

**Disclaimer:**

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.