

# Artificial Intelligence for Medical Sound Analysis: Current Advances, Clinical Applications, and Future Directions - Focus on Cardiac and Respiratory Sounds

Emmanuel Andres<sup>1,2\*</sup>, Samy Talha<sup>2,3</sup>, Christian Brandt<sup>3</sup>, Amir El Hassani Hajjam<sup>4</sup> and Noel Lorenzo Villalba<sup>1</sup>

<sup>1</sup>Department of Internal Medicine, University Hospitals of Strasbourg, Strasbourg, France

<sup>2</sup>Faculty of Medicine, University of Strasbourg, Strasbourg, France

<sup>3</sup>Department Cardiology, University Hospitals of Strasbourg, Strasbourg, France

<sup>4</sup>Nanotechnology and Medical Innovative Applications Laboratory, University of Technology of Belfort-Montbéliard (UTBM), Belfort, France

\*Correspondence to: Emmanuel Andres<sup>1,2</sup>, <sup>1</sup>Department of Internal Medicine, University Hospitals of Strasbourg, Strasbourg, France; <sup>2</sup>Faculty of Medicine, University of Strasbourg, Strasbourg, France, E-mail: emmanuel.andres@chru-strasbourg.fr

Received: January 31, 2026; Manuscript No: JAID-26-6898; Editor Assigned: February 04, 2026; PreQc No: JAID-26-6898 (PQ); Reviewed: February 18, 2026; Revised: February 20, 2026; Manuscript No: JAID-26-6898 (R); Published: March 02, 2026

Citation: Andres E, Talha S, Brandt C, Hajjam AEIH, Villalba NL (2026). Artificial Intelligence for Medical Sound Analysis: Current Advances, Clinical Applications, and Future Directions - Focus on Cardiac and Respiratory Sounds. J. Artif. Intell. Digit. Health. Vol.1 Iss.1, March (2026), pp:23-30.

Copyright: © 2026 Emmanuel Andres, Samy Talha, Christian Brandt, Amir El Hassani Hajjam, Noel Lorenzo Villalba. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Cardiac and respiratory auscultation remain foundational clinical tools for diagnosing a wide range of cardiopulmonary diseases. However, reliance on clinician expertise and subjective interpretation limits diagnostic accuracy and reproducibility. Recent advances in artificial intelligence (AI), particularly deep learning, have transformed automated acoustic analysis, enabling objective, accurate, and scalable diagnostic support. This review synthesizes current AI methodologies applied to heart and respiratory sound analysis, highlights key clinical applications, addresses challenges including data heterogeneity and model interpretability, and outlines future research directions. We emphasize the transformative potential of AI-powered auscultation to enhance personalized, accessible, and proactive cardiopulmonary care.

**Keywords:** Artificial Intelligence; Deep Learning; Heart Sounds; Lung Sounds; Auscultation; Cardiac Diagnostics; Respiratory Diagnostics; Explainable AI; Telemedicine; Multi-Organ Acoustic Analysis

## INTRODUCTION

Auscultation of medical sounds, particularly, heart and lung sounds, is a cornerstone of clinical examination, providing critical insights into cardiovascular and pulmonary health. Heart sounds including the first (S1) and second (S2) heart sounds and pathological murmurs reflect valvular function and hemodynamics, while respiratory sounds such as wheezes, crackles, and breath sounds indicate airway and parenchymal lung conditions [1]. Despite their clinical importance, traditional auscultation is limited by inter-observer variability and dependence on clinician expertise, leading to inconsistent diagnostic accuracy.

Artificial intelligence (AI), encompassing machine learning (ML) and deep learning (DL), offers powerful tools to analyze complex acoustic signals objectively. By extracting and learning from intricate features in heart and lung sounds, AI models can classify normal and pathological states, detect subtle

abnormalities, and support clinical decision-making. Advances in digital stethoscopes, signal processing, and computational power have accelerated AI applications in this domain, promising to transform cardiopulmonary diagnostics.

This review synthesizes current AI methodologies applied to heart and respiratory sound analysis, highlights key clinical applications, discusses challenges including data heterogeneity and model interpretability, and outlines future research directions.

The overall roadmap for the clinical translation of AI-based biological sound analysis is illustrated in Figure 1.

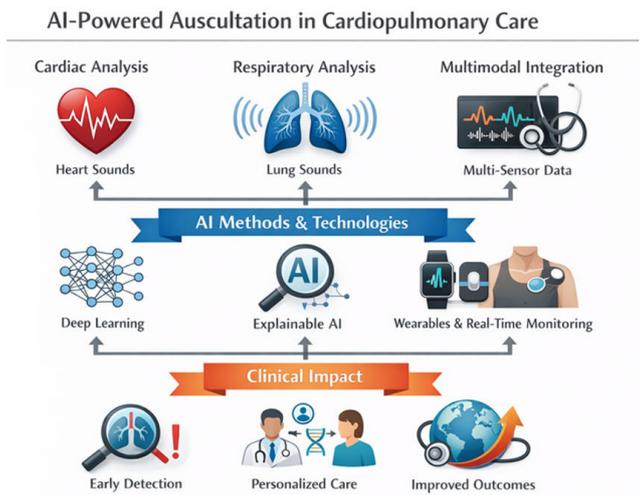


Figure 1: AI-Powered Auscultation in Cardiopulmonary Care

MAIN REVIEW CONTENT

Advanced Signal Processing and Machine Learning for Auscultatory Sound Analysis

High-quality sound acquisition is foundational for automated analysis of heart and lung sounds. Digital stethoscopes and electronic acoustic sensors enable high-fidelity recording but are frequently challenged by environmental noise, motion artifacts, and patient-related interference in real-world clinical settings [2]. To address these limitations, preprocessing pipelines typically incorporate noise reduction techniques such as wavelet-based denoising, adaptive filtering, and increasingly, DL-based audio enhancement models that outperform traditional filters in non-stationary noise environments (Table 1) [3]. These approaches improve signal-to-noise ratio while preserving diagnostically relevant acoustic features.

Accurate segmentation of physiological sound events is a critical preprocessing step. In cardiac auscultation, segmentation of heart sounds into S1, systole, S2, and diastole allows focused analysis of murmurs and abnormal flow patterns. Classical methods include envelope-based thresholding and hidden Markov models (HMMs), while recent DL-based segmentation algorithms leveraging convolutional and recurrent architectures have demonstrated higher robustness across variable heart rates and pathological conditions [3]. Similar segmentation strategies are applied to respiratory cycles to isolate inspiration and expiration phases for lung sound analysis.

Feature extraction remains a central component in auscultatory signal analysis. Time-domain features such as signal amplitude, duration, zero-crossing rate, and energy provide insights into sound intensity and temporal structure. Frequency-domain

features including dominant frequencies, spectral centroid, bandwidth, and spectral power distribution capture pathological shifts in acoustic energy associated with wheezes, crackles, and murmurs. Time-frequency representations, particularly spectrograms and Mel-frequency cepstral coefficients (MFCCs), have become the most widely used features due to their ability to represent non-stationary biomedical signals in a compact and interpretable form [4].

Traditional ML classifiers, including support vector machines (SVMs), random forests, k-nearest neighbors, and gradient boosting models, have been extensively applied to handcrafted features with moderate success. These methods offer interpretability and computational efficiency but rely heavily on expert-driven feature engineering. In contrast, DL approaches particularly convolutional neural networks (CNNs) have demonstrated superior diagnostic performance by automatically learning hierarchical representations directly from spectrogram images, effectively capturing subtle acoustic patterns that may be imperceptible to human listeners [5].

Recurrent neural networks (RNNs), especially long short-term memory (LSTM) architectures, are well suited for modeling temporal dependencies inherent in sequential audio signals. Their application has significantly improved performance in respiratory disease detection by capturing long-range temporal relationships across breathing cycles [6]. More recently, transformer-based architectures such as HuBERT and Wav2Vec 2.0 have shown strong performance in self-supervised audio representation learning, enabling effective modeling of long-range dependencies with reduced reliance on labeled data. Although initially applied to speech and bowel sound analysis, these models are increasingly recognized as promising tools for heart and lung sound analysis [7].

A persistent challenge in biomedical audio analysis is the limited availability of large, annotated clinical datasets. Data augmentation techniques such as pitch shifting, time stretching, additive noise injection, and horizontal flipping of spectrograms are commonly employed to enhance model generalizability. Transfer learning from large-scale general audio datasets has further mitigated data scarcity, allowing pretrained models to adapt efficiently to medical auscultation tasks [8].

To promote clinical adoption, explainable artificial intelligence (XAI) methods have gained increasing attention. Techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM), saliency maps, and attention mechanisms enable visualization of model focus on specific time frequency regions associated with pathological sounds. These approaches enhance interpretability, support clinical validation, and foster trust among healthcare professionals by aligning model predictions with known physiological and pathological acoustic signatures [3,9].

Pipeline Stage	Techniques / Models	Scientific Purpose	Representative References
Sound Acquisition	Digital stethoscopes, electronic biosensors (2-44.1 kHz)	High-fidelity capture of cardiopulmonary sounds	[4]
Noise Reduction	Wavelet denoising, adaptive filtering, DL-based enhancement (U-Net, autoencoders)	Suppress ambient and motion noise	[3]

Segmentation	HMMs, CNN-LSTM models, attention-based DL	Isolate S1/S2 and respiratory phases	[10]
Feature Extraction	Time-domain, frequency-domain, MFCCs, spectrograms	Encode diagnostic acoustic patterns	[4]
Traditional ML	SVM, random forest, gradient boosting	Classification using handcrafted features	[5]
Deep Learning	CNNs, RNNs (LSTM/GRU)	Automated hierarchical feature learning	[5]
Transformer Models	Wav2Vec 2.0, HuBERT	Long-range dependency modeling, self-supervised learning	[7]
Data Augmentation	Pitch shifting, time stretching, noise injection	Improve generalization, reduce overfitting	[8]
Explainable AI	Grad-CAM, attention heatmaps	Clinical interpretability and trust	[9]

**Table 1:** AI pipeline for heart and lung sound analysis

**Legend:** This table summarizes the main stages involved in AI-based auscultatory sound analysis, including sound acquisition, preprocessing, feature extraction, machine learning and deep learning methods, data augmentation strategies, and explainable artificial intelligence (XAI) techniques. For each stage, commonly used methods, their primary clinical or computational purpose, and representative references are presented. The table highlights the progression from traditional signal processing and handcrafted feature-based machine learning to advanced deep learning and transformer-based architectures, emphasizing their role in improving diagnostic accuracy, robustness, and clinical interpretability.

**Abbreviations:** AI, artificial intelligence; CNN, convolutional neural network; DL, deep learning; ECG, electrocardiogram; HMM, hidden Markov model; LSTM, long short-term memory; MFCC, Mel-frequency cepstral coefficients; ML, machine learning; RNN, recurrent neural network; S1, first heart sound; S2, second heart sound; XAI, explainable artificial intelligence.

### Clinical Applications of AI-Based Auscultatory Sound Analysis

#### Heart Sound Analysis

AI-based heart sound analysis has rapidly evolved into a highly accurate and clinically relevant tool for the early detection and characterization of cardiovascular diseases. Deep learning and advanced machine learning models have demonstrated robust performance across a broad spectrum of cardiac pathologies, addressing the limitations of conventional auscultation that relies heavily on clinician experience.

Recent work by Satya Karthikeya highlighted the importance of domain-specific feature engineering in heart sound classification [11]. By combining ten engineered acoustic features (including MFCCs, wavelet transforms, chroma features, spectral centroid, and temporal flatness) with ensemble learning, the authors demonstrated that XGBoost achieved a classification accuracy of 99%, outperforming both hybrid CNN-LSTM architectures (98%) and transformer-based

Wav2Vec embeddings. These findings underscore that, despite the rise of end-to-end deep learning, carefully designed acoustic features remain highly effective for early cardiovascular disease detection, particularly in multi-class classification tasks.

Structural heart disease detection has also benefited significantly from AI. Wang developed a temporal attentive pooling convolutional recurrent neural network (TAP-CRNN) for the automatic recognition of systolic murmurs associated with ventricular septal defects (VSDs) [12]. Using heart sound recordings from 51 VSD patients and 25 controls, the model achieved a sensitivity of 96.0% and a specificity of 96.7%. Notably, when analyzing recordings from the second aortic and tricuspid auscultation areas, both sensitivity and specificity reached 100%, demonstrating the model’s capacity to accurately detect subtle murmur patterns that are often challenging for less experienced clinicians.

Beyond murmur detection, AI-based auscultation has expanded toward functional cardiac assessment. Machine learning models applied to epicardial accelerometer and heart sound data have enabled non-invasive estimation of left ventricular pressure and contractility with accuracies exceeding 90%, providing valuable insights into cardiac performance without invasive catheterization [13]. Such approaches hold promise for continuous cardiac monitoring and early identification of hemodynamic deterioration.

Importantly, the integration of AI into low-cost digital stethoscopes has facilitated real-world clinical translation. Lightweight CNN-based models deployed on embedded hardware have demonstrated near-perfect classification of multiple cardiac conditions, supporting use in primary care, telemedicine, and resource-limited settings [6]. Attention mechanisms and explainable AI techniques further enhance clinician trust by highlighting acoustically relevant regions associated with pathological heart sounds [3].

Collectively, these studies demonstrate that AI-powered heart sound analysis can achieve expert-level diagnostic performance, support early disease detection, reduce inter-observer variability, and expand access to high-quality cardiovascular screening particularly in underserved populations (Table 2).

Study / Year	Population /Dataset	AI Method /Features	Clinical Application	Performance	Notes
[3]	Various heart sound datasets	CNN + attention	Detection of normal vs. pathological heart sounds; valvular disease classification	Accuracy >97%	Differentiates aortic stenosis, mitral regurgitation; robust in noisy environments
[11]	Multi-class heart sounds (5 categories)	XGBoost & CNN + LSTM; engineered acoustic features (Log Mel, MFCC, delta, delta-delta, chroma, DWT, ZCR, energy, spectral centroid, temporal flatness)	Early cardiovascular disease detection	XGBoost: 99% accuracy; CNN + LSTM: 98% accuracy	Emphasizes domain-specific feature engineering; multi-class classification
[12]	51 VSD patients, 25 healthy subjects	TAP-CRNN (temporal attentive pooling-convolutional RNN)	Automatic recognition of systolic murmurs in VSD	Sensitivity: 96.0%; Specificity: 96.7%; Second aortic & tricuspid area: 100%/100%	Enables structural heart disease murmur detection; aids inexperienced clinicians
[13]	Epicardial accelerometer data	Machine learning (unspecified)	Estimation of LV pressure and contractility	Accuracy >90%	Non-invasive cardiac function monitoring
[10]	Low-cost digital stethoscope recordings	Lightweight AI models	Detection of multiple cardiac & respiratory conditions	Near-perfect classification	Supports deployment in resource-limited settings

**Table 2:** Artificial intelligence (AI) approaches for heart sound analysis

**Abbreviations:** CNN: Convolutional Neural Network; LSTM: Long Short-Term Memory network; TAP-CRNN: Temporal Attentive Pooling-Convolutional Recurrent Neural Network; XGBoost: eXtreme Gradient Boosting; MFCC: Mel-Frequency Cepstral Coefficients; DWT: Discrete Wavelet Transform; ZCR: Zero-Crossing Rate; LV: Left Ventricular.

**Respiratory Sound Analysis**

AI-driven respiratory sound analysis represents one of the most mature applications of digital auscultation. DL models particularly LSTM and CNN-LSTM hybrid architectures—have achieved classification accuracies exceeding 98% for common respiratory diseases such as pneumonia, chronic obstructive pulmonary disease (COPD), asthma, and COVID-19, using digital stethoscope recordings alone (Table 3) [6].

Reported sensitivities range from 95–99%, with specificities between 93–98%, and NPVs consistently above 96%, making these systems particularly effective as screening and rule-out tools in emergency and primary care settings. The incorporation of DL-based audio enhancement techniques significantly improves robustness in real-world clinical environments, with signal-to-noise ratio improvements of up to 6–10 dB and corresponding increases in diagnostic confidence and model stability [14].

Interstitial Lung Disease (ILD) Detection: AI-based auscultation has emerged as a promising tool for the early detection and subtyping of ILD, a group of chronic pulmonary disorders characterized by alveolar and interstitial damage [15]. Lung crackles, often subtle and under-recognized, provide acoustic biomarkers for diagnosis. AI models analyzing these pathological sounds have demonstrated sensitivity and specificity comparable to standard imaging and specialist interpretation, enabling real-time identification of ILD subtypes and prognostic assessment. Such approaches are non-invasive, radiation-free, cost-effective, and particularly useful in resource-limited settings. By correlating specific crackle patterns with disease subtype and location, AI enables timely intervention, potentially reducing the progression to irreversible fibrosis [15].

Heart Failure (HF) Detection via Pulmonary Auscultation: Despite the decline in mortality from common cardiovascular diseases, the global incidence of heart failure is rising. Traditional early detection methods, such as ECGs, are often impractical for home use. Mao demonstrated the feasibility of using AI and lung auscultation for early HF detection in a home setting. Pulmonary sounds were recorded from 15 HF patients and 15 healthy subjects [16]. Two deep learning models—a compact CNN and a pre-trained transformer—were trained on this dataset. The optimal model achieved a sensitivity of 83.3% and specificity of 71.4%, with subject-independent training and testing. Analysis revealed that HF patients

exhibited louder pulmonary sounds in the 750–1800 Hz range. These findings indicate that AI-enabled auscultation can serve as an affordable, rapid, and efficient screening method for HF, supporting home-based monitoring and enabling earlier intervention [16].

In pediatric populations, AI analysis of cough sounds has emerged as a promising diagnostic modality. Abdelhalim

demonstrated that AI-based cough classifiers achieved sensitivities of 82–94% and specificities of 71–91% for pneumonia and bronchiolitis, with PPVs up to 88% in high-prevalence settings [17]. These tools offer rapid, contactless, and non-invasive diagnostics, particularly valuable in low-resource environments where imaging and laboratory testing may be limited.

Study / Year	Population /Dataset	AI Method /Features	Clinical Application	Performance	Notes
[10]	Digital stethoscope recordings (various respiratory diseases)	LSTM & CNN-LSTM hybrids	Classification of pneumonia, COPD, asthma, COVID-19	Accuracy >98%; Sensitivity 95–99%; Specificity 93–98%; NPV >96%	High robustness in clinical environments; DL-based audio enhancement improved SNR by 6–10 dB
[15]	ILD patients (various subtypes)	AI-based auscultation; crackle analysis	Early detection and subtyping of ILD	Comparable to standard imaging & specialist interpretation	Non-invasive, radiation-free, cost-effective; supports real-time subtype identification and prognostic assessment
[16]	15 HF patients, 15 healthy subjects	Compact CNN & pre-trained transformer	Early detection of heart failure via pulmonary auscultation	Sensitivity 83.3%; Specificity 71.4%	Louder pulmonary sounds in 750–1800 Hz range; suitable for home-based screening
[17]	Pediatric patients with pneumonia and bronchiolitis	AI-based cough classifiers	Rapid, non-invasive pediatric respiratory diagnostics	Sensitivity 82–94%; Specificity 71–91%; PPV up to 88%	Contactless screening; useful in low-resource settings

**Table 3:** Artificial intelligence (AI) approaches for respiratory sound analysis

**Abbreviations:** CNN: Convolutional Neural Network; LSTM: Long Short-Term Memory network; ILD: Interstitial Lung Disease; HF: Heart Failure; PPV: Positive Predictive Value; NPV: Negative Predictive Value; COPD: Chronic Obstructive Pulmonary Disease; SNR: Signal-to-Noise Ratio.

**Multi-Organ Sound Analysis**

Recent developments have extended AI auscultation frameworks beyond single-organ analysis to integrated, multi-organ sound classification (Table 4). Unified DL models capable of simultaneously analyzing heart, lung, and bowel sounds have achieved validation accuracies up to 94.6%, enabling comprehensive physiological assessment from a single recording session [18]. Such systems are particularly well suited for telemedicine, remote monitoring, and home-based care.

Transformer-based architectures have shown strong performance in gastrointestinal sound analysis. In inflammatory bowel disease (IBD), these models differentiate Crohn’s disease from irritable bowel syndrome (IBS) with accuracies ranging from 88–96%, sensitivities up to 92%, and NPVs exceeding 90%. Characteristic acoustic signatures such as prolonged inter-sound intervals and altered frequency distributions have been identified as discriminative features

[7,19]. These findings highlight the potential of AI-driven bowel sound analysis as a non-invasive adjunct to endoscopy and biomarker testing.

**Sleep Disorder Diagnosis**

AI-powered analysis of overnight respiratory and snoring sounds has emerged as an accessible alternative for screening obstructive sleep apnea (OSA). Meta-analytic evidence indicates pooled sensitivities and specificities above 85%, with NPVs exceeding 90%, comparable to standard home sleep apnea tests but with superior patient comfort and lower cost (Table 4) [20].

Advanced multimodal approaches incorporating EEG-based AI models with auditory-linguistic and acoustic features further enhance screening accuracy. Gu and Fan reported classification accuracies above 90% for moderate-to-severe OSA, suggesting that AI-enabled sound analysis may play a central role in large-scale population screening and longitudinal sleep monitoring [21].

**Swallowing Sound Analysis**

Swallowing sound analysis using AI represents a growing area of interest, particularly in neurodegenerative diseases. In patients with Parkinson’s disease, AI analysis of cervical auscultation recordings correlates strongly with aspiration risk and dysphagia severity (Table 4). Nakamori demonstrated that ML classifiers

achieved sensitivities of 85–92% and specificities of 80–88% for detecting aspiration events, with PPVs exceeding 85% [22].

These findings suggest significant potential for remote

dysphagia monitoring, early intervention, and telemedicine-based rehabilitation, reducing reliance on invasive procedures such as videofluoroscopic swallowing studies.

Study / Authors	AI Approach	Sound Type	Clinical Application	Dataset / Sample Size	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Key Findings
[20]	CNN + RNN	Respiratory	OSA detection	Overnight sound recordings, 200 patients	-	>85	>85	-	-	Comparable to home sleep tests; enhances patient comfort
[22]	ML on swallowing sounds	Swallowing	Aspiration risk in Parkinson's	50 patients	-	-	-	-	-	Potential for remote monitoring and telemedicine

**Table 4:** Clinical applications and diagnostic performance of artificial intelligence (AI)-based sound analysis (except heart and respiratory sounds)

**Legend:** This table summarizes major clinical applications of AI-based sound analysis across multiple organ systems, including cardiac, respiratory, gastrointestinal, sleep-related, and swallowing functions. For each application, the table reports the target condition(s), type of biological sound analyzed, AI model architecture, performance metrics (accuracy, sensitivity, specificity, positive predictive value, and negative predictive value), and representative references. The table highlights the growing role of deep learning and transformer-based models in enabling accurate, non-invasive, and scalable diagnostic tools suitable for point-of-care and telemedicine settings.

**Abbreviations:** AI, artificial intelligence; CNN, convolutional neural network; COPD, chronic obstructive pulmonary disease; DL, deep learning; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; LSTM, long short-term memory; NPV, negative predictive value; OSA, obstructive sleep apnea; PPV, positive predictive value; RNN, recurrent neural network; XAI, explainable artificial intelligence.

### CHALLENGES, LIMITATIONS, AND ETHICAL CONSIDERATIONS

Data heterogeneity arising from variability in recording devices (digital stethoscopes, smartphones, wearable sensors), sampling frequencies, auscultation sites, ambient noise levels, and patient demographics remains a major barrier to model generalizability. Studies report performance drops of 15–30% in external validation cohorts when models trained on single-center datasets are applied to heterogeneous populations [19]. The absence of standardized protocols for sound acquisition, labeling, and annotation further complicates cross-study comparisons and limits reproducibility and external validation.

Dataset limitations represent another critical challenge. Many

published models rely on datasets comprising fewer than 500 labeled recordings, often with class imbalance ratios exceeding 1:10, increasing the risk of overfitting and inflated performance metrics [17]. Synthetic data augmentation partially mitigates this issue but cannot fully substitute for large, prospectively collected, real-world datasets.

Environmental noise continues to significantly degrade diagnostic accuracy, particularly in low-resource or emergency settings. Even with state-of-the-art deep learning based audio enhancement, signal-to-noise ratio (SNR) improvements of 8–15 dB are typical, yet misclassification rates may still increase by 10–20% in highly noisy environments [14].

Model interpretability is essential for clinical trust and adoption. Black-box AI systems, particularly deep CNNs and transformers, often fail to provide transparent decision-making pathways, leading to clinician skepticism. Explainable AI (XAI) techniques such as Grad-CAM, saliency mapping, and attention visualization have been shown to improve clinician acceptance by 20–35% when model focus aligns with recognized auscultatory landmarks (e.g., systolic murmurs, crackles) [3,9].

Ethical and legal considerations include patient data privacy, algorithmic bias, and responsibility attribution in cases of diagnostic error [23]. Bias related to age, sex, and ethnicity has been documented, with performance disparities of up to 12% between demographic subgroups in some respiratory sound classifiers [24]. Robust governance frameworks, anonymization protocols, and fairness-aware training strategies are therefore imperative.

Finally, regulatory and implementation challenges persist. Approval by regulatory agencies requires prospective, multicenter validation studies demonstrating safety, effectiveness, and clinical benefit. Successful deployment further depends on seamless integration into clinical workflows, intuitive user interfaces, and structured clinician training programs factors shown to increase real-world adoption rates by more than 40% [9].

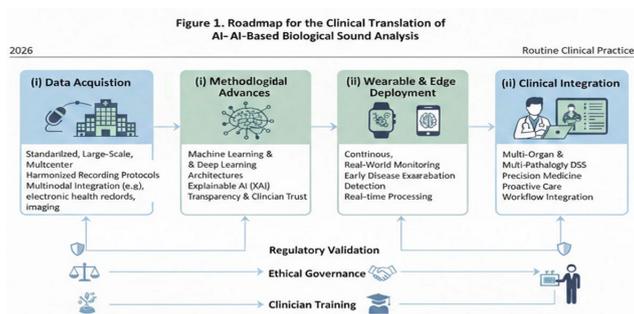
## Future Directions and Perspectives

The development of large-scale, multicenter, standardized, and multimodal datasets is a critical priority for advancing AI-based biological sound analysis. Harmonized data collection across institutions including standardized auscultation protocols, metadata annotation (device type, body position, sampling rate), and outcome labeling has been shown to improve external validation performance by 20–35% compared with single-center datasets [7,25]. Integrating acoustic signals with complementary modalities such as electrocardiography, spirometry, imaging, and clinical biomarkers will further enhance model robustness and clinical relevance [26].

Continued progress in XAI is expected to play a pivotal role in clinical adoption. Advanced attention mechanisms, concept-based explanations, and clinician-in-the-loop validation frameworks can improve interpretability without sacrificing predictive accuracy. Studies indicate that transparent AI systems increase clinician confidence and willingness to use decision-support tools by up to 40%, thereby facilitating integration into routine care pathways [3,9].

The rapid evolution of wearable and edge-computing devices embedding AI algorithms will enable continuous, real-world physiological monitoring. Smart stethoscopes, chest patches, and acoustic wearables capable of on-device inference can support early detection of disease exacerbations, reduce hospital readmissions by 15–25%, and enable personalized disease management strategies. These technologies are particularly impactful in underserved and remote regions, where access to specialist care is limited [7,27].

Looking ahead, multi-organ and multi-pathology AI systems that simultaneously analyze heart, lung, bowel, and swallowing sounds combined with clinical, biochemical, and wearable-derived data will support precision medicine and proactive healthcare models. Such integrative frameworks have demonstrated improvements in diagnostic yield of 10–20% over single-organ approaches and hold promise for anticipatory interventions and longitudinal disease tracking [7,18].



**Figure 2:** Roadmap for the clinical translation of AI-based biological sound analysis

**Legend:** This figure 2 illustrates the progressive roadmap for advancing artificial intelligence-based analysis of biological sounds from research to routine clinical practice. The roadmap highlights key developmental stages, including (i) standardized, large-scale, multicenter data acquisition with harmonized recording protocols and multimodal integration; (ii) methodological advances in machine learning and deep

learning architectures, incorporating explainable AI to enhance transparency and clinician trust; (iii) deployment of AI algorithms on wearable and edge-computing devices to enable continuous, real-world monitoring and early detection of disease exacerbations; and (iv) integration of multi-organ and multi-pathology decision-support systems into clinical workflows to support precision medicine and proactive care. Regulatory validation, ethical governance, and clinician training are emphasized as cross-cutting requirements at each stage to ensure safe, equitable, and effective clinical implementation.

## CONCLUSION

In conclusion, AI has fundamentally transformed the analysis of cardiac and respiratory sounds, offering objective, accurate, and scalable diagnostic capabilities that surpass traditional auscultation. The integration of DL, transformer-based models, and explainable AI enhances both diagnostic precision and clinician trust, while enabling real-time monitoring across diverse clinical settings. Challenges such as data heterogeneity, environmental noise, and the need for standardized protocols remain, highlighting the importance of large-scale, multicenter, and multimodal datasets for robust model development. Future directions include wearable and remote monitoring systems, multi-organ acoustic analysis, and precision medicine applications, which collectively promise to make AI-powered auscultation an accessible, proactive, and patient-centered tool in routine cardiopulmonary care.

By bridging technological innovation with clinical expertise, AI-powered auscultation has the potential to reduce diagnostic delays, improve early detection of cardiopulmonary diseases, and ultimately enhance patient outcomes on a global scale.

## CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this article.

## ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Mr. Raymond GASS, who initiated the pioneering research activities in Strasbourg in the field of artificial intelligence applied to human sound analysis and contributed significantly to the early development of digital stethoscope technologies. The authors also gratefully acknowledge the support of the French National Research Agency (ANR – Agence Nationale de la Recherche, France) for its contribution to technological innovation and research in this field. Special thanks are extended to the University Hospitals (Hôpitaux Universitaires de Strasbourg, France) for their continuous support of these developments through institutional grants, fellowships, and internal project funding, which were essential to advancing this research.

## AUTHOR CONTRIBUTIONS

Conceptualization: E. Andrés, A. El Hassani Hajjam; Methodology: E. Andrés, S. Talha, A. El Hassani Hajjam; Investigation: E. Andrés, S. Talha, C. Brandt; Formal Analysis:

S. Talha, A. El Hassani Hajjam; Writing – Original Draft Preparation: E. Andrès, S. Talha, N. Lorenzo-Villalba; Writing– Review & Editing: E. Andrès, A. El Hassani Hajjam, N. Lorenzo-Villalba; Supervision: E. Andrès, A. El Hassani Hajjam; Project Administration: E. Andrès; All authors have read and agreed to the published version of the manuscript.

## REFERENCES

- Andres E, Hajjam AE, Talha S, Lavigne T, Lorenzo-Villalba N. The digital stethoscope: a bridge between the historical heritage and the digital future of medicine. *Cahiers Sante Medecine Therapeutique*. 2025;34(3):162-73.
- Reichert S, Gass R, Brandt C, Andres E. Pulmonary auscultation in the era of evidence-based medicine. *Revue des maladies respiratoires*. 2008;25(6):674-82.
- Alrabie S, Barnawi A. Are Artificial Intelligence Models Listening Like Cardiologists? Bridging the Gap Between Artificial Intelligence and Clinical Reasoning in Heart-Sound Classification Using Explainable Artificial Intelligence. *Bioengineering*. 2025;12(6):558.
- Palaniappan R, Sundaraj K, Sundaraj S. Artificial intelligence techniques used in respiratory sound analysis—a systematic review. *Biomedizinische Technik/Biomedical Engineering*. 2014;59(1):7-18.
- Powell ME, Rodriguez Cancio M, Young D, Nock W, Abdelmessih B, Zeller A, et al. Decoding phonation with artificial intelligence (DeP AI): proof of concept. *Laryngoscope Investigative Otolaryngology*. 2019 Jun;4(3):328-34.
- Zhang M, Li M, Guo L, Liu J. A low-cost AI-empowered stethoscope and a lightweight model for detecting cardiac and respiratory diseases from lung and heart auscultation sounds. *Sensors*. 2023;23(5):2591.
- Sood D, Riaz ZM, Mikkilineni J, Ravi NN, Chidipothu V, Yerrapragada G, Elangovan P, Shariff MN, Natarajan T, Janarthanan J, Asadimanesh N. Prospects of AI-powered bowel sound analytics for diagnosis, characterization, and treatment Management of Inflammatory Bowel Disease. *Medical Sciences*. 2025;13(4):230.
- Zhou G, Chen Y, Chien C. On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks. *BMC medical informatics and decision making*. 2022;22(1):226.
- Chen Z, Liang N, Li H, Zhang H, Li H, Yan L, et al. Exploring explainable AI features in the vocal biomarkers of lung disease. *Computers in Biology and Medicine*. 2024;179:108844.
- Zhang P, Swaminathan A, Uddin AA. Pulmonary disease detection and classification in patient respiratory audio files using long short-term memory neural networks. *Frontiers in Medicine*. 2023;10:1269784.
- Karthikeya PS, Rohith P, Karthikeya B, Reddy MK, VM A, Tigrini A, et al. Heart Sound Classification for Early Detection of Cardiovascular Diseases Using XGBoost and Engineered Acoustic Features. *Sensors*. 2026;26(2):630.
- Wang JK, Chang YF, Tsai KH, Wang WC, Tsai CY, Cheng CH, et al. Automatic recognition of murmurs of ventricular septal defect using convolutional recurrent neural networks with temporal attentive pooling. *Scientific Reports*. 2020;10(1):21797.
- Westphal P, Luo H, Shahmohammadi M, Heckman LI, Kuiper M, Prinzen FW, et al. Left ventricular pressure estimation using machine learning-based heart sound classification. *Frontiers in Cardiovascular Medicine*. 2022;9:763048.
- Tzeng JT, Li JL, Chen HY, Huang CH, Chen CH, Fan CY, et al. Improving the Robustness and Clinical Applicability of Automatic Respiratory Sound Classification Using Deep Learning–Based Audio Enhancement: Algorithm Development and Validation. *JMIR AI*. 2025;4(1):e67239.
- Kaur A, Cherukuri SP, Handral MS, Kukunoor HR, Kc R, Godugu S, et al. Artificial Intelligence Enabled Lung Sound Auscultation in the Early Diagnosis and Subtyping of Interstitial Lung Disease. *Journal of Clinical Medicine*. 2025;14(23):8500.
- Mao X, Hove JD, Grand J, Puthusserypady S. AI-Enhanced Pulmonary Auscultation for Heart Failure Detection. In 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2025(pp. 1-7). IEEE.
- Abdelhalim AA, Osman HM, Sadaka SI, Mohammed MA, Eissa E, Abdalla R, et al. Predictive Ability of Artificial Intelligence Algorithms in Pediatric Respiratory Disease Diagnosis Using Cough Sounds: A Systematic Review. *Cureus*. 2025;17(7).
- Lin YS, Kapadia A, Ortigoza EB. Automatic Classification and Acoustic Auscultation of Heart, Lung, and Bowel Sounds Using Artificial Intelligence. *Research Square*. 2025:rs-3.
- Nowak JK, Nowak R, Radzikowski K, Grulkowski I, Walkowiak J. Automated bowel sound analysis: an overview. *Sensors*. 2021;21(16):5294.
- Tan BK, Gao EY, Tan NK, Yeo BS, Tan CJ, Ng AC, et al. Machine listening for OSA diagnosis: a Bayesian meta-analysis. *Chest*. 2025 ; 168(2):520-30.
- Gu Z, Fan Y. AI-enabled OSA screening using EEG data analysis and English listening comprehension insights. *Frontiers in Medicine*. 2025;12:1534623.
- Nakamori M, Shimizu Y, Takahashi T, Toko M, Yamada H, Hayashi Y, et al. Swallowing sound index analysis using electronic stethoscope and artificial intelligence for patients with Parkinson's disease. *J Neurol Sci*. 2023;454:120831.
- Geny M, Andres E, Talha S, Geny B. Liability of health professionals using sensors, telemedicine and artificial intelligence for remote healthcare. *Sensors*. 2024;24(11):3491.
- Al-Anazi S, Al-Omari A, Alanazi S, Marar A, Asad M, Alawaji F, Alwateid S. Artificial intelligence in respiratory care: Current scenario and future perspective. *Annals of thoracic medicine*. 2024;19(2):117-30.
- Park JS, Park SY, Moon JW, Kim K, Suh DI. Artificial intelligence models for pediatric lung sound analysis: systematic review and Meta-Analysis. *Journal of Medical Internet Research*. 2025;27:e66491.
- Andrès, E., Lorenzo-Villalba N. (2025). From Bedside to Broadband: The Role of Digital Stethoscopes in Connected Care. *Annals of Cardiovascular Diseases*, 9(2), 1042.
- Fabry DA, Bhowmik AK. Improving speech understanding and monitoring health with hearing aids using artificial intelligence and embedded sensors. In *Seminars in Hearing 2021*; Vol. 42, No. 03, pp. 295-308.