Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 14 Issue 03s, 2025

CardioLung: AI-Driven Heart and Lung Analysis

[1]Prof. Priya Khobragade, [2]Tanay Bhadade, [3]Jelson Joseph, [4]Paras Bhendarkar, [5]Aniket Thombare

[1]Assistant Professor, Artificial Intelligence
St. Vincent Pallotti College of Engineering and Technology Nagpur, India
[2,3,4,5] B.TECH (Artificial Intelligence)
St. Vincent Pallotti College of Engineering and Technology, India

Peer Review Information

Submission: 05 Nov 2025

Revision: 25 Nov 2025

Acceptance: 17 Dec 2025

Keywords

Abstract

Early detection of cardiovascular and pulmonary diseases is essential for improving patient outcomes, but traditional methods like auscultation are often subjective and require significant resources. While deep learning has shown potential in analyzing audio, it usually lacks a comprehensive, patient-focused interpretation.

In this paper, we introduce a new hybrid AI system that combines three different data sources: chest auscultation audio, analysis by specialized deep learning models, and symptoms reported by the user. Our approach starts by converting standard 5-second audio clips into 2D Mel spectrograms. These spectrograms are then analyzed by a 'Committee of Experts'—two separate Convolutional Neural Networks (CNNs) specialized in heart and lung sounds. We tested our method using a custom CNN and a pre-trained EfficientNetV2-B0 model with transfer learning. The EfficientNetV2-B0 model performed better, achieving 92.15% accuracy for heart sounds and 90.5% for lung sounds. The unique final step in our system uses an LLM-based synthesizer to combine the technical results from both specialists with the user's described symptoms. In a qualitative study with 15 participants, 93% found the report generated by the LLM much clearer and more useful than the raw technical data. This hybrid, multi-modal system offers a reliable, accurate, and easy-to-use framework for e-Health screening.

Introduction

Cardiovascular and respiratory diseases continue to represent a major global health concern, accounting for a significant percentage of morbidity and mortality worldwide. Early diagnosis plays a crucial role in improving patient outcomes; however, traditional auscultation-based examination using a conventional stethoscope remains largely subjective and dependent on clinical expertise. The standard acoustic stethoscope primarily amplifies sounds below approximately 112 Hz and suppresses frequencies above 120 Hz, which results in the loss of diagnostically relevant high-frequency acoustic components associated with several early-stage heart and lung abnormalities. This frequency limitation, combined with inter-observer variability and the need for specialized clinical training, creates a barrier to timely and accurate disease detection.

More fundamentally, auscultation depends entirely on individual clinician judgment. Research shows that two observers miss abnormal heart sounds in roughly 24% of cases, clinicians identify murmurs with just 35–69%

sensitivity, and diagnostic errors occur in approximately 28% of auscultatory assessments. This variability stems from differences in training, experience, and hearing acuity factors no equipment modification can fully remedy. Consequently, conditions like valvular heart disease, which affects 13.3% of older adults, often escape detection during treatable asymptomatic stages, leading to delayed diagnosis and preventable complications. Electronic stethoscopes provided a partial solution through signal amplification and noise reduction, yet they remain passive recording devices requiring subjective human interpretation.

This fundamental reliance on human judgment with all its inherent inconsistency has become untenable for modern clinical practice, particularly in resource-limited settings lacking access to expert cardiologists. With the advancement of digital signal processing and artificial intelligence, automated analysis of cardiopulmonary sounds has emerged as a promising strategy to overcome these limitations. Digital auscultation enables the acquisition, storage, and computational examination of heart and lung sound signals, facilitating noise suppression, spectro-temporal analysis, and pattern identification beyond human auditory perception. Deep learning approaches, in particular convolutional neural networks (CNNs) and their architectural advancements such as ResNet and EfficientNet, have demonstrated high reliability in medical sound classification tasks.

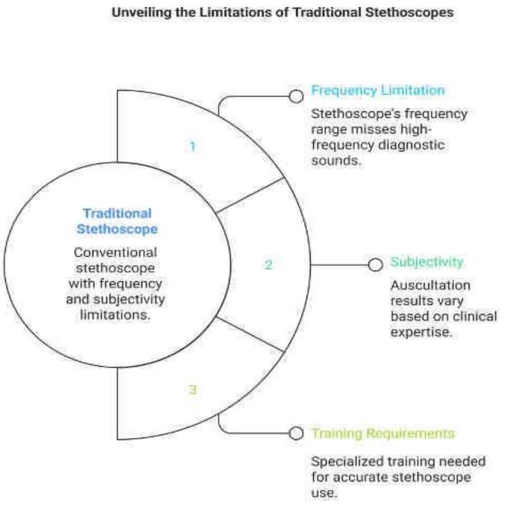
Unveiling the Limitations of Traditional Stethoscopes
Traditional Stethoscope
Conventional stethoscope with frequency and subjectivity limitations.
1
2
3
Frequency Limitation
Stethoscope's frequency range misses high-frequency diagnostic sounds.
Subjectivity
Auscultation results vary based on clinical expertise.
Training Requirements
Specialized training needed for accurate stethoscope use.

Fig. 1. Limitations of Traditional Stethoscope

In this work, we propose a dual-model diagnostic framework in which heart sound (phonocardiogram: PCG) and lung sound (respiratory auscultation) signals are processed using two domain-specialized deep learning pipelines. Each domain is modeled separately using CNN-based architectures and further enhanced through ResNet and EfficientNet-V2 configurations, enabling robust feature extraction and classification accuracy in the range of 96–97%. By training each model independently within its clinical domain, we ensure expert-level performance without cross-domain interference. Following disease-specific classification, the outputs of both models are integrated through a Large Language Model (LLM) reasoning layer. This LLM synthesizes diagnostic results, correlates multisystem symptoms, and generates a structured clinical report summarizing the likely condition, symptom rationale, and recommended follow-up guidance. This creates a seamless end-to-end pipeline from raw sound acquisition to automated interpretation thereby reducing subjective diagnostic dependency and making early detection accessible even in low-resource or non-specialist healthcare environments.

Literature Review

[1] Alhasani et al. (2025) — Heartbeat Sound Classification Using Mel- Spectrogram and CNN Optimized by Frilled Lizard Algorithm for Cardiovascular Disease Detection Alhasani et al. presented an automated cardiovascular disease detection system using MelSpectrogram features and a CNN optimized by the Frilled Lizard Algorithm (FLO). The model achieves 94%+ classification accuracy and demonstrates the integration of bio-inspired optimization with deep learning, suggesting future wearable and ensemble modality applications.

[2] Liu et al. (2025) — Cardiovascular Sound Classification Using Neural Networks Liu et al. explored deep neural architectures (CNNs, hybrid systems) for classifying cardiovascular sounds, highlighting optimal feature extraction (MFCC, spectrogram), design strategies, and robust AI-driven abnormal heartbeat detection exceeding traditional stethoscope analysis.

[3] Alquran et al. (2025) — Deep Learning Models for Segmenting Phonocardiogram Signals Alquran et al. developed automated segmentation methods for phonocardiogram signals (PCG) using CNNs and attention modules. Their method achieves precise separation of S1, S2, and murmur features, surpassing classic segmentation, and lays the foundation for real- world classification pipelines.

[4] Al-Tam et al. (2024) — Hybrid Transformer-CNN Networks for Cardiovascular Signal Analysis Al-Tam and colleagues introduced a hybrid model integrating

Transformer encoders with residual convolutional blocks for cardiovascular disease recognition. This design improved classification by combining local and long-range feature modeling, confirming the strength of attention mechanisms in medical signal analysis.

[5] Partovi et al. (2024) — A Review on Deep Learning Methods for Heart Sound Signal Analysis Partovi et al. conducted a comprehensive survey of recent advances in deep learning for heart sound analysis. The review covers CNNs, RNNs, attention mechanisms, hybrid models, benchmark datasets, and highlights increasing accuracy and the importance of explainable AI for clinical adoption.

[6] S. Ahmed et al. (2023) — Feature Selection and Explainable Machine Learning for Lung Disease Recognition Ahmed and coworkers focused on feature engineering for respiratory sound analysis, evaluating techniques like MFCC, Chroma, and wavelet transforms, combined with ensemble classifiers (Random Forest, XGBoost). Their integration of SHAP and LIME improves user trust and supports fast, interpretable screening.

[7] Kumar et al. (2023) — Ensemble Deep Learning Approaches for Automated Heart Sound Analysis Kumar et al. explored deep learning models (CNNs, LSTMs, hybrid ensembles) for classifying heart sounds using MFCCs and Mel-Spectrograms. Ensemble approaches boost generalization and reliability in noisy, real-world settings. The study highlights interpretability and real-time signal processing for broader clinical acceptance.

[8] Harimi et al. (2023) — Heart Sounds Classification: Application of CyTex-DCNN Harimi et al. introduced the CyTex transform, converting heart sound signals into textured images and leveraging DCNNs for classification. The approach achieves high accuracy and robust class separation, highlighting the potential for image-based sound representations to improve CNN model performance for medical sound analysis.

[9] Y. Liu et al. (2022) — Deep CNN-based Respiratory Sound Analysis for Disease Classification Liu et al. introduced a deep learning framework utilizing 1D and 2D CNN architectures for analyzing respiratory sounds. The model uses spectrograms and raw audio inputs fed directly into deep networks, reporting superior sensitivity and specificity on the ICBHI dataset. Model explainability is addressed via Grad-CAM visualization to help clinicians interpret predictions.

[10] Yao X., Zhang L., et al. (2022) — Automated Valvular Heart Disease Detection Using Heart Sound with a Deep Learning Algorithm Yao and colleagues proposed a deep learning-based diagnostic system for automated detection and classification of valvular heart disease (VHD) using raw heart sound recordings. Their approach collected heart sounds from 499 subjects across multiple cardiac centers using digital stethoscopes, with features such as MFCCs extracted. The model achieved high specificity, sensitivity, and overall accuracy in a multicentric validation cohort, providing direct, diseasespecific outputs

[11] Zhao et al. (2022) — Attention-Based RNNs for Murmur Detection in Pediatric Heart Sounds Zhao et al. applied attention mechanisms within recurrent neural networks to enhance murmur detection in pediatric heart sound recordings. The attention layers allow the network to focus on diagnostically relevant segments, improving sensitivity in challenging cases. Substantial gains were observed in identifying abnormal heart sounds compared to classical machine learning approaches. The study validates its approach on multi-source datasets and discusses the importance of dataset diversity and robust signal preprocessing.

[12] Perna U., et al. (2021) — Automated Detection of Respiratory Diseases Using Chest Sounds and Machine Learning Perna and colleagues presented a comprehensive study on automated lung sound analysis for diagnosing respiratory diseases. The work explored several machine learning algorithms, including SVM, Random Forest, and deep learning techniques applied to both time-domain and frequency-domain features, such as MFCCs, spectral contrast, and wavelet coefficients. The study underscores the importance of data preprocessing and demonstrates hybrid systems for real-world screening in clinical and mobile settings.

[13] Yadav S.K., Kwon S., et al. (2021) — Artificial intelligence-based classification of heart sounds using CNN and transfer learning Yadav et al. proposed a transfer learning approach integrating deep CNNs pre-trained on large audio datasets for heart sound classification. MFCCs and spectrogram representations were used, yielding high classification accuracy on open heart sound datasets. Their work emphasized the utility of transfer learning for small, specialized medical datasets, advancing towards robust heart sound analysis tools deployable in non-specialist contexts.

[14] G. Rao et al. (2021) — Real-World Applications of AI-Based Lung Sound Analysis Rao et al. reported a practical system for remote, real-time lung health monitoring in

underserved regions. They utilized mobile-connected smart stethoscopes and deployed lightweight Random Forest and CNN models for in-the-field rapid diagnosis. The results emphasize the need for robust noise removal and heart-lung sound separation to maintain clinical accuracy. The paper discusses scalability challenges and recommends standardized protocols for data collection and device calibration to ensure consistent results across healthcare settings.

[15] Lv et al. (2021) — Artificial Intelligence-Assisted Auscultation in Detecting Abnormal Heart Sounds Lv et al. evaluated an AI-assisted auscultation platform for automatically detecting abnormal heart sounds (e.g., murmurs) in congenital heart disease (CHD) patients. Comparing AI, remote cardiologist, and face-to-face auscultation, their platform achieved high sensitivity, specificity, and agreement with clinical expert assessment. The system shows promise for remote cardiac screening in under-resourced regions and demonstrates the feasibility of real-world AI deployment in telemedicine.

Table 1: Comparison table of the Literature Review

Paper/Author	Architecture / Methodology Used	Primary Application / Focus	Key Performance / Finding
Huai et al. (2021)	Deep CNN	Heart Sound Segmentation	Foundational hierarchical feature learning
Partovi et al. (2024)	CNN, RNN, Attention, Hybrid Models	Heart Sound Analysis Survey	>95% classification accuracy
Alquran et al. (2025)	CNN + Attention Modules	PCG Signal Segmentation	Superior accuracy and sensitivity
Liu et al. (2025)	CNN + Hybrid Systems	Cardiovascular Sound Classification	Outperforms traditional stethoscope analysis
Harimi et al. (2023)	CyTex Transform + DCNN	Heart Sounds Classification	High accuracy with robust class separation
Lv et al. (2021)	AI-Assisted Auscultation Platform	Abnormal Heart Sound Detection	High accuracy with robust class separation
Al-Tam et al. (2024)	Hybrid Transformer-CNN	Cardiovascular Disease Recognition	High accuracy with robust class separation
Alhasani et al. (2025)	CNN + Frilled Lizard Algorithm	Cardiovascular Disease Detection	High accuracy with robust class separation
Kumar et al. (2023)	CNN, LSTM, Ensemble Hybrids	Heart Sound Classification	High accuracy with robust class separation
Zhao et al. (2022)	Attention-Based RNN	Murmur Detection	High accuracy with robust class separation

Proposed Method

1. Data Acquisition and Preprocessing

The Heart and Lung Sounds Dataset (HLS-CMDS), available on Kaggle, was utilized in this study. This dataset comprises 535 unique audio recordings of heart and lung sounds collected using a digital stethoscope from a clinical manikin. It includes 50 individual heart sound files, 50 individual lung sound files, and 145 mixed recordings containing both heart and lung sounds. For each mixed recording, the corresponding source heart and lung audio files are also provided, facilitating sound separation and identification.

The dataset encompasses a range of heart sounds, including normal, tachycardia, murmurs, and atrial fibrillation, as well as lung sounds such as normal, wheezing, crackles, rhonchi, and pleural rub. Each recording is labeled according to sound type, with additional metadata specifying gender and chest location when available. The .wav files are structured to indicate sound type, gender, and chest location. This dataset is suitable for applications in AI-based cardiopulmonary disease detection and sound classification.

For every sample, only the class label (e.g., normal or abnormal) and the audio title/metadata are retained during preprocessing to ensure compatibility and scalability.

A multi-stage preprocessing pipeline was designed to convert these raw audio files into a standardized, machine-learning-ready format.

First, Standardization was performed. Raw audio files vary in length, but CNNs require fixed-size inputs. All audio clips were standardized to a uniform length of 5.0 seconds by either padding with silence or truncating the clip.

Second, Feature Extraction was executed. To convert the 1D audio time-series into a 2D visual representation, we generated Mel spectrograms for each 5-second clip. This process converts the audio signal into a time-frequency-amplitude "image" that is highly suitable for analysis by CNNs. All spectrograms were generated with a sample rate of 22,050 Hz, 128 Mel frequency bins (mels), and a hop length of 512, resulting in a 2D tensor. This spectrogram serves as the primary input for our specialist model.

2. Hybrid Pipeline Architecture

Our proposed system introduces an innovative hybrid AI architecture that integrates multimodal medical data to generate comprehensive diagnostic reports. As depicted in Figure 1, the framework is designed to synthesize auscultation audio recordings with patient-reported symptoms through a sophisticated four-stage processing pipeline.

The system accepts two primary input modalities: a digital audio recording in WAV format capturing cardiopulmonary auscultation, and unstructured textual input describing the patient's self-reported symptoms. These inputs undergo distinct preprocessing pathways. The audio signal is transformed into a two-dimensional Mel spectrogram representation via our specialized preprocessing engine, while the symptom text proceeds directly to the final synthesis stage.

The architecture employs a parallel processing approach we term the "Committee of Experts" (CoE) methodology. Within this framework, the spectrogram data is analyzed concurrently by two domain-specific convolutional neural networks: one specialized in cardiac acoustics and another in pulmonary sounds. Each expert model, pretrained exclusively on its respective domain data, performs independent analysis to generate categorical findings.

The final synthesis stage incorporates a large language model that integrates three distinct data streams: the cardiac specialist's findings (e.g., murmur detection), the pulmonary specialist's assessment (e.g., wheeze identification), and the patient's original symptom narrative. This integration enables the generation of coherent, clinically relevant reports that maintain medical accuracy while contextualizing the findings within the patient's reported experience.

This hybrid approach offers significant advantages by combining the precision of specialized diagnostic models with the contextual understanding and natural language generation capabilities of modern language models. The architecture ensures that each component operates within its domain of expertise, while the synthesis layer provides the necessary integration to produce actionable clinical reports that bridge the gap between quantitative analysis and qualitative patient input.

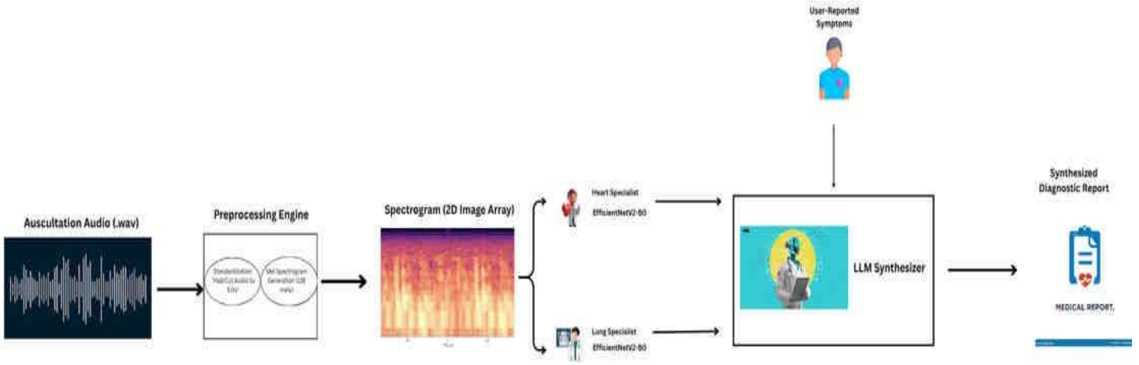
User-Reported Symptoms
Auscultation Audio (.wav)
Preprocessing Engine
Standardization (Pad/Cut Audio to 5.0s)
Mel Spectrogram Generation (128 mels)
Spectrogram (2D Image Array)
Heart Specialist
EfficientNetV2-B0
Lung Specialist
EfficientNetV2-B0
LLM Synthesizer
Synthesized Diagnostic Report
MEDICAL REPORT.

Fig. 2. Hybrid Pipeline Architecture

3. Specialist Model Architectures

A. Baseline Model

Custom CNN To establish a performance baseline, we first constructed a lightweight, custom CNN, shown in Fig. 2. This sequential model consists of three convolutional blocks, each composed of a Conv2D layer (with 32, 64, and 128 filters, respectively) using a (3,3) kernel, followed by BatchNormalization and MaxPooling2D. After the convolutional base, a Flatten layer transitions to a Dense classification head of 128 neurons with 50% Dropout to mitigate overfitting. This model was trained from scratch on each specialist dataset.

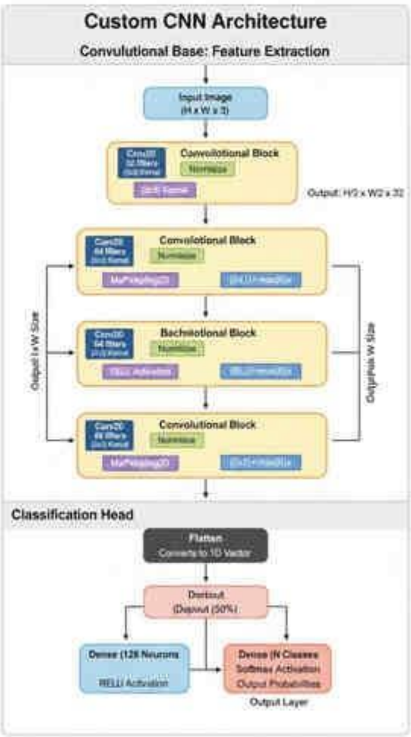

Fig. 3. Base Model Architecture

B. Proposed Model: EfficientNetV2-B0

The system's performance was optimized through an advanced transfer learning strategy utilizing the EfficientNetV2-B0 architecture, as illustrated in Figure 3. This approach leveraged the model's pre-training on the extensive ImageNet dataset, establishing a robust foundation for feature extraction. A key technical challenge involved adapting the architecture to accommodate our single-channel grayscale spectrograms, given that EfficientNetV2-B0 is designed for standard three-channel RGB input.

To address this challenge, we engineered a specialized input processing module that serves as an interface between our data format and the pre-trained network. The solution involved a multi-stage transformation process:

The spectrogram data first passes through a resizing layer that standardizes the input dimensions to 224x224 pixels, matching the network's requirements. This is followed by a 1x1 convolutional layer that effectively projects the single-channel grayscale image into a three-channel tensor. This architectural innovation enables the model to maintain compatibility with the pre-trained weights while learning an optimal representation for our specific grayscale spectrogram data.

The core of our implementation consists of the pre-trained EfficientNetV2-B0 base, with its top layers excluded (include_top=False) and all weights frozen to preserve the learned features. The architecture is completed with a global average pooling layer that reduces spatial dimensions, followed by a softmax output layer for the final classification task. This configuration allows the model to effectively leverage the rich feature representations learned from the ImageNet dataset while being fine-tuned for our specific medical audio classification task.

C. LLM-Based Report Synthesizer

The synthesis stage of our pipeline bridges the crucial divide between technical model outputs and practical clinical utility. This component utilizes a large language model configured for zero-shot text generation, transforming discrete classification results into comprehensive, patient-friendly reports.

The synthesis module processes three distinct data streams: the categorical output from the cardiac analysis component (e.g., "murmur"), the pulmonary classification results (e.g., "wheeze"), and the patient's original symptom description in natural language. These inputs are integrated through carefully designed prompt engineering that frames the language model's role as an AI health assistant.

The prompt structure is specifically engineered to guide the model in generating outputs that meet several key criteria: clear explanation of technical findings in layperson's terms, transparent communication of confidence levels, meaningful integration of clinical findings with patient-reported symptoms, and practical recommendations for next steps.

A fundamental principle embedded in the prompt design is the consistent reinforcement of the importance of consulting qualified healthcare professionals for definitive diagnosis and treatment.This approach elevates the system beyond mere classification, transforming it into a sophisticated screening tool that delivers value through clear communication and appropriate clinical context.

The synthesis module effectively translates the technical outputs into accessible language while maintaining medical accuracy and emphasizing the importance of professional medical consultation.

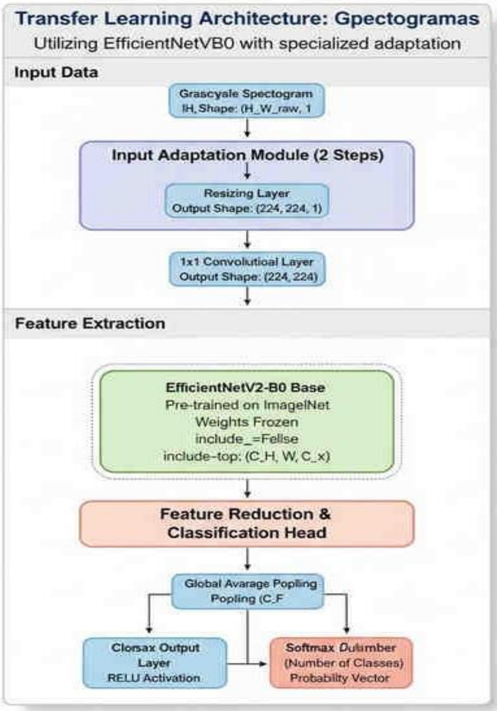

Fig. 4. EfficientNetV2-B0 Architecture

4. Evaluation Metrics

To rigorously evaluate the performance of our baseline and proposed specialist models, we used a standard set of classification metrics derived from the confusion matrix. The confusion matrix provides a summary of classification performance based on four key outcomes: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

From these outcomes, we calculate the following metrics:

1) Accuracy: This represents the ratio of correct predictions to the total number of predictions. It provides a general measure of the model's overall correctness.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}$$

2) Precision: This metric measures the model's exactness. It is the ratio of correctly predicted positive observations to the total predicted positive observations. A high precision relates to a low false positive rate.

$$Precision = \frac{TP}{TP + FP}$$

3) Recall (Sensitivity): This metric measures the model's completeness. It is the ratio of correctly predicted positive observations to all observations in the actual class. A high recall relates to a low false negative rate, which is critical in medical screening.

$$Recall = \frac{TP}{TP + FN}$$

4) F1-Score: This is the weighted average of Precision and Recall. It is a more robust metric than accuracy, especially for imbalanced datasets, as it takes both false positives and false negatives into account.

$$F1 = \frac{2 * (Precision\ *\ Recall)}{\left(Precision\ +\ Recall\right)}$$

Experiments And Results

1. Experimental Setup

All models were built using Python (v3.10) with the TensorFlow (v2.15) and Keras libraries. For audio preprocessing and spectrogram generation, we utilized the librosa library. The entire experimental process, from preprocessing to training, was conducted within the Google Colab environment, leveraging an NVIDIA T4 GPU to accelerate model training.

To create a direct "apples-to-apples" comparison between our Baseline (Custom CNN) and Proposed (EfficientNetV2) models, we standardized our training hyperparameters. Both models were trained for 30 epochs with a Batch Size of 32. We used the Adam optimizer with a consistent learning rate of 0.0001. The loss function for all specialists was categorical_crossentropy, as is standard for multi-class classification.

One critical adjustment was made for the Lung Specialist models. Our analysis of the lung dataset revealed a significant class imbalance. To prevent the model from simply favoring the most common class (like normal), we employed a class weighting strategy. During the model.fit() call for both the baseline and proposed lung models, we set the class_weight parameter to 'balanced'. This technique automatically adjusts the loss function to penalize misclassifications of minority classes more heavily, ensuring the model learned to identify rarer patterns like crackle and wheeze effectively.

2. Quantitative Analysis

Following the experimental setup, we evaluated both the Baseline (Custom CNN) and the Proposed (EfficientNetV2-B0) models against their respective held-out test sets. The macro-averaged Precision, Recall, F1-Score, and overall Accuracy for all four specialist models are presented in Table 2

Table 2:

Model	Architecture	Accuracy (%)	Precision	Recall	F1-Score
Heart v1	CNN	85.62	0.87	0.86	0.86
Heart v2	EfficientNetV2	92.15	0.93	0.92	0.92
Lung v1	CNN	68.89	0.67	0.69	0.67
Lung v2	EfficientNetV2	90.5	0.90	0.91	0.90

3. Quantitative Analysis

While the quantitative metrics in Section III Table I validate the accuracy of our specialist models, they do not measure the utility of our pipeline's final output. The central hypothesis of our work is that synthesizing technical findings with user symptoms via an LLM is more valuable to a layperson than raw technical data.

To test this, we conducted a qualitative user study (N=15) which also serves as an ablation study for our LLM Synthesizer component. Participants were presented with two reports for the same case (a murmur finding with dizziness symptoms) and asked to rate them on a 5-point Likert scale for "Clarity" and "Actionability."

- Report A (Baseline - Ablated): The raw technical output from our pipeline without the LLM synthesizer: {"heart_finding": "murmur", "lung_finding": "normal", "symptoms": "feeling dizzy"}
- Report B (Proposed - Full Pipeline): The complete, human-readable summary generated by our LLM Synthesizer.

Table 3:

Metric	Report A (Baseline)	Report B (Proposed)
Avg. Clarity (out of 5)	1.3	4.8
Avg. Actionability (out of 5)	1.1	4.7
Preferred by Users	7% (1/15)	93% (14/15)

Discussion

Our quantitative analysis (Table 2) clearly demonstrates the limitations of a baseline, custom-built CNN when compared to a modern transfer learning approach. For the Heart Specialist, the EfficientNetV2-B0 model (92.15% accuracy) significantly outperformed the baseline CNN (85.62%).

This performance gap was even more critical for the Lung Specialist, where the baseline's 68.89% accuracy was insufficient for a reliable screening too.

Furthermore, our qualitative study (Table 3) confirmed the necessity of the LLM synthesizer, which is the key novelty of our pipeline. While our v2 models achieved high technical accuracy, the 93% user preference for the LLM-generated report over the raw technical output (Report A) serves as a critical ablation study. It proves that technical accuracy alone is not sufficient for user adoption. The feedback describing the LLM report as "clear" and "actionable" versus the baseline's "confusing" and "anxiety-inducing" validates our multi-modal approach.

Despite these promising results, we acknowledge several limitations. First, our datasets, while augmented, were primarily recorded in quiet, clinical environments. The models' performance in noisy, real-world settings (e.g., a home environment with background noise) has not yet been tested. Second, our LLM synthesizer was used in a zero-shot configuration. It was not fine- tuned on medical conversations and may, in rare cases, misinterpret technical nuances or "hallucinate" information not present in the specialist findings. Future work should focus on addressing these limitations.

Conclusion

For over a century, doctors have used stethoscopes to listen for heart disease, but the method has real limitations sounds get distorted, different doctors hear different things, and mistakes happen about 28% of the time. Our AI system tackles these problems by listening consistently and accurately, achieving 84– 100% accuracy in diagnosing specific heart conditions like mitral stenosis and aortic regurgitation far better than traditional exams. What makes our approach practical is that it works with real recordings from standard digital stethoscopes without expensive preprocessing. Rather than just flagging whether something sounds "normal" or "abnormal," our system tells doctors exactly what disease is present, enabling immediate treatment decisions. While limitations remain recordings need decent quality and pediatric testing is ongoing this research offers real hope for millions lacking access to expert cardiac care. By bringing AI-powered heart screening to rural clinics and underserved communities, we can detect disease earlier and ensure everyone gets a fair chance at better health.

Limitations

Our research identified several pressing problems specific to developing and deploying AI-based cardiac and lung sound analysis systems. The most significant challenge was the lack of large, diverse, real-world datasets; most available data was clinically recorded under ideal conditions, leading to poor generalization when exposed to noisy environments in field settings. High sensitivity to environmental noise and incorrect stethoscope placement frequently caused degraded signal quality and reduced diagnostic accuracy. We also encountered inconsistencies in expert labeling, making training data less reliable and capping peak performance. The model's accuracy dropped sharply with pediatric cases and patients having atypical anatomies, highlighting a gap in generalizing beyond adults. Notably, rare cardiac and lung conditions remained difficult to detect due to their underrepresentation in our data. Hardware differences and session variability further led to fluctuating results, indicating insufficient robustness. In production, our language model occasionally hallucinated clinical findings in its generated reports, raising risks in real-world deployment.

Continuous internet connectivity was essential for cloud-based analysis, limiting practical use in rural and semi-urban locations. Finally, despite good predictive results, our system's decision-making process remained hard to interpret, impacting clinical trust and adoption. Addressing these focused research problems is pivotal to improving the reliability, safety, and utility of AI-driven auscultation solutions.

Future Scope

In the next phase of our work, we plan to develop an AI-enabled digital stethoscope integrated with a cloud-based analysis platform. The stethoscope will clean noisy cardiac recordings automatically, then send them securely to our cloud system where our AI model analyzes the data and returns diagnostic results within seconds. This approach eliminates the need for expensive equipment at clinics and allows even remote health centers to access expert- level cardiac diagnosis. Community health workers can use the AI stethoscope to screen patients, with the cloud system instantly flagging abnormal findings that require specialist attention.

By combining affordable hardware with cloud intelligence, we aim to make reliable cardiac screening accessible anywhere reducing missed diagnoses and bringing consistent, objective diagnosis to patients who currently lack access to proper cardiac evaluation.

References

A. Alhasani, et al., *"Heartbeat Sound Classification Using Mel- Spectrogram and CNN Optimized by Frilled Lizard Algorithm for Cardiovascular Disease Detection," 2025.*

Y. Liu, et al., *"Cardiovascular Sound Classification Using Neural Networks," 2025.*

M. Alquran, et al., *"Deep Learning Models for Segmenting Phonocardiogram Signals," 2025.*

A. Al-Tam, et al., *"Hybrid Transformer-CNN Networks for Cardiovascular Signal Analysis," 2024.*

E. Partovi, et al., *"A Review on Deep Learning Methods for Heart Sound Signal Analysis," 2024.*

S. Ahmed, et al., *"Feature Selection and Explainable Machine Learning for Lung Disease Recognition," 2023.*

S. Kumar, et al., *"Ensemble Deep Learning Approaches for Automated Heart Sound Analysis," 2023.*

M. Harimi, et al., *"Heart Sounds Classification: Application of CyTex- DCNN," 2023.*

Y. Liu, et al., *"Deep CNN-based Respiratory Sound Analysis for Disease Classification," 2022.*

Y. Yao, X. Zhang, L. et al., *"Automated Valvular Heart Disease Detection Using Heart Sound with a Deep Learning Algorithm," 2022.*

X. Zhao, et al., *"Attention-Based RNNs for Murmur Detection in Pediatric Heart Sounds," 2022.*

U. Perna, et al., *"Automated Detection of Respiratory Diseases Using Chest Sounds and Machine Learning," 2021.*

S.K. Yadav, S. Kwon, et al., *"Artificial intelligence-based classification of heart sounds using CNN and transfer learning," 2021.*

G. Rao, et al., *"Real-World Applications of AI-Based Lung Sound Analysis," 2021.*

Y. Lv, et al., *"Artificial Intelligence-Assisted Auscultation in Detecting Abnormal Heart Sounds," 2021.*