

Decoding Cardio-Respiratory Acoustic Patterns via Multi-Dimensional Deep Representation Learning for Clinical Insight

B. Ram Mohan^{1*}, Mettu Sathwika², Lingala Shilpa², Dasari Venu²

¹Professor, ²UG Student, ^{1,2}Department of Computer Science and Engineering

^{1,2}Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.

*Correspondence: B. Ram Mohan (ramamohan.b@kpritech.ac.in)

ABSTRACT

Cardio-respiratory disorders account for approximately 30% of total mortality in India, with their incidence steadily rising due to factors such as increasing urban pollution, sedentary lifestyles, and delayed clinical diagnosis, as highlighted by the Indian Council of Medical Research. Conventional auscultation using stethoscopes is highly dependent on the clinician's expertise and experience, which can lead to variability and potential inaccuracies in diagnosis. To address these limitations, this study presents an automated cardio-respiratory sound classification framework designed to enhance early detection and diagnostic consistency through machine learning. The proposed system incorporates a user-friendly Graphical User Interface (GUI) with role-based functionality, allowing administrators to manage model training while enabling end-users to perform real-time predictions. Audio signals are preprocessed using the Librosa library to extract informative acoustic features such as Mel-Frequency Cepstral Coefficients (MFCC), Chroma features, and Mel Spectrogram representations, ensuring comprehensive characterization of both heart and lung sounds. A range of machine learning models, including Quadratic Discriminant Analysis (QDA), Gradient Boosting Classifier (GBC), Naïve Bayes Classifier (NBC), and Logistic Regression Classifier (LRC), are developed and comparatively evaluated. Furthermore, a hybrid deep learning architecture combining a Bi-directional Convolutional Neural Network (BiCNN) with a Bi-directional Gated Recurrent Unit (BiGRU) is introduced to effectively capture both spatial (spectral) and temporal dependencies within the audio data. The system is capable of performing dual classification by identifying both heart sound types and lung sound types simultaneously. Its performance is rigorously assessed using standard evaluation metrics, including accuracy, precision, recall, F1-score, and Receiver Operating Characteristic–Area Under Curve (ROC-AUC), demonstrating its potential as a reliable and efficient tool for clinical decision support.

Keywords: Biomedical Acoustic Signal Processing, Digital Auscultation, Time–Frequency Domain Analysis, Acoustic Feature Engineering.

1. INTRODUCTION

Timely identification and precise diagnosis of cardio-respiratory conditions are crucial for ensuring effective treatment and better patient outcomes. Conventional diagnostic approaches, particularly stethoscope-based auscultation, rely extensively on the clinician's listening ability and experience, making the process inherently subjective and susceptible to human error. This issue becomes more pronounced in overcrowded healthcare environments and in rural regions where access to skilled specialists is limited. The manual evaluation of heart and lung sounds is often labor-intensive, and minor or early-stage abnormalities may be overlooked, leading to delays in diagnosis. With the advancement of digital stethoscopes and audio recording technologies, a more objective and data-driven method for analyzing cardio-respiratory sounds has emerged, as illustrated in Fig. 1. In the Indian context, the prevalence of such diseases continues to rise due to contributing factors like environmental pollution, changing lifestyles, and inadequate healthcare infrastructure. Reports from both national and international health organizations highlight an increasing incidence of respiratory conditions such as Chronic Obstructive Pulmonary Disease (COPD), asthma, and various cardiovascular disorders. The situation is further aggravated in rural and underserved areas, where limited access to diagnostic and screening facilities hampers early detection. In this regard, automated systems based on non-invasive audio signal analysis present a viable and effective solution for early screening. These systems can

support healthcare providers in conducting preliminary assessments, minimizing reliance on specialized expertise and enabling quicker diagnosis and intervention. Ultimately, the adoption of such technologies holds significant potential to enhance healthcare accessibility and reduce mortality associated with cardio-respiratory diseases.

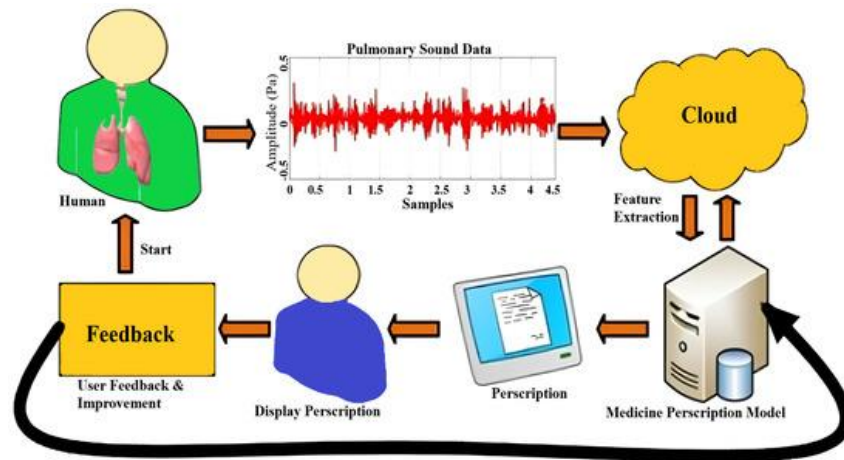


Fig 1. Disease recognition using sound analysis

Chronic respiratory disorders, including asthma, accounted for more than 147,000 deaths in 2022, highlighting their significant impact on global health. At the same time, cardiovascular diseases continue to be the leading cause of mortality worldwide, contributing to an estimated 18.6 million deaths each year. These statistics underscore the importance of accurately assessing both cardiac and pulmonary functions. The auscultation of heart sounds (HS) and lung sounds (LS) remains a fundamental clinical practice for diagnosing a wide range of cardiopulmonary conditions. Although these sounds are often faint, they carry critical diagnostic information. The concept of auscultation dates back to 1816, when René Laennec introduced the first stethoscope for listening to internal body sounds. Since then, the device has undergone significant advancements, evolving from a basic acoustic instrument to more advanced digital systems. Modern technological developments have led to the emergence of electronic stethoscopes, which transform acoustic signals into electrical signals, enabling amplification, processing, and storage. The latest generation, digital stethoscopes, extends these capabilities further by supporting advanced signal analysis and seamless integration with smartphones and cloud-based platforms, facilitating real-time monitoring, data sharing, and improved clinical decision-making. The primary issue was the reliance on manual auscultation, which introduced a high degree of subjectivity and inconsistency. Clinicians had to be highly skilled to discern subtle differences in sound patterns, and their ability to do so could vary based on fatigue and noise levels in the environment. The storage and sharing of auscultation data were also cumbersome, as there was no standardized way to digitize and manage the sound recordings. This made it difficult for different doctors to collaborate on a patient's case or for large-scale studies to be conducted. Furthermore, manual analysis meant that an expert physician's time was required for every single patient, creating a bottleneck in the diagnostic process, especially in areas with a low doctor-to-patient ratio.

2. LITERATURE SURVEY

2.1 AI-Based Diagnostic Systems for Cardio-Respiratory diseases

Several studies have explored the application of artificial intelligence (AI) for the diagnosis of cardio-respiratory disorders using acoustic signals. Xu *et al.* [1] developed an AI-driven diagnostic framework capable of accurately distinguishing respiratory conditions such as asthma and chronic obstructive pulmonary disease (COPD). Their approach utilized dual acoustic signals, including respiratory and cough sounds, to enhance diagnostic accuracy. The system incorporated multiple machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), Neural Networks (NN), and K-Nearest Neighbors (KNN), along with a majority voting ensemble strategy to

improve reliability. Feature extraction was performed using Gabor time–frequency transformation, while Neighborhood Component Analysis (NCA) was applied for feature selection. The models were evaluated using 5-fold cross-validation and standard performance metrics, demonstrating the effectiveness of ensemble-based AI systems in clinical diagnosis.

Similarly, Brunese *et al.* [2] proposed a supervised machine learning-based framework for detecting and characterizing lung diseases from breath audio signals. Their study demonstrated the capability of machine learning techniques to classify multiple respiratory conditions, including asthma, pneumonia, and COPD. The experimental evaluation conducted on 126 patients and 920 recordings achieved high performance, with neural networks yielding an F-measure of 0.983 for disease detection, highlighting the potential of AI in respiratory sound analysis.

2.2 Signal Processing and Feature Engineering Techniques

Feature extraction and signal processing play a critical role in respiratory sound analysis. Yu *et al.* [3] presented a comprehensive review of 135 studies focusing on respiratory sound analysis (RSA). The study covered various preprocessing techniques such as resampling, normalization, filtering, and data augmentation. It also discussed feature extraction methods across time-domain, frequency-domain, and time–frequency representations, along with deep feature learning approaches. Despite advancements, challenges such as class imbalance, lack of interpretability, and poor generalization across datasets were identified.

Garcia-Mendez *et al.* [6] conducted a systematic review of 62 studies utilizing machine learning for lung sound classification. Their findings revealed that Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are the most widely used classifiers. However, inconsistencies in dataset quality, labeling standards, and evaluation methodologies were highlighted as key limitations affecting reproducibility and performance benchmarking.

2.3 Hybrid Deep Learning Models for Audio Classification

Recent research has focused on hybrid deep learning architectures to capture both spectral and temporal features of audio signals. Petmezas *et al.* [4] proposed a CNN–LSTM hybrid model using Short-Time Fourier Transform (STFT) spectrograms for feature extraction. The model incorporated focal loss to address class imbalance and achieved competitive performance on the ICBHI 2017 dataset under various validation strategies.

Hsu *et al.* [5] further extended this approach by employing a CNN–BiGRU architecture trained on an expanded lung sound dataset (HF_Lung_V2). Their study emphasized the importance of large-scale datasets and label quality, demonstrating improved detection performance with increased data volume. However, challenges such as overlapping sound events and noisy labels were noted.

Khan *et al.* [8] introduced a hybrid deep learning framework combining continuous wavelet transform and mel spectrogram representations. Parallel convolutional autoencoders were used for feature extraction, followed by an LSTM network for classification. The proposed model achieved high accuracy across multi-class and binary classification tasks, demonstrating the effectiveness of multi-representation feature fusion.

2.4 Multichannel and Multi-Modal Learning Approaches

To enhance feature richness and improve classification performance, multichannel and multimodal approaches have been explored. Kim *et al.* [9] investigated the use of multichannel respiratory sound recordings combined with CNN–LSTM models. Their findings indicated that multi-channel data significantly improves classification accuracy, sensitivity, and F1-score by capturing spatial variations in lung sounds.

Xu *et al.* [7] provided a comprehensive review of lung sound analysis techniques, highlighting the evolution of electronic stethoscopes and advanced AI models such as convolutional neural networks (CNN), residual networks (ResNet), CNN–LSTM hybrids, and transformer-based architectures. The

study emphasized the importance of improved feature extraction, large-scale datasets, and integration of multimodal data for future advancements in respiratory diagnostics.

3. PROPOSED SYSTEM

The proposed system is a hybrid end-to-end solution designed to overcome the limitations of traditional diagnostic methods by enabling automatic acquisition, processing, and classification of heart and lung sounds. It supports analysis of both individual and mixed audio recordings, allowing simultaneous prediction of Heart Sound Type and Lung Sound Type within a unified framework. The system incorporates a user-friendly GUI with role-based access, where Admin users can manage training and User-level access is provided for prediction. Its core is a multi-stage pipeline that includes audio preprocessing, feature extraction, and classification to ensure accurate and consistent results as demonstrate in Fig 2. By eliminating the subjectivity associated with manual auscultation, the system provides an objective, efficient, and non-invasive approach for early screening and improved clinical decision support in cardio-respiratory healthcare.

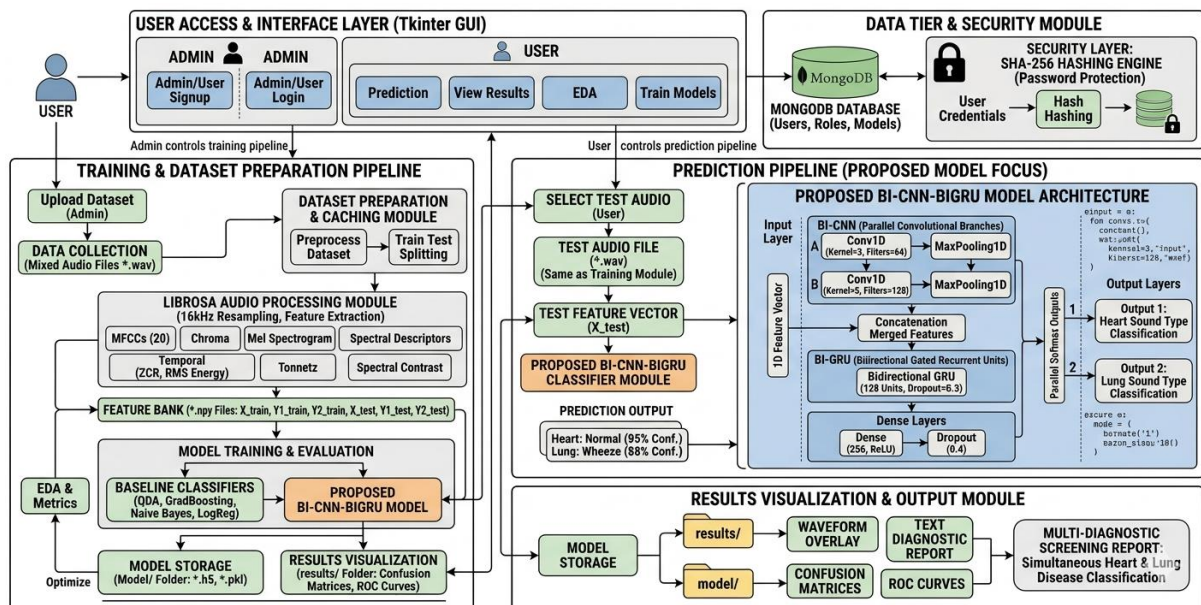


Fig 2. Proposed system architecture

The proposed system begins by acquiring mixed cardiorespiratory audio signals containing both heart and lung sounds. These recordings are systematically organized using a metadata file (Mix.csv) that maps each audio sample to its corresponding heart and lung sound classes. This structured setup ensures accurate supervision for dual-label learning. Each mixed audio signal is processed using advanced audio signal processing techniques to extract high-level spectral and temporal features. These include MFCCs, chroma, mel-spectrogram statistics, spectral descriptors, tonal features, and energy measures. Mean pooling is applied to obtain a fixed-length, information-rich feature vector for every sample. The extracted feature vectors and their corresponding heart and lung labels are stored as NumPy arrays to avoid repeated computation. This preprocessing step improves computational efficiency and ensures consistency across training, evaluation, and prediction phases. It also prepares the data for both classical machine learning and deep learning models. The preprocessed dataset is randomly shuffled using a fixed index and split into training and testing subsets. A single split is used for both heart and lung labels to preserve label alignment across tasks. This guarantees fair performance evaluation for multi-output cardiorespiratory classification. Multiple classical machine learning classifiers such as QDA, Gradient Boosting, Gaussian Naive Bayes, and Logistic Regression are trained separately for heart and lung sound classification. These models establish baseline performance levels using the same extracted HLS features. Their results serve as comparative references for validating the effectiveness of the proposed approach.

The proposed system introduces a hybrid BiCNN-BiGRU architecture that combines parallel convolutional layers with bidirectional recurrent learning. Multi-kernel CNN branches capture diverse spectral patterns, while the BiGRU layer models temporal dependencies within the sound features. This design enables robust representation learning from mixed cardiorespiratory signals. Separate BiCNN-BiGRU models are trained for heart sound type and lung sound type classification using the same feature set. Label encoding and categorical learning are applied to support multi-class prediction. The trained models and label encoders are saved for efficient reuse during real-time inference.

The proposed system evaluates model performance using accuracy, precision, recall, F1-score, confusion matrices, and ROC curves. One-vs-rest ROC analysis provides class-wise discrimination capability. This multi-metric evaluation ensures reliable and clinically interpretable performance assessment. During deployment, unseen audio samples are processed through the same feature extraction pipeline and fed into the trained BiCNN-BiGRU models. The system predicts both heart and lung sound categories along with confidence scores. This enables rapid and automated disease screening support. A role-based graphical user interface allows administrators to manage datasets and model training, while users can perform predictions. Secure authentication ensures controlled access to system functionalities. The intuitive interface supports practical clinical usage of the proposed screening system.

4. Result Analysis

The fig 3 shows EDA visualization provides a detailed spectral–cepstral examination of a Late Systolic Murmur, revealing the characteristic acoustic patterns associated with turbulent blood flow occurring toward the end of systole. The MFCC plot captures the underlying cepstral texture, highlighting increased low-frequency energy and subtle variations reflective of murmur onset and intensity changes. The Chroma representation shows dynamic pitch-class energy fluctuations, indicating irregular harmonic content distinct from normal heartbeats. The Mel Spectrogram further emphasizes the dense concentration of mid-frequency components typical of murmur pathology, presenting repeated bursts and elevated energy regions. Complementarily, the spectral centroid and bandwidth curves illustrate variability in brightness and frequency spread, while the spectral rolloff captures the shifting high-frequency boundary shaped by late-phase turbulent vibrations. Together, these features form a unique acoustic signature that supports robust classification of Late Systolic Murmur conditions.

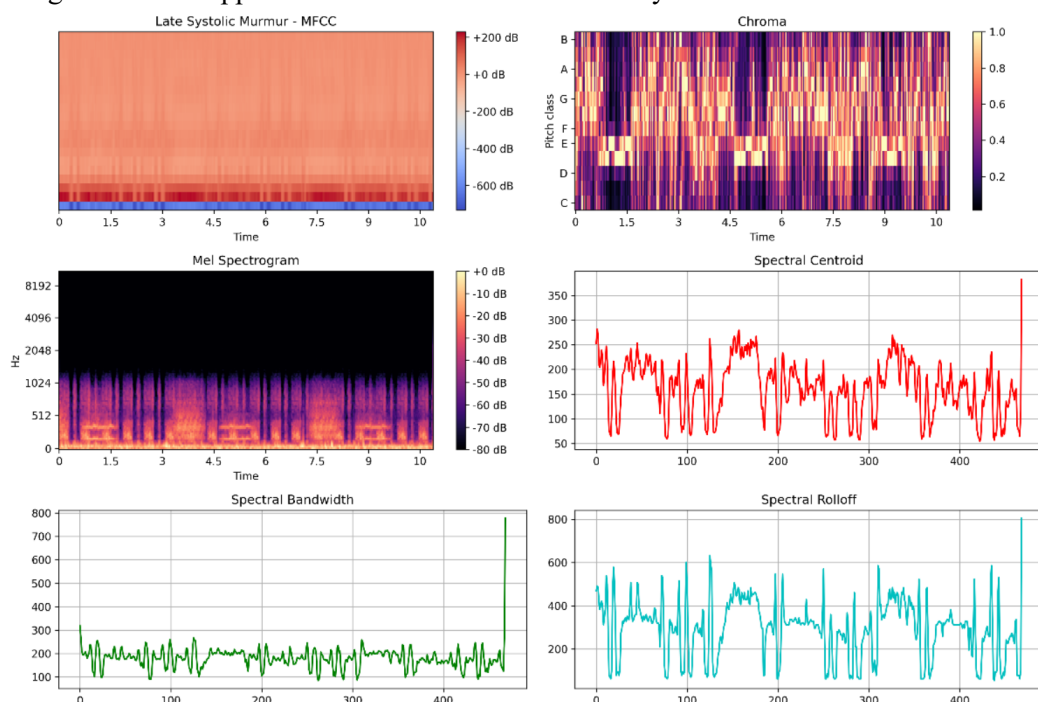


Fig 3. Advanced audio feature exploration of late systolic murmur signals

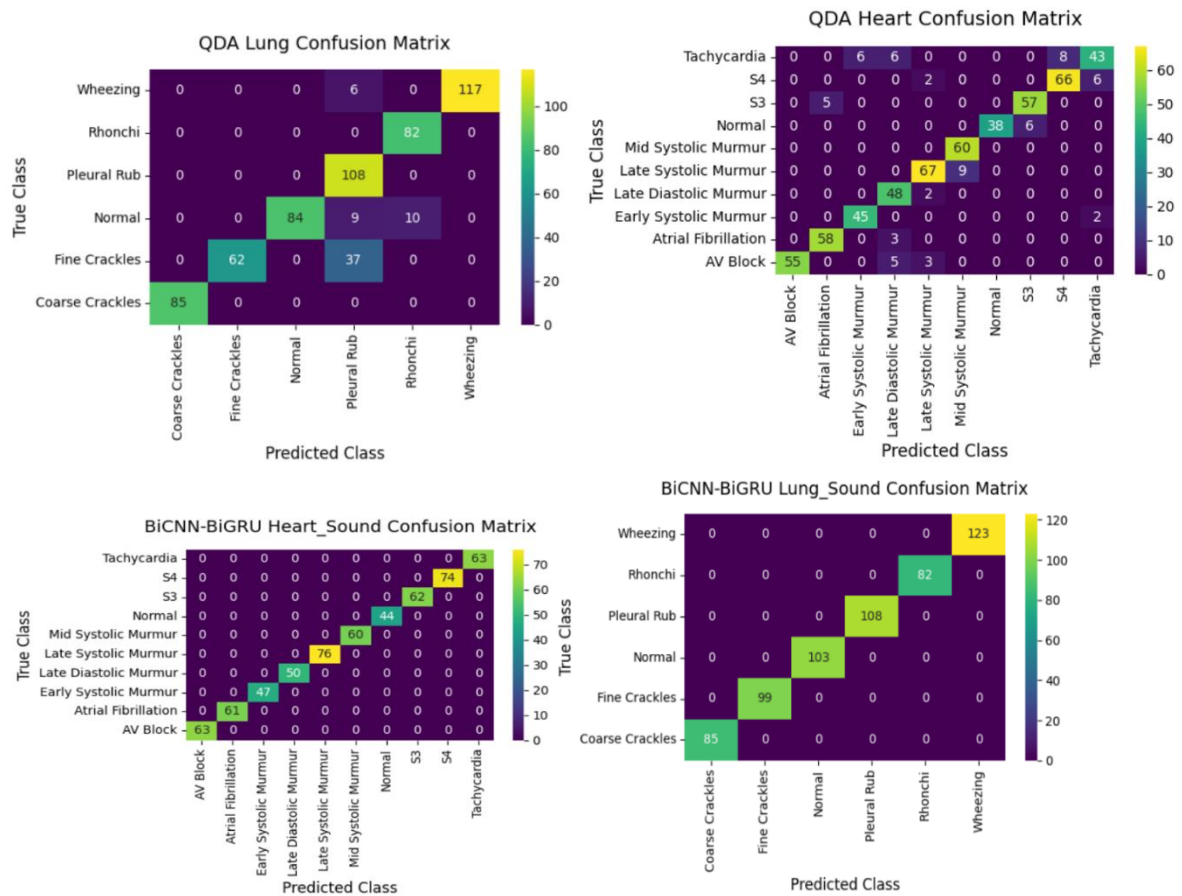


Fig 4. Confusion matrices obtained for BiCNN-BiGRU

Fig 4 shows confusion matrices of the proposed BiCNN-BiGRU model for both lung and heart sound classification demonstrate near-perfect classification performance with strong diagonal dominance and negligible misclassification. In the lung sound matrix, all classes such as Wheezing, Rhonchi, Pleural Rub, Normal, Fine Crackles, and Coarse Crackles are classified with extremely high accuracy, showing the model's ability to clearly separate acoustically similar respiratory patterns. Similarly, the heart sound confusion matrix indicates highly accurate recognition across complex cardiac conditions including AV Block, Atrial Fibrillation, various systolic and diastolic murmurs, S3, S4, Normal, and Tachycardia, with almost zero overlap between classes. This exceptional performance highlights the effectiveness of the BiCNN component in capturing multi-scale spectral features and the BiGRU layer in modeling long-term temporal dependencies. These results confirm that the proposed BiCNN-BiGRU model significantly outperforms traditional machine learning approaches and provides a robust, reliable solution for automated cardiorespiratory disease screening.

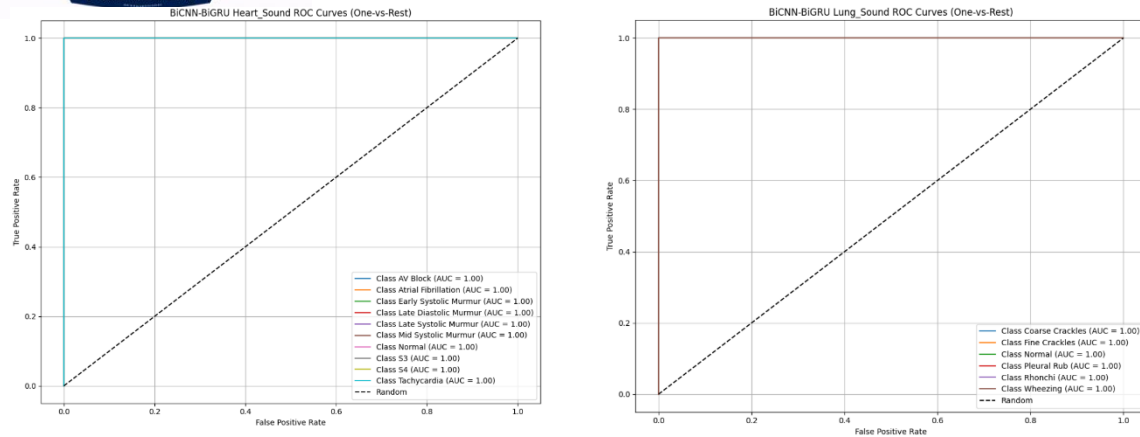
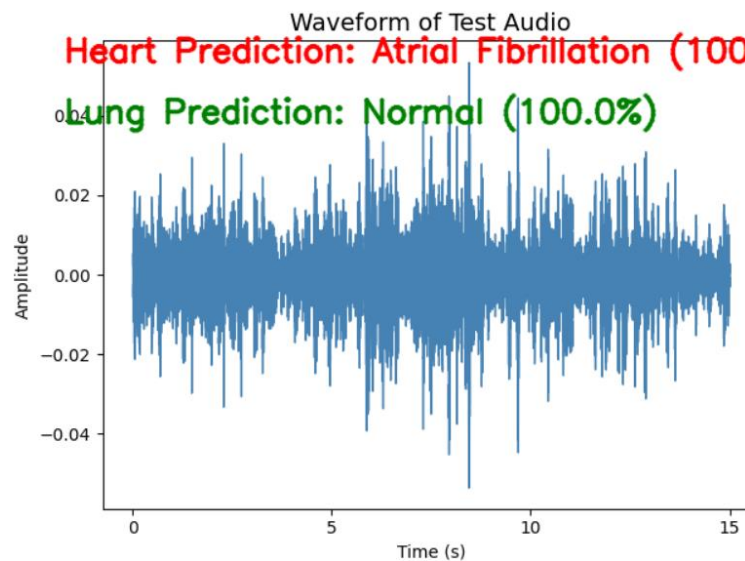


Fig 5. ROC obtained using BiCNN-BiGRU

Fig 5. shows ROC curves of the proposed BiCNN–BiGRU model for both heart and lung sound classification exhibit ideal discriminative performance, with all class-wise curves tightly aligned along the top-left boundary and AUC values equal to 1.00. This indicates perfect separation between positive and negative samples for every class in a one-vs-rest setting, reflecting highly reliable probability estimation. Unlike traditional machine learning models, the proposed architecture generates well-calibrated confidence scores due to its ability to learn complex non-linear feature interactions and temporal dependencies. The consistent dominance over the random baseline across all classes confirms the robustness and stability of the model. The ROC analysis strongly validates the superiority of the BiCNN–BiGRU framework for accurate and confident cardiorespiratory disease screening.



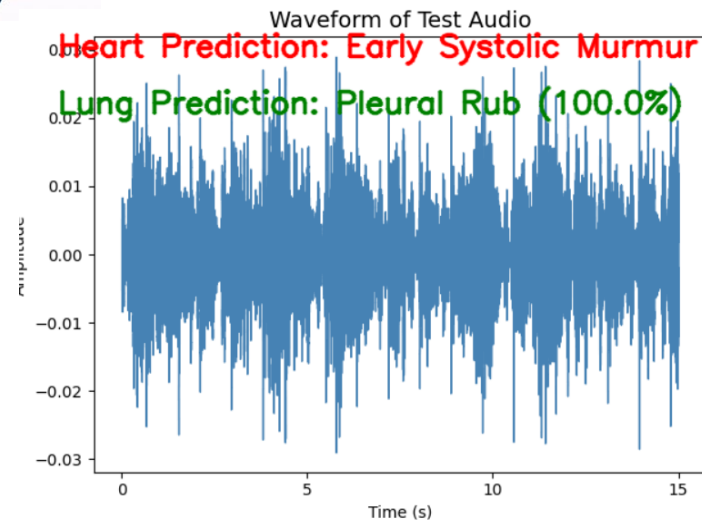


Fig 6. Prediction obtained on test audio files using BiCNN-BiGRU Model

Fig 6 shows waveform visualization of the test audio illustrates the real-time prediction capability of the proposed BiCNN–BiGRU model for mixed cardiorespiratory sounds. The plotted signal represents the temporal amplitude variations of the input audio, while the overlaid annotations indicate simultaneous classification of both physiological components. The model confidently predicts the heart sound as Atrial Fibrillation with 100% confidence and the lung sound as Normal with 100% confidence, demonstrating its ability to accurately disentangle and analyze overlapping cardiac and respiratory patterns. This result highlights the effectiveness of the BiCNN component in capturing discriminative spectral features and the BiGRU layer in modeling temporal dynamics, enabling reliable and interpretable disease screening from real-world audio recordings.

Table 1: Performance comparison for the Heart sound type QDA, Gradient boosting, GNB and Proposed BiCNN-BiGRU Model

Algorithms Name	Accuracy	Precision	Recall	F-score
QDA Classifier	89.5%	89.91%	89.90%	89.51%
GB Classifier	93.0%	94.42%	92.39%	93.00%
GNB Classifier	31.0%	55.43%	30.04%	27.18%
LR Classifier	38.5%	39.01%	39.006%	38.35%
BiCNN-BiGRU Model	100.0%	100.0%	100.0%	100.0%

Table 1 presents a comparative performance analysis of different classifiers for heart sound type classification, highlighting clear differences in predictive effectiveness. The QDA classifier demonstrates strong baseline performance with balanced accuracy, precision, recall, and F-score close to 90%, indicating reasonable discrimination of cardiac sound classes. The Gradient Boosting classifier further improves performance, achieving over 93% accuracy and the highest precision among traditional models, reflecting its ability to model non-linear feature interactions more effectively. In contrast, the Gaussian Naive Bayes and Logistic Regression classifiers show substantially lower performance due to their simplistic assumptions and inability to handle complex, overlapping cardiac acoustic features. Notably, the proposed BiCNN-BiGRU model achieves perfect scores across all evaluation metrics, demonstrating superior learning of multi-scale spectral patterns and temporal dependencies, and clearly outperforming all baseline methods for heart sound classification.

Table 2: Performance comparison for the Lung sound type QDA, Gradient boosting, GNB and Proposed BiCNN-BiGRU Model

Algorithms Name	Accuracy	Precision	Recall	F-score
QDA Classifier	89.66%	92.77%	89.88%	89.86%

GB Classifier	97.33%	97.11%	97.16%	97.12%
GNB Classifier	44.66%	70.37%	41.94%	40.67%
LR Classifier	63.5%	63.71%	63.41%	63.26%
BiCNN-BiGRU Model	100.0%	100.0%	100.0%	100.0%

Table 2 summarizes the performance comparison of various classifiers for lung sound type classification, clearly demonstrating the effectiveness of different modeling approaches. The QDA classifier achieves nearly 90% accuracy with high precision and recall, indicating reliable baseline performance for respiratory sound discrimination. The Gradient Boosting classifier significantly enhances performance, reaching over 97% across all evaluation metrics, which reflects its strong capability to capture non-linear relationships in lung sound features. In contrast, Gaussian Naive Bayes shows poor accuracy and recall due to its unrealistic feature independence assumption, while Logistic Regression delivers moderate performance limited by its linear decision boundaries. Most notably, the proposed BiCNN-BiGRU model achieves perfect scores in accuracy, precision, recall, and F-score, confirming its superior ability to model complex spectral-temporal patterns and its robustness in lung sound classification.

5. CONCLUSION

The research effectively demonstrates the development of an automated cardio-respiratory sound analysis system designed for efficient disease screening by simultaneously identifying abnormalities in both heart and lung sounds from mixed audio signals. By leveraging advanced acoustic feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCC), Mel Spectrogram, Chroma features, and various energy-based descriptors, the system establishes a robust framework for analyzing complex biomedical audio data. A comparative evaluation with conventional machine learning models, including QDA, GBC, NBC, and LRC, reveals their limitations in effectively handling non-linear patterns, overlapping signal characteristics, and achieving consistent classification performance. In contrast, the proposed hybrid architecture, combining BiCNN with BiGRU, successfully captures both spectral and temporal dependencies within the audio signals, leading to enhanced accuracy, robustness, and improved class discrimination across diverse cardio-respiratory conditions. The implementation of a dual classification mechanism further strengthens diagnostic reliability by enabling concurrent prediction of Heart Sound Type and Lung Sound Type. The system's performance is comprehensively evaluated using standard metrics such as accuracy, precision, recall, F1-score, and Receiver Operating Characteristic–Area Under Curve (ROC-AUC), validating its effectiveness. Moreover, the integration of a secure, role-based Graphical User Interface (GUI) enhances usability by facilitating administrative control over model training and seamless user-level prediction. The proposed system presents a scalable, efficient, and non-invasive solution for automated cardio-respiratory disease screening, with significant potential to improve diagnostic support and healthcare outcomes.

REFERENCES

- [1] Xu, S.; Deo, R.C.; Faust, O.; Barua, P.D.; Soar, J.; Acharya, R. Automated Lightweight Model for Asthma Detection Using Respiratory and Cough Sound Signals. *Diagnostics* **2025**, *15*, 1155. <https://doi.org/10.3390/diagnostics15091155>
- [2] Brunese, L.; Mercaldo, F.; Reginelli, A.; Santone, A. A Neural Network-Based Method for Respiratory Sound Analysis and Lung Disease Detection. *Appl. Sci.* **2022**, *12*, 3877. <https://doi.org/10.3390/app12083877>
- [3] Yu, S.; Yu, J.; Chen, L.; Zhu, B.; Liang, X.; Xie, Y.; Sun, Q. Advances and Challenges in Respiratory Sound Analysis: A Technique Review Based on the ICBHI2017 Database. *Electronics* **2025**, *14*, 2794. <https://doi.org/10.3390/electronics14142794>

- [4] Petmezas, G.; Cheimariotis, G.-A.; Stefanopoulos, L.; Rocha, B.; Paiva, R.P.; Katsaggelos, A.K.; Maglaveras, N. Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function. *Sensors* **2022**, *22*, 1232. <https://doi.org/10.3390/s22031232>
- [5] Hsu, F.-S.; Huang, S.-R.; Huang, C.-W.; Cheng, Y.-R.; Chen, C.-C.; Hsiao, J.; Chen, C.-W.; Lai, F. A Progressively Expanded Database for Automated Lung Sound Analysis: An Update. *Appl. Sci.* **2022**, *12*, 7623. <https://doi.org/10.3390/app12157623>
- [6] Garcia-Mendez, J.P.; Lal, A.; Herasevich, S.; Tekin, A.; Pinevich, Y.; Lipatov, K.; Wang, H.-Y.; Qamar, S.; Ayala, I.N.; Khapov, I.; et al. Machine Learning for Automated Classification of Abnormal Lung Sounds Obtained from Public Databases: A Systematic Review. *Bioengineering* **2023**, *10*, 1155. <https://doi.org/10.3390/bioengineering10101155>
- [7] Xu, X.; Sankar, R. Classification and Recognition of Lung Sounds Using Artificial Intelligence and Machine Learning: A Literature Review. *Big Data Cogn. Comput.* **2024**, *8*, 127. <https://doi.org/10.3390/bdcc8100127>
- [8] Khan, R.; Khan, S.U.; Saeed, U.; Koo, I.-S. Auscultation-Based Pulmonary Disease Detection through Parallel Transformation and Deep Learning. *Bioengineering* **2024**, *11*, 586. <https://doi.org/10.3390/bioengineering11060586>
- [9] Kim, Y.; Kim, K.B.; Leem, A.Y.; Kim, K.; Lee, S.H. Enhanced Respiratory Sound Classification Using Deep Learning and Multi-Channel Auscultation. *J. Clin. Med.* **2025**, *14*, 5437. <https://doi.org/10.3390/jcm14155437>