Multi-Model Cardiac Screening Using PCG And ECG Signals

Shridhani M [1] and V Subapriya [2*]

Department of Computer Science and Engineering,
Sathyabama Institute of Science and Technology, Chennai, India

sdhani2020@gmail.com,subapriya.cse@sathyabama.ac.in

Abstract. Cardiac evaluation at initial stages depends on precise handling of biological data. A method involving separate models appears here, examining PCG and ECG signals through distinct computational paths. Instead of merging information early, processing stays independent, allowing use even if only one signal type exists. From the CirCor collection, heart audio recordings feed into the PCG pathway. After signal processing ,features tied to timing and sound properties undergo extraction. Following this step, classification relies on a Gradient Boosting framework. This part identifies abnormal sounds, estimate valve location, assess severity levels, and calculates murmur likelihood scores. Analysis of heart signals uses data drawn from the PTB-XL collection, applying a Random Forest method to assess irregularities and perform analysis across rhythm, waveform form, subclass, superclass, and severity levels. Insights into model decisions emerge via SHAP-driven examination, scrutiny of biological markers, graphical displays of waveforms, together with spectrogram visualizations for PCG and R-R-interval analysis for ECG signals. Access occurs through an online platform requiring only a patient ID, producing distinct reports: one for phonocardiogram findings, another for electrocardiographic results. With focus placed on clarity and separation of sensing channels, the framework supports usable, supporting automated and interpretable cardiac screening.

Keywords: Phonocardiogram (PCG), Electrocardiogram (ECG), Gradient Boosting, Random Forest, Multi-Label Classification, SHAP interpretability, biomarker analysis.

1 INTRODUCTION

Despite being a major global cause of death, cardiovascular conditions strain medical services unevenly across regions. Detection at an early stage often influences health results more than later interventions. Availability of clear, trustworthy diagnostics varies widely, especially where resources are limited. Electrical patterns of the heart appear through ECG recordings, one tool among several. Mechanical vibrations produced by valves and blood flow show up in PCG data instead. One does not replace the other - they inform distinct aspects of function. Interpretation without human experts remains difficult even with automation attempts. Combined insights exist, though processing them clearly is still complex. Each signal type brings value that stands apart from mere technical detail.

Available at scale, datasets like PTB-XL [1] for ECG signals and CirCor DigiScope [2] for PCG recordings have pushed progress in machine learning applied to heart signal classification. Performance stands out especially when deep networks handle more than one condition at once within ECG readings [10]–[12], [20]. In parallel, work on detecting murmurs through PCG relies increasingly on refined features and modern network designs, allowing systems to spot valve issues and irregular sounds automatically [14]–[17]. Even so, moving into actual use demands methods that do not only predict well but also remain stable across diverse clinical environments while offering clarity in their decisions. Lately, research has looked into combining ECG and PCG signals to capture broader heart activity details [4]–[8]. Rather than simple merging, many methods apply complex models such as convolutions, transformers, or grouped network structures [6], [7], [8]. Despite gains in accuracy, blending dissimilar data too soon might restrict adaptability, raise processing demands, while obscuring insights. At the same time, alternative training setups - federated or decentralized - have emerged aiming at better privacy handling and system growth potential [9].

In medicine, demand grows for transparent artificial intelligence, especially where decisions affect health outcomes. Because clarity matters in clinics, methods like SHAP - developed by Lundberg and Lee [3] - offer mathematically sound ways to trace how inputs shape outputs. Rather than accept predictions without insight, recent studies embed interpretability within models that analyze heart signals such as ECG and PCG [4], [10], [13], [17]. Instead of opaque results, these systems reveal which features drive conclusions, supporting understanding over blind reliance. Thus,

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,
https://doi.org/10.2991/978-94-6239-693-7_12

progress shifts focus: prediction alone is insufficient when reasoning must also be visible. A further key issue in categorizing heart signals lies in uneven class distribution, especially within multi-label ECG assessments where uncommon medical conditions appear infrequently [18], [19]. When data skews heavily toward certain labels, measures like recall and macro averages often decline - this calls for precise assessment methods alongside resilient model design.

Driven by such difficulties, this work introduces a segmented system for heart assessment where ECG and PCG inputs follow distinct analytical routes instead of merging at an initial stage. Through the ECG route, information drawn from PTB-XL datasets [1] feeds into a multi-output Random Forest model designed to identify patterns across multiple labels. Alongside it, the PCG channel applies Gradient Boosting methods on CirCor samples [2] with focus on recognizing murmurs. To clarify decision logic, insights emerge via SHAP-driven examination [3], making output reasoning accessible without oversimplification. Separating these modalities during computation allows adaptability to remain intact, aids clarity in results, and aligns with real-world integration within web platforms organized around individual patient identifiers. Clarity and room to grow define the intent behind this method meant for automatic evaluation of cardiac health.

The main contributions of the work are

- A modular cardiac screening framework that processes ECG and PCG signals through independent pipelines instead of early multimodal fusion.
- Implementation of modality-specific interpretable machine learning models (Random Forest for ECG and Gradient Boosting for PCG) within independent processing pipelines.
- Incorporation of SHAP-based explainability combined with clinically meaningful waveform biomarkers.
- Multi-level diagnostic output including superclass, subclass, rhythm, waveform form, and severity grading.
- Deployment through a patient ID-based web interface for practical and flexible clinical usage

2 RELATED WORK

Progress in automatic heart condition detection now includes systems that process one data type or several together. Rather than relying solely on ECGs, some techniques merge ECG with phonocardiogram inputs - this pairing uses separate but related physiological traits to boost accuracy [4]–[9]. Deep network designs like convolutions, transformers, and grouped predictors appear frequently; these typically link signal types near the start of analysis. Even though blending early can strengthen internal representations, it sometimes limits adaptability while making results harder to trace. At the same time, many studies target multi-output prediction from ECG alone, drawing on broad collections such as PTB-XL [1], where transparent structures and refined layouts were introduced across works [10]–[13], [20]. In like manner, methods using PCG for detecting murmurs have evolved via machine learning together with deep learning approaches [14]–[17], focusing on pulling out features while applying attention to catch unusual sounds. Still, a number of these PCG setups work only in two-class settings or rely solely on dense end-to-end models lacking modularity in how they deploy. On top of this, skewed distributions within multi-label ECG data keep affecting results at higher aggregation levels, pushing demand for training plus assessment tactics mindful of imbalance [18], [19]. Even though tools meant for explanation - SHAP being one example [3] - have entered workflows analyzing heart signals, solutions combining flexibility with separation of processing paths, all the while staying clear in their reasoning, are not common. Absence in this area sets direction toward the method put forward here.

3 METHODOLOGY

A structure built in separate parts handles ECG and PCG signals using distinct paths for computation. Rather than combining inputs at an initial stage, individual analysis maintains unique traits of each signal while supporting use even if just one kind exists. Through a browser-accessible portal, records tied to patients pull matching ECG or PCG files from organized repositories. Based on what signal arrives, routing directs it to the correct processing unit.

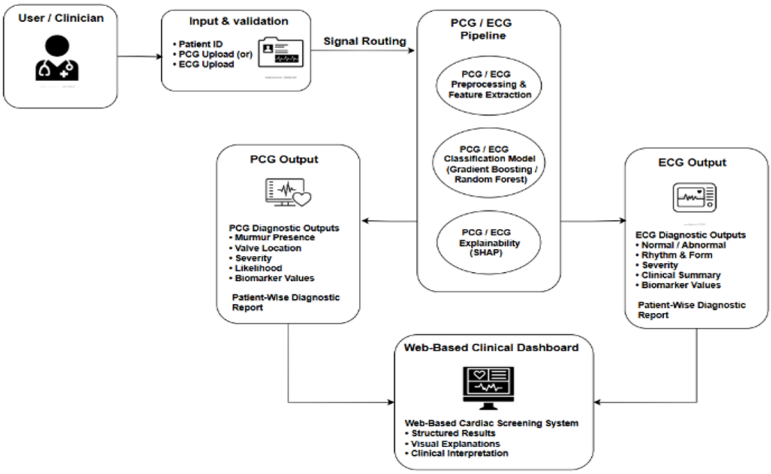

Fig. 1. Overall architecture of the proposed multi-modal cardiac screening framework.

3.1 PCG Processing Pipeline

From the CirCor DigiScope dataset come the PCG recordings. To maintain uniformity, every recording gets resampled at 4 kHz. Suppression of low-end movement interference and upper-range disturbances occurs through a fourth-order Butterworth band-pass filter. Following filtering, normalization adjusts the amplitude levels before features are extracted. Mel-Frequency Cepstral Coefficients serve as the basis for extracting acoustic properties. For every recording, thirteen of these coefficients are determined; meanwhile, measures like average and spread describe changes over time. Valve position signals join those values as inputs. Into a Gradient Boosting model flows this combined data set.

Employed to detect murmurs and assess their intensity, the Gradient Boosting model functions through structured probability outputs. Depending on threshold values that mirror medical evaluation standards, it assigns categories - Absent, Mild, Moderate, or Severe - not by label alone but through calculated likelihoods. Prediction of existence comes paired with graded judgment, shaped entirely by data-driven boundaries. Categorization emerges not from assumption, but alignment with diagnostic norms embedded in training outcomes. What results is a dual output: one indicating occurrence, another reflecting magnitude, both rooted in incremental learning processes.

For clearer clinical meaning, biomarkers taken from waveforms undergo calculation. Among them: Murmur Energy Index (MEI), along with Spectral Turbulence Index (STI), appear through Welch-based methods. High-Frequency Ratio (HFR) emerges next; so does Temporal Variability (TVAR). Interpretation of model outputs relies on SHAP values - these reflect how each MFCC feature plays a role. Diagnostic clarity finds aid via visuals - waveform displays, Mel-spectrograms included. Such images follow only after data processing completes.

3.2 ECG Processing Pipeline

From the PTB-XL dataset , ECG readings come with multi-label diagnostics sorted into a tiered system. To maintain uniformity in feature size, signal values undergo normalization via standard scaling. Layered analysis of heart irregularities becomes possible due to label organization - split between broader superclasses and more specific subclasses. A single tree does not stand alone; instead, a collection of decision paths forms predictions through aggregated votes within labeled data frameworks. Outputs emerge along several health indicators - rhythm irregularities appear alongside shape distortions, broad groupings connect with finer subtypes, while intensity degrees unfold gradually. Clinical reasoning finds reflection here: layered judgments replace simple yes-or-no conclusions. Structure guides meaning, much like how medical experts weigh evidence step by step.

Starting with clarity in mind, SHAP analysis measures how each feature influences single predictions. Instead of simple listings, visual results show ECG waveforms where irregular parts stand out clearly. Rhythm evaluation appears through inspection of R–R intervals, presented apart from broader summaries. Feature relevance, guided by SHAP logic, forms part of these displays. Without relying on opaque outcomes, visuals assist clinical review while aligning with organized diagnostic reports.

In addition to classification, waveform-derived biomarkers are computed to support clinical interpretation. These include spectral and temporal measures such as Murmur Energy Index (MEI), Spectral Turbulence Index (STI), High-Frequency Ratio (HFR), and Temporal Variability (TVAR), derived from power spectral analysis.

3.3Web-Based Deployment

The entire setup runs through a Flask-powered web platform. Instead of uploading files by hand, the system pulls existing ECG or PCG data via patient IDs. From each type of signal, a distinct analysis report emerges - featuring diagnosis labels, severity ratings, alongside clarity-focused visuals. Adaptation to live clinical settings happens smoothly due to its segmented design, yet openness and ease of use remain intact.

4 Experimental Setup

4.1 Datasets

Testing began with two open-access collections of heart signals. From one source, electrical activity patterns came through the PTB-XL group, offering layered medical tags like broad diagnosis types, detailed subgroups, timing features, and shape descriptors. Another set contributed acoustic data via the CirCor DigiScope collection - recordings focused on individual valve sounds, each marked for murmurs. Each measurement followed structured labeling protocols tied to clinical findings.

From Lead I, ECG signals sampled at 100 Hz served as raw inputs. To maintain consistent size, each recording was zero-padded- ending precisely at 2000 samples. Regarding PCG data, adjustment occurred: sample rates shifted to 4 kHz. This shift aligned timing structure regardless of valve type recorded.

4.2 Preprocessing and Feature Extraction

Beginning at the lower edge of biological relevance, ECG data passed through a fourth-order Butterworth configuration, confined within 0.5 to 45 hertz. Following this stage, disruption from rapid fluctuations diminished via wavelet-driven soft threshold application. Prior to isolating traits, amplitude scaling brought all instances into uniform range. From each processed stream, a fixed-format array emerged - embedding averages, spread metrics, root-mean-square values, shape indicators like asymmetry and peakiness. Energy distribution across frequencies formed part of the profile, alongside transitions crossing baseline level. Each segment concluded with counts capturing those shifts.

Filtered PCG signals applied a fourth-order Butterworth method, range set between 20 and 1000 Hz. From these, thirteen Mel-Frequency Cepstral Coefficients emerged, each contributing mean alongside standard deviation - twenty-six features total. Valve position details followed, transformed via one-hot logic before joining the existing vector.

4.3 Data Splitting Strategy

The ECG and PCG data got separated into training and test portions following an 80–20 division pattern. Splitting of the PCG records involved a method that kept label proportions steady across groups. With each run, consistency came through locking the randomness source at value forty-two.

4.4 Model Configuration

A Gradient Boosting classifier, set with 200 estimators, a learning rate at 0.05, and trees limited to depth three, handled murmur detection in PCG data. Before any model fitting occurred, feature values were rescaled via StandardScaler. Despite fixed structure, variation emerged through parameter interplay across iterations.

Thirty-five decision trees, each allowed to grow up to depth thirty-two, formed the core structure applied here. Balancing class distribution shaped the weighting scheme during training phases. Prediction targets included rhythm patterns alongside waveform morphology without sequential separation. Normal versus abnormal labeling occurred at the same time as broader diagnostic groupings. Multiple outputs emerged together due to the framework's design choice. Equal emphasis went to superclass assignments and more specific subclass distinctions.

4.5 Evaluation Metrics

When assessing PCG murmur detection, evaluation relied on Accuracy alongside ROC-AUC. One measure considered was Precision–Recall AUC, while MCC also contributed to the analysis. Performance metrics were selected carefully - each adding distinct insight. The approach avoided redundancy by including complementary indicators across different aspects of classification behavior. In evaluating ECG multi-label tasks, calculations included Micro F1-score along with Macro F1-score, Hamming Accuracy, and Subset Accuracy. To examine how class imbalance affects results, a comparison emerged between micro and macro variants of the F1 metric.

4.6 Implementation Details

All tests implemented via Python, relying on tools such as scikit-learn, librosa, SciPy, WFDB, PyWavelets, and SHAP. Training and validation occurred within an ordinary desktop setup. A Flask-powered web platform, which enabled analysis tied to patient identifiers.

5 Results and Discussion

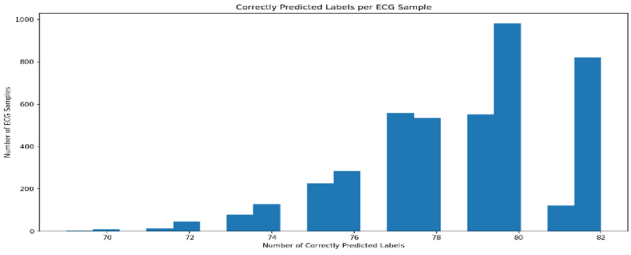

Fig. 2. Distribution of correctly predicted labels per ECG sample

TABLE 1. Performance metrics of the ECG abnormality classification model

Metric	Value
Hamming Accuracy	0.9597
Micro F1-score	0.5885
Macro F1-score	0.0603

In ECG multi-label tasks, the Random Forest delivered a Hamming Accuracy of 0.9595, alongside a Micro F1-score of 0.5865, with Subset Accuracy at 0.1883. Notable differences between micro-level and macro-level scores suggest imbalance among diagnostic labels - typical within PTB-XL datasets. Per-sample counts of accurate predictions are shown through the histogram presented in Figure 2. The performance metrics of the ECG abnormality classification model are summarized in Table 1. Steady output emerges for prevalent classes when examining multi-output labeling behavior.

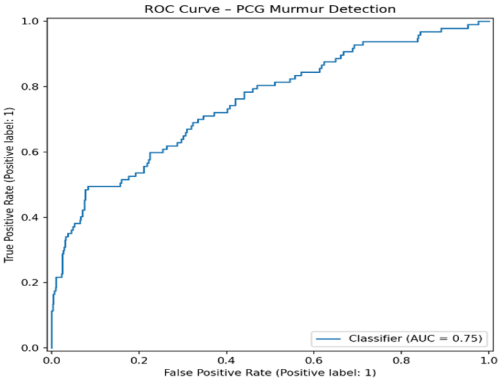

Fig. 3. ROC curve of the Gradient Boosting model for PCG murmur detection

TABLE 2 Class-wise performance of the PCG murmur detection model

Class	Precision	Recall	F1-score
Murmur Absent	0.8648	0.9798	0.9187
Murmer Present	0.6774	0.2165	0.3281
Overall Accuracy: 0.8550			

At 85.50% accuracy, the PCG-based murmur identifier performed moderately well; its ROC–AUC reached 0.7530 as shown in Figure 3, alongside an MCC of 0.3263. Though absent murmurs were detected with high consistency - evidenced by an F1-score of 0.9187 - the system struggled when identifying present ones. This gap arises partly due to uneven distribution in the CirCor data, where normal instances outnumber abnormal. The class-wise performance of the PCG murmur detection model is summarized in Table 2.

Compared with fusion-based deep learning approaches, the suggested system processes ECG and PCG data separately, allowing adaptation if just a single signal type exists. Although accuracy ranks similarly, emphasis shifts toward structure - modular layout stands out, clarity via SHAP follows, inclusion of clinical markers appears next, then report delivery online completes the sequence.

6 Conclusion

This study presented a modular multi-modal cardiac screening framework that independently processes PCG and ECG signals using interpretable ensemble learning techniques. The PCG module achieved an accuracy of 85.50% with a ROC–AUC of 0.7530 for murmur detection, while the ECG module demonstrated strong label-wise agreement with a Hamming Accuracy of 0.9595 on the PTB-XL dataset.

Unlike conventional fusion-based architectures, the proposed system maintains signal-specific pipelines, enabling operation when only a single modality is available. The integration of SHAP-based interpretability, biomarker computation, hierarchical diagnostic outputs, and web-based visualization enhances transparency and clinical relevance.

The results demonstrate that competitive predictive performance can be achieved while preserving modularity and explainability in automated cardiac screening systems.

7 Future work

Future work will focus on improving minority-class sensitivity through advanced imbalance handling and threshold optimization strategies. Cross-dataset validation and clinical expert evaluation will be conducted to further assess generalization. Additionally, integrating lightweight deployment for portable screening environments and exploring hybrid interpretable-deep learning approaches may enhance robustness while maintaining transparency.

REFERNCES

[1] P. Wagner et al., "PTB-XL, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, no. 1, 2020.

[2] B. Oliveira et al., "The CirCor DigiScope phonocardiogram dataset," *PhysioNet*, 2022.

[3] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[4] B. Oliveira et al., "Explainable multimodal deep learning for heart sounds and ECG classification," in *Proc. IEEE EMBC*, 2024.

[5] N. T. Vinh, B. T. Tung, and N. V. Ha, "Multimodal deep learning for ECG heartbeat classification with SHAP-based interpretability," 2025.

[6] A. Calzoni et al., "Bimodal ECG and PCG cardiovascular disease detection: Exploring the potential and modality contribution," *Journal of Medical Systems*, vol. 49, no. 1, 2025.

[7] E. Kalatehjari et al., "Advanced ensemble learning-based CNN-BiLSTM network for cardiovascular disease classification using ECG and PCG signal," *Biomedical Signal Processing and Control*, vol. 108, 2025.

[8] O. Aiadi, M. Ad, and I. Boublal, "Fusion CNN and ViT with self-attention for multi-label ECG classification," 2025.

[9] Federated multimodal ECG/PCG classification framework, 2025.

[10] Y. Cheng et al., "Multi-label few-shot classification of abnormal ECG signals using metric learning," *Circuits, Systems, and Signal Processing*, vol. 44, no. 10, 2025.

[11] A. E. M. Atwa et al., "Interpretable deep learning models for arrhythmia classification based on ECG signals using PTB-XL dataset," *Diagnostics*, vol. 15, no. 15, 2025.

[12] X. Yuan et al., "Enhancing multi-label ECG classification via task-guided lead correlations in Internet of Medical Things," *IEEE Internet of Things Journal*, 2025.

[13] A. Santamónica et al., "Explainable AI for automatic heart disease diagnosis using ECG features," *Biomedical Signal Processing and Control*, 2026.

[14] S. Basak and U. Bhattacharya, "Deep determination of cardiac condition from phonocardiograms," *Neural Computing and Applications*, vol. 37, no. 31, 2025.

[15] B. Althaph and N. P. Challa, "Explainable attention-based deep learning for classification and interpretation of heart murmurs using phonocardiograms," *Scientific Reports*, vol. 15, 2025.

[16] M. Kalimuthu and C. Hemanth, "A comparative analysis of machine learning and deep learning approaches for phonocardiogram classification using dataset integration," *IEEE Access*, 2025.

[17] A review on PCG valvular disease detection and murmur analysis, 2025.

[18] J. Tiffany and Q. Zhang, "Robust multilabel ECG classification under class imbalance using augmentation and adaptive iterative training," *Biomedical Signal Processing and Control*, 2026.

[19] B. Zhao et al., "DRL-ECG-HF: Deep reinforcement learning for enhanced automated diagnosis of heart failure with imbalanced ECG data," *Biomedical Signal Processing and Control*, vol. 107, 2025.

[20] T. Yang et al., "Lead analysis for the classification of multi-label cardiovascular diseases and neural network architecture design," *Electronics*, vol. 14, no. 16, 2025.