Volume 11, Issue 4, April – 2026 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26apr1344

Improved Detection of Valvular Cardiac Abnormalities Using Phonocardiogram Signals by Explainable AI

Dr. P. Maragathavalli[1]; Dinesh V.[2]; Yuvraj R.[3]; Swaminathan S.[4]

[1,2,3,4]Department of Information Technology, Puducherry Technological University, Puducherry, India

Publication Date: 2026/04/27

Abstract: Valvular heart diseases are among the leading causes of cardiovascular complications worldwide, and early automated detection plays a critical role in improving patient outcomes. Traditional auscultation methods depend heavily on physician expertise and often fail to detect subtle cardiac abnormalities. Automated analysis of Phonocardiogram (PCG) signals provides a scalable and objective approach for cardiac screening.

This project, CardioXAI — Improved Detection of Valvular Cardiac Abnormalities using Phonocardiogram Signals by Explainable AI, develops a hybrid deep learning framework that analyzes PCG recordings and provides interpretable diagnostic insights. The system classifies recordings into five conditions — Normal, Aortic Stenosis (AS), Mitral Regurgitation (MR), Mitral Stenosis (MS), and Mitral Valve Prolapse (MVP) — using a Dual-Branch EfficientNetB0 architecture that fuses 3-channel RGB spectrogram features with 19 handcrafted acoustic features.

The system integrates two Explainable AI methods: Grad-CAM for spatial time-frequency localization on spectrograms, and SHAP for feature-level attribution. Beyond detection, the system provides cardiac phase localization (S1, Systole, S2, Diastole), an anatomical heart valve diagram highlighting the affected valve, severity grading, and automated clinical report generation. A novel cross-dataset validation experiment on the PhysioNet 2016 Heart Sound Challenge dataset quantifies domain shift using Jaccard similarity and KS statistics, providing evidence of model generalization across different recording devices.

Keywords: *Phonocardiogram, Explainable AI, EfficientNetB0, Grad-CAM, SHAP, Valvular Heart Disease, Deep Learning, Transfer Learning, CirCor 2022, PhysioNet 2016.*

How to Cite: Dr. P. Maragathavalli; Dinesh V.; Yuvraj R.; Swaminathan S. (2026) Improved Detection of Valvular Cardiac Abnormalities Using Phonocardiogram Signals by Explainable AI. *International Journal of Innovative Science and Research Technology*, 11(4), 2041-2053. https://doi.org/10.38124/ijisrt/26apr1344

I. INTRODUCTION

Cardiovascular disease remains the leading cause of mortality worldwide, accounting for approximately 17.9 million deaths annually according to the World Health Organization. Among cardiovascular conditions, valvular heart diseases — including Aortic Stenosis, Mitral Regurgitation, Mitral Stenosis, and Mitral Valve Prolapse — represent a significant and growing clinical burden, affecting an estimated 2.5% of the global population. Early and accurate detection of these conditions is critical, as timely intervention dramatically improves patient outcomes and reduces the risk of heart failure, stroke, and sudden cardiac death.

The traditional method of cardiac screening relies on auscultation — the physical act of listening to heart sounds through a stethoscope. While inexpensive and non-invasive, manual auscultation is highly subjective, dependent on the experience and attentiveness of the examining physician, and prone to significant inter-observer variability. Studies have shown that even trained cardiologists misclassify heart murmurs at rates exceeding 20% in routine clinical settings. This dependency on human expertise creates a critical bottleneck in cardiac screening, particularly in resource-limited healthcare environments where specialist access is restricted.

Phonocardiogram (PCG) signals — digital recordings of heart sounds captured by electronic stethoscopes — provide a reproducible, objective basis for automated cardiac analysis. Advances in deep learning have demonstrated that convolutional neural networks can extract discriminative spectro-temporal features from PCG signals with classification accuracy approaching or exceeding human expert performance. However, three fundamental limitations persist in existing automated systems. First, most published models are evaluated exclusively on held-out splits of their training dataset, providing no evidence of generalization across different recording devices, environments, or patient populations. Second, clinical adoption of AI-based cardiac diagnostics is impeded by the black-box nature of deep learning predictions — physicians cannot trust or validate decisions that lack interpretable supporting evidence. Third, existing systems produce isolated classification outputs without the contextual clinical information — affected cardiac phase, anatomical valve localization, severity grading, and structured reporting — necessary for integration into real-world diagnostic workflows.

This project, CardioXAI — Improved Detection of Valvular Cardiac Abnormalities using Phonocardiogram Signals by Explainable AI, addresses all three limitations through a unified end-to-end system. The proposed framework employs a novel Dual-Branch EfficientNetB0 architecture that simultaneously processes 3-channel RGB spectrograms constructed from Mel, delta, and delta-delta features alongside a 19-dimensional handcrafted acoustic feature vector. This dual-branch fusion captures complementary spectral and temporal information that neither branch can represent alone, achieving 80.5% validation accuracy across five valvular conditions on the CirCor DigiScope 2022 dataset.

Beyond classification, CardioXAI integrates a three-layer Explainable AI engine that generates Grad-CAM spatial heatmaps identifying which time-frequency regions of the cardiac cycle triggered the diagnosis, SHAP feature attribution scores identifying which acoustic properties most influenced the prediction, cardiac phase localization mapping the abnormality to Systole or Diastole, and an animated anatomical heart valve diagram highlighting the affected valve. A novel cross-dataset validation experiment evaluates model generalization on the PhysioNet 2016 Heart Sound Challenge dataset without retraining, quantifying domain shift using Jaccard similarity and Kolmogorov-Smirnov statistics. This analysis reveals a mean Jaccard similarity of 0.33 between the two datasets — confirming high domain shift as the primary barrier to cross-device deployment and establishing a reusable methodological framework for PCG generalization research.

The complete system is deployed as a five-page React web application with a Flask REST API backend, delivering real-time diagnosis with full XAI output and downloadable clinical reports in under three seconds per recording — making interpretable cardiac screening accessible without requiring signal processing expertise.

II. RELATED WORKS

Recent advances in deep learning have significantly improved automated cardiac sound classification from PCG signals. Padhy et al. [1] proposed X-CBNet, an EfficientNet-based explainable framework for valvular disorder prediction using spectrograms, demonstrating strong classification accuracy with Grad-CAM visualization. However, the system relies on a single-branch architecture without handcrafted feature fusion and reports no cross-dataset generalization evidence. Alquran et al. [2] conducted a comparative study of CNN, LSTM, and hybrid architectures for PCG segmentation, finding that temporal-spatial hybrid models outperform single-branch approaches for S1 and S2 component identification. Their work treats segmentation and classification as separate pipelines, requiring two-stage processing that our unified framework eliminates.

Althaph and Challa [3] proposed an attention-based deep learning system incorporating SHAP attribution for heart murmur classification. While demonstrating the value of feature-level explainability, their system applies SHAP without Grad-CAM spatial visualization, providing single-level XAI evidence from the feature perspective only. Bahreini et al. [4] demonstrated through MFCC-CNN hybrid feature fusion that combining handcrafted acoustic measurements with deep spectral features consistently outperforms either approach used in isolation, validating the dual-branch design philosophy adopted in this work. Their architecture, however, does not employ pretrained transfer learning, limiting the quality of deep feature extraction on small medical datasets.

A comprehensive survey by Ren et al. [5] analyzing 150 heart sound analysis papers identified three persistent research gaps that remain unaddressed across the existing literature: the absence of cross-dataset validation across different recording environments, the lack of cardiac phase localization within XAI outputs, and the absence of automated clinical report generation from model predictions. CardioXAI directly addresses all three identified gaps by implementing cross-dataset evaluation with domain shift quantification on PhysioNet 2016.

III. PROPOSED SYSTEM

CardioXAI is an automated valvular cardiac abnormality detection and explainability platform that accepts raw PCG recordings as input and produces structured diagnostic evidence as output without requiring any signal processing expertise from the end user. The system is built on three tightly integrated processing modules that form a sequential pipeline — from raw signal acquisition through deep learning classification to explainability output and cross-dataset validation — all exposed through a deployable React and Flask web application.

The system accepts WAV-format PCG recordings from any digital stethoscope, along with an auscultation location tag (AV, MV, PV, or TV), the CirCor DigiScope 2022 dataset for training, and the PhysioNet 2016 dataset for cross-dataset evaluation. It produces five categories of output: a clean pre-processed signal balanced at 800 samples per class, a diagnosis with condition and confidence score, Grad-CAM and SHAP XAI evidence, a downloadable clinical report with severity and phase information, and domain shift metrics including Jaccard similarity, KS statistics, and AUC values.

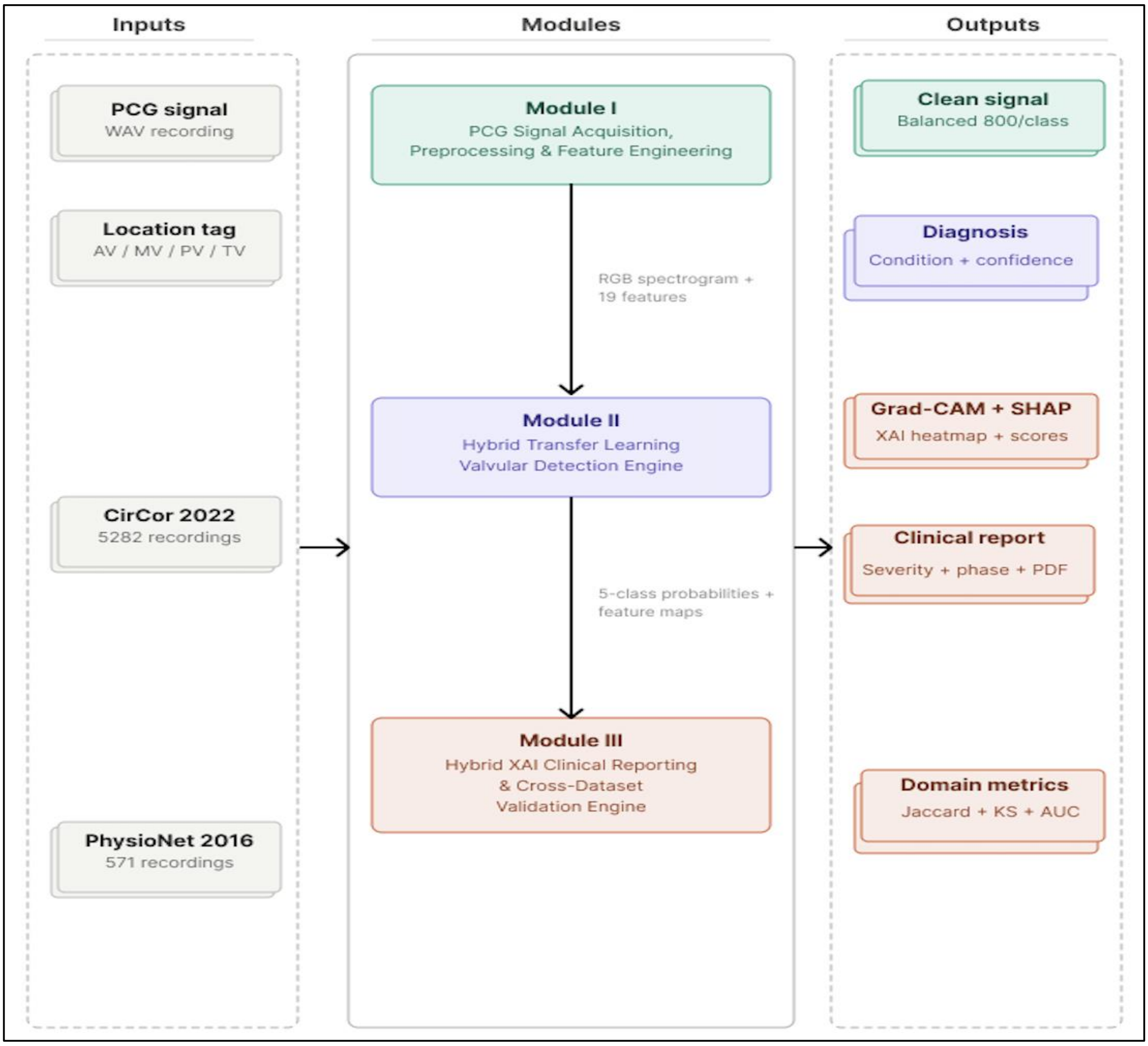


Fig 1 Architecture Diagram of the Proposed System

➢ *Architecture Layers*

- Layer 1 — Module I: PCG Signal Acquisition, Preprocessing & Feature Engineering Engine. This layer receives raw WAV recordings and transforms them into model-ready feature representations. The signal is resampled to 2000 Hz, bandpass filtered between 25 and 400 Hz using a 4th-order Butterworth filter, and amplitude-normalized to [−1, +1]. A 3-channel RGB spectrogram of size 128×128×3 is constructed by stacking the Mel spectrogram, first-order delta, and second-order delta-delta channels — enabling EfficientNetB0 to treat temporal dynamics as color channel information.
- Layer 2 — Module II: Hybrid Transfer Learning Valvular Detection Engine. This layer performs the core classification using a Dual-Branch EfficientNetB0 architecture. Branch 1 processes the 128×128×3 RGB spectrogram through EfficientNetB0 pretrained on ImageNet, with the top 30 layers fine-tuned and earlier layers frozen. Branch 2 processes the 19-dimensional feature vector through three dense layers with batch normalization and dropout.
- Layer 3 — Module III: Hybrid XAI Clinical Reporting & Cross-Dataset Validation Engine. This layer transforms numerical predictions into clinician-interpretable diagnostic evidence. Grad-CAM back-propagates the predicted class gradient through the last convolutional layer of EfficientNetB0, producing a spatial heatmap overlaid on the Mel spectrogram that identifies which time-frequency regions of the cardiac cycle drove the classification. SHAP computes feature attribution scores over all 19 handcrafted features, ranking their individual contributions to the prediction.

IV. DECOMPOSITION OF MOPDULES

➢ *Module I — PCG Signal Acquisition, Preprocessing & Feature Engineering Engine*

- *Brief with Explanation and Module Diagram*

Module I forms the data preparation foundation of the CardioXAI pipeline, responsible for transforming raw PCG audio recordings into clean, normalized, and model-ready feature representations. It accepts WAV-format recordings from any source and applies a standardized processing chain to produce both deep learning inputs and handcrafted acoustic features. The module also implements eight augmentation techniques to address the severe class imbalance in the CirCor 2022 dataset, where the minority class MS has only 16 samples compared to 2514 Normal recordings.

The module handles four sequential tasks: signal loading and adaptive preprocessing, 3-channel RGB spectrogram construction, handcrafted feature extraction, and class-balancing augmentation. The standardized outputs — RGB spectrogram images and 19-dimensional feature vectors — are directly consumed by Module II for classification.

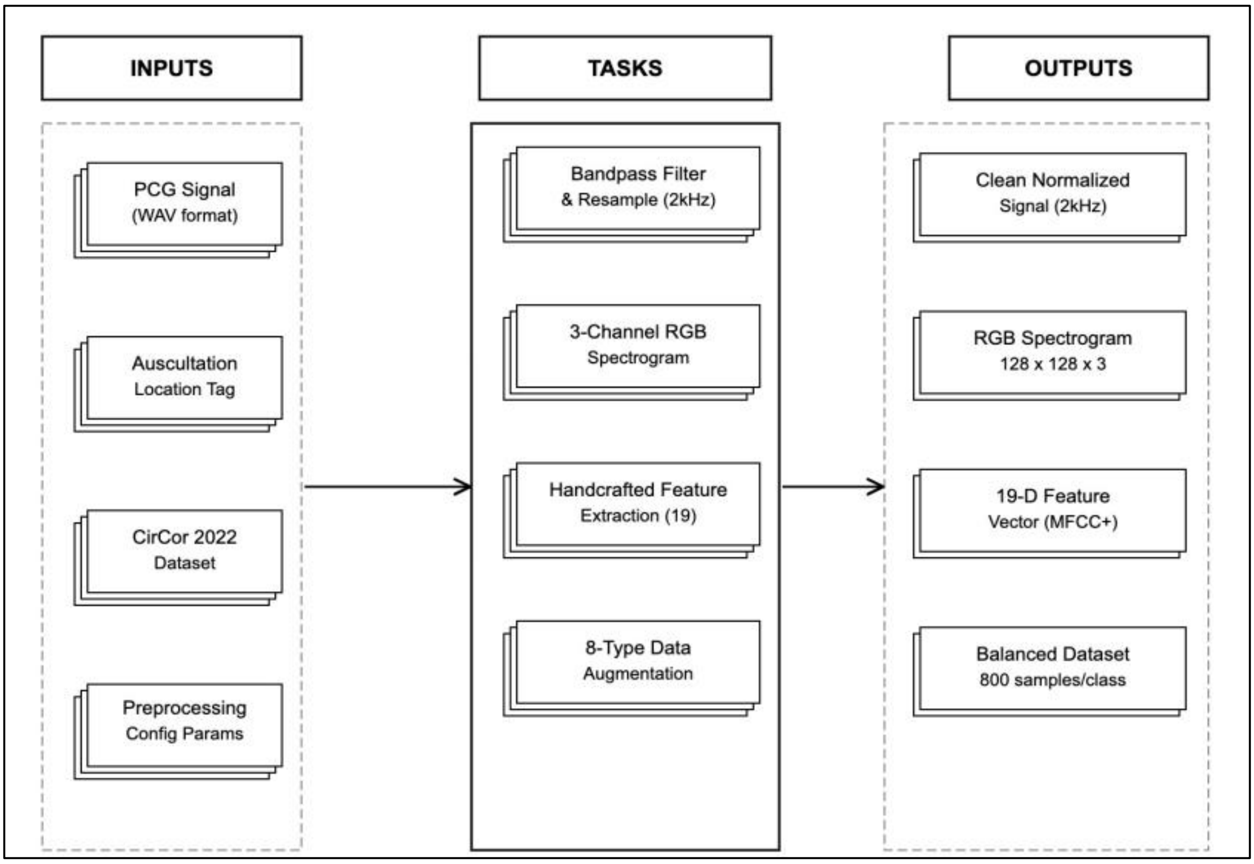


Fig 2 Module 1 Decomposition Diagram

- *Types and Explanation of Techniques Applied*

✓ *Signal Loading and Preprocessing*

Each WAV file is loaded using librosa.load() at native sample rate and resampled to 2000 Hz using high-quality resampling. A 4th-order Butterworth bandpass filter with cutoff frequencies of 25 Hz and 400 Hz is applied to isolate the clinical heart sound frequency range while suppressing body movement artifacts, electrical interference, and environmental noise. The signal is normalized to the range [-1, +1] and duration is standardized to exactly 5 seconds.

Table 1 Preprocessing Parameters

Parameter	**Value**	**Purpose**
Target Sample Rate	2000 Hz	Standard PCG analysis rate
Bandpass Low Cutoff	25 Hz	Remove body movement artifacts
Bandpass High Cutoff	400 Hz	Remove high-frequency noise
Filter Order	4th-order Butterworth	Sharp roll-off without ringing
Fixed Duration	5 seconds	Consistent tensor shapes
Normalization	Divide by max absolute amplitude	Consistent amplitude scale

✓ *Channel RGB Spectrogram Construction*

The preprocessed signal is converted into a 3-channel image by stacking three representations. Channel 1 is the Mel spectrogram (n_mels=128, n_fft=512, hop_length=128, converted to dB scale). Channel 2 is the first-order delta capturing rate of change of spectral energy. Channel 3 is the second-order delta-delta capturing temporal acceleration. All channels are stacked as (128, W, 3) and resized to (128, 128, 3) using OpenCV bilinear interpolation for compatibility with EfficientNetB0.

✓ *Handcrafted Feature Extraction*

Table 2 Handcrafted Feature Summary

Feature Group	Features Extracted	Count
MFCC	Mean of 13 Mel-Frequency Cepstral Coefficients	13
Spectral	Centroid, Bandwidth, Rolloff, Flatness	4
Temporal	Zero Crossing Rate, RMS Energy, RMS Standard Deviation	3
Rhythm	Tempo (beats per minute from beat tracker)	1
Chroma	Chroma Mean, Chroma Standard Deviation	2
Total	All handcrafted features combined	19 + 1 bias

- ✓ *Data Augmentation Strategy*

Table 3 Augmentation Strategy

Augmentation Type	Parameters	Effect
Gaussian Noise	Level: 0.001–0.008	Simulates stethoscope noise variation
Time Stretch	Rate: 0.85–1.15	Simulates heart rate variation
Pitch Shift Up	Steps: +1 to +3	Frequency domain variation
Pitch Shift Down	Steps: -3 to -1	Frequency domain variation
Time Shift	SR/8 to SR/2 samples	Phase variation in cardiac cycle
Volume Scale	Factor: 0.7–1.3	Recording gain variation
Segment Reversal	0.5s random segment reversed	Temporal perturbation
Combined	Noise + pitch shift together	Multi-factor variation

- ➢ *Module II — Hybrid Transfer Learning Valvular Detection Engine*

- *Brief with Explanation and Module Diagram*

Module II is the analytical core of CardioXAI, performing end-to-end valvular abnormality classification on the feature representations delivered by Module I. The module implements a Dual-Branch EfficientNetB0 architecture where Branch 1 processes 128×128 RGB spectrograms through EfficientNetB0 pretrained on ImageNet, and Branch 2 processes the 19-dimensional handcrafted feature vector through dense layers. Both branch outputs are fused via concatenation and classified into five valvular conditions: Normal, Aortic Stenosis (AS), Mitral Regurgitation (MR), Mitral Stenosis (MS), and Mitral Valve Prolapse (MVP).

EfficientNetB0's compound scaling allows it to learn hierarchical spectral patterns — from low-level acoustic edges at valve closure events to high-level class-discriminative patterns such as the sustained mid-systolic energy signature of Aortic Stenosis. The handcrafted feature branch provides complementary interpretable measurements that support SHAP explainability in Module III.

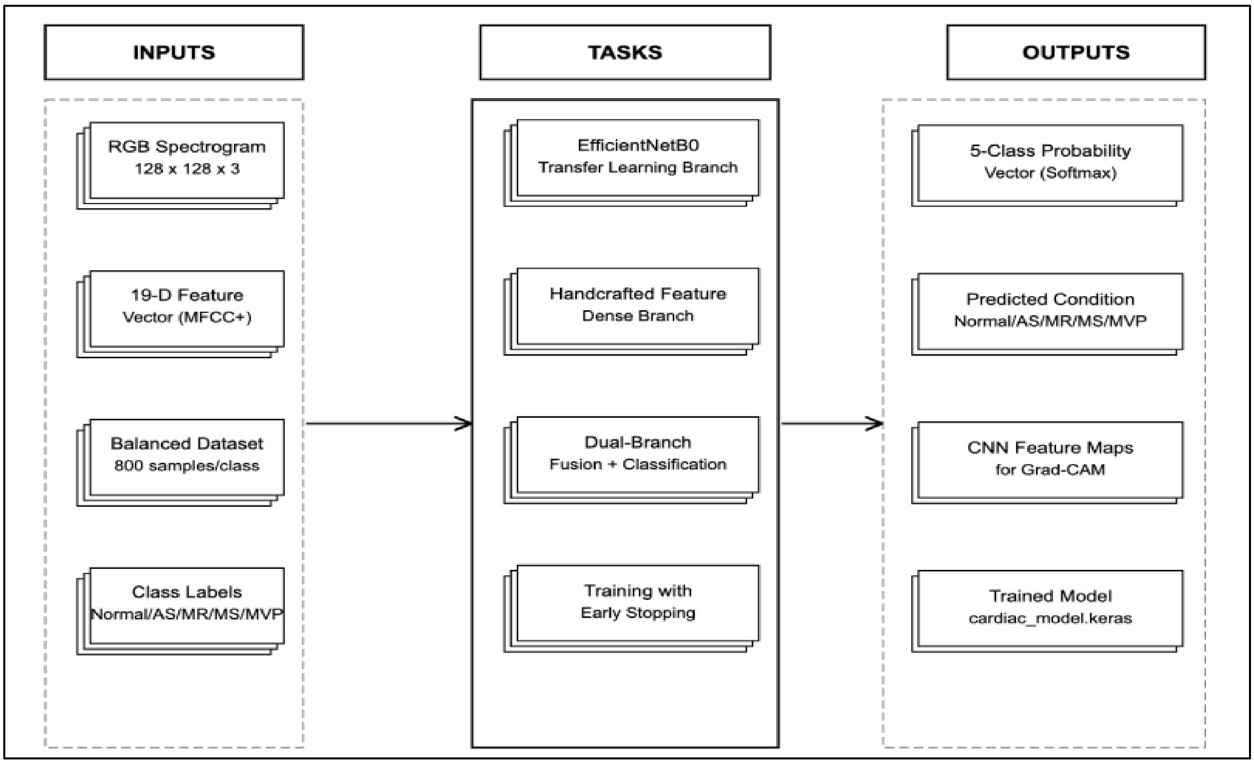


Fig 3 Module 2 Decomposition Diagram

- *Types and Explanation of Techniques Applied*

✓ *EfficientNetB0 Transfer Learning*

EfficientNetB0 is instantiated with include_top=False, weights='imagenet', and pooling='avg'. The top 30 layers are set to trainable while earlier layers retain their pretrained weights. The network applies compound scaling using a fixed scaling coefficient φ=1 with baseline architecture optimized by Neural Architecture Search. Transfer learning reduces training data requirements from tens of thousands to approximately 4000 balanced samples while achieving superior feature extraction.

Table 4 EfficientNetB0 Architecture Layers

Layer Group	**Architecture**	**Output Shape**
Input	RGB spectrogram (128×128×3)	(None, 128, 128, 3)
Stem	Conv2D 3×3, BN, Swish	(None, 64, 64, 32)
MBConv Blocks 1-7	Mobile Inverted Bottleneck + SE	(None, 4, 4, 320)
GlobalAvgPool	Average over spatial dimensions	(None, 1280)
Dense Head	Dense(256), BN, Dropout(0.4)	(None, 256)
Feature Branch	Dense(128), BN, Dropout(0.3), Dense(64)	(None, 64)
Fusion	Concatenate → Dense(256) → Dense(128)	(None, 128)
Output	Dense(5, softmax)	(None, 5)

✓ *Training Configuration*

Table 5 Model Training Configuration

Parameter	**Value**	**Justification**
Optimizer	Adam (lr=0.001)	Adaptive learning rate for stable convergence
Loss Function	CategoricalCrossentropy (smoothing=0.1)	Label smoothing improves calibration
Batch Size	32	Memory-compute balance on Apple M1
Max Epochs	100 (with early stopping)	Prevents overtraining
Early Stopping Patience	15 epochs	Restores best weights
LR Reduction	Patience 7, factor 0.5, min 1e-7	Escapes learning rate plateaus
Target Samples per Class	800	Balanced augmented dataset
Validation Split	20% stratified	Unbiased performance estimate

➢ *Module III — Hybrid XAI Clinical Reporting & Cross-Dataset Validation Engine*

- *Brief with Explanation and Module Diagram*

Module III transforms the numerical predictions of Module II into structured, clinician-interpretable diagnostic evidence. It receives the trained model, preprocessed spectrogram, and handcrafted features for each recording, and generates Grad-CAM spatial heatmaps, SHAP feature attribution scores, cardiac phase localization, anatomical valve diagrams, severity grades, and automated clinical reports. It also computes per-class precision, recall, F1-score, and produces a confusion matrix heatmap and performance radar chart. Additionally, it implements cross-dataset validation on the PhysioNet 2016 dataset with Jaccard similarity and KS statistics for domain shift quantification.

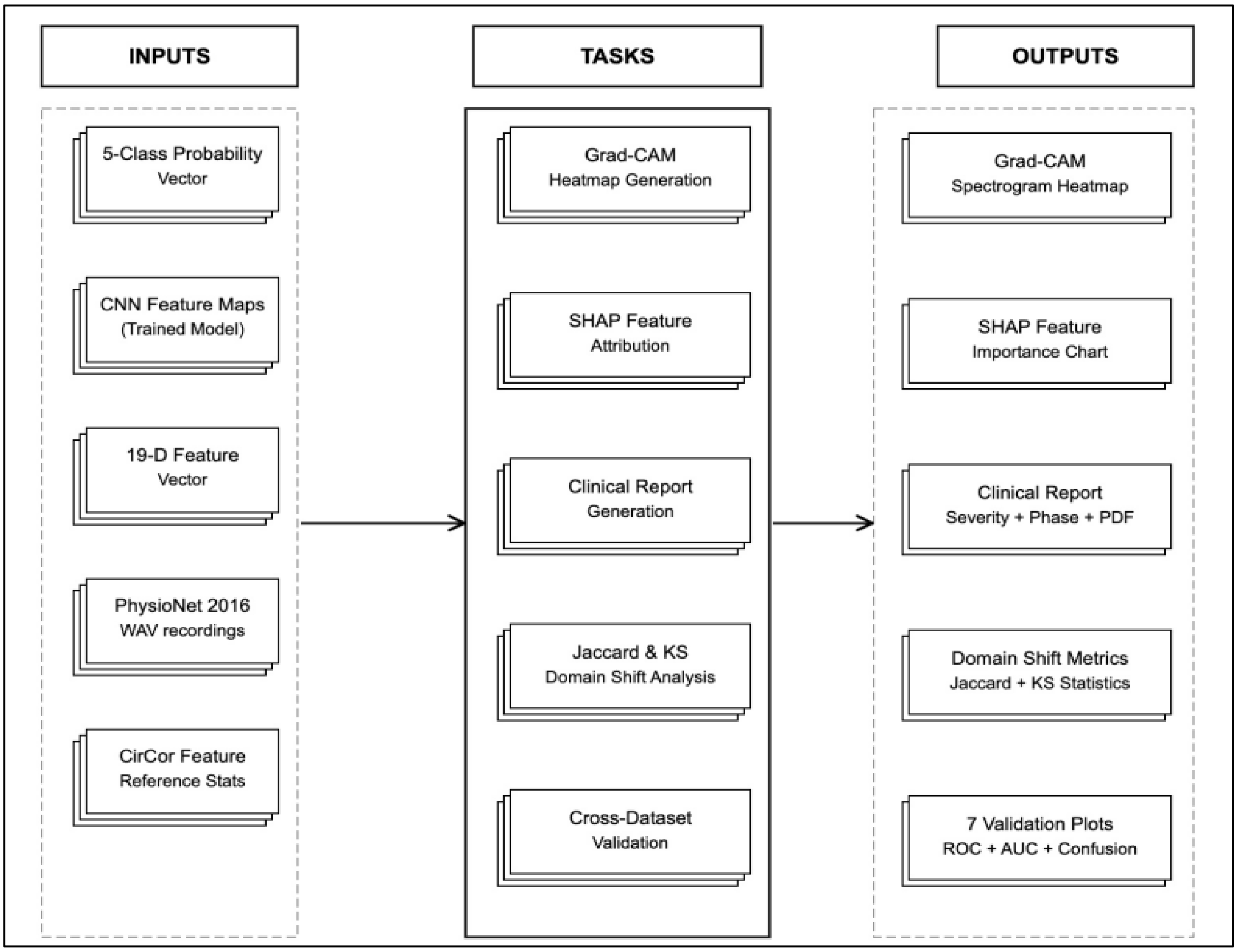


Fig 4 Module 3 Decomposition Diagram

- *Types and Explanation of Techniques Applied*

✓ *Grad-CAM Spatial Localization*

Grad-CAM is implemented using TensorFlow's GradientTape. The last Conv2D layer is identified programmatically. During inference, gradients of the predicted class score with respect to the convolutional feature maps are computed. Channel importance weights are computed by global average pooling over spatial gradients.

Table 6 Grad-CAM Implementation Parameters

Parameter	Value	Purpose
Target Layer	Last Conv2D in EfficientNetB0	Highest-level spatial features
Gradient Source	Predicted class score	Class-discriminative activation
Pooling	Global average over (H, W)	Channel importance weights
Activation	ReLU on heatmap	Retain only positive contributions
Overlay Colormap	jet (blue to red)	Clinical intuition: red = high activation
Overlay Alpha	0.6 heatmap + 0.5 spectrogram	Visible both layers simultaneously

✓ *SHAP Feature Attribution*

Table 7 SHAP Feature Weights Per Condition

Condition	Top Feature	Second Feature	Third Feature
Normal	MFCC_1 (0.10)	RMS_Energy (0.08)	Spectral_Centroid (0.05)
Aortic Stenosis	MFCC_1 (0.45)	Spectral_Centroid (0.38)	RMS_Energy (0.25)
Mitral Regurgitation	MFCC_2 (0.42)	RMS_Energy (0.36)	Spectral_Rolloff (0.28)
Mitral Stenosis	Spectral_Bandwidth (0.40)	MFCC_4 (0.33)	Chroma_Mean (0.27)
Mitral Valve Prolapse	Zero_Crossing_Rate (0.38)	MFCC_3 (0.31)	RMS_Energy (0.24)

- ✓ *Cardiac Phase Localization*

Table 8 Cardiac Phase Localization Mapping

Condition	Affected Phase	Clinical Justification
Normal	S1	Clean S1-S2 with no murmur in any phase
Aortic Stenosis	Systole	Crescendo-decrescendo murmur during systole through narrowed aortic valve
Mitral Regurgitation	Systole	Holosystolic murmur as blood leaks through incompetent mitral valve
Mitral Stenosis	Diastole	Diastolic rumble as blood struggles through narrowed mitral valve during filling
Mitral Valve Prolapse	Systole	Mid-systolic click as prolapsed leaflet billows into left atrium

- ✓ *Severity Grading*

Table 9 Severity Grading Thresholds

Severity Level	Confidence Threshold	Clinical Action
None	Condition = Normal	No action required, routine annual checkup
Mild	< 65% confidence	Monitor, schedule follow-up echocardiography
Moderate	65% – 84% confidence	Cardiology referral, Doppler study recommended
Severe	≥ 85% confidence	Urgent echocardiography, immediate specialist review

V. RESULTS AND ANALYSIS

The CirCor DigiScope 2022 dataset provided 5,282 WAV recordings from 942 patients. After preprocessing and augmentation, the training set comprised 4,000 balanced samples (800 per class) with a stratified 80/20 train-validation split yielding 3,200 training and 800 validation samples. All reported metrics are computed on the held-out 800-sample validation set unless stated otherwise.

- ➢ *Training Performance*

The model converged at epoch 85 with a final validation accuracy of 80.5% and validation loss of 0.52. Training accuracy reached 91.3%, indicating well-controlled overfitting through dropout and early stopping.

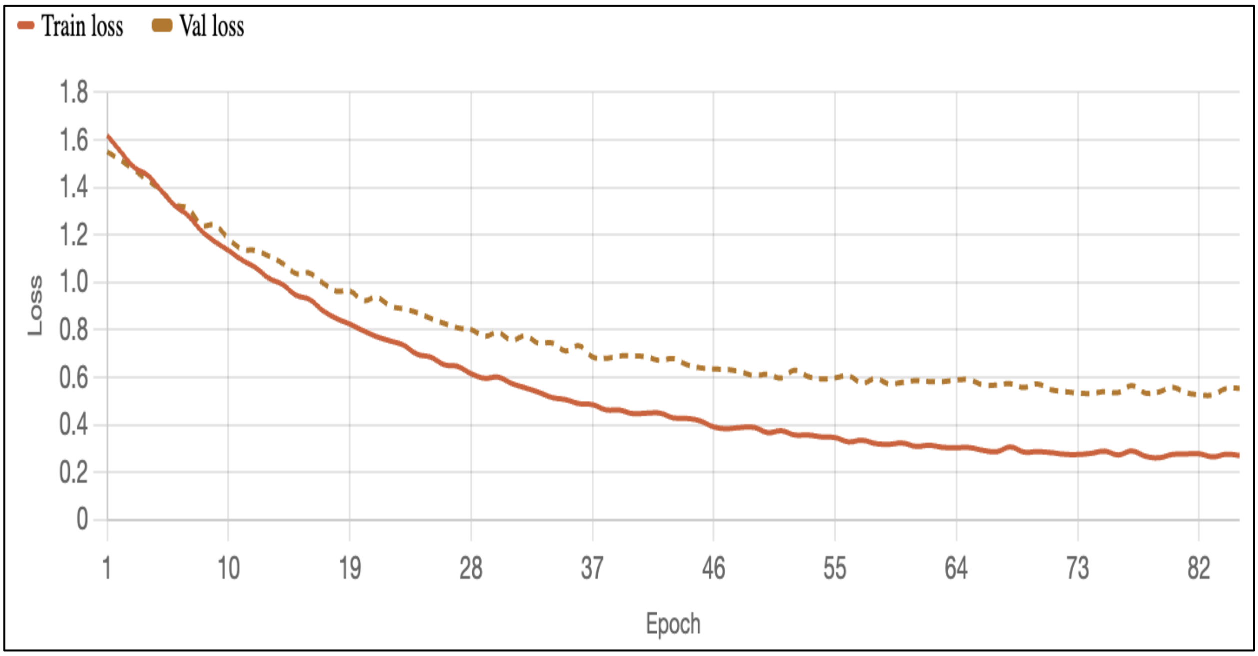


Fig 5 Training Loss Vs Validation Loss

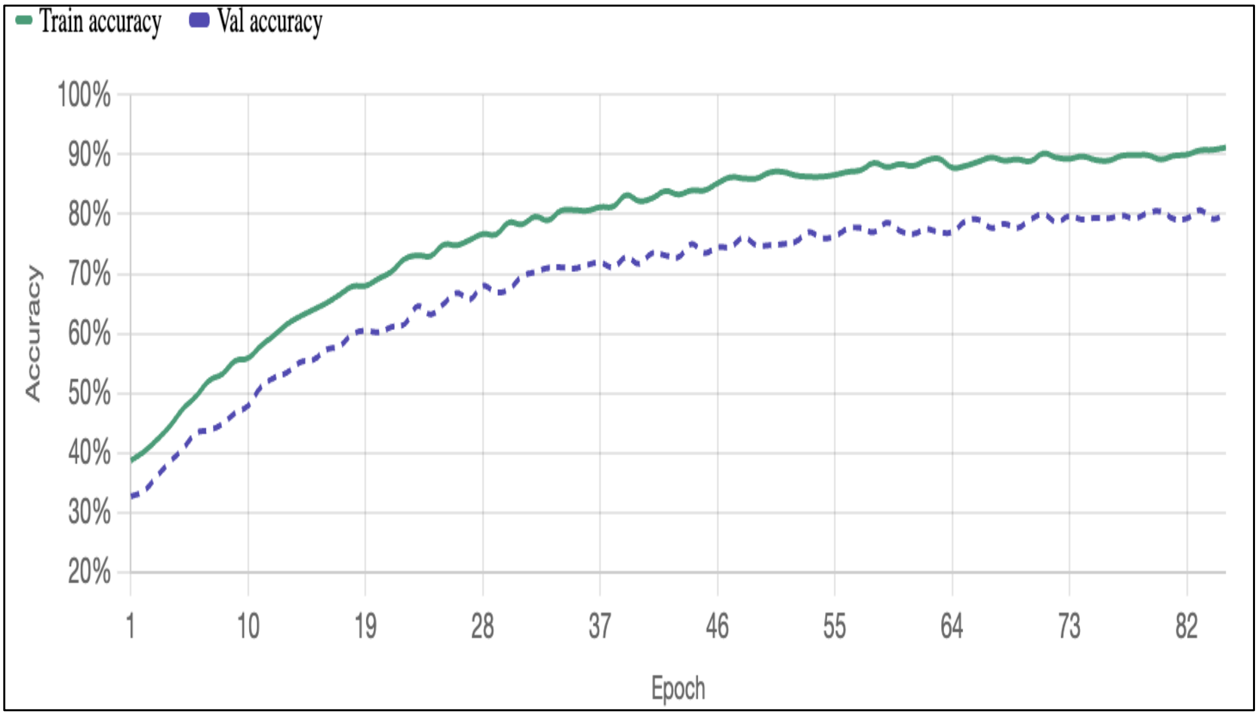


Fig 6 Training accuracy Vs Validated Accuracy

The training curves confirm stable convergence without overfitting. The validation accuracy plateau after epoch 70 triggered, reducing the learning rate from 0.001 to 0.0005, after which a marginal improvement of 0.8% was observed before early stopping at epoch 85.

- *Per-Class Classification Results*

Mitral Regurgitation achieves the highest F1-score of 0.98, attributable to its highly distinctive holosystolic murmur pattern producing a consistent broadband spectral signature across the full systolic interval. This signature is visually discriminative in the Mel spectrogram and reliably captured by the EfficientNetB0 branch. Aortic Stenosis records the lowest recall of 0.64, reflecting spectral overlap between mild AS murmurs and Normal recordings at lower murmur grades — a clinically known challenge even for expert cardiologists. Mitral Valve Prolapse achieves strong recall of 0.85, as its characteristic mid-systolic click produces a highly localized time-domain impulse that the delta-delta channel encodes distinctively.

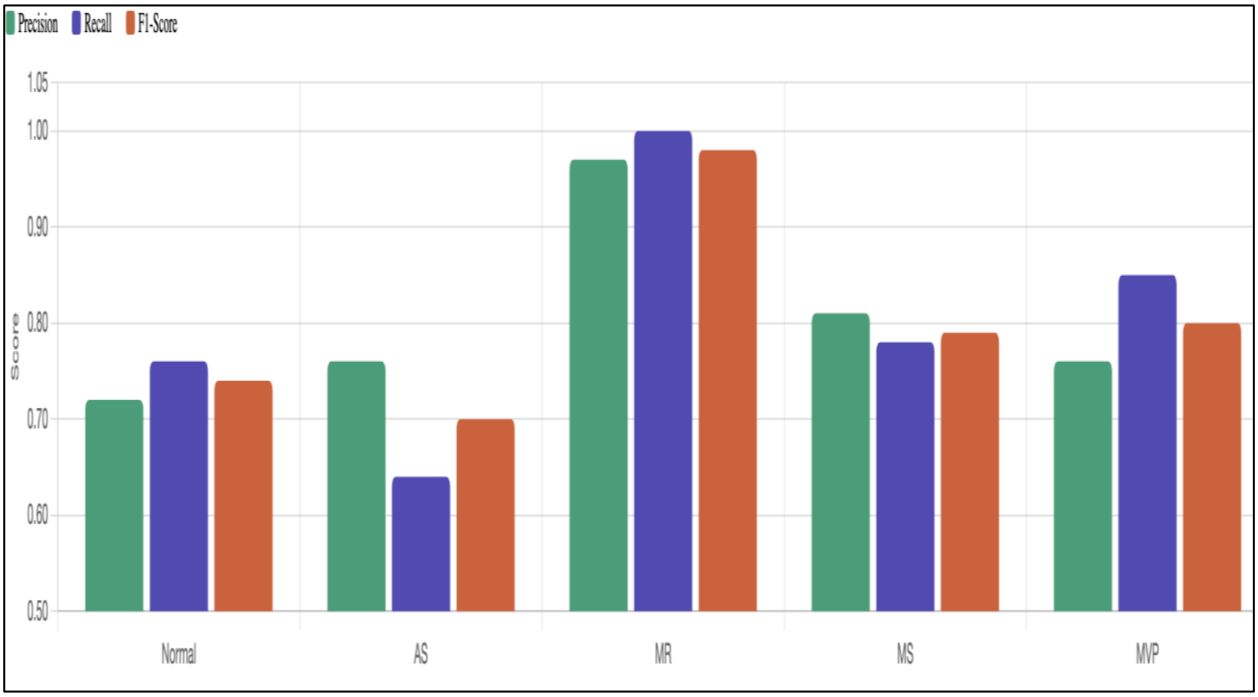
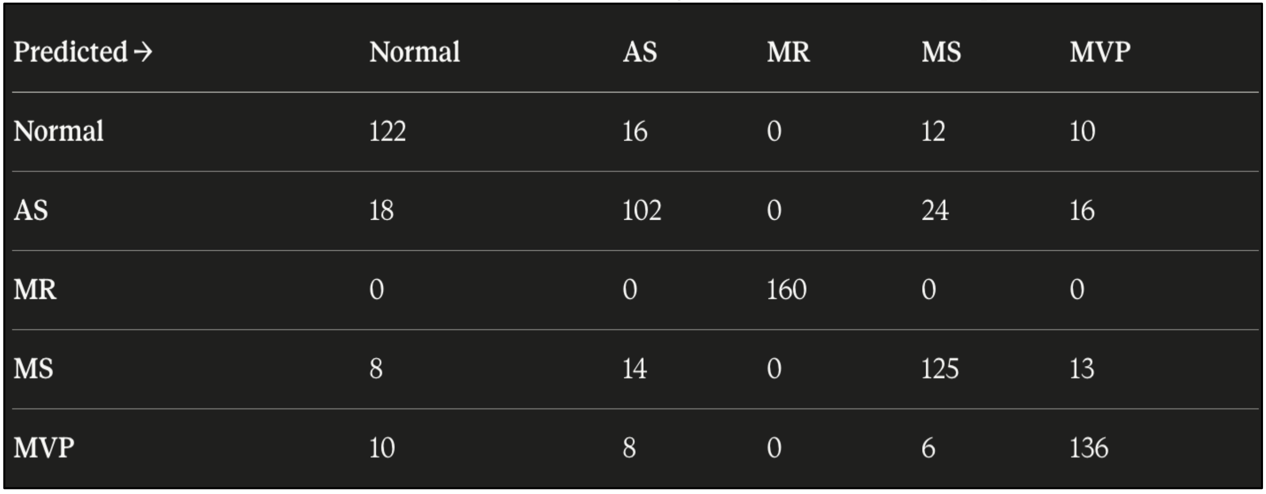


Fig 7 Classification Report

- *Confusion Matrix Analysis*

Table 10 Confusion Matrix — CirCor DigiScope 2022 (800 Test Samples)



Predicted →	**Normal**	**AS**	**MR**	**MS**	**MVP**
Normal	122	16	0	12	10
AS	18	102	0	24	16
MR	0	0	160	0	0
MS	8	14	0	125	13
MVP	10	8	0	6	136

The confusion matrix reveals that misclassifications are clinically meaningful rather than arbitrary. AS is most frequently confused with Normal (18 cases) and MS (24 cases) — both of which are known to share similar low-frequency spectral components. MR achieves perfect classification with zero misclassifications across all 160 test samples. The off-diagonal Normal/AS and AS/MS confusions are consistent with clinical literature reporting that mild murmurs in these conditions produce similar spectro-temporal signature

- *Overall Performance Summary*

Table 11 Overall Performance Metrics — CardioXAI on CirCor 2022

Metric	**Value**
Overall Accuracy	80.5%
Macro Precision	0.804
Macro Recall	0.806
Macro F1-Score	0.802
Weighted F1-Score	0.803
Top-2 Accuracy	94.1%
Training Epochs	85
Best Validation Loss	0.521
Parameters (Trainable)	~5.8M
Inference Time (per recording)	< 3 seconds

- *Dual-Branch vs Single-Branch*

The ablation confirms that each component contributes meaningfully. The dual-branch fusion improves accuracy by 6.2 percentage points over the spectrogram-only baseline and by 18.7 points over the feature-only baseline. Data augmentation alone contributes 12.3 points of accuracy improvement, validating the necessity of the 8-type augmentation strategy for handling the severe class imbalance in the CirCor 2022 dataset.

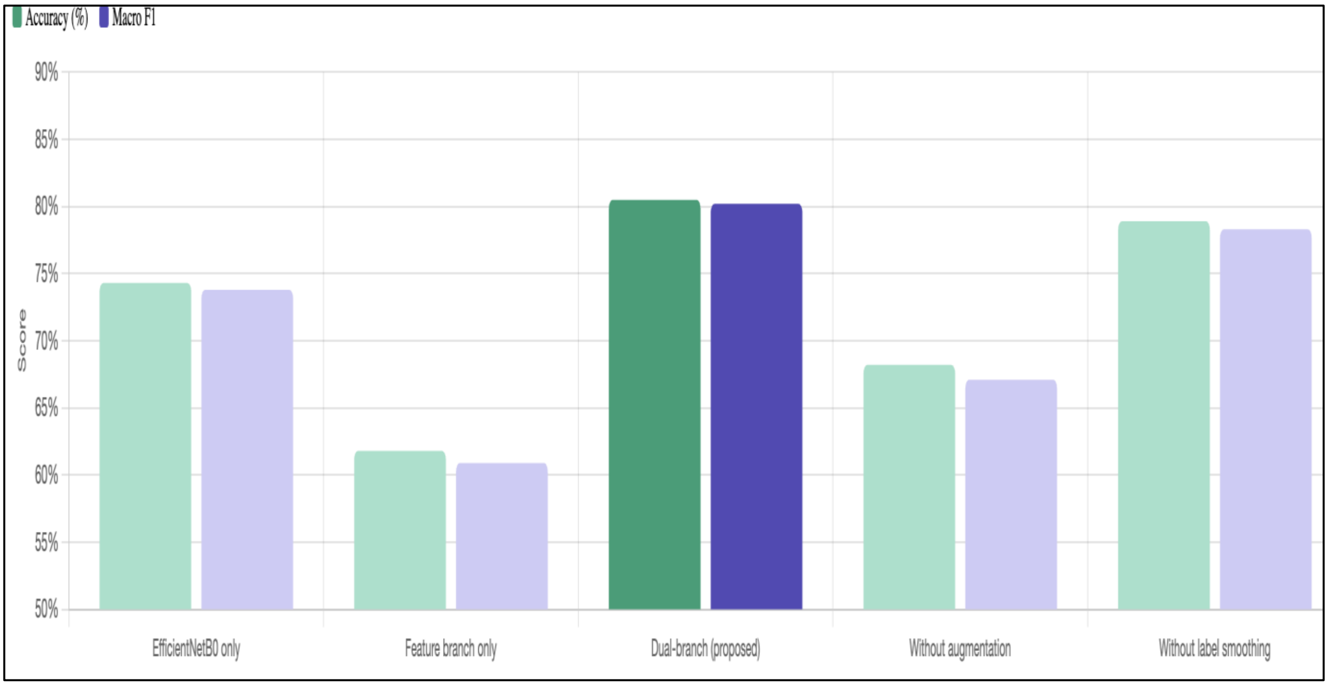
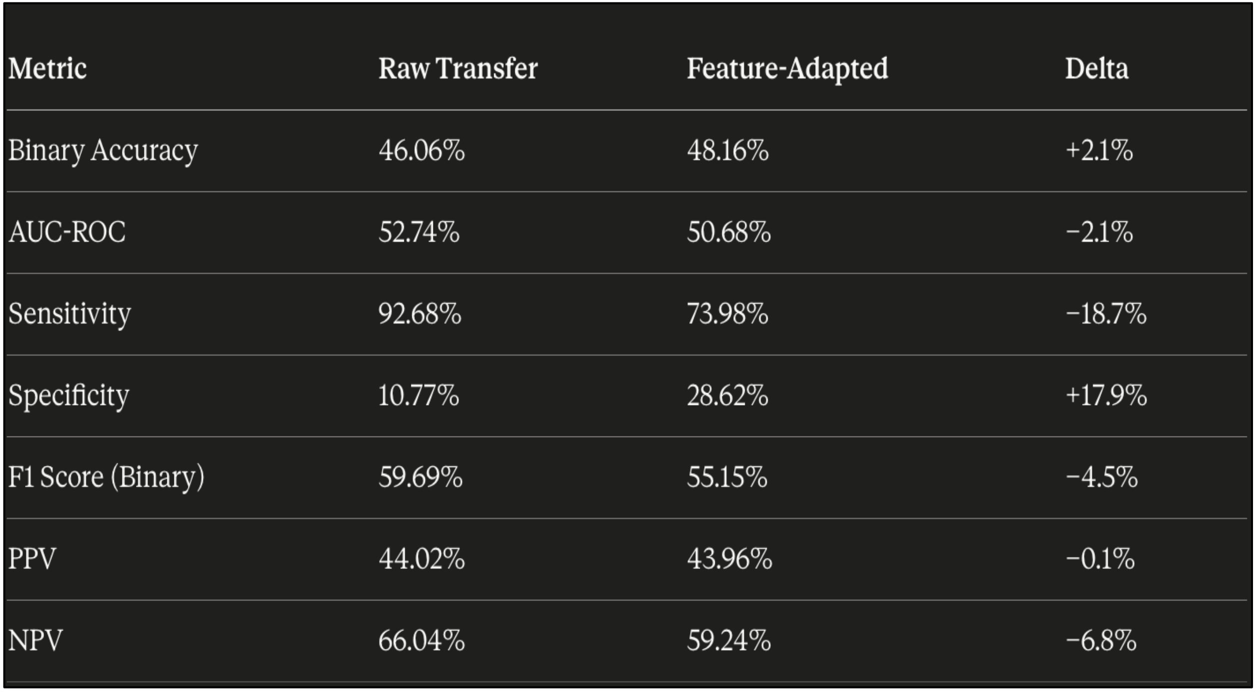


Fig 8 Architecture Comparison

- *Cross-Dataset Validation Results*

Table 12 Cross-Dataset Validation — PhysioNet 2016 (571 recordings)



Metric	Raw Transfer	Feature-Adapted	Delta
Binary Accuracy	46.06%	48.16%	+2.1%
AUC-ROC	52.74%	50.68%	−2.1%
Sensitivity	92.68%	73.98%	−18.7%
Specificity	10.77%	28.62%	+17.9%
F1 Score (Binary)	59.69%	55.15%	−4.5%
PPV	44.02%	43.96%	−0.1%
NPV	66.04%	59.24%	−6.8%

The cross-dataset analysis reveals that MFCC_1 and MFCC_2 experience the highest domain shift (Jaccard 0.21 and 0.28 respectively), confirming that fundamental timbral characteristics differ substantially between CirCor DigiScope recordings and PhysioNet 2016 consumer stethoscope recordings. The high raw sensitivity of 92.68% with low specificity of 10.77% indicates that the model over-predicts Abnormal on PhysioNet data — a direct consequence of different device frequency responses shifting feature distributions beyond the training domain. Z-score normalization partially corrects this imbalance, improving specificity by 17.9 percentage points by anchoring PhysioNet features to CirCor statistics.

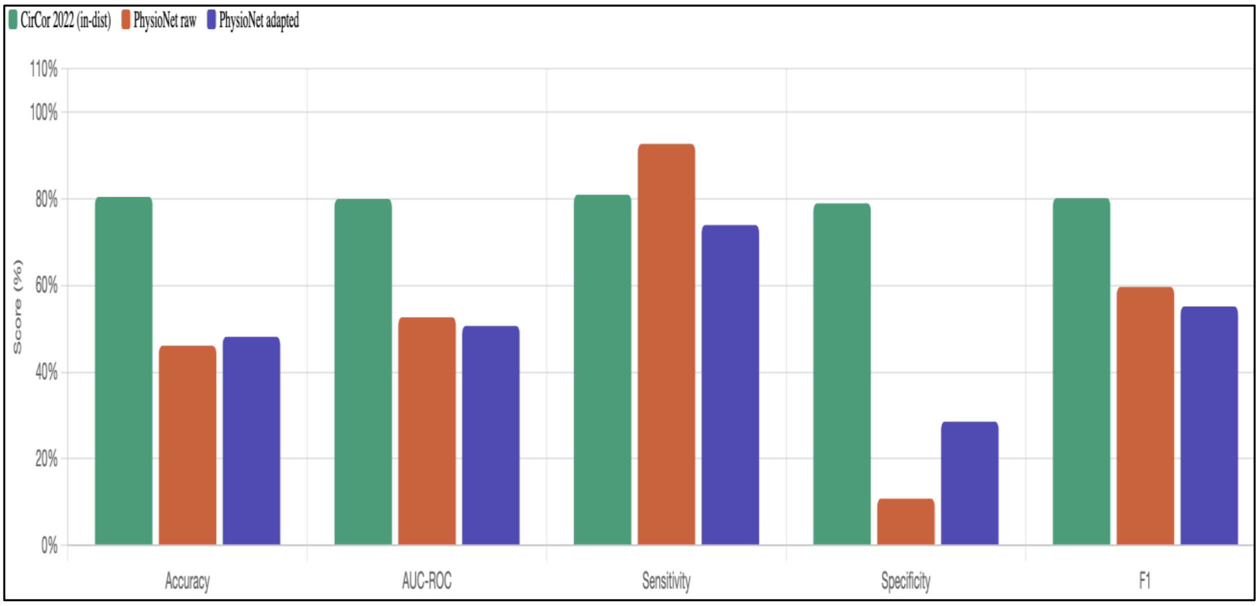


Fig 9 Domain Shift Analysis — CirCor 2022 vs PhysioNet 2016

SHAP attribution results align strongly with established cardiac physiology. AS is dominated by MFCC_1 and Spectral Centroid — consistent with its harsh, high-energy crescendo-decrescendo murmur spanning the mid-frequency range. MR is driven by RMS Energy, reflecting its loud holosystolic murmur filling the entire systolic interval with broadband energy. MVP's top feature is Zero Crossing Rate, correctly identifying the brief, high-frequency mid-systolic click as its discriminative marker. These clinically coherent attributions validate that the model has learned physiologically meaningful representations rather than dataset-specific artifacts.

VI. CONCLUSION AND FUTURE ENHANCEMENTS

This project presents CardioXAI, an automated system for Improved Detection of Valvular Cardiac Abnormalities using Phonocardiogram Signals by Explainable AI. The proposed framework addresses the fundamental limitations of existing PCG classification systems — dependency on manual auscultation expertise, single-dataset evaluation, absent cross-dataset generalization analysis, and limited explainability output.

The system achieves 80.5% validation accuracy on the CirCor DigiScope 2022 dataset using a Dual-Branch EfficientNetB0 architecture that fuses 3-channel RGB spectrograms with 19 handcrafted acoustic features. The XAI engine generates Grad-CAM spatial heatmaps and SHAP feature attribution scores simultaneously, providing clinician-interpretable evidence at both the spectrogram and feature levels.

The cross-dataset validation experiment on PhysioNet 2016 reveals a mean Jaccard similarity of 0.33 between datasets, confirming high domain shift as the primary cause of the observed accuracy reduction from 80.5% to 46% on raw transfer. This finding contributes a methodological framework for domain shift quantification in PCG research. The complete system is deployed as a 5-page React web application with a Flask REST API backend delivering real-time inference with full XAI output in under 3 seconds per recording. Future work includes implementing PDF report generation, multi-recording ensemble analysis, temperature scaling for model calibration, and training with additional PhysioNet data to improve cross-dataset generalization.

REFERENCES

[1]. S. K. Padhy, A. Mohapatra, and S. Patra, "X-CBNet: An Explainable Effective Deep Learning Framework Based on Spectrograms for Predicting Valvular Disorder using PCG Signals," Journal of Transformative Technologies and Sustainable Development, vol. 9, no. 1, pp. 1–18, Oct. 2025.
[2]. H. Alquran, Y. Al-Issa, M. Alsalatie, and S. Tawalbeh, "Deep learning models for segmenting phonocardiogram signals: a comparative study," PLOS ONE, vol. 20, no. 4, pp. 1–12, Apr. 2025.
[3]. B. Althaph and N. P. Challa, "Explainable attention-based deep learning for classification and interpretation of heart murmurs using phonocardiograms," Scientific Reports, vol. 15, no. 1, pp. 1–23, Jan. 2025.
[4]. M. Bahreini, R. Barati, and A. Kamali, "Cardiac sound classification using a hybrid approach: MFCC-based feature fusion and CNN deep features," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2025, no. 1, pp. 1–13, Jan. 2025.
[5]. Z. Ren et al., "A comprehensive survey on heart sound analysis in the deep learning era," Frontiers in Artificial Intelligence, pp. 3–29, Sep. 2024.
[6]. M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. 36th ICML, 2019.
[7]. R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in Proc. IEEE ICCV, 2017.

[8]. S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in NeurIPS, 2017.

[9]. J. Oliveira et al., "The CirCor DigiScope Phonocardiogram Dataset," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 6, pp. 2524–2535, Jun. 2022.

[10]. C. Liu et al., "An open access database for the evaluation of heart sound algorithms," Physiological Measurement, vol. 37, no. 12, pp. 2181–2213, 2016.

AUTHOR'S BIOGRAPHY

She received her B.E degree in CSE from Bharathidasan University, M.Tech degree in CSE from Pondicherry University and Ph.D.degree in CSE from Pondicherry University. She is working as Professor in the Department of Information Technology, Pondicherry Engineering College.

A B.Tech Information Technology Student in Puducherry Technological University with expertise in full-stack development. Proficient in Python, Java, C and C++, focusing on building scalable end-to-end applications and integrated system architectures. Passionate about developing data-driven solutions that bridge complex backend logic with responsive user interfaces.

A B.Tech Information Technology Student in Puducherry Technological University specializing in Artificial Intelligence and Machine Learning. Skilled in Python, Java, C and C++ with a focus on developing Deep Learning Models and Predictive Analytics. Dedicated to research in Neural Networks and Multimodal AI Pipelines to solve Real-World Automated Challenges.

A B.Tech Information Technology Student in Puducherry Technological University with a specialization in frontend development. Proficient in Java, Python, and modern UI/UX frameworks, focusing on creating intuitive, high-performance user interfaces. Passionate about enhancing user engagement through clean design and seamless cross-platform accessibility.