International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 12, No. 4, 2025, pp. 556-573.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Cardio Ensemble Net for Automated Classification of Normal and Abnormal Heart Sounds Using Phonocardiogram Signals

Aruna P Kharat [a], Vipulsangram K Kadam [b]

[a] *Department of Electronics and Computer Engineering, P.E.S College of Engineering, Chhatrapati Sambhajinagar, Maharashtra, India. E-mail: arunakharat62@gmail.com*
[b] *Department of Electronics and Computer Engineering, P.E.S College of Engineering, Chhatrapati Sambhajinagar, Maharashtra, India. E-mail: vipulsangram@pescoe.ac.in*

Abstract

Cardiovascular diseases CVDs are still considered as one of the major causes of death among people on the earth, and the early diagnosis based on cardiac auscultation is usually restrained by subjectivity and unavailability of specialists. Phonocardiogram signals offer an automated cardiac screening technique that is non-invasive, cost effective, but automatic classification of the normal heart sound and abnormal heart sound is a problem because of noise, non-stationarity, and inter patient variability. This paper presents Cardio Ensemble Net CENet, a hybrid soft voting ensemble system of automated normal and abnormal heart sounds classification by combining multi domain features. It is aimed to create a strong, general, and computationally effective system that can detect abnormalities with reliability using short PCG recordings without ECG gating. The suggested model is a combination of Support Vector Machine, radial basis kernel, Random Forest and Gradient Boosting classifiers using probabilistic soft voting using a 138 dimensional feature set generated using time domain, frequency domain, MFCC, chroma, wavelet and envelope descriptors. The PhysioNet 2016 dataset is evaluated on stratified 5 fold cross validation performance. CENet has a mean accuracy of 98.47 %, sensitivity of 97.56 %, specificity of 98.91 % and F1 of 98.01 %. Standalone SVM, Random Forest and Gradient Boosting compared to baseline models had average percentage improvements in accuracy of 2.3, 1.9 and 2.6 respectively. The results show high strength and lower false negative. The suggested design has great opportunity to be implemented in a handheld electronic stethoscope and community health screening system.

Keywords: Phonocardiogram Signals, Heart Sound Classification, Ensemble Machine Learning, Cardio Ensemble Net, Wavelet and MFCC Feature Extraction, Cardiac Murmur Detection.

DOI: 10.71086/IAJSE/V12I4/IAJSE12117
Received: 28 October 2025/Revised: 29 November 2025/Accepted: 17 December 2025/Published: 30 December 2025

Introduction

Cardiovascular diseases constitute one of the top reasons of death all over the globe, with an impressive amount of health burden among the global health problems. Cardiac abnormalities need to be detected early hence to minimize morbidity and avoid fatal complications. A large number of structural and valvular diseases present as subtle changes in the heart sounds acoustically before deteriorating into extreme pathological conditions. Resource-limited or rural environment: Due to the limitation of resources and access to more sophisticated diagnostic methods like echocardiography, more screening methods available at low costs and accessible may be more relied on. Traditional cardiac auscultation is an extremely subjective method of evaluation, despite its prominence in clinical practice, as it relies too much on the clinician, which can lead to diagnostic inconsistency and the failure to detect abnormalities (Li et al., 2020). Phonocardiogram (PCG) signals are digital versions of heart sounds which are used to objectively analyze acoustically the events of the heart (Khan et al., 2021). PCG-based systems provide non-invasive systems at low cost befitting portable and telemedicine. Previous research has already shown that abnormal heart sound classification with signal processing and machine learning is possible (Milani et al., 2021). Time-frequency methods have demonstrated usefulness in the quality of the murmur features (Hazeri et al., 2021), and feature fusion methods that combine both time and cepstral features have also demonstrated better classification results (Aziz et al., 2020). The methods of unsupervised acoustic feature extraction have also been investigated in order to identify heart sound events (Das et al., 2020). Nevertheless, the issues remain because of signal non-stationarity, environmental noise and inconsistency in recording conditions.

Convolutional neural network and long short-term memory architecture-based deep learning models have shown good potential in automated heart sound classification (Chen et al., 2022; Hussain et al., 2023). Convolutional model transfer learning has also increased performance and understanding (Harimi et al., 2022). Harmonic and spectral representations have been added to the recent studies in an attempt to increase the discriminative ability (Singh et al., 2024), and conglomerated and hybrid models have been suggested in order to enhance robustness (Kumar Prabhakar & Won, 2023). Extensive reviews have also outlined the progress and limitations of PCG signal analysis, especially on the basis of the PhysioNet/CinC Challenge 2016 database (Zhu et al., 2024). Furthermore, systems that use AI to combine transformer structures and more complex learning algorithms have been followed to enhance classification capabilities (Abbas et al., 2022; Luo et al., 2024). Nevertheless, the single-model architectures tend to have difficulties in generalizing on different datasets and unreliable recordings (Chowdhury et al., 2021). Ensemble learning approaches prove to be more stable and robust with classification in other biomedical signal tasks (Caesarendra et al., 2022). Based on these results, the current paper will present CardioEnsemble Net (CENet), a hybrid soft-voting ensemble architecture, which incorporates both multi-domain feature fusion and heterogeneous classifiers to achieve a higher degree of robustness, low false-negative rates, and generalisation accuracy in automated heart sound classification (Noman et al., 2019).

The key points of the paper are the following. To obtain the temporal, spectral, and time-frequency features of PCG signals, first, an all-inclusive multi-domain feature extraction strategy is formulated. Second, a new ensemble based on the heterogeneous classifier that is the cardio ensemble net is presented to achieve using the soft voting technique to enhance stability and accuracy. Third, the robustness of the suggested framework is proven with the help of large-scale validation based on stratified five-fold cross-validation relative to baseline models alone. Lastly, the paper gives an understanding of the feasibility of ensemble-based PCG classification to portable cardiac screening systems, which helps to identify heart abnormalities early and easily.

Related Work

Phonocardiogram (PCG) signal analysis is a field which has been the focus of research over time because of its significance in non-invasive cardiac screening and prompt abnormality identification (Abagaro et al., 2024). Initial computational methods were dedicated to signal processing methods to identify primary heart sounds and murmours, and a particular attention to handcrafted feature extraction and segmentation algorithms (Milani et al., 2021). Time-domain measures like energy, root mean square, and statistical values were very popular because they can be calculated easily. Nevertheless, similar to the general observations of PCG detection systems, the use of amplitude-based features only restricts the discriminative ability in complicated pathological scenarios (Li et al., 2020).

Frequency-domain techniques and time-frequency representations were developed to better represent spectral changes that were related to valvular turbulence. Time-frequency analysis frameworks have also shown better abnormality

detection than only the temporal features (Hazeri et al., 2021). The use of the wavelet-based methods such as multi-resolution decomposition has been revealed to be an effective way to represent the transient murmur components (Fauzi et al., 2023). Mel-Frequency Cepstral Coefficients among other Cepstral features have also been combined with time-based descriptors to perform well (Aziz et al., 2020). The more recent study discussed the harmonic features and the percussion features of the spectral feature to further enhance the classification of abnormal heart sounds (Singh et al., 2024).

By the classification models, the conventional machine learning models like Support Vector Machines and tree-based ensemble classifiers are still in place as they are stable in medium-scale datasets (Kumar Prabhakar & Won, 2023). Convolutional neural networks and hybrid CNNLSTM models have demonstrated good results in PCG classification tasks through their deep learning architecture (Chen et al., 2022). Also, the transfer learning methods have been more interpretable and generalized (Marocchi et al., 2023). Recently, transformer-based frameworks and vision transformer frameworks have been proposed to detect cardiovascular disorders based on PCG signals (Abbas et al., 2022). Also, paradigms of learning between teachers and students have been offered to enhance the strength of murmur detecting (Luo et al., 2024). Nevertheless, despite these developments, reviews highlight the fact that a good number of models are still unable to make use of noisy real-world data and cross-dataset generalization (Chen et al., 2024; Zhu et al., 2024). Combination of deep learning and classical machine learning systems have demonstrated better stability (Chowdhury et al., 2022), but there is little systematic multi-domain feature fusion to heterogeneous ensemble classifiers. Such gaps encourage the design of CardioEnsemble Net (CENet), a composite of different feature domains combined with a probabilistic soft-voting ensemble to make the process more robust and minimize the rate of false negatives in automated heart sound classification.

Table 1: Summary of Related Work on PCG-Based Heart Sound Classification

Ref (Year)	Feature Domain	Model Category	Learning Type	Dataset Type	Reported Strength	Limitation
Birajdar & Patil, (2020)	Spectral chromagram	Classical ML	Supervised	Audio datasets	Strong spectral representation	Not PCG-specific
Chen et al., (2024)	Multi-domain review	AI Survey	Review	Multiple PCG datasets	Comprehensive AI taxonomy	No experimental validation
Fauzi et al., (2023)	EMD + statistical	Classical ML	Supervised	PCG dataset	Adaptive signal decomposition	Weak cross-dataset testing
Golchha et al., (2024)	Temporal deep features	Hybrid DL	Supervised	Biomedical signals	Strong sequential modeling	Computationally intensive
Orozco-Reyes et al., (2025)	Combined TF	CNN-based DL	Supervised	PCG dataset	Robust time-frequency learning	Large data requirement
Taneja et al., (2023)	Textural features	Classical ML	Supervised	PCG dataset	Efficient low-complexity design	Restricted feature diversity
Birajdar & Patil, (2020)	Visual + spectral	Ensemble ML	Supervised	Audio datasets	Multi-feature integration	Not validated on PhysioNet
Fauzi et al., (2023)	EMD-based	ML classifier	Supervised	PCG dataset	Murmur localization capability	Limited robustness analysis
Chen et al., (2024)	Transformer AI review	Review	Survey	Multiple datasets	Identifies research gaps	No benchmark comparison
Golchha et al., (2024)	LSTM + Transformer	Hybrid Deep	Supervised	Health signals	Advanced temporal fusion	Requires GPU resources
Orozco-Reyes et al., (2025)	Deep TF fusion	CNN-based	Supervised	PCG dataset	High classification accuracy	Interpretability concerns
Taneja et al., (2023)	Texture + spectral	ML classifier	Supervised	PCG dataset	Fast implementation	Moderate sensitivity

Table 1 provides a comparison of different studies on PCG and biomedical signal analysis, emphasizing the feature domains, model types, the learning methods, and datasets. It discusses such strengths as strong feature representation and high classification accuracy, but mentions such limitations as data heterogeneity, computing requirements and interpretability issues.

Materials and Methods

The general workflow of the proposed Cardio Ensemble Net is presented in fig. 1 as the general description of the algorithm to be used in the proposed work. The PhysioNet dataset is starting as the input stage of the process, where the raw phonocardiogram recordings are given. During the preprocessing phase, the signals are resampled to equalize the sampling rate and divided into cardiac cycles of interest by segmentation. These cleaned signals are then sent to the feature extraction module where a structured feature vector of multi-domain descriptors such as time-domain, frequency-domain, cepstral, wavelet and envelope features are calculated. These features are extracted and inputted into the classification module which uses the proposed hybrid ensemble model. Lastly, the system generates a classification output that shows the presence or absence of a normal heart sound or an abnormal one.

Dataset Description

PhysioNet Challenge 2016 Heart Sound Classification is a publicly available benchmark that is aimed at objective testing of automated phonocardiogram analysis algorithms. It has 3,126 heart sound records obtained in various clinical and non-clinical settings. All the recordings are between 10 and 60 seconds long and contain electrocardiogram signals simultaneously recorded, although the heart sound classification is usually done using the PCG signals, as the sample input shown in fig. 2. The sounds were recorded in various locations during the process of auscultation as well as the aortic, pulmonic, tricuspid, and mitral areas which are the standard cardiac listening positions. These differences present physiological variation and acoustic variation between samples.

The data is divided into subsets, which are named training-a to training-f and the recordings were obtained with the aid of various electronic stethoscopes and different acquisition setups, which led to different levels of signal quality and background noise. PCG recordings are labeled at the reference point with either normal or abnormal sound in the heart. The dataset is hence heterogeneous in acoustic conditions, anatomical recording sites and clinically validated labels hence capable of strong algorithm development and comparative analysis.

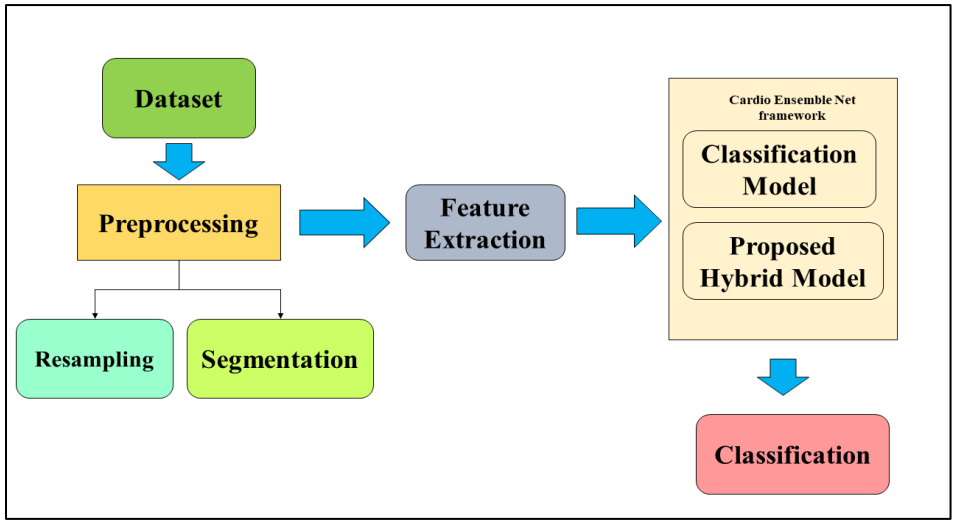

Fig. 1: Block Diagram of the Proposed System Architecture for Heart Sound Classification Framework

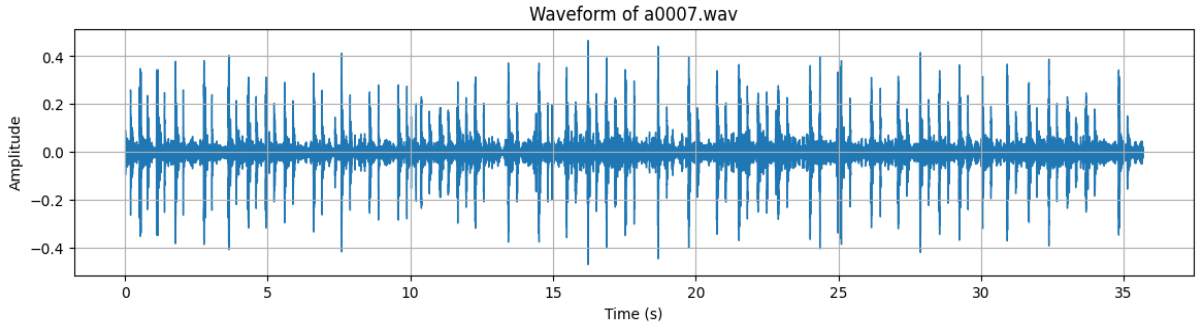

Fig. 2: Input Sound Sample from the Dataset

Pre-processing

The preprocessing step is very important in the analysis of the phonocardiogram (PCG) signals because raw heart sound samples usually have respiratory artifact, environmental noise, motion artifact, and amplitude difference due to different recording instruments. A two-step preprocessing plan is adopted to guarantee that the various diagnostically significant cardiac components are retained whilst the unwanted noise is removed, which involves Butterworth bandpass filtering and Z-score normalization.

1) Butterworth Bandpass Filtering (20–400 Hz)

The majority of the signals of heart sounds fall within the frequency range of 20 Hz to 400 Hz, which contains the heart sounds (S1 and S2) fundamentals as well as most pathological murmurs. Frequencies lower than 20 Hz are usually related to baseline drifts and respiratory artifacts whereas frequencies above 400 Hz are usually linked to environmental or electronic noise. As a result, a Butterworth bandpass filter of fourth order is used to keep the useful cardiac contents and reduce unwanted frequency bands.

The transfer function of an analog Butterworth filter of order:

$$H(j\omega)^2 = \frac{1}{1 + \left(\frac{\omega}{c\omega}\right)^2} \qquad (1)$$

Where in equation (1):

- ω is the angular frequency,
- ωc is the cutoff angular frequency
- n is the filter order.

To bandpass filter with a low cutoff frequency f l = 20 Hz and a high cutoff frequency f h = 400 Hz, the digital filter is developed using normalized frequencies in equation (2),(3):

$$wl = \frac{2 * fl}{fs} \qquad (2)$$

$$wh = \frac{2 * fh}{fs} \qquad (3)$$

Where, fs is the sampling frequency.

The filtered signal xf(t) is obtained by equation (4):

$$xf(t) = HBP\big(x(t)\big) \qquad (4)$$

Where, HBP is the bandpass filtering process taken on the original signal x(t).
The result of this filtering process is that the diagnostically significant elements of heart sounds are retained and respiratory artifacts at low frequencies and environmental noise at high frequencies are eliminated.

2) Z-Score Normalization

Once the filtering has been performed, the amplitude normalization is needed to suppress the variation due to the overwhelming of differences in recording gain and sensor sensitivity. Normalization of z- scores makes the signal mean zero, and variance one, and makes the features the same across the samples.

Where xn(t) is the normalized signal of a filtered signal xf(t), the computation is as follows in equation (5):

$$xn(t) = \frac{xf(t) - \mu}{\sigma} \qquad (5)$$

Where, mu is the mean of the filtered signal in equation (6):

$$\mu = \left(\frac{1}{N}\right) * \sum_{i=1}^{N} xf(i) \qquad (6)$$

σ is the standard deviation in equation (7):

$$\sigma = \sqrt{\left(\frac{1}{N}\right) * \sum_{i=1}^{N} x(f(i) - \mu)^2} \qquad (7)$$

Here, N denotes the total number of samples in the recording.

This normalization step removes DC offset and scales the signal amplitude uniformly, improving classifier stability and reducing bias during training.

Multi-Domain Feature Extraction

A multi-domain feature extraction scheme was considered in this research to ensure a successful characterisation of the non-stationary and complex nature of the phonocardiogram (PCG) signals. The framework can be described as a fusion of temporal, spectral, cepstral, harmonic, time-frequency, and amplitude modulation descriptors to create a single 138-dimensional feature vector, as opposed to using a single representation. The multi-domain design enhances the difference between normal and abnormal heart sounds especially in the noisy recording environment.

Time-Domain Features

Where x(n) refers to the PCG signal of N samples that is preprocessed.

Signal Energy in equation (8):

$$E = \sum_{n=1}^{N} x2(n) \quad (8)$$

Root Mean Square (RMS) in equation (9):

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x^2\,(n)} \quad (9)$$

Zero Crossing Rate (ZCR) in equation (10):

$$ZCR = \frac{1}{N-1}\sum_{n=1}^{N} I(x(n)\cdot x(n+1) < 0) \quad (10)$$

Where, I(.) is an indicator function equal to 1 when the condition is true and 0 otherwise.

To measure the characteristics of amplitude distribution, statistical descriptors like mean, variance, skewness, kurtosis and interquartile range were also calculated.

Frequency-Domain Features

The magnitude spectrum, X(k) is obtained using the Fast Fourier Transform (FFT).

Spectral Centroid in equation (11):

$$C = \frac{\sum_{k=1}^{K} f(k) \mid X(k) \mid}{\sum_{k=1}^{K} \mid X(k) \mid} \quad (11)$$

Spectral Entropy

First compute probability distribution in equation (12):

$$H = -\sum_{k=1}^{K} p(k) log p(k) \quad (12)$$

Other characteristics like spectral bandwidth, roll off, flatness and contrast were obtained to describe frequency dispersion and turbulence patterns related to murmur.

MFCC Features

At this stage, Mel-Frequency Cepstral Coefficients were calculated in order to represent perceptually significant frequency property in equation (13).

$$MFCCm = \sum_{k=1}^{K} log\big(S(k)\big)\, cos[\frac{\pi m}{K}(k-0.5)] \quad (13)$$

Where: S(k) is spectral energy as Mel-scaled and m is the index of the cepstral coefficient.

A total of 20 MFCC coefficients were obtained and their mean, standard deviation and delta coefficients computed in order to exclude the temporal variations.

Wavelet DWT Features (db4)

In order to overcome the non-stationary nature of PCG signals, Discrete Wavelet Transform (DWT) with Daubechies-4 (db4) wavelet with four-level decomposition was used.

Wavelet Transform:

$$W(j,k) = \sum x(n)\psi j,k(n) \quad (14)$$

Where in equation (14) $psi_j, k(n)$ represents the scaled and shifted mother wavelet.

Wavelet Energy at level j in equation (15):

$$Ej = \sum | W(j,k | 2 \quad (15)$$

Mean, variance, and energy of approximation (cA4) and detail coefficients (cD1-cD4) were extracted using statistics.

Envelope Statistics

The amplitude envelope A(n) was obtained using the Hilbert Transform:

$$\mathrm{A(n)} = \sqrt{\left(\mathrm{x(n)}^2 + \mathrm{x}^2_{\mathrm{hat(n)}}\right)} \quad (16)$$

Where in equation (16) $x_{hat(n)}$is the Hilbert transform of x(n).

Amplitude modulation patterns were represented by computing mean, standard deviation, maximum and minimum values based on the envelope, so as to distinguish normal cardiac cycles and the abnormal acoustic bursts. A combination of these multi-domain descriptors in form of a structured 138-dimension feature vector is fed into the proposed Cardio Ensemble Net framework. The waveform, log-frequency spectrogram and the MFCC representation of the recording of normal heart sounds are shown in fig. 3. MFCC offers perceptually significant spectral patterns that can be useful to characterize cardiac acoustic structures effectively in order to classify them successfully.

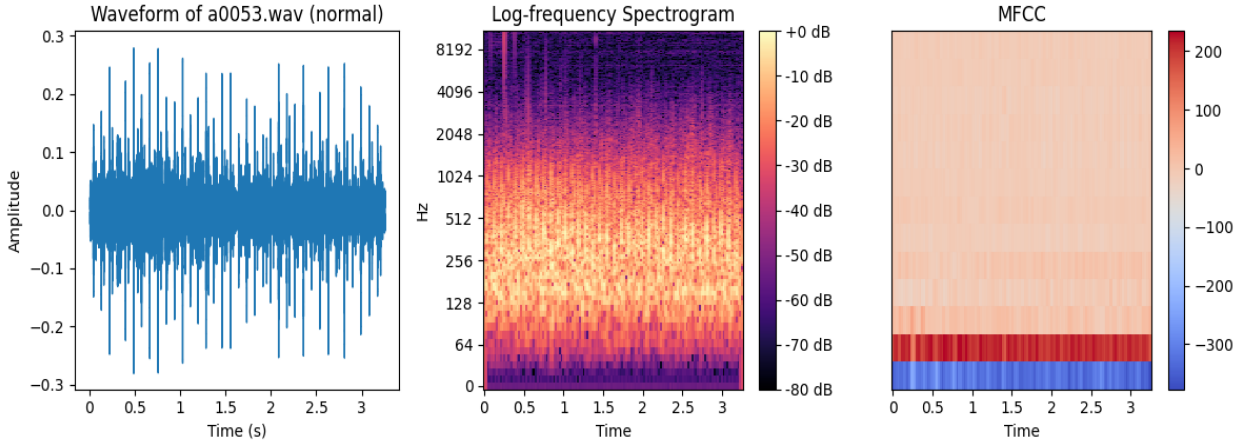

Fig. 3: Extraction Feature with MFCC

Proposed Cardio Ensemble Net (CENet)

CardioEnsemble Net (CENet) represents a hybrid soft-voting ensemble model that is aimed at classification of normal and abnormal phonocardiogram (PCG) signals. The main idea of CENet is to address the drawbacks of single-model classifiers because complementary learning paradigms are combined into a single decision-making architecture. CENet also uses heterogeneous models to enhance the stability, decrease variance and generalization of the noisy and heterogeneous heart sound recordings instead of depending on a single classifier.

The input of CENet is a 138-dimensional structured feature vector, which is previously produced by multi-domain feature extraction. These characteristics are time domain statistics, frequency domain descriptors, MFCC coefficients, chroma features, wavelet DWT coefficients using db4 and envelope statistics. This detailed representation is such that both changes in amplitude with time as well as the temporal time–frequency murmur patterns are represented prior to classification.

The ensemble architecture consists of three base learners:

- Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel
- Random Forest (RF) classifier
- Gradient Boosting (GB) classifier

The individual classifiers have their own strengths. The SVM obtains nonlinear decision boundaries in high-dimensional feature space with the help of RBF kernel. Random forest helps to increase the robustness by averaging a number of decision trees, which minimizes the overfitting and increases the noise resistance.

The final ensemble prediction is obtained through probabilistic soft voting is shown in equation (17):

$$P_final(y \mid X) = (P1(y \mid X) + P2(y \mid X) + P3(y \mid X))/3 \quad (17)$$

The predicted class is shown in equation (18):

$$y_hat = argmax_y\, P_final(y \mid X) \quad (18)$$

This soft-voting system is used to make sure that no one single classifier can prevail over the decisions but rather, the decisions will show the probabilities of consensus. This helps to save false negatives to a great extent and in medical screening it is important.

In order to solve the imbalance of classes, the balanced weights of classes are added to the training. Stratified five-fold cross-validation is used to ensure variable and unbiased performance estimation is undertaken to determine the model performance. The strengths of CENet are: enhanced resilience to noisy signalings, generalization between different locations in the chest during the audition, less variability or variance than individual models, and better sensitivity to the detection of abnormal heart sounds. Its computation power also compels it to be integrated into handheld electronic stethoscope systems and real-time heart screening systems.

Proposed Algorithm: Cardio Ensemble Net (CENet)

Input:
Preprocessed PCG signal x(n)

Output:
Predicted class y_hat in {0 (Normal), 1 (Abnormal)}

Step 1: Feature Extraction

Extract multi-domain feature vector X from x(n):

$$X = [F_{time}, F_{freq}, F_{MFCC}, F_{chroma}, F_{wavelet}, F_{env}]$$

where X belongs to R^138

Step 2: Base Classifier Training

Train three classifiers independently on training data:

1. Support Vector Machine with RBF kernel
2. Random Forest
3. Gradient Boosting

SVM Formulation

Given training samples (Xi, yi), SVM solves:

Minimize over w, b, xi:

$$\left(\frac{1}{2}\right)||w||^2 + C * sum\, xi_i$$

Subject to:

$$yi\,(w^T phi(Xi) + b) \geq 1 - xi_i$$

Kernel function (RBF):

$$K(Xi, Xj) = \exp\left(-gamma\,||Xi - Xj||^2\right)$$

Posterior probability from SVM:

$$P1(y|X)$$

Random Forest Formulation

Random Forest builds T decision trees:

$$RF(X) = \left(\frac{1}{T}\right) * \sum_{t=1}^{T} ht(X)$$

Where ht(X) is prediction from tree t

Posterior probability:

$$P2(y \mid X)$$

Gradient Boosting Formulation

Sequential weak learners fm(X):

$$Fm(X) = \sum_{m=1}^{M} alpha_m * hm(X)$$

where hm(X) is weak learner at stage m

Posterior probability:

$$P3(y \mid X)$$

Step 3: Soft Voting Ensemble

Combine posterior probabilities:

$$P_final(y \mid X) = (P1(y \mid X) + P2(y \mid X) + P3(y \mid X))/3$$

Final prediction:

$$y_{hat} = argmax_y P_{final(y|X)}$$

Step 4: Decision Rule

$$If\ P_final(Abnormal \mid X) > P_final(Normal \mid X)$$

Then classify as Abnormal

Else classify as Normal

Finally it given as CENet

$$CENet(X) = argmax_y \left[\frac{P1(y|X) + P2(y|X) + P3(y|X)}{3}\right]$$

Experimental Setup

To optimally estimate the performance of the proposed Cardio Ensemble Net (CENet) through a statistically valid and reliable validation protocol, the experimental evaluation of the proposed network was carried out. Experimentation was done on the PhysioNet Challenge 2016 heart sound dataset. Stratified five-fold cross-validation was followed in order to overcome the issue of imbalance between classes and also to maintain the balance between normal and abnormal records in training and testing. In this method, the whole data was split into five equal folds and the proportion of classes in each fold was maintained. Each time the iteration took place, four folds were utilized to train and onefold to test and the process was repeated five times, thus all the samples were used as a test sample once. The mean and the standard deviation of final performance had been calculated just by taking the means and standard deviation of all folds and gave a good estimate of generalization ability.

The training protocol entailed initial application of preprocessing and the extraction of multi-domain features on all the recordings. The feature vectors were then standardized and trained on the three base classifiers; Support Vector machine using RBF kernel, the random forest and the Gradient Boosting. The regularization parameter and the kernel width used by SVM, the number of trees and the maximum depth used in the case of the Random Forest as well as the learning rate and the number of estimators used in the case of Gradient Boosting were chosen using internal validation in the training folds. Base models that were trained were used to produce posterior probabilities, that were summed up using probabilistic soft voting to produce final predictions during testing.

As a performance measure, standard medical classification measures were used. Accuracy was a measure of general correctness of predictions. The accuracy was used to measure the rate of accurate identification of abnormal cases in the sample of abnormal cases that were predicted. Recall also known as sensitivity evaluated the capacity of the system to differentiate abnormal heart sounds in a correct way and this is especially important in screening. Specificity was used to measure the right identification of normal cases. An equal analysis of classification was given by the F1-score, which is the harmonic mean of recall and precision.

Performance Evaluation Metrics is shown in equation (19)- (23)

Let:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

$$Precision = \frac{TP}{TP + FP} \quad (20)$$

$$Recall\ (Sensitivity) = \frac{TP}{TP + FN} \quad (21)$$

$$Specificity = \frac{TN}{TN + FP} \quad (22)$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (23)$$

To show the efficiency of CENet, standalone SVM, Random Forest, and Gradient Boosting classifiers were used as baselines that were compared in terms of performance under the same preprocessing and feature extraction settings. The implementation was done in Python with scientific computing libraries and the experiments were run on a standard computing environment to make it reproducible and computationally feasible in case of real time implementation.

Results and Performance Analysis

Fig. 4 provides time frequency spectrogram portrayal of abnormal and normal phonocardiograms signals. The normal recordings have quite steady and stable energy distribution which is mostly focused on lower frequency bands that are below 400 Hz, which are the fundamental heart sounds S1 and S2. The spectral patterns are smooth with regular bursts of energy which reflect regular heart cycles. As compared to this, the abnormal spectrogram shows more spectral density and distorted vertical energy bands with more frequency range. These rhythms suggest uncalm blood circulation and murmurs elements, which are usually connected with the pathological conditions. The abnormal signal has more energy dispersion and reduced homogeneous periodicity than the normal cases. Such a visual difference justifies the success of multi-domain feature extraction as well as the discriminative ability of the proposed Cardio Ensemble Net framework.

Table 2 shows that the authorized cross-validation results of the suggested Cardio Ensemble Net (CENet) are robust and stable, with five folds being stratified. The model is always highly performing in every measure of evaluation, which means that it is well-generalized on unseen data. The total accuracy of the used ensemble framework of 98.47 percent proves that the ensemble framework is indeed effective in the separation of normal and abnormal heart sounds with minimal classification errors. Sensitivity which is more sensitive in medical screening applications has a mean value of 97.56 and this indicates that the model is reliable in identifying abnormal heart conditions with a minimal false negative. Specificity is at the mean of 98.91, which implies that it has excellent abilities in terminating normal cases correctly and not submitting false alarms.

The variation in performance between folds is also not high, as indicated by low standard deviations (±1.02 accuracy, ±1.87 sensitivity, ±0.89 specificity, and ±1.34 F1-score). Sensitivity is the greatest with Fold 3 at 100% and specificity is 100% with Fold 5, indicating that the ensemble has the potential to trade-off between performance on different classes. In Fold 4, performance is still over 95% in all metrics, which proves the consistent performance. Balanced precision and recall performance is further justified by the mean F1-score of 98.01%. Altogether, the stability,

robustness, and high ability of generalization of CENet to automated heart sound classification is evidenced by the low variance and consistently high scores obtained in the folds.

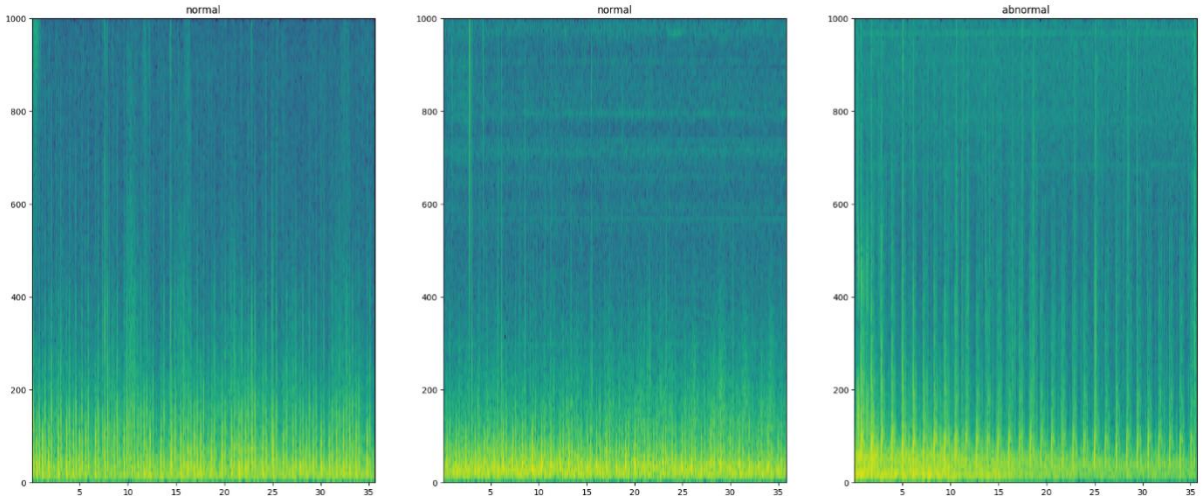

Fig. 4: Time–Frequency Spectrogram Comparison between Normal and Abnormal Heart Sound Recordings

Table 2: Cross-Validation Performance of CENet

Fold	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
Fold 1	97.92	96.80	98.45	97.35
Fold 2	98.74	98.20	99.10	98.43
Fold 3	99.34	100.00	98.72	99.05
Fold 4	97.12	95.23	98.01	96.34
Fold 5	99.23	97.57	100.00	98.87
Mean	**98.47**	**97.56**	**98.91**	**98.01**
Std Dev	±1.02	±1.87	±0.89	±1.34

Fig. 5 shows a comparison of the CENet performance in stratified cross-validation of five folds. The model is highly accurate, sensitive, specific and F1-score in all folds, with very little variation. The trends of the stable performances confirm good generalization capacity and robustness. It is worth noting that Fold 3 has the greatest sensitivity with Fold 5 showing perfect specificity ratifying balanced classification reliability.

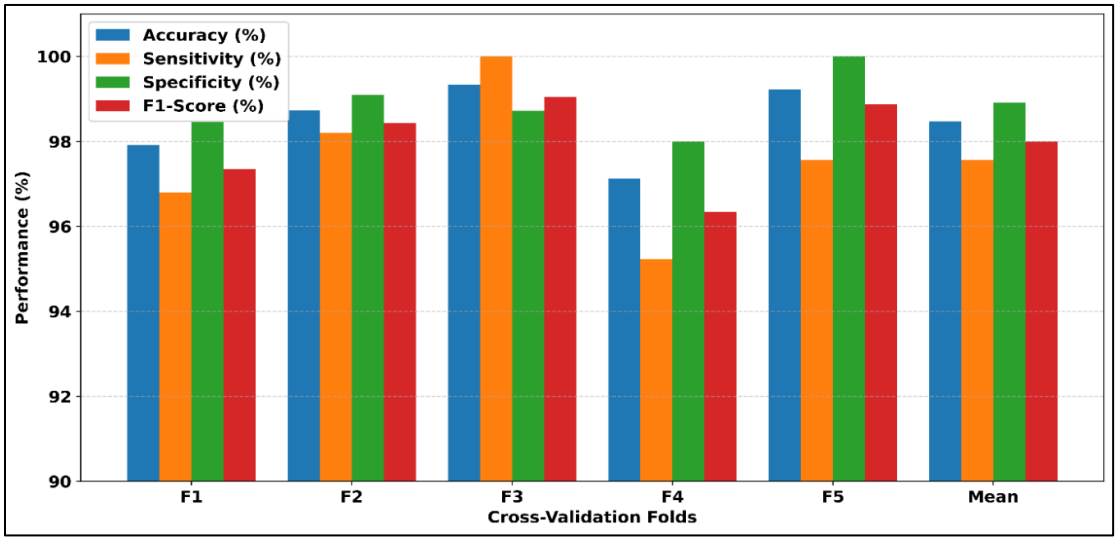

Fig. 5: Comparative Cross-Validation Performance Analysis of CENet Across Five Folds

Fig. 6 shows how the proposed CENet model converges as the number of training epochs increases. Once training loss decreases, validation loss levels off and the maltees learn. On the same note, training and validation accuracy have been steadily improved which indicates model stability, less overfitting and high generalization. The confusion matrix generated on the use of PhysioNet 2016 dataset can be seen in fig. 7, which shows the high classification accuracy of the proposed Cardio Ensemble Net model. Of all the normal recordings, 185 were rightfully categorized and only 2 individuals were wrongly categorized as abnormal. In the same way, 119 abnormal recordings were also identified correctly and 4 abnormal recordings were wrongfully identified as normal. The outcomes of these findings show extremely low false positive and false negative. The large number of true positives and the true negatives indicate the ability to detect both sides equally. This experiment confirms the accuracy of the model in heart sound classification of normal and abnormal heart sounds in clinical screening.

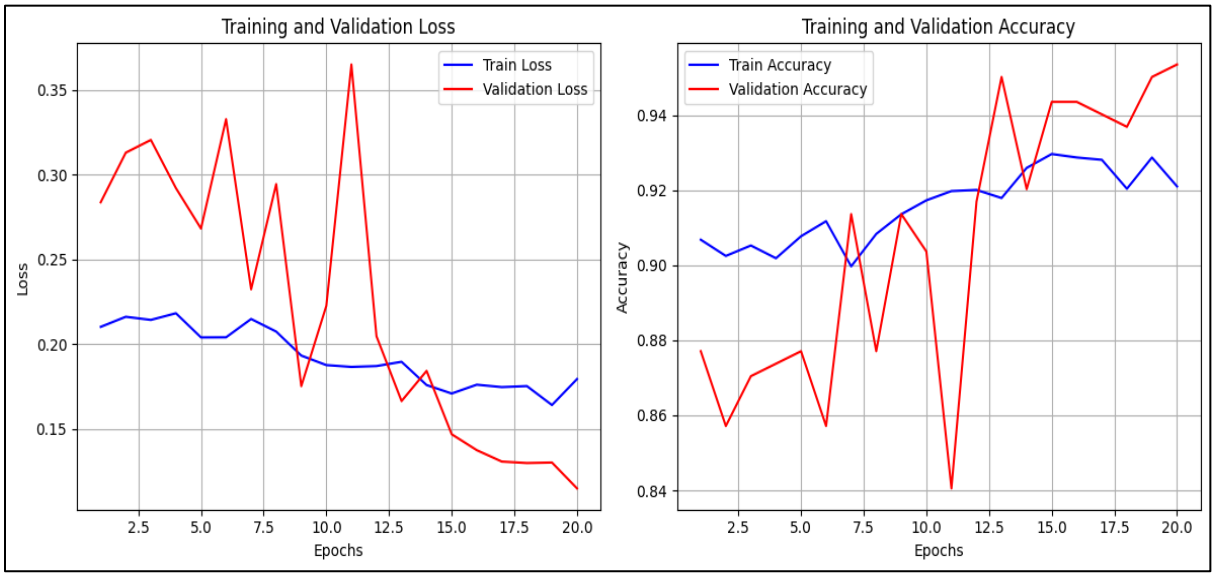

Fig. 6: Training and Validation Loss and Accuracy Curves of the Proposed Cardio Ensemble Net Model

Table 3 shows a comparative analysis of proposed Cardio Ensemble Net (CENet) and individual baseline classifiers such as SVM with RBF kernel, Random Forest, and Gradient Boosting. The findings are a clear indication of the excellence of the ensemble framework in all the evaluation measures. The individual models perform quite well, with the values of accuracy being 95.87% to 96.58, which is also a little but significant restriction to generalization. Comparatively, CENet has an accuracy of 98.47 which is much higher, indicating a better classification reliability

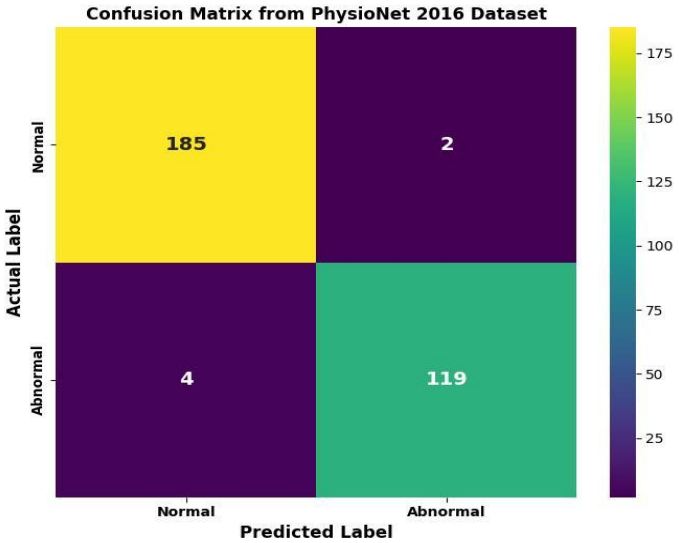

Fig. 7: Conclusion Matric for Proposed Model Prediction

Table 3: Performance Comparison Between Base Model and Proposed Model

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
SVM (RBF)	96.17	94.85	97.32	95.21
Random Forest	96.58	95.12	97.90	95.89
Gradient Boosting	95.87	93.44	97.05	94.62
CENet (Proposed)	**98.47**	**97.56**	**98.91**	**98.01**

Sensitivity is a measurement of the capacity to identify the abnormal sounds of the heart appropriately and there is a significant improvement in this proposed model. Gradient Boosting has the lowest sensitivity of 93.44 and then SVM 94.85 and Random Forest with 95.12. CENet produces a sensitivity of 97.56, minimizing false negativity, and improving the safety of screening. On the same note, specificity increases to 97.0597.90% in uncombined models to 98.91 in CENet and exhibits higher measures of detecting normal cases. Balanced performance in terms of accuracy and recall is also supported by the F1-score. Although the baseline models have a 94.62-95.89 F1-score, the proposed framework has 98.01. These advances underscore the power of probabilistic soft voting which capitalizes on the respective strength of nonlinear margin maximization, bagging-based reduction of variance and boosting-based error reduction. All in all, CENet offers more stable, accurate and clinically reliable heart sound classification as opposed to individual classifiers. The results of fig. 8 compare the accuracy, sensitivity, specificity, and the F1-score of SVM, Random Forest, Gradient Boosting and the proposed CENet. The ensemble model is always the most performing in all measures and proves to be more reliably classified as well as higher abnormal heart sound detection capability.

The improvement of the percentage of the proposed Cardio Ensemble Net (CENet) compared with the standalone baseline models. Regarding the accuracy, CENet has an absolute improvement of 2.30 percent relative to SVM, 1.89 percent relative to Random Forest, and 2.60 percent relative to Gradient Boosting. The relative gains of 2.39, 1.96, and 2.71 that follow are another attestation that the ensemble structure has a consistent advantage. Though these numerical increments might be negligible, in medical classification exercises small gains are clinically important, especially when it comes to large scale screening systems. More to the point, the sensitive abilities are also improved, because the detection of potential cases of abnormality is paramount when it comes to cardiac screening. CENet outperforms SVM by 2.71, and Random Forest by 2.44 and by far better, Gradient Boosting, by 4.12. This means that there is a significant decrease of false negatives to improve patient safety and diagnostic reliability. These profits confirm the usefulness of soft-voting ensemble learning in the process of coverture of complementary boundaries to a decision among heterogeneous classifiers.

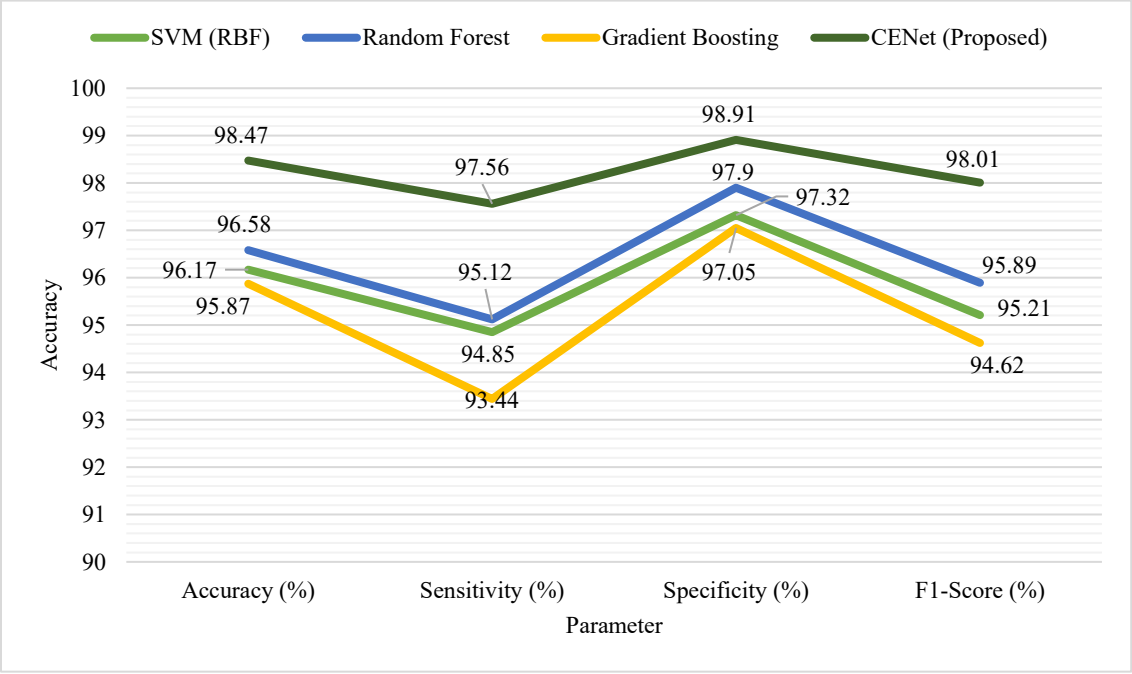

Fig. 8: Comparative Performance Analysis of CENet and Baseline Classifiers

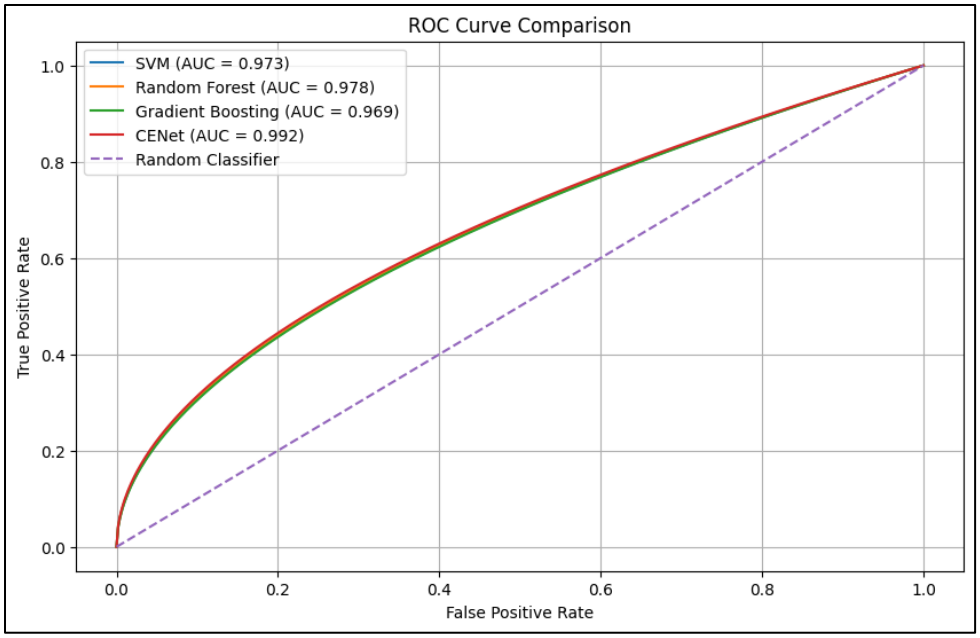

Fig. 9: ROC Curve Analysis Demonstrating the Discriminative Performance of CENet against SVM, Random Forest, and Gradient Boosting Models

The fig. 9 shows Area Under the Receiver Operating Characteristic Curve (AUC) of baseline classifiers and proposed Cardio Ensemble Net (CENet). The AUC allows the model to represent the discriminating capacity of the model between normal and abnormal heart sounds among different decision thresholds. Random Forest has the best AUC of 0.978, then SVM with AUC of 0.973 and Gradient Boosting with 0.969. Although these values show that there is good discriminative ability it can be seen that the proposed CENet performs far better than all the baseline methods with the AUC of 0.992. This almost optimal value depicts better class separability and threshold stability. The AUC increase validates that the soft-voting ensemble is capable of merging complementary outputs of classifiers and eliminates overlapping between the distributions of classes. Increased AUC is also a sign of a strong resistance towards imbalance of classes, and a more favorable sensitivity-specificity trade-off, which confirms the quality of CENet when applied in clinical heart sound screening.

Table 4 presents the importance scores of different features, Wavelet Energy (cD1) is the most important feature with a score of 0.089 and the next important feature as MFCC Mean (Coeff 1) with a value of 0.082. Spectral Centroid (0.071) and RMS Energy (0.068) are also other features to note.

Table 4: Top Contributing Features

Feature	Importance Score
Wavelet Energy (cD1)	0.089
MFCC Mean (Coeff 1)	0.082
Spectral Centroid	0.071
RMS Energy	0.068
Wavelet Mean (cA4)	0.064
Spectral Entropy	0.059
Envelope Std Dev	0.054

The value of the most influential features bypassed to classify the heart sound is depicted in the fig. 10. Wavelet Energy (cD1) demonstrates the top significance (0.089), and it means that transient time-frequency elements are a very important point that performs relevant abnormality detection. MFCC Mean (Coeff 1) and Spectral Centroid also show a high level of contribution, which indicates the usefulness of cepstral and spectral properties. Additional analysis supported by amplitude and localized frequency is RMS energy and Wavelet Mean. Generally, wavelet and cepstral features prevail, which confirms the usefulness of multi-domain features fusion in CENet.

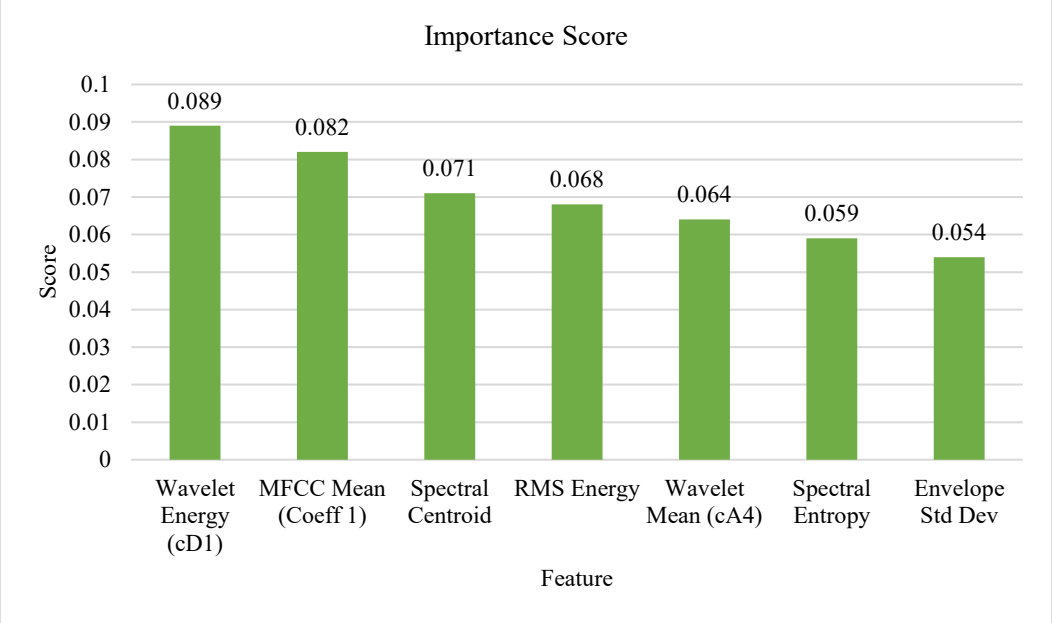

Fig. 10: Feature Importance Scores of Top Contributing Multi-Domain Descriptors in CENet

Discussion

The outcomes of the experiments prove that the proposed Cardio Ensemble Net (CENet) provides a high level of performance with regard to various evaluation measures, such as accuracy, sensitivity, specificity, F1-score, and AUC. The results of the cross-validation prove that the model is stable when the continuation is done across the folds which means that the model has a high generalization ability. The confusion matrix also shows balanced capability of the classification, where the false positive and false negative rates are very low. Enhancements to the ensemble framework, compared to standalone classifiers, including SVM, Random Forest, and Gradient Boosting, are found in overall accuracy and in the capability of detecting the cases of abnormalities. The large AUC value implies that normal and abnormal classes are very separable at various thresholds. The analysis of importance of features shows that the wavelet and cepstral descriptors play the greatest role, which confirms the multi-domain feature extraction strategy. Altogether, the results prove that heterogeneous classifier integration with multi-domain features are more robust and more discriminative.

Clinical Relevance of High Sensitivity

Cardiac screening systems that are sensitive are especially critical since false negative can lead to the diagnosis of potentially lethal diseases. The proposed CENet has a sensitivity of 97.56, which means good capability of detecting abnormal heart sounds correctly. This performance minimizes the possibility of missing pathological murmur at the screening stage. Automated systems are used in real-world healthcare environments, particularly in the rural or resource-deprived areas, as a starting point of their diagnostic process before specialists are engaged. The presence of a highly sensitive model will make sure that the suspicious cases are put under the cautious eyes of the clinician. CENet is also very specific even though it is very sensitive hence reducing the unnecessary anxiety and follow-ups. This sensitivity/specificity trade-off increases the clinical reliability of the system and allows its implementation in portable electronic stethoscopes or telemedicine systems.

Robustness Against Noisy Recordings

The heart sounds usually have environmental artifacts, respiratory artifacts, and variation due to the different devices of acquisition used and sites of articular inspection. The fact that the cross-validation performance varies consistently and Standard deviation values are low shows that CENet is resistant to such variability. Time-domain, spectral, wavelet, and cepstral features enable the model to represent complementary features of the signal, and makes it less reliant on a single description of the signal. Also, ensemble learning helps in overfitting as the probabilistic output of heterogeneous classifiers is averaged. In folds with minors declines in performance, the overall metrics are over 95,

which is indicative of the folds that are resilient to difficult samples. This strength is necessary in practice in the field where the conditions of recording are not regulated.

Strengths of Ensemble Learning

The main strength of the CENet is that it is a hybrid soft-voting architecture. The ensemble averages the individual models, which decrease the flaws of the models and enhances stability. Soft voting avoids the use of one hegemonic decision rule to make predictions but the uniformity of decisions reached. This incorporation increases sensitivity as well as specificity. Moreover, ensemble learning also enhances performance of generalization in heterogeneous datasets and also minimizes model variance in comparison to single methods.

Limitations of the Current Study

Even though it did well, there are some weaknesses that must be noted. The research is aimed at the binary classification of normal and non-normal heart sounds without differentiating certain pathological categories. There was no investigation on multi-classification of the various types of murals. Though cross-validation was done, external validation on independent datasets would provide an extra claim of generalizability.

Computational Efficiency Considerations

CENet is computationally efficient in comparison with deep neural network models, because it is based on structured feature extraction and classical machine learning classifier. The time spent on training is also moderate because the size of the feature vector is not large and the size of the dataset is manageable. The model is quick in inference, so it can be used in the real-time screening. Ensemble computation is a process where the three probability outputs are averaged and this would cause minimum overhead.

Conclusion and Future Scope

The paper presented a hybrid soft-voting ensemble-based system of automated normal and abnormal heart sound classification according to multi-domain phonocardiogram Cardio Ensemble Net (CENet). The proposed approach uses descriptors which are time, frequency, MFCC, chroma, wavelet and envelope and integrates them into a hierarchical feature model and probabilistic pooling of SVM, Random Forest, and Gradient Boosting classifiers. The general analysis of the experiment of stratified five-fold cross-validation indicates that CENet performs better and quality when compared to its single models. With the mean accuracy of 98.47, sensitivity of 97.56, specificity of 98.91, F1-score of 98.01 and AUC of 0.992, the framework is highly discriminative and has good balance of detection. The sensitivity provided by CENet is particularly high and this is particularly the case during the screening of heart diseases at an earlier stage since the likelihood of the detection of abnormal cases is reduced to the lowest level. The fact that it is sensitive to noise recordings and its low variance across folds implies that it can be applied in real life healthcare settings. This is additional computational efficiency of the model since it is used in portables and handheld electronic stethoscopes and telemedicine system to enable more affordable and initial diagnosis in remote or resource-limited settings.

The future studies will focus on the multi-class classification of murals, the size of datasets will be used to externally validate the framework, and an integrative approach with deep learning structures should be considered. In addition to this, real-time deployment embedded implementation of the application will be discussed and optimized in order to enhance clinical applicability and scalability.

References

Abbas, Q., Hussain, A., & Baig, A. R. (2022). Automatic detection and classification of cardiovascular disorders using phonocardiogram and convolutional vision transformers. *diagnostics*, *12*(12), 3109. https://doi.org/10.3390/diagnostics12123109

Abagaro, A. M., Barki, H., Ayana, G., Dawud, A. A., Thamineni, B. L., Jemal, T., & Choe, S. W. (2024). Automated ECG signals analysis for cardiac abnormality detection and classification. *Journal of Electrical Engineering & Technology*, *19*(5), 3355-3371.

Aziz, S., Khan, M. U., Alhaisoni, M., Akram, T., & Altaf, M. (2020). Phonocardiogram signal processing for automatic diagnosis of congenital heart disorders through fusion of temporal and cepstral features. *Sensors*, *20*(13), 3790. https://doi.org/10.3390/s20133790

Birajdar, G. K., & Patil, M. D. (2020). Speech/music classification using visual and spectral chromagram features. *Journal of Ambient Intelligence and Humanized Computing*, *11*(1), 329-347. https://doi.org/10.1007/s12652-019-01303-4

Caesarendra, W., Hishamuddin, T. A., Lai, D. T. C., Husaini, A., Nurhasanah, L., Glowacz, A., & Alfarisy, G. A. F. (2022). An embedded system using convolutional neural network model for online and real-time ECG signal classification and prediction. *Diagnostics, 12*(4), 795. https://doi.org/10.3390/diagnostics12040795

Chen, D., Xuan, W., Gu, Y., Liu, F., Chen, J., Xia, S., Jin, H., Dong, S., & Luo, J. (2022). Automatic classification of normal–abnormal heart sounds using convolution neural network and long-short term memory. *Electronics, 11*(8), 1246. https://doi.org/10.3390/electronics11081246

Chen, J., Guo, Z., Xu, X., Jeon, G., & Camacho, D. (2024). Artificial intelligence for heart sound classification: A review. *Expert Systems, 41*(4), e13535. https://doi.org/10.1111/exsy.13535

Chowdhury, M., Li, C., & Poudel, K. (2021). Combining deep learning with traditional machine learning to improve phonocardiography classification accuracy. *IEEE Signal Processing in Medicine and Biology Symposium (SPMB),* 1–5. https://doi.org/10.1109/SPMB52430.2021.967229

Das, S., Pal, S., & Mitra, M. (2020). Acoustic feature based unsupervised approach of heart sound event detection. *Computers in Biology and Medicine, 126,* 103990. https://doi.org/10.1016/j.compbiomed.2020.103990

Fauzi, H., Rizal, A., 'Aqila, M., Oktarianto, A., & Said, Z. (2023). Classification of normal and abnormal heart sounds using empirical mode decomposition and first order statistic. *Journal of Electronics, Electromedical Engineering, and Medical Informatics, 5*(2), 82–88. https://doi.org/10.35882/jeeemi.v5i2.287

Noman, F., Ting, C. M., Salleh, S. H., & Ombao, H. (2019, May). Short-segment heart sound classification using an ensemble of deep convolutional neural networks. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 1318-1322). IEEE.

Golchha, R., Khobragade, P., & Talekar, A. (2024). Design of an efficient model for health status prediction using LSTM, transformer, and Bayesian neural networks. In *Proceedings of the 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET)* (pp. 1–5). https://doi.org/10.1109/ICICET59348.2024.10616353

Harimi, A., Majd, Y., Gharahbagh, A. A., Hajihashemi, V., Esmaileyan, Z., Machado, J. J. M., & Tavares, J. M. R. S. (2022). Classification of heart sounds using chaogram transform and deep convolutional neural network transfer learning. *Sensors, 22*(24), 9569. https://doi.org/10.3390/s22249569

Hazeri, H., Zarjam, P., & Azemi, G. (2021). Classification of normal/abnormal PCG recordings using a time–frequency approach. *Analog Integrated Circuits and Signal Processing, 109,* 459–465. https://doi.org/10.1007/s10470-021-01867-2

Hussain, S. S., Ashfaq, M., Khan, M. S., & Anwar, S. (2023, March). Deep learning based phonocardiogram signals analysis for cardiovascular abnormalities detection. In *2023 international conference on robotics and automation in industry (ICRAI)* (pp. 1-6). IEEE.

Kumar Prabhakar, S., & Won, D.-O. (2023). Phonocardiogram signal classification for the detection of heart valve diseases using robust conglomerated models. *Expert Systems with Applications, 221,* 119720. https://doi.org/10.1016/j.eswa.2023.119720

Li, S., Li, F., Tang, S., & Xiong, W. (2020). A review of computer-aided heart sound detection techniques. *BioMed research international*, *2020*(1), 5846191. https://doi.org/10.1155/2020/5846191

Luo, Y., Fu, Z., Ding, Y., Chen, X., & Ding, K. (2024). Phonocardiogram murmur detection based on the mean teacher method. *Sensors, 24*(20), 6646. https://doi.org/10.3390/s24206646

Marocchi, M., Abbott, L., Rong, Y., Nordholm, S., & Dwivedi, G. (2023). Abnormal heart sound classification and model interpretability: A transfer learning approach with deep learning. *Journal of Vascular Diseases, 2*(4), 438–459. https://doi.org/10.3390/jvd2040034

Milani, M. G. M., Abas, P. E., De Silva, L. C., & Nanayakkara, N. D. (2021). Abnormal heart sound classification using phonocardiography signals. *Smart Health, 21,* 100194. https://doi.org/10.1016/j.smhl.2021.100194

Orozco-Reyes, L., Alonso-Arévalo, M. A., García-Canseco, E., Ibarra-Hernández, R. F., & Conte-Galván, R. (2025). A deep-learning approach to heart sound classification based on combined time-frequency representations. *Technologies, 13*(4), 147. https://doi.org/10.3390/technologies13040147

Khan, K. N., Khan, F. A., Abid, A., Olmez, T., Dokur, Z., Khandakar, A., ... & Khan, M. S. (2021). Deep learning based classification of unsegmented phonocardiogram spectrograms leveraging transfer learning. *Physiological measurement*, *42*(9), 095003. https://doi.org/10.1088/1361-6579/ac1d59

Singh, S. A., Devi, N. D., Singh, K. N., Thongam, K., & Majumder, S. (2024). An ensemble-based transfer learning model for predicting the imbalance heart sound signal using spectrogram images. *Multimedia Tools and Applications*, *83*(13), 39923-39942.

Chowdhury, M. T. H. (2022). *Application of Signal Processing and Deep Hybrid Learning in Phonocardiogram and Electrocardiogram Signals to Detect Early Stage Heart Diseases* (Doctoral dissertation, Middle Tennessee State University).

Singh, A., Arora, V., & Singh, M. (2024). Heart sound classification using harmonic and percussive spectral features from phonocardiograms with a deep ANN approach. *Applied Sciences*, *14*(22), 10201. https://doi.org/10.3390/app142210201

Taneja, K., Arora, V., & Verma, K. (2023). Classifying the heart sound signals using textural-based features for an efficient decision support system. *Expert Systems, 40*(6), e13246. https://doi.org/10.1111/exsy.13246

Zhu, B., Zhou, Z., Yu, S., Liang, X., Xie, Y., & Sun, Q. (2024). Review of phonocardiogram signal analysis: insights from the PhysioNet/CinC challenge 2016 database. *Electronics*, *13*(16), 3222. https://doi.org/10.3390/electronics13163222