Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(RESEARCH ARTICLE)

Check for updates

Machine learning based system for early heart disease detection and classification using audio signal processing approach

Ekikereabasi Effiong Nkereuwem *, Godwin Okon Ansa, Uduak Augustine Umoh, Ubong David Essien, Effiong Edet Nkereuwem and Mfoniso Polycarp Asuquo

Department of Computer, Faculty of Physical Science, Akwa Ibom State Univeristy Ikot Akpaden, Mkpat Enin, Akwa Ibom State, Nigeria.

Global Journal of Engineering and Technology Advances, 2024, 21(01), 087–104

Publication history: Received on 19 August 2024; revised on 28 September 2024; accepted on 30 September 2024

Article DOI: https://doi.org/10.30574/gjeta.2024.21.1.0178

Abstract

Cardiovascular diseases (CVDs) remain a leading cause of mortality globally, accounting for approximately 17.9 million deaths annually. Traditional diagnostic methods, though useful, have limitations such as invasiveness, high cost, and delays in detecting early-stage heart conditions. This study presents a novel machine learning-based system for early heart disease detection using audio signal processing, specifically leveraging phonocardiogram (PCG) data. Features were extracted using Mel-Frequency Cepstral Coefficients (MFCCs), Delta MFCCs, and Delta-Delta MFCCs, followed by dimensionality reduction via Principal Component Analysis (PCA). Support Vector Machine (SVM) and XGBoost classifiers were used to analyze the extracted features, and their performance was optimized through an ensemble model using the Moth Flame Optimization (MFO) algorithm. The model was rigorously evaluated using accuracy, precision, recall, and F1-score metrics. The ensemble model achieved an accuracy of 99.13%, precision of 98.94%, recall of 95.05%, and an F1-score of 97.46%. The application of SMOTE for data augmentation significantly improved classification performance, highlighting its effectiveness in addressing class imbalance. The proposed system provides a non-invasive, cost-effective solution for heart disease detection and holds potential for improving diagnostic access, particularly in resource-limited settings.

Keywords: Phonocardiogram; Support Vector Machine; XGBoost; Moth Flame Optimization; SMOTE; Audio signal processing

1. Introduction

Cardiovascular diseases (CVDs) represent a formidable global health challenge, accounting for approximately 17.9 million deaths annually and placing a significant burden on healthcare systems worldwide [20]. The prevalence of heart disease has witnessed an alarming surge in recent decades, reflecting a complex interplay of lifestyle factors, genetic predispositions, and an aging population [16]. This escalating trend underscores the critical need for innovative approaches to early detection and management of cardiovascular conditions. Current methods employed for heart disease detection encompass a multifaceted approach, integrating clinical assessments, advanced imaging techniques, and traditional diagnostic tools. However, these methods are not without limitations, including their reliance on invasive procedures, cost constraints, and occasional lack of sensitivity in early disease stages. The confluence of these challenges has catalysed a transformative shift in healthcare, propelled by the burgeoning applications of artificial intelligence (AI) and machine learning (ML), particularly in cardiovascular health. Machine learning models applied in this field utilize a variety of data sources, including electrocardiograms (ECGs), medical imaging, and, increasingly, audio signals derived from heart sounds. These technologies demonstrate a remarkable ability to detect subtle patterns within

* Corresponding author: Ekikereabasi Nkereuwem

signals, offering a non-invasive and potentially more accessible method of diagnosing and monitoring heart diseases [3,7, 21]

Recent research has shown promising results in this area, with a focus on leveraging various signal processing techniques and machine learning algorithms. Srivastava et al. [18] achieved a classification accuracy of 95% using Mel-frequency cepstral coefficients (MFCCs) and Support Vector Machines (SVM). Their approach involved the application of a Butterworth filter for noise removal, contributing to signal refinement. Rath et al. [14] developed ensemble models combining discrete wavelet transform (DWT) and MFCC features, achieving accuracy, F1 score, and AUC values of 89.53%, 0.9, and 0.95, respectively. Their study addressed the important issue of imbalanced datasets in heart sound analysis, proposing novel ensemble models that outperformed existing systems.

Zeinali and Niaki [21] made significant strides in the diagnosis of heart diseases through the analysis of heart sounds obtained via a stethoscope. They focused on categorizing heart sounds into four groups (S1 to S4), with S1 and S2 representing normal heart sounds and S3 and S4 indicating abnormal sounds associated with specific heart diseases. Their methodology achieved a notable accuracy rate of 87.5% and 95% for multiclass data (3 classes) and 98% for binary classification (normal vs. abnormal) scenarios. Other notable contributions have further advanced the field. Muhammad et al. [10] conducted a comprehensive study focused on early and accurate detection of heart disease using intelligent computational models. Their work showcased the limitations of invasive techniques like angiography and highlighted the effectiveness of non-invasive, intelligent computational approaches. Nabih-Ali et al. [11] proposed an intelligent system using phonocardiogram (PCG) data, achieving an impressive overall accuracy of 97% using an artificial neural network (ANN) classifier. Their study emphasized the potential of combining PCG data with intelligent algorithms for improved healthcare outcomes.

Arora et al. [1] made important contributions to heart sound classification by utilizing digital phonocardiogram (PCG) signals for screening heart conditions. Their study employed the XGBoost method on unsegmented heart sounds, outperforming 18 existing methodologies with a mean score of 92.9 and sensitivity and specificity ratings of 94.5 and 91.3, respectively. This work underscored the importance of developing timely projections for heart health to aid specialists in risk stratification and decision-making. Advancing the field further, presented a comprehensive investigation into the diagnosis of congenital heart disease (CHD), focusing on Ventricular Septal Defects (VSDs) and Arterial Septal Defects (ASDs) as presented in [4]. Their three-stage processing model, incorporating empirical mode decomposition (EMD) for denoising raw PCG signals and feature extraction using one-dimensional local ternary patterns (1D-LTPs) and MFCCs, achieved a remarkable mean accuracy of 95.24% in classifying ASD, VSD, and normal subjects. Building upon these advancements, our research aims to design and implement a machine learning-based system for early heart disease detection and classification using audio signal processing. We seek to address the limitations of current diagnostic methods by leveraging the power of artificial intelligence and the rich information contained in heart sounds. To achieve this, we have formulated the following specific objectives:

To design and implement a detailed system framework using a combination of Mel-frequency cepstral coefficients (MFCC), Delta MFCC, and Delta-delta MFCC features, with principal component analysis (PCA) for dimensionality reduction.

To utilize and compare the performance of two machine learning classifiers, Support Vector Machine (SVM) and XGBoost, ensembled by Moth Flame Optimization (MFO).

To evaluate the system's performance using metrics such as accuracy, precision, ROC-AUC, and confusion matrix.

These objectives guide our novel approach, which combines advanced feature extraction techniques with state-of-the-art machine learning classifiers. Our feature extraction approach incorporates Mel-frequency cepstral coefficients (MFCC), Delta MFCC, and Delta-delta MFCC features, with principal component analysis (PCA) for dimensionality reduction. This approach builds upon and extends the work of [18], who successfully used MFCC as their primary feature for heart sound analysis. While [18] demonstrated the effectiveness of MFCC in this context, our study expands on this by incorporating Delta and Delta-delta MFCC features to capture additional temporal dynamics of the heart sounds. The inclusion of Delta and Delta-delta MFCC features is inspired by their success in speech and audio processing tasks, where they have been shown to provide valuable information about the rate of change of spectral characteristics. By incorporating these additional features, we aim to capture more nuanced aspects of heart sounds that may be indicative of cardiac abnormalities. The use of PCA for dimensionality reduction is motivated by the work of [17], who successfully applied this technique in heart disease diagnosis to enhance model efficiency and performance.

Our approach also innovates in the classification stage, where we utilize and compare the performance of two machine learning classifiers: Support Vector Machine (SVM) and XGBoost. The choice of SVM is informed by its effective use in heart sound classification, as demonstrated by [18] and [4]. The inclusion of XGBoost is motivated by its strong performance in related studies, such as the work of [1]. To potentially enhance the overall performance of the system, we introduce a novel element of ensembling these classifiers using Moth Flame Optimization (MFO), inspired by the success of ensemble methods in addressing complex classification tasks, as demonstrated by [14] and [6]. This study aims to contribute to the early identification of cardiac abnormalities, potentially leading to more timely interventions and improved patient outcomes. The cost-effectiveness and non-invasive nature of the proposed system could democratize access to advanced cardiac diagnostics, particularly in resource-limited settings where traditional methods may be prohibitively expensive or unavailable [4].

The rest of this paper is organized as follows. Section 2 details of the proposed method, including the PhysioNet/CinC Challenge 2022 heart sound database, MFCC, Delta MFCC and Detla-Delta MFCC feature extraction and selection with PCA, SVM and Xgboost Classifier which then ensemble with Moth Flame Optimization. Section 3 presents experimental results. Sections 3 and 4 give some discussions and conclusions, respectively and conclusions, respectively.

2. Proposed method

The study describes a novel approach for distinguishing between normal and abnormal heart sound signals using nonlinear dynamics of phonocardiogram (PCG) systems. The method involves multiple stages: data acquisition, preprocessing (noise reduction and normalization), feature extraction (using MFCC, Delta MFCC, and Delta-Delta MFCC), data augmentation (using SMOTE), feature selection (using PCA), and classification. Two classifiers, Support Vector Machines (SVM) and XGBoost, are employed, with their outputs combined through an ensemble approach optimized by Moth Flame Optimization (MFO). The process is visually represented in Fig. 1, showcasing the interconnections between each stage. This comprehensive approach aims to capture subtle differences in PCG system dynamics, advancing automated cardiac auscultation and anomaly detection. The system is implemented to integrate efficiently with telemedicine systems such as the work of [12].

2.1. Data Collection

The CirCor DigiScope Phonocardiogram Dataset represents a significant contribution to the fields of pediatric cardiology and digital auscultation. This comprehensive resource comprises 5,272 heart sound recordings obtained from 1,568 subjects aged 0-21 years, collected from four primary cardiac auscultation sites. The recordings exhibit a duration range of 4.8 to 80.4 seconds, amassing over 33.5 hours of high-fidelity phonocardiogram data. Each recording is accompanied by meticulous annotations detailing cardiac murmurs and is supplemented with pertinent subject metadata. This extensive dataset, derived from two large-scale screening campaigns conducted in Brazil between 2014 and 2015, serves as the cornerstone for the George B. Moody PhysioNet Challenge 2022, which focuses on heart murmur detection [13]. The study under consideration utilized a subset of 3,163 records from the aforementioned dataset, demonstrating a near-equal distribution between normal (n=1,632) and abnormal (n=1,531) recordings. To mitigate potential class imbalance and enhance the robustness of subsequent analyses, the Synthetic Minority Over-sampling Technique (SMOTE) was employed.

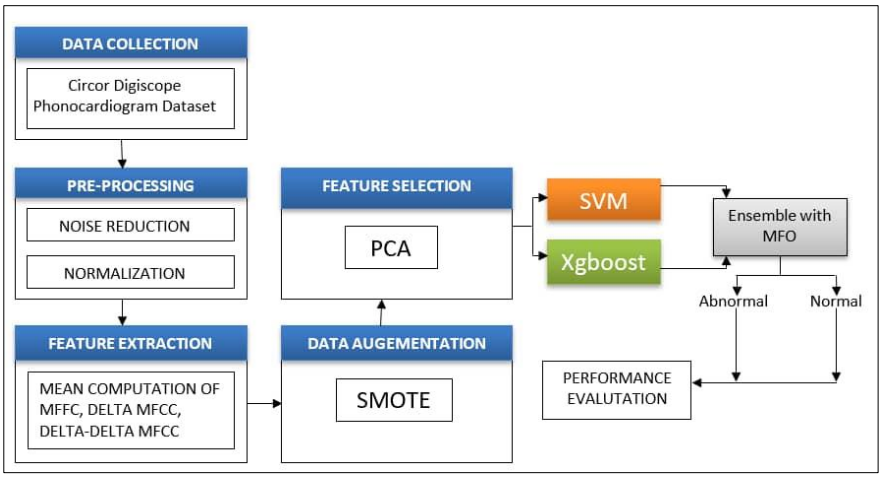

Figure 1 System architecture of the proposed system

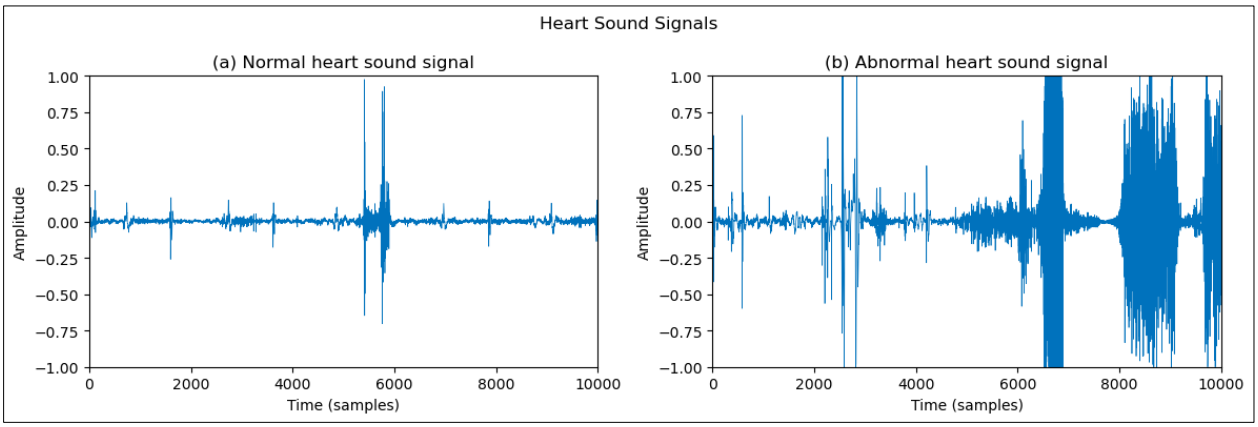

Figure 2 The waveforms of heart sound signals

The composition of the dataset is detailed in Table 1, which presents the distribution of heart sound recordings across various auscultation locations and their variants. This data augmentation strategy resulted in an expanded dataset of 50,000 recordings, ensuring an equitable distribution. The recordings were acquired from various auscultation sites, including the Aortic, Mitral, Pulmonary, and Tricuspid valves, as well as additional locations, with multiple variants for numerous sites. This diverse sampling approach facilitates a comprehensive representation of heart sounds across different anatomical locations, potentially encompassing a wide spectrum of normal and pathological cardiac acoustic phenomena, as illustrated in Fig. 2, which depicts examples of normal and abnormal heart sound waveform signals.

Table 1 Presents a comprehensive overview of the heart sound recording distribution in the CirCor DigiScope Phonocardiogram Dataset

Valve Types	**Original Dataset**		**Dataset After SMOTE**	
	Normal	**Abnormal**	**Normal**	**Abnormal**
AV	394	393	6287	6151
AV_1	2	4	57	40
AV_2	2	4	51	46
AV_3	1	0	3	13
MV	423	426	6637	6811
MV_1	3	3	49	48
MV_2	3	3	50	47
PV	411	351	6078	5944
PV_1	0	2	20	13
PV_2	0	2	19	14
Phc	0	2	17	15
Phc_1	0	1	8	8
Phc_2	0	1	10	6
TV	383	337	5666	5700
TV_1	5	1	56	40
TV_2	5	1	62	34
Total	3163		50,000	

Ratio of abnormal to normal	1.066	0.994

2.2. Data Preprocessing

The data preprocessing phase is an important phase in the research for the model's validity and performance [2]. The data preprocessing for phonocardiogram (PCG) signals encompasses two crucial steps: noise reduction and normalization.

Noise reduction: is implemented through a Butterworth bandpass filter, which is configured to transmit frequencies between 15 Hz and 130 Hz, effectively isolating the relevant cardiac acoustic information. The lower bound of 15 Hz helps eliminate unwanted low-frequency components, while the upper limit of 130 Hz ensures the retention of primary spectral components of heart sounds. This meticulous preprocessing approach enhances the quality and standardization of the audio data, making it more suitable for subsequent analysis and interpretation in cardiac diagnostics [9].

2.2.1. Butterworth Bandpass Filter

The design of a Butterworth bandpass filter is based on the transfer function in the s-domain. For an nth-order Butterworth filter, the transfer function $H(s)$ is expressed as:

$$H(s) = K \cdot \frac{s^n}{[(s^2 + \omega_1^2)(s^2 + \omega_2^2)]^{n/2}}$$

Where:

K represents the filter gain.
n is the order of the filter (in this study, $n = 5$)
$\omega_1 = 2\pi f_1$ where f_1 is the lower cutoff frequency (15 Hz).
$\omega_2 = 2\pi f_2$ where f_2 is the is the upper cutoff frequency (130 Hz).

2.2.2. Digital Filter

To implement the filter in digital signal processing, the analogy transfer function is transformed to the digital domain using the bilinear transform:

$$s = \frac{2}{T} \cdot \frac{z-1}{z+1}$$

where T is the sampling period, given by $T = \frac{1}{f_s}$ and f_s being the sampling frequency.

2.2.3. Normalized Frequencies

The cutoff frequencies are normalized with respect to the Nyquist frequency $\frac{f_s}{2}$ as follows:

$$\omega_{\text{low}} = \frac{2\pi \cdot f_s}{2f_1} = \frac{2\pi \cdot f_s}{2 \cdot 15}$$

$$\omega_{\text{high}} = \frac{2\pi \cdot f_s}{2f_2} = \frac{2\pi \cdot f_s}{2 \cdot 130}$$

These normalized frequencies are utilized in the design of the digital filter.

2.2.4. Filter Application

The bandpass filter is applied to the input signal $x[n]$ through a difference equation. The output signal $y[n]$ is computed as:

$$y[n] = b_0 \cdot x[n] + b_1 \cdot x[n-1] + \cdots + b_5 \cdot x[n-5] - a_1 \cdot y[n-1] - a_2 \cdot y[n-2] - \cdots - a_5 \cdot y[n-5]$$

where b_i and a_i represent the filter coefficients, which are derived from the digital transfer function.

2.2.5. Frequency Response

The frequency response of the Butterworth bandpass filter is characterized by its magnitude response $|H(j\omega)|$ which is given by:

$$|H(j\omega)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_0}\right)^{2n}}}$$

Where $\omega_0 = \sqrt{\omega_1 \cdot \omega_2}$ is the centre frequency of the passband.

Normalization: standardizes data to a common range, ensuring comparable magnitudes and distributions for reliable analysis. For phonocardiogram signals, the process involves peak normalization to scale the signal to a maximum absolute value of 1.0, followed by amplification with a gain factor of 3.0. This methodology standardizes signal amplitudes across recordings and optimizes feature extraction. The approach enhances signal quality for both visual and auditory analysis [8].

2.2.6. Signal Normalization

Post-filtering, the signal is normalized to ensure that the maximum amplitude of the output signal does not exceed 1. This normalization is defined as:

$$y_{\text{normalized}}[n] = \frac{|y[n]|}{\max(|y[n]|)}$$

2.2.7. Amplification

Finally, the normalized signal is amplified by a constant gain factor gain. The amplified signal $y_{\text{amplified}}[n]$ is computed as:

$$y_{\text{amplified}}[n] = \text{gain} \cdot y_{\text{normalized}}[n]$$

where the gain factor is set to 3.0 in this study.

2.3. Feature extraction

Feature extraction in phonocardiogram (PCG) analysis transforms raw heart sound recordings into meaningful representations for machine learning tasks, focusing on Mel-Frequency Cepstral Coefficients (MFCCs) and their derivatives. These features effectively capture spectral and dynamic characteristics of audio signals, distinguishing between normal and abnormal heart sounds. The extraction process involves optimized parameters like 13 MFCCs, a 25ms frame size, and a 10ms hop length, balancing dimensionality and information retention. Features are averaged across frames to produce fixed-length vectors suitable for classification tasks. Fig. 2 visualizes the acoustic characteristics of normal and abnormal heart sounds across different feature representations, illustrating differences in spectro-temporal patterns.

2.3.1. MFCC Computation

Given a PCG signal $x[n]$ the MFCC computation involves the following steps:

- Short-Time Fourier Transform (STFT):

$$X[k, m] = \sum_{n=0}^{N-1} x[n + mH]w[n]e^{-j\frac{2\pi kn}{N}}$$

Where $w[n]$ is a Hamming window of length N, H is the hop length, and m is the frame index.

- Mel-filterbank Energy

$$E[l,m] = \sum_{k=0}^{N/2} \lvert X[k,m] \rvert^2 H_l[k]$$

Where $H_l[k]$ represents the $l-$ th triangular mel-filter.

- Logarithmic Compression:

$$S[l,m] = \log(E[l,m])$$

- Discrete Cosine Transform (DCT):

$$c[i,m] = \sum_{l=0}^{L-1} S[l,m] \cos\left(\frac{L\pi i(l+0.5)}{L}\right)$$

where L is the number of mel-filters and $i = 0, \dots, C-1$, with C being the number of cepstral coefficients.

2.3.2. Delta and Delta-Delta Computation

The first and second-order temporal derivatives of the MFCCs are computed as:

- Delta coefficients

$$\Delta c[i,m] = \frac{2\sum_{\theta=1}^{\Theta} \theta^2 \sum_{\theta=1}^{\Theta} \theta(c[i,m+\theta] - c[i,m-\theta])}{\sum_{\theta=1}^{\Theta} \theta^2}$$

- Delta-Delta coefficients

$$\Delta^2 c[i,m] = \frac{2\sum_{\theta=1}^{\Theta} \theta^2 \sum_{\theta=1}^{\Theta} \theta(\Delta c[i,m+\theta] - \Delta c[i,m-\theta])}{\sum_{\theta=1}^{\Theta} \theta^2}$$

2.3.3. Feature Aggregation

The final feature vector for each PCG signal is obtained by computing the mean of each MFCC and its derivatives across all frames:

$$\mu_c[i] = \frac{1}{M} \sum_{m=0}^{M-1} c[i,m]$$

$$\mu_{\Delta c}[i] = \frac{1}{M} \sum_{m=0}^{M-1} \Delta c[i,m]$$

$$\mu_{\Delta^2 c}[i] = \frac{1}{M} \sum_{m=0}^{M-1} \Delta^2 c[i,m]$$

where M is the total number of frames.

The complete feature vector f for a PCG signal is then formed by concatenating these mean values:

$$f = \left[\mu_c[0], \dots, \mu_c[C-1], \mu_{\Delta c}[0], \dots, \mu_{\Delta c}[C-1], \mu_{\Delta^2 c}[0], \dots, \mu_{\Delta^2 c}[C-1]\right]$$

2.3.4. Frequency Warping

The mel-scale warping used in the MFCC computation is defined as:

$$\text{mel}(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

This warping approximates the frequency perception of the human auditory system, which is especially useful for assessing heart sounds that are normally auscultated by medical practitioners.

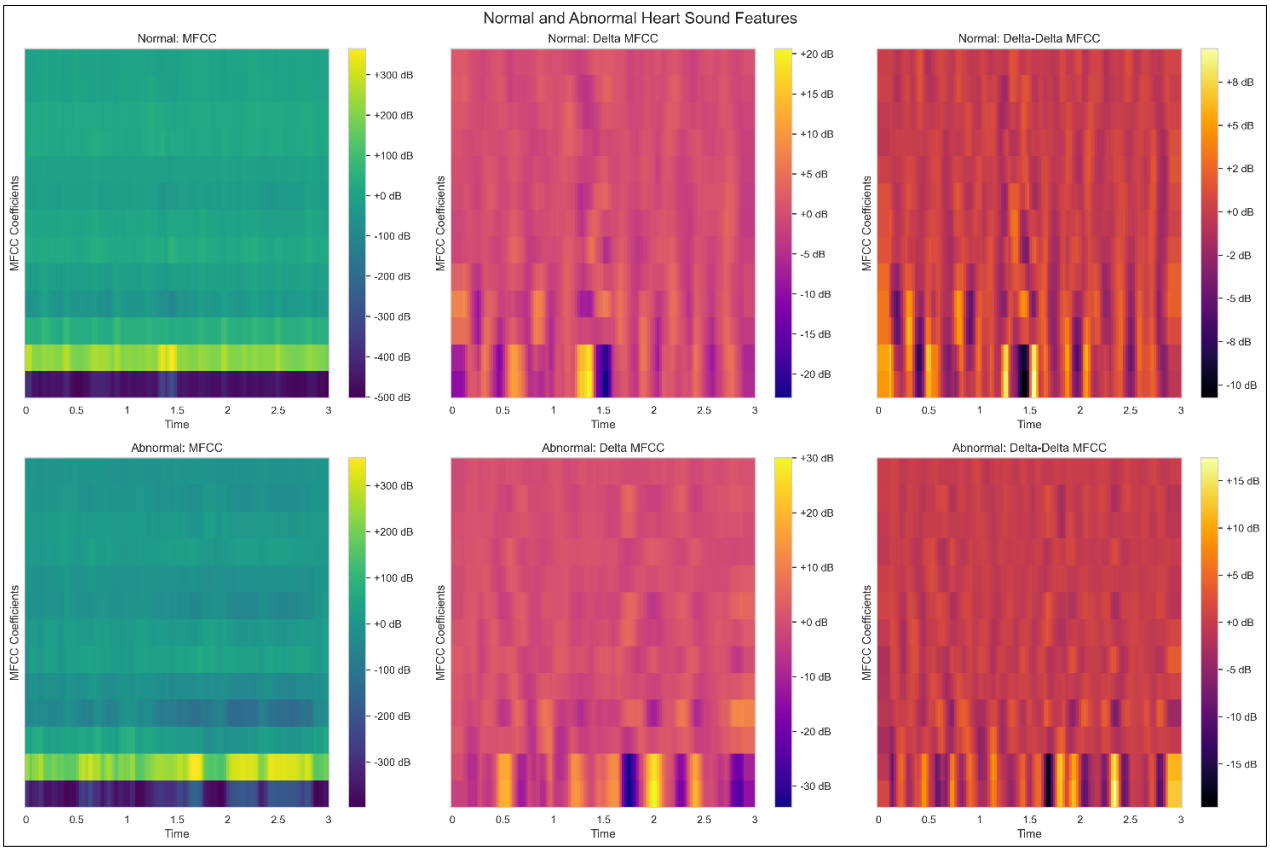

Figure 3 Acoustic Features for Normal and Abnormal Heart Sounds

2.4. Data Augmentation

Class imbalance in phonocardiogram (PCG) classification was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) to enhance model robustness. As shown in Table 1, the original dataset of 3,163 samples exhibited a slight imbalance with an abnormal to normal ratio of 1.066. SMOTE was applied to the training data using the imbalanced-learn library, generating synthetic samples to balance classes across all valve types. This approach is theoretically justified for its ability to enhance model performance, improve generalization, and preserve essential signal characteristics in PCG analysis [15]. Post-SMOTE application, the dataset expanded to 50,000 samples with a near-perfect balance ratio of 0.994, ensuring equal representation of normal and abnormal heart sounds for unbiased model training.

2.4.1. Data Preparation

Let $X \in R^{n \times d}$ be the feature matrix of n samples and d features, $y \in \{0,1\}^n$ and be the corresponding binary labels. The dataset is defined as:

$$D = \{(x_i, y_i) \mid x_i \in X, y_i \in y, i = 1, \ldots, n\}$$

2.4.2. Feature Standardization

Each feature is standardized using:

$$z_{ij} = \sigma_j(x_{ij} - \mu_j)$$

Where x_{ij} is the $j-$th feature of the $i-$th sample, and μ_j and σ_j are the mean and standard deviation of the feature, j − th respectively.

2.4.3. SMOTE Algorithm

Let $S_{\min} = \{(x_i, y_i) \in D \mid y_i = 1\}$ be the set of minority class samples. For each $x_i \in S_{\min}$

- Find the k-nearest neighbours of x_i in $S_{\min}$, denoted as $NN_k(x_i)$
- Randomly select a neighbour x_j from $NN_k(x_i)$
- Generate a synthetic sample x_{new}

$$x_{\text{new}} = x_i + \lambda(x_j - x_i)$$

Where $\lambda \in [0,1]$ is a random number

2.4.4. Iterative SMOTE Application

Let n_{target} be the desired total number of samples. The iterative SMOTE process is defined as:

$$\text{while } |D| < n_{\text{target}}:$$

$$S_{\text{combined}} = D$$

$$S_{\text{augmented}} = \text{SMOTE}(S_{\text{combined}})$$

$$n_{\text{additional}} = \min(n_{\text{target}} - |D|, |S_{\text{augmented}}| - |S_{\text{combined}}|)$$

$$D = D \cup \{\text{first } n_{\text{additional}} \text{ samples from } S_{\text{augmented}}\}$$

2.4.5. Nearest Neighbour Computation

The k-nearest neighbors are determined using Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^{d} (x_{il} - x_{jl})^2}$$

2.4.6. Sampling Strategy

The number of synthetic samples to generate for the minority class is determined by:

$$n_{\text{synthetic}} = \lceil \alpha \times (n_{\text{majority}} - n_{\text{minority}}) \rceil$$

where n_{majority} and n_{minority} are the number of samples in the majority and minority classes, respectively, and α is the sampling ratio (default $\alpha = 1$ for balanced classes).

2.4.7. Dataset Expansion

The final augmented dataset D_{aug} is constructed as:

$$D_{\text{aug}} = D \cup S_{\text{synthetic}} \quad (7)$$

Where $S_{\text{synthetic}}$ is the set of all generated synthetic samples.

2.5. Feature Selection

Principal Component Analysis (PCA) is employed in phonocardiogram (PCG) analysis to address challenges posed by high-dimensional features, reducing the feature set from 39 to 12 components. Table 2 presents a comprehensive statistical analysis of the 12 most important features obtained through PCA, including mean, standard deviation, skewness, and kurtosis for each principal component. The features are ranked by standard deviation, indicating their relative importance in explaining dataset variance, while correlation coefficients and p-values for normal and abnormal cases provide insights into each component's relationship with the classification outcome. This analysis aids in understanding which principal components contribute most significantly to distinguishing between normal and abnormal heart sounds, potentially enhancing early heart disease detection capabilities.

2.5.1. Data Representation

Let $X \in R^{n \times d}$ be the feature matrix of n samples and d features. Each row $x_i \in R^d$ represents a single PCG sample in the feature space.

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

2.5.2. Centering the Data

The data is centered by subtracting the mean of each feature:

$$X_c = X - \mu$$

Where $\mu \in R^d$ is the mean vector of with X

$$\mu_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}.$$

2.5.3. Covariance Matrix Computation

The covariance matrix Σ is computed as:

$$\Sigma = \frac{1}{n-1} X_c^T X_c$$

Where $\Sigma \in R^{d \times d}$ is a symmetric positive semi-definite matrix.

2.5.4. Eigendecomposition

PCA involves the eigendecomposition of the covariance matrix:

$$\Sigma = V \Lambda V^T$$

Where $V \in R^{d \times d}$ is the matrix of eigenvectors, and $\Lambda \in R^{d \times d}$ is a diagonal matrix of corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0$

2.5.5. Principal Component Selection

The first k eigenvectors corresponding to the k largest eigenvalues are selected to form the transformation matrix $W \in R^{d \times k}$:

$$W = [\, v_1 \mid v_2 \mid \ldots \mid v_k \,]$$

Where v_i is the i −th column of V

2.5.6. Dimensionality Reduction

The original data X is projected onto the k −dimensional subspace:

$$Y = X_c W$$

Where $Y \in R^{n \times k}$ is the reduced-dimensional representation of X

2.5.7. Variance Explained

The proportion of variance explained by the first k principal components is given by:

$$\rho_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{d} \lambda_i}$$

2.5.8. Reconstruction

An approximation of the original data $\hat{X}$ can be obtained by:

$$\hat{X} = YW^T + \mu$$

Table 2 Statistical Analysis of Principal Components after Feature Extraction

Important Feature	Mean	Standard Deviation	Skewness	Kurtosis	Normal Correlation	Abnormal Correlation	p-value
Feature_1	1.73E-08	3.07457	1.00779	48.4592	0.029148	0.028991	0.479
Feature_2	1.94E-08	2.705915	-2.2176	13.3407	0.116442	0.158887	<0.001
Feature_3	-1.9E-08	2.386499	2.042394	32.60357	-0.20404	-0.23548	<0.001
Feature_4	-9.8E-09	1.979426	2.778216	31.5085	-0.02198	-0.05588	<0.001
Feature_5	-9.4E-09	1.903309	0.017382	13.44376	0.101153	0.071927	<0.001
Feature_6	-8.5E-09	1.621159	1.515172	92.10209	0.096498	0.058554	<0.001
Feature_7	-1.1E-08	1.359701	0.108716	0.734679	-0.062	-0.0764	<0.001
Feature_8	3.77E-09	0.968578	2.713994	83.99344	0.008183	-0.04407	<0.001
Feature_9	-2.4E-09	0.885049	-1.54659	52.10939	0.044081	0.067078	0.001
Feature_10	-9.5E-11	0.861617	-6.05595	134.0834	-0.05401	0.020494	<0.001
Feature_11	3.3E-09	0.642755	-1.59392	14.79416	-0.03234	0.038926	<0.001
Feature_12	-2.2E-09	0.544546	-0.41278	25.29352	-0.01363	-0.00071	<0.001

2.6. Support Vector Machine (SVM)

The Support Vector Machine (SVM) classification model for heart disease detection functions by identifying an optimal hyperplane that maximizes the margin between classes, utilizing kernel functions such as Radial Basis Function (RBF) and polynomial to handle non-linear decision boundaries[19]. Hyperparameter optimization, including parameters like the regularization parameter C, kernel coefficient γ, and kernel type, is performed using a grid search strategy with cross-validation to ensure robust performance across various data split.

2.6.1. SVM Formulation

Given a training dataset $\{(x_i, y_i)\}_{i=1}^{n}$, where $x_i \in R^d$ are the input features and $y_i \in \{-1,1\}$ are the corresponding labels, the SVM aims to find the optimal hyperplane that maximizes the margin between classes.

The primal form of the SVM optimization problem is:

$$\text{minimize}_{w,b,\xi} \left(\frac{1}{2}|w|^2 + C \sum_{i=1}^{n} \xi_i \right)$$

$$\text{subject to} \quad y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \quad (1)$$

where w is the weight vector, b is the bias term, ξ_i are slack variables, $C > 0$ is the regularization parameter, and $\phi(\cdot)$ is a feature mapping function.

2.6.2. Kernel Functions

The kernel trick allows for implicit feature mapping through the kernel function $K(x_i, x_j) = \phi(x_i)^T\phi(x_j)$

We consider two kernel functions:

- Radial Basis Function (RBF) Kernel

$$K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2) \quad (2)$$

- Polynomial Kernel

$$K(x_i, x_j) = (\gamma\langle x_i, x_j\rangle + r)^d \quad (3)$$

where $\gamma > 0$ is the kernel coefficient, $r \geq 0$ is the intercept term, $d \in N$ and is the polynomial degree

2.6.3. Dual Formulation

The dual form of the SVM optimization problem is:

$$\text{maximize } \alpha\left(\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i, x_j)\right)$$

$$\text{subject to } 0 \leq \alpha_i \leq C, \forall i \text{ and } \sum_{i=1}^{n}\alpha_i y_i = 0$$

where α_i are the Lagrange multipliers.

2.6.4. Decision Function

The decision function for classification is given by:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n}\alpha_i y_i K(x_i, x) + b\right)$$

2.6.5. Class Weighting

To address potential class imbalance, we introduce class weights ω^+ and ω^- for the positive and negative classes, respectively. The weighted SVM formulation becomes:

$$\text{minimize}_{w,b,\xi}\left(\frac{1}{2}|w|^2 + C\left(\sum_{i:y_i=1}\xi_i + \sum_{i:y_i=-1}\xi_i\right)\right)$$

$$\text{subject to } y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i$$

2.6.6. Hyperparameter Optimization

We define a hyperparameter space $\Theta = \{C, \gamma, \text{kernel}, \text{class\_weight}\}$ and seek the optimal configuration that maximizes the cross-validation accuracy:

$$\theta^* = \arg\max_{\theta \in \Theta} \left(\sum_{k=1}^{K} Acc_k(\theta) \right)$$

Where K is the number of cross-validation folds and $Acc_k(\theta)$ is the accuracy on the $k-th$ fold using hyperparameters θ

2.7. XGBoost

XGBoost, an advanced implementation of gradient boosting machines, combines multiple weak learners to create a robust predictive model for heart disease detection using audio signal processing [5]. The implementation leverages advanced computational techniques and carefully tuned parameters to optimize performance and efficiency. XGBoost optimizes a regularized objective function, employing a level-wise tree growth strategy for parallelization and improved accuracy in heart disease detection.

2.7.1. Ensemble Model

The XGBoost model predicts the output $\hat{y}_i$ for the $i-th$ instance as an ensemble of K trees:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i)$$

Where $f_k \in \mathcal{F}$ and $\mathcal{F}$ is the space of all possible regression trees.

2.7.2. Objective Function

The model is trained by minimizing the regularized objective:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where l is the loss function, y_i is the true label, and Ω is the regularization term

2.7.3. Additive Training

The model is trained additively. At step t we optimize:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \widehat{y_i^{(t-1)}} + f_t(x_i) \right) + \Omega(f_t)$$

Where $(\widehat{y_i^{(t-1)}}$ is the prediction at step $t-1$.

2.7.4. Taylor Expansion of Loss Function

Applying the second-order Taylor expansion to the loss function:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[l\left(y_i, \widehat{y_i^{(t-1)}} \right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

Where g_i and h_i are the first and second-order gradients of the loss function with respect to $\widehat{y_i^{(t-1)}}$

2.7.5. Tree Structure Score

For a tree structure (x) , the quality score is defined as

$$L_{Spilt} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} \right] - \gamma$$

Where I_L and I_R are the instance sets of left and right nodes after the split, λ is the L2 regularization term, and γ is the minimum loss reduction required for a split.

2.7.6. Regularization Term

The regularization term Ω is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

Where T is the number of leaves in the tree, w_j is the score on the j −th leaf, γ is the complexity control parameter, and λ is the L2 regularization term.

2.7.7. Multiclass Classification

For multiclass classification with M classes, we use the softmax objective:

$$l(y_i, \hat{y_i}) = -\sum_{m=1}^{M} y_{im} \log(p_{im})$$

Where y_{im} is 1 if the i −th instance belongs to the m −th class and 0 otherwise, and p_{im}

is the predicted probability of the i −th instance belonging to the m −th class.

2.7.8. Feature Importance

The importance score for a feature is calculated as:

$$\text{Score}(F) = \sum_{k=1}^{K} \sum_{j \in J} I_j^2$$

where J is the set of nodes in k tree that split on feature F_i and I_j is the improvement in the loss function resulting from this split.

2.8. Ensemble classification model

The ensemble classification model presented in this study combines Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost) classifiers, with their weights optimized using the Moth Flame Optimization (MFO) algorithm. The ensemble employs a weighted average method, where the final prediction is a weighted sum of the individual model predictions, with the objective function incorporating both accuracy and a regularization penalty to promote balanced weighting.

2.8.1. Ensemble Classification Model

Let M_1 and M_2 be two base classifiers, in this case, Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost), respectively. The ensemble model E is defined as a weighted combination of these classifiers:

$$E(x) = w_1 \cdot M_1(x) + w_2 \cdot M_2(x)$$

Where x is the input feature vector, w_1 and w_2 are the weights assigned to each classifier such that $w_1 + w_2 = 1$, and $E(x)$ is the ensemble prediction.

2.8.2. Objective Function

The objective function $f(w)$ to be minimized is defined as:

$$f(w) = -A\big(E(X)\big) + \lambda R(w)$$

2.8.3. Moth Flame Optimization (MFO)

Let S be the search space defined by the bounds of the weight parameters. The MFO algorithm can be formalized as follows:

Initialization

Generate a population of n moths $M = \{m_1, m_2, \dots, m_n\}$ where each $m_i \in S$. Initialize a set of n flames $F = \{f_1, f_2, \dots, f_n\}$ where initially $f_i = m_i$

2.8.3.2 Iteration Process

For each iteration $t = 1, 2, \dots, T$

For each moth m_i

$$m_i^{(t+1)} = m_i^{(t)} + \alpha_t \cdot \frac{\left(f_{\text{best}} - m_i^{(t)}\right)}{d} + \beta_t \cdot r \quad (3)$$

Evaluate $f\left(m_i^{(t+1)}\right)$ for each updated moth position. Update F by selecting the n best positions among $M \cup F$

Update exploration factor:

$$\alpha_{(t+1)} = \gamma \cdot \alpha_t$$

where γ is a decay factor. The algorithm terminates after T iterations, returning the best solution found.

3. Experimental result

The experiments were conducted using a high-performance computing environment provided by Kaggle, featuring dual NVIDIA Tesla T4 GPUs with 16 GB GDDR6 memory and 2,560 CUDA cores each. This configuration enabled efficient parallel processing for data augmentation, feature extraction, and model training tasks, utilizing Python 3.8 with scikit-learn 0.24.2 and XGBoost 1.4.2 libraries. GPU acceleration was employed for both Support Vector Machine (SVM) and XGBoost implementations, facilitating rapid prototyping and iteration of models, particularly beneficial for computationally intensive processes such as SMOTE and SVM hyperparameter optimization.

We assign feature vector sequences for all the normal and abnormal heart sound signals in the PhysioNet/CinC Challenge 2022 heart sound database. According to the method described in Sect. 2.3 and 2.4, we extract features and applied feature selection (PCA), which means the input feature vector $x_i = [MFCC_i^T, Delta_i^T, DeltaDelta_i^T]^T$. Model performance was rigorously evaluated using 5-fold stratified cross-validation to ensure robust and generalizable results. The primary evaluation metrics included accuracy, precision, recall, and F1-score. Additionally, we monitored training and inference times to assess the computational efficiency of our proposed method. The performance metrics of our individual classifiers (SVM and XGBoost) and the ensemble model optimized with Moth Flame Optimization (MFO) on the test set are presented in Table 3.

Table 3 Performance Comparison of Classifiers

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.9893	0.9926	0.9864	0.9895
Xgboost	0.9597	0.9607	0.9601	0.9604
Ensemble	0.9913	0.9894	0.9505	0.9746

The SVM model demonstrated excellent performance, achieving an accuracy of 98.93% on the test set. This high accuracy is further substantiated by impressive precision (99.26%), recall (98.64%), and F1-score (98.95%). The optimal hyperparameters for the SVM model were found to be C=100, class_weight='balanced', gamma=0.1, and kernel='rbf'. The XGBoost model achieved a solid accuracy of 95.97% on the test set, with balanced precision (96.07%),

recall (96.01%), and F1-score (96.04%). The ensemble model, combining SVM and XGBoost, showed superior results with the highest overall accuracy of 99.13% on the test set. This ensemble approach demonstrated high precision (98.94%), recall (95.05%), and F1-score (97.46%).

Table 4 Effect of Data Augmentation on Ensemble Model Performance

Metric	Without SMOTE	With SMOTE
Accuracy	0.5753	0.9913
Precision	0.5628	0.9894
Recall	0.5468	0.9505
F1-Score	0.5547	0.9746

The application of SMOTE led to a dramatic transformation in the model's performance. Accuracy increased by 41.6 percentage points to 99.13%, while precision improved to 98.94%, indicating a significant reduction in false positives. The recall rose to 95.05%, demonstrating a substantial enhancement in the model's capacity to identify positive cases.

4. Discussion

The ensemble model's superior performance can be attributed to its ability to combine the strengths of both SVM and XGBoost. The SVM's capability to find complex decision boundaries in high-dimensional space complements XGBoost's strength in handling non-linear relationships and feature interactions.

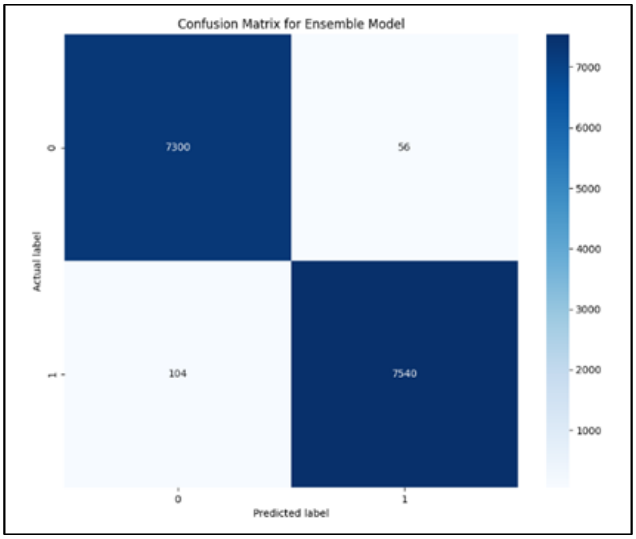

Figure 4 Confusion Matrix for Ensemble Model

The confusion matrix in Fig. 4 provides a detailed breakdown of the ensemble model's performance. With 7300 true negatives and 7540 true positives, the model demonstrates high accuracy in classifying both negative and positive cases. The relatively low number of false positives (56) and false negatives (104) further underscores the model's reliability. This visual representation reinforces the high precision and recall values reported in Table 4. The high precision of the ensemble model (98.94%) is particularly valuable in clinical settings where minimizing false positives is crucial to avoid unnecessary stress and further testing for patients. The confusion matrix visually confirms this, showing only 56 false positives out of 7356 predicted positive cases. However, the slightly lower recall compared to the SVM model alone suggests that it might miss some positive cases, a trade-off that needs careful consideration in a medical context.

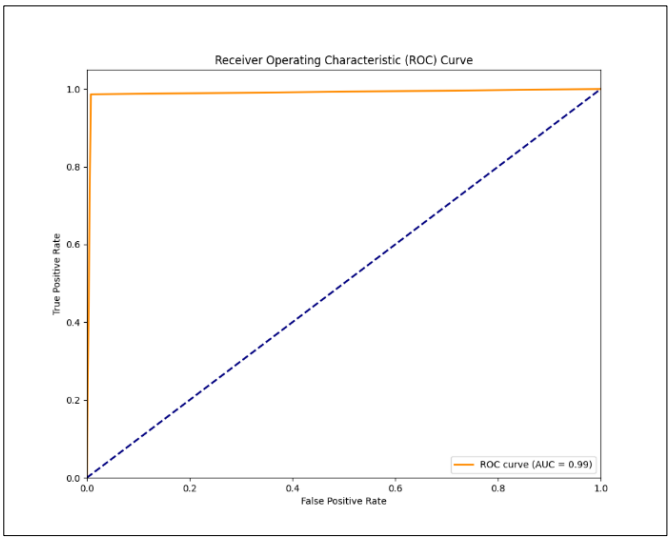

Figure 5 Receiver Operating Characteristic (ROC) Curve

The Receiver Operating Characteristic (ROC) curve (Fig. 5) further validates the ensemble model's exceptional performance. The curve hugs the top-left corner of the plot, indicating a near-perfect classification ability. The reported Area Under the Curve (AUC) of 0.99 quantifies this outstanding performance, where 1.0 represents a perfect classifier. This high AUC value suggests that the model maintains its high performance across various classification thresholds, demonstrating its robustness and reliability. The stark contrast in performance metrics before and after applying SMOTE underscores the critical importance of appropriate data preprocessing in machine learning, particularly for medical applications. Even with a relatively mild class imbalance, the right preprocessing technique can have a profound impact on model performance. The dramatic improvement in accuracy from 57.53% to 99.13% after applying SMOTE (as shown in Table 4) is likely reflected in the highly favourable confusion matrix and ROC curve we observe.

Future research will focus on validating the ensemble model's performance across diverse datasets and clinical settings to ensure generalizability. Exploration of additional ensemble techniques or incorporation of deep learning models could potentially enhance recall while maintaining high precision. In-depth analysis of misclassified cases may provide insights for refining the feature engineering process and improving overall model performance. Furthermore, optimizing the model for specific clinical needs, such as minimizing false negatives in certain medical contexts, could be achieved through threshold adjustment or cost-sensitive learning techniques.

5. Conclusion

This research demonstrates the potential of machine learning, particularly through the combination of SVM and XGBoost classifiers, for the accurate and early detection of heart disease using audio signals. By employing MFCC-based features and leveraging data augmentation through SMOTE, the study effectively addressed class imbalance issues, improving model accuracy and generalizability. The ensemble model, optimized using MFO, achieved superior results in terms of accuracy and precision, proving its efficacy in distinguishing between normal and abnormal heart sounds. This approach offers a non-invasive, cost-effective alternative to traditional diagnostic methods, particularly in resource-constrained environments. Future work should focus on external validation, refining feature extraction methods, and exploring other ensemble techniques to enhance model robustness and further reduce false positives in clinical settings.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Arora V, Leekha R, Singh R, Chana I. Heart sound classification using machine learning and phonocardiogram. Mod Phys Lett B. 2019;33(26):1950321.

[2] Asuquo M, Umoren I. A Hybrid Machine Learning Model for Clustering and Prediction of Closing Price of Cryptocurrency. Int J Netw Commun Res. 2024;8(1):1-22.

[3] Attia ZI, Kapa S, Yao X, Lopez-Jimenez F, Mohan TL, Pellikka PA. Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. J Cardiovasc Electrophysiol. 2019;30(5):668-74.

[4] Aziz S, Khan MU, Alhaisoni M, Akram T, Altaf M. Phonocardiogram Signal Processing for Automatic Diagnosis of Congenital Heart Disorders through Fusion of Temporal and Cepstral Features. Sensors (Basel). 2020;20(13):3790.

[5] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13-17; San Francisco, CA, USA. New York: ACM; 2016. p. 785-94.

[6] Essien UD, Ansa GO, Akpnobong A. EnsembleForge: A Comprehensive Framework for Simplified Training and Deployment of Stacked Ensemble Models in Classification Tasks. J Inf Syst Inform. 2024;6(1):68-82.

[7] Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. Eur Heart J. 2020;41(3):407-77.

[8] Latif S, Usman M, Rana R, Qadir J. Phonocardiographic Sensing using Deep Learning for Abnormal Heartbeat Detection [Preprint]. 2018 [cited 2024 Sep 16]. Available from: https://arxiv.org/abs/1801.08322

[9] Liu C, Springer D, Clifford GD. Performance of an open-source heart sound segmentation algorithm on eight independent databases. Physiol Meas. 2017;38(8):1730-45.

[10] Muhammad Y, Tahir M, Hayat M, Chong KT. Early and accurate detection and diagnosis of heart disease using intelligent computational model. Sci Rep. 2020;10(1):19747.

[11] Nabih-Ali M, El-Dahshan EA, Yahia AS. Heart diseases diagnosis using intelligent algorithm based on PCG signal analysis. Circuits Syst. 2017;8(7):184-90.

[12] Nkereuwem E, Ansa G. Enhancing Data Integrity in Telemedicine System Using Blockchain Approach. Res J Sci Technol. 2023;3(2):55-67.

[13] Oliveira J, Renna F, Costa PD, Nogueira M, Oliveira C, Ferreira C. The CIRCOR DigiScope Dataset: From murmur detection to murmur Classification. IEEE J Biomed Health Inform. 2022;26(6):2524-35.

[14] Rath A, Mishra D, Panda G, Pal M. Development and assessment of machine learning based heart disease detection using imbalanced heart sound signal. Biomed Signal Process Control. 2022;76:103730.

[15] Rivera WA, Xanthopoulos P. A priori synthetic over-sampling methods for increasing classification sensitivity in imbalanced data sets. Expert Syst Appl. 2016;66:124-35.

[16] Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol. 2017;70(1):1-25.

[17] Shah SMS, Shah FA, Hussain SA, Batool S. Support Vector Machines-based Heart Disease Diagnosis using Feature Subset, Wrapping Selection and Extraction Methods. Comput Electr Eng. 2020;84:106628.

[18] Srivastava N, Bhatnagar M, Yadav A, Dutta MK, Travieso CM. Machine learning based improved automatic diagnosis of cardiac disorder. In: Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks; 2019 Feb 14-15; Tiruchirappalli, India. New York: ACM; 2019. p. 1-6.

[19] Lin CJ. Welcome to Chih-Jen Lin's home page [Internet]. [cited 2024 Sep 16]. Available from: http://www.csie.ntu.edu.tw/~cjlin

[20] World Health Organization. Cardiovascular diseases (CVDs) [Internet]. 2021 Jun 11 [cited 2024 Sep 16]. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[21] Zeinali Y, Niaki STA. Heart sound classification using signal processing and machine learning algorithms. Mach Learn Appl. 2022;7:100206