Received: 5 July 2024 | Revised: 18 October 2024 | Accepted: 23 January 2025 | Published online: 17 February 2025

Journal of Data Science and Intelligent Systems
2025, Vol. 3(4) 312–323
DOI: 10.47852/bonviewJDSIS52023774
BON VIEW PUBLISHING

RESEARCH ARTICLE

Enhancing Cardiac Health Diagnoses Through Machine Learning Analysis of Phonocardiograms (PCG)

Popal Khan Popalzai[1], Khurram Shehzad Khattak[1], Anwar Mehmood Sohail[1,2,*] and Zawar Hussain Khan[3]

[1]*Department of Computer System Engineering, University of Engineering and Technology, Pakistan*

[2]*Department of Information Systems, University of Malaya, Malaysia*

[3]*Department of Electrical and Computer Engineering, University of Victoria, Canada*

Abstract: Phonocardiograms (PCG) provide a non-invasive approach to analyzing heart sounds, making them vital for the early detection of cardiac issues. However, identifying the most effective machine learning models and feature extraction techniques for classifying PCG signals remains a challenge. This study aims to determine the most efficient and accurate combinations of machine learning models and feature engineering techniques for classifying PCG signals, with the overarching goal of enhancing diagnostic capabilities in heart health. Seven machine learning algorithms—Logistic Regression, Decision Tree, Random Forest, Naive Bayes, AdaBoost, XGBoost, and Support Vector Machine (SVM)—were evaluated. Feature extraction methods such as Mel-frequency cepstral coefficients (MFCC), Linear Predictive Coding (LPC), and Short-Time Fourier Transform (STFT) were applied. Model performance was assessed using metrics including accuracy, precision, recall, and $F1$-score. The study found that advanced models like XGBoost and Random Forest, particularly when combined with STFT and MFCC features, consistently outperformed others. These models demonstrated superior accuracy and $F1$-scores, although they also introduced higher computational complexity. The results suggest that sophisticated model-feature combinations, particularly involving XGBoost and Random Forest with STFT and MFCC, hold promise for improving cardiac diagnostics.

Keywords: phonocardiogram, machine learning, cardiac health diagnostics, signal analysis, feature extraction

1. Introduction

Cardiovascular diseases (CVDs) remain one of the leading global health challenges, with millions of lives lost annually, as reported by the World Health Organization (WHO) [1, 2]. Early and accurate diagnosis is critical in reducing the mortality and morbidity associated with CVDs. Among various diagnostic techniques, electrocardiography (ECG) is a widely recognized non-invasive method that provides a viable alternative to more invasive procedures such as coronary angiography. In recent years, Phonocardiograms (PCG) have gained attention for their ability to visually represent heart sounds captured through auscultation, thereby extending the diagnostic capabilities beyond the limitations of human hearing [3, 4]. PCG signals, obtained using a sensitive microphone or phonocardiographic transducer placed on the chest, capture the timing, intensity, and frequency of heart sounds, providing valuable insights into cardiac function and potential abnormalities such as murmurs or valve disorders.

Research Context: The clinical relevance of PCG lies in its non-invasive nature and its utility in the early detection of various cardiac conditions, including valvular heart diseases, septal defects, and heart failure. Its versatility allows for use in a variety of clinical settings, from advanced inpatient cardiac units to outpatient clinics, making it accessible even in resource-limited environments and telemedicine [3]. However, the interpretation of PCG signals traditionally requires clinical expertise, which may limit its widespread adoption. This challenge presents an opportunity for innovation through the integration of Machine Learning (ML) techniques [5–10]. By leveraging ML's computational power, vast amounts of acoustic data can be processed into actionable insights, enabling the early detection of cardiac abnormalities that might otherwise be missed. As healthcare increasingly moves towards precision medicine, enhancing the diagnostic capabilities of PCG with ML becomes increasingly important.

Before ML models can be applied to PCG data, signal segmentation and feature extraction are essential steps. Feature extraction, in particular, is crucial as it transforms raw PCG signals into analytically meaningful components [2, 11–14]. Various studies have employed different feature extraction techniques for analyzing PCG data [15–18]. For instance, Debbal

***Corresponding author:** Anwar Mehmood Sohail, Department of Computer System Engineering, University of Engineering and Technology, Pakistan and Department of Information Systems, University of Malaya, Malaysia. Email: shl.ktk2@uetpeshawar.edu.pk

and Hamza [19] as well as Aziz et al. [20] emphasized the importance of time-frequency analysis and wavelet transforms, which provide insights into both time and frequency domains. Mel-frequency cepstral coefficients (MFCCs), widely used to capture the spectral characteristics of sounds, have proven effective in differentiating between normal and abnormal cardiac sounds [18]. Techniques like discrete wavelet transform (DWT) and Shannon entropy have also been employed to uncover hidden patterns in PCG data, particularly in the detection of critical conditions like myocardial infarction (MI) [16].

Scope and Significance: This study takes a comprehensive approach by exploring various feature extraction techniques to capture the diverse characteristics of heart sounds. The primary objective is to identify the most efficient and accurate combinations of machine learning models and feature extraction methods for classifying PCG signals, with the ultimate goal of improving heart health diagnostics. By disseminating the results to the medical community and healthcare professionals, this study aims to accelerate the adoption of ML-enhanced PCG analysis in clinical practice. The findings could have a significant impact on the implementation of non-invasive diagnostic tools in hospitals and clinics, particularly in resource-limited settings where advanced diagnostic instruments may not be readily available.

The remainder of this manuscript is structured as follows: Section 2 reviews the related works, Section 3 presents the research methodology employed in this study, Section 4 analyzes and discusses the results, and Section 5 concludes the study.

2. Literature Review

The PCG serves as a crucial tool for capturing the intricate valve signals of the heart, providing insight into its inner workings akin to a conductor orchestrating life's symphony. However, decoding these complex vibrations has historically posed a challenge. Recent strides in advanced ML techniques have emerged as a groundbreaking solution, revolutionizing the classification of PCG signals. This literature review explores the landscape of PCG classification utilizing ML methodologies, addressing methodological approaches, overcoming challenges, and envisioning a promising future for cardiovascular health.

Signal Processing methods, like time-frequency analysis and wavelet transform, are vital for extracting valuable information from PCG data [21]. PCG signals are segmented before input into ML models. MFCCs capture sound spectrum characteristics, aiding in distinguishing regular and abnormal cardiac [22]. Spectral features analyze energy distribution across frequencies, while time-domain analysis provides insights into cardiac events.

Recent advancements in signal processing and analysis have unlocked the hidden beats of the human heart through PCG signals, fostering profound interest in the medical community [2, 23]. In a study by Hernández-Ibarra et al. [11], spectral and sparse methods were compared for identifying anomalies in PCG signals, crucial for cardiovascular disease detection. They employed a Random Forest classifier, incorporating features like MFCCs, LPC, and MP, which showed improved diagnostic accuracy over individual techniques, particularly with suboptimal recordings. Findings underscored the necessity of data balancing techniques like SMOTE for precise categorization in imbalanced datasets.

Movahedi et al. [6] introduced a novel method using hidden Markov models to efficiently segment ECG and PCG signals, enabling precise quantitative analysis and diagnosis by accurately separating cardiac signals from PCG data. Khan et al. [24] researched classification algorithms for detecting cardiac abnormalities, emphasizing variables in frequency and time domains and their interactions. The study meticulously examines sound aspects, including regular heart rhythm and murmurs' rapid fluctuations. In 2022, Yang et al. [13] introduced a novel feature extraction method from PCG data using fuzzy logic, selectively identifying valuable notes while filtering out noise. Four machine learning algorithms were compared to evaluate the features, enhancing cardiac problem diagnosis.

Various machine learning and deep learning algorithms have been explored for classifying PCG signals, demonstrating their efficacy in identifying PCG signals despite intricate patterns [5, 7, 17, 25–27].

The emergence of CNNs and other deep learning techniques has revolutionized PCG classification, as they possess exceptional ability in detecting subtle patterns, akin to virtuosos [14, 28]. In a study by Mukherjee et al. [7], a novel U-Net deep learning framework is introduced to mitigate noise interference in PCG signals, enabling the elimination of common real-world noises such as sneezing and coughing. These advancements enable a more comprehensive understanding of the cardiac symphony, improving both human and computer decision-making and diagnoses in cardiac health assessment. Li et al. [25], on the other hand, employed CNNs for automatic classification, showcasing remarkable performance in detecting hierarchical characteristics in PCG signals. Hassanuzzaman et al. [26] utilized deep learning algorithms to detect and diagnose congenital heart disease (CHD) by analyzing faint heart sounds in youngsters. They developed a remarkable one-dimensional convolutional neural network (1D-CNN) trained on phonocardiogram recordings from pediatric patients. This approach shows promise for early identification and management of congenital heart abnormalities in children. Altaf et al. [5] provide a thorough analysis of ML and DL applications in PCG classification, offering valuable insights into cardiovascular diagnosis. They highlight the potential of these technologies to revolutionize the field by providing datasets, algorithms, and feature extraction strategies.

This paper addresses a critical gap in existing literature, representing one of its most significant contributions. Until now, there has been a notable absence of comprehensive analyses comparing the performance of multiple ML models trained on various PCG signal features selection techniques. By offering detailed insights into the most effective model-feature combinations, this study not only fills a crucial void in current knowledge but also provides valuable guidance for future research and practical applications in the field of biological signal processing.

3. Research Methodology

This section outlines a systematic approach using sophisticated machine learning techniques to analyze and categorize PCG data, focusing on the Physionet Challenge 2016 dataset [29]. It involves stages like data preparation, feature extraction, model selection and training, and validation. Precision is ensured through procedures like normalization, denoising and feature extraction, leading to model assessment via cross-validation.

3.1. Dataset description

This work relies on the PhysioNet Challenge 2016 dataset [29], which contains extensive heart sound recordings. The dataset includes PCG signals from various ages, health conditions and environments, recorded in clinical and home settings using different equipment. The diversity of heart sounds makes distinguishing between normal and abnormal sounds challenging, reflecting real clinical scenarios. The dataset includes annotations and metadata with contextual information on recording conditions, participants' health, and specialist evaluations. Making it ideal for training robust machine learning models for PCG signal interpretation.

The database comprises 4,430 recordings from 1,072 subjects, resulting in a total of 233,512 heart sound samples, collected from both healthy individuals and patients with various conditions, including heart valve disease and coronary artery disease. These recordings were gathered using diverse equipment in both clinical settings and nonclinical environments, such as in-home visits. The duration of each recording ranged from a few seconds to several minutes. Additional data provided includes subject demographics (age and gender), details of the recordings (number per patient, body location, and duration), synchronously recorded signals (such as ECG), sampling frequency, and the type of sensor used. Participants were asked to classify the recordings as normal, abnormal, or not possible to evaluate due to noise or uncertainty.

3.2. Data preparation and preprocessing

The audio recordings from the dataset are imported in waveform format and converted into a numerical representation for analysis. Standardizing the data ensures uniformity, and normalization reduces the impact of variations in recording equipment or conditions. Annotations from the PhysioNet dataset categorize the recordings as "normal" or "abnormal." To maintain uniformity for effective pattern recognition and feature extraction, the PCG signals are segmented into equal lengths. Precise data preparation is essential for accurate analysis and classification of PCG signals. Using the PhysioNet Challenge 2016 dataset enhances the work's applicability in real-life clinical environments, adding authenticity and complexity.

3.3. Feature extraction

Feature extraction is a crucial aspect of this inquiry, since it is employed to convert unprocessed PCG signals into components that are analytically valuable. This study adopts a comprehensive approach by gathering all the different characteristics of heart sounds through seven different feature extraction techniques. These feature extraction techniques are MFCCs, LPC, STFT, Chroma features, Spectral contrast, Tonnetz and Pitch features. The integration of all the methodologies synergistically enhances data analysis by providing a comprehensive perspective.

3.4. ML models training and validation

The core of this study involves selecting and training seven machine learning models to classify PCG signals, with each stage meticulously addressing key aspects of the machine learning process.

Each model undergoes a rigorous training procedure designed to differentiate between normal and abnormal PCG signals. The models are trained independently on the dataset, optimizing their internal parameters—such as neural network weights or decision tree split criteria—to minimize prediction errors. To ensure robust model performance, efforts are made to expose each model to a balanced distribution of normal and abnormal signals. Model performance is continuously monitored using metrics like training loss and accuracy, with adjustments made as needed to address issues like overfitting, such as modifying the training data or altering the learning rate.

A 5-fold cross-validation approach is employed to comprehensively evaluate each model. The training dataset is divided into five equal parts, and in each iteration, a different subset is used for validation while the remaining portions are used for training. This process is repeated five times, ensuring that the model's efficacy is assessed across different subsets of the dataset.

Hyperparameter tuning is the final step in the training and validation process. Specific model parameters are fine-tuned using methods like grid search, systematically exploring different parameter combinations. Each model is evaluated with various parameter values to identify the most effective combination that maximizes accuracy and generalizability. This rigorous and iterative process of training, validation, and hyperparameter tuning is crucial for ensuring the accuracy and reliability of the study's findings. Important hyperparameters of each ML model have been presented in Table 1.

Logistic Regression (LR): Logistic Regression is chosen for its simplicity, speed, and interpretability, making it a fundamental model for binary classification [30, 31]. A high max_iter parameter ensures sufficient iterations for complex datasets, while the random_state parameter ensures reproducibility, aiding in the assessment of more sophisticated models' effectiveness.

Table 1
Hyperparameters for ML models

Model	Hyperparameters	Setting
LR	C	1.0
	solver	lbfgs
	max_iter	100
	penalty	l2
KNN	n_neighbors	5
	Weights	uniform
	Algorithm	Auto
	p	2
DT	criterion	Gini
	max_depth	None
	min_samples_split	2
	random_state	none
RF	n_estimators	100
	criterion	Gini
	max_depth	None
	min_sample_split	2
	min_sample_leaf	1
	Bootstrap	True
	random_state	None
NB	var_smoothing	1e-9
AdaBoost	n_estimators	50
	learning_rate	1.0
	random_state	175
	base_estimator	
XGBoost	n_estimators	100
	learning_rate	0.1
	subsample	1.0
	colsample_bytree	1.0
	gamma	0
	reg_alpha	0
	reg_lambda	1
	random_state	175
	n_jobs	1
	max_depth	6

K-Nearest Neighbors (KNN): K-Nearest Neighbors is a non-parametric model, ideal for feature-based classification when parametric models cannot establish linear relationships [32–34]. The n_neighbors parameter balances variance and bias, fine-tuning the model to avoid overfitting or underfitting.

Decision Tree (DT): Decision Tree models excel in visually representing complex decision-making processes and are well-suited for handling intricate data patterns and non-linear correlations [35, 36]. The random_state parameter ensures consistent decision tree generation, providing stable results throughout the study.

Random Forest (RF): Random Forest, an ensemble of decision trees, reduces overfitting and enhances interpretability compared to individual decision trees [37, 38]. The model's randomness in training enhances accuracy and resilience, while the n_jobs parameter optimizes CPU resource usage for large datasets.

Gaussian Naive Bayes (NB): Gaussian Naive Bayes is a probabilistic approach particularly effective in high-dimensional feature spaces like audio signal processing [39, 40]. Although it assumes feature independence—which may not always hold in real-world data—it remains a fast and effective model for the task at hand.

AdaBoost: AdaBoost improves the performance of weak classifiers by iteratively adjusting the weights of misclassified instances [17, 41]. The random_state parameter ensures consistency in the sequential processes, leading to steady model performance.

XGBoost: XGBoost is known for its exceptional performance, scalability, and efficiency, particularly in handling large datasets with numerous features [42, 43]. The use_label_encoder parameter enhances the handling of categorical data, while the eval_metric parameter is used to assess training effectiveness. This model excels in dealing with sparse data, performing rapid gradient boosting, and executing regularized boosting.

Each model was carefully selected and configured to address the specific challenges posed by the PCG dataset. The subsequent training and validation stages leveraged the strengths of each model, laying a solid foundation for the study's findings.

3.5. ML models evaluation

To rigorously evaluate the effectiveness of each machine learning model with various feature types, we employed several key performance metrics. Class 0 and Class 1 used in the tables indicate normal and abnormal PCG signals respectively.

1) Accuracy metrics

Cross-Validation Accuracy: This represents the average accuracy across multiple training and testing iterations, ensuring a more robust evaluation of the model's performance.

Test Set Accuracy: This is the percentage of correctly classified instances in a dedicated test set, providing a final assessment of the model's generalization ability.

Validation Set Accuracy: Similar to the test set, this is used to evaluate the model's performance on unseen data during the training process, helping to tune hyperparameters.

2) Precision and recall metrics

Test Set Precision (Class 0/1): Precision measures how many of the instances predicted as class 0 (or 1) are actually class 0 (or 1). A high precision indicates few false positives.

Test Set Recall (Class 0/1): Recall measures how many of the actual class 0 (or 1) instances were correctly predicted. A high recall indicates few false negatives.

3) F1-score metrics

Test Set *F*1-Score (Class 0/1): The *F*1-score is the harmonic mean of precision and recall, providing a balanced measure of both. A high *F*1-score indicates good performance in terms of both precision and recall.

Weighted Avg *F*1-Score (Test Set): This is the average *F*1-score across all classes, weighted by the number of instances in each class. It provides an overall evaluation of the model's performance, considering the class distribution.

These metrics are essential for evaluating the performance of a machine learning model, especially in classification tasks. They help to assess the model's accuracy, precision, recall, and overall effectiveness in predicting the correct class for given instances.

4. Results

In this section, we conduct a comprehensive analysis of seven machine learning models for the identification and classification of cardiac acoustic signals. Each model was systematically evaluated using seven distinct feature selection techniques, with their performance rigorously compared in an exhaustive study.

4.1. ML models with MFCC

Mel-Frequency Cepstral Coefficients (MFCCs) play a pivotal role in processing auditory information within this investigation [18, 44, 45]. The PCG signal is first segmented into short frames, and for each frame, the periodogram estimate of the power spectrum is calculated. The mel scale, a perceptual scale where pitches are perceived as equidistant, is then applied to assign spectral intensities. To derive the MFCCs, the logarithm of the power values at each mel frequency is taken, followed by a discrete cosine transform. This method effectively captures the detailed tonal and textural characteristics necessary for distinguishing between normal and abnormal cardiac sounds.

We evaluated various machine learning models for the classification of heart sounds using MFCC features. Among the models, Random Forest achieved the highest accuracy, exceeding 89% across different evaluation sets, and demonstrated exceptional performance in identifying normal heart sounds. Decision Tree and K-Nearest Neighbors (KNN) also produced strong results, with accuracies exceeding 80%. Interestingly, while Naive Bayes exhibited a lower overall accuracy, it was particularly effective in detecting abnormal heart sounds. This finding suggests that different models may offer specific advantages depending on the classification task. A summary of model performance with MFCC features is presented in Table 2.

4.2. ML models with LPC

Linear Predictive Coding (LPC) aims to estimate the coefficients of a linear filter that closely replicates the original signal when applied [39, 46, 47]. LPC effectively captures the primary spectral characteristics, or formants, within cardiac sounds, emphasizing resonance frequencies that are crucial for identifying heart problems. K-Nearest Neighbors (KNN) demonstrated the highest test accuracy (86.15%), highlighting its effectiveness in leveraging Linear Predictive Coding (LPC) features for heart sound

Table 2
MFCC feature-based models PCG signal classification

Model	Cross-Validation accuracy (%)	Test set accuracy (%)	Validation set accuracy (%)	Test set precision (Class 0)	Test set precision (Class 1)	Test set recall (Class 0)	Test set recall (Class 1)	Test set $F1$-score (Class 0)	Test set $F1$-score (Class 1)	Weighted Avg $F1$-score (Test set) (%)
LR	82.46	83.38	74.07	0.88	0.66	0.91	0.57	0.89	0.61	83
KNN	83.96	84.43	69.54	0.85	0.82	0.97	0.41	0.91	0.55	82
DT	82.99	82.81	97.62	0.89	0.62	0.89	0.62	0.89	0.62	83
RF	89.20	89.97	97.28	0.91	0.84	0.96	0.70	0.94	0.76	90
NB	78.26	81.28	71.23	0.95	0.56	0.80	0.85	0.87	0.68	82
AdaBoost	85.15	86.91	71.12	0.90	0.75	0.94	0.64	0.92	0.69	87
XGBoost	89.94	89.88	97.28	0.92	0.81	0.95	0.73	0.94	0.77	90

Table 3
LPC feature-based models for PCG signal classification

Model	Cross-Validation accuracy (%)	Test set accuracy (%)	Validation set accuracy (%)	Test set precision (Class 0)	Test set precision (Class 1)	Test set recall (Class 0)	Test set recall (Class 1)	Test set $F1$-score (Class 0)	Test set $F1$-score (Class 1)	Weighted Avg $F1$-score (Test set) (%)
LR	77.99	78.03	53.23	0.79	0.59	0.97	0.14	0.87	0.22	77
KNN	86.03	86.15	76.44	0.89	0.75	0.94	0.60	0.91	0.66	86
DT	80.59	80.80	96.94	0.89	0.57	0.86	0.63	0.87	0.60	81
RF	86.17	87.68	97.51	0.89	0.82	0.96	0.60	0.92	0.69	87
NB	26.53	26.93	53.11	0.82	0.23	0.07	0.95	0.12	0.37	27
AdaBoost	80.87	81.38	67.16	0.84	0.67	0.94	0.38	0.89	0.48	81
XGBoost	85.95	87.20	97.51	0.90	0.75	0.94	0.66	0.92	0.70	87

classification. Random Forest (RF) and Decision Tree (DT) also performed well, with strong validation set scores, indicating their ability to capture the spectral and temporal characteristics of LPC for accurate classification. Conversely, Naive Bayes (NB) showed poor performance, achieving only 26.93% accuracy, likely due to a mismatch between its underlying assumptions and the nature of LPC data. Table 3 provides a summary of model performance using LPC features.

4.3. ML models with Chroma

Chroma features, commonly used in music analysis, are instrumental in identifying harmonic patterns within cardiac sounds. These features simplify complex spectral shapes into a small number of bins, each representing one of the twelve semitones. This reduction process helps minimize background noise, enabling clearer differentiation of heart sounds and their harmonic characteristics.

In our analysis, Logistic Regression (LR) and K-Nearest Neighbors (KNN) achieved moderate test accuracy, approximately 78%, indicating room for improvement in leveraging Chroma features for heart sound classification. In contrast, Random Forest (RF) and Decision Tree (DT) performed exceptionally well on validation sets, demonstrating their capacity to effectively learn and generalize from Chroma features. This underscores the potential of Chroma features to capture subtle variations in heart sound harmonics, which can be valuable in medical diagnosis.

On the other hand, Naive Bayes (NB) performed poorly, with an accuracy around 50%, suggesting that simpler models may struggle with the complexity inherent in Chroma features for heart sound classification. Table 4 provides a summary of model performance using Chroma features.

4.4. ML models with spectral contrast

Spectral contrast measures the variation within the sound spectrum by calculating the disparity between its peaks and troughs [9, 21]. Peaks typically represent harmonic components, while troughs correspond to noise or less prominent frequencies. Analyzing spectral contrast provides insights into the textural fluctuations in heart sounds, making it a valuable tool for detecting cardiac anomalies.

The evaluation of machine learning models using Spectral Contrast features yielded two key findings. First, Decision Tree (DT) and Random Forest (RF) excelled, achieving validation set accuracies exceeding 96%.

This indicates their strong capability to leverage Spectral Contrast, which effectively captures the differences between high and low points in the frequency spectrum, allowing for precise differentiation between various heart sounds. Second, Logistic Regression (LR) and K-Nearest Neighbors (KNN) also performed well, with KNN reaching a test set accuracy of 86.25%. This suggests that these models hold potential for accurate heart sound classification using Spectral Contrast features.

Table 4
Chroma feature-based models for PCG signal classification

Model	Cross-Validation accuracy (%)	Test set accuracy (%)	Validation set accuracy (%)	Test set precision (Class 0)	Test set precision (Class 1)	Test set recall (Class 0)	Test set recall (Class 1)	Test set $F1$-score (Class 0)	Test set $F1$-score (Class 1)	Weighted Avg $F1$-score (Test set) (%)
LR	76.14	76.50	51.53	0.79	0.31	0.96	0.07	0.86	0.12	70
KNN	77.20	78.13	51.76	0.79	0.44	0.98	0.05	0.88	0.09	71
DT	66.94	66.76	96.38	0.80	0.27	0.76	0.32	0.78	0.30	68
RF	77.64	78.80	95.92	0.79	0.61	0.99	0.05	0.88	0.09	71
NB	50.08	50.72	54.02	0.86	0.27	0.44	0.75	0.58	0.36	54
AdaBoost	75.58	76.70	57.30	0.80	0.38	0.94	0.13	0.86	0.20	72
XGBoost	77.00	77.84	96.39	0.82	0.56	0.95	0.26	0.87	0.35	77

Table 5
Spectral contrast feature-based models for PCG signal classification

Model	Cross-Validation accuracy (%)	Test set accuracy (%)	Validation set accuracy (%)	Test set precision (Class 0)	Test set precision (Class 1)	Test set recall (Class 0)	Test set recall (Class 1)	Test set $F1$-score (Class 0)	Test set $F1$-score (Class 1)	Weighted Avg $F1$-score (Test set) (%)
LR	84.68	84.91	67.61	0.89	0.68	0.92	0.58	0.91	0.62	85
KNN	85.59	86.25	69.42	0.88	0.78	0.96	0.50	0.92	0.61	86
DT	79.73	78.99	96.72	0.87	0.51	0.86	0.52	0.87	0.52	79
RF	86.20	86.34	97.06	0.89	0.72	0.94	0.59	0.92	0.65	86
NB	77.05	77.46	65.69	0.93	0.49	0.77	0.81	0.84	0.61	77
AdaBoost	84.42	83.86	67.84	0.88	0.64	0.91	0.61	0.90	0.62	84
XGBoost	87.05	87.11	97.73	0.91	0.72	0.93	0.65	0.92	0.68	87

Overall, the results indicate that both complex and simpler models can effectively utilize Spectral Contrast for heart sound classification. Table 5 summarizes the performance of the models with Spectral Contrast features.

4.5. ML models with Tonnetz

Tonnetz features capture the harmonic relationships in an audio stream by utilizing tonal centroid characteristics [48, 49]. These features are particularly useful for detecting subtle alterations in the harmonic structure of cardiac sounds, which may indicate underlying heart conditions.

In our analysis, the Random Forest (RF) model excelled in classifying heart sounds using Tonnetz features, achieving a high accuracy of 95.81%. In contrast, simpler models such as Logistic Regression (LR) and K-Nearest Neighbors (KNN) encountered difficulties, suggesting that the complexity of Tonnetz features may require the more sophisticated capabilities of advanced models like RF to be effectively leveraged. Table 6 provides a summary of model performance using Tonnetz features.

4.6. ML models with pitch

Pitch features analyze the fundamental frequency of heart sounds [50]. Variations in this frequency can signify normal or abnormal cardiac conditions, making pitch characteristics essential for accurately recognizing these sounds.

In our evaluation, Logistic Regression (LR) and K-Nearest Neighbors (KNN) demonstrated limited performance with pitch features, achieving accuracies of 74% and 78%, respectively. In contrast, Random Forest (RF) and XGBoost exhibited stronger performance, indicating their enhanced capability to handle pitch variations critical for diagnosing heart conditions. This suggests that more complex models may be better suited for effectively utilizing pitch information in heart sound analysis. Table 7 summarizes the performance of the models using pitch features.

4.7. ML models with STFT

Short-Time Fourier Transform (STFT) analyzes the frequency and phase characteristics of Phonocardiogram (PCG) signals over brief, overlapping time intervals [51–53]. This method effectively captures the dynamic fluctuations in heart sounds, providing a critical representation for identifying cardiac abnormalities.

When evaluating machine learning models for heart sound classification, STFT features, which represent sound in both the time and frequency domains, produced the most robust results. Both advanced models like XGBoost and Random Forest (RF), as well as simpler models, demonstrated significant accuracy

Table 6
Tonnetz feature-based models for phonocardiogram (PCG) signal classification

Model	Cross-Validation accuracy (%)	Test set accuracy (%)	Validation set accuracy (%)	Test set precision (Class 0)	Test set precision (Class 1)	Test set recall (Class 0)	Test set recall (Class 1)	Test set $F1$-score (Class 0)	Test set $F1$-score (Class 1)	Weighted Avg $F1$-score (Test set) (%)
LR	77.07	78.41	49.26	0.78	0.00	1.00	0.00	0.88	0.00	69
KNN	77.01	78.41	50.74	0.79	0.50	1.00	0.01	0.88	0.02	69
DT	69.04	69.15	96.49	0.81	0.29	0.80	0.31	0.80	0.30	69
RF	77.18	78.32	95.81	0.79	0.47	0.99	0.04	0.88	0.07	78
NB	40.81	43.08	50.28	0.83	0.24	0.34	0.75	0.49	0.36	43
AdaBoost	75.56	76.70	57.30	0.80	0.38	0.94	0.13	0.86	0.20	72
XGBoost	78.85	79.66	96.49	0.82	0.56	0.95	0.26	0.88	0.35	80

Table 7
Pitch feature-based models for phonocardiogram (PCG) signal classification

Model	Cross-validation accuracy (%)	Test set accuracy (%)	Validation set accuracy (%)	Test set precision (Class 0)	Test set precision (Class 1)	Test set recall (Class 0)	Test set recall (Class 1)	Test set $F1$-score (Class 0)	Test set $F1$-score (Class 1)	Weighted Avg $F1$-score (Test set) (%)
LR	72.53	74.21	56.51	0.79	0.26	0.92	0.11	0.85	0.15	70
KNN	76.70	78.23	51.42	0.79	0.47	0.98	0.06	0.88	0.11	71
DT	69.02	69.25	77.80	0.79	0.24	0.83	0.19	0.81	0.21	68
RF	76.46	77.46	77.69	0.79	0.40	0.96	0.09	0.87	0.14	71
NB	75.05	76.03	49.49	0.79	0.29	0.95	0.08	0.86	0.12	70
AdaBoost	77.11	78.03	50.74	0.79	0.42	0.98	0.05	0.88	0.09	71
XGBoost	76.72	78.41	65.12	0.79	0.50	0.98	0.07	0.88	0.12	71

Table 8
STFT feature-based models for phonocardiogram (PCG) signal classification

Model	Cross-Validation accuracy (%)	Test set accuracy (%)	Validation set accuracy (%)	Test set precision (Class 0)	Test set precision (Class 1)	Test set recall (Class 0)	Test set recall (Class 1)	Test set $F1$-score (Class 0)	Test set $F1$-score (Class 1)	Weighted Avg $F1$-score (Test set) (%)
LR	81.35	83.38	91.28	0.90	0.61	0.89	0.65	0.89	0.63	84
KNN	87.08	88.63	72.59	0.90	0.81	0.96	0.62	0.93	0.70	88
DT	83.65	83.38	96.26	0.90	0.61	0.89	0.62	0.89	0.62	83
RF	87.68	88.44	97.28	0.90	0.82	0.96	0.59	0.93	0.69	88
NB	69.44	71.16	63.87	0.92	0.41	0.69	0.79	0.79	0.54	74
AdaBoost	86.32	86.44	72.71	0.90	0.71	0.93	0.64	0.91	0.67	86
XGBoost	90.71	91.31	98.30	0.93	0.84	0.96	0.73	0.95	0.78	91

improvements when utilizing STFT features. This underscores the effectiveness of STFT in time-frequency analysis, making it a powerful tool for heart sound categorization. Table 8 provides a summary of model performance using STFT features.

The impact of feature selection on model performance in classifying PCG signals is evident across various feature types. Complex models such as XGBoost and RF consistently performed well, while simpler models like Logistic Regression (LR) and K-Nearest Neighbors (KNN) struggled with certain features. This highlights the importance of selecting the appropriate features, as they can significantly enhance model accuracy, particularly in the intricate domain of biological signal processing, such as heart sound analysis.

5. Discussion

This section provides a detailed analysis of the performance of various machine learning models in classifying heart sounds using different feature sets. By critically evaluating the strengths and limitations of each model and feature combination, we aim to identify the most effective strategies for heart sound classification. The discussion highlights key insights into model behavior, feature selection, and their implications for future research and clinical applications in cardiac diagnostics.

5.1. Comparative analysis of ML models

This study conducted a comprehensive comparison of machine learning models for heart sound classification using various feature sets. Key findings are summarized as follows:

Logistic Regression (LR): LR performed well overall, particularly with Short-Time Fourier Transform (STFT) features, which leverage time-frequency analysis. However, it struggled with more complex features such as Tonnetz and Pitch. This indicates that LR's linear nature may limit its ability to capture the intricate relationships inherent in these feature sets.

K-Nearest Neighbors (KNN): KNN showed strong performance with STFT and Spectral Contrast features. However, it exhibited overfitting issues with Chroma and Pitch features, evidenced by a noticeable drop in accuracy from the validation to the test set. This suggests that KNN's generalizability may be limited when applied to these specific features.

Decision Tree (DT): DT achieved high accuracy on validation sets using features like Linear Predictive Coding (LPC), Tonnetz, and STFT. However, a reduction in accuracy on the test sets suggests potential overfitting. Despite this, DT effectively captured fundamental patterns in PCG signals for these specific features.

Random Forest (RF): RF consistently performed well across various feature types, though it may be prone to overfitting when dealing with more complex features. RF's ensemble approach provided robustness, making it a reliable choice for PCG signal classification.

Naive Bayes (NB): NB exhibited inconsistent performance across different feature types. While it excelled in detecting abnormal signals in certain cases, it struggled with LPC and Chroma features. This inconsistency likely stems from NB's assumption of feature independence, which may not hold true for the complex structure of PCG data.

AdaBoost and XGBoost: Both models demonstrated superior performance across all feature sets, with XGBoost particularly excelling with STFT features due to its strength in time-frequency analysis. These advanced models effectively combine weaker learners and iteratively correct errors, resulting in highly accurate predictions.

Table 9 presents the precision of each machine learning model when applied to the seven feature types: MFCC, LPC, Chroma, Spectral Contrast, Tonnetz, Pitch, and STFT. The results underscore the critical impact of feature selection on model performance in PCG signal classification.

The analysis revealed that certain models, such as XGBoost and RF, were highly robust and accurate across multiple feature types. For instance, RF achieved 89.97% accuracy with MFCCs and 88.44% with STFT, while XGBoost reached 89.88% with MFCCs and 91.31% with STFT. Conversely, simpler models like NB showed lower accuracy, particularly with LPC (26.93%) and Tonnetz (43.08%). This highlights the importance of selecting the appropriate machine learning models based on the feature sets to achieve accurate PCG signal classification.

5.2. Comparative analysis of features

Figure 1 illustrates the comparative analysis of different feature types in relation to various machine learning models for PCG signal classification. Several key trends and patterns emerged from this investigation:

MFCC Features: These features consistently provided strong classification performance across multiple models, particularly with ensemble methods like RF (89.97% with MFCCs) and XGBoost (91.31% with STFT). MFCC features effectively captured essential audio characteristics, proving reliable across diverse machine learning models.

LPC Features: While NB struggled with LPC features (26.93% accuracy), DT (80.80%) and RF (87.68%) excelled. This suggests that certain models are better equipped to extract relevant information from LPC features for differentiating cardiac sounds.

Chroma Features: These yielded moderate performance, with LR achieving 76.50% accuracy and KNN 78.13%. However, when combined with Spectral Contrast features, DT (78.99%) and RF (86.34%) delivered outstanding results, leveraging spectral variations in audio signals effectively.

Tonnetz Features: While most models struggled with these features, RF (78.32%) and DT (69.15%) maintained relatively high accuracy, suggesting their potential for extracting tonal information crucial for heart sound classification.

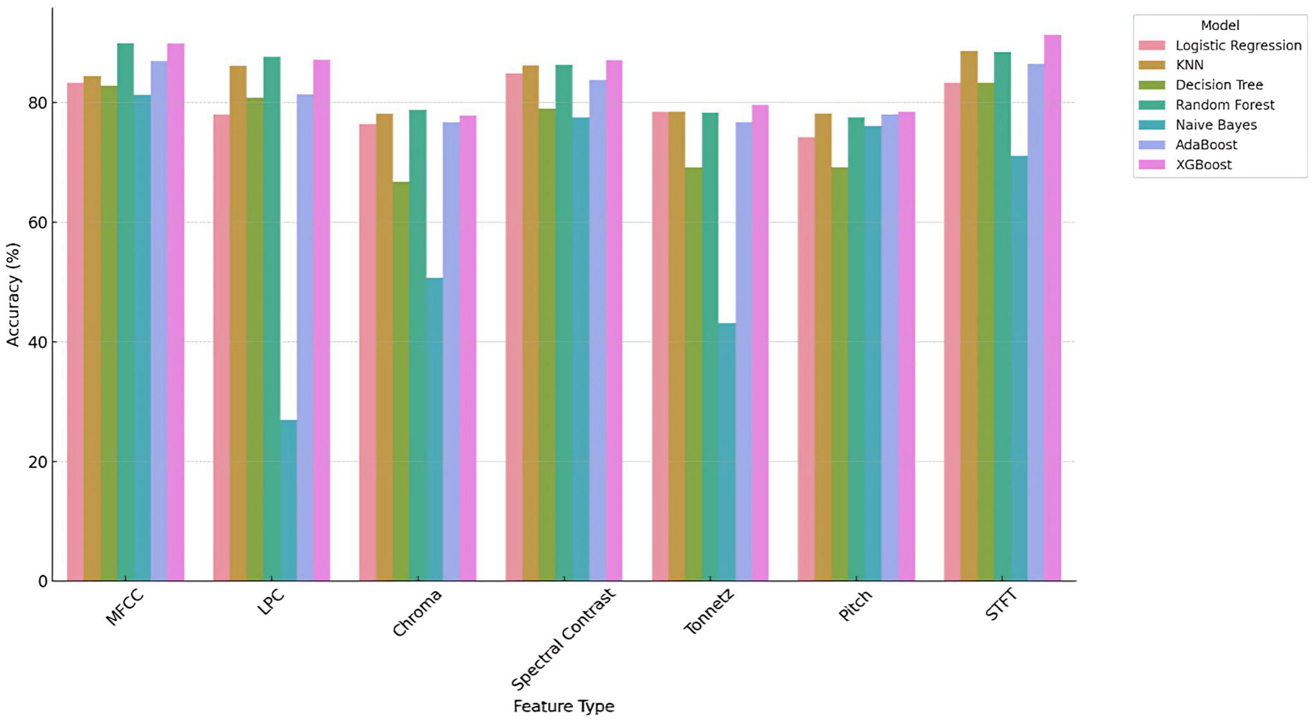
Pitch Features: Generally, pitch characteristics led to reduced accuracy across models, except for RF (77.46%) and XGBoost (78.41%), which performed exceptionally well, indicating their suitability for handling pitch variations critical in diagnosing heart conditions.

STFT Features: These were effective across all models, particularly XGBoost (91.31%), emphasizing their value in time-frequency representation for heart sound classification.

Table 9
Test set accuracies across different features

Model	MFCC	LPC	Chroma	Spectral contrast	Tonnetz	Pitch	STFT
LR	83.38	78.03	76.50	84.91	78.41	74.21	83.38
KNN Neighbors	84.43	86.15	78.13	86.25	78.41	78.23	88.63
DT	82.81	80.80	66.76	78.99	69.15	69.25	83.38
RF	89.97	87.68	78.80	86.34	78.32	77.46	88.44
NB	81.28	26.93	50.72	77.46	43.08	76.03	71.16
AdaBoost	86.91	81.38	76.70	83.86	76.70	78.03	86.44
XGBoost	89.88	87.20	77.84	87.11	79.66	78.41	91.31

Figure 1
Test accuracies across various ML models



This study concludes that RF and XGBoost models with carefully selected features, particularly STFT, holds significant promise for developing precise and reliable classification systems in biomedical signal processing. These findings pave the way for future research into advanced machine learning algorithms paired with tailored feature engineering, which could enhance the precision and reliability of PCG signal classification systems, ultimately advancing medical diagnostics and healthcare.

Key insights include the high efficacy of XGBoost and RF models, especially with features like STFT and MFCC, which consistently improved model performance. Conversely, simpler models like NB struggled with complex features. The study underscores the importance of strategic feature selection, noting that while STFT and MFCC were universally effective, features like LPC, Chroma, and Tonnetz were beneficial only for specific models. These findings highlight the potential of machine learning to enhance cardiac disease diagnosis and inform the development of non-invasive diagnostic tools in cardiology, emphasizing the critical role of feature selection and model choice.

6. Conclusion

This paper presents a comprehensive evaluation of machine learning models for the classification of PCG signals, utilizing a variety of feature extraction methods. The findings underscore the critical role of feature selection in optimizing model performance, highlighting the importance of strategic feature engineering in achieving accurate and reliable classification outcomes. Notably, Random Forest and XGBoost models with carefully selected features, particularly the Short-Time Fourier Transform, emerged as a promising approach for developing robust classification systems in biomedical signal processing.

While this study offers valuable insights, it also acknowledges certain limitations. The potential variation in model performance across different datasets necessitates further testing and validation to ensure generalizability. Additionally, the computational demands of advanced models, although beneficial, pose challenges, particularly in resource-constrained environments. Future research should aim to develop more computationally efficient techniques that maintain high accuracy without compromising feasibility.

In building upon the findings of this study, several avenues for future exploration emerge. Investigating hybrid or ensemble models could harness the strengths of multiple approaches, leading to more reliable and accurate classification systems. Moreover, developing methods for real-time analysis of PCG signals could enhance the applicability of machine learning models in clinical settings, enabling timely and precise diagnostics. Expanding the evaluation of models on larger and more diverse datasets, representing various patient populations, will improve their robustness and generalizability. Additionally, further research into sophisticated feature extraction and selection techniques could significantly enhance model accuracy and the diagnostic capabilities of heart sound classification systems.

By addressing these areas, future research has the potential to advance the field of PCG signal classification, ultimately improving clinical outcomes and contributing to the development of non-invasive diagnostic tools in cardiology.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Popal Khan Popalzai: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation. **Khurram Shehzad Khattak:** Conceptualization, Methodology, Formal analysis, Writing – review & editing, Supervision. **Anwar Mehmood Sohail:** Conceptualization, Writing – original draft, Writing – review & editing. **Zawar Hussain Khan:** Methodology, Supervision.

References

[1] Gaidai, O., Cao, Y., & Loginov, S. (2023). Global cardiovascular diseases death rate prediction. *Current Problems in Cardiology*, *48*(5), 101622. https://doi.org/10.1016/j.cpcardiol.2023.101622

[2] Vijean, V., Prathaban, S. N., Palaniappan, R., Chin, L. C., Abdullah, R., Maniraj, V., . . . , & Abdelaziz, K. M. H. (2024). Investigation of valvular heart defects through phonocardiogram signals. *International Journal of Integrated Engineering*, *16*(1), 328–336. https://doi.org/10.30880/ijie.2024.16.01.028

[3] Zheng, Y., Guo, X., Yang, Y., Wang, H., Liao, K., & Qin, J. (2023). Phonocardiogram transfer learning-based CatBoost model for diastolic dysfunction identification using multiple domain-specific deep feature fusion. *Computers in Biology and Medicine*, *156*, 106707. https://doi.org/10.1016/j.compbiomed.2023.106707

[4] Kazemnejad, A., Karimi, S., Gordany, P., Clifford, G. D., & Sameni, R. (2024). An open-access simultaneous electrocardiogram and phonocardiogram database. *Physiological Measurement*, *45*(5), 055005. https://doi.org/10.1088/1361-6579/ad43af

[5] Altaf, A., Mahdin, H., Alive, A. M., Ninggal, M. I. H., Altaf, A., & Javid, I. (2023). Systematic review for phonocardiography classification based on machine learning. *International Journal of Advanced Computer Science and Applications*, *14*(8), 806–817. https://doi.org/10.14569/IJACSA.2023.0140889

[6] Movahedi, M. M., Shakerpour, M., Mousavi, S., Nori, A., Dehkordi, S. H. M., & Parsaei, H. (2023). A hardware-software system for accurate segmentation of phonocardiogram signal. *Journal of Biomedical Physics & Engineering*, *13*(3), 261–268. https://doi.org/10.31661/jbpe.v0i0.2104-1301

[7] Mukherjee, A., Banerjee, R., & Ghose, A. (2023). A novel U-Net architecture for denoising of real-world noise corrupted phonocardiogram signal. *arXiv Preprint:2310.00216*. https://doi.org/10.48550/arXiv.2310.00216

[8] Hamza, M. F. A. B., & Sjarif, N. N. A. (2024). A comprehensive overview of heart sound analysis using machine learning methods. *IEEE Access*, *12*, 117203–117217. https://doi.org/10.1109/ACCESS.2024.3432309

[9] Chambi, E. M., Cuela, J., Zegarra, M., Sulla, E., & Rendulich, J. (2024). Benchmarking time-frequency representations of phonocardiogram signals for classification of valvular heart diseases using deep features and machine learning. *Electronics*, *13*(15), 2912. https://doi.org/10.3390/electronics13152912

[10] de Fazio, R., Spongano, L., de Vittorio, M., Patrono, L., & Visconti, P. (2024). Machine learning algorithms for processing and classifying unsegmented phonocardiographic signals: An efficient edge computing solution suitable for wearable devices. *Sensors*, *24*(12), 3853. https://doi.org/10.3390/s24123853

[11] Hernández-Ibarra, R. F., Alonso-Arévalo, M. Á., & García-Canseco, E. del C. (2023). Comparison of spectral and sparse feature extraction methods for heart sounds classification. *Revista Mexicana de Ingenieria Biomedica*, *44*(4), 6–22. https://doi.org/10.17488/rmib.44.4.1

[12] Putra, W. R., Mandala, S., & Pramudyo, M. (2021). Study of feature extraction methods to detect valvular heart disease (VHD) using a phonocardiogram. In *2021 International Conference on Intelligent Cybernetics Technology & Applications*, 122–127. https://doi.org/10.1109/ICICyTA53712.2021.9689119

[13] Yang, W., Xu, J., Xiang, J., Yan, Z., Zhou, H., Wen, B., . . . , & Li, W. (2022). Diagnosis of cardiac abnormalities based on phonocardiogram using a novel fuzzy matching feature extraction method. *BMC Medical Informatics and Decision Making*, *22*(1), 230. https://doi.org/10.1186/s12911-022-01976-6

[14] Ramadhan, M. R., Mandala, S., Ullah, R., Yafooz, W. M., & Qomaruddin, M. (2024). Enhanced identification of valvular heart diseases through selective phonocardiogram features driven by convolutional neural networks (SFD-CNN). *Jurnal Nasional Teknik Elektro*, *13*(1), 20–35. https://doi.org/10.25077/jnte.v13n1.1184.2024

[15] Ozkan, I., & Yilmaz, A. (2023). Performance of using Mel-frequency cepstrum based features in nonlinear classifiers for phonocardiography recordings. In *2023 31st European Signal Processing Conference*, 1190–1194. https://doi.org/10.23919/EUSIPCO58844.2023.10289832

[16] Malik, A., Mandala, S., & Pramudyo, M. (2023). Study of feature extraction method to detect myocardial infraction using a phonocardiogram. *Jurnal Media Informatika Budidarma*, *7*(3), 1238–1246.

[17] Nehary, E. A., & Rajan, S. (2024). Phonocardiogram classification by learning from positive and unlabeled examples. *IEEE Transactions on Instrumentation and Measurement*, *73*, 1–14. https://doi.org/10.1109/TIM.2024.3372221

[18] Suryady, Z., Rahman, A. W. A., & Oktarina, D. (2024). Human emotion classification based on phonocardiography signals (PCG). *AIP Conference Proceedings*, *2729*(1), 020002. https://doi.org/10.1063/5.0168132

[19] Debbal, S. M., & Hamza, C. (2021). Heart sounds analysis using the three wavelet transform versions the continuous wavelet transform (CWT), the discrete wavelet transform (DWT) and the wavelet packet transforms (PWT). *Journal of Cardiology Interventions*, *1*(1), 1–9. http://doi.org/03.2020/1.1003

[20] 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

[21] Jang, Y., Jung, J., Hong, Y., Lee, J., Jeong, H., Shim, H., & Chang, H. J. (2024). Fully convolutional hybrid fusion network with heterogeneous representations for identification of S1 and S2 from phonocardiogram. *IEEE Journal of Biomedical and Health Informatics*, *28*, 7151–7163. https://doi.org/10.1109/JBHI.2024.3431028

[22] Chua, K. C., Chandran, V., Acharya, U. R., & Lim, C. M. (2010). Application of higher order statistics/spectra in biomedical signals—A review. *Medical Engineering & Physics*, *32*(7), 679–689. https://doi.org/10.1016/j.medengphy.2010.04.009

[23] Rebiai, M., Berkani, M. R. A., Bentegri, M. Y., & Seghir, M. T. (2024). Mathematical morphology-based envelope detection for cardiac disease classification from phonocardiogram signals. In *2024 8th International Conference on Image and Signal Processing and their Applications*, 1–5. https://doi.org/10.1109/ISPA59904.2024.10536802

[24] Khan, F. A., Abid, A., & Khan, M. S. (2020). Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features. *Physiological Measurement*, *41*(5), 055006. https://doi.org/10.1088/1361-6579/ab8770

[25] Li, F., Tang, H., Shang, S., Mathiak, K., & Cong, F. (2020). Classification of heart sounds using convolutional neural network. *Applied Sciences*, *10*(11), 3956. https://doi.org/10.3390/app10113956

[26] Hassanuzzaman, M., Hasan, N. A., Al Mamun, M. A., Ahmed, K. I., Khandoker, A. H., & Mostafa, R. (2023). A deep learning model for recognizing pediatric congenital heart diseases using phonocardiogram signals. In *2023 Computing in Cardiology*, *50*, 1–4. https://doi.org/10.22489/CinC.2023.146

[27] Alkhodari, M., Hadjileontiadis, L. J., & Khandoker, A. H. (2024). Identification of congenital valvular murmurs in young patients using deep learning-based attention transformers and phonocardiograms. *IEEE Journal of Biomedical and Health Informatics*, *28*(4), 1803–1814. https://doi.org/10.1109/JBHI.2024.3357506

[28] Ma, S., Chen, J., & Ho, J. W. (2024). An edge-device-compatible algorithm for valvular heart diseases screening using phonocardiogram signals with a lightweight convolutional neural network and self-supervised learning. *Computer Methods and Programs in Biomedicine*, *243*, 107906. https://doi.org/10.1016/j.cmpb.2023.107906

[29] Liu, C., Springer, D., Li, Q., Moody, B., Juan, R. A., Chorro, F. J., . . . , & Clifford, G. D. (2016). An open access database for the evaluation of heart sound algorithms. *Physiological Measurement, 37*(12), 2181. https://doi.org/10.1088/0967-3334/37/12/2181

[30] Chen, Y., Su, B., Zeng, W., Yuan, C., & Ji, B. (2023). Abnormal heart sound detection from unsegmented phonocardiogram using deep features and shallow classifiers. *Multimedia Tools and Applications*, *82*(17), 26859–26883. https://doi.org/10.1007/s11042-022-14315-8

[31] Larsen, B. S., Winther, S., Nissen, L., Diederichsen, A., Bøttcher, M., Renker, M., . . . , & Schmidt, S. E. (2022). Improved pre-test likelihood estimation of coronary artery disease using phonocardiography. *European Heart Journal-Digital Health*, *3*(4), 600–609. https://doi.org/10.1093/ehjdh/ztac057

[32] Iqtidar, K., Qamar, U., Aziz, S., & Khan, M. U. (2021). Phonocardiogram signal analysis for classification of coronary artery diseases using MFCC and 1D adaptive local ternary patterns. *Computers in Biology and Medicine*, *138*, 104926. https://doi.org/10.1016/j.compbiomed.2021.104926

[33] Toulni, Y., Nsiri, B., & Belhoussine Drissi, T. (2023). Heart problems diagnosis using ECG and PCG signals and a k-nearest neighbor classifier. In *IoT Based Control Networks and Intelligent Systems: Proceedings of 3rd ICICNIS 2022*, 547–560. https://doi.org/10.1007/978-981-19-5845-8_38

[34] Tiwari, S., & Maheshwari, P. (2023). Heartbeat abnormality detection in phonocardiogram signals using wavelet time scattering and optimized KNN classification. In *2023 24th International Arab Conference on Information Technology*, 1–6. https://doi.org/10.1109/ACIT58888.2023.10453787

[35] Dastagir, J., Khan, F. A., Khan, M. S., & Khan, K. N. (2021). Computer-aided phonocardiogram classification using multidomain time and frequency features. In *2021 International Conference on Artificial Intelligence*, 50–55. https://doi.org/10.1109/ICAI52203.2021.9445235

[36] Hasan, A., & Bahri, Z. (2023). Comparative study on heart anomalies early detection using phonocardiography (PCG) signals. *International Journal of Computing and Digital Systems*, *14*(1), 1023–1040. http://dx.doi.org/10.12785/ijcds/140180

[37] Puspasari, I., Mengko, T. L., Setiawan, A. W., Adiono, T., & Pramudyo, M. (2024). Automated classification of myocardial infarction based on auscultation position using random forest. In *2024 International Conference on Artificial Intelligence in Information and Communication*, 275–280. https://doi.org/10.1109/ICAIIC60209.2024.10463493

[38] Roy, T. S., Roy, J. K., & Mandal, N. (2023). Conv-random forest based IoT: A deep learning model based on CNN and random forest for classification and analysis of valvular heart diseases. *IEEE Open Journal of Instrumentation and Measurement*, *2*, 2500717. https://doi.org/10.1109/OJIM.2023.3320765

[39] Swaminathan, S., Krishnamurthy, S. M., Gudada, C., Mallappa, S. K., & Ail, N. (2024). Heart sound analysis with machine learning using audio features for detecting heart diseases. *International Journal of Computer Information Systems and Industrial Management Applications*, *16*, 131–147.

[40] Gadde, Y., & Kumar, T. K. (2023). Prediction of heart abnormality using heart sound signals. In *Machine Intelligence Techniques for Data Analysis and Signal Processing: Proceedings of the 4th International Conference MISP 2022*, *1*, 669–680. https://doi.org/10.1007/978-981-99-0085-5_54

[41] Ramani, R. G., Ganesh, A., Balasubramanian, R., & Ramkumar, A. S. (2023). Application of phonocardiogram and electrocardiogram signal features in cardiovascular abnormality recognition. In *Computer, Communication, and Signal Processing. AI, Knowledge Engineering and IoT for Smart Systems: 7th IFIP TC 12 International Conference*, 196–209. https://doi.org/10.1007/978-3-031-39811-7_16

[42] Kong, L., Barnova, K., Jaros, R., Mirjalili, S., Snasel, V., Pan, J. S., & Martinek, R. (2024). Analysis on fetal

phonocardiography segmentation problem by hybridized classifier. *Engineering Applications of Artificial Intelligence*, *135*, 108621. https://doi.org/10.1016/j.engappai.2024.108621

[43] Moridani, M. K. (2024). *Efficient automated cardiovascular disease detection using machine learning. Research Square Preprint:rs.3.rs-4404419/v1. https://doi.org/10.21203/rs.3.rs-4404419/v1*

[44] Neili, Z., & Sundaraj, K. (2024). Addressing varying lengths in PCG signal classification with BiLSTM model and MFCC features. In *2024 8th International Conference on Image and Signal Processing and Their Applications*, 1–5. https://doi.org/10.1109/ISPA59904.2024.10536851

[45] Netto, A. N., Abraham, L., & Philip, S. (2024). HBNET: A blended ensemble model for the detection of cardiovascular anomalies using phonocardiogram. *Technology and Health Care*, *32*(3), 1925–1945. https://doi.org/10.3233/THC-231290

[46] Cao, Y., Cai, C., Li, F., Chen, Z., & Luo, J. (2025). Enabling passive user authentication via heart sounds on in-ear microphones. *IEEE Transactions on Dependable and Secure Computing, 22*(2), 1195–1209. https://doi.org/10.1109/TDSC.2024.3429574

[47] Gantert, L., Zeffiro, T., Sammarco, M., & Campista, M. E. M. (2024). Multiclass classification of faulty industrial machinery using sound samples. *Engineering Applications of Artificial Intelligence*, *136*, 108943. https://doi.org/10.1016/j.engappai.2024.108943

[48] Ranipa, K., Zhu, W.-P., & Swamy, M. N. S. (2024). A novel feature-level fusion scheme with multimodal attention CNN for heart sound classification. *Computer Methods and Programs in Biomedicine*, *248*, 108122. https://doi.org/10.1016/j.cmpb.2024.108122

[49] Roy, T. S., Roy, J. K., & Mandal, N. (2023). Early screening of valvular heart disease prediction using CNN-based mobile network. In *2023 International Conference on Computer, Electrical & Communication Engineering*, 1–8. https://doi.org/10.1109/ICCECE51049.2023.10085513

[50] Rajeshwari, B. S., Patra, M., Sinha, A., Sengupta, A., & Ghosh, N. (2023). Detection of phonocardiogram event patterns in mitral valve prolapse: An automated clinically relevant explainable diagnostic framework. *IEEE Transactions on Instrumentation and Measurement*, *72*, 4001709. https://doi.org/10.1109/TIM.2023.3240995

[51] Singh, S. A., Devi, N. D., Singh, K. N., Thongam, K., Balakrishna Reddy, D., & 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. https://doi.org/10.1007/s11042-023-17186-9

[52] Damodaran, H. K., Tripathy, R. K., & Pachori, R. B. (2024). Time-frequency-domain deep representation learning for detection of heart valve diseases using PCG recordings for IoT-based smart healthcare applications. In R. K. Tripathy & R. B. Pachori (Eds.), *Signal processing driven machine learning techniques for cardiovascular data processing* (pp. 149–165). Academic Press. https://doi.org/10.1016/B978-0-44-314141-6.00015-3

[53] Deng, E., Jia, Y., Zhu, G., & Zhou, E. (2024). Heart sound signals classification with image conversion employed. *Electronics*, *13*(7), 1179. https://doi.org/10.3390/electronics13071179

How to Cite: Popalzai, P. K., Khattak, K. S., Sohail, A. M., & Khan, Z. H. (2025). Enhancing Cardiac Health Diagnoses Through Machine Learning Analysis of Phonocardiograms (PCG). *Journal of Data Science and Intelligent Systems*, *3*(4), 312–323. https://doi.org/10.47852/bonviewJDSIS52023774