Machine Learning Analysis of Heartbeat Sounds Using Weka for Classification

Rafeeq Muhammad
School of Computing and Informatics
University of Louisiana at Lafayette
Lafayette, USA
rafeeq.muhammad1@louisiana.edu

T. Maheswara Reddy Yenumula
School of Computing and Informatics
University of Louisiana at Lafayette
Lafayette, USA
C00538416@louisiana.edu

Tanner Mergist
School of Computing and Informatics
University of Louisiana at Lafayette
Lafayette, USA
tanner.mergist1@louisiana.edu

Tytiana James
School of Computing and Informatics
University of Louisiana at Lafayette
Lafayette, USA
tytiana.james1@louisiana.edu

Ali Harimi
School of Computing and Informatics
University of Louisiana at Lafayette
Lafayette, USA
ali.harimi1@louisiana.edu

Kasem Khalil
Electrical and Computer Engineering
University of Mississippi
Mississippi, USA
kmkhalil@olemiss.edu

Ashok Kumar
School of Computing and Informatics
University of Louisiana at Lafayette
Lafayette, USA
ashok.kumar@louisiana.edu

2025 IEEE International systems Conference (SysCon) | 979-8-3315-0818-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/SYSCON64521.2025.11014657

Abstract—**Classifying heartbeat sounds is a critical step in improving cardiovascular diagnostics, especially as reliance on traditional stethoscope assessments can be limited by human subjectivity. In this study, we explored the application of machine learning techniques to the PhysioNet Heartbeat Sound Database, aiming to differentiate normal from abnormal heart sounds with both accuracy and consistency. Our approach integrates unsupervised K-means clustering, which identifies natural groupings within the data, with supervised classification models to validate these groupings and enhance predictive power. Key acoustic features—such as pitch, RMS energy, MFCCs, and formants—were extracted to provide a comprehensive representation of the heart sounds. To streamline classification, we applied the InfoGain algorithm, selecting the 50 most significant features from an initial set of 238. This process ensured that the models focused on the most relevant information, improving efficiency and interpretability. We evaluated four classifiers—SVM (SMO), Decision Tree (J48), Naive Bayes, and k-NN (IBK)—using a 10-fold cross-validation approach. SMO and J48 demonstrated strong performance, achieving accuracies of 97.4% and 96.02%, respectively, with high Kappa statistics, showcasing their reliability in distinguishing heart sound categories. These results highlight the potential of machine learning in addressing critical healthcare challenges, particularly when labeled data is scarce. Future work will focus on expanding the dataset, incorporating neural network architectures, and exploring additional clustering methods to further improve generalizability and diagnostic accuracy.**

Index Terms—**Heart Sound Classification, Machine Learning, Cardiac Diagnostics, Acoustic Feature Extraction, K-means Clustering, Support Vector Machine (SVM), Decision Tree (J48), PhysioNet Heartbeat Sound Database.**

I. Introduction

The assessment of heartbeat sounds is a crucial practice in evaluating cardiovascular health. Clinicians traditionally rely on stethoscopes to evaluate these sounds, but this process is often hindered by subjectivity and variability in human interpretation, particularly when distinguishing complex or overlapping conditions. Studies have demonstrated significant intra-observer and inter-observer variability in the manual assessment of heart sounds, leading to inconsistent diagnoses [1]. Given the increasing prevalence of cardiovascular diseases, there is a pressing need for more reliable and scalable diagnostic tools that can handle large volumes of data objectively and efficiently.

Machine learning presents a promising solution by enabling the automated analysis of heart sounds. ML models can systematically extract detailed acoustic features, facilitating a level of diagnostic consistency and accuracy that surpasses human perception. These models can differentiate normal heart sounds from abnormal ones, revealing underlying patterns that might otherwise go unnoticed, thus offering substantial potential for improving diagnostic practices. The use of machine learning can overcome the limitations of conventional stethoscope-based assessment, providing a more objective and consistent approach to heart sound analysis.

In this study, we utilized the PhysioNet Heartbeat Sound Database, which includes a wide range of both normal and abnormal heart sounds, to investigate the capabilities of ML in classifying heartbeat sounds. Our approach combines unsupervised clustering, using K-means to identify natural groupings in the data, with supervised classification methods to assess model performance in heart sound categorization. This integration of clustering and classification techniques highlights the value of ML in developing more effective and reliable

Authorized licensed use limited to: University of Louisiana at Lafayette. Downloaded on September 12,2025 at 16:35:37 UTC from IEEE Xplore. Restrictions apply.

diagnostic tools for cardiac health assessment, with potential applications in real-world clinical settings where labeled data might be limited or unavailable.

II. Related Work

Heart sound analysis using machine learning is increasingly valuable for enhancing diagnostic accuracy and consistency in cardiovascular health. Although much of the prior work focuses on supervised techniques, there is a notable gap in leveraging unsupervised methods, which can identify meaningful patterns without requiring labeled data—a critical advantage in healthcare settings where labeled data is often limited.

Research has demonstrated machine learning's potential for classifying heart sounds. For example, Mastracci et al. [2] used the PhysioNet database with an 80/20 training-validation split and supervised classifiers to differentiate normal from abnormal sounds. Building on this, our study applies a 90/10 split and integrates unsupervised K-means clustering, allowing us to observe natural groupings within unlabeled data. Furthermore, we assess classifier reliability using the Kappa statistic, which goes beyond accuracy to evaluate the consistency of predictions.

Unsupervised approaches have been pursued in other studies, though often on different datasets. Amiriparian et al. [3], for instance, used the HSS corpus dataset, which comprises 845 recordings—significantly fewer than the 3,500 recordings available in our PhysioNet dataset. The larger dataset facilitates a more comprehensive evaluation of clustering effectiveness and enhances the generalization of our findings. While some studies [7], [10] incorporate labeled data for pre-clustering, we chose a fully unsupervised approach to better simulate real-world scenarios where labeled samples may be unavailable.

Recent research underscores the growing relevance of unsupervised learning in heart sound analysis. For example, Li et al., used Beta Variational Auto-Encoders to model normal phonocardiogram signals, achieving an AUC of 0.91 for anomaly detection, showcasing the potential of unsupervised methods in detecting cardiac irregularities[11]. Similarly, Tsai et al., developed a periodicity-coded deep auto-encoder to separate heart and lung sounds without labeled data, highlighting the flexibility of unsupervised learning in complex biomedical signal processing[12]. These studies validate the effectiveness of unsupervised methods in biomedical applications, aligning with our objective of uncovering naturally occurring patterns in heart sounds.

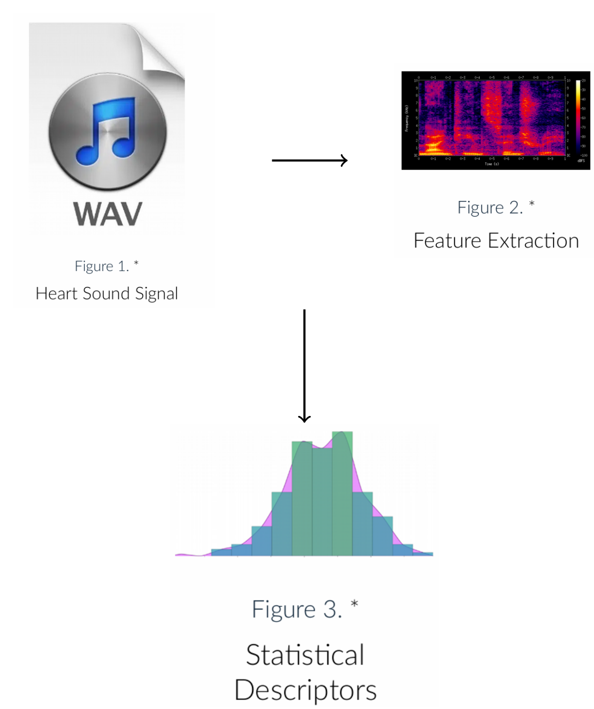
Feature extraction approaches also vary across studies. For instance, Zhu et al. [8] examined fractal dimension features to capture heartbeat complexity, while Hossein-Nejad and Nasri [9] used discrete wavelet transformation with sparse representation for time-frequency feature extraction. In contrast, we use established acoustic features—pitch, MFCCs, RMS energy, and formants—alongside statistical descriptors to achieve a robust yet interpretable analysis.

To streamline classification, we use InfoGain for feature selection, reducing the initial set of 238 features to the top 50. This selection process enhances clustering and classification performance, a crucial benefit for real-time diagnostic tools.

By combining unsupervised clustering with targeted feature selection, our study offers an innovative approach to heart sound classification, contributing to a foundation for automated cardiac diagnostics and potentially improving diagnostic reliability in healthcare.

III. Feature Extraction and Statistical Descriptors

In analyzing heartbeat sounds, the selection of acoustic features is essential for effective classification. Each feature provides insight into different sound characteristics, allowing machine learning models to interpret and distinguish between healthy and potentially abnormal cardiac patterns.



Pitch captures the fundamental frequency of a heartbeat, reflecting the dominant vibration within the sound signal. Variations in pitch can indicate irregularities in the cardiac cycle, as abnormal sounds may display unique frequency profiles, offering clues to structural anomalies or turbulent blood flow.

Root Mean Square (RMS) Energy measures sound intensity or loudness. Since RMS Energy relates to the physical strength of a heartbeat, it serves as an indicator of murmurs or other structural irregularities within the heart. Both higher and lower-than-average RMS values can suggest abnormal heart function, adding a valuable perspective beyond typical auditory analysis.

Mel-Frequency Cepstral Coefficients (MFCCs) represent the tonal and textural aspects of sounds, capturing differences that might otherwise be subtle. In heartbeat sound analysis, MFCCs reveal variances in timbre and frequency distribution, helping

Authorized licensed use limited to: University of Louisiana at Lafayette. Downloaded on September 12,2025 at 16:35:37 UTC from IEEE Xplore. Restrictions apply.

to identify patterns indicative of structural or functional abnormalities.

Formants (F1, F2, F3) represent resonant frequencies within heart sounds, helping to map sound energy distribution across frequencies. These resonances are essential for understanding the structural characteristics of the heart as they reflect how heart valves and chambers contribute to sound production.

To deepen the analysis, statistical descriptors provide a comprehensive view of each feature's distribution across recordings:

- Maximum and Minimum values set boundaries, aiding in detecting extreme variations that may indicate abnormal conditions.
- Range and Arithmetic Mean offer a summary of typical values and variability, while Root Mean Square (RMS) provides insights into average intensity.
- Standard Deviation, Skewness, and Kurtosis reveal distribution characteristics, helping to detect fluctuations and anomalies within sound patterns.
- Percentiles and Interquartile Range (IQR) offer a detailed view of data spread and typical values, minimizing the impact of outliers.

  This combination of features and descriptors enables models to interpret both fundamental and complex patterns within heartbeat sounds, providing a robust framework for identifying cardiac anomalies effectively across varied datasets.

IV. Methodology

This study employs a machine learning approach to classify heartbeat sounds, integrating both unsupervised clustering and supervised classification. The methodology is structured into three key stages: clustering, feature selection, and classification—each selected for its capability to enhance the analysis of heart sound data.

A. Clustering with K-means

We began by applying K-means clustering to identify natural groupings within the heartbeat sound data, a beneficial strategy when labeled data is limited.

Given that the dataset includes two primary categories (normal and abnormal sounds), we set the number of clusters, k=2. K-means was chosen for its efficiency and scalability, enabling the model to uncover underlying structures in extensive diagnostic datasets effectively.

B. Feature Selection through Information Gain

To streamline the analysis, we used an information gain-based feature selection technique to identify the most relevant attributes. This approach prioritized the top 50 features from an initial set of 238, selecting those that significantly differentiate between normal and abnormal heart sounds. By focusing on these high-value features, the classification process becomes more targeted and capable of detecting subtle variations in sound characteristics.

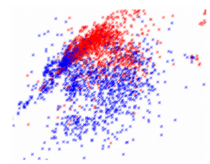


Figure 4. *

Clustering

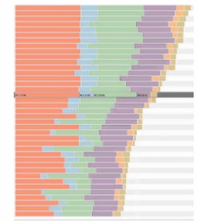


Figure 5. *

Feature Selection

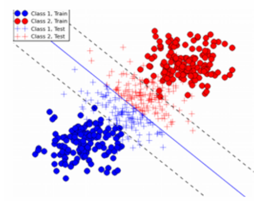
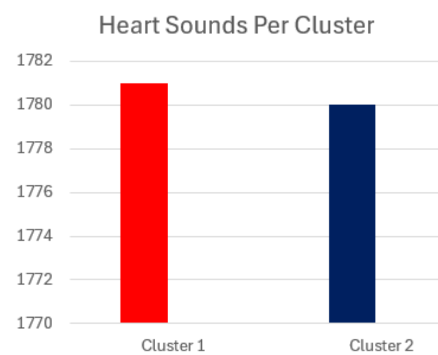


Figure 6. *

Classification



C. Classification with Supervised Models

After feature selection, we assessed four supervised classifiers, evaluating their accuracy and reliability in distinguishing heartbeat sound categories. A 10-fold cross-validation approach was applied to each classifier to obtain a robust performance estimate. 10-fold cross-validation is a well- established technique that helps ensure the reliability and generalizability of the classification models. By repeatedly training and evaluating the models on different subsets of the data, this method provides a more comprehensive and unbiased assessment of the models' performance, mitigating the risk of overfitting and enhancing the models' ability to generalize to new, unseen data. Each classifier was selected for its unique strengths in handling acoustic data:

- **Support Vector Machine (SVM) - SMO:** Implemented with Sequential Minimal Optimization (SMO), SVM is adept at managing high-dimensional data, making it suitable for identifying complex acoustic distinctions in heart

Authorized licensed use limited to: University of Louisiana at Lafayette. Downloaded on September 12,2025 at 16:35:37 UTC from IEEE Xplore. Restrictions apply.

Attribute	InfoGain Score
formant_2_min	0.791056
formant_2_percentile_90	0.791056
formant_2_percentile_25	0.791056
formant_2_percentile_10	0.791056
formant_2_rms	0.791056
formant_2_mean	0.791056
formant_2_max	0.791056
formant_2_percentile_75	0.791056
formant_2_percentile_50	0.791056
formant_1_mean	0.441398
formant_1_percentile_75	0.441398
formant_1_percentile_90	0.441398
formant_1_min	0.441398
formant_1_percentile_50	0.441398
formant_1_rms	0.441398
formant_1_max	0.441398
formant_1_percentile_25	0.441398
formant_1_percentile_10	0.441398
mfcc_3_percentile_10	0.357228
mfcc_3_mean	0.350974
mfcc_3_percentile_25	0.34747
mfcc_3_rms	0.342134
mfcc_3_percentile_50	0.330012
mfcc_3_min	0.305323
mfcc_3_percentile_75	0.294518
mfcc_3_max	0.268336
mfcc_10_percentile_50	0.186855
mfcc_10_mean	0.181587

Fig. 8. Key Features selected

sounds.

- **Decision Tree - J48:** The J48 algorithm offers an interpretable model structure, which is advantageous in medical applications where transparent decision-making is essential. To reduce the risk of overfitting, we set a minimum leaf size of two instances.
- **k-Nearest Neighbors (k-NN) - IBK:** Configured with k=5, the k-NN classifier uses proximity to classify cases, balancing sensitivity to local patterns with overall model stability. This technique can effectively capture subtle differences in sound data, as it considers similarities among neighboring cases.
- **Naive Bayes:** Employed with default settings, Naive Bayes provides a simple yet effective baseline model. Its computational efficiency is valuable, especially in scenarios where feature independence assumptions reasonably hold.

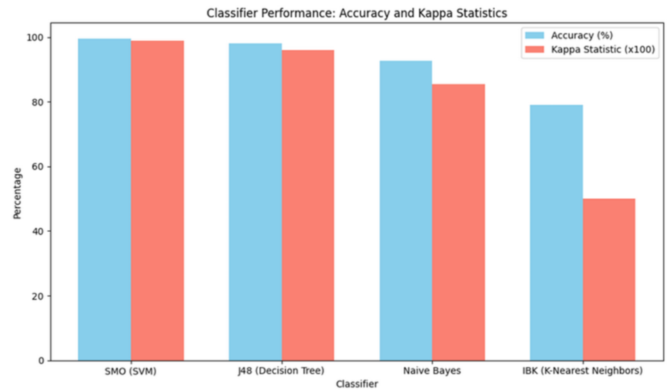


Fig. 9. Classifier Performance accuracy and Kappa statistics

Each classifier's performance was assessed not only by accuracy but also by the Kappa statistic, which measures consistency beyond chance. This dual evaluation offers a well-rounded view of each model's reliability in classifying heartbeat sounds accurately.

V. Results

Our analysis shows that the SMO and J48 classifiers stand out in terms of accuracy and reliability, with SMO achieving an accuracy of 98.43% and J48 close behind at 97.02%. These classifiers also show high Kappa statistics, indicating a strong alignment between predictions and actual classifications.

The Kappa statistic is a measure of classifier consistency, evaluating the agreement between predicted labels and true labels, while accounting for chance agreement. The high Kappa values of 0.9787 for SMO and 0.9505 for J48 demonstrate that these models make consistently reliable predictions, providing further validation of their strong performance. Kappa values range from -1 to 1, with 0 indicating no agreement beyond chance and 1 indicating perfect agreement. The Kappa values obtained for the SMO and J48 classifiers are very close to 1, indicating almost perfect agreement between the model predictions and the true labels, which confirms the models' reliability and consistency.

In comparison, the Naive Bayes and IBK (k-NN) classifiers demonstrated more moderate performance. Naive Bayes attained an accuracy of 92.71% and a Kappa statistic of 0.8542, providing a reasonable baseline but not quite reaching the robustness of SVM and Decision Tree. The k-NN classifier achieved the lowest accuracy, with 79.02% and a Kappa statistic of 0.5013. This lower performance may be attributed to the sensitivity of distance-based models like k-NN to complex or overlapping clusters within high-dimensional data, where it can struggle to separate classes effectively. Implementing dimensionality reduction methods such as Principal Component Analysis (PCA) in future work could help improve k-NN's effectiveness by reducing data complexity.

Authorized licensed use limited to: University of Louisiana at Lafayette. Downloaded on September 12,2025 at 16:35:37 UTC from IEEE Xplore. Restrictions apply.

One limitation observed is the potential overlap in the unsupervised clusters, which may introduce classification challenges. Certain heart sound patterns, whether normal or abnormal, may share acoustic similarities, making boundaries difficult to distinguish—particularly for k-NN. Addressing this overlap could involve more sophisticated feature selection or ensemble techniques to blend model strengths and improve classification consistency across boundary cases.

Compared with existing studies, our approach offers a nuanced view by combining unsupervised clustering with supervised classification, allowing us to work with both labeled and unlabeled data. This approach aligns well with real-world medical scenarios where data labels may not always be accessible. Further research might benefit from testing on larger, more diverse datasets to enhance the generalizability of these models for various cardiac conditions, which would support broader applications in clinical settings.

VI. Conclusion and Future Work

This research underscores the potential of integrating clustering and classification methods to enhance the reliability of heartbeat sound analysis, which is essential for supporting consistent cardiovascular diagnostics. Our approach of pairing unsupervised clustering with classification models, especially SVM and Decision Tree, demonstrated effective accuracy and reliability in distinguishing normal and abnormal heart sounds.

Looking ahead, further validation of the unsupervised clustering method through comparisons with labeled data can provide deeper insights. By employing metrics like the Adjusted Rand Index and Silhouette Score, we aim to assess how closely the clusters reflect actual heart sound categories, refining the clustering approach to better align with natural data patterns.

Expanding the dataset with broader and more diverse heart sound recordings is another priority, as this will enable our models to generalize more effectively across different cardiac conditions, strengthening their diagnostic applicability. Additionally, techniques for dimensionality reduction, such as Principal Component Analysis (PCA), could help simplify the feature space, optimizing processing for distance-based classifiers like k-NN.

Lastly, exploring more advanced machine learning architectures, such as neural networks, may reveal complex patterns within heart sounds that simpler models might overlook. These future directions reinforce our goal of creating dependable, automated tools for cardiovascular monitoring, which can aid clinicians in making timely, data-informed assessments.

References

[1] Emmanuel, B. S. (2012). A review of signal processing techniques for heart sound analysis in clinical diagnosis. Journal of Medical Engineering Technology, 36(6), 303–307. https://doi.org/10.3109/03091902.2012.684831

[2] N. Mastracci, F. Derakhshan, E. R. Sykes and D. Khan, "Classification of Heart Sounds Using Machine Learning," 2023 IEEE International Conference on Digital Health (ICDH), Chicago, IL, USA, 2023, pp. 205-207, doi: 10.1109/ICDH60066.2023.00037.

[3] S. Amiriparian, M. Schmitt, N. Cummins, K. Qian, F. Dong and B. Schuller, "Deep Unsupervised Representation Learning for Abnormal Heart Sound Classification," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, pp. 4776-4779, doi: 10.1109/EMBC.2018.8513102.

[4] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in Proceedings of the 14th Python in Science Conference, 2015, pp. 18-25.

[5] The pandas development team, "pandas-dev/pandas: Pandas," Zenodo, vol. 1, Feb. 2020. Available: https://doi.org/10.5281/zenodo.35091

[6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10-18, 2009.

[7] S. K. Padhy, A. Mohapatra, and S. Patra, "Heart Sound Classification Based on MFCC Feature Extraction and Long-Short Term Neural Networks," 2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC), Bhubaneswar, India, 2023, pp. 1-6, doi: 10.1109/ICAIHC59020.2023.10431453.

[8] C. Zhu et al., "Less is More: A Novel Feature Extraction Method for Heart Sound Classification via Fractal Transformation," 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 2023, pp. 1-4, doi: 10.1109/EMBC40787.2023.10340710.

[9] Z. Hossein-Nejad and M. Nasri, "Heart Sound Classification based on Discrete Wavelet Transform and Group-based Sparse Features of PCG Signal," 2023 14th International Conference on Information and Knowledge Technology (IKT), Isfahan, Iran, 2023, pp. 218-222, doi: 10.1109/IKT62039.2023.10433012.

[10] H. Zhang, P. Zhang, Z. Wang, L. Chao, Y. Chen, and Q. Li, "Multi-Feature Decision Fusion Network for Heart Sound Abnormality Detection and Classification," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, pp. 1386-1397, March 2024, doi: 10.1109/JBHI.2023.3307870.

[11] S. Li, et al., "Unsupervised Heart Abnormality Detection Based on Phonocardiogram Analysis with Beta Variational Auto-Encoders," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 8353-8357.

[12] K.-H. Tsai, et al., "Blind Monaural Source Separation on Heart and Lung Sounds Based on Periodic-Coded Deep Autoencoder," IEEE Journal of Biomedical and Health Informatics, vol. 24, 2020, pp. 3203-3214.

[13] PhysioNet/CinC Challenge 2016, "Heart Sound Classification Dataset," PhysioNet, 2016. [Online]. Available: https://physionet.org/content/challenge-2016/

Authorized licensed use limited to: University of Louisiana at Lafayette. Downloaded on September 12,2025 at 16:35:37 UTC from IEEE Xplore. Restrictions apply.