Rhythm of Life: Revolutionizing cardiac diagnostics with Aldriven lub-dub analysis

Aryaman Kedia¹ and Nirupma Singh²

¹Grade 11, Dhirubhai Ambani International School, Mumbai, Maharashtra, 400098.

²Department of Biological Sciences and Engineering, Netaji Subhas University of Technology (Formerly NSIT), Dwarka, Delhi, India.

KEYWORDS. Heart sounds, Machine learning, MFCC, CVDs, LSTM.

BRIEF. This study leverages machine learning algorithms to analyze "lub-dub" heart sounds from digital stethoscopes, achieving 98% accuracy in detecting abnormal heartbeat-related heart diseases, offering a breakthrough in non-invasive cardiac diagnostics.

ABSTRACT. Heart disease is the leading cause of global mortality. Use of capable and robust machine learning (ML) algorithms could help to greatly lower mortality rates due to these disorders. While most current research focuses on electrocardiogram (ECG) data, this work addresses the gap in the analysis of "lub-dub" heartbeat sound patterns for diagnosis. This research study aims to improve cardiac diagnosis by leveraging ML algorithms to analyze heartbeat sounds collected from digital stethoscopes. It accurately identifies abnormal heartbeat related disorders such as Valvular heart disease, septal defects, congenital heart defects, Arrhythmias, electrolyte imbalances, Heart failure (S3), hypertension and left ventricular hypertrophy (S4), by analyzing the "lub-dub" heart sounds through machine learning, offering higher accuracy and early disease detection. This study utilizes heart sound datasets and employs noise reduction, exploratory data analysis (EDA), and feature extraction, focusing on Mel-frequency cepstral coefficients (MFCC) features. Several ML models, SVM, Random Forest, and boosting, were trained and evaluated for robustness on the basis of different metrics like accuracy (fraction of correctly classified instances out of all instances), precision (fraction of correct predicted positives), recall (fraction of actual positives correctly predicted by the model), f1-score (harmonic mean of precision and recall) and auc-roc score (probability that a random chosen positive instance is ranked higher than a negative instance). Additionally, further refinement with the use of advanced models such as Long Short-Term Memory (LSTM) improved accuracy greatly to 98%. The findings of this study demonstrate that ML algorithms can significantly enhance abnormal heartbeat related heart disease detection, offering a promising step forward in non-invasive cardiac diagnostics.

INTRODUCTION.

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, accounting for approximately 17.9 million deaths annually. Early detection is crucial for effective management and treatment of these conditions ¹. Traditional auscultation, relying on physicians' auditory skills to interpret heart sounds, has been a cornerstone in diagnosing heart anomalies. However, this method is inherently subjective and susceptible to variability in clinical experience and environmental noise, potentially leading to misdiagnoses ².

Advancements in digital stethoscope technology have revolutionized cardiac auscultation by enabling the recording and digitization of heart sounds. These devices facilitate the application of machine learning (ML) algorithms to analyze heart sounds, offering a more objective and accurate assessment of cardiac function ³. By capturing high-fidelity audio data, digital stethoscopes allow for detailed analysis of heart sounds, including the fundamental 'lub-dub' (S1 and S2) sounds associated with valve closures ⁴. Machine learning models, particularly deep learning architectures, have demonstrated significant potential in classifying heart sounds and detecting abnormalities. The integration of digital stethoscopes with ML algorithms offers several advantages such as Noise Reduction⁵; Objective Analysis⁵; Early Detection⁶; Remote Monitoring⁷. generalizability of these models. For example, a study on heart sound classification in noisy environments proposed a combination of linear and logarithmic spectrogram-image features input into a residual convolutional neural network (ResNet). This approach achieved an area under the ROC curve (AUC) of 91.36% and an F1 score of 84.09%, demonstrating improved performance in challenging conditions⁸. Despite these advancements, challenges persist, particularly in developing robust algorithms capable of accurately analyzing and multi-classifying heart sounds across diverse populations and clinical settings. The fusion of digital stethoscope technology with machine learning algorithms represents a significant advancement in cardiac diagnostics. By leveraging high-quality heart sound data and sophisticated analytical models, this approach holds promise for enhancing the accuracy and reliability of heart disease detection, ultimately improving patient outcomes ^{9,10}.

This research addresses a critical gap in cardiac diagnostics by utilizing digital stethoscope data in combination with machine learning (ML) algorithms to improve the accuracy of heart disease detection. By offering a non-invasive, highly sensitive, and practical diagnostic tool, this work presents significant advancements over traditional methods. The proposed approach mitigates such limitations through advanced ML algorithms such as LSTM, trained on large datasets, enabling the detection of subtle patterns that are imperceptible to the human brain. Through the construction of a dedicated dataset, implementation of noise reduction techniques, exploratory data analysis (EDA), and the application of advanced ML and LSTM models, this research successfully bridges existing gaps in cardiac diagnostics. The high accuracy of 98% achieved by analyzing "lub-dub" heart sounds highlights the potential of combining digital stethoscope technology with state-of-the-art ML algorithms to enhance cardiovascular disease detection. This integration represents a transformative step forward in non-invasive cardiac healthcare, demonstrating the potential to significantly improve diagnostic accuracy and patient outcomes.

MATERIALS AND METHODS.

Data Collection.

The dataset of different types of heartbeat sounds: normal, murmur, abnormal, artifact, extrastole, and extrahls, was sourced from various sources, including the Pascal Classifying Heart Sounds Challenge Dataset, The Heart Sounds and Murmur Library (HSML), iStethoscope Pro Dataset, Cardiac Auscultatory Sounds (CASC) Dataset, and Physionet. The collected data was of lub-dub heartbeat sounds recorded from a digital stethoscope or the iStethoscope Pro iPhone application and stored in the WAV file format. These recordings were categorized into different classes, such as normal, murmur, abnormal, artifact, extrastole, and extrahls as shown in Fig 1.



Figure 1. A pie chart showing distribution of different categories of audio files.

In total, the dataset comprised of 835 audio files with 587 labelled audio files and 248 unlabelled audio files.

Data processing and augmentation.

Initially, noise reduction was performed for all audio files to eliminate sounds of all frequencies other than the required heart sounds, ensuring the clarity of the heart sounds. Subsequently, exploratory data analysis was carried out on each type of heartbeat sound category. Three different visualizations were observed for the heart beat sound audio files, i.e. audio waveform, audio spectrum, and audio spectrogram. Sound is the pressure of air propagates to our ear. The waveform graph shows the wave's displacement, and how it changes over time, gotten from a sound sensor that can detect sound waves and converts it to electrical signals. The waveform graph of normal and abnormal heartbeat sound is shown in Fig 2. A sound spectrum is a representation of a short sample of a sound in terms of the amount of vibration at each individual frequency. It is usually presented as a graph of either power or pressure as a function of frequency. It expresses the frequency composition of the sound and is obtained by analyzing the sound. The spectrum difference of normal and abnormal heartbeat sound is shown in Fig 3. Class weighting was incorporated in the dataset in the model training to address potential class imbalance and to compensate for the underrepresented categories. Furthermore, to enhance model generalization and reduce class imbalance effects, data augmentation techniques were applied, including: noise addition, time stretching and pitch shifting. These augmentation techniques helped improve classification performance, particularly for underrepresented heartbeat categories.



Figure 2. Audio waveform graphs of Normal and Abnormal heartbeat sound waves. (A) It represents the audio waveform of normal heartbeat sound. (B) It represents the audio waveform of abnormal heartbeat sound. The x axis represents the time and y axis represents the amplitude of the sound.



Figure 3. Audio spectrum graphs of Normal and Abnormal heartbeat sound waves. (A) It represents the audio spectrum of normal heartbeat sound. (B) It represents the audio spectrum of abnormal heartbeat sound. The x axis represents the frequency and y axis represents the magnitude.

Feature Engineering.

The raw audio signal cannot be taken as input to the ML model because there will be a lot of noise in the audio signal. It was observed that extracting features from the audio signal and using it as input to the base model will produce much better performance than directly considering raw audio signal as input. Mel-frequency cepstral coefficient (MFCC) is the widely used technique for extracting the features from the audio signal. In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. MFCCs are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-aspectrum"). The difference between the cepstrum and the melfrequency cepstrum is that in the MFC, the frequency bands are equally spaced on the mel scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal spectrum. This frequency warping can allow for better representation of sound. The audio files were in the range of 0.2 - 0.9 seconds of duration. 52 MFCC features were extracted for all the audio files. The MFCC representation of normal and abnormal labelled audio files is shown in Fig 4.



Figure 4. A representation of MFCC features. (A) The MFCC feature graph of normal audio file. (B) The MFCC feature graph of abnormal audio file. The x axis represents the time and y-axis represents the mfcc coefficients.

Model Development.

Python programming language and its libraries were used in Spyder IDE. Different machine learning algorithms, namely Support Vector Machines (SVM), Logistic Regression (LR), Random Forests (RF), Adaptive boosting (ADB) and XGBoost (XGB), were built using scikit learn library of Python to train classification and regression models. For classification, SVM, RF, XGboost, and AdaBoost models were built. For regression, LR, RF, SVM, XGboost and AdaBoost models were built. The models were initially built on with three classes labels (normal, murmur, and artifact) followed by model building with five classification Iabels (normal, murmur, and artifact, extrastole and extrahls). The classification ML models were followed by building of LSTM model using TensorFlow framework. The LSTM model is part-

Table 1. Evaluation scores of all the classification ML models and LSTM models for both 3 and 5 categories of heart-beat sound label

Model Name /	SVM		XGB		ADB		RF		LSTM	
Evaluation Metrics	3 Classes	5 classes	3 clas-	5 clas-						
			ses							
Accuracy	0.82	0.74	0.93	0.89	0.66	0.55	0.90	0.85	0.98	0.95
Precision	0.83	0.743	0.93	0.90	0.59	0.36	0.90	0.87	0.94	0.92
Recall	0.82	0.74	0.93	0.89	0.66	0.55	0.90	0.85	0.94	0.92
F1 Score	0.80	0.72	0.92	0.89	0.60	0.42	0.90	0.84	0.94	0.92
AUC ROC	N/A	N/A	0.98	0.98	0.65	0.58	0.97	0.97	N/A	N/A

Table 2. Evaluation scores of all the regression ML models for both 3 and 5 categories of heart-beat sound labels

Model Name / Evaluation Metrics	SVM		XGB		ADB		RF		LR	
	3 Classes	5 classes								
MAE	0.54	0.54	0.32	0.32	0.61	0.61	0.36	0.36	0.40	0.40
RMSE	0.81	0.81	0.49	0.49	0.77	0.77	0.55	0.55	0.82	0.82

icularly suited for time-step based sequential data like heart sounds, which exhibit temporal dependencies. Evaluation metrics such as accuracy, precision, F1 score, recall, ROC AUC were computed to evaluate classification models that provided insights into the models' ability to accurately classify and distinguish between heart sound categories.. Mean absolute error (MAE), and root mean squared error (RMSE) were computed to evaluate the performance of regression models to quantify the models' performance in predicting risk scores.

RESULTS.

Classification model performance.

For the classification models focusing on 3-categories of heart sounds (normal, abnormal, and device-related errors), the XGBoost classifier achieved an accuracy of 93.1%, demonstrating its effectiveness in distinguishing heart sound categories as shown in Table 1. However, when the classification task was extended to 5-categories (normal, abnormal, and specific abnormalities such as murmurs, arrhythmias, and device-related errors), a slight drop in accuracy was observed. The XGBoost model achieved an accuracy of 89%, maintaining its position as the best-performing model in this category. To ensure statistical robustness, a 95% confidence interval (CI) was computed for the reported 98% AUC-ROC, yielding a range of [95.46%, 100%]. Additionally, 5-fold cross-validation was conducted, resulting in a mean AUC-ROC of 97.11% (±1.16%), demonstrating consistent performance across training subsets. These measures confirm the model's reliability on unseen data.

Regression model performance.

For regression models, designed to predict a risk score for heart diseases, the XGBoost model performed best with a mean absolute error (MAE) of 0.33, indicating its robustness in predicting numerical risk scores with high precision as shown in Table 2. No significant changes were observed in the MAE values for the regression models under the 5-category setup. The findings highlight that while the 3-category classification is reliable for basic differentiation, the 5-category classification provides a more detailed analysis of specific heart sound abnormalities at a slight trade-off in accuracy. The evaluation scores of each ML model for classification and regression models are shown in Table 1 and 2, respectively.

LSTM Model Architecture. The deep learning model used in this study is a hybrid CNN-LSTM network designed to extract both spatial and temporal features from heart sound data. The architecture consists of three 1D convolutional layers (2048, 1024, and 512 filters, kernel size = 5, ReLU activation), each followed by max pooling and batch normalization to enhance feature extraction and stability. These extracted spatial features are then processed through two stacked LSTM layers (256 and 128 units), enabling the model to capture temporal dependencies in the heart sound sequences. The fully connected layers consist of 64 and 32 dense units, each followed by dropout (0.5 rate) to prevent overfitting. The final softmax output layer classifies heart sounds into three categories (normal, abnormal, and device error). The model contains 14,130,371 total parameters, of which 14,123,203 are trainable. This architecture significantly contributed to the model's 98.66% accuracy for 3-category classification and 95.77% for 5-category classification, as shown in Table 1.

DISCUSSION.

This study explored the application of ML and LSTM models for the analysis of "lub-dub" heart sounds, aiming to detect and classify abnormal heartbeat-related disorders such as valvular heart disease, septal defects, congenital heart defects, arrhythmias, electrolyte imbalances, heart failure (S3), hypertension, and left ventricular hypertrophy (S4)¹¹. The primary objective was to enhance diagnostic accuracy by leveraging advanced computational techniques to analyze heart sound patterns, a critical yet underutilized area in cardiovascular diagnostics. The methodology involved extracting MFCC features from heart sound recordings during EDA and employing ML models such as SVM, RF, and XGB, alongside an LSTM model. These models were evaluated for their robustness using standard metrics such as accuracy, precision, recall, F1-score, and AUC-ROC.

The superior performance of the LSTM model underscores the importance of leveraging temporal dependencies and sequential patterns in heart sound data, which are often overlooked by traditional ML methods ¹². Integrating this algorithm with digital stethoscopes could enable clinicians to directly analyze heart sounds during routine check-ups, enhancing diagnostic accuracy and efficiency. Such an approach addresses the limitations of traditional stethoscopes, which are prone to subjective variability depending on the physician's hearing acuity ¹³. The ability to detect CVDs at an early stage is particularly crucial, as timely intervention can prevent the progression of these conditions and reduce mortality. For instance, early detection of arrhythmias or valvular disorders could prompt immediate medical attention, avoiding life-threatening complications ¹⁴.

The regression model predicted a continuous output by mapping extracted acoustic features (MFCCs, spectral, and time-domain properties) to a severity indicator, even though predefined risk scores were not provided during training. This enabled the model to infer relative risk levels based on learned feature distributions from labeled heartbeat sound data. Thus, the regression model provided a continuous risk estimation based on extracted acoustic features, complementing the classification framework by enabling severity assessment of abnormal heart sounds. Future research of this project is focused on designing a hardware easily accessible to common people for autonomous data collection and analysis of heart-beat sounds with the incorporation of the model in the hardware which can help people to get the risk score of disease at home. Thus, this study demonstrates the potential of machine learning and deep learning models, particularly LSTM, in transforming cardiac diagnostics. By leveraging heart sound data collected through digital stethoscopes, the proposed framework achieves exceptional accuracy and reliability in detecting cardiovascular diseases. While challenges remain, the integration of such models into clinical workflows holds immense promise for improving early disease detection, enhancing patient outcomes, and addressing global disparities in healthcare access.

ACKNOWLEDGMENTS.

AK would like to acknowledge Aashna Saraf, Founder of CreatED for providing valuable feedback and guidance throughout the project.

REFERENCES

- 1 WHO. Cardiovascular diseases (CVDs). (2021), doi:https://www.who.int/health-topics/cardiovascular-diseases#tab= tab 1.
- 2 Krittanawong, C., Zhang, H., Wang, Z., Aydar, M. & Kitai, T. Artificial Intelligence in Precision Cardiovascular Medicine. J Am Coll Cardiol 69, 2657-2664 (2017).
- 3 Leng, S. *et al.* The electronic stethoscope. *BioMedical Engineering OnLine* **14**, 66, (2015).
- 4 Giacobbe, D. R. *et al.* Early Detection of Sepsis With Machine Learning Techniques: A Brief Clinical Perspective. *Front Med* (*Lausanne*) 8, 617486 (2021).
- 5 Prince, J. *et al.* Deep Learning Algorithms to Detect Murmurs Associated With Structural Heart Disease. *J Am Heart Assoc* **12**, e030377 (2023).
- 6 Partovi, E., Babic, A. & Gharehbaghi, A. A review on deep learning methods for heart sound signal analysis. *Frontiers in Artificial Intelligence* 7 (2024).

- 7 Chang, V., Bhavani, V. R., Xu, A. Q. & Hossain, M. A. An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Analytics* 2, 100016 (2022).
- 8 Li, F., Zhang, Z., Wang, L. & Liu, W. Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning. *Front Physiol* 13, 1084420 (2022).
- 9 Zhao, Q. *et al.* Deep Learning in Heart Sound Analysis: From Techniques to Clinical Applications. *Health Data Science* **4**, 0182
- 10 Omarov, B. *et al.* Digital Stethoscope for Early Detection of Heart Disease on Phonocardiography Data. *International Journal of Advanced Computer Science and Applications* **14** (2023).
- 11 Reed, T. R., Reed, N. E. & Fritzson, P. Heart sound analysis for symptom detection and computer-aided diagnosis. *Simulation Modelling Practice and Theory* 12, 129-146 (2004).
- 12 Xie, F. *et al.* Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. *Journal of Biomedical Informatics* **126** (2022).
- 13 Maddula, R. *et al.* The role of digital health in the cardiovascular learning healthcare system. *Front Cardiovasc Med* **9**, 1008575 (2022).
- 14 Clur, S. A. & Bilardo, C. M. Early detection of fetal cardiac abnormalities: how effective is it and how should we manage these patients? *Prenat Diagn* **34**, 1235-1245 (2014).



Aryaman Kedia is a student at Dhirubhai Ambani International School in Mumbai, Maharashtra. He participated in the research internship through the CreatED Research Program.