ROBUST HYBRID MACHINE LEARNING MODEL FOR HEART SOUND SIGNAL CLASSIFICATION

Dingari Jahnavi
CSE-AIE
Amrita School of Computing, Chennai
jahnavidingari04@gmail.com

Mona A
CSE-AIE
Amrita School of Computing, Chennai
monaananth07@gmail.com

Sandeep Pulata
CSE-AIE
Amrita School of Computing, Chennai
sandeeppulata@gmail.com

Sasank Sami
CSE-AIE
Amrita School of Computing, Chennai
sasanksami21@gmail.com

Bharadwaj V
CSE-AIE
Amrita School of Computing, Chennai
bharadwaj2193@gmail.com

Dr. Muthulakshmi M.
ECE
Amrita School of Engineering, Chennai
m_muthulakshmi@ch.amrita.edu

***Abstract*— The field of cardiac sound signals is critical to the timely diagnosis of cardiac problems in modern medicine. Existing techniques, such as Random Forests and Support Vector Machines, face limitations in capturing complex temporal correlations and hierarchical features present in these signals. Their issues include a failure to adjust to minute changes, a problem to deal with non-linear connections, and a challenge to cope with the dynamic character of cardiac sounds.The purpose of this technique is to deliberately address the limitations of traditional machine learning (ML) models by combining Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNN). Heart sound waves can be found to include complex patterns thanks to the initial steps of Exploratory Data Analysis (EDA), which include waveform, spectrum, and spectrogram visualization. MFCCs, or robust Mel-frequency cepstral coefficients, are essential to feature extraction.**

The hybrid model combines Bidirectional LSTM layers, Conv1D layers, and extra dense layers to optimize the advantages of both architectures. The model's capacity to classify cardiac sound signals is improved by this integration, which allows it to differentiate between extrahls, murmurs, artifacts, normal, extra sole, and unlabeled data. A model with 70% accuracy is the product of rigorous training and thorough accuracy assessments. In summary, this method significantly contributes to the field of medical signal processing based on heart sound signals. These highly precise models serve as valuable tools for improving cardiac healthcare diagnostics and treatment actions, particularly in the realm of heart sound signal processing.

<u>Keywords</u> :-

Convolutional neural networks (CNN), long short-term memory (LSTM) networks, exploratory data analysis (EDA), Mel-frequency cepstral coefficients (MFCCs), Bidirectional LSTM layers, Conv1D layers, spatial information.

I. INTRODUCTION

Cardiovascular illnesses consistently rank among the leading causes of death worldwide, posing a serious threat to global health. This ongoing problem highlights the essential need for cutting-edge diagnostic equipment and prompt solutions

In addressing the intricate task of classifying cardiac sound signals, traditional machine learning models, such as Support Vector Machines (SVM) and Random Forests, have confronted inherent limitations [1]. These conventional models often struggle to capture the complex temporal dependencies and hierarchical features present in heart sound signals, hindering their ability to provide accurate and nuanced classifications. Limitations include a lack of adaptability to subtle variations in signal patterns, inefficiencies in handling non-linear relationships within the data, and challenges in accommodating the dynamic nature of cardiac sounds.

This study sets out on a determined quest to create a reliable and accurate heart sound signal classification model in answer to these urgent problems. Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs) were developed as a powerful hybrid model that seamlessly combine Machine Learning (ML) and Signal Processing at the heart of methodology [2]. uses the exceptional Mel-frequency cepstral coefficients (MFCCs) for feature extraction, while radix-2 Fast Fourier Transform (FFT) for spectrum production and Short-Time Fourier Transform (STFT) are used to gain spectral insights.

The main contribution, the development of a hybrid model combining LSTM and CNN, is a sign of hope for the field of cardiovascular healthcare. This combination of LSTM and CNN has the potential to change the field of early cardiovascular abnormality detection [3]. This hybrid model has a significant impact. Healthcare practitioners now have a powerful tool for the early detection of cardiovascular diseases thanks to the ability to detect small irregularities within heart sound signals.

The heart sound dataset, a wealth of acoustic recordings that reflect the complex tapestry of clinical practice, is essential to the success of the project. This dataset demonstrates the dedication to recording the whole range of heart sound patterns, including normal heart sounds, murmurs, artifacts, extrasystoles, extrahls and more. Dataset curation is a labor of love with the goal of ensuring that the model can meet the demands of real-world clinical scenarios with uncompromising accuracy and understanding. This journey encompasses an in-depth exploration of the dataset, the intricacies of MFCC-based feature extraction, the architecture of the hybrid model, and a rigorous evaluation that underscores the model's reliability.

The CNN model, the LSTM model, and the hybrid LSTM-CNN model are three different models that are compared in the project [4]. Each model has its benefits and drawbacks. The CNN model performs well in extracting spatial characteristics from heart sound data, making it useful in recognising patterns associated with particular heart sound problems. The LSTM model, in contrast, is built to capture temporal relationships within the signals, improving its capacity to spot subtle patterns over time. The hybrid model, which combines LSTM and CNN, intends to fully utilize the advantages of both architectures for heart sound categorization. This comparative analysis provides important new insights to the field of cardiovascular healthcare diagnostics by highlighting trade-offs and potential synergies between different models.

In conclusion, cardiovascular illnesses continue to be a global health concern that calls for creative diagnostic approaches. Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs), which form the foundation of methodology, combine the strengths of Signal Processing and Machine Learning (ML). Mel-frequency cepstral coefficients (MFCCs) are used for feature extraction, and the Short-Time Fourier Transform (STFT) and spectrum generating techniques are used to reveal spectral information. A significant development is the hybrid LSTM-CNN model, which enables the early diagnosis of cardiovascular irregularities in heart sound signals.

II . LITERATURE REVIEW

1) The paper ‘Artificial Intelligence Algorithm for Heart Disease Diagnosis using Phonocardiogram Signals’ initiatives greatly by progressing AI-based cardiovascular diagnostics [5]. It relates as it prioritizes MFCCs and EDA while integrating ML, signal processing, LSTM, and CNN for heart sound classification. It utilizes RBF and BPN techniques to achieve high accuracy when diagnosing cardiac problems from PCG signals using ANN algorithms, exhibiting a variety of AI methodologies for early diagnosis. From this paper we got a complete idea of using MFCC in feature extraction.

2) The paper ‘ECG Arrhythmia Classification Using STFT’-based Spectrogram and Convolutional Neural Network [6]. As we are examining a cutting-edge hybrid model for classifying heart sound signals while highlighting the interplay between machine learning, signal processing, and neural networks. It demonstrates growing intelligent algorithms in cardiovascular health monitoring by introducing a highly accurate 2D-CNN model for ECG arrhythmia classification using time-frequency. spectrograms. By reading this paper, we have got to know about the STFT algorithm which we have used to plot a spectrogram.

3) The paper ‘Heart Diseases diagnosis using an intelligent algorithm based on PCG signal analysis’ uses machine learning and enhanced signal processing for the identification of cardiac disease [7]. While it proposes a four-stage technique including DWT for feature extraction and an ANN with 97% accuracy, exhibiting their potential to advance cardiac healthcare diagnostics, the first stresses a hybrid model with CNN and LSTM networks, utilizing MFCCs and EDA. We got to know about CNN and LSTM models and their usages.

4) The paper ‘Detection of Valvular Heart Diseases Using Fourier Transform and Simple CNN Model’ uses a hybrid model fusing ML, signal processing, LSTM, and CNN for heart sound classification, and the second suggests a low-complexity, automated system using a multiclass CNN model for diagnosing cardiac illness via PCG signals [10]. Both emphasize automated cardiac health diagnosis. Both emphasize how cutting-edge technology may increase the precision and effectiveness of cardiac health diagnostics. From this paper we got to know about hybrid models.

5) The paper ‘MFCC Feature Extraction and KNN Classification in ECG Signals’ uses a hybrid model that combines machine learning, signal processing, and CNN for precise heart sound categorization, focusing on EDA and MFCCs [11]. The importance of feature-rich techniques in cardiovascular diagnostics is highlighted by the second study, which focuses on ECG data and uses MFCC, Discrete Wavelet transformation, and KNN for accurate diagnosis of normal and pathological states.

6) The paper ‘Heartbeat Sound Signal Classification Using Deep Learning’ utilizes cutting-edge tools for cardiac signal processing, including CNNs and LSTMs [14]. Together, they advance the early diagnosis of cardiovascular illnesses. The first highlights a hybrid model with EDA and MFCCs for heart sound classification, while the second uses band filtering, down-sampling, and RNN with LSTM.

7) The paper ‘ECG Classification Using Wavelet Packet Entropy and Random Forests’ uses a hybrid machine learning model to identify cardiovascular disorders, placing special emphasis on the accurate classification of heart sound signals using MFCCs [16]. The second on the other hand, highlights several methods for evaluating cardiac health and concentrates on ECG-based heart disease detection using WPE and RF with an inter-patient scheme.

III . METHODOLOGY

To create an efficient model, we have adhered to a unique process at every stage. A number of crucial steps are included in the process in order to fully analyze heart sound information. To comprehend the nature and properties of the data, the procedure first focuses on exploratory data analysis, or EDA. Subsequently, Mel-frequency Cepstral Coefficients (MFCCs) are utilized in feature extraction procedures to obtain the key acoustic characteristics of the heart sounds. Then, a hybrid model is utilized, utilizing the advantages of both Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures in capturing temporal correlations and extracting spatial features.

A. Model Architecture

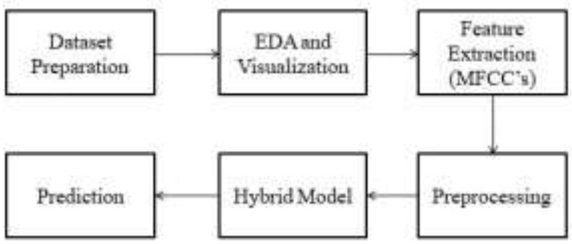

Fig .1: Block diagram

The approach is characterized by a well-organized framework created to handle the particular difficulties at hand.

The model architecture for this method is shown in Fig. 1. To comprehend the nature and qualities of the data, the process begins with exploratory data analysis (EDA) (Fig. 1, Step 2). Subsequently, Mel-frequency Cepstral Coefficients (MFCCs) are used for feature extraction, effectively capturing the principal acoustic attributes of cardiac sounds (Fig. 1, Step 3). Next, the model design adopts a hybrid strategy that utilizes the benefits of both Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architectures to efficiently extract spatial data and record temporal correlations (Fig. 1, Step 5). In the final step, it helps to classify the heart beat sound (Fig. 1, Step 6).

B. Dataset Exploration

The dataset encompasses a diverse range of heart sound signals, comprising normal heart sounds, murmurs, artifacts, extrasystoles, and extrahls. Additionally, a subset of the data is designated as unlabeled, specifically intended for testing the robustness and generalization capability of the developed model. Fig.2 represents the pie chart of labeled and unlabeled data:

Pie chart representation of our labeled and unlabeled data:

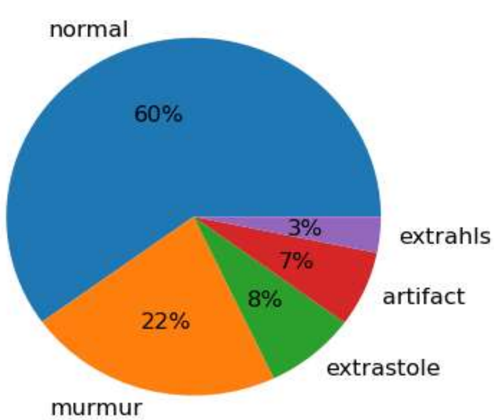

Fig . 2 : Data visualization

C. Exploratory Data Analysis (EDA) for Heart Sound Signals: Understanding the Heart Sound Dataset

For enhanced insights and a more comprehensive understanding, signals have been subjected to various visualization techniques.

1. **Waveform Analysis:**

Waveform Analysis basically shows the properties of cardiac sound signals in the time domain. The study of the properties of the audio signals over time is aided by this visualization. Fig.3, represents the waveform of a sample sound signal.

Example:

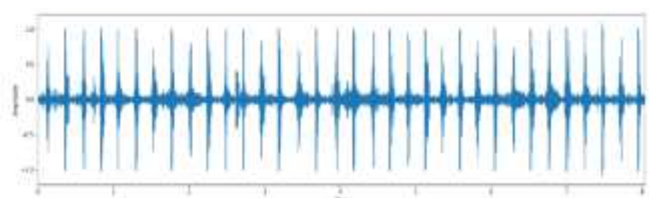

Fig.3:Waveform

2. **Spectral Analysis:**

This technique analyses the characteristics of cardiac sound waves in the frequency domain by using Radix-2 FFT for spectrum analysis. Studying the properties of audio signals at various frequencies is made easier by it. A sample sound signal's spectrum is shown in Fig. 4, which provides an insight on the cardiac signal's frequency components.

Example:

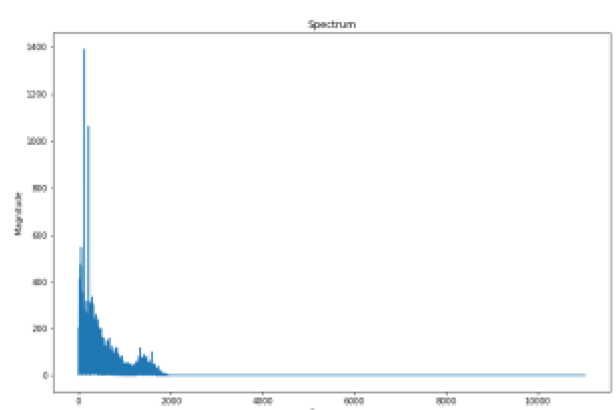

Fig.4: spectral analysis

3. **Spectrogram Representation:**

This approach aims to investigate the features of cardiac sound signals in the frequency domain by using Short-Time Fourier Transform (STFT) for spectrogram analysis. This makes it easy to analyze the properties of audio signals at various frequencies. Fig 5 shows the spectrogram of a sample sound signal, which provides proper insight on the frequency components of the cardiac signal across time.

Example:

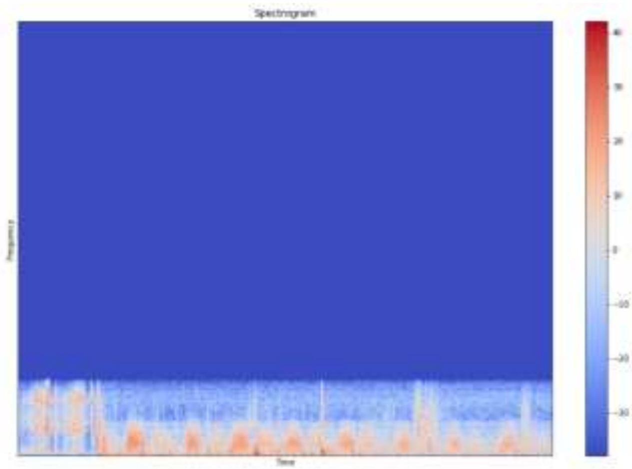

Fig.5:spectrogram

D. Feature Extraction Using Mel-frequency Cepstral Coefficients (MFCCs): Capturing Distinctive Acoustic Characteristics

One of the key techniques for identifying and displaying the acoustic characteristics present in audio signals is the computation of Mel-frequency cepstral coefficients, or MFCCs. Using the Short-Time Fourier Transform (STFT) and certain parameters like hop length, window size, and sampling rate, the described process takes an audio sample as input and extracts its MFCCs. The resulting MFCCs are then displayed as a spectrogram, as seen in Fig 6.
Example:

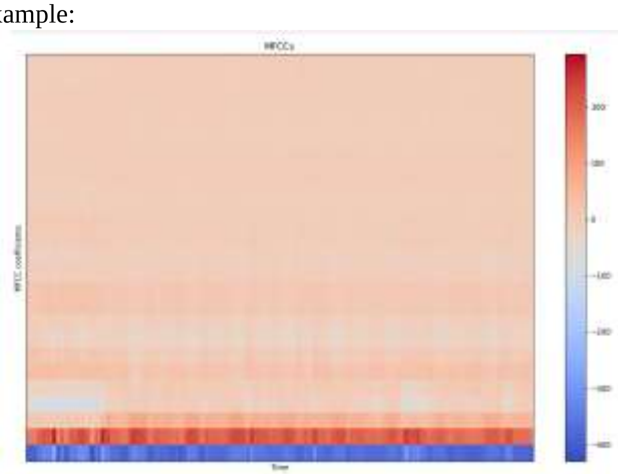

Fig.6 : Mfcc

E. Hybrid Model (CNN+ LSTM) : Modeling Temporal Dependencies

The classification method is based on a hybrid model that makes use of the advantages of both the LSTM and CNN architectures, providing a comprehensive solution for the categorization of heart sound signals:

CNN:

Convolution and pooling are two common functions carried out by a convolutional neural network (CNN) architecture, which normally consists of a series of layers. These layers are specifically designed to maintain spatial information by identifying latent features regardless of positional fluctuations by using weight sharing and local receptive fields.

Max-pooling layers, dropout techniques intended to prevent overfitting, Conv-blocks, and a sequence of Conv1D layers are all included in the model. The Conv1D layers use filters with widths of 32, 64, and 128. ReLU activation is done for non-linearity, and each filter has a kernel size of 3 and strides of 3. After testing, the best performance was indicated by a dropout rate of 0.25.

The model optimizes via Adam, using parameters for beta 1, beta 2, and learning rate. The goal function is binary cross-entropy. The main goal of the model's construction is to identify characteristics in the input data so that it can distinguish between normal and pathological heart sounds. Even though it is superficial, this model works well, especially when combined with its Conv-block procedures, which are designed to reduce input dimensions and explore feature relationships—basically, feature selection in signal processing. Utilizing Conv1D and MaxPooling1D structures, the feature maps from the final Conv1D layer serve as the foundation for sound categorization and feed into Dense layers. By using the knowledge gained from the acquired features, this method seeks to perform accurate classification.Fig.7 represents the CNN model architecture in the form of block diagram.

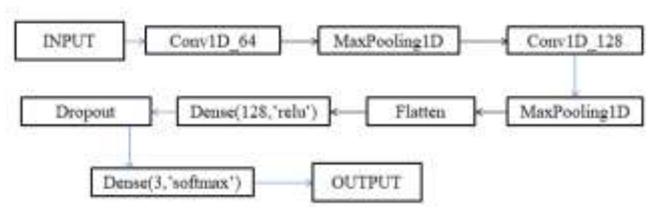

Fig.7:CNN architecture

LSTM:

The architecture uses a sequence of Bidirectional Long Short-Term Memory (LSTM) layers with a focus on sequential data analysis. By learning both forward and backward temporal patterns, the bidirectional LSTM layer which consists of 128 units performs exceptionally well in capturing temporal dependencies. Its dropout rates for input and recurrent connections are 0.05 and 0.20, respectively. Repaired linear unit (ReLU) activation is used by the next two dense layers two with 128 to the learned features. The model's last component is a Softmax output layer that divides input data into three groups.units and two with 64 units to add non-linearity

Adam optimization with a predetermined learning rate is used for efficient model training, and binary cross-entropy functions are selected as the objective criterion. With an emphasis on strong temporal feature extraction, this architecture leverages the strengths of LSTM layers to accurately categorize heart sounds. It is designed for sequential data analysis. Fig.8 represents the LSTM model architecture in the form of block diagram.

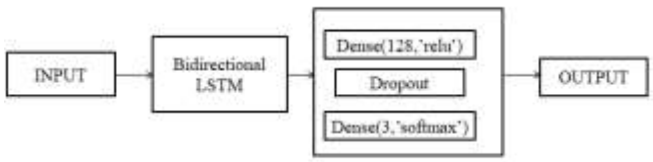

Fig.8:LSTM architecture

Hybrid Model: CNN + LSTM

The model starts with a Bidirectional Long Short-Term Memory (LSTM) layer with 128 units. For input and recurrent connections, dropout rates are set at 0.05 and 0.20, respectively. Through the learning of both forward and backward temporal patterns within sequential data, this bidirectional LSTM setup efficiently captures temporal interdependence. It is made to be exceptionally good at deciphering intricate temporal correlations.

The rectified linear unit (ReLU) activation function is used in the Conv1D layer, which comes after the LSTM layer and adds non-linearity to the feature extraction process. It has 64 filters and a kernel size of 3. Then, in order to decrease the dimensionality of the feature map and facilitate the extraction of important data, a MaxPooling1D layer downsampled the data.

The design uses ReLU activation and has many dense layers, two of which have 128 units each and two of which have 64 units each. The non-linear transformations of the collected features from the LSTM and Conv1D layers are intended to be further improved by these layers.

The model culminates in a Softmax output layer designed to partition input data into three independent classes. For efficient training, it makes use of the binary cross-entropy loss function and Adam optimization with a certain learning rate. Specifically designed for sequential data analysis, this novel architecture exploits the advantages of long short-term memory (LSTM) layers to extract robust temporal features with an emphasis on heart sound classification. The complete block diagram which consists of hybrid model architecture is shown in fig.9.

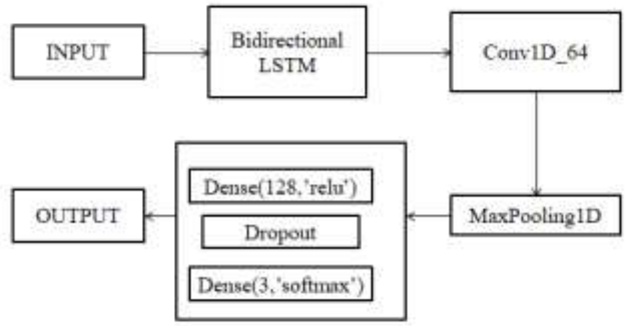

fig9.:Hybrid architecture

F. Rigorous Model Evaluation: Ensuring Accuracy and Precision

Comprehensive assessments of accuracy were performed to Comprehensive assessments of accuracy were performed to gauge the model's efficacy. To accurately categorize cardiac sound waves as normal or unhealthy, the model conducted extensive testing on hitherto unexplored data. In order to guarantee the model's accuracy and dependability in practical applications, several accuracy criteria were used.

Neural networks with efficient strategies are trained in an organized manner as part of the operational process. The validation loss is continuously tracked and recorded for further analysis during the training phase. By employing a dynamic learning rate adjustment technique, the procedure entails a 0.8 reduction factor for each epoch, which helps to boost the learning rate adaptively over time. Training is stopped in order to prevent overfitting if, after three consecutive epochs, the validation loss does not indicate improvement. To ensure efficient and thorough learning, training is done using the supplied training data (x_train, y_train) with a batch size of three over 30 epochs. Class weights are assigned to the model in order to address the imbalance between classes. The efficiency of the neural network is greatly increased by utilizing the Model Checkpoint, Learning Rate Scheduler, and Early Stopping callbacks.

Accuracy of hybrid model: 70%

IV. Result Analysis

The model uses a structured table style without explicit function references to deliver predictions for audio files inside a given directory. It arranges file names, serial numbers, and the corresponding predictions consistently. The data is produced for model predictions by processing just ".wav" audio files and utilizing secret preprocessing. From the analyzed audio, the model predicts class labels by utilizing a dictionary called "class_mapping" to convert numeric outputs (0, 1, or 2) into intuitive labels ('abnormal' or 'normal'). This strategy makes sure the results are comprehensible, which helps in understanding the anticipated classes.

Prediction on a Sample Audio File:

```
The prediction for the provided audio file is: normal
```

Fig.10: output for a single signal

The first normal sample predicted which gives the output accurately is shown in Fig.10.

Prediction on a sample directory:

```
S.NO  File Name                     Prediction
-------------------------------------------------
1     artifact__201106131834.wav    artifact
2     Aunlabelledtest__201106061215.wavartifact
3     artifact__201106101955.wav    artifact
4     normal__201101151127.wav      normal
5     artifact__201106220340.wav    artifact
6     artifact__201105190800.wav    artifact
7     normal__201102201230.wav      normal
8     Aunlabelledtest__201108222241.wavmurmur
9     murmur__201102052338.wav      normal
10    Aunlabelledtest__201106191034.wavartifact
11    artifact__201106040722.wav    artifact
12    artifact__201106070949.wav    artifact
13    artifact__201106221254.wav    artifact
14    normal__201104122156.wav      normal
15    normal__201108011112.wav      normal
```

Fig.11:Output

The outputs of the testing audio samples which gives us accurate solutions is shown in the Fig .11

V. Future Scope

Future work on the project will focus on one main area: Improving Model Accuracy, which will entail optimizing the model architecture, discovering and utilizing cutting-edge algorithms, and fine-tuning hyperparameters in order to improve the model's predictive power.

To sum up, this study presents a thorough method for classifying heart sounds that involves data exploration using EDA, feature extraction with MFCCs, deep learning based on LSTM, and careful model validation. Beyond its technological significance, the research could transform the early detection of cardiovascular problems, improving patient outcomes and healthcare delivery. The LSTM model is a promising tool for enhancing cardiac healthcare diagnostics and therapy because of its high accuracy.

VI. REFERENCES

[1] Wei Zeng, Zixiang Lin, Chengzhi Yuan, Qinghui Wang, Fenglin Liu,Ying Wang , " A new learning and classification framework for the detection of abnormal heart sound signals using hybrid signal processing and neural networks".

[2] Yaseen, Gui-Young Son and Soonil Kwon, " Classification of Heart Sound Signal Using Multiple Features".

[3] Wen-Chung Kao, Chih-Chao Wei, Jen-Jui Liu, Pei-Yung Hsiao, "Automatic Heart Sound Analysis with Short-Time Fourier Transform and Support Vector Machines".

[4] Shivnarayan Patidar, Ram Bilas Pachori, " Classification of cardiac sound signals using constrained tunable-Q wavelet transform".

[5] Ibrahim Abdel-Motaleb and Rohit Akula, "Artificial Intelligence Algorithm for Heart Disease Diagnosis using Phonocardiogram Signals".

[6] JINGSHAN HUANG, BINQIANG CHEN, BIN YAO AND WANGPENG HE, "ECG Arrhythmia Classification Using STFT-based Spectrogram and Convolutional Neural Network".

[7] Mohammed Nabih-Ali, EL-Sayed A. El-Dahshan and Ashraf S. Yahia, " Heart Diseases diagnosis using intelligent algorithm based on PCG signal analysis".

[8] Juan P. Dominguez-Morales, Angel F. Jimenez-Fernandez, Manuel J. Dominguez-Morales, and Gabriel Jimenez-Moreno, " Deep Neural Networks for the Recognition and Classification of Heart Murmurs Using Neuromorphic Auditory Sensors".

[9] Matias Brusco and Homer Nazeran, "Digital Phonocardiography: A PDA-based Approach".

[10] Wafaa N. Al-Sharu, Ali Mohammad Alqudah Member, IAENG, Shoroq Qazan, Amin Alqudah, " Detection of Valvular Heart Diseases Using Fourier Transform and Simple CNN Model ".

[11] Siti Agrippina Alodia Yusuf and Risanuri Hidayat, " MFCC Feature Extraction and KNN Classification in ECG Signals".

[12] Wei Chen, Qiang Sun, Xiaomin Chen, Gangcai Xie, Huiqun Wu and Chen Xu , " Deep Learning Methods for Heart Sounds Classification: A Systematic Review".

[13] Muqing Deng, Tingting Meng, Jiuwen Cao, Shimin Wang , Jing Zhang , Huijie Fan, Heart sound classification based on improved MFCC features and convolutional recurrent neural networks".

[14] Ali Raza , Arif Mehmood , Saleem Ullah, Maqsood Ahmad, Gyu Sang Choi and Byung-Won On, " Heartbeat Sound Signal Classification Using Deep Learning".

[15] Fan Li , Hong Tang, Shang Shang, Klaus Mathiak and Fengyu Cong, " Classification of Heart Sounds Using ConvolutionalNeuralNetwork".

[16] Taiyong Li, Min Zhou, " ECG Classification Using Wavelet Packet Entropy and Random Forests".

[17] Emran M. Tamil, Nor Hafeezah Kamarudin, Rosli Salleh,M. Yamani Idna Idris,Noorzaily M.Noor,Azmi Mohd Tamil., " Heartbeat Electrocardiogram (ECG) Signal Feature Extraction Using Discrete Wavelet Transforms (DWT)".

[18] Muqing Deng, Tingting Meng, Jiuwen Cao, Shimin Wang, Jing Zhang, Huijie Fan, " Heart sound classification based on improved MFCC features and convolutional recurrent neural networks".

[19] Methaq A. Shyaa, Ayat S. Hasan, Hassan Ibrahim & Weam Saadi Hamza , " Automatic Classification of Heart Sounds Utilizing Hybrid Model of Convolutional NeuralNetworks".

[20] Tharindu Fernando, Houman Ghaemmaghami, Simon Denman, Sridha Sridharan, Nayyar Hussain, and Clinton Fookes, " Heart Sound Segmentation using Bidirectional LSTMs with Attention".