Analysis of Audio Signals and Hyper-Parameter Optimization of Convolutional Neural Network for Heartbeat Anomaly Detection

Gandotra Ekta[1*], Gupta Deepak[1], Mahajan Ria[1,2], Sharma Parul[1] and Kumar Harish[3]
1. Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Waknaghat, Himachal Pradesh, INDIA
2. Department of Computer and Information Science and Engineering, University of Florida, Gainesville, USA
3. Department of Computer Science and Engineering, SDGI Global University, Ghaziabad, Uttar Pradesh, INDIA
*ekta.gandotra@gmail.com

Abstract

Cardiovascular diseases hold significant importance as global health concerns due to their alarmingly high mortality rates. These diseases are characterized by the heart's inability to supply sufficient blood to the body's other organs. This inability of the heart can have severe repercussions on a patient's health. Cardiovascular diseases encompass conditions such as coronary artery disease, heart failure, stroke, hypertension and various other disorders that affect the heart and blood vessels. A timely and accurate diagnosis is essential for a patient's survival as well as for averting further loss.

This study presents an efficient method that helps in identifying irregular heartbeats. The proposed method uses an audio signal dataset that has been compiled from general public as well as clinical studies. It uses time-frequency heatmaps and deep convolutional neural network to automate the classification of heartbeat audio signals. An extensive process of hyperparameter tuning to optimize the learning rate, batch size and the number of epochs is used to enhance the performance of the model. The experimental results of the study show the effectiveness of the proposed model in identifying irregular heartbeats. After doing hyperparameter tuning, the proposed model obtains an accuracy of 96% on the validation dataset.

Keywords: Cardiovascular diseases, Convolutional neural network, Hyperparameter tuning, Mel-Frequency cepstral coefficients, Phonocardiogram signals, Spectrograms.

Introduction

Cardiovascular disease (CVD) is one of the leading causes of death worldwide. It resulted in over 17.9 million deaths in 2019[31]. The primary reasons for CVD related mortality are heart attacks and strokes. The low and middle income nations bear the brunt of this burden. The increasing prevalence of CVDs is attributed to various factors including urbanization, poor diet, use of tobacco and restricted access to healthcare amenities. Effective preventative and controlling measures are therefore urgently needed. These high rates are mostly the result of the disease's delayed diagnosis and the failure of high-risk patients to follow preventive advice from their healthcare professionals. Anomalies in heartbeat are very important to detect since they may indicate underlying heart conditions requiring treatment.

Although electrocardiograms (ECG) and echocardiograph are most commonly used ways to monitor cardiovascular health, they can be time-consuming, inconvenient and may require expertise. Therefore, convolutional neural networks (CNNs) with deep learning (DL) abilities offer a feasible solution to this problem by improving the speed and accuracy of ECG interpretation[6].

DL has revolutionized various fields by facilitating machines to learn and make intelligent decisions from large and complex data. These have the ability to extract meaningful patterns from the raw data leading to advanced capabilities in various applications and domains[11]. DL models can be trained on physiologic time series data in the medical field to find anomalies and patterns that can help in early diagnosis. DL has done a tremendous job in the field of audio processing by utilizing spectrograms, which is a widely used technique for representing sound. It eliminates the requirement of conventional audio processing methods by relying on standard data preparation rather than manually creating feature vector.

The architecture of CNN as explained by Goodfellow et al[11], leverages spatial hierarchy and local connectivity to automatically learn and extract meaningful features from grid-like data, such as images. Nowadays, these are extensively being used in various applications including image classification[9], object detection[8], image segmentation[15], face recognition[2], assessing the risk of Android applications[5,7], detecting improper face mask[4] and healthcare imaging[27,28] etc.

By training on labeled datasets, CNNs excel at tasks like distinguishing objects within images, localizing objects, segmenting images into meaningful regions, identifying and verifying individuals based on facial features and aiding in medical imaging analysis. These capabilities highlight the potential of CNNs in advancing medical diagnostics, particularly in identifying irregularities in cardiovascular functions. The components of CNN, their functions and all other significant problems were discussed in detail by Albawi et al[1]. They also listed the factors that affect CNN's

effectiveness. The absence of training samples with annotations is the fundamental obstacle to the DL-based classification of medical images. Wang et al[29] showed how, for small training samples, fine-tuning greatly increased the classification accuracy of liver lesions.

Based on the electroencephalography (EEG) spectrogram data, Mandhouj et al[21] created a deep CNN model that was effective in identifying and categorizing epilepsy seizures. The experimental findings demonstrated the effectiveness of the suggested strategy, which had an average accuracy rate of 98.22% in identifying EEG signals. The approach for classifying ECG arrhythmias proposed by Huang et al[16] used two-dimensional (2D) deep CNN. Using the short-term Fourier transform (STFT), five different forms of ECG time-domain data were first transformed into time-frequency spectrograms. Finally, ECG arrhythmia types were identified and classified using spectrograms of five different arrhythmia types as inputs to the 2D-CNN.

Kaur et al[20] proposed a unique framework based on grid-search optimization to develop a DL model to predict the early onset of Parkinson's disease, where several hyperparameters had to be established and tweaked for evaluation of the resulting DL model. Audio signal processing is the process of applying complex algorithms and methods to work with audio signals. Audio signals (both digital and analog) are used to represent sound. Binary representations contain digital signals, while electrical signals contain analog signals. The time-frequency bands must be balanced and unwanted noise must be removed using this technique. Audio signal processing has a focus on computational methods for manipulating sounds. It minimizes or eliminates undesired noise like echo and over-modulation by utilizing a variety of techniques.

Ming et al[22] focused on the modeling and de-noising of audio signals and applied broadly a variety of fundamental concepts from digital signal processing before performing spectrum analysis and applying filters to the audio signals. The processing of audio signals in this work was realized using the fundamental concepts of digital signal science and the processing of speech signals was accomplished extensively using signal extraction, amplitude-frequency transform, Fourier transform, filtering and other methods. A review of contemporary deep learning techniques for handling audio signals was carried out by Purwins et al[24]. Hershey et al[14] classified the soundtracks of a dataset of 70M training videos (5.24 million hours) using 30,871 video-level labels utilizing various CNN architectures. The research investigated fully connected deep neural network (DNN), ResNet, AlexNet, VGG and Inception.

Rong[25] suggested a novel machine learning (ML) method for categorizing audio. The author discussed the four-layer hierarchical structure of audio data as well as the three different categories of audio data features: short time energy, zero crossing rate and mel-frequency cepstral coefficients (MFCCs) which were further extracted to create the feature vector. The research finally discussed the use of the support vector machine classifier with a Gaussian kernel to categorize audio data.

Numerous studies have employed various methods, including segmentation, down sampling, feature extraction and classification to forecast heart disease. Identifying abnormal behaviors in a certain context is both a challenge and a necessity since we can quickly address a problem by detecting an anomaly. A typical heartbeat has a distinct pattern of "lub dub, dub lub," with the time between each beat being longer than the time between each beat and typically beats between 60 and 100 times per minute (lub and dub). With symptoms of numerous heart diseases, a murmur heartbeat sound has a noise pattern that whooshes, roars, rumbles, or turbulent fluid between lub to dub or dub to lub. Omarov et al[23] implemented ML techniques for the detection of heart disease using phonocardiogram (PCG) signals.

Gomes et al[10] described a system for categorizing cardiac sounds in the PASCAL challenge. The S1 and S2 heart sounds, where S1 is lub and S2 is dub, were found using an algorithm. To reduce the noise in such signals, they first employed a band-pass filter after using MATLAB decimate function on the original sound stream. The average Shannon energy was then applied, which helped to clearly detect the peaks of the heart sound signal. The segmentation of the heartbeat sound was achieved using an algorithm in which the maxima and minima locations of the sound signal were located. The model was trained using the J48 and multilayer perceptron method. Similarly, Jadhav et al[17] suggested a novel method of categorizing heart sounds into the normal or murmur category using a peak identification approach based on Shannon energy envelope calculation and neural networks.

A cutting-edge classifier for the early diagnosis of cardiac diseases was developed by Wolk et al[30] using CNN and multipart interactive training. Over 93% accuracy was achieved in the experiments utilizing the ResNet pre-trained network.

Kaisti et al[19] proposed a technique that employed miniaturized inertial sensors to assess the chest's pre-cordial translational and rotational motions. The algorithm eliminated motion artifacts chose the optimal axis from the multi-axial accelerometer and gyroscope data sets and located beats using two detection concepts based on the signal envelope and signal morphology. The authors compared the detection performance between the sensor modalities in two study groups: (i) healthy subjects and (ii) heart disease patients. The research also took account of the beat-to-beat detection accuracy and estimated the heart rate. For healthy people, the average true and positive prediction rates of beat detection were 99.9% and 99.6% whereas for heart disease patients, these rates ranged from 95.9% and 95.3%.

Although high-accuracy beat detection was accomplished for heart disease patients, location matching was found to be less accurate in such individuals when compared to healthy subjects. Rubin et al[26] described an automated heart sound classification technique that blends deep CNNs and time-frequency heatmap representations.

From the literature review, it is clear that various DL methods are used for the classification of heartbeat sounds; however, these are limited by things such as large, annotated and high-dimensional datasets. The proposed method overcomes this drawback by combining spectrogram analysis and DL for extracting features. Further, the proposed approach combines clinical and manually gathered data (via a mobile application), thereby augmenting accessibility for a wider demographic. The proposed approach demonstrates efficacy even with poor quality audio signals. It overcomes the constraint of data accessibility encountered by earlier research.

The main aim of this study is to create an accurate, automated approach for detecting heartbeat anomalies using time-frequency heatmaps and CNN model. The primary contributions include:

- Creation of a system that can precisely classify irregular heartbeats using time-frequency heatmaps and CNN.
- Examining the best accuracy that is attained by adjusting hyperparameters.

Material and Methods

The workflow used in this methodology is depicted in figure 1. Four separate phases comprise the entire methodology: data collection, data preparation, feature extraction and model building. This is followed by hyperparameter tuning and result analysis. The description of each phase is given as follows:

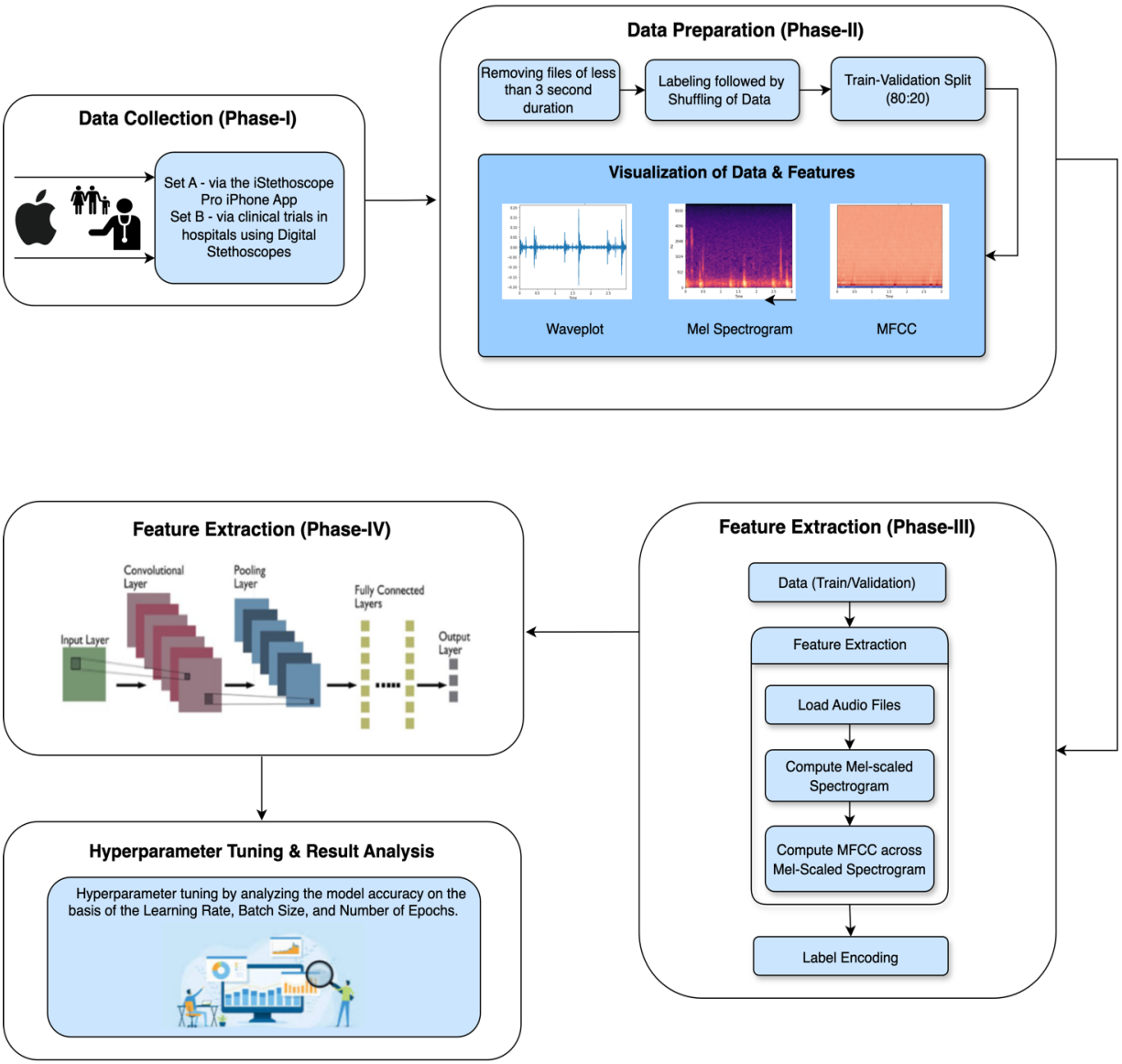

Fig. 1: Workflow of the proposed methodology.

Data Collection: The heartbeat sound dataset used in this work is obtained from Kaggle[18]. It was first used in a ML competition for heartbeat sound classification[3]. According to the source of the data, it is divided into two sets. The first set A consists of 176 audio signals which are collected from the general population using the I-Stethoscope Pro iPhone application, whereas the second set B is made up of 656 audio signals collected from a clinical trial carried out in hospitals using the DigiScope digital stethoscope.

Data Preparation: To ensure stable data representation and to retain contextual information, this study used files that are longer than 4 seconds. This approach means that the length of every sample in the dataset is uniform, which helps to make data analysis easier and more efficient by keeping ML and DL models training consistent at all times. The reason for choosing a 4-second segment is to ensure that training and analysis are not burdened by an unmanageable amount of data, while at the same time including enough contextual information in audio signals. For longer segments, more computational resources are needed. A too short segment would not produce accurate data for analysis.

To reduce noise resulting from physical contact between the microphone and the body, an offset computation is utilized to account for the extra space in audio files.

The labels provided in the dataset are used to classify the audio files as 'normal' or 'abnormal' after the proper segment size has been chosen. This labeling helps to classify heartbeat sounds as either indicative of a healthy heart or exhibiting an anomaly. Subsequently, the data is shuffled to avoid any bias or pattern in the sequence. The shuffling step helps in ensuring the model's robustness and generalization performance, as the model does not learn any particular sequence or pattern in the input data. The resulting dataset at this stage contains two types of audio data: 148 normal and 96 abnormal audio files. The data is subsequently split into training and validation data in the ratio of 80% (195 audio files) and 20% (49 audio files) respectively. The 1D time series input segments are transformed into 2D spectrograms (heatmaps) using MFCCs, which display the distribution of signal energy over time and frequency.

Feature Extraction: Mel spectrograms and MFCCs are widely used in audio signal processing and analysis, particularly for speech and music recognition tasks. Both representations capture essential information from the audio signal, transforming it into a format that facilitates pattern recognition and classification. Mel spectrograms are computed from the STFT of an audio signal, which converts the time-domain signal into a time-frequency representation. The STFT is performed by applying a sliding window to the signal and computing the discrete Fourier transform (DFT) of the windowed segments. The Fourier transform allows us to represent the signal as a sum of sinusoids with different frequencies and amplitudes. Mathematically, the DFT is defined using eq. (1):

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi kn}{N}} \quad (1)$$

where $x(n)$ denotes the input signal, N represents the quantity of samples, j is an imaginary unit ($j^2 = -1$) and $X(k)$ serves as the frequency-domain portrayal of the said signal.

Upon computation of the STFT, the resultant spectrogram is converted to the Mel scale, a scale that is perceptually motivated and strategically designed to more accurately portray human auditory perception. The Mel scale emphasizes lower frequencies while compressing higher frequencies, simulating the non-linear frequency resolution of the human auditory system. MFCCs are derived from the Mel spectrogram. These are obtained through the application of the discrete cosine transform on the logarithmic Mel-scaled power spectrum. This process decorrelates the spectral coefficients, resulting in a compact and robust representation of the audio signal. MFCCs capture the spectral envelope of the signal, which is particularly useful for distinguishing different phonemes in speech and identifying musical characteristics such as timbre.

In this study, a custom function is defined to process the audio data by extracting MFCC features. This function takes an audio file's path and an offset as input parameters, loads the audio file with duration of 4 seconds starting at the given offset, computes the Mel spectrogram and subsequently extracts 40 MFCC features. This function is applied to both the training and validation datasets, obtaining feature arrays and their corresponding labels. The audio data, stored in the Waveform Audio File Format (.wav), represents 1D time series signals. Transforming these signals into 2D heatmaps like Mel-spectrograms and MFCCs, captures the time-frequency patterns of the signals. It makes them more informative for analysis purpose.

Acquiring data using different tools like digital stethoscope or a cell phone microphone, can result in variations in amplitude ranges. The process of converting data into the frequency domain yields more consistent and reliable results. To facilitate the efficient use of CNNs for classification tasks, the original time series audio data in this study is converted into 2D heatmaps. When compared to using raw time series data alone, this leads to the extraction of more relevant and discriminative features from the audio signal which improves classification performance. The study intends to improve the efficacy of the presented framework by developing an accurate and robust classifier for differentiating between usual and anomalous cardiac sounds by using CNNs on Mel spectrograms and MFCCs.

Model Building: The 40×130 MFCC heatmaps that represent cardiac sound segments, are precisely analyzed and classified by the CNN architecture used in this work. Convolutional layers, fully connected layers and SoftMax classifiers are the components of this CNN that work together harmoniously to produce a binary classification

output that indicates whether the input segment corresponds to a normal or abnormal heart sound.

Four convolutional layers make up the model and they are in charge of gathering and extracting local attributes from the input MFCC heatmaps. Four max-pooling layers are added after these convolutional layers to reduce the feature maps' spatial dimensions and improve the model's capacity to manage translation variation. To reduce overfitting and improve the model's ability to generalize, the architecture additionally has four dropout layers. A global average pooling layer efficiently condenses the obtained characteristics into a fixed-size vector by combining the feature maps. Lastly, a dense layer is used to perform the classification task using the aggregated feature vector.

The convolutional layers use the rectified linear unit (ReLU) activation function. ReLU is preferred because it can solve the vanishing gradient problem and is computationally efficient in both the training and validation stages. ReLU's selection as the activation function is a key factor in improving the CNN's general resilience and performance which helps it to distinguish between normal and abnormal heart sounds.

Adam is selected to be the optimizer for this study because of its widespread use in DNN optimization. Adam can handle big datasets and parameters because it is computationally efficient in terms of both time and memory. Moreover, Adam is insensitive to gradient scaling, which leads to increased stability and reduced responsiveness to the selection of hyperparameters.

Hyperparameter Tuning: In order to achieve optimal performance, the study explores various sets of hyperparameters.

- **Learning Rate:** One crucial hyperparameter that controls the number of iterations needed for the model to minimize the loss function is the learning rate. Selecting a larger value of learning rate allows the model to learn faster. The drawback is that there is a chance that it will surpass the minimum loss function. On the other hand, a lower value of learning rate increases the likelihood of obtaining the minimum loss function though it requires more memory.
- **Batch Size:** The batch size is very important in determining the number of subsamples of the input training data used in each iteration. A smaller value of batch size accelerates the learning process but it may lead to variability in the accuracy of the validation dataset. On the other hand, a larger value of batch size delays the learning process while declining the variance in the accuracy of the validation dataset.
- **Epochs:** Small value of epochs can result in underfitting which means that the model has not learned enough from the training dataset. Too many epochs can result in overfitting indicate that the model performs well on the training dataset but struggles to generalize to new data. The optimal number of epochs must be determined to achieve the best results.

Evaluation Metrics Used: The proposed model is evaluated using the evaluation metrics, namely, Precision, Recall, F1-Score, Accuracy and Loss. These are computed from the fields of the confusion matrix i.e. True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). TP refers to the correct classification of abnormal heartbeat sounds as abnormal. FP denotes the incorrect classification of normal heartbeat sounds as abnormal. TN represents the accurate classification of normal heartbeat sounds falling within the normal category, while FN signifies the mistaken classification of abnormal heartbeat sounds as normal. By utilizing these conventions, the ensuing discussion provides insights into the classification outcomes and their implications for accurately identifying normal and abnormal heartbeat sounds.

- **Precision:** Precision is defined as the ratio of TP predictions (i.e. instances correctly identified as belonging to a specific class) to the total instances predicted as belonging to that class. It measures the accuracy of positive predictions made by the classifier. Mathematically, precision is defined using eq. (2):

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- **Recall (Sensitivity):** It is defined as the ratio of TP predictions (i.e. correctly identified instances of a class) to the total actual occurrences of that class. It measures the classifier's ability to identify all relevant instances of a class. Mathematically, recall is defined using eq. (3):

$$Recall = \frac{TP}{FN+TP} \quad (3)$$

- **F1-Score:** It is the harmonic mean of precision and recall. It is particularly useful when working with imbalanced datasets because it provides a single metric that balances the trade-off between precision and recall. Mathematically, F1-Score is defined using eq. (4):

$$F1 - Score = \frac{2*(Precision*Recall)}{(Precision+Recall)} \quad (4)$$

- **Accuracy:** The correct number of predictions to total instances in the dataset is called accuracy. Although it gives a broad picture of the classifier's performance, it can be misleading when working with datasets that are unbalanced. Mathematically, accuracy (%) is defined using eq. (5):

$$Accuracy(\%) = \frac{TP+TN}{TP+TN+FP+FN} *100 \quad (5)$$

- **Loss:** The metric of loss measures the discrepancy between the true value of a single sample and the value

projected by the model. In this study, the categorical cross-entropy loss function is utilized. Mathematically, loss is defined using eq. (6):

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)] \quad (6)$$

where N stands for the sample size, $y_i \in \{0,1\}$ is the true label of the i^{th} sample and $\hat{y}_i$ is the predicted probability of the positive class for that sample.

Results and Discussion

A 64-bit operating system and Intel® CoreTM i5-10300H processor together with 8 GB of installed memory (RAM) along with Python (version 3.7.15) with Numpy (version 1.21.6), Pandas (version 1.3.5), Librosa (version 0.8.1), Sklearn (version 1.0.2) and Keras (version 2.9.0) libraries are used while conducting the experiments. In the proposed 2D CNN model, learning rate, batch size and the number of epochs are three critical parameters. Optimizing these model parameters is essential to achieve the best classification performance for PCG heartbeat signals. A series of experiments are conducted to assess the impact of learning rate and batch size on the performance of the proposed 2D CNN model. Initially, trials are performed with different learning rates while keeping the batch size and number of epochs constant. Table 1 presents the variations in the model's accuracy with changes in the learning rate.

Figure 2 illustrates the variation in the model's validation accuracy as the learning rate is altered from 0.0001 to 1.0 while keeping the batch size constant at 128 and training the model for 300 epochs. It is evident that the validation accuracy increases as the learning rate is raised. The validation accuracy reaches its peak value of 95.91% at a learning rate of 0.01, which is a commonly used value in the literature. However, beyond this point, the validation accuracy declines sharply to 57.14% when the learning rate is set to 1.0. Consequently, a learning rate of 0.01 emerges as the most optimal setting for the model, providing the highest validation accuracy.

Table 1
Validation accuracy with varying learning rates for batch size = 128 and epochs = 300

Learning Rate	Validation Accuracy
0.0001	0.71428
0.001	0.79591
0.01	0.95918
1.0	0.57142

After determining the optimal learning rate for the model, the next step is to experiment with different batch sizes. The model is trained with epochs = 300, learning rate = 0.01 and batch sizes ranging from 16 to 256. The results of these experiments, with a fixed learning rate and epochs with varying batch sizes, are presented in table 2. As shown in figure 3, the model achieves highest validation accuracy of 93.87% at a batch size of 16. The validation accuracy gradually decreases up to a batch size of 64. There is a sudden increase in validation accuracy at a batch size of 128, but it drops significantly to 69.38% when the batch size is increased to 256. Based on these findings, a batch size of 16 is determined to be the optimal choice for the model.

After determining the optimal learning rate and batch size, the accuracy of the proposed model is evaluated by gradually increasing the number of epochs during training. Table 3 presents the validation accuracy with varying epochs (from 50 to 400) for batch size set to 16 and learning rate set to 0.01. By comparing the outcomes of different parameter configurations, the goal is to identify the most effective combination of learning rate and batch size, which would improve the overall performance and generalizability of the 2D CNN model in heartbeat anomaly detection.

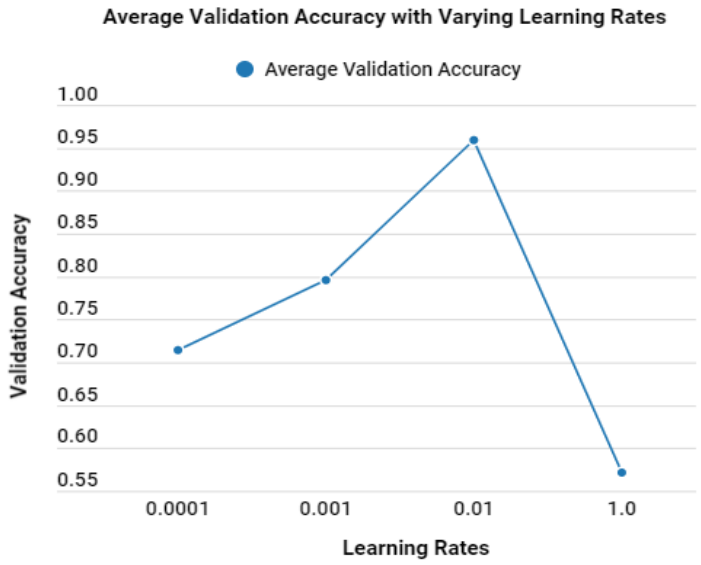

Fig. 2: Average validation accuracy with varying learning rates for batch size = 128 and epochs = 300

Table 2
Validation accuracy with varying batch sizes for learning rate = 0.01 and epochs = 300

Batch Size	Validation Accuracy
16	0.93877
32	0.83673
64	0.75510
128	0.85714
256	0.69387

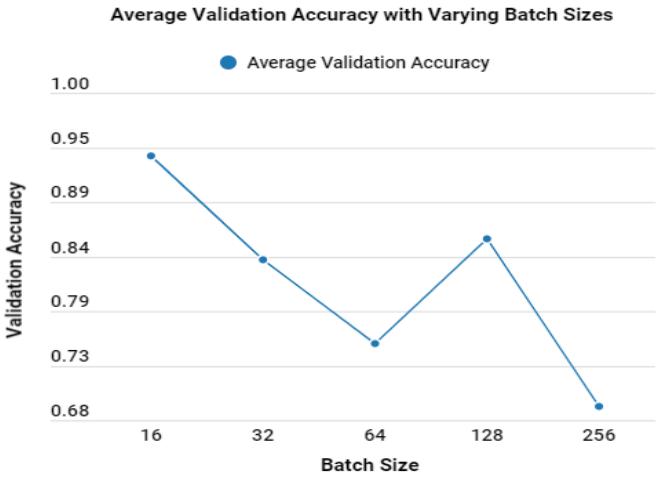

Fig. 3: Average validation accuracy with varying batch sizes for learning rate = 0.01 and epochs = 300

Table 3
Validation accuracy with varying epochs for batch size = 16 and learning rate = 0.01

Epochs	Validation Accuracy
50	0.83691
100	0.85714
150	0.77551
200	0.89795
250	0.91836
300	0.95918
350	0.93877
400	0.8776

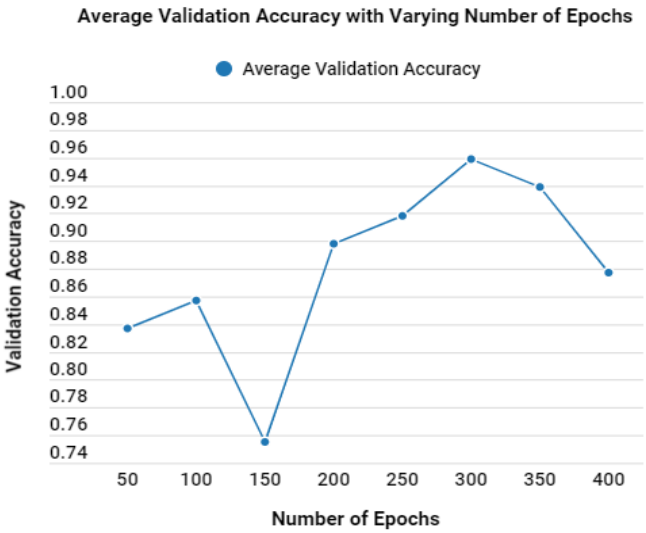

Fig. 4: Average validation accuracy with varying number of epochs for batch size = 16 and learning rate = 0.01

The change in model accuracy with varying epochs is visualized in figure 4. It shows that the model achieves the best accuracy of 95.92% at 300 epochs.

After performing hyperparameter tuning, the optimal values for the parameters are identified. These values are used to build the final CNN model, which is then trained and evaluated. The chosen parameters for the final model are shown in table 4.

Table 4
Parameters of the final proposed CNN model.

Parameters	Values
Learning Rate	0.01
Optimiser	Adam
Batch Size	16
Epochs	300

With these hyperparameters, the model is trained. Loss and accuracy for training and validation parts of the dataset are recorded over the epochs ranging from 0 to 300 as shown in figure 5 and figure 6 respectively. Figure 5 shows that the model starts with an initial training and validation loss of approximately 5.5. A significant drop is observed in the first 50 epochs where both losses reduce to around 1.0, marking a rapid learning phase. Beyond 50 epochs, the losses stabilize, showing minor fluctuations while remaining below 1.0. By the final epoch, the training loss reaches approximately 0.4 and the validation loss is around 0.5 demonstrating effective convergence, minimal generalization error and strong performance of the model with optimal hyperparameters.

Figure 6 illustrates that the model starts with an initial accuracy of approximately 60% for both training and validation parts of the dataset. During the first 50 epochs, accuracy improves significantly, reaching around 83%. By the final epoch, training accuracy reaches close to 98% while validation accuracy stabilizes at approximately 96%, indicating excellent generalization and strong performance.

A thorough evaluation of the classifier's capacity to distinguish between normal and abnormal heart sounds for the validation data is provided in the classification report in figure 7. It provides an assessment of the proposed CNN model to detect heartbeat anomalies. A number of significant metrics that highlight the performance of the model, are shown in the report.

The model provides a precision of 0.93 for the abnormal class, which assesses the accuracy of positive predictions. This shows that when the model correctly predicts an abnormal heartbeat sound, it does so 93% of the time. The precision is 1.00 for the normal class which demonstrates a perfect accuracy in predicting normal heartbeat sounds. The proposed model identifies all real abnormal heartbeat sounds with a recall value of 1.00 for the abnormal class. Recall value for the normal class is 0.90 which means that only 90% of the actual heartbeat sounds are captured by the model. The F1-Score is a metric that combines precision and recall. For abnormal class, its value is 0.97 and for the normal class it is 0.95. The support metric reveals that there are 28 samples of abnormal and 21 samples of normal heartbeat sounds. The overall accuracy provided by the proposed model is 96%.

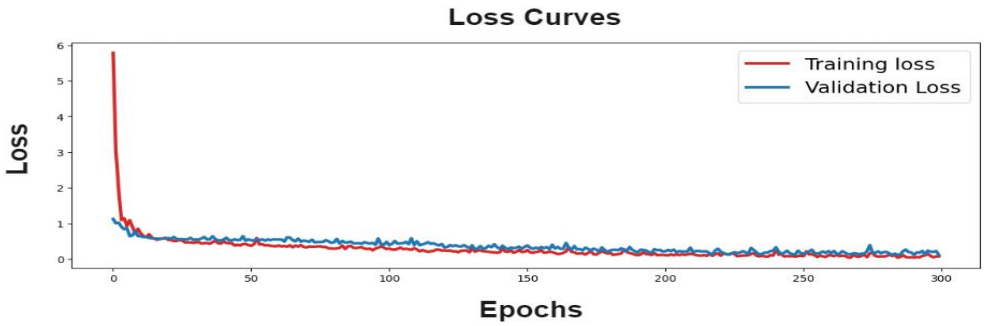

Fig. 5: Loss curves for proposed CNN model.

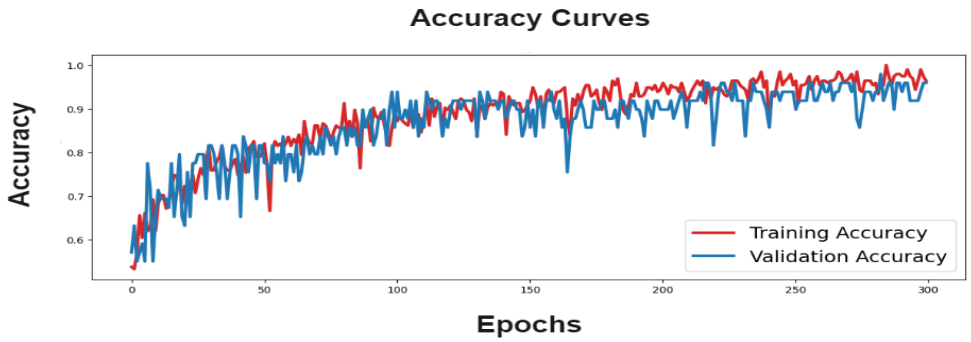

Fig. 6: Accuracy curves for proposed CNN model.

	precision	recall	f1-score	support
abnormal	0.93	1.00	0.97	28
normal	1.00	0.90	0.95	21
accuracy			0.96	49
macro avg	0.97	0.95	0.96	49
weighted avg	0.96	0.96	0.96	49

Fig. 7: Classification report for validation data using proposed CNN model

Conclusion

This study proposed a time-frequency heatmap and deep CNN-based approach for automating heartbeat anomaly detection. The proposed approach demonstrated the ability of CNN to effectively classify recordings of normal and abnormal heartbeats. The proposed CNN model provided an overall accuracy of 96% with precision, recall and F1-Score as 0.96. The proposed work proved its efficacy in finding patterns in heart sound data by using MFCCs for feature extraction and creating a customized CNN architecture. The efficacious execution of proposed approach holds the potential to exert a substantial influence in the field of medicine, enabling medical practitioners to promptly identify and manage cardiac complications, thereby enhancing patient outcomes and elevating their standard of living.

In future, we intend to focus to encompass a plethora of distinct domains in order to further augment the efficacy and generalizability of the heartbeat anomaly detection paradigm. To complement the MFCCs and to enhance the classifier's ability to distinguish normal heart sounds from abnormal heart sounds, additional characteristics such as spectral and temporal aspects of the audio data will be considered. The dataset will be expanded to encompass more diverse and extensive recordings derived from a variety of sources, demographics and medical conditions using big data techniques[12,13]. It will help the classifier capturing the heart sounds' variability more effectively, thus providing more robust and accurate results.

References

1. Albawi S., Mohammed T.A. and Al-Zawi S., Understanding of a convolutional neural network, In 2017 international conference on engineering and technology (ICET), IEEE, Antalya, Turkey, 1-6 **(2017)**

2. Ben Fredj H., Bouguezzi S. and Souani C., Face recognition in unconstrained environment with CNN, *The Visual Computer*, **37**, 217-26 **(2021)**

3. Bentley P., Nordehn G., Coimbra M., Mannor S. and Getz R., The Heart Challenge, http://www.peterjbentley.com/heartchallenge, Accessed 16 August, 2023 **(2011)**

4. Bhaik A., Singh V., Gandotra E. and Gupta D., Detection of improperly worn face masks using deep learning–A preventive measure against the spread of COVID-19, *International Journal of Interactive Multimedia and Artificial Intelligence*, **7**, 14-25 **(2021)**

5. Chauhan K. and Gandotra E., Risk Analysis of Android Applications using Static Permissions and Convolutional Neural Network, In Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing, 269-273 **(2023)**

6. Darmawahyuni A., Nurmaini S., Rachmatullah M.N., Tutuko B., Sapitri A.I., Firdaus F., Fansyuri A. and Predyansyah A., Deep learning-based electrocardiogram rhythm and beat features for heart abnormality classification, *Peer J Computer Science*, **8**, e825 **(2022)**

7. Dhalaria M. and Gandotra E., Convolutional Neural Network for Classification of Android Applications Represented as Grayscale Images, *International Journal of Innovative Technology and Exploring Engineering*, **8**, 835-843 **(2019)**

8. Dhillon A. and Verma G.K., Convolutional neural network: a review of models, methodologies and applications to object detection, *Progress in Artificial Intelligence*, **9(2)**, 85-112 **(2020)**

9. Dhruv P. and Naskar S., Image classification using convolutional neural network (CNN) and recurrent neural network (RNN): A review, In Swain D., Pattnaik P. and Gupta P., eds., Machine Learning and Information Processing. Advances in Intelligent Systems and Computing, Springer, Singapore, **1101**, 367-81 **(2020)**

10. Gomes E.F. and Pereira E., Classifying heart sounds using peak location for segmentation and feature construction, In Workshop Classifying Heart Sounds, 480-92 **(2012)**

11. Goodfellow I., Bengio Y. and Courville A., Deep learning, MIT Press **(2016)**

12. Gupta D. and Rani R., A study of big data evolution and research challenges, *Journal of Information Science*, **45**, 322-340 **(2019)**

13. Gupta D. and Rani R., Improving malware detection using big data and ensemble learning, *Computers & Electrical Engineering*, **86**, 106729 **(2020)**

14. Hershey S., Chaudhuri S., Ellis D.P., Gemmeke J.F., Jansen A., Moore R.C., Plakal M., Platt D., Saurous R.A., Seybold B. and Slaney M., CNN architectures for large-scale audio classification, In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP), New Orleans, LA, USA, 131-135 **(2017)**

15. Hesamian M.H., Jia W., He X. and Kennedy P., Deep learning

techniques for medical image segmentation: achievements and challenges, *Journal of Digital Imaging*, **32**, 582-96 **(2019)**

16. Huang J., Chen B., Yao B. and He W., ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network, *IEEE Access*, **7**, 92871-80 **(2019)**

17. Jadhav A.R., Ghontale A.G. and Ganesh A., Heart sounds segmentation and classification using adaptive learning neural networks, In 2017 International Conference on Signal Processing and Communication (ICSPC), IEEE, Coimbatore, India, 33-38 **(2017)**

18. Kaggle Heartbeat Sounds Classification-Analysis, https://www.kaggle.com/code/brsdincer/heartbeat-sounds-classification-analysis. Accessed 16 August 2023 **(2023)**

19. Kaisti M., Tadi M.J., Lahdenoja O., Hurnanen T., Saraste A., Pänkäälä M. and Koivisto T., Stand-alone heartbeat detection in multidimensional mechanocardiograms, *IEEE Sensors Journal*, **19(1)**, 234-42 **(2018)**

20. Kaur S., Aggarwal H. and Rani R., Hyper-parameter optimization of deep learning model for prediction of Parkinson's disease, *Machine Vision and Applications*, **31**, 1-5 **(2020)**

21. Mandhouj B., Cherni M.A. and Sayadi M., An automated classification of EEG signals based on spectrogram and CNN for epilepsy diagnosis, *Analog Integrated Circuits and Signal Processing*, **108**, 101-110 **(2021)**

22. Ming B. and Wu P., Research on Audio Signal Denoising and Simulation Processing, In 2019 International Conference on Communications, Information System and Computer Engineering (CISCE), Haikou, China, IEEE, 192-194 **(2019)**

23. Omarov B., Gamry K., Batyrbekov A., Alimzhanova Z., Dauytova Z. and Seitbekova G., Detection of heartbeat abnormalities from phonocardiography using machine learning 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 843-847 **(2021)**

24. Purwins H., Li B., Virtanen T., Schlüter J., Chang S.Y. and Sainath T., Deep learning for audio signal processing, *IEEE Journal of Selected Topics in Signal Processing*, **13(2)**, 206-19 **(2019)**

25. Rong F., Audio classification method based on machine learning, In 2016 International conference on intelligent transportation, big data & smart city (ICITBS), Changsha, China, 81-84 **(2016)**

26. Rubin J., Abreu R., Ganguli A., Nelaturi S., Matei I. and Sricharan K., Recognizing abnormal heart sounds using deep learning, *arXiv*, https://doi.org/10.48550/arXiv.1707.04642 **(2017)**

27. Saravanan N., Photocatalytic degradation of methylene blue dye from aqueous solution using TiO2 doped Activated carbon, *Res. J. Chem. Environ.*, **28(1)**, 38-42 **(2024)**

28. Sharma D., Shelly K., Gandotra E. and Gupta D., Diagnosis of Covid-19 using Deep Learning, In Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing, 388-395 **(2022)**

29. Wang W., Liang D., Chen Q., Iwamoto Y., Han X.H., Zhang Q., Hu H., Lin L. and Chen Y.W., Medical image classification using deep learning, In Chen Y.W. and Jain L., eds., Deep Learning in Healthcare. Intelligent Systems Reference Library, Springer, Cham, **171**, 33-51 **(2020)**

30. Wołk K. and Wołk A., Early and remote detection of possible heartbeat problems with convolutional neural networks and multipart interactive training, *IEEE Access*, **7**, 145921-7 **(2019)**

31. World Health Organization, Cardiovascular diseases (CVDs), https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), Accessed on 30 August, 2023 **(2019)**.

(Received 17th November 2024, accepted 18th January 2025)