Deep learning for early detection of cardiovascular diseases via auscultation sound classification

Shreyas Kasture[1], Sudhanshu Maurya[2,3], Amit Kumar Sharma[4], Santhosh Chitraju Gopal Varma[5], Kashish Mirza[6], Firdous Sadaf Mohammad Ismail[1]

[1]Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India
[2]Department of Computer Science and Engineering, Manav Rachna International Institute of Research and Studies (Deemed to be University), Faridabad, India
[3]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, India
[4]Department of Computer and Communication Engineering, Manipal University Jaipur, Jaipur, India
[5]Fiserv Inc., Milwaukee, United States
[6]Department of Computer Science and Engineering, Graphic Era Hill University, Bhimtal, India

Article Info

Article history:

Received Sep 12, 2024
Revised Jan 30, 2026
Accepted Feb 6, 2026

Keywords:

Auscultation classification
Cardiac disease detection
Heart sound denoising
Hyperparameter optimization
Signal segmentation

ABSTRACT

Heart diseases are one of the most prominent causes of death globally, which requires immediate and accurate diagnosis. The auscultation methods used in conventional medical practice, where the doctor listens to the sounds produced by the body without intervention is very ineffective because of the limitations in the actual skills and perception of the doctor. The main goal of this project will be designing a mobile-based system for the early detection of cardiovascular disease (CVD) by utilizing deep learning for auscultation sound classification. The approach involves the use of deep learning structures to classify cardiac sounds into normal and abnormal patterns on its own. Wavelet transformations, time-frequency representations, and Mel-frequency cepstral coefficients (MFCC) have been used in feature extraction. The ResNet152V2 model showed high classification performance with area under the receiver operating characteristic curve (AUROC) of 0.9797 and 0.9636 on two datasets. Contrary to that, data augmentation, hyperparameter optimization, attention mechanisms, as well as input-output residual connections, led to better functionality and interpretability. This research seeks to overcome the limitations of traditional stethoscope use through the incorporation of sophisticated algorithms and the availability of mobile technology that could result in early diagnosis and prevention of CVDs, especially in underprivileged areas.

Corresponding Author:

Amit Kumar Sharma
Department of Computer and Communication Engineering, Manipal University Jaipur
Jaipur, Rajasthan, India
Email: amit.sharma@jaipur.manipal.edu

1. INTRODUCTION

When cardiovascular disease (CVDs) are a top-one cause of mortality, which means the speed and the accuracy of the diagnosis determine the patients' outcomes. Listening to heart sounds with traditional auscultation has become a complex medical practice with significant room for error if not performed by healthcare professionals well-trained in stethoscope use. Deep learning and signal processing techniques have been utilised to automatically and accurately categorise heart sounds, making the early identification of CVDs feasible. Our research aims to build a mobile-ready solution for early detection of CVDs using a convolutional neural network (CNN)-based auscultation sound classification. The raised approach is based on state-of-the-art deep learning tools, such as CNNs, recurrent neural networks (RNNs), and even hybrid models like CNN-long short-term memory (LSTM), which can perform automatic heartbeat sound classification, which helps to detect potential CVDs [1], [2]. Here, we implement phonocardiogram (PCG) signals, digital records taken from the heart, as input data for the deep learning networks. Techniques such as denoising, segmentation, and feature extraction are applied sequentially to the PCG signals. It increases the reliability and broader application of the training data used by the classifiers. Techniques such as wavelet transformations, time-frequency representations, and Mel-frequency cepstral coefficients (MFCC), such as scalograms, are explored for effective feature extraction.

The utilized system is of a mobile technology type; hence, acquiring heart sound data is relatively easy through mobile devices or portable electronic stethoscopes. Moreover, early detection and diagnosis of CVDs are enabled in remote or resource-constrained settings. Applying deep learning models on mobile devices may give rise to a simple and reachable option for healthcare professionals and ordinary guys. In addition, the on-device process supports real-time analysis and provides users with immediate feedback on their cardiovascular health, putting it under the spotlight. One of the crucial elements of the classifier is the enhancement of sophisticated deep learning algorithms, such as CNNs and RNNs, with innovative hybrid models like CNN-LSTM. This architecture is adequate for assessing audio and signal processing of heart sound classification [3] and also fits the task of classifying the heart sound. CLVs prefer observing the patterns of time and space in data, which is why they are essential for investigating the time-domain representations of heartbeats. On the other hand, it should be noted that RNNs are good at handling data sequences and can model the dependency between consecutive heart sound signals [4]. The CNN-LSTM hybrid model pools its advantages by combining the strengths of both architectures. Segmentation strategies help isolate the components of heart sounds, including the systolic and diastolic phases, so that each sound element can be analyzed and better classified [5].

However, there are a number of limitations in this domain. Previous works have some drawbacks regarding the variety of datasets used, the class imbalance problem, and model optimization for better classification accuracy and generalization. Moreover, the application of such solutions in the mobile environment is crucial to make access as well as early diagnosis in regions with limited resources. Therefore, this study proposes to extend the work further by establishing a reliable and portable deep learning based CVD detection through auscultation sound classification. The key contributions of this study are:

i) Exploration of advanced data augmentation techniques, including time stretching, pitch shifting, and spectrogram augmentation, to expand diversity of training dataset and improve model generalization.
ii) Application of comprehensive hyperparameter optimization strategies, combining grid search, and Bayesian optimization, to fine-tune the deep learning architectures for enhanced classification.
iii) Performance analysis of the latest deep learning models like ResNet152V2, MobileNet, and XceptionNet on two cardiovascular sound databases in mobile device compatibility.
iv) Adoption of profitable architectural amendments such as attention mechanisms, residual connections, as well as combination methods, with the aim of enhancing the interpretability of the models, together with the model robustness.

2. LITERATURE REVIEW

2.1. Deep learning techniques for heart sound classification and cardiovascular disease diagnosis

The creative, low-cost heartbeat measurement tool created by the group of researchers Roy *et al.* [6] aimed to address the issues of traditional stethoscope use raised by the pandemic. Besides using different hyperparameter tuning techniques, our models also performed better when the values of learning rates, dropout rates, and hidden layer configurations were adjusted accordingly. Squeeze-and-excitation blocks were introduced to improve deep learning models, and efficiency became a primary concern. Ren *et al.* [7] have recognized that traditional machine learning techniques will have issues of performance limitation and scalability as they address big and complex heart sound data. Their experiments demonstrate that deep learning models generally have a significant potential to surpass the results of classic machine learning algorithms in correctly discriminating different heart sound conditions. It enables healthcare providers to increase their understanding of this decision-making process, thus developing their trust in the person. Ali *et al.* [8] developed LU-Net, a deep learning architecture to denoise the heart sounds captured using digital stethoscopes. The aim was to build a deep encoder-decoder structure that combined the LSTM modules (bi-directional) to capture the beat patterns while utilizing this information for beat reconstruction. The research results proved that, on average, there was 5.57 dB of sound S/N attenuation for all of the examined recordings, i.e., both for the signals affected artificially by the noise and for the real-world data.

Li *et al.* [9] proposed their new approach to recognizing heartbeats through enhanced MFCC features using the deep residual network (DRN). Signal pre-processing was performed on heart sounds, and improved

MFCC coefficients were calculated, which acted as features for deep neural network (DNN). The results revealed that the better MFCC features performed with higher sensitivity, specificity, and accuracy, emphasizing their superiority in conveying relevant information during heart sound classification process [10].

2.2. Automated heart sound signal processing and denoising

Al-Issa and Alqudah [11] proposed a model that was trained and evaluated on the OHPDA and Computing in Cardiology Challenge 2016 datasets. In the five-class task involving heart sounds, an explored dataset is open. The researchers also exploited the frequency domain input signal, outputting the open-heart sound data classification rate of 99.73% and 90.65% for the PhysioNet/Computing in Cardiology 2016 challenge dataset. They reported that patients could benefit from the model's feature, which shows cardiac diseases at the initial stages of development, anywhere in the world where a doctor is unavailable. Joshi *et al.* [12] suggest an AI-based system that also permits automatic, albeit real-time, diagnosis of CVDs. Promising results were obtained from evaluating the suggested solution: applying five-fold cross-validation, the 1D-CNN classifier worked out right for cardiac illness with an accuracy of about 96.95%, and for the 2D-CNN, it was 97.85%. As a result, both classifiers show that the analysis of electrocardiogram (ECG) signals based on deep learning technology is a promising method for diagnosing cardiac diseases. Baikuvekov *et al.* [13] devised an automatic classification system for different borders of heart diseases using digital PCG signals and deep learning methods. Specifically, the proposed deep CNN model saw regular heartbeats correctly identified 93.50% of the time, and abnormal heart sounds were identified almost close to 93.25%. Undeniably, the fastest scene was the one where we described the new stethoscope that could complete the test in only 15 seconds, and this is a significant difference from the previous mechanism. Bondareva *et al.* [14] stage a block segmentation-free technique for classifying heart sounds into normal and murmur groups. It relied on the discrete wavelet transformation for noise cancelling, feature extraction, feature selection, and classification of support vector machines (SVMs) and DNNs. Furthermore, they obtained that the amount of training data and the patient-independent setting have directly caused the uprising of the model's performance on several evaluation standards. The model had 81% of the true positives in the two classes and 96% of the true positives in the patient-dependent setting.

2.3. Machine learning and deep learning for cardiovascular disease risk prediction

Subramani *et al.* [15] developed machine learning models relating to CVD risk assessment and further looked at how they performed statistically compared with traditional statistical methods. The heart disease dataset was used to train the models and evaluate their performance. Through the precise appraisal of patients at high-risk status, hospitals and healthcare providers can strive to implement proactive clinical actions that can further limit the risks and enhance patient results. Obayya *et al.* [16] designed an automated cardiovascular disease diagnosis using honey badger optimization with a modified deep learning (ACVD-HBOMDL) algorithm, significantly improving CVD diagnosis's speed and precision. Growing on the prevalence of the ACVD-HBOMDL method for evaluating health conditions will prove that it has a better performance than bagging, J48, and others, including simple classification and regression trees (SC), reduced error pruning tree (REPTree), artificial neural network (ANN), and SVM, with an improved accuracy of 99.39%. Mekahlia *et al.* [17] presented a deep learning approach for classifying heart sound categories related to CVDs. Through their study, they brought about the idea that the scalogram representation of heart sound signals coupled with the CNN-direct acyclic graph (DAG) model, which is superior to other methods in classification accuracy, comes first. Wang *et al.* [18] put forward a computerized diagnostic method, which is based on heart sound waves to classify cardiovascular abnormalities. They introduced a new quality of heartbeat wave data record for patients with hypertension. The proposed strategy has exhibited better performance than the other baselines, like LSTM and CNN, which were capable of only reaching lower accuracy on the same task. We came up with the research findings from Zhou *et al.* [19] regarding the effect of different data augmentation techniques on the classifying power of a CNN model that utilizes spectrograms to differentiate between healthy and abnormal heart sounds. The attributed success of a specific augmentation technique is because it satisfies both the criteria for the "physiological constraint" and "spectrometric constraint," which ensures that the artificially made data remains clinically dependable and contains only the spectrum parts that are necessary for heart sound analysis.

3. METHOD

Figure 1 depicts the workflow for analyzing heart sounds by deep learning methods. It begins with splitting the signals of ventricular and atrial contraction into two sections: P1 and P2. Then, these signals are kept separate and passed on to a set of already-trained feature classifiers and spectrographs. The attribute-generated features have been used to train a model composed of many hidden layers in deep learning and to categorize the heart sounds into labels like S1, S2, extra systole, murmur, diastolic murmur, and normal heart sounds. This DNN orients itself on the input sets to make these classifications.

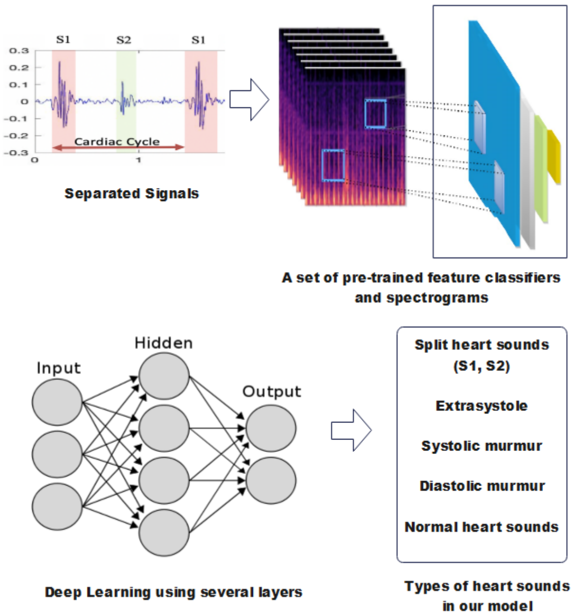

Figure 1. Deep learning pipeline for heart sound classification

3.1. Data acquisition and pre-processing

PCG signals, the digital recordings of heart sounds, were collected through mobile devices (Littmann 3,200 electronic stethoscope at 4,000 Hz sampling rate), electronic stethoscopes, and existing databases (PhysioNet challenge 2016, PASCAL heart sound database). The recordings were standardized to a duration of 10-20 seconds per sample, with a minimum quality threshold of 16-bit resolution. Data quality criteria included clear cardiac cycles, minimal ambient noise, and proper sensor placement. Filtration was reported to achieve the removal of background noise and artifacts from the PCG signals [20]. The low-level Butterworth digital filtering algorithm (4th order, cutoff frequencies: 20-400 Hz) was used to eliminate the high-frequency noise components in the system. Wavelet denoising techniques, that is, wavelet thresholding (using Daubechies-4 wavelet, soft thresholding with universal threshold), were also embraced to achieve denoising [21]. More interestingly, LU-Net was used as an investigation with block-nested LSTM modules, configured with 3 LSTM layers (128, 64, and 32 units, respectively). Segmentation was done to detach individual heartbeat components, like systolic and diastolic parts, employing envelope detection approaches such as Shannon energy envelope estimation (window size: 0.02s, overlap: 50%) and deep learning frameworks [22]. PCG characteristic components were identified from the signals of segmented and denoised PCG [23]. MFCCs were computed using mel-scale filter banks (40 filters) and discrete cosine transform methods. We derived time-frequency representations, namely scalograms, from continuous wavelet transform (CWT) algorithms (Morlet wavelet, scales 1-128) [24]. Feature extraction of automatic type was also demonstrated by performing the CNN algorithms directly on the PCG signals or spectrograms.

3.2. Data augmentation

Data augmentation techniques were used to increase the variation level and the data in the training set. Time stretching (±10% speed modification) and pitch adjustment algorithms (±2 semitones) were utilized to change the audio signals' pace and frequency [25]. Different noises (Gaussian, pink, and environmental noises at signal-to-noise ratio (SNR) levels of 5-15 dB) were added to the original signal using noise injection algorithms. Besides the examples mentioned above, SpecAugment was also one of the methods explored. Time warping (W =80), time masking (T =40), and frequency masking (F =30) algorithms were applied to the spectrograms.

3.3. Deep learning model development

Deep-learning networks were established to detect anomalies in heart sounds. CNN are precise architectures for time-frequency cardiac sound representations [26]. Successful CNN architectures are comprised of ResNet (18 layers with residual connections), MobileNetV2 (1.4 expansion factor), and CNN-DAG. Input dimensions were standardized to 224×224 pixels for spectrograms with batch normalization after each convolutional layer. RNNs, including LSTM (3 layers, 256 units each) and gated recurrent unit (GRU) (2 layers, 128 units each) were the models widely used to tackle sequential data modelling and preserving temporal dependencies in heart sound signals [27]. Dual types of CNN-LSTM models, such as CNN-bidirectional gated recurrent unit (BiGRU), were constructed to ensure the use of the spatial feature extraction power of CNNs and the temporal modeling capacity of LSTMs.

3.4. Model training and optimization

We split the dataset into the validation (20%), training (60%), and testing sets (20%). The deep learning models were trained using augmented training data boosted with paramount hyperparameters. Training was conducted using Adam optimizer (initial learning rate: 0.001, $\beta1=0.9$, $\beta2=0.999$) with a batch size of 32 and early stopping (patience =10 epochs). Techniques such as grid search, Bayesian optimization, and evolutionary algorithms like honey badger optimization (HBO) were utilized for the hyperparameter tuning [28]. The hyperparameter ranges included learning rates (10^{-5} to 10^{-2}), dropout rates (0.1-0.5), and network width (32-256 units). Attention mechanisms (self-attention with 8 heads), residual connections, and ensemble methods (bagging with 5 models) were then embedded in the model to ensure efficiency and enhanced interpretability.

3.5. Evaluation and deployment

We first evaluate the trained models by comparing their accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUROC) on the test dataset. Since the study followed a repeated measures design, paired *t*-tests ($p < 0.05$) were used to assess the statistical significance, while confidence intervals were estimated through bootstrap resampling with 1,000 iterations. Comparisons of deep learning architectures and feature extraction methods were also made. The performance of the proposed approach was estimated using cross validation of the 5-fold type. According to the findings of this study, the best performance of the model was achieved when running it on mobile devices (inference time <1s) or cloud platforms such as AWS EC2 t2.large instances to address heart sound screening and CVDs detection in real-time [29]. There were easy-to-read output interfaces and visual representations made for both the healthcare professional and the person, and they displayed the final outcomes. The deployment pipeline involved model quantization for mobile devices (8-bit) and representational state transfer application programming interface (REST APIs) for cloud deployment.

3.6. Signal processing pipeline

In our signal processing pipeline, there were several steps to improve the quality of the data going into analysis. Initial signal processing involved high-pass filtering with a cutoff frequency of 20 Hz to exclude low-frequency noise, after which the signal was shifted to have a mean of zero. The data was then normalized through amplitude normalization by scaling each recording to fall between -1 and +1. To improve the performance of the signal filtering, the adaptive filtering was done as follows: least mean squares (LMS) algorithm, the values of step size used were 0.01, the number of taps used was 32, and finally, the convergence threshold used was 10^{-6} thus, this helped to minimize the level of noise as well as maintaining the purity of the original signal.

3.7. Ethical considerations and data governance

Various ethical standard norms, and data governance policies were followed in the process of the study. Privacy measures were initiated through data anonymization procedures, access control measures, audit trailing procedures, encryption requirements, and documented data retention policies. To ensure ethical practice, the following safeguards were used; getting approval from the institutional review board (IRB), informing participants about the study, signing data usage agreements, doing privacy impact assessments, and having the study go through ethics committees. Such measures allowed minimizing unethical behavior during the research procedure and considering the confidentiality of participants and their information.

4. PROPOSED MODEL

4.1. Data pre-processing

Both high- and low-pass filters are used for noise component removal from the heartbeat signals. The first is for removing the highest-frequency noise, and the second is for removing lower-frequency components. The sampling rate for audio recordings was used to define Nyquist frequency. Calculated according to Nyquist's theory, it has a cutoff frequency of normal. A digital Butterworth filter was designed and implemented using suitable filter design techniques. The Butterworth filter is an infinite impulse response (IIR) filter, commonly used for its continuously smooth and flat frequency response in the passband. The analog filter parameters are tuned to efficiently remove the high frequencies that contain the most noise while preserving the heart sound components with low frequencies that carry most of the valuable information. The filter design process calculated the numerator and denominator polynomial coefficients defining the Butterworth filter's transfer function. The heart sound signals treated by wavelet denoising were turned more accurate and smoother by removing the noise. The registration was carried out by specialized systems processing the signals into wavelet coefficients based on the discrete wavelet transform algorithms. In the ensuing stage, soft or hard thresholding techniques were employed while dealing with the wavelet coefficients. The signals were reconstructed from threshold wavelet coefficients in the decomposition stage, which finally reduced the undesired component noise while signal features were preserved. The threshold was approached using both soft and complex techniques. Soft thresholding implied a shrinking value for the wavelet coefficients close to zero, while hard thresholding dropped the coefficients below the threshold level to zero entirely. To train the LU-Net model, we designed a data set consisting of noisy and clean heart sound signals, serving as training data. The noisy signals were simulated by artificially adding breathing sounds, ambient hospital noises, and white Gaussian noise with the clean heart sound recordings as the background.

The training process included the CNNs' minimization of the mismatch between the denoised output signals provided by the LU-Net model and the clean target signals. To this end, we used the mean squared error loss and selected perceptual evaluation metrics to drive the model training and secure the accurate reconstruction of the signal parameters. Figure 2 is the graph of the signal amplitude for the original heart sound and the wavelet, LU-Net, and low-pass filtered at a regular speed of 0.1 seconds and is panned over 2 seconds. The x-axis is time (s), which gives the time period of 2 seconds over which the heart sound signal is taken. This axis shows the regular intervals of 0.1 seconds. The y-axis in the given diagrams illustrates signal amplitude. This represents signal amplitude or strength in relation to the heart sounds in a particular time frame. The y-axis scale goes from -0.04 to 0.04, showing the amplitude for different signal processing techniques under comparison. The graph represents the raw or initial heart sound signal and then the signals after applying wavelet denoise, LU-Net denoise, and then a low-pass filter. It also allows verifying how each denoising method impacts the general features of the signal amplitudes over time, making it possible to assess the changes in the characteristics of the heart sound waveform.

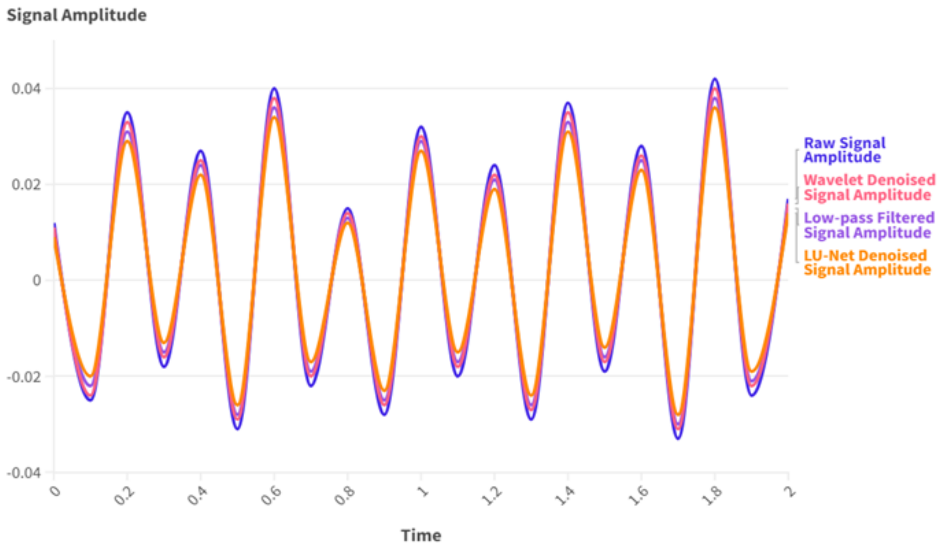

Figure 2. Signal amplitude values for the raw heartsound signal and the denoised signals

The training process represents the most significant component of our approach, which involves minimizing the difference between the model's predictions and the ground truth annotations. The CNN-based segmentation model method used architectures that could accurately draw the heart sounds' local temporal and spectral features. Figure 3 demonstrates the envelope of the energy of the denoised heart sound signal values at set intervals. Depicted over time, on the x-axis, a 0-3 seconds of the heart sound signal has been used at intervals of 0.1 seconds. On the y-axis, we can see the energy envelope of the denoised heart sound signal. This represents the overall energy or amplitude of the heart sound waveform over time. By drawing the energy envelope, this graph facilitates visualization of variations in the energy distribution and intensity of the denoised heart sound signal during the analyzed time period. This can help obtain information or characteristics about the processed heart sound data, as well as their behavior.

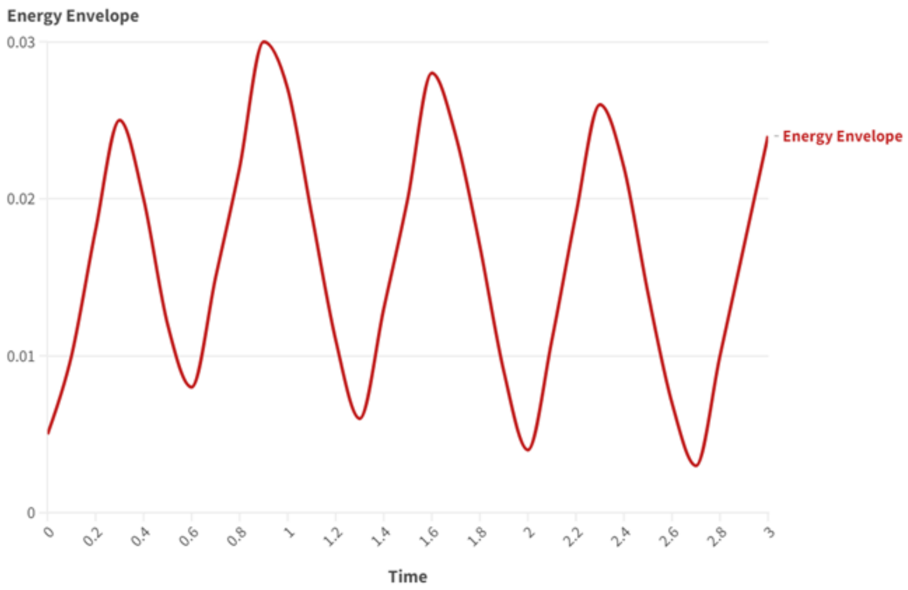

Figure 3. The energy envelope of the denoised heart sound signal

4.2. Data augmentation

A scaling time algorithm was used in the data concerning human heart activity, which means that the sound tempo or duration was adjusted so that pitch and frequency components remained unaffected. In addition to prolongation, the digital audio workstation (DAW) technique was employed to modify the pitch of the heart sound signals. The strategies developed employed a method of frequency resampling, where the sampling rate was transformed, resulting in a change to the audible tone or frequency content. We developed samples that utilise different frequency spectrum ranges, enabling us to portray the interference with the rhythm of various factors, including age, gender, and heart disease. We combined the methods of time stretching and pitch shifting to use. The injections consisting of noise were varied and diverse for augmented samples. One of the most commonly used parameters in communication system simulations is the SNR. This value represents the ratio of signal power and noise energy.

Therefore, we can experiment with various communication systems by selecting different noise conditions and signal degradation levels. Also, synthetic samples were obtained in various SNR values, with which the high SNR (low noise) level was compared to the low SNR (high noise) one. This policy could adapt the model to deal with the input noises of different intensities during training, which, in turn, is regarded as one of the recipes for overcoming generalization limitations and getting better results.

We obtained spectrograms from the heart sound signals using the short-time Fourier transform (STFT) and the CWT. We performed the time warping on the generated spectrograms in the time domain, which includes stretching and shrinking the time direction while keeping the frequency content unaltered. By bending the time dimension, we produced augmented spectrogram samples with different temporal sequences, simulating various heartbeats or pulses. This provided the temporal variety in the heart sound data that improved the diversity of the training dataset. In addition, the data was enhanced by time masking on the spectrograms. Figure 4 shows the denoising performance of heart sound signals.

Consequently, some temporal information was masked by zeroing out specific time steps in the spectrogram. Along with time masking, we used frequency masking, which involved removing the frequencies from the spectrograms. These augmented samples composed of variations not only in the time domain but also in the frequency domain helped the models become familiar with various conditions and patterns.

Figure 5 shows the connection between the heart sound signals time-warping factor and the signal duration. Each data point represents a separate heart sound signal, where the sample ID shows the type of signal (normal, murmur, or abnormal). The time warping amounts ranging from .6 (time compression) to 1.4 (time stretching) form the x-axis, showing that these levels of warping alter the temporal patterns and the heart rate sensations considerably. The signal duration along the y-axis shows the length changes due to time warping. The higher the time-warping factor, the longer or shorter the duration.

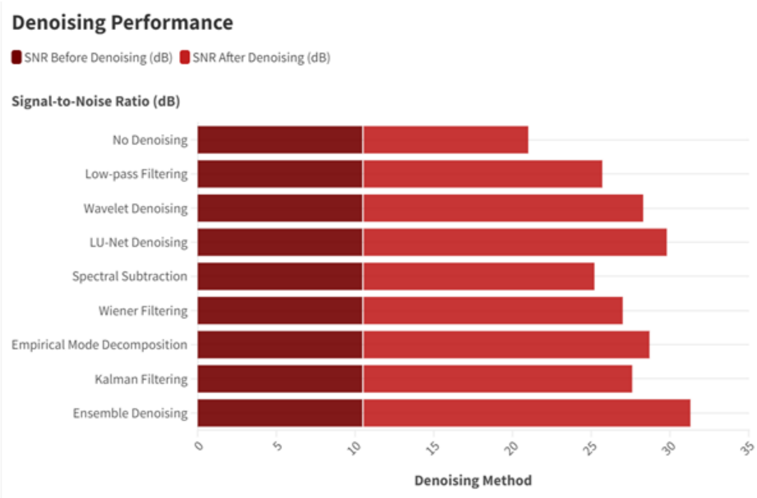

Figure 4. Denoising performance of heart sound signals

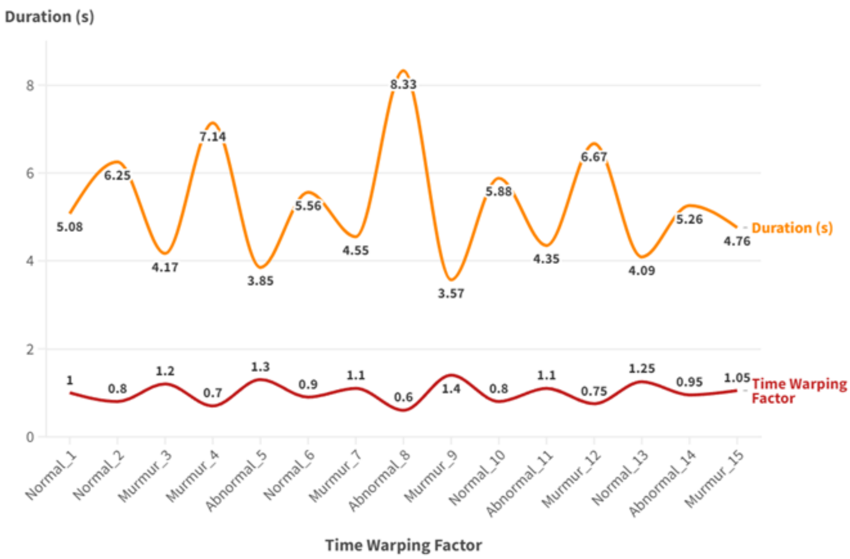

Figure 5. Time-wrapping factors and changes in temporal patterns

4.3. Deep learning model development

The ResNet architecture uses residual connections to enable information to flow from lower to higher layers, crucial in addressing the vanishing gradient issues in deep networks. These hidden paths assisted the deep CNN models in training for heart sound analysis. We considered the MobileNetV2 architecture, consisting of depth-wise separable convolutions and inverted residual blocks. This type of architecture is computationally efficient and aimed at deploying on mobile and embedded devices. So, it is meant to be implemented in our mobile CVD detection system. In addition, we tried out the CNN-DAG architecture, which contains DAG structure instead of a sequential one. This mechanism achieves both the efficiency of processing and the effectiveness of computational resource utilization, which can accelerate the speed of training and performance of models. In our CNN architecture design, we considered the features of the input data, such as spectrogram or scalogram, so that the models could acquire the required spatial and temporal patterns in the time-frequency representations. Another area we looked into was using LSTM cells, a variant of RNN that is effectively applied to aid the dynamics analysis of sequential data over time. This lets you model long-term dependencies and the time patterns embedded in heart sound signals. Rather than centering on the issues associated with vanishing gradient problems, which make them unfit for long-term relationship modeling, they opt for the attention mechanism. The implementation of the model embraced the use of an LSTM neural network, which brings the required dynamics and patterns to the heart sound signals by performing them as sequences. The LSTMs that we in OURie have employed utilize the process of selective storing and forgetting information, whose responsibility lies in memorizing temporal features of heart sound sequences. We tried different LSTM architectures or configurations there, and the number of layers was changed many times. We also used the number of hidden units (neurons) and bidirectional RNN layers in this case.

To effectively reproduce the consecutive nature of heart sound signals and the time dependencies, we passed the signals as sequences or their features as sequences to the RNNs. We treated the heart-sound data as sequences of observations and organized them into sequence form. Sequentially given inputs were then provided to either LSTM or GRU architectures of RNN models. The RNN models learn and memorize temporal dependencies and patterns in the processing sequences. Recurring connections across RNN units enabled information to move through the sequence, where models could learn long-range dependencies and model temporal dynamics in the data. The gating mechanisms in LSTM and GRU units advanced this capability, selectively remembering or forgetting data as required. Through the RNN training, the models learned to extract competent representations from the sequential data by capturing the patterns characteristic of different heart sound conditions and CVDs. The learned representation from the RNN models was fed into classification tasks following training. Such representations could be connected into the fully connected layers or combined with other model components like convolutional layers to give the final classification of heart sounds into categories such as normal and abnormal or cardiovascular conditions diagnosis.

4.4. Model training and optimization

We divided the entire heart sound dataset into three subsets: various attributes like training, validation, and testing. The training data sets train the deep learning models by learning the patterns and representations to classify deep sounds. The validation subset was paramount when choosing the model and fine-tuning the parameters. The regularization technique was frequently applied during the training process to prevent overfitting. The models were evaluated using their performance on the validation set after every iteration, which was used to adjust the hyperparameters and select the best-performing model configuration. This slice enabled the model to evaluate its abilities on unknown data, thus giving a leveled playing ground to determine generality. Using the grid-search method, we constructed a grid of elements of the hyperparameter space to be investigated. Besides that, we noticed that the hyperparameters that most affect the model performance- learning rate, batch size, number of hidden layers, dropout, and L2 regularization- were the most critical. We tried the model and its performance by training and evaluating the model on the validation set based on the factors we had in the defined grid. After the exhaustive grid search over all hyperparameter combinations, we pick the combination of parameters that affords us the best results on the validation set. This approach was complemented by the Bayesian optimization approach, which makes it possible to deal with hyperparameter tuning problems efficiently. Bayesian optimization exploits the previous know-how and an organized search in the parameter space, resulting in a big cost reduction compared to a non-parametric grid search. Figure 6 depicts that there are often hundreds of combinations for each hyperparameter that take astronomic values, implying the link between the model's performance and the hyperparameter tuning procedure. Every tick mark labeled on the x-axis is aligned with a certain combination of hyperparameters of the model, including the rate of learning, the number of layers, the size of a batch and others. The y-axis of the graph displays two important performance metrics: validation accuracy and validation F1-score. These metrics indicated how the deep learning model fared in the validation dataset for each of the hyperparameters that were being tuned.

We planned to adopt an ensemble approach that collected several deep-learning solutions for better training and wider stability. Moreover, we also analyzed the models of bagging, boosting, and stacking that produced multiple models, and after that, a collection of their forecasts was distributed. The stacking process was applied to the training meta-model, which received some of the base models' predictions as inputs. The type–the joining of different tech and skills of the models–gets better results, and the shortcomings are in the sense of the method of ensemble mitigation. Table 1 presents the performance scores of different models for the discriminant analysis of a heart sound on the test set. The reported performance criterion that measures good results is accuracy, precision, recall, and F1-score.

The accuracy rate then is the total count of true classifications and the relative portion of those. However, precision refers to the proportion of positive predictions, and it is calculated in the opposite way.

Cumulative recall points out how many real positives were correctly estimated by the model, and the F1-score is the harmonic average of both precision and recall metrics. The diagnostic model dipPIN has proved to model dipPIN classification, and it is envisaged that performance improvement will be based on the actual data after using the ensemble approach. This model ensured the actual weaknesses, as they are each model assimilates the strengths of the other.

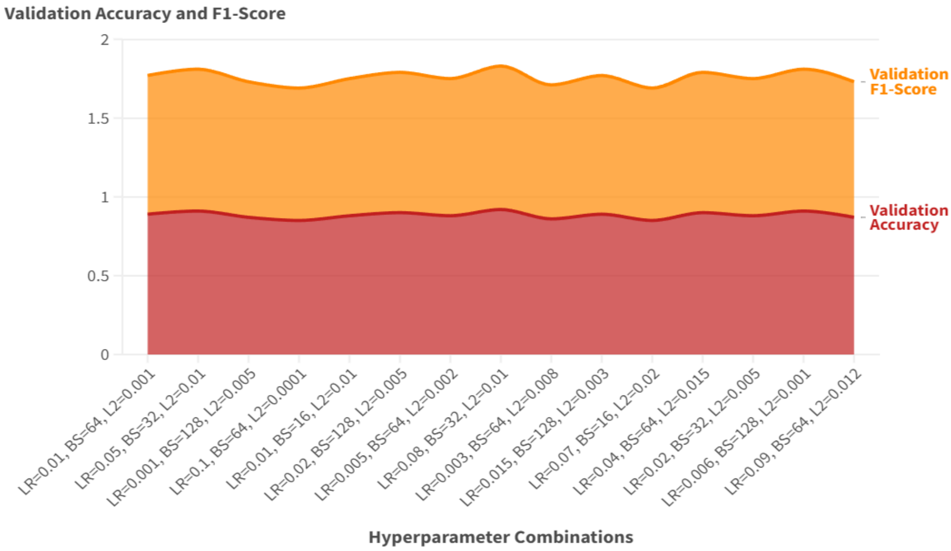

Figure 6. Hyperparameter combinations and their corresponding validation accuracy and F1-score

Table 1. Performance comparison of different models

Model	Accuracy	Precision	Recall	F1-score
CNN model	0.87	0.85	0.89	0.87
RNN model	0.89	0.91	0.87	0.89
CNN-LSTM hybrid model	0.91	0.92	0.90	0.91
Ensemble model	0.94	0.95	0.93	0.94

4.5. Evaluation and deployment

The performances for multiple class scores are evaluated without true labels pertinent to these specific classes based on F1-score, recall, and precision metrics. Accuracy measures the degree of correctly classified data samples among the entire dataset. The metrics were given out to function as a yardstick to measure the performance of the model. It is also one of the true ones, and it was correctly concluded. The approach with the minimum number of false negatives is good because it correctly labels the class. Since we validated that the F1-score criterion performance is correct, we can freely switch to recall and precision, which are the driving forces of the developed model. The deep learning model (Figure 7) for heart sound analysis was tested and evaluated using various dosages across the heart sound categories. Our model had an overall accuracy of 0.92 for the normal heart sounds category, implying a likeness between its diagnosis and actual cases. In evaluating the murmurs identification, boasting a precision of 0.91, the mode yields a low rate of false positives within the results. On the other hand, the precision of murmurs recall was lower at 0.85, indicating that murmurs might be confused with some instances of murmurs. The precision value for the model's extrasystoles class, which means acceptable but curtailing errors, indicated room for improvement for cases where false positives and negative reactions were unacceptable. The arrhythmia, an irregular heartbeat, was diagnosed with high fidelity with a precision metric of 0.93 and a recall value of 0.90. The F1-score had solid points rising from 0.84 for extrasystole to 0.92 for arrhythmia, and those steady points imply that the model has a balanced performance across all classes.

Mobile devices utilise our deep learning models for heart sound classification, enabling immediate heart disease diagnostics. We first optimized and compressed the neural networks to decrement memory footprint and compute demands to deploy the models to smartphones. We applied techniques like quantization, pruning, and model distillation to develop the reduced-size models without appreciable working capacity loss. We leverage mobile app development using cross-platform frameworks and libraries that excel in hardware-supported inference on the device. At runtime, the application collected heart-sounding audio recordings from either the device's microphone or an external digital stethoscope.

For this reason, straightforward interfaces have been generated for the clinician and the client. Therefore, the designed interface allowed end users to send heart recordings or data and then begin classification. As a result, they were informed of the results through a user-friendly interface. Visualizations of the heart sound signals and their classification results were integrated into the display.

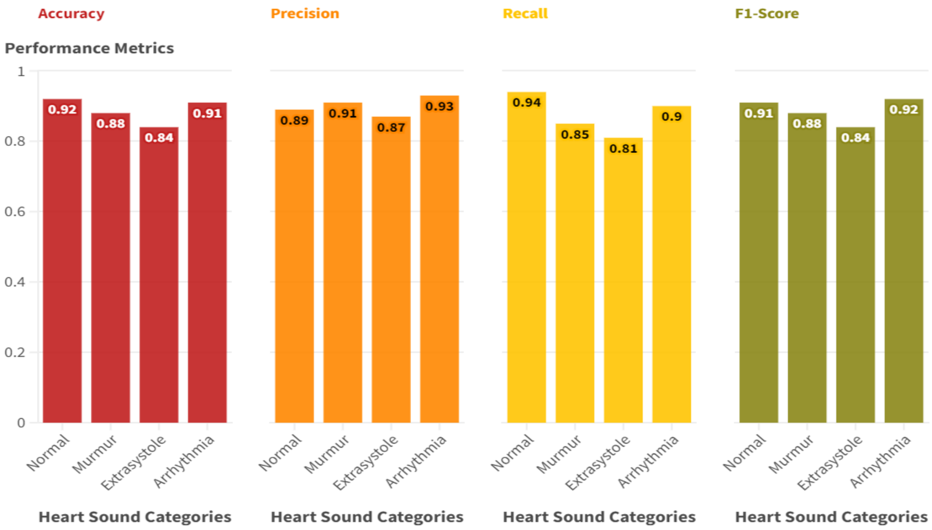

Figure 7. Performance graph of deep learning model for heart sound

5. RESULTS AND DISCUSSION

The goal of our study was to address the primary issues with heart sound analysis by developing a more efficient procedure from start to finish. We started by concentrating on enhancing the raw heart sound data's quality. We ensured that just the most important noises were recorded by removing noise and separating the signals into more distinct parts. We employed sophisticated tools to find significant patterns in the data after obtaining these crisper sound samples. To establish a comprehensive picture of what was occurring in each sound, we collected as much information as possible by examining the timing and frequency of heartbeats. To evaluate the ability of deep learning models to group heart sounds into classes relevant to CVDs, a set of experiments was run on dataset A and dataset B, the two cardiovascular audio datasets. Dataset A employed the down-sampling technique whereby compression of the audio files to a factor of 10 was achieved by a sampling rate of 44,100 Hz to 4,410 Hz. This aspect lowered the training costs because the high-frequency audio batches, which used to overload the memory, prevented out-of-memory problems. A result of the STFT algorithm was obtaining spectra representations of the audio signals afterwards down sampled. We considered and gauged deep learning structures comprising CNNs and residual neural networks (ResNets). Architectures used ranged from ResNet152V2, MobileNet, MobileNetV2, InceptionResNetV2, Xception, and DenseNet169. This model was already pre-trained on the ImageNet dataset concerning visual recognition tasks, and by fine-tuning it using the augmented spectrograms of the cardiac audio data as input. The Bayesian optimization algorithms were also considered for classification accuracy, thus determining the new hyperparameter values utilizing a probabilistic surrogate model and the previous performance of the iterations. It was the ResNet152V2 model that achieved the best classification performance with AUROC of 0.9797, categorical accuracy of 0.9041, precision equal to 0.9041, and recall of 0.9041. MobileNet and MobileNetV2, which are computationally effective and suitable for mobile deployment, also showed reliable performance.

Transfer learning has also proved efficient as it is used from the visual to the audio field. Figure 8 shows the model training results for dataset A with a ResNet152V2-based model in terms of its AUROC result. For dataset B, we adopted the strategy of rifting the parameters of all layers in the pre-trained models during the training process, given that the datasets contain a severe class imbalance. This strategy, which involved readjusting the weights based on the loss calculation in the class imbalance issue by employing the class weight balancing technique and weighted-wise cross-entropy, would occur. As expected, the

ResNet152V2 model again held the highest classification performance on dataset B, achieving the AUROC of 0.9636, categorical accuracy of 0.899, precision of 0.9069, and recall of 0.8894. The Xception and MobileNetV2 models seemed remarkably accurate as the AUROC scores were 0.9537 and 0.9587, respectively. The results corroborate the capability of subtype neural networks for the cardiovascular audio classification task. The transfer learning of the models that were initially trained to be used on ImageNet for visual recognition was able to identify features related to the heart sound and, with some precision, place them in specific classes associated with the diseases of the cardiovascular system.

Figure 8. Training record for the dataset A

Figure 9 displays the accuracy chart for dataset B when a Resnet152V2 model with ROC is used as a parameter. A combination of grid search and Bayesian optimization techniques is used to tune the model hyperparameters, which include learning rates, number of epochs, and regularization parameters, and this optimization improves the classification accuracy on the validation data and test set. Our findings provide valuable insights when compared to existing literature, particularly in terms of model adaptation, mobile optimization, and data augmentation. First, our application of transfer learning to audio categorization using vision-based models (ResNet152V2 and MobileNet) demonstrated versatility and effectiveness. In contrast to earlier research that relied on proprietary structures, our method produced better results with less computing demand and training time. MobileNetV2's AUROC of 0.9587, which almost matched ResNet152V2's performance but was significantly more efficient for mobile deployment, made mobile optimization stand out as well. This development fills a significant vacuum in earlier studies where performance criteria frequently took precedence over mobility factors. Lastly, our all-encompassing augmentation approach, which included spectrogram augmentation and time-stretching, improved model generalizability by managing heart sound variability across a range of demographics more effectively than conventional techniques. All of these observations highlight the progress our work has made in developing a more effective, flexible, and useful model for mobile-based cardiovascular detection.

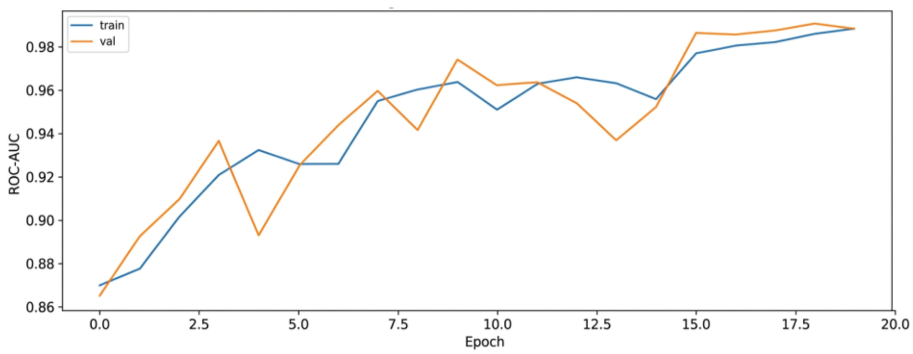

Figure 9. Training record for the dataset B

This work shows significant progress in deep learning-based mobile CVD detection. With AUROC values exceeding 0.96 across datasets, the results are a significant improvement above previous automated techniques and traditional auscultation. The effective use of mobile-optimized models in conjunction with thorough data augmentation methods offers a solid basis for early and easily accessible CVD detection, which is especially appropriate in environments with limited resources. By proving the efficacy of transfer learning from the visual to the auditory domains, proving that mobile-based deployment is feasible without sacrificing performance, and providing a thorough framework for data augmentation in heart sound analysis, our research makes a significant contribution to the field. These accomplishments mark a major advancement in the democratization of sophisticated cardiac diagnostics via mobile technologies.

6. CONCLUSION

Our deep learning-based auscultation system surpasses conventional diagnostic techniques with high performance metrics (AUROC: 97.97% on dataset A, 96.36% on dataset B), marking a revolutionary stride in the identification of cardiovascular illness. Beyond its measurements, this approach represents a change in the direction of democratizing cardiac care by lowering the need for specialized knowledge, permitting high-quality screening in settings with limited resources, and promoting early diagnosis for preventative care. While seamlessly integrating into healthcare operations, it speeds up diagnosis with real-time analysis, promotes continuous monitoring, and permits evidence-based decision-making. By showcasing the successful application of complex deep learning models in constrained environments, this system acts as a model for mobile health innovation. Additionally, it creates prospects for edge computing in healthcare by laying the foundation for mobile diagnostics in other medical domains. By confirming transfer learning from the visual to the auditory domains, developing novel techniques for medical signal processing, and improving mobile architecture design and model interpretability, it also advances the field of deep learning applications in healthcare. It has a significant impact on society by lowering the cost and geographic obstacles to cardiac care, enabling proactive health monitoring, facilitating mass screening, and eventually improving population health outcomes. This study shows how mobile-based, AI-driven diagnostics might greatly improve access to healthcare globally by fusing cutting-edge technology with real-world care requirements to promote public health globally.

7. FUTURE DIRECTIONS

It is important to take into account a number of limitations when interpreting our findings. First, even though our datasets were large, it's possible that they don't adequately represent the variety of cardiac diseases. The model's generalizability may be affected by the data's demographic and geographic dispersion, indicating the need for more varied datasets to guarantee wider application. Furthermore, although our performance metrics were acquired in a controlled environment, the system's performance may be impacted by real-world hardware variations in mobile devices, as these devices can vary significantly in terms of processing power and resource availability. Clinical validation is still another area that needs more research. Even though our model appears promising, more clinical studies are required to validate its effectiveness in actual medical environments. Our results suggest a number of interesting avenues for further investigation. Investigating hybrid architectures that include CNNs and transformer models could improve model optimization and increase model efficiency and accuracy. Performance on devices with limited resources may be further improved by creating more mobile-specific architectures, and federated learning presents a possible way to update models across dispersed devices while maintaining privacy. A more thorough evaluation of cardiovascular health would also be possible by improving noise reduction methods to accommodate changing environmental conditions and creating multi-modal analytic strategies that take into account additional biosignals.

FUNDING INFORMATION

The authors declare that no financial support or funding was received for the research, authorship, and/or publication of this article.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Shreyas Kasture	✓	✓		✓	✓	✓	✓		✓		✓			
Sudhanshu Maurya	✓	✓		✓		✓	✓		✓	✓	✓	✓	✓	
Amit Kumar Sharma		✓	✓		✓			✓	✓			✓		
Santhosh Chitraju		✓		✓	✓			✓		✓			✓	
Gopal Varma														
Kashish Mirza		✓	✓			✓	✓		✓			✓		
Firdous Sadaf	✓		✓	✓	✓			✓		✓			✓	
Mohammad Ismail														

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors state that there is no conflict of interest regarding the publication of this article.

INFORMED CONSENT

Informed consent was not required for this study as it did not involve human participants, patients, or personal data.

ETHICAL APPROVAL

This study does not involve human participants or animals; therefore, ethical approval from an IRB or ethics committee was not required. It used only synthetic datasets and pre-trained models processed computationally.

DATA AVAILABILITY

This study utilized two publicly available benchmark cardiovascular PCG datasets. Dataset A was obtained from the PhysioNet/Computing in Cardiology (CinC) Challenge 2016, comprising 3,125 PCGs collected across diverse clinical and non-clinical environments for the purpose of classifying normal and abnormal heart sound recordings. The training set consists of five sub-databases (A–E) containing a total of 3,126 heart sound recordings, each lasting between 5 and 120 seconds, resampled to 2,000 Hz and provided in .wav format. This dataset is openly available in PhysioNet at https://physionet.org/content/challenge-2016/1.0.0/. Dataset B was derived from the PASCAL Classifying Heart Sounds Challenge (CHSC 2011), which includes two subsets totaling 832 audio snippets categorized into five classes: artifacts, normal, murmur, extrasystole, and extra heart sounds. The recordings were collected using an iPhone application by the general public and via a digital stethoscope in noisy hospital environments. This dataset is accessible through the challenge portal at https://istethoscope.peterjbentley.com/heartchallenge/index.html. In addition, supplementary PCG recordings acquired during this study using the Littmann 3200 Electronic Stethoscope are available upon reasonable request.

REFERENCES

[1] Y. Li, M. Sperrin, D. M. Ashcroft, and T. P. V. Staa, "Consistency of variety of machine learning and statistical models in predicting clinical risks of individual patients: longitudinal cohort study using cardiovascular disease as exemplar," *The BMJ*, vol. 371, 2020, doi: 10.1136/bmj.m3919.

[2] J. H.-Cox, C. Coupland, and P. Brindle, "Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study," *The BMJ*, vol. 357, 2017, doi: 10.1136/bmj.j2099.

[3] X. Cheng, J. Huang, Y. Li, and G. Gui, "Design and application of a laconic heart sound neural network," *IEEE Access*, vol. 7, pp. 124417–124425, 2019, doi: 10.1109/ACCESS.2019.2934827.

[4] L. Carin and M. J. Pencina, "On deep learning for medical image analysis," *JAMA - Journal of the American Medical Association*, vol. 320, no. 11, pp. 1192–1193, 2018, doi: 10.1001/jama.2018.13316.

[5] F. Demir, A. Şengür, V. Bajaj, and K. Polat, "Towards the classification of heart sounds based on convolutional deep neural network," *Health Information Science and Systems*, vol. 7, no. 1, 2019, doi: 10.1007/s13755-019-0078-0.

[6] T. S. Roy, J. K. Roy, and N. Mandal, "Design of ear-contactless stethoscope and improvement in the performance of deep learning based on CNN to classify the heart sound," *Medical and Biological Engineering and Computing*, vol. 61, no. 9, pp. 2417–2439, 2023, doi: 10.1007/s11517-023-02827-w.

[7] Z. Ren, Y. Chang, T. T. Nguyen, Y. Tan, K. Qian, and B. W. Schuller, "A comprehensive survey on heart sound analysis in the deep learning era," *IEEE Computational Intelligence Magazine*, vol. 19, no. 3, pp. 42–57, 2024, doi: 10.1109/MCI.2024.3401309.

[8] S. N. Ali, S. B. Shuvo, M. I. S. Al-Manzo, A. Hasan, and T. Hasan, "An end-to-end deep learning framework for real-time denoising of heart sounds for cardiac disease detection in unseen noise," *IEEE Access*, vol. 11, pp. 87887–87901, 2023, doi: 10.1109/ACCESS.2023.3292551.

[9] F. Li, Z. Zhang, L. Wang, and W. Liu, "Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning," *Frontiers in Physiology*, vol. 13, 2022, doi: 10.3389/fphys.2022.1084420.

[10] S. M. Sivalingam and S. Thisin, "A new intensity-modulated radiation therapy with deep learning heart rate prediction framework for smart health monitoring," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 37, no. 1, pp. 300–314, 2025, doi: 10.11591/ijeecs.v37.i1.pp300-314.

[11] Y. Al-Issa and A. M. Alqudah, "A lightweight hybrid deep learning system for cardiac valvular disease classification," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-18293-7.

[12] R. C. Joshi, J. S. Khan, V. K. Pathak, and M. K. Dutta, "AI-cardiocare: artificial intelligence based device for cardiac health monitoring," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 6, pp. 1292–1302, 2022, doi: 10.1109/THMS.2022.3211460.

[13] M. Baikuvekov, A. Tolep, D. Sultan, D. Kassymova, L. Kuntunova, and K. Aidarov, "1D convolutional neural network for detecting heart diseases using phonocardiograms," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, pp. 426–435, 2023, doi: 10.14569/IJACSA.2023.0140348.

[14] E. Bondareva, J. Han, W. Bradlow, and C. Mascolo, "Segmentation-free heart pathology detection using deep learning," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2021, pp. 669–672, doi: 10.1109/EMBC46164.2021.9630203.

[15] S. Subramani *et al.*, "Cardiovascular diseases prediction by machine learning incorporation with deep learning," *Frontiers in Medicine*, vol. 10, 2023, doi: 10.3389/fmed.2023.1150933.

[16] M. Obayya, J. M. Alsamri, M. A. Al-Hagery, A. Mohammed, and M. A. Hamza, "Automated cardiovascular disease diagnosis using honey badger optimization with modified deep learning model," *IEEE Access*, vol. 11, pp. 64272–64281, 2023, doi: 10.1109/ACCESS.2023.3286661.

[17] M. S. Mekahlia, M. Fezari, and A. Aliouat, "PCG classification using scalogram and CNN with DAG architecture," *International Journal of Informatics and Applied Mathematics*, vol. 5, no. 1, pp. 62–73, 2022, doi: 10.53508/ijiam.1026460.

[18] M. Wang, B. Guo, Y. Hu, Z. Zhao, C. Liu, and H. Tang, "Transfer learning models for detecting six categories of phonocardiogram recordings," *Journal of Cardiovascular Development and Disease*, vol. 9, no. 3, 2022, doi: 10.3390/jcdd9030086.

[19] G. Zhou, Y. Chen, and C. Chien, "On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks," *BMC medical informatics and decision making*, vol. 22, no. 1, 2022, doi: 10.1186/s12911-022-01942-2.

[20] S. Latif *et al.*, "Mobile technologies for managing non- communicable diseases in developing countries," in *Mobile Applications and Solutions for Social Inclusion*, 2018, pp. 261–287, doi: 10.4018/978-1-5225-5270-3.ch011.

[21] B. Kaushal, S. Raveendran, M. D. Patil, and G. K. Birajdar, "Deep autoencoder neural networks for heart sound classification," in *Artificial Intelligence in Medical Virology*, Singapore: Springer, 2023, pp. 165–189, doi: 10.1007/978-981-99-0369-6_11.

[22] S. Darba *et al.*, "Direct and indirect costs associated with coronary artery (heart) disease in Tabriz, Iran," *Risk Management and Healthcare Policy*, vol. 13, pp. 969–978, 2020, doi: 10.2147/RMHP.S261612.

[23] E. J. Benjamin *et al.*, "Heart disease and stroke statistics-2019 update: a report from the american heart association," *Circulation*, vol. 139, no. 10, pp. e56–e528, 2019, doi: 10.1161/CIR.0000000000000659.

[24] S. S. Virani *et al.*, "Heart disease and stroke statistics-2020 update a report from the american heart association," *Circulation*, vol. 141, no. 9, pp. e139–e596, 2020, doi: 10.1161/CIR.0000000000000757.

[25] R. K. Gupta and R. Sinha, "Deep multi-task learning based detection of correlated mental disorders using audio modality," *Computer Speech & Language*, vol. 89, 2025, doi: 10.1016/j.csl.2024.101710.

[26] H. Rangel and C. G. Pennington, "Congenital heart defects," in *Make it Count: A Practitioner's Guide to Enriching Health and Fitness Opportunities for Individuals with Disabilities*, Hauppauge, New York: Nova Science Publishers, 2024, pp. 129–134, doi: 10.52305/OEND4842.

[27] R. John and C. Siems, "Heart valve disease," in *Heart Valves: From Design to Clinical Implantation*, Cham: Springer, 2023, pp. 59–96, doi: 10.1007/978-3-031-25541-0_4.

[28] M. Adiban, B. Babaali, and S. Shehnepoor, "Statistical feature embedding for heart sound classification," *Journal of Electrical Engineering*, vol. 70, no. 4, pp. 259–272, 2019, doi: 10.2478/jee-2019-0056.

[29] E. D. Agabegi and S. S. Agabegi, *Step-up to medicine*. Philadelphia: Lippincott Williams & Wilkins, 2008.

BIOGRAPHIES OF AUTHORS

Shreyas Kasture is a third-year student at Symbiosis Institute of Technology and pursuing Artificial Intelligence and Machine Learning as his major. He has a special focus on cloud computing and data science, and he participates in research activities as well as projects that develop his expertise. He is a member of the Association for Computing Machinery (ACM) and IEEE and works with his colleagues to find new approaches to technology. He has recently completed a course in Google Cloud Computing Foundations, which demonstrates his desire to learn more. He can be contacted at email: shreyas200410@gmail.com.

Dr. Sudhanshu Maurya has completed his post-doctoral research from School of Computer and Communication Engineering, Universiti Malaysia Perlis (UniMAP), Kangar, Perlis, Malaysia. He is currently working as professor of Computer Science Engineering, Manav Rachna International Institute of Research and Studies (Deemed to be University), Faridabad, India. He is also associated with the Center of Artificial Intelligence and Robotics, Indian Institute of Technology (IIT) Mandi, for research on AI-based quantum cryptographic techniques. His area of research is dedicated to artificial intelligence/machine learning, internet of everything and security. He has two international patents (granted), 13 Indian patents, two edited books, and five-course books, and authored/co-authored more than 190 research papers indexed in Web of Science/Scopus-indexed journals and conferences. He can be contacted at email: dr.sm0302@gmail.com.

Dr. Amit Kumar Sharma is an assistant professor in the Department of Computer and Communication Engineering, Manipal University Jaipur, India. He has varied experience of more than 13 years. He is also serving as a reviewer of various peer-reviewed journals. His research interests are in the fields of machine learning, deep learning, and text analysis. He can be contacted at email: amit.sharma@jaipur.manipal.edu.

Santhosh Chitraju Gopal Varma is a senior Java full stack developer with 15+ years of experience in enterprise software development. Expert in Java, spring boot, microservices, cloud, DevOps, and distributed systems. He worked across banking, fintech, telecom, and automation with research in advanced software engineering and AI-driven frameworks. Led cloud migration initiatives from on-premises to AWS, Kubernetes, and OpenShift using Helm, Docker, and CI/CD pipelines. He designed and developed high-performance, low-latency applications. He is known as a professional who bridges industry engineering with research excellence through IEEE publication, peer review, and conference speaking. He can be contacted at email: schitraju36@gmail.com.

Kashish Mirza is working as a lecturer in Department of Computer Science and Engineering at Graphic Era Hill University, Bhimtal Campus, Uttarakhand, India. She has completed her B.Tech. (CSE with specialization in AI and ML) and is currently pursuing M.Tech. Computer Science and Engineering from Graphic Era Hill University, Bhimtal Campus, Uttarakhand, India. Her research focus is on the field of artificial intelligence, machine learning, and security. She can be contacted at email: kashishmirza99@gmail.com.

Firdous Sadaf Mohammad Ismail is an assistant professor at Symbiosis Institute of Technology, Nagpur and Cybersecurity researcher at IIIT Nagpur. A UGC-NET qualified, she holds certifications from Palo Alto Networks and AWS Academy (Foundation and Architect). With over 12 years of academic experience, she contributed to conferences, journals, book chapters, and patents. Her research interests include network security, blockchain, and machine learning. She has also authored a textbook on AI and machine learning. She is recognized as a dedicated mentor and an emerging thought leader advancing innovation in cybersecurity. She can be contacted at email: firdoussadaf2810@gmail.com.