IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 December 2023, accepted 30 December 2023, date of publication 19 January 2024, date of current version 30 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3356075

RESEARCH ARTICLE

Raw Waveform-Based Custom Scalogram CRNN in Cardiac Abnormality Diagnosis

KODALI RADHA[1], MOHAN BANSAL[1,2], (Senior Member, IEEE), AND RAJEEV SHARMA[1], (Member, IEEE)

[1]School of Electronics Engineering, VIT-AP University, Amaravati, Andhra Pradesh 522237, India
[2]Indian Institute of Information Technology Sonepat, Sonipat 131001, India

Corresponding author: Rajeev Sharma (rajeev.sharma@vitap.ac.in)

ABSTRACT Cardiovascular disease is a significant cause of death worldwide, emphasizing the crucial need for timely detection and diagnosis of heart abnormalities. This study presents a new approach that utilizes deep learning models to diagnose cardiac issues by analyzing raw phonocardiogram (PCG) signals. The proposed method introduces a novel technique called custom scalogram-based convolutional recurrent neural network (CS-CRNN). Diverging from conventional techniques, this model directly handles the raw PCG signals. These signals undergo a transformation into scalogram images within the initial layer of the CRNN architecture, without incorporating any learnable parameters. The results obtained from the CS-CRNN model are compared with traditional feature-based recurrent neural network (RNN) models. The comparison demonstrates comparable performance in both binary classification (normal and abnormal categories) and multiclass classification (5 categories). The CS-CRNN model directly handles raw PCG data and employs data augmentation to enhance performance on small datasets. It achieves an accuracy of 99.6% for binary classification and 98.6% and 99.7% before and after optimization for multiclass classification on the augmented dataset. The results show that the CS-CRNN model offers comparable performance to traditional methods, making it a promising tool for diagnosing cardiac abnormalities.

INDEX TERMS Phonocardiogram signals, cardiovascular disease, recurrent neural networks, wavelet scattering transform, custom scalogram-based CRNN, Bayesian optimization.

I. INTRODUCTION

The progress made in machine learning (ML) algorithms for signal processing has led to improved detection of a range of diseases through the analysis of biomedical signals generated by the human body. These signals serve as indicators of physiological characteristics of human organs [1] and can be utilized in a non-invasive manner to diagnose various diseases with the help of ML algorithms. One such signal, the phonocardiogram (PCG), represents the sound produced by the mechanical movements of cardiac components and has great potential in diagnosing cardiovascular disease (CVD). Heart disease, also known as CVD, is the leading cause of death worldwide. The World Health Organization (WHO) reports that 17.9 million deaths annually are caused by cardiovascular diseases, about 32% among all deaths, globally [2]. Early detection of CVD is critical to initiate timely and appropriate treatment. During cardiac auscultation, the heart's sounds must be observed in order to diagnose particular cardiac diseases [3]. Phonocardiography, a common method used to perform this procedure, is a diagnostic technique that records and analyzes the sounds and murmurs of the heart throughout the cardiac cycle. Differences in the PCG signal's temporal and spectral characteristics can be observed between a healthy heart and an abnormal heart [4]. Furthermore, the type of heart disease can also affect the phonocardiography-observed sound characteristics of the heart. This study investigates the possibility of employing ML algorithms to identify different forms of cardiac diseases via phonocardiography signals. In the assessment of cardiac health, phonocardiography assumes a crucial role as it analyzes the acoustic signals produced by the

The associate editor coordinating the review of this manuscript and approving it for publication was Mostafa M. Fouda.

heart's four chambers–the atria and ventricles [5]. Through this technique, the functioning of these chambers, which are responsible for the efficient pumping and distribution of blood, can be closely monitored and evaluated.

The heart comprises four valves, namely the aortic, pulmonary, mitral, and tricuspid valves [6]. These valves open and close in rhythm with every heartbeat, facilitating the flow of blood between the chambers of the heart. The efficient functioning of the chambers and valves is essential for a healthy heart and optimal mechanical performance. Heart valve disease (HVD) is a common type of heart disease characterized by the malfunctioning of one or more of the heart valves [7]. HVD is typically classified into two types: valvular stenosis (VS) and valvular regurgitation (VR). VS occurs when the valve leaflets become stiff, reducing their ability to open fully, leading to conditions like aortic stenosis, pulmonary stenosis, mitral stenosis, and tricuspid stenosis. On the other hand, VR, also known as a leaky valve, occurs when the valve fails to close tightly, resulting in conditions such as aortic regurgitation, pulmonary regurgitation, mitral regurgitation, and tricuspid regurgitation. Mitral valve prolapse is a common cause of mitral regurgitation and involves the valve that connects the left atrium to the left ventricle. In this condition, the mitral valve leaflets become floppy and prolapse backward into the left atrium, allowing blood to flow back from the lower to the upper chamber. Aortic and mitral valves are the most commonly affected valves in HVDs. Therefore, the most common forms of HVDs include aortic stenosis (AS), mitral stenosis (MS), mitral regurgitation (MR), and mitral valve prolapse (MVP) [8]. The proper diagnosis and treatment of HVDs are crucial for maintaining a healthy heart and preventing complications. A PCG signal is a graphical presentation of the cardiac rhythm collected using a stethoscope.

Moreover, early detection and diagnosis of cardiac abnormalities is critical for ensuring timely treatment and better patient outcomes. In many cases, cardiac abnormalities can be asymptomatic or present with subtle symptoms, which can make them difficult to diagnose in the early stages. Delayed diagnosis can lead to serious complications, including heart failure, arrhythmias, and even sudden cardiac death. Machine and deep learning approaches, such as convolutional neural networks (CNNs), and recurrent neural networks (RNNs) including long short-term memory networks (LSTMs) and gated recurrent units (GRUs) along with ensemble methods, offer several advantages for detecting and diagnosing cardiac abnormalities [9]. These neural network architectures, particularly LSTMs and GRUs, are well-suited for capturing temporal dependencies in cardiac data, making them valuable tools in improving the accuracy of cardiac abnormality detection and diagnosis.

The details of the article are as follows: Section II offers a comprehensive literature review of recent state-of-the-art models used for the detection of HVDs in the last decade. Section III proposes the methodology, including the dataset, feature extraction, RNN, and custom scalogram-based convolution recurrent neural network (CS-CRNN) models for classification. Bayesian optimization is introduced in Section IV for model optimization. The experimental setup and results are presented in Sections V and VI respectively. Finally, Section VII concludes with recommendations for future work to improve the proposed methodology and patient outcomes.

II. RELATED WORK

Several experiments have been done for the detection of heart valve disorder in the last decade. A detailed review of the several methods for HVD detection is presented [10]. Transthoracic echocardiography (TTE) is one of the most often utilized procedures for detecting HVDs [11]. The TTE is the low-cost and extensive examination duration method [12]. Even with proper training, medical practitioners may struggle to assess recorded PCG signals and diagnose abnormalities. In such circumstances, computer-aided automated systems always outperform traditional methods [13]. The current automated systems use machine learning techniques to process the recorded PCG signals as inputs and classify them in different classes of HVDs. The performance of the signal classification heavily depends on the proper ML model selection and feature extraction of the signals [14]. A recent book chapter explores the application of IoT technology in detecting heart valve disorders by employing a novel amplitude and frequency-modulated signal model [15]. This aims to leverage IoT capabilities for accurate and real-time monitoring of heart valve health, potentially enhancing early detection and medical intervention for patients.

For the HVD detection, time-domain (TD), frequency-domain (FD), and time-frequency (TF) domain techniques are widely used to extract tempo-spectral features from PCG signals in literature [15], [16], [17], [18], [19]. The fast Fourier transform (FFT) [20], [21], short-time Fourier transform (STFT) [22], and TF decomposition (TFD) [23] approaches are used for HVD detection by utilizing spectral features. A time-frequency-domain deep neural network (TFD-DNN) method is used for automated heart sound activity detection using PCG signals [24], [25]. Some wavelet transform techniques [26], [27], [28], [29], [30] are also used for PCG classification to overcome the limitation of time-frequency resolution in STFT. Combined features are also used by some researchers to improve the accuracy of the classification. Son et al. [31] used the combined features of mel frequency cepstral coefficients (MFCC) and discrete wavelet transform (DWT) to classify both normal and abnormal PCG signals with the help of deep neural network (DNN).

Recently, another paper introduced a novel DNN, DsaNet, that can classify PCG signals without requiring complex feature engineering [32]. To boost its performance, the authors utilize a distinctive training technique called two-stage training, which includes randomly cropping the data

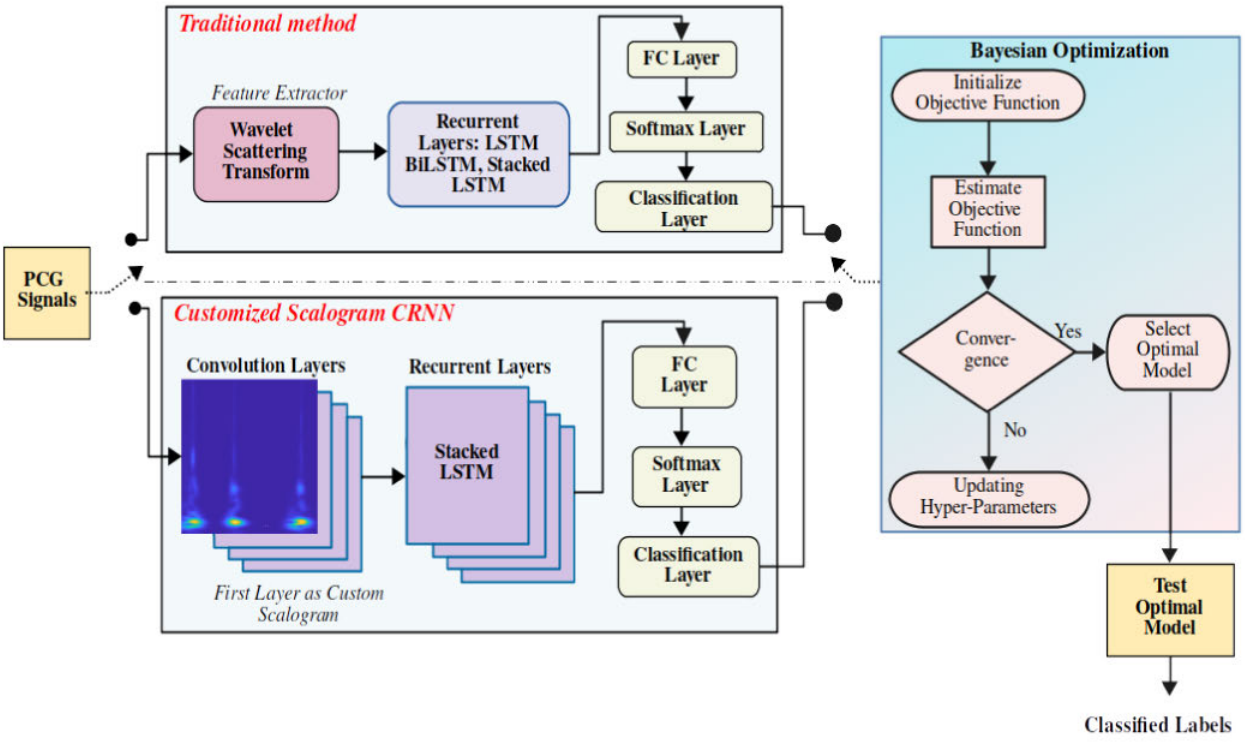


FIGURE 1. Flow chart of advanced diagnostic approach for cardiac abnormalities using traditional multi-scale RNN and CS-CRNN with Bayesian optimization.

to enhance its diversity. The experimental results, obtained on the public 2016 PhysioNet/CinC Challenge dataset, demonstrate that DsaNet outperforms seven other models, thus affirming its efficiency and effectiveness in solving the PCG signal classification challenge. In another study, the researchers investigate the classification of PCG signals into five categories using ML techniques. They use Hilbert-Huang transform (HHT) to decompose PCG signals into intrinsic mode functions (IMFs), extract MFCC features from each mode, and apply a genetic algorithm as a feature selection method [33]. They compare the performance of four ML classifiers and demonstrate that the DNN model achieves the highest accuracy. Moreover, the latest article introduces a technique that utilizes high-resolution spectrum generation, spectrogram conversion, and multi-round training to address variations in analyzing PCG signals. The experimental results demonstrate that the proposed technique using a Chirplet Z-transform-based spectrogram with multiple rounds of training achieves high accuracy in multiclass classification while maintaining low computational cost [34]. Furthermore, the methodology was validated using multiple datasets with varying signal characteristics.

The proposed customization of the scalogram-based CRNN is inspired by previous studies that incorporated log spectrogram layers as the initial layer in their networks [35], [36], [37]. The main motivation behind this research is to leverage the benefits observed in those studies and further enhance the performance and capabilities of the CS-CRNN model. It offers the flexibility of customizable scalogram computations directly in the first layer, eliminating the need for hand-crafted features. By combining these elements, the CS-CRNN demonstrates enhanced capabilities in capturing intricate patterns and relationships within the input data.

Although, recent studies have proposed various techniques for diagnosing cardiovascular diseases using phonocardiography. Many of these approaches suffer from limitations in terms of efficiency and performance, particularly in classifying binary and multiclass PCG sound signals in both balanced and imbalanced datasets. As such, there remains a gap in the development of effective and efficient models for PCG signal classification. Therefore, the proposed methodology aims to address the gap of efficiency in the binary or multiclass classification of PCG sound signals, even in datasets that are balanced or imbalanced. However, the proposed methodology for diagnosing cardiac abnormalities using PCG signals offers several important contributions:

- The article introduces a new approach called CS-CRNN. Its purpose is to employ deep learning models in the diagnosis of cardiac issues by analyzing raw PCG signals.
- The study compares two methods for diagnosing cardiac issues: one using traditional wavelet scattering features and different RNN models, and the other utilizing a more advanced approach called raw waveform-based CS-CRNN.

- The CS-CRNN model directly analyzes raw PCG signals without extracting features. It converts the signals into scalogram images at the beginning of the architecture, which can potentially improve accuracy and efficiency in diagnosis.
- The CS-CRNN model incorporates non-learnable parameters in its custom-scalogram layer, reducing the complexity of the model. This leads to a reduction in training time as the model deals with a limited number of features.
- Additionally, Bayesian optimization is employed to improve the approach's effectiveness by fine-tuning the CS-CRNN model's parameters. This technique enhances performance in diagnosing cardiac issues, allowing for more precise analysis of raw PCG signals.

III. PROPOSED SYSTEM OVERVIEW

A novel approach is proposed in the article for diagnosing cardiac abnormalities using a CS-CRNN and compared to a wavelet scattering transformed features-based RNN modeling. The proposed approach involves utilizing wavelet scattering to transform the raw PCG waveform into multi-scale features [38], which are then inputted into various RNN models such as LSTM, BiLSTM, and stacked LSTM. These models learn the temporal dependencies and classify the presence of cardiac abnormalities. Compared to feature-based RNN models, the proposed method significantly reduces system complexity by directly utilizing raw PCG waveform data with CS-CRNN, eliminating the need for manual feature extraction, as depicted in Figure 1. Bayesian optimization is employed to optimize the RNN and CS-CRNN models' performance by iteratively exploring different hyperparameter configurations. By combining the strengths of custom scalograms and deep learning, this novel approach shows promise in improving the accuracy and efficiency of cardiac abnormality detection systems using PCG signals.

A. DATASET

The dataset used in the study, discussed in [31], consists of 1000 sound files representing 5 classes of heart sound signals: AS, MR, MS, MVP, and Normal (N) heart sounds. Each class contains 200 sound files, providing a comprehensive set of data for analysis. The PCG serves as a valuable diagnostic tool for recording the sounds produced by the heart. The signals in this dataset were recorded at a sampling frequency of 8 KHz, ensuring an accurate representation of the heart sounds. By analyzing the recorded sounds, specific heart conditions can be identified. For example, AS is characterized by a narrowing of the aortic valve, resulting in a distinct ''whooshing'' sound in the PCG. Similarly, MR is indicated by a faulty mitral valve, causing a notable ''blowing'' sound. MS, on the other hand, manifests as a constricted mitral valve, producing a characteristic ''clicking'' sound. MVP is identified by the backward bulging of the mitral valve, resulting in a distinct clicking sound. Additionally, the PCG can record normal heart sounds, which serve as a reference for a healthy heart. By utilizing the PCG in conjunction with the dataset, physicians can diagnose heart conditions accurately and plan appropriate treatment. The PCG, with its ability to provide detailed insights into the heart's health and function, plays a crucial role in the diagnosis and management of cardiac conditions.

B. WAVELET SCATTERING TRANSFORM

The wavelet scattering transform (WST) is a potent characteristic extraction technique that is particularly valuable in handling time-series signals such as PCG, electroencephalogram (EEG), and electrocardiogram (ECG) [39]. The WST decomposes a signal into different frequency components, capturing features at multiple scales or frequency bands [38]. However, it also includes an extra layer of non-linearity that allows for the extraction of more robust and distinctive features from the signal at each scale. However, its true strength lies in the analysis of complex time-series signals such as medical data, where it can reveal important insights about heart and brain activity, among other things.

The WST is composed of two main steps: the wavelet transform and the scattering transform. The wavelet transform is used to decompose the signal into different frequency bands, and the scattering transform is used to extract features from the decomposed signal [40]. In the wavelet transform step, signals are convolved with a series of band-pass filters, each of which is centered at a different frequency. The outcome of this phase is a group of wavelet coefficients, which portray the distinct frequency constituents of the signal. Mathematically, the wavelet scattering transform of the signal can be represented as,

$$W_I[f_1, f_2, \ldots, f_n]X = \left|\ldots\left|\left|X * \psi_{f_1}\right| * \psi_{f_2}\right| \ldots \psi_{f_n}\right| * \alpha_I \tag{1}$$

where, the wavelet scattering transform of n layers is represented by W_I, and the input signal is denoted by X. The symbol ψ_{f_n} refers to a dilated wavelet that is centered at the frequency f_n of a bandpass filter. There is also a translation invariant given by α_I, and the convolution operator is denoted by $*$. In the scattering transform step, the wavelet coefficients are then convolved with a low-pass filter, which is used to average the coefficients over a certain time window, as shown in Figure 2. This step is repeated for different scales and orientations, resulting in a set of scattering coefficients that represent the signal at different scales and orientations [41].

The effectiveness of the WST feature-based method in automatically diagnosing heart conditions is investigated in this section. The study focuses on extracting features from PCG signals to train an RNN model classifier capable of identifying various heart conditions. Specifically, the extraction of first and second-order scattering coefficients from both normal and abnormal PCG heart sounds is performed in this feature-based analysis. Wavelet scattering coefficients provide a comprehensive representation of signal characteristics,

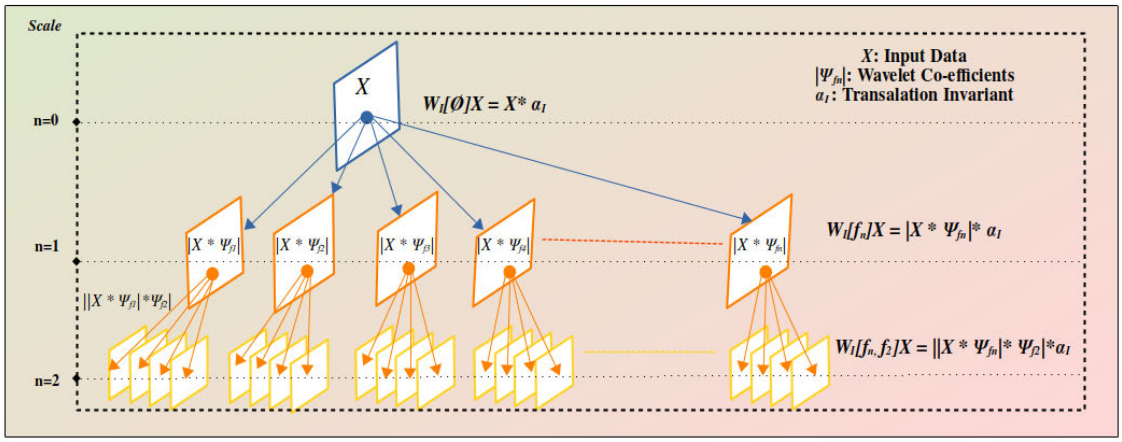


FIGURE 2. Signal decomposition using wavelet scattering network.

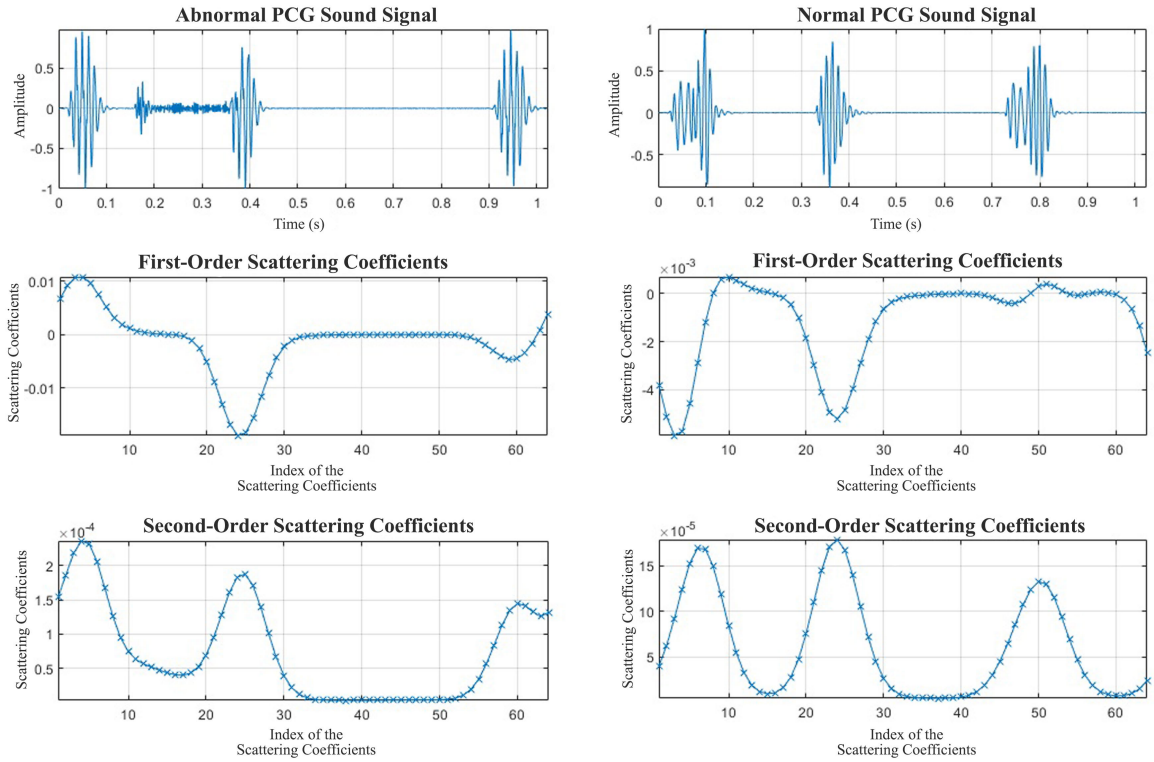


FIGURE 3. Time-domain representations of first-order and second-order filter bank scattering coefficients for abnormal and normal PCG signals.

with first-order coefficients computed through initial wavelet transforms, capturing fundamental frequency content across scales. Subsequently, second-order coefficients, obtained by applying additional wavelet transforms to the first-order coefficients, reveal intricate interactions and modulations among distinct frequency components. The analysis of these coefficients reveal significant noise distortion present in the abnormal PCG signals, particularly between 0.1 to 0.4 seconds. This distinctive distortion is visually illustrated in Figure 3. By utilizing first and second-order features, the study demonstrates their ability to discriminate abnormal heart conditions based on the observed noise distortion patterns in the PCG signals.

C. RECURRENT NEURAL NETWORKS (RNN)

Recurring neural networks are specifically appropriate for handling time series signals, like PCG, which are employed in the identification of heart irregularities. Unlike traditional feedforward neural networks, RNNs have loops in their architecture that allow them to maintain an internal memory of previous inputs [42]. This memory enables RNNs to learn patterns and dependencies in sequential PCG data,

making them a powerful tool for tasks such as heart sound segmentation, feature extraction, and classification [43]. By processing PCG signals using RNNs, it is possible to improve the accuracy and efficiency of cardiac abnormality diagnosis. Various RNN models were employed, including LSTM, Bi-LSTM, and stacked LSTM (LSTM+LSTM) models. These RNN models were selected due to their effectiveness in processing sequential data and learning temporal dependencies [44]. By integrating these RNN models into the analysis, the goal was to improve the accuracy of cardiac abnormality diagnosis using PCG signals.

1) LONG SHORT-TERM MEMORY (LSTM)

LSTM, a category of RNN, is an ideal fit for managing time-series information. It is specifically developed to overcome the issue of the vanishing gradient that may arise during the training of conventional RNNs. In traditional RNNs, the gradient signal can become very small as it is propagated back through time, leading to slow learning or even the inability to learn long-term dependencies [9], [45]. LSTMs address this problem by introducing a gating mechanism that allows them to selectively remember or forget information from previous time steps. In the architecture of an LSTM cell, information flow is regulated by three gates: the input gate, the forget gate, and the output gate [45]. The input gate manages incoming information, the forget gate governs outgoing information, and the output gate manages the information that remains within the cell. The activation functions of these gates are sigmoidal, enabling the LSTM to control them and decide which data to retain or exclude. The LSTM cell also has a cell state, which serves as the internal memory of the cell. The cell state can be updated or reset using the input gate and forget gate, allowing the LSTM to selectively remember or forget information from previous time steps [46].

2) BIDIRECTIONAL LONG SHORT-TERM MEMORY (Bi-LSTM)

In addition to LSTM, another commonly used RNN architecture for processing sequential data is the Bi-LSTM network. Like LSTM, Bi-LSTM includes memory cells and gates that allow the network to selectively remember or forget information over time [47]. However, Bi-LSTM also incorporates a second set of memory cells and gates that process the input sequence in reverse order, providing a complementary perspective on the temporal dependencies in the data [48]. In PCG signal analysis, Bi-LSTM can be used to capture both forward and backward dependencies in the signal, which may improve the accuracy of diagnosis for certain cardiac abnormalities [49]. The Bi-LSTM architecture operates by taking in a sequence of PCG signal samples and processing them in a bi-directional manner, i.e., both forward and backward. Once the computations are complete, the outputs of both the forward and backward layers are merged via concatenation to generate a final prediction.

3) STACKED LSTM

The stacked LSTM structure is formed by placing multiple layers of LSTM on top of each other. However, in stacked LSTM, there are typically two or more LSTM layers, allowing for the modeling of even more complex and higher-level representations of the sequence. The idea behind stacked LSTM is to use multiple layers of LSTM cells to learn multiple levels of abstraction in the sequence. Each layer in the stack receives the output from the preceding layer as input, with each layer focusing on learning a unique abstraction level. The input sequence is first processed by the initial layer of LSTM cells.

In a two-layer stacked LSTM network, the output of the first layer serves as the input to the second layer. The second layer can learn to build on the representations learned by the first layer, allowing for the modeling of even more complex sequences. At each time step in each LSTM layer, several operations are performed on the input and the previous state to produce a new hidden state and a new cell state. These operations include computing the forget gate, input gate, output gate, and cell state values using a set of learnable weights, recurrent weights, and biases.

In order to perform a mathematical examination of a stacked LSTM model with two layers, it is essential to compute the gradients of the loss function with respect to the weights and biases of each individual layer. These gradients can be determined by utilizing the backpropagation through the time algorithm, which entails repeatedly applying the chain rule of differentiation. The final sequence prediction is generated by the output of the last layer of LSTM cells. It is worth noting that the input to each layer of the model is derived from the output of the preceding layer, rather than from the initial input sequence, as depicted in Figure 4(a).

In stacked LSTM, each layer can have a different number of hidden units. Typically, the number of hidden units in each layer is gradually reduced as the stack moves up, allowing for the network to capture high-level features in the sequence. One advantage of stacked LSTM is that it allows for the modeling of very complex sequences, as each layer in the stack is able to learn different features of the sequence. Stacked LSTM has been shown to perform well on a wide range of sequence modeling and prediction tasks, including natural language processing, speech recognition, and time-series analysis.

D. CUSTOM SCALOGRAM LAYERED CRNN (CS-CRNN)

The CS-CRNN is an innovative approach that combines the advantages of convolutional and recurrent neural networks. Conventional approaches often require the manual engineering of features, which can be time-consuming. By integrating the scalogram computation within the network itself, the CS-CRNN eliminates the need for hand-crafted features, allowing for a more automated and data-driven approach to feature extraction.

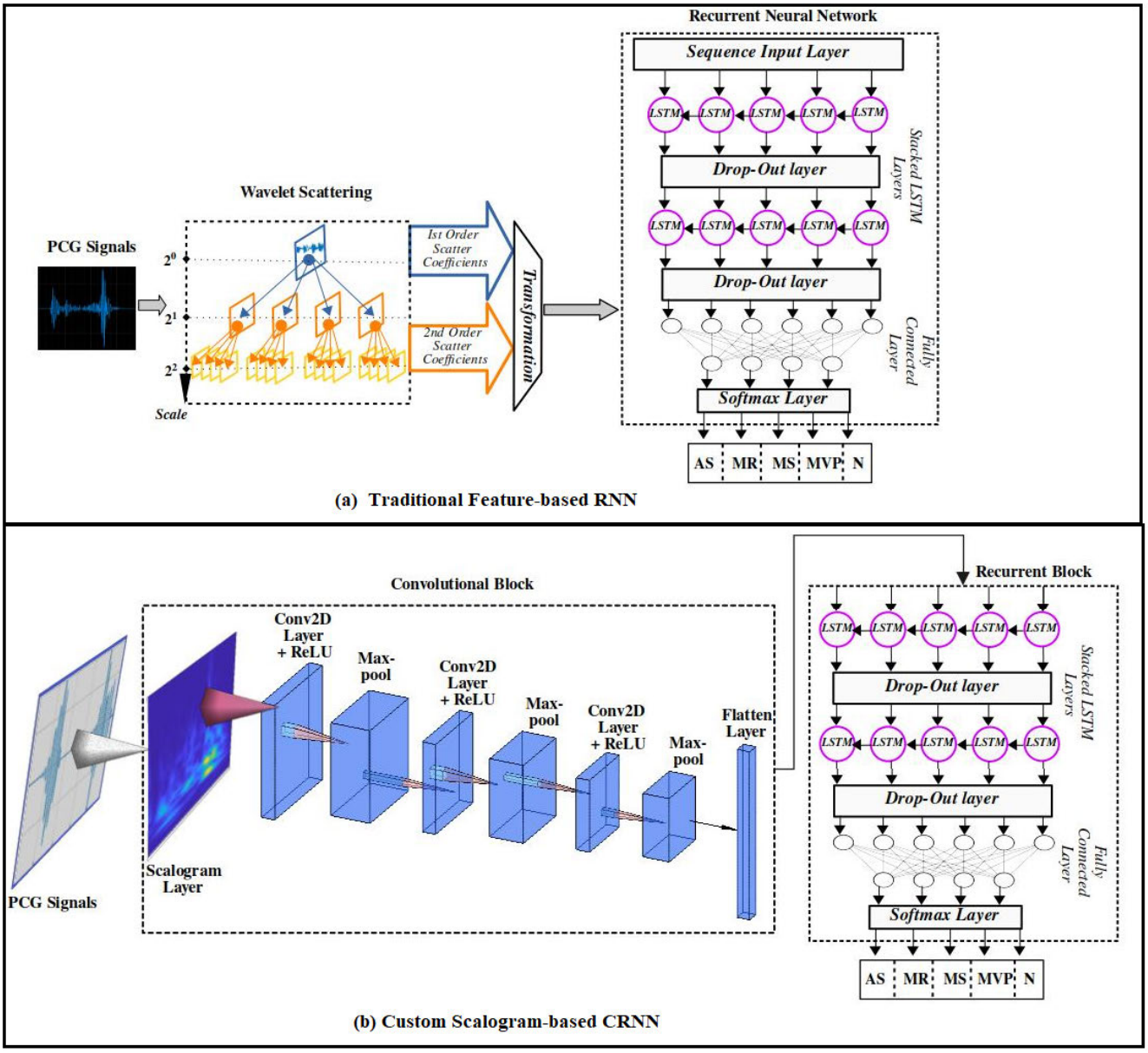


FIGURE 4. Architectural differences of feature-based RNN and CS-CRNN in heart sound classification.

1) CUSTOMIZATION OF THE SCALOGRAM LAYER

The scalogram computation involves convolving the input signal with the wavelet function at various scales and analyzing the resulting magnitude or power values. This process provides a time-frequency representation of the input signal, highlighting the presence of different frequency components at different time intervals. The scalogram layer, being a non-learnable transformation, does not possess any learnable parameters. However, it is important to note that in the CRNN architecture, as illustrated in Figure 4(b), the subsequent convolutional and recurrent layers do include learnable parameters. These parameters are optimized during the training process, as highlighted in Table 1. By carefully choosing suitable wavelet functions, adjusting scale resolutions, and optimizing other parameters like the number of scales, overlap ratio, and normalization techniques, the CS-CRNN model can effectively adapt to the specific characteristics of the time-frequency representations of PCG data. Figure 5 depicts the extracted scalograms at the first layer of CS-CRNN for all five classes of PCG signals.

2) CRNN

The CRNN architecture is a powerful model that combines the strengths of both convolutional and recurrent neural networks, making it particularly suitable for sequential data analysis, such as in the case of PCG signals. The scalogram, generated by the customized scalogram (CS) layer, provides a time-frequency representation of the PCG

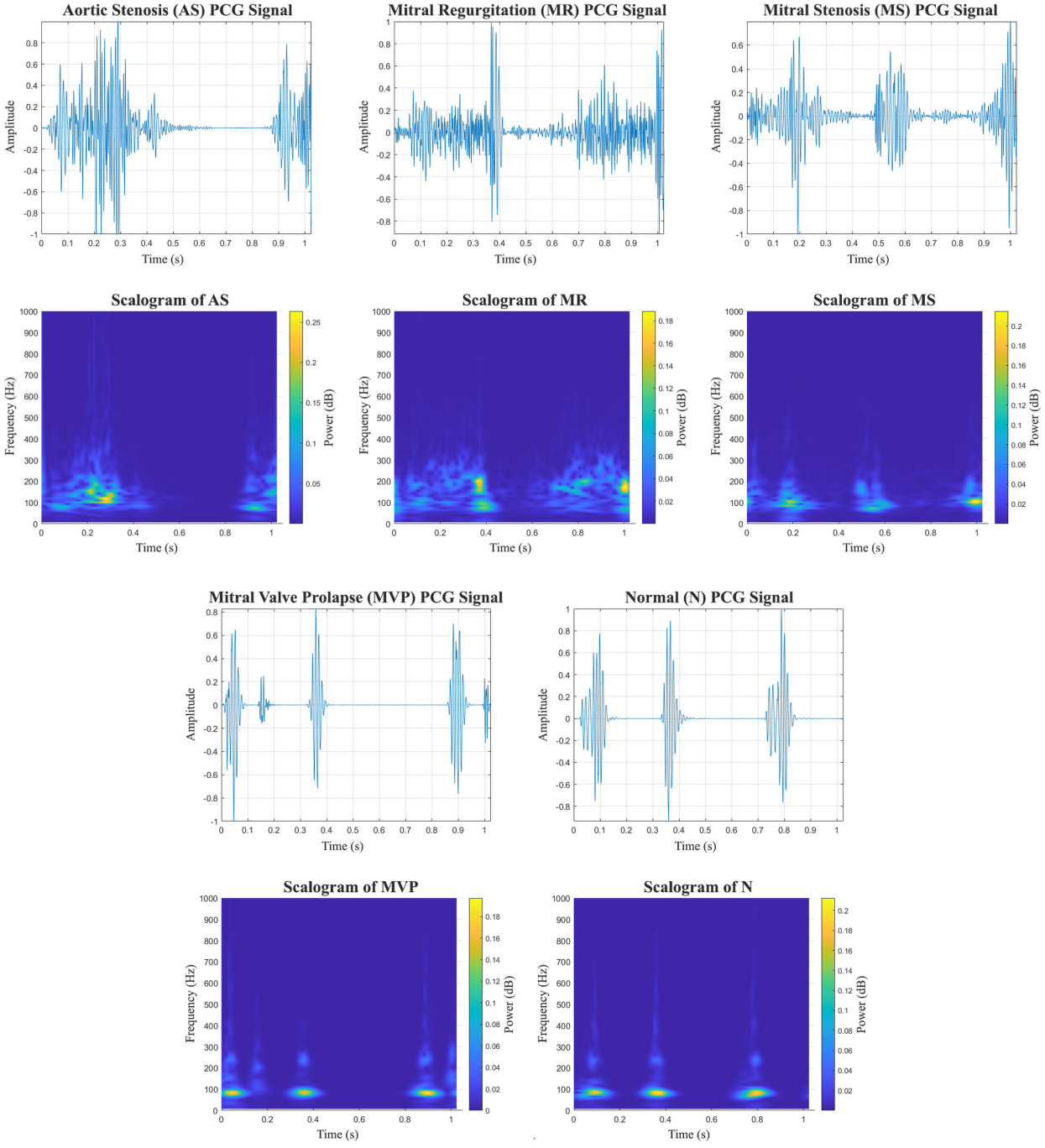


FIGURE 5. Five categories of PCG Signals and their corresponding scalograms from the first layer of CRNN.

signal. By applying convolutional filters, the convolutional layers are able to capture relevant spatial patterns in the scalogram, enabling the model to identify discriminative features related to cardiac abnormalities. On the other hand, the recurrent layers in the CRNN architecture are responsible for capturing temporal dependencies and long-term patterns in the PCG signals. In the context of cardiac abnormality detection, it is important to consider not only local features but also the temporal dynamics of the signals. The recurrent layers, typically implemented using LSTM units, are capable of modeling the sequential nature of the PCG signals and learning meaningful representations of long-term dependencies. This enables the CRNN model to effectively discriminate between normal and abnormal cardiac patterns, as it can identify patterns that evolve over time and are indicative of specific cardiac conditions.

As described in Table 1, the CS-CRNN architecture consists of multiple layers, including the CS layer, convolutional

TABLE 1. Description of layers and learnable parameters in the CS-CRNN Model.

Name	Layers	Learnable Parameters
Custom-Scalogram Layer	Image input Scalogram	-
Convolution Layer (× 3)	Conv2D layer (5×5) with stride [1 1] Batch Norm ReLU Max-Pool layer (3×3) with stride [2 2] Flatten Layer	Weights & Bias Offset & Scale - - -
Recurrent Layer (× 2)	LSTM with 512 hidden states Dropout of 20%	Input weights, Recurrent weights, & Bias -
Fully Connected (FC) Layer	5 FC layer	Weights & Bias
Soft-Max Layer	-	-
Classification Layer	-	Cross-entropy with output labels

layers, recurrent layers, fully connected (FC) layers, and the final softmax and classification layers. Each layer has its own set of learnable parameters, such as weights, biases, and dropout rates. These parameters are optimized during the training process to minimize the classification loss and improve the model's ability to detect cardiac abnormalities.

IV. BAYESIAN OPTIMIZATION

Bayesian optimization is a powerful technique used in deep learning approaches to optimize complex functions that are expensive to evaluate, such as hyperparameter optimization.The objective of this article is to identify the ideal combination of hyperparameters that can deliver optimal outcomes for detecting cardiac irregularities using PCG signals. The basic idea of Bayesian optimization is to use a probabilistic model, such as a Gaussian process (GP), to estimate the performance of a model as a function of its hyperparameters. The model is trained on a small subset of the hyperparameter space and then used to predict the performance of the model at unexplored points in the hyperparameter space. After making predictions, an acquisition function is employed to determine the next point for evaluation. This function balances exploration, which involves testing new hyperparameters, and exploitation, which involves trying out the hyperparameters that are deemed most likely to produce satisfactory results. The general flow of Bayesian optimization involves several steps:

- *Define the objective function:* The objective function is defined as the function to be optimized, which is usually a metric of interest for a given model and hyperparameter set. In the proposed study, the objective function is the accuracy of the RNN model in classifying PCG signals.
- *Estimate the objective function:* The objective function is estimated by training the model on a small subset of the hyperparameter space, which generates a set of data points that can be used to train a probabilistic model. In Bayesian optimization, a GP is typically used as a probabilistic model to estimate the objective function.
- *Evaluate the model at unexplored points:* The probabilistic model is used to predict the performance of the model at unexplored points in the hyperparameter space. The model is evaluated at the point that the acquisition function suggest optimizing exploration and exploitation trade-offs.
- *Convergence check:* The model's performance is evaluated at multiple points, and a convergence check is done to determine whether the model has converged or not.
- *Update hyperparameters:* If the model has not converged, the hyperparameters are updated and the process continues.
- *Select optimal model:* The set of hyperparameters that produces the highest score on the objective function is chosen after the network has converged, providing the optimum model for the current classification task.

Additionally, Bayesian optimization is employed to discover the optimal set of hyperparameters for a specific machine-learning task. It is efficient and effective in exploring a large hyperparameter space without having to test all possible options. By using a probabilistic model to predict the performance of a model based on its hyperparameters, it saves time and resources.

V. EXPERIMENTAL SETUP

The article conducted two experiments to assess the effectiveness of a proposed method. In the first experiment, a traditional approach was used, where wavelet scattering features were extracted from the signals. These features were based on first-order and second-order wavelet coefficients. RNN models were then trained using these extracted features. Multiple RNN models were trained, and the best-performing model was selected for further analysis. In the second experiment, a more advanced technique was employed, involving the development of a custom model known as the scalogram layered CRNN model. This model eliminated the need for a separate feature extraction step. Instead, it directly processed the original PCG signals, which were transformed into scalogram images in the first layer of CRNN. Scalogram images represent the frequency content of the signals over time. They provide a visual representation of how the signal's frequency components change throughout the duration. The CRNN model utilized the strengths of convolutional layers to extract spatial features from the scalogram images. This allows the model to capture important patterns and structures in the time-frequency domain. Additionally, recurrent layers in the model were incorporated to capture temporal dependencies, enabling the network to recognize sequential patterns and dynamics present in the data.

The aforementioned experiments were further conducted to diagnose cardiac abnormalities through two distinct classification tasks: binary classification and multiclass classification. In the binary classification task, the objective was to accurately classify heart sounds into two categories, typically representing normal and abnormal cardiac conditions. Similarly, the multiclass classification task aimed to

TABLE 2. Train and test split in binary-class and multiclass data without (W/o) and with augmentation.

Augmentation (Total Samples)	Data Split	Binaryclass-Data		Multiclass-Data				
		Abnormal	Normal	AS	MR	MS	MVP	N
(W/o)	Train	640	160	160	160	160	160	160
(1000)	Test	160	40	40	40	40	40	40
With	Train	3200	800	800	800	800	800	800
(5000)	Test	800	200	200	200	200	200	200

AS: Aortic Stenosis, MR: Mitral Regurgitation, MS: Mitral Stenosis, MVP: Mitral Valve Prolapse, N: Normal

classify heart sounds into five categories, each representing different types of cardiac abnormalities or conditions. The dataset was partitioned into training and testing sets at an 80:20 ratio for both experiments, and the partitioning details are presented in Table 2. The dataset for the binary classification task was imbalanced. The training set consisted of 640 abnormal heart sounds and 160 normal heart sounds, whereas the testing set contained 160 abnormal heart sounds and 40 normal heart sounds. Various feature-based RNN models, including LSTM, Bi-LSTM, stacked LSTM, and raw waveform-based CS-CRNN models, were trained and evaluated using this dataset.

In the multiclass classification task, the dataset was well-balanced, with equal numbers of heart sounds in each of the five categories, including AS, MR, MS, MVP, and N. Moreover, Bayesian optimization was applied to optimize the hyperparameters of the RNN models to achieve better performance. Accuracy was the only metric used to evaluate the multiclass classification models since the dataset was well-balanced. However, the experiments aimed to develop accurate heart sound classification models using different types of RNN and CRNN models, coupled with Bayesian optimization, for both binary and multiclass classification tasks. The utilization of innovative CS-CRNN models and optimization techniques represents a cutting-edge approach to the diagnosis of cardiac abnormalities.

VI. RESULTS AND DISCUSSIONS

The diagnosis of cardiac abnormalities using PCG signals involved two experiments, which included binary class and multiclass classification. The pre-existing datasets that were used contained five classes of heart sound signals, and wavelet scattering was employed to extract features. Subsequently, multi-scale RNN and advanced CS-CRNN models were trained and optimized using Bayesian optimization, showcasing their potential in the diagnosis of cardiac conditions. The resulting RNN and CS-CRNN models were found to be highly effective in accurately diagnosing PCG signals and predicting the presence of specific cardiac abnormalities. The forthcoming sections provide a detailed description of the proposed methodology employed for PCG signal analysis, which encompassed feature-based RNN and raw waveform-based CRNN models. These models were rigorously evaluated through binary and multiclass classification experiments, offering valuable insights into the utilization of AI in medical diagnosis. It is imperative to underscore the significance of conducting thorough testing and validation before deploying these models in real-world clinical settings, ensuring their reliability and suitability for practical applications.

A. TRADITIONAL FEATURE EXTRACTION

The WST is a vital phase in the process of feature extraction in the proposed method for detecting cardiac abnormalities from PCG signals. This approach employs the Morlet wavelet, which has a depth of 2 and is structured with two layers. The first layer consists of eight Morlet wavelets per octave, while the second layer includes one Morlet wavelet per octave, represented as quality factors $Q = [8\ 1]$. The utilization of a low-pass scaling function by wavelets results in robust representations of PCG signals. This, in turn, guarantees consistency and accuracy in the analysis.

After standardizing the wavelet coefficients, they are subjected to a logarithmic function to make the data more suitable for additional processing. This results in a scattering network, which has 411 paths and 25 scattering time windows for each PCG signal. The use of wavelet scattering transforms extends the mel filter bank representation while preserving the valuable information contained in the PCG signals. This reliable feature extraction process enhances the accuracy and robustness of the method in diagnosing cardiac abnormalities.

B. RAW WAVEFORM-BASED CUSTOMIZED SCALOGRAM LAYER

The raw waveform-based CS layer is a novel approach used in deep learning models for signal analysis, particularly in the field of medical diagnosis. This layer is specifically designed to extract meaningful features from raw waveform data, such as ECG and PCG signals, which are crucial in detecting and classifying cardiac abnormalities. Unlike traditional methods that rely on handcrafted features or pre-defined transformations, the CS layer offers a data-driven approach by learning and adapting to the unique characteristics of the input signals. It leverages the concept of the scalogram, which provides a time-frequency representation of the signal, allowing for better capturing of both temporal and spectral information. The CS layer is customized to suit the specific requirements of the raw waveform data in cardiac signal analysis. It integrates techniques such as time-frequency analysis, wavelet transforms, or scalograms to transform the raw waveform into a multi-scale representation. By incorporating multiple scales, the CS layer enables the

TABLE 3. Comparison of hyper-parameters before and after Bayesian optimization for proposed models.

Hyper-parameters	Before Optimization		After Optimization	
	RNN	CS-CRNN	RNN	CS-CRNN
Learning rate	1e-4	1e-3	[1e-5, 1e-1]	[1e-5, 1e-1]
Dropout rate	0.2	0.2	[0.1, 0.5]	[0.1, 0.5]
Activation function	ReLU	ReLU	ReLU	ReLU
Regularization strength	1e-4	5e-4	[1e-6, 1e-3]	[1e-6, 1e-3]
Batch size	30	50	30	50
Number of epochs	300	300	300	300
Initialization method	NA	NA	GP	GP
Optimizer type	Adam	Adam	Adam	Adam
Gradient decay rate	NA	NA	0.90	0.90
Best observed objective function	NA	NA	0.015	0.021
Best estimated objective function	NA	NA	0.012	0.018

model to capture both local and global patterns, providing a more comprehensive understanding of the underlying signal dynamics.

At each scale, s, the CS layer performs a convolution operation between the input signal $x(t)$ and a wavelet filter $\psi_s(t)$ as represented by the convolution integral:

$$CS_s(t) = \int x(u) * \psi_s(t-u)du \tag{2}$$

Here, $CS_s(t)$ represents the output at scale s and time index t. The convolution operation captures the local features of the input signal at the given scale.

To obtain a multi-scale representation, the CS layer repeats this convolution operation for a range of scales s, typically spanning multiple octaves. Each scale captures different frequency bands and provides a different level of time-frequency resolution. The resulting multi-scale representation $CS_s(t)$ can be further processed by subsequent layers in the deep learning model for the classification of cardiac abnormalities. Since the CS layer does not have any learnable parameters, its purpose is to extract useful features from the raw waveform signal without altering the underlying characteristics of the signal. Its ability to capture both temporal and spectral information at multiple scales empowers deep learning models to make accurate diagnoses and predictions, paving the way for advancements in medical diagnosis and patient care. It has also demonstrated improved training speed, robustness, and generalization capabilities compared to traditional feature-based approaches.

C. NEED OF DATA AUGMENTATION

The necessity for data augmentation in the proposed study stemmed from the inherent limitations of the original dataset. Initially, the dataset consisted of a relatively small number of heart sound signals, potentially restricting the CS-CRNN model's ability to comprehend the diverse patterns in the PCG signals. However, this may lead to limited generalization of the model to unseen variations in heart sound signals, crucial for real-world applications. Therefore, to assess the generalization ability of the complex CS-CRNN model, audio augmentation was employed, resulting in an expansion of the dataset to five times its original size. By introducing variations through time shift, noise, and pitch shift techniques, a diverse dataset of 5000 signals was generated. Time shift allowed the simulation of different cardiac cycles, noise injection replicated real-world interference, and pitch shift emulated variations in heart sound frequencies. Within the augmentation process, pitch shift was randomly applied to the signals throughout, with a probability of 0.5, spanning from 0 semitone to 1 semitone. Time shift, ranging between -0.3 seconds and 0.3 seconds, was applied to the signals with a probability of 1. Furthermore, the signals underwent noise injection with a probability of 1, resulting in signal-to-noise ratio (SNR) values varying from -10 dB to 10 dB. This augmentation process led to the creation of an additional 4000 PCG signals, randomly derived from the original signals as shown in Table 2. It is worth mentioning that the experiment was conducted using MATLAB R2022b.

D. TRAINING NETWORK CONFIGURATION

The process of diagnosing heart anomalies using PCG signals encompasses several crucial steps. This section outlines the key steps involved in this diagnostic process, including the utilization of PCG datasets, preprocessing of stationary signals through segmentation, analysis of signals by extracting features using wavelet scattering for RNN modeling, and raw waveform modeling for CS-CRNN. Additionally, Bayesian optimization is employed to optimize the training parameters for enhanced performance.

When it comes to selecting the RNN architecture, several methods can be used including LSTM, Bi-LSTM, and stacked LSTM. Each of these approaches has its advantages and disadvantages, and the decision of which to choose will depend on the requirements and the nature of the PCG data being analyzed. For instance, LSTMs are recognized for their capability to grasp long-term dependencies and thus suit tasks involving sequential data. The stacked LSTM variations allow for deeper network configurations, which provide improved performance for complex tasks. Bi-LSTMs, meanwhile, make use of both forward and backward sequences, making them appropriate for capturing both past and future context.

TABLE 4. Confusion matrices of multi-scale RNN and CS-CRNN models in the classification of abnormal (Ab) and normal (N) PCG heart sound signals.

Output Class		Multi-Scale RNN (WST Features)						CS-CRNN (Raw Waveform)					
		LSTM		Bi-LSTM		Stacked LSTM		W/o Augmentation		With Augmentation (Imbalanced Test Data)		With Augmentation (Balanced Test Data)	
		Ab	N	Ab	N	Ab	N	Ab	N	Ab	N	Ab	N
	Ab	159	3	156	1	159	0	155	2	800	4	199	4
	N	1	37	4	39	1	40	5	38	0	196	1	196
		Target Class											

TABLE 5. Comparative evaluation results of multi-scale RNN and CS-CRNN models in the classification of abnormal and normal (binary-class) PCG heart sound signals.

Front-End	Model	Acc (%)	Sens (%)	Spec (%)	Pr (%)	F1 (%)
Feature-Based	LSTM	98	98.1	97.4	99.3	98.6
	Bi-LSTM	97.5	99.4	90.7	97.5	98.3
	Stacked LSTM	99.5	100	97.6	99.3	99.6
Raw Waveform-Based	CS-CRNN (W/o Augmentation)	96.5	98.7	95	96.8	97.7
	CS-CRNN (With Augmentation & Imbalanced Test Data)	99.6	99.5	100	100	99.7
	CS-CRNN (With Augmentation & Balanced Test Data)	98.8	98.0	99.5	99.5	98.7

The neural networks were trained using the Adam optimizer with carefully chosen hyperparameters. The initial learning rate was set to 0.0001, the maximum number of epochs was 300, and the mini-batch size was 50. To ensure that the network learned from the full sequence of data, the sequence length was set to 'shortest', and the training data was shuffled every epoch. In order to minimize computational overhead, the verbose option was turned off. The training was conducted on a GPU with the number of hidden units set to 512. The selection of hyperparameters, both before and after optimization, aimed to strike a favorable balance between training time and performance on the PCG sound signals utilized in this study, as outlined in Table 3.

E. PERFORMANCE ANALYSIS OF THE MODELS IN BINARY CLASSIFICATION

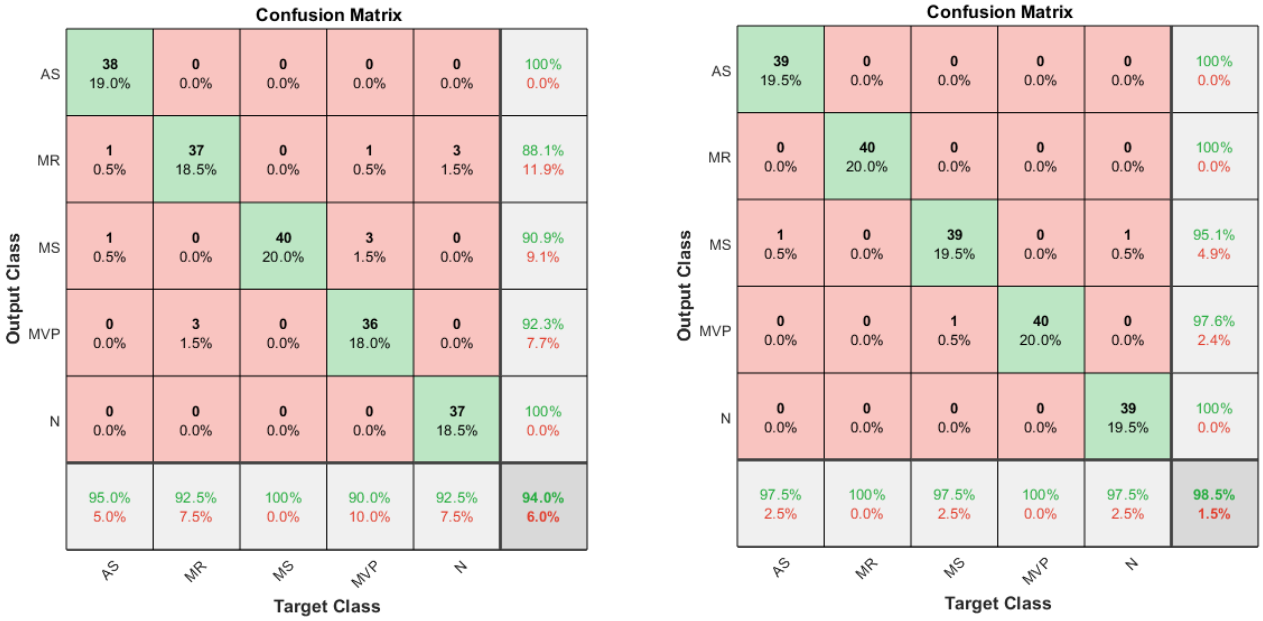
1) RNN MODELS

In this section, we evaluate the performance of several RNNs using a feature-based front-end approach for diagnosing cardiac abnormalities using PCG signals, including LSTM, Bi-LSTM, and stacked LSTMs. As previously mentioned, we used an imbalanced dataset with 160 abnormal/unhealthy signals and 40 normal/healthy PCG sound signals to test the binary classifier model. During the testing phase, the LSTM model achieved a high level of performance, achieving a 98% accuracy (Acc), 98.1% sensitivity (Sens), 97.4% specificity (Spec), 99.3% precision (Pr) and an F1-score (F1) of 98.6%. However, there were a few instances where normal sounds were incorrectly classified as abnormal, and vice versa.

Subsequently, the Bi-LSTM model attained an Acc of 97.5%, Sens of 99.4%, Spec of 90.7% and an F1 of 98.3%. However, it is worth noting that it misclassified 4 abnormal sounds as normal, as demonstrated in Table 4. Finally, stacked RNN models were implemented, which resulted in a significant improvement in accuracy. On comparison, the stacked LSTM model demonstrated the highest Acc and F1, reaching 99.5% and 99.6%, respectively. Additionally, its Sens and Spec were recorded at 100% and 97.6%, respectively, as outlined in Table 5. Moreover, by using stacked RNN models, the network can learn hierarchical representations of the input data. Each layer can learn to focus on different aspects of the input, with the lower layers learning low-level features and the higher layers learning more abstract representations. This hierarchical approach to learning helps the model identify and extract the most relevant features in the input data, leading to better accuracy in classification tasks.

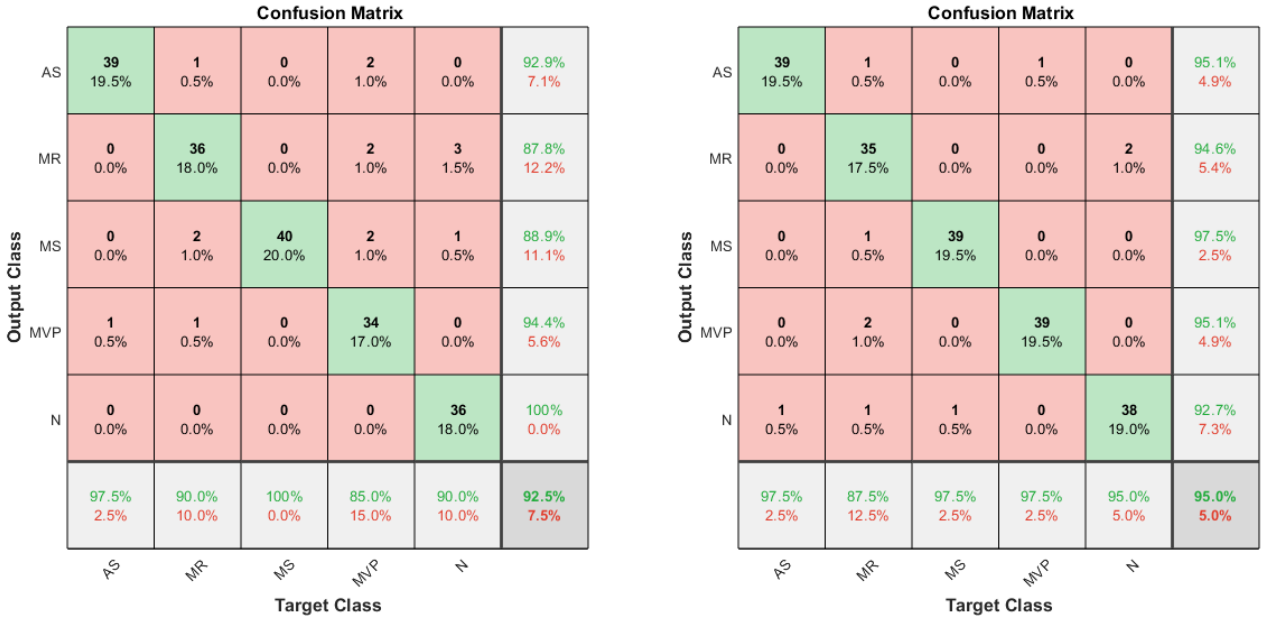
2) CS-CRNN MODEL

The raw waveform-based CS-CRNN was evaluated on the test data of binary classification, revealing that 5 classes of abnormal sounds were misclassified as normal sounds, while 2 normal sounds were misclassified as abnormal, as shown in the confusion matrix of Table 4. As a result, the model achieved an Acc of 96.5%, Sens of 98.7%, Spec of 95% and F1 score of 97.7%. The slightly lower accuracy of the raw waveform-based CS-CRNN compared to feature-based models may be attributed to a few factors. The engineered features can provide explicit information about the underlying abnormalities, making it easier for the model to distinguish between normal and abnormal sounds. In contrast, the raw waveform-based approach relies solely on the model to learn and extract



(a) Before optimization (b) After optimization

FIGURE 6. Confusion matrix of LSTM model without augmentation.



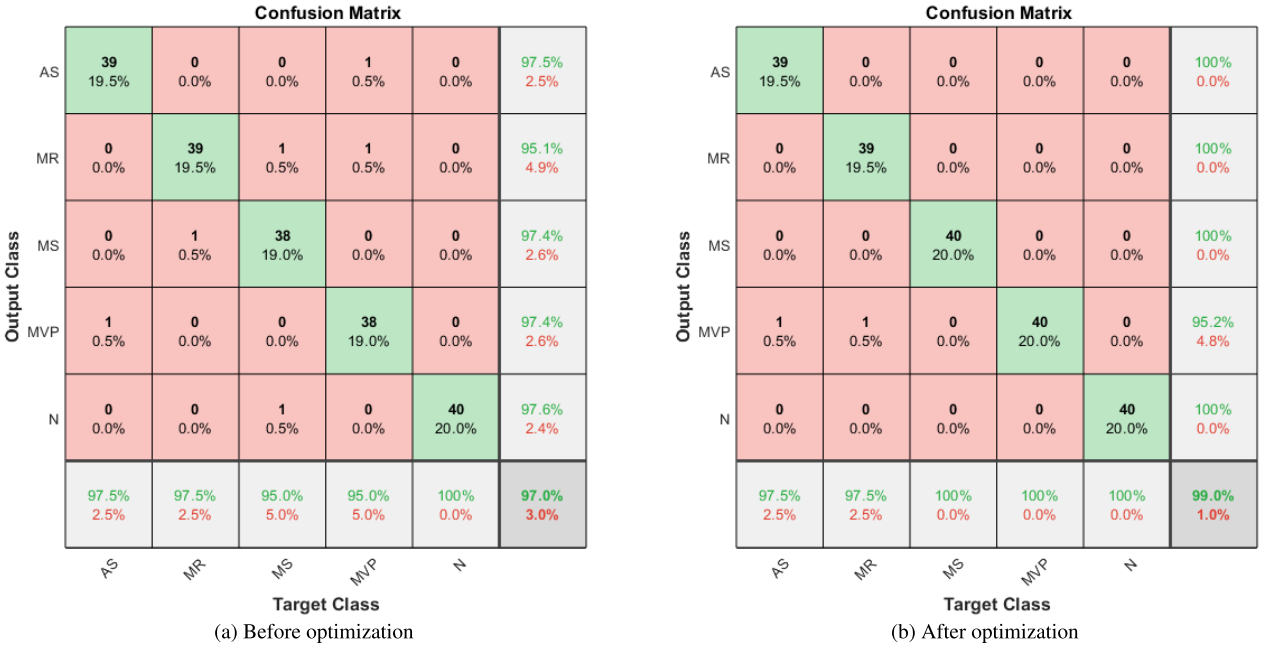
(a) Before optimization (b) After optimization

FIGURE 7. Confusion matrix of Bi-LSTM model without augmentation.

relevant features from the raw data, which can be more challenging and may result in slightly more misclassifications. However, the raw waveform-based CS-CRNN demonstrates the highest accuracy following augmentation. To assess potential biases in the model's performance across different cases, the training dataset's imbalance was addressed by balancing the test dataset. Within the augmented binary-class dataset, a random selection of 200 abnormal samples (50 samples from each of the four classes) from the test data was made. These balanced samples were merged with the test data representing the normal class. Subsequently, the performance of this balanced dataset on the pre-trained CS-CRNN model was evaluated, resulting in an Acc of 98.8%. In the subsequent section, the evaluation of the CS-CRNN model involves the use of multiclass datasets.

TABLE 6. Evaluation results of multi-scale RNN models in the classification of multiclass PCG heart sound signals with and without Bayesian optimization.

Front-End	Model	Acc (%) without optimization	Acc (%) with optimization	Relative Improvement (RI)%
Feature-Based	LSTM	94	98.5	4.78
	Bi-LSTM	92.5	95	2.7
	Stacked LSTM	97	99	2.06
Raw Waveform-Based	CS-CRNN (W/o Augmentation)	94.5	98.5	4.23
	CS-CRNN (with Augmentation)	98.6	99.7	1.11



(a) Before optimization

(b) After optimization

FIGURE 8. Confusion matrix of stacked LSTM model without augmentation.

F. PERFORMANCE ANALYSIS OF THE MODELS IN MULTICLASS CLASSIFICATION

1) RNN MODELS

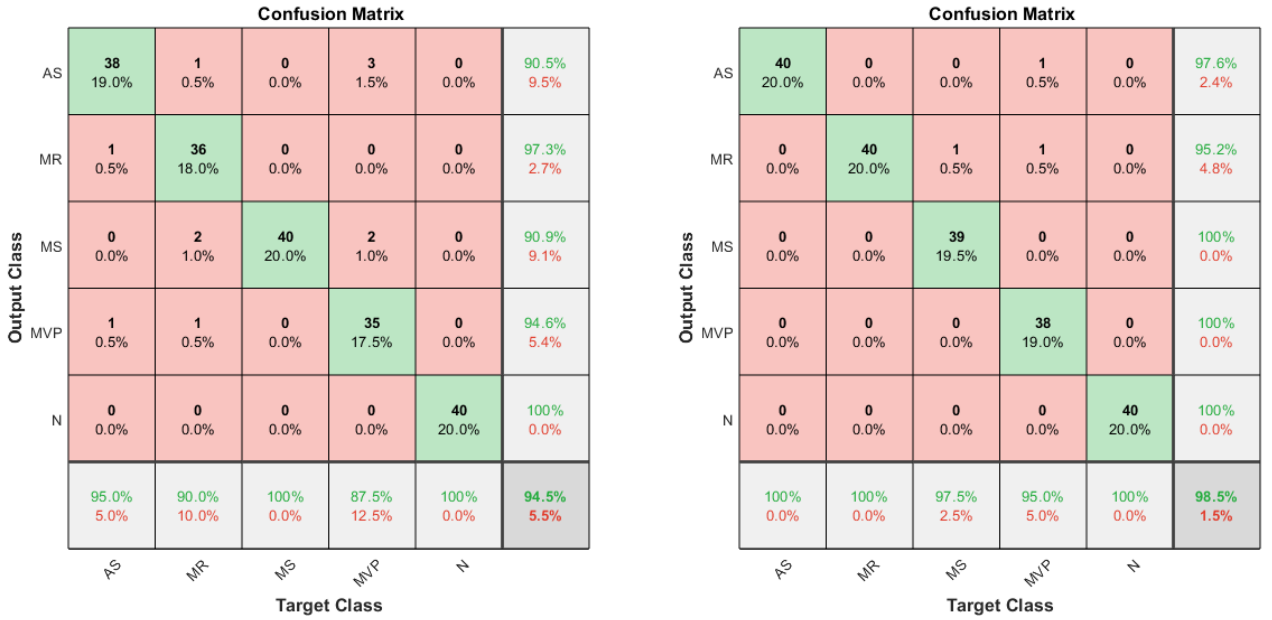
The goal of this experiment was to evaluate the performance of various RNN models in the multiclass classification of PCG heart sound signals. The datasets were perfectly balanced and, hence, accuracy was considered the performance metric. The models were trained and tested without the use of Bayesian Optimization. Table 6 summarizes the results obtained from this experiment, where the accuracy of each model is reported.

In the multiclass classification, the LSTM model achieved an Acc of 94%, with 6% misclassified instances. The misclassification is mainly due to the confusion between MR and MVP sound signals, and normal instances are misclassified as MR, as shown in Figure 6(a). On the other hand, the Bi-LSTM model achieved an Acc of 92.5% with most misclassifications occurring in the MR, MVP, and normal classes as shown in Figure 7(a). However, stacked RNN models performed better in multiclass classification, similar to binary classification. The stacked LSTM model also achieved a similar Acc of 97%, with misclassification mostly due to one instance in each class being misclassified as another class, except for normal sound signals as demonstrated in the confusion matrix of Figure 8(a).

Furthermore, in Section VI-G, the multiclass experiments were enhanced by incorporating Bayesian optimization, aiming to improve the model's performance specifically in cases of misclassification.

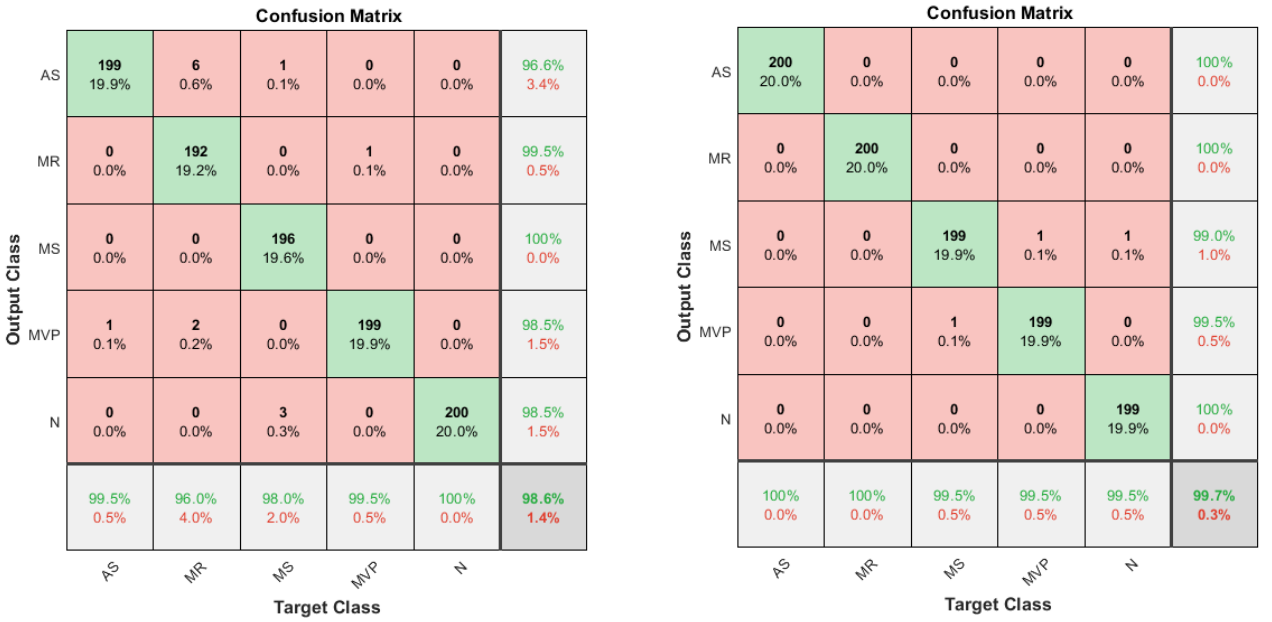
2) CS-CRNN MODEL

During the performance analysis of the models for multiclass classification using CS-CRNN, certain misclassifications were observed, particularly in the MR and MVP classes, where they were mistakenly classified as MS. Additionally, two classes from AS were also misclassified as MR and MVP, resulting in an overall Acc of 94.5% for the raw waveform-based CS-CRNN model, as shown in Figure 9(a). However, after augmentation, the CS-CRNN model achieved an Acc of 98.6%, reflecting a significant RI of 4.33%. The confusion matrix for augmented data is presented in Figure 10(a). These misclassifications can be attributed to various factors,



(a) Before optimization (b) After optimization

FIGURE 9. Confusion matrix of CS-CRNN model without augmentation.



(a) Before optimization (b) After optimization

FIGURE 10. Confusion matrix of CS-CRNN model with augmentation.

including the variability in signal characteristics within classes, overlapping features between different classes, the complexity and variability of cardiac abnormalities, potential limitations of the model architecture, and the adequacy of training data. To address these challenges, Bayesian optimization was employed to properly tune the hyperparameters, resulting in improved Acc. The use of Bayesian optimization ensures that the CS-CRNN model achieves comparable accuracy to the feature-based RNN models. Therefore, by refining the model, incorporating data augmentation techniques, and exploring advanced signal processing methods, the accuracy of multiclass classification using raw waveform data can be further improved. Moreover, the computational complexity of the CS-CRNN method, in comparison to traditional feature-based RNN models, highlights an intriguing observation. In the CS-CRNN

TABLE 7. Comparative analysis of existing models with a proposed model in the diagnosis of cardiac abnormalities using PCG sound signal dataset.

Author/ Reference/ Year	Datasets	Front-End Apporach	Model	Performance
Yaseen et al. [31] 2018	1000 samples (200 samples in each class of AS, MR, MS, MVP, N)	MFCC, DWT	SVM, DNN, Centroid KNN	The SVM model achieved a highest accuracy of 97.9% when both MFCC and DWT features were fused.
Ghosh et al. [17] 2019	Only 800 samples were considered (200 samples in each class of AS, MR, MS, N)	Wavelet Synchrosqueezing transform (WSST)	Random Forest (RF)	The RF classifier was able to classify 4 classes with a 95.12% accuracy
Ghosh et al. [51] 2020	1000 samples (200 samples in each class of AS, MR, MS, MVP, N)	Chirplet transform (CT) (local energy and local entropy) features	Multiclass composite classifier (MCC)	By utilizing CT features, the MCC model attained a 98.33% overall accuracy.
Alkhodari et al. [52] 2021	1000 samples (200 samples in each class of AS, MR, MS, MVP, N)	Maximal overlap discrete wavelet transform (MODWT) & z-score normalization	CNN, Bi-LSTM	Using a CNN-Bi-LSTM network resulted in a maximum accuracy of 99.30%.
Oh et al. [53] 2020	1000 samples (200 samples in each class of AS, MR, MS, MVP, N)	–	WaveNet (6 residual blocks with gated activation)	The Wavenet model attained a 97% accuracy through 10-fold cross validation.
Xiao et al. [54] 2020	3153 samples (PhysioNet/CinC 2016) (2488 normal and 665 abnormal samples)	–	1D-CNN with attention mechanism	Separable convolutions model attained a 93% accuracy with 0.19M parameters.
Yazan et al. [55] 2022	1000 samples (200 samples in each class of AS, MR, MS, MVP, N)	–	CNN+LSTM	Audio augmentation: 99.8% (time) and 99.7% (frequency). No augmentation: 98.4% (time) and 95.4% (frequency).
Ismail et al. [34] 2023	1000 samples (200 samples in each class of AS, MR, MS, MVP, N)	Chirplet Z transform (CZT) and spectrograms	Rounded-base transfer learning	Achieved a highest accuracy of 98%.
Zang et al. [56] 2023	1000 samples (200 samples in each class of AS, MR, MS, MVP, N)	–	AmtNet	Achieved a highest accuracy of 100% using both attention and temporal pyramid pool.
Wang et al. [57] 2023	1000 samples (200 samples in each class of AS, MR, MS, MVP, N)	2^{nd} order spectral analysis	PCTMF-Net	Achieved a highest accuracy of 99.4%.
Shuvo et al. [58] 2023	1000 samples (200 samples in each class of AS, MR, MS, MVP, N)	CQT, CWT, STFT, MFCC	NRC-Net	Achieved a highest accuracy of 99.7% using CWT.
Proposed Model	Binary-class (Ab & N) Multiclass (AS, MR, MS, MVP, N)	Wavelet scattering transform	Stacked LSTM	Highest accuracy of 99.5% & F1-score of 99.6%. Highest accuracy of 99% in multiclass after optimization.
	Binary-class (Ab & N) Multiclass (AS, MR, MS, MVP, N)	Raw waveform-based	CS-CRNN (with augmentation)	Highest accuracy of 99.6% & F1-score of 99.5%. Highest accuracy of 99.7% in multiclass after optimization.

approach, a non-learnable feature learner is integrated as the initial scalogram layer, directly processing raw data. This design choice eliminates the necessity for the model to adapt and learn features during training, resulting in a reduction in computational complexity. On the other hand, traditional RNN models, require the learning and adaptation of features, such as first and second-order wavelet coefficients during training, potentially leading to increased computational demands.

G. PERFORMANCE ANALYSIS OF THE MODELS AFTER BAYESIAN OPTIMIZATION IN MULTICLASS CLASSIFICATION

After analyzing the performance of RNN models in multiclass classification, Bayesian optimization was used to improve their performance. The Bayesian optimization was applied to all RNN models, including the best-performing LSTM, Bi-LSTM and stacked LSTM models. Hyperparameters such as the number of layers, learning rate, and regularization etc., were detailed in Table 3. The optimized models were evaluated using the same test dataset and compared with the non-optimized models. Table 6 shows that all models achieved a significant improvement in accuracy after optimization, with stacked LSTM achieving the highest accuracy of 99% and CS-CRNN achieving an accuracy of 98.5%. Following augmentation, the accuracy on the optimized CS-CRNN model demonstrates a notable increase to 99.7%, as evidenced in Figure 10(b).

Furthermore, the optimized models showed a reduction in misclassification errors, particularly in the MR and MVP classes of stacked LSTM and CS-CRNN models, as demonstrated in the confusion matrices of Figure 8(b) and

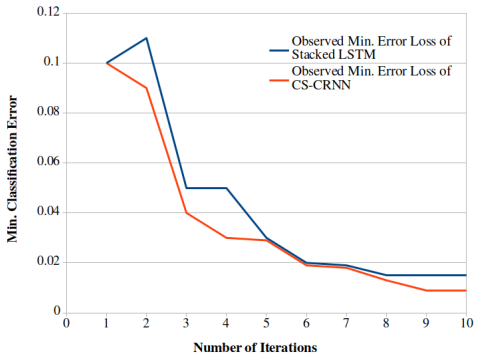


FIGURE 11. Minimum loss function.

Figure 9(b) respectively. These misclassification errors were reduced by 2% and 4%, respectively.

The Bayesian optimization technique demonstrated its effectiveness in enhancing the performance of RNN and CS-RNN models in multiclass classification. Among these models, the stacked LSTM model achieved the highest accuracy, which is nearly comparable to CS-CRNN. Figure 11 visualizes the minimum classification error of optimized stacked LSTM and augmented CS-CRNN models. Notably, the optimized stacked LSTM model exhibited a sudden decrease in the loss function values during training, indicating significant improvements in parameter updates. In contrast, CS-CRNN displayed a smooth curve, suggesting a more gradual convergence process. The minimum observed objective (MOO) [50] values were calculated for the RNN and CS-CRNN models. The RNN model achieved a MOO of 0.015, while the CS-CRNN model had a MOO of 0.009. These values represent the lowest recorded loss function values during the training process. By utilizing Bayesian optimization and fine-tuning the hyperparameters, exceptional results were obtained in the multiclass classification of cardiac abnormalities. The RNN model achieved an impressive accuracy rate of 99%, while the CS-CRNN model achieved a noteworthy accuracy of 99.7% on augmented data. The effectiveness of Bayesian optimization in optimizing the models and achieving high accuracy levels was demonstrated in this study. However, one limitation of the current study lies in the utilization of non-learnable parameters in the scalogram layer within the CS-CRNN model. This layer employs fixed transformations that remain static and do not adapt to the training data, potentially limiting the model's adaptability and transparency. To address this limitation in future research, one potential approach is to incorporate learnable parameters into the scalogram layer. This would optimize the model for feature learning, potentially enhancing its performance, adaptability to various data distributions, and overall transparency.

H. COMPARATIVE ANALYSIS OF PROPOSED SYSTEM WITH EARLIER WORK

This section presents a comparison of a cardiac abnormality diagnostic model with existing approaches, using the PCG sound signal dataset. Utilizing the CS-CRNN model with augmentation, the method demonstrates superior performance. In binary classification (800 abnormal, 200 normal), it achieves 99.6% accuracy and a 99.7% F1-score. The multiclass classification (AS, MR, MS, MVP, N) reaches an optimized accuracy of 99.7%, surpassing the 98.6% obtained by comparison models before optimization. The efficacy of this approach outperforms prior studies, as summarized in Table 7.

VII. CONCLUSION AND FUTURE WORK

This article presents a novel CS-CRNN approach for diagnosing cardiac abnormalities using PCG signals. The CS-CRNN model exhibited strong performance in both binary and multiclass classification tasks. In binary classification, the model already demonstrated high accuracy. In the more complex multiclass classification task, through the application of Bayesian optimization, the CS-CRNN model's accuracy was improved, resulting in reduced error rates. The CS-CRNN model adeptly processes raw PCG data, augmented for improved performance on smaller datasets. It achieves remarkable accuracies of 99.6% for binary classification and 98.6% and 99.7% before and after optimization for multiclass classification on the augmented dataset. These findings emphasize the CS-CRNN model's effectiveness in enhancing accuracy and its role as a robust and reliable tool for diagnosing cardiac abnormalities. However, it is important to acknowledge that one limitation of this study lies in the utilization of non-learnable parameters within the scalogram layer, which may impact the model's adaptability and transparency. In the future, researchers can explore the potential advantages of utilizing raw waveform data in diagnosing cardiac abnormalities, with a specific focus on the incorporation of learnable parameters of the customized deep learning models. Addressing this limitation and incorporating learnable parameters could further enhance the model's adaptability and transparency, ultimately advancing its potential applications in the field of cardiac diagnostics.

Declaration of Interests: The authors declare no conflict of interest.

REFERENCES

[1] L. J. Nowak and K. M. Nowak, "An experimental study on the role and function of the diaphragm in modern acoustic stethoscopes," *Appl. Acoust.*, vol. 155, pp. 24–31, Dec. 2019.
[2] S. Mendis, I. Graham, and J. Narula, "Addressing the global burden of cardiovascular diseases; need for scalable and sustainable frameworks," *Global Heart*, vol. 17, no. 1, p. 48, Jul. 2022.
[3] A. K. Abbas and R. Bassam, "Phonocardiography signal processing," *Synth. Lectures Biomed. Eng.*, vol. 4, no. 1, pp. 1–194, Jan. 2009.
[4] T. H. Chowdhury, K. N. Poudel, and Y. Hu, "Time-frequency analysis, denoising, compression, segmentation, and classification of PCG signals," *IEEE Access*, vol. 8, pp. 160882–160890, 2020.
[5] A. F. M. Moorman and V. M. Christoffels, "Cardiac chamber formation: Development, genes, and evolution," *Physiological Rev.*, vol. 83, no. 4, pp. 1223–1267, Oct. 2003.
[6] C. Rostagno, "Heart valve disease in elderly," *World J. Cardiol.*, vol. 11, no. 2, pp. 71–83, Feb. 2019.
[7] M. S. Sacks and A. P. Yoganathan, "Heart valve function: A biomechanical perspective," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 363, no. 1502, p. 2481, Jul. 2008.

[8] R. B. Hinton and K. E. Yutzey, "Heart valve structure and function in development and disease," *Annu. Rev. Physiol.*, vol. 73, no. 1, pp. 29–46, Mar. 2011.

[9] R. Ghosh, S. Phadikar, N. Deb, N. Sinha, P. Das, and E. Ghaderpour, "Automatic eyeblink and muscular artifact detection and removal from EEG signals using k-nearest neighbor classifier and long short-term memory networks," *IEEE Sensors J.*, vol. 23, no. 5, pp. 5422–5436, Mar. 2023.

[10] A. K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, "Algorithms for automatic analysis and classification of heart sounds—A systematic review," *IEEE Access*, vol. 7, pp. 8316–8345, 2019.

[11] S. A. Matulevicius, A. Rohatgi, S. R. Das, A. L. Price, A. de Luna, and S. C. Reimold, "Appropriate use and clinical impact of transthoracic echocardiography," *JAMA Internal Med.*, vol. 173, no. 17, p. 1600, Sep. 2013.

[12] S. B. Malik, N. Chen, R. A. Parker, and J. Y. Hsu, "Transthoracic echocardiography: Pitfalls and limitations as delineated at cardiac CT and MR imaging," *RadioGraphics*, vol. 37, no. 2, pp. 383–406, Mar. 2017.

[13] S. Li, F. Li, S. Tang, and W. Xiong, "A review of computer-aided heart sound detection techniques," *BioMed Res. Int.*, vol. 2020, pp. 1–10, Jan. 2020.

[14] K. Radha and M. Bansal, "Feature fusion and ablation analysis in gender identification of preschool children from spontaneous speech," *Circuits, Syst., Signal Process.*, vol. 42, no. 10, pp. 6228–6252, 2023.

[15] M. Bansal, R. Sharma, and A. Dagar, "IoT-based heart valve disorder detection using an amplitude and frequency modulated signal model," in *Internet of Things in Biomedical Sciences: Challenges and Applications*. Bristol, U.K.: IOP Publishing Bristol, 2023, pp. 1–8.

[16] H.-L. Her and H.-W. Chiu, "Using time-frequency features to recognize abnormal heart sounds," in *Proc. Comput. Cardiology Conf. (CinC)*, Sep. 2016, pp. 1145–1147.

[17] S. K. Ghosh, R. K. Tripathy, R. N. Ponnalagu, and R. B. Pachori, "Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features," *IEEE Sensors Lett.*, vol. 3, no. 12, pp. 1–4, Dec. 2019.

[18] M. Á. Goda and P. Hajas, "Morphological determination of pathological PCG signals by time and frequency domain analysis," in *Proc. Comput. Cardiology Conf. (CinC)*, Sep. 2016, pp. 1133–1136.

[19] R. B. Pachori, *Time-Frequency Analysis Techniques and Their Applications*. Boca Raton, FL, USA: CRC Press, 2023.

[20] P. Gopika, V. Sowmya, E. A. Gopalakrishnan, and K. P. Soman, "Performance improvement of deep learning architectures for phonocardiogram signal classification using fast Fourier transform," in *Proc. 9th Int. Conf. Adv. Comput. Commun. (ICACC)*, Nov. 2019, pp. 290–294.

[21] A. Yadav, M. K. Dutta, C. M. Travieso, and J. B. Alonso, "Automatic classification of normal and abnormal PCG recording heart sound recording using Fourier transform," in *Proc. IEEE Int. Work Conf. Bioinspired Intell. (IWOBI)*, Jul. 2018, pp. 1–9.

[22] A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Dominguez, "Selection of dynamic features based on time-frequency representations for heart murmur detection from phonocardiographic signals," *Ann. Biomed. Eng.*, vol. 38, no. 1, pp. 118–137, Jan. 2010.

[23] M. Nabhan Homsi and P. Warrick, "Ensemble methods with outliers for phonocardiogram classification," *Physiological Meas.*, vol. 38, no. 8, pp. 1631–1644, Jul. 2017.

[24] S. K. Ghosh, R. N. Ponnalagu, R. K. Tripathy, G. Panda, and R. B. Pachori, "Automated heart sound activity detection from PCG signal using time-frequency-domain deep neural network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.

[25] B. Satyasai, R. Sharma, and M. Bansal, "A gammatonegram based abnormality detection in PCG signals using CNN," in *Proc. 3rd Int. Conf. Artif. Intell. Signal Process. (AISP)*, Mar. 2023, pp. 1–5.

[26] S. Babaei and A. Geranmayeh, "Heart sound reproduction based on neural network classification of cardiac valve disorders using wavelet transforms of PCG signals," *Comput. Biol. Med.*, vol. 39, no. 1, pp. 8–15, Jan. 2009.

[27] S. Patidar and R. B. Pachori, "A continuous wavelet transform-based method for detecting heart valve disorders using phonocardiograph signals," in *Proc. Int. Conf. Hybrid Inf. Technol.* Cham, Switzerland: Springer, 2012, pp. 513–520.

[28] H. Uğuz, "Adaptive neuro-fuzzy inference system for diagnosis of the heart valve diseases using wavelet transform with entropy," *Neural Comput. Appl.*, vol. 21, no. 7, pp. 1617–1628, Oct. 2012.

[29] V. Nivitha Varghees, K. I. Ramachandran, and K. P. Soman, "Wavelet-based fundamental heart sound recognition method using morphological and interval features," *Healthcare Technol. Lett.*, vol. 5, no. 3, pp. 81–87, Jun. 2018.

[30] S. Patidar and R. B. Pachori, "Classification of cardiac sound signals using constrained tunable-Q wavelet transform," *Exp. Syst. Appl.*, vol. 41, no. 16, pp. 7161–7170, Nov. 2014.

[31] G.-Y. Son and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, p. 2344, Nov. 2018.

[32] G. Tian, C. Lian, Z. Zeng, B. Xu, Y. Su, J. Zang, Z. Zhang, and C. Xue, "Imbalanced heart sound signal classification based on two-stage trained DsaNet," *Cognit. Comput.*, vol. 14, no. 4, pp. 1378–1391, Jul. 2022.

[33] E. Ghaderpour, S. D. Pagiatakis, and Q. K. Hassan, "A survey on change detection and time series analysis with applications," *Appl. Sci.*, vol. 11, no. 13, p. 6141, Jul. 2021.

[34] S. Ismail and B. Ismail, "PCG signal classification using a hybrid multi round transfer learning classifier," *Biocybernetics Biomed. Eng.*, vol. 43, no. 1, pp. 313–334, Jan. 2023.

[35] M. Istiaq Ansari and T. Hasan, "SpectNet : End-to-end audio signal classification using learnable spectrograms," 2022, *arXiv:2211.09352*.

[36] Z. Ren, K. Qian, F. Dong, Z. Dai, W. Nejdl, Y. Yamamoto, and B. W. Schuller, "Deep attention-based neural networks for explainable heart sound classification," *Mach. Learn. Appl.*, vol. 9, Sep. 2022, Art. no. 100322.

[37] M. T. Nguyen, W. W. Lin, and J. H. Huang, "Heart sound classification using deep learning techniques based on log-mel spectrogram," *Circuits, Syst., Signal Process.*, vol. 42, no. 1, pp. 344–360, Jan. 2023.

[38] K. Radha and M. Bansal, "Closed-set automatic speaker identification using multi-scale recurrent networks in non-native children," *Int. J. Inf. Technol.*, vol. 15, no. 3, pp. 1375–1385, Mar. 2023.

[39] A. Sepúlveda, F. Castillo, C. Palma, and M. Rodriguez-Fernandez, "Emotion recognition from ECG signals using wavelet scattering and machine learning," *Appl. Sci.*, vol. 11, no. 11, p. 4945, May 2021.

[40] I. J. Brown, "A wavelet tour of signal processing: The sparse way," *Investigacion Operacional*, vol. 30, no. 1, pp. 85–87, 2009.

[41] J. Andén, "Scattering transforms and deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 1–8.

[42] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," 2017, *arXiv:1801.01078*.

[43] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaei, "CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings," *IEEE Access*, vol. 9, pp. 36955–36967, 2021.

[44] A. Mahmoud and A. Mohammed, "A survey on deep learning for time-series forecasting," in *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*. Cham, Switzerland: Springer, 2021, pp. 365–392.

[45] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.

[46] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[47] Z. Wei and Z. Dai, "BiLSTM with novel feature matrix predicts the binding affinity between MHC-I and peptides," in *Proc. 5th Int. Conf. Big Data Technol.*, Sep. 2022, pp. 351–356.

[48] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.

[49] M. Wang, B. Guo, Y. Hu, Z. Zhao, C. Liu, and H. Tang, "Transfer learning models for detecting six categories of phonocardiogram recordings," *J. Cardiovascular Develop. Disease*, vol. 9, no. 3, p. 86, Mar. 2022.

[50] M. A. Gelbart, J. Snoek, and R. P. Adams, "Bayesian optimization with unknown constraints," 2014, *arXiv:1403.5607*.

[51] S. K. Ghosh, R. N. Ponnalagu, R. K. Tripathy, and U. R. Acharya, "Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals," *Comput. Biol. Med.*, vol. 118, Mar. 2020, Art. no. 103632.

[52] M. Alkhodari and L. Fraiwan, "Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105940.

[53] S. L. Oh, V. Jahmunah, C. P. Ooi, R.-S. Tan, E. J. Ciaccio, T. Yamakawa, M. Tanabe, M. Kobayashi, and U. Rajendra Acharya, "Classification of heart sound signals using a novel deep WaveNet model," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105604.

[54] B. Xiao, Y. Xu, X. Bi, J. Zhang, and X. Ma, "Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption," *Neurocomputing*, vol. 392, pp. 153–159, Jun. 2020.
[55] Y. Al-Issa and A. M. Alqudah, "A lightweight hybrid deep learning system for cardiac valvular disease classification," *Sci. Rep.*, vol. 12, no. 1, p. 14297, Aug. 2022.
[56] J. Zang, C. Lian, B. Xu, Z. Zhang, Y. Su, and C. Xue, "AmtNet: Attentional multi-scale temporal network for phonocardiogram signal classification," *Biomed. Signal Process. Control*, vol. 85, Aug. 2023, Art. no. 104934.
[57] R. Wang, Y. Duan, Y. Li, D. Zheng, X. Liu, C. T. Lam, and T. Tan, "PCTMF-Net: Heart sound classification with parallel CNNs-transformer and second-order spectral analysis," *Vis. Comput.*, vol. 39, no. 8, pp. 3811–3822, Aug. 2023.
[58] S. B. Shuvo, S. S. Alam, S. U. Ayman, A. Chakma, P. D. Barua, and U. R. Acharya, "NRC-Net: Automated noise robust cardio net for detecting valvular cardiac diseases using optimum transformation method with heart sound signals," 2023, *arXiv:2305.00141*.

KODALI RADHA received the B.Tech. degree in Electronics and Communication Engineering and the M.Tech. degree in VLSI from Jawaharlal Nehru Technological University, Kakinada, India, in 2012 and 2015, respectively. She is currently pursuing the Ph.D. degree with the School of Electronics Engineering (SENSE), VIT-AP University, Amaravati, Andhra Pradesh, India. She has been an Assistant Professor with the ECE Department, JNTUK, since March 2015. She is also an Assistant Professor with the VR Siddhartha Engineering College, Kanuru, Vijayawada, India. Her research interests include speech processing, machine learning, and deep learning. She is also actively engaged in research related to next-generation AI technologies on children's speech recognition and paralinguistic classification.

MOHAN BANSAL (Senior Member, IEEE) received the B.E. degree in Electronics and Communication Engineering from Rajiv Gandhi Technological University (RGTU), Bhopal, India, in 2008, the M.Tech. degree from the National Institute of Technology (NIT) Kurukshetra, India, in 2011, and the Ph.D. degree in Electrical Engineering from the Indian Institute of Technology (IIT) Kanpur, India, in 2020. He is currently an Assistant Professor with the Indian Institute of Information Technology Sonepat, Sonipat, India. His research interests include digital signal processing, speech signal processing, parametric modeling of the signal, machine learning, and deep learning.

RAJEEV SHARMA (Member, IEEE) received the B.E. degree in Electronics and Communication Engineering from Rajiv Gandhi Technological University, Bhopal, India, in 2009, the M.Tech. degree in Electronic Instrumentation from the National Institute of Technology, Warangal, India, in 2011, and the Ph.D. degree in Electrical Engineering from the Indian Institute of Technology Indore, Indore, India, in 2017. He is currently an Associate Professor with the School of Electronics Engineering (SENSE), VIT-AP University, Amaravati, Andhra Pradesh, India. His research interests include biomedical signal analysis, pattern recognition, deep neural networks, and machine learning.