Advanced Ensemble Learning-Based CNN-BiLSTM Network for Cardiovascular Disease Classification Using ECG and PCG Signal

Ehsan Kalatehjari[1], Mohammad Mehdi Hosseini[2*], Ali Harimi[3], Vahid Abolghasemi[4]

[1] Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran, IRAN, ehsankl@yahoo.com

[2]Department of Computer Engineering, Shahrood Branch, Islamic Azad University, Shahrood, IRAN, Hosseini_mm@yahoo.com

[3]Department of Electrical Engineering, Shahrood Branch, Islamic Azad University, Shahrood, IRAN, a.harimi@gmail.com

[4]School of Computer Science and Electronic Engineering (CSEE), University of Essex, UK, v.abolghasemi@essex.ac.uk

Abstract:

Cardiovascular disease (CVD) is a well-known leading cause of death worldwide. This highlights the need for an effective and efficient diagnostic-therapeutic path for the diagnosis and risk stratification of coronary artery disease (CAD) patients. However, it is inaccurate to investigate CAD only based on either electrocardiogram (ECG) or phonocardiogram (PCG) recordings. Several studies have attempted to use a combination of both signals in the early prediction and diagnosis of CAD. Considering the strong capability of deep learning models in feature extraction this research explores the efficiency of a hybrid CNN-BiLSTM ensemble approach that combines ECG and PCG signals to determine cardiac health status. Inspired by the significant performance of ensemble learning techniques in combining multiple base models to enhance overall prediction accuracy, a hybrid network architecture is suggested. The proposed CNN-BiLSTM model is considered as a baseline for both ECG and PCG signal prediction. Then, a bilinear layer combines both predictions of individual models to obtain a final accurate and robust prediction. It applies a bilinear transformation to incoming outputs from two base models to make the final output. The proposed architecture shows considerable improvement in prediction accuracy compared to using both ECG and PCG signals separately. Employing the well-known PhysioNet/Computing in Cardiology (CinC) Challenge 2016 Database, the proposed method has achieved 97% diagnosis accuracy, which how improvement over comparable methods and various other existing techniques.

Keywords: Coronary Artery Disease, Ensemble Learning, Convolutional Neural Networks, Long Short-Term Memory Networks, Electrocardiogram, Deep Learning

Highlights:

- Novel Hybrid Model: Developed a CNN-BiLSTM ensemble architecture for improved coronary artery disease (CAD) diagnosis.

- Dual Signal Integration: Effectively combined ECG and PCG signals using a bilinear layer to enhance diagnostic accuracy.

- Superior Prediction Accuracy: Achieved a 97% accuracy rate in CAD detection, outperforming existing methodologies.

- Advanced Feature Extraction: Utilized deep learning for precise feature extraction, leading to robust cardiac risk stratification.

- Ensemble techniques: Achieving superior results in the proposed method using ensemble methods.

1- Introduction:

With the increasing prevalence of coronary artery disease (CAD) and its impact on human life, early detection and accurate diagnosis of the disease have become more important than ever. The current cardiovascular disease diagnosis and treatment technologies are based on single-modal signals and the diagnostic process is complex, leading to inconvenience for patients. Furthermore, the only existing way that could provide a full understanding of the health status of the patient's heart is through invasive approaches, which are inconvenient, and require sophisticated equipment. However, artificial intelligence-assisted auscultation, while a low-cost cardiac diagnostic technology, relies heavily on the clinician's experience, which may lead to diagnostic errors, especially in cases of complex heart diseases [1]. Although ECG examinations are common in modern cardiovascular diagnosis, considering the widely accessible and noninvasive characteristics of the ECG method, may still lack sensitivity in detecting some diseases. As a result, patients may experience delayed diagnosis and treatment, leading to disease progression and the loss of the best opportunity for treatment. Hence, clinicians rely on multi-modal detection data, such as PCG, ECG, and cardiac color Doppler ultrasound, for diagnosis. For instance, in a study by Stepanov et al., cardiac hemodynamic parameters were determined by using the combined features of ECG, PCG, and cardiac impedance signals, and the relevant parameters of cardiac impedance were obtained with the help of ECG and PCG signals [1]. By considering the morphological changes in the ECG signal, such as ST-segment elevation, ST-segment depression, T-wave inversion, and T-wave flatness, can be considered as indications of myocardial ischemia or infarction resulting from coronary artery disease (CAD) [2]. Moreover, previous research has demonstrated that the PCG may contain subtle high-frequency murmurs generated by turbulence in coronary arteries that have become narrowed [3]. Similar to ECG, a photoplethysmogram (PPG) can be utilized for monitoring various cardiovascular conditions. Over the years, numerous computerized systems have been proposed for the automatic analysis of ECG [4, 5], PPG [6], and other modalities in combination with these signals [7]. The ECG, PPG, and PCG signals are all cyclostationary, meaning that while the signal statistics may vary, they repeat periodically within a specific period. ECG and PCG signals are highly correlated [8], and they are known to contain more information than the PPG signal. While ECG provides information about the electrical activity of the heart, PCG records the acoustic properties of the heart. This feature gives PCG a distinct advantage over ECG and PPG signals. The PCG signal, in addition to its acoustic properties, is well-suited for detecting murmurs, which are abnormal heart sounds [9]. Moreover, the PCG signal has an advantageous starting trigger in the form of the S1 wave [10, 11].

In recent years, convolutional neural networks (CNNs) have achieved remarkable success in many applications including computer vision and multidimensional data analysis. Furthermore, researchers have become interested in applying CNNs to the classification of ECG and PCG signals [12-14]. The convolutional neural network (CNN) can learn high-level features automatically by building multiple hidden layers and convolution operations, simplifying the

process of feature extraction by avoiding the need for expert knowledge-based feature engineering. While CNN-based techniques have shown effectiveness in ECG or PCG classification, there is still room for improvement in their performance. In practical applications, relying solely on single-domain features may not be enough to provide sufficient information for accurate classification. The research on multiple decision-making methods has highlighted the benefits of multi-modal signal analysis and has overcome the limitations of single-model signal diagnosis [15-17]. Despite advancements in using CNN and BiLSTM models for ECG and PCG signal analysis, current methods still struggle with accurately detecting coronary artery disease due to the reliance on single-domain features. The integration of multi-modal signals in a robust, unified framework remains underexplored. Although some recent efforts have been made to benefit from multimodal signal advantages, the integration of these diverse data sources using deep learning techniques holds great potential for enhancing the performance of signal analysis, enabling more comprehensive insights and improved decision-making across various applications, such as healthcare diagnostics.

This study addresses the need for improved accuracy and reliability in CAD detection through a hybrid CNN-BiLSTM ensemble approach, combining ECG and PCG signals. To this end, the paper aims to utilize ensemble methods to combine PCG and ECG-based disease evaluation techniques and overcome individual analyses' limitations. The goal is to develop techniques to enable the assessment of cardiac health status based on the analysis of both ECG and PCG signals simultaneously. In contrast to the existing ensemble methods which use a homogeneous database, here, we deal with two different signals which are inhomogeneous. Additionally, the proposed deep learning model is straightforward yet delivers promising performance. Inspired by ensemble learning methods, the model first makes predictions for each ECG and PCG input signal using individual CNN-BiLSTM models. A final network, which includes a bilinear layer, then combines the outputs of these two models. Specifically, the bagging strategy is employed as a fusion solution to integrate the different outputs from the multimodal database, enhancing the overall prediction accuracy. Although deep learning techniques have been in considerable use for CAD detection, most related studies to date have the key limitation of relying on single-modality data, such as ECG or PCG. Most of these models poorly exploit the available multimodal data; they also lack sophisticated strategies for the fusion of heterogeneous signal sources. Moreover, many models have placed a major concentration on ECG-based class discriminations, foregoing several diagnostic details that may be possible from other signals related to spectrograms, VCG, and even EMRs.

Some of these efforts use publicly available data, while others use internal datasets to achieve their objectives. Samiul et al [18] suggested a new framework for Cardiovascular Disease Classification called CardioXNet. CardioXNet is a lightweight, end-to-end CRNN architecture specifically designed for the automatic detection of five classes of cardiac auscultation using raw PCG signals. The architecture operates in two learning phases: representation learning and sequence residual learning. During representation learning, the model efficiently extracts time-invariant features and converges quickly. In the sequence residual learning phase, it captures temporal features without the need for explicit feature extraction. This architecture achieves superior performance, with accuracy reaching up to 99.60%, precision at 99.56%, recall at 99.52%, and an F1-score of 99.68%. CardioXNet not only has accuracy, and improvement over comparable previous methods but also is highly suitable for point-of-care CVD screening in low-resource environments using memory-constrained mobile devices. While CardioXNet achieves high

accuracy with PCG signals, multi-modal classification using both PCG and ECG signals can enhance robustness and comprehensiveness. It is reported in [19], that cardiovascular disease poses a significant threat to human life and health. Holter surveillance has transformed ECG monitoring into remote cardiac monitoring. A new wireless ECG patch was developed using deep learning frameworks. Existing models struggled to differentiate two main heartbeat types, resulting in low accuracy. A semi-supervised method was proposed, achieving an average accuracy of 91.2%. In a similar work, [20] proposed a method that aims to develop a deep-learning-based system to predict cardiovascular diseases like arrhythmia and heart failure from abnormalities in ECG signals. The system uses a model combining BiLSTM networks and CNNs. Experiments using ECG data from MIT-BIH and BIDMC databases showed the proposed approach outperformed existing methods in both scenarios. Another work [21] A deep-learning-based approach is proposed to classify ECG signals into sixteen arrhythmia classes, enabling the diagnosis of cardiovascular diseases. The method uses continuous wavelet transform, D-CNN, attention block, and a clump of features framework. The proposed framework achieves 99.84% accuracy, 100% sensitivity, and 99.6% specificity, outperforming state-of-the-art techniques in accuracy, F1-score, and sensitivity. Pankaj et.al [22], This work propose an optimized deep learning model that can estimate Systolic Blood Pressure (SBP) and diastolic blood pressure (DBP) by categorizing the stages of BP through a single-channel PPG signal. In doing so, it utilizes the CNN framework along with the superset transform method to convert the 1-D PPG signal into a 2-D super-resolution time-frequency spectrogram in a less complex manner that is highly feasible on wearable devices bound by battery resources. The model produced a mean absolute error of 2.71 mmHg and classification accuracy of 96.79% for SBP prediction and 98.94% for DBP.

Numerous studies have leveraged PCG signals for the classification of various heart conditions. For instance, in [23], an in-house dataset comprising 150 PCG signals was employed to classify different types of cardiac arrhythmias. Utilizing mel-frequency cepstral coefficients (MFCC) for feature extraction and the k-nearest neighbors (k-NN) classifier, the study achieved an accuracy of 94.23% across three classes (Normal, Atrial Fibrillation, and Other Arrhythmias). As shown in the [24], Heart valve disorders (HVDs) are caused by damage to heart valves and can lead to congestive heart failure, hypertrophy, and stroke. Early detection using PCG signals is crucial to minimize cardiac complications. This article proposes a time-frequency-domain deep learning (TFDDL) framework for the automatic detection of HVDs using PCG signals. The approach uses a deep convolutional neural network (CNN) model to detect four types of HVDs. The TFDDL model has achieved 99% and 99.48% accuracy in detecting HVDs and 85.16% accuracy in classifying normal and abnormal heart sound classes. However, the literature reveals significant advancements in the automatic detection and classification of cardiovascular diseases (CVDs) and heart valve disorders (HVDs) using various machine learning techniques and signal processing methods. Despite these advancements, challenges such as improving accuracy, reducing computational complexity, and enhancing real-time applicability remain. The recent approaches to CVD and HVD automatic detection have demonstrated remarkable development concerning high accuracy, more complex architectures in machine learning techniques such as CRNN, CNN, BiLSTM, and wavelet transforms, and the capability to work in real-time, especially in resource-limited environments. In addition, large datasets have been utilized for robust training. Despite the significant advancements in the automatic detection of CVD and HVD using machine learning techniques, several gaps and limitations exist in contemporary research. Generally, most of the existing methods are characterized by high computational complexity, limiting their direct application on memory-constrained mobile devices used in point-of-care settings. The

generalization and scalability across different datasets remain limited as most of the methods require high-quality labeled data, which is not available for many datasets. Moreover, the black-box nature of deep learning models compromises the interpretability and transparency of the diagnosis, posing a problem for clinical acceptance. Although high accuracy scores have been reported with some models using single modalities like PCG or ECG signals, such approaches do not always capture the full repertoire of conditions and may miss potentially important diagnostic information. The limitations can be addressed by developing more resource-lean and interpretable multi-modal approaches that can be integrated into real-time, resource-constrained healthcare environments. Here, a new approach is introduced that attempts to address these limitations by jointly classifying PCG and ECG signals using multi-modal classification. The aim is to integrate the strengths of each modality in clinical reasoning to improve the robustness and coverage of the classifier.

a novel co-learning-assisted progressive dense fusion network (CL-PDFN) is proposed for CVD detection with both ECG and PCG signals [25]. A progressive dense fusion strategy has been presented in the proposed methodology to extract effectively the complementary features from both ECG and PCG signals and enhance the robustness of CVD detection. The co-learning framework enables iterative improvements in the feature representations through knowledge sharing across the two modalities of signals. Experimental results verify that the CL-PDFN significantly raises the detection accuracy bar compared to its traditional single-modality and early fusion counterpart methods. The best performance, especially for complex cases, shows the great potentiality of the model toward a reliable and precise diagnosis of CVD. [26] presents some efficient pooling convolutional models for the multi-classification of ECG-PCG signals, which helps enhance the detection of cardiovascular diseases. After preprocessing the signals, the authors constructed several structural blocks, including a convolutional-max-pooling-stacked block called MCM and a residual block called REC. All models can handle different sampling rates by adjusting the number of structural blocks. In the final tests, the model using MCM block reached an accuracy of 98.70% and 92.58% for ECG-PCG fusion datasets with the best performances compared to other variations. Also, the MCM block outperforms the model based on REC by 0.02% on ECG and by 4.30% on PCG datasets. This research has been validated by testing on many different ECG and PCG datasets and was compared with other various published works. Finally, to verify the generalizability of the model, an experiment was conducted using a synchronized ECG-PCG dataset. Seven levels of fatigue were tagged in this dataset based on physical activity performed by healthy subjects. In this experiment, again the highest accuracy was achieved by the MCM-based model for proving its robust performance on different datasets and applications. In [27], focus on the concomitant use of the ECG and PCG signals for the detection of cardiovascular diseases. Both ECG and PCG signals carry significant, complementary, and non-invasive information about heart function from two different perspectives. Combining these modalities to detect CVD poses a challenge due to the complexity involved in extracting discriminative features without losing critical information. To handle this, we propose DDR-Net, which is a dual-scale deep residual network that automatically extracts features from raw ECG and PCG signals. A dual-scale feature aggregation module integrates low-level features across different scales. Then, Support Vector Machine combined with Recursive Feature Elimination with Cross-Validation follows the selection of the most important features, then the use of SVM for the final

classification. Our approach was evaluated on the "training-a" set from the 2016 PhysioNet/CinC Challenge database. The comparison of performance verifies that our approach outperforms not only the single-modal approaches based on ECG or PCG but also the existing multimodal approaches, achieving an accuracy of 91.6% and an AUC of 0.962. We go further in a detailed analysis of the feature importance of ECG and PCG in CVD detection. [28] proposes a transformer-based multi-feature fusion network, MF-CADNet, to detect CAD and arrhythmias. Our model combines information from ECG, its spectrograms, vectorcardiograms (VCG), and electronic medical records (EMR). It uses convolutional kernels with different receptive fields to capture features across various segments in the signal and inter-lead and inter-module attention mechanisms for extracting more holistic features. We tested MF-CADNet on the publicly available PTB Diagnostic ECG Database and our self-developed PKUSZ Diagnostic CAD Database, respectively. In these two datasets used, the average F1 score for the diagnosis was 98.41% and 93.49%, respectively. These results underline the effectiveness of MF-CADNet and present a promising approach for non-invasive CAD screening. One of the most profound limitations within the literature is the under-exploitation of multi-modal sources since most models fail to seamlessly integrate diverse signal types. Most works so far used traditional convolutional networks or hybrid CNN-BiLSTM architectures, which were doing well but suffered due to a lack of better mechanisms for feature extraction. Proposed models lack the attention mechanism for full exploitation of inter-lead and inter-module relationships within multi-model data.

2- Materials & Method

2-1 Dataset Description

The PhysioNet/Computing in Cardiology (CinC) Challenge 2016 dataset [29] has been employed in this study. This dataset originated from the PhysioNet/CINC challenge held in 2016, specifically focusing on heart sound classification and hosted on the PhysioNet website. The PhysioNet/Computing in Cardiology Challenge 2016 comprises ECG and PCG signals acquired from a heterogeneous sample of the population, with equal gender distribution and a range of ages between 18 and 85 years. It is expected that this demographic distribution will support the generalization of the model for different cardiac profiles. The datasets, labeled as *training-a* to *training-f*, were gathered from various research institutions by PhysioNet. Notably, the *training-a* subset, comprising 409 samples, was contributed by MIT and features a combination of ECG and PCG signals. Out of these samples, 405 contain both ECG and PCG signals, while the remaining 4 contain just PCG signals. Within this dataset, 113 signals are categorized as normal, while 292 signals are considered abnormal. All the signals were initially recorded at a sampling rate of 44.1 kHz, but they were subsequently resampled at 2000 Hz during post-processing. Figure 1 shows the distribution of data.

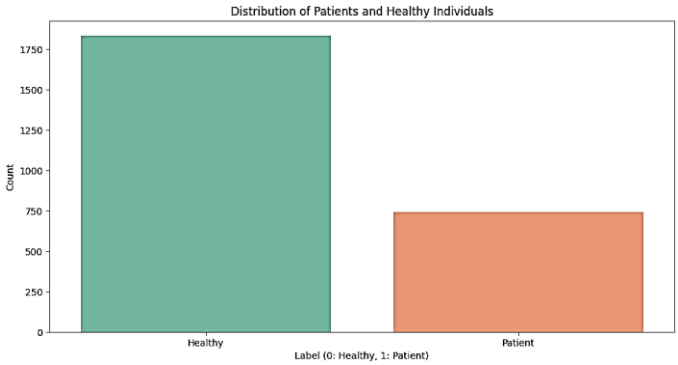


Figure 1. the distribution of data, healthy individuals compared to patients

For the study, 70% of the signals were designated for training, and the remaining 30% were allocated for testing. Figure 2 displays the sample signals taken from the PhysioNet dataset and typical waveforms regarding PCG and ECG recordings. These signals are of great importance in the diagnosis of cardiovascular diseases; ECG captures the electrical activity of the heart, while PCG captures its acoustic properties like heart sounds and murmurs. This figure highlights the importance of integrating both signals for comprehensive cardiac analysis in the proposed CNN-BiLSTM model.

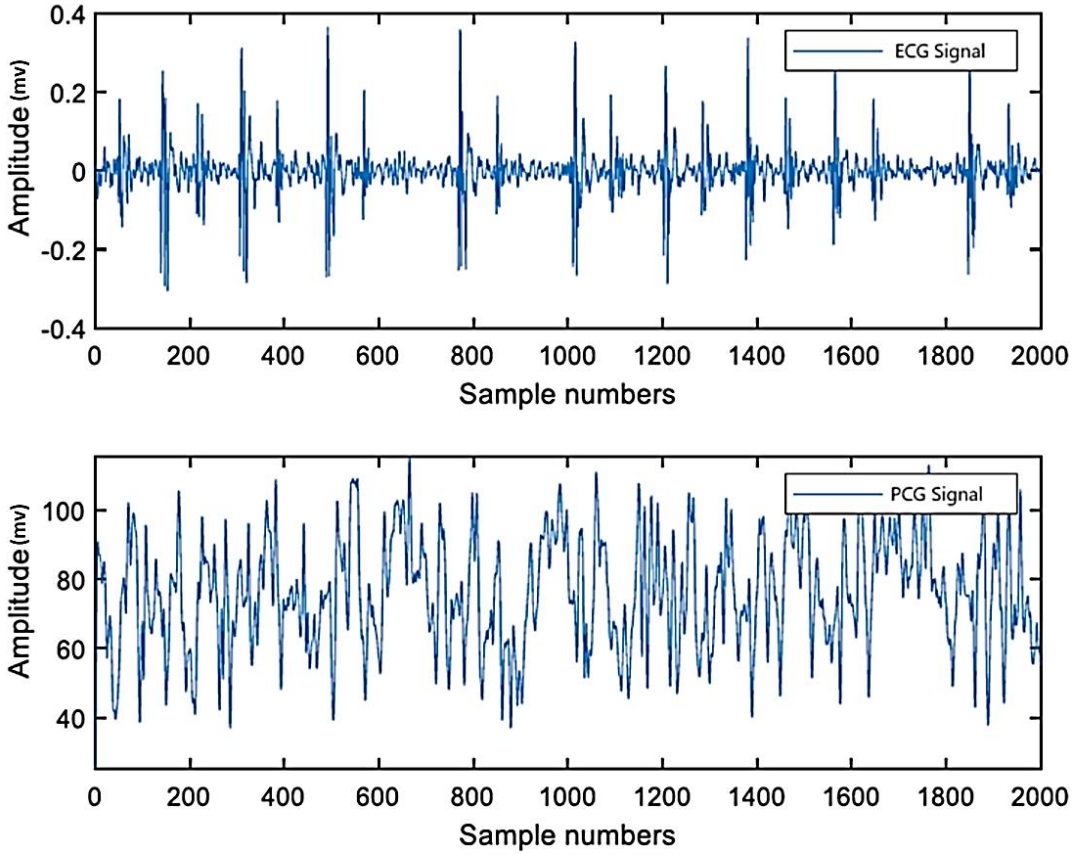
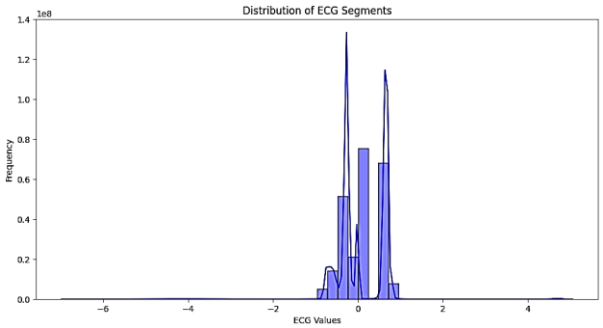
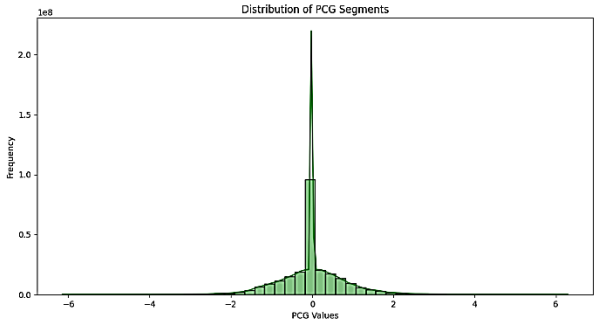


Figure 2. Examples of PCG and ECG Signal in physionet dataset [25]

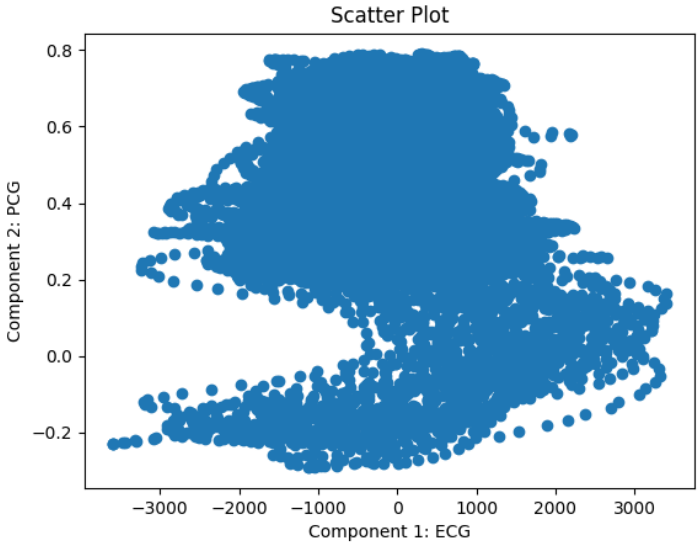
Figures 3 collectively describe the statistical behavior and relationships of ECG and PCG data. Figure 3-1 was used to depict the distribution of ECG segment values. The histogram with a superimposed density curve indicated that the range of ECG values falls between a very small interval around zero. This reflects limited variability with intermittent extreme values. Figure 3-b shows the distribution of PCG segment values. As with the ECG data, the PCG values are densely clustered around zero, but the distribution appears more symmetric and smoother than the ECG. Figure 3-c demonstrates a scatter plot of ECG values against PCG values. As shown in this plot, the plot of Component 1 versus Component 2 displays a non-linear relationship between the two components, as suggested by the high-density cluster at the center and scattered points at various edges that demonstrate signal alignment variability.



a) the distribution of ECG segment values



b) the distribution of PCG segment values



c) a scatter plot of ECG values (Component 1) versus PCG values (Component 2)

Figure 3. The statistical behavior and relationships of ECG and PCG data. Figure a, elaborates on the distribution of ECG signal amplitudes between the range -1 to -4 and its appropriate frequencies ranging from -0.8 to 0.0. This visualization highlights the variety and concentration of ECG amplitudes. Plot b is more informative about the amplitude characteristic and variability present within the PCG signals. Figure c shows the scatter plot for ECG versus PCG. The amplitudes for ECG and PCG correspond to the x and y axes respectively. It will be useful to observe any possible correlation or trend between these two physiological signals.

2-2 Proposed Method

In this study, a method based on the combination of ECG and PCG signals, utilizing BiLSTM-CNN and ensemble learning, is proposed for predicting and classifying signals in the database. These preprocessed signals are then fed into two CNN-BiLSTM structures. The proposed architecture consists of two configurations. In the first configuration, CNN and LSTM are used sequentially, where the output of the CNN is fed into the LSTM. In the second configuration, CNN and LSTM are used in parallel, and their outputs are combined. the outputs of two hybrid networks are fed into an ensemble network. For both steps, the ECG and PCG signals are preprocessed, mainly involving filtering procedures and segmentation. Finally, the processed signals are forwarded for further analysis and prediction. The architecture of the proposed ensemble model is represented in Figure 4, representing CNN-BiLSTM networks integrated to concurrently process ECG and PCG signals. This architecture leverages the strengths of both networks: the extraction of spatial features with CNN and temporal dependencies with BiLSTM in the data. This integrated approach improves classification performance related to the diagnosis of coronary artery disease by combining predictions obtained from both types of signals.

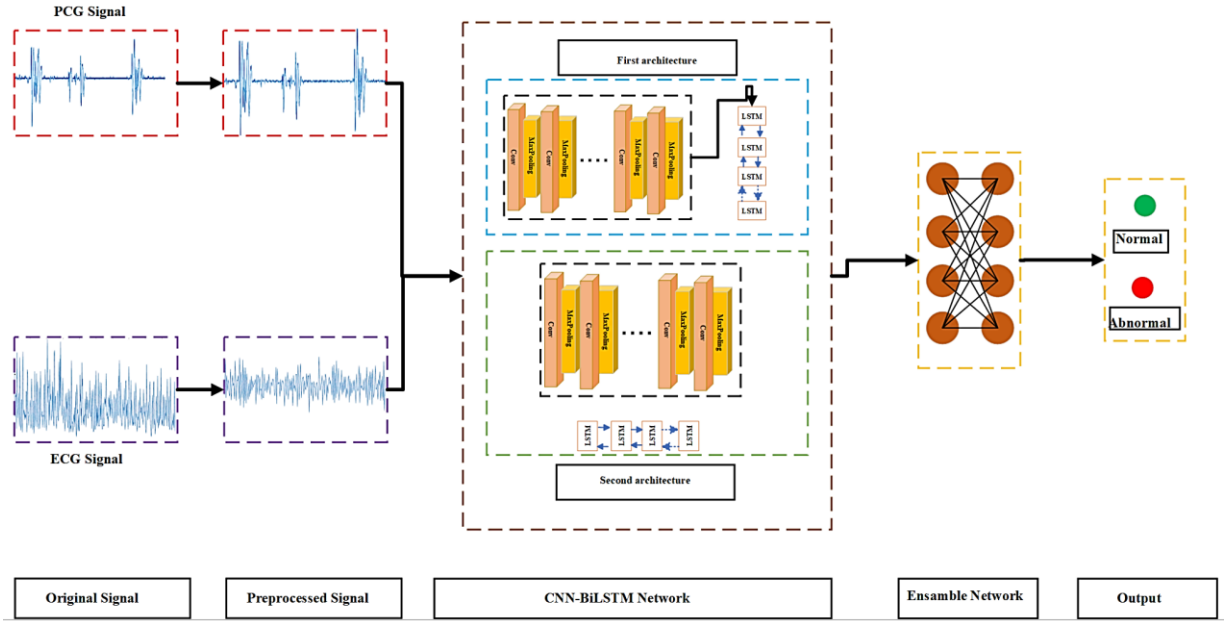


Figure 4. The Proposed System Architecture

2-2-1 Preprocessing: denoising and segmentation

The pre-processing phase involves two steps: firstly, denoising PCG and ECG signals, and secondly, signal segmentation to enhance the model's accuracy. During the denoising step, low-pass filtering cleans noise and artifacts in the ECG signal. On the other hand, band-pass filtering is applied to the PCG signal for the removal of interference and baseline wander. At this point, the denoised signals are segmented into subsections of interest, usually targeting individual heartbeats or other specific cardiac cycles. Peak detection, thresholding, and windowing are used for segmentation and are introduced to ensure that a model is trained or used only on meaningful data segments. The filtering of ECG and PCG signals was performed using constant cutoff frequencies to provide normalization across different signals. The cutoff values were selected to effectively reduce noise while retaining critical diagnostic information relevant to cardiovascular analysis.

These low-pass and band-pass filters are, in particular, IIR Butterworth filters, chosen for their smooth frequency response and ability to minimize noise without excessive distortion. Unlike FIR filters, the IIR Butterworth filters used here offer computational efficiency with carefully designed recursive structures to ensure stability for denoising applications. The frequency response of the Butterworth filters ensures no significant ripples in the passband or stopband, preserving the physiological integrity of the ECG and PCG signals. The difference equations and transfer functions for these filters are detailed in Appendix A.

As shown in Figure 5, this entire preprocessing procedure ensures that the signals are well-prepared for further analysis. the pipeline of preprocessing steps carried out on raw ECG and PCG signals before their use in feeding the CNN-BiLSTM model. Among others, preparatory steps include the removal of noise and frequencies outside the range of interest through low-pass and band-pass filters, respectively, followed by the segmentation of cardiac cycles. In this way, only the most meaningful parts of the signals have been used for training, which intrinsically enhanced the performance and accuracy to predict cardiovascular abnormalities from the model.

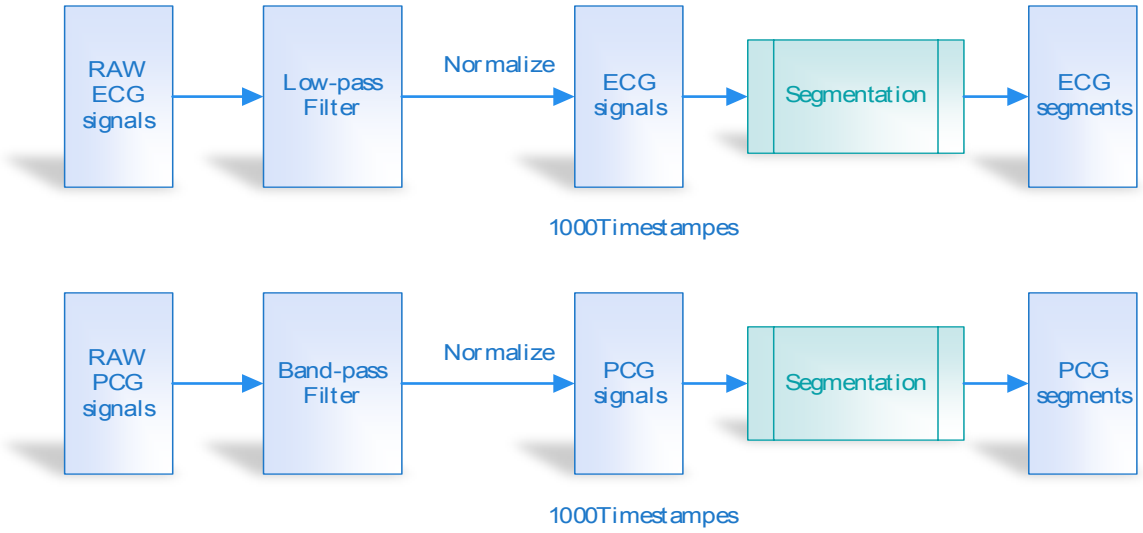


Figure 5. The procedure of preprocessing of RAW input signals

For denoising, the input ECG and PCG signals undergo a filtering process for purification from noise in the original signals before being fed into the network. By noise here, we mean any unwanted signal masking or distorting the useful information contained in the ECG and PCG signals. The useful information of PCG signals is mainly concentrated below 500Hz and above 25Hz. For ECG signals, the effective information is already concentrated below 20Hz. Therefore, a low-pass filter with a cutoff frequency of 20 Hz was applied to ECG signals and a bandpass filter with a low cutoff frequency of 25 Hz and a high cutoff frequency of 400 Hz was applied to PCG signals. This initial step aims to enhance the quality of the signals by removing noise and irrelevant information. Figure 6 illustrates the comparison before and after filtering. In (a) can notice that the original ECG signal contains significant noise and baseline wandering that could mask several important diagnostic features. After applying a low-pass filter (in b), the signal is now clearer and

shows main interest features like the P, QRS, and T waves in and ready for an accurate diagnosis of coronary artery diseases.

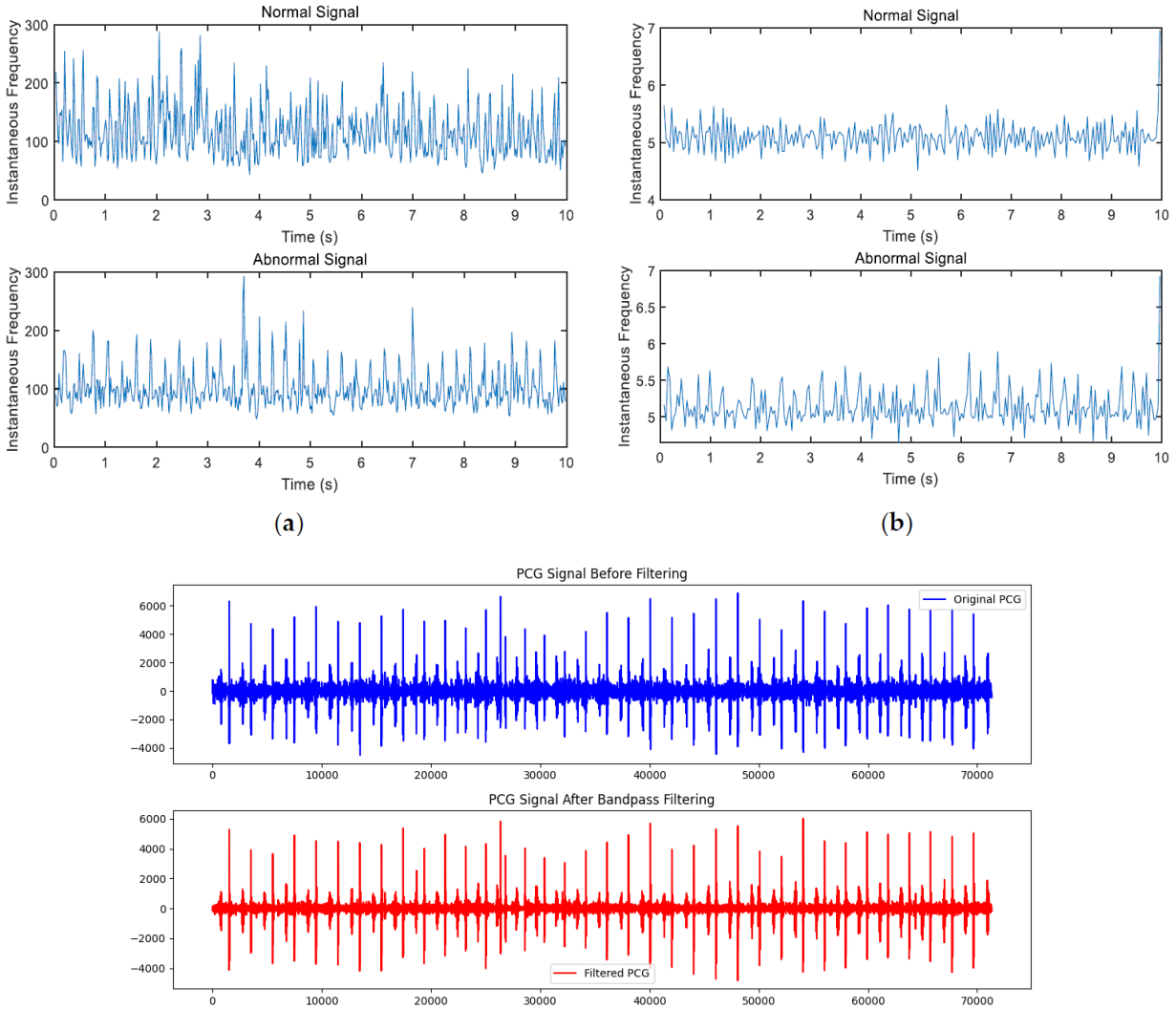


Figure 6. Comparison before and after filtering: (**a**) Original ECG; (**b**) ECG signal after filter (**c**) Original PCG and after applying filter

Figure 7 shows the distributions of ECG and PCG segment values both before and after preprocessing. The left plot of ECG distribution before processing shows that values are concentrated narrowly around zero, but slight asymmetry and outliers can be observed. The right plot in the figure shows a distribution after preprocessing and, thus, it has a more normalized structure with fewer outliers, improving data uniformity. These comparisons illustrate that preprocessing techniques enhance data by yielding a better outcome for subsequent analyses with improved statistical characteristics.

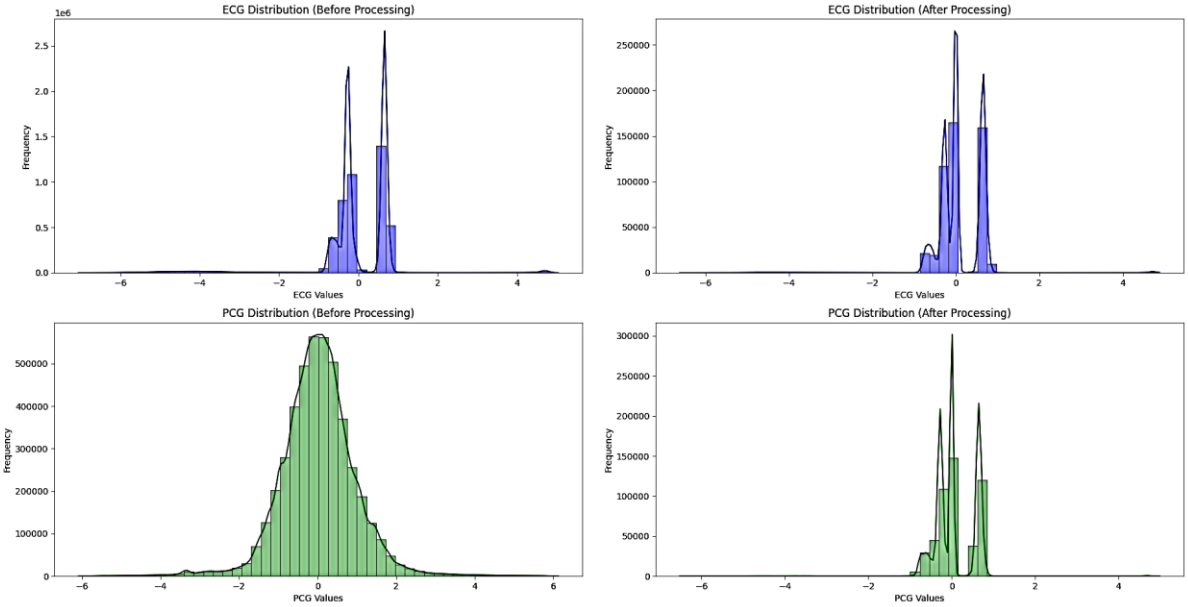
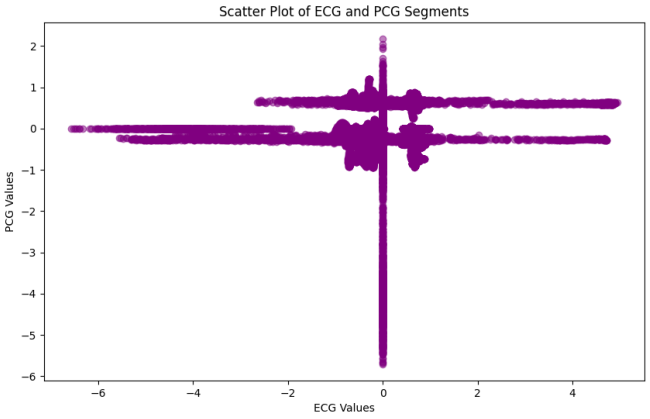
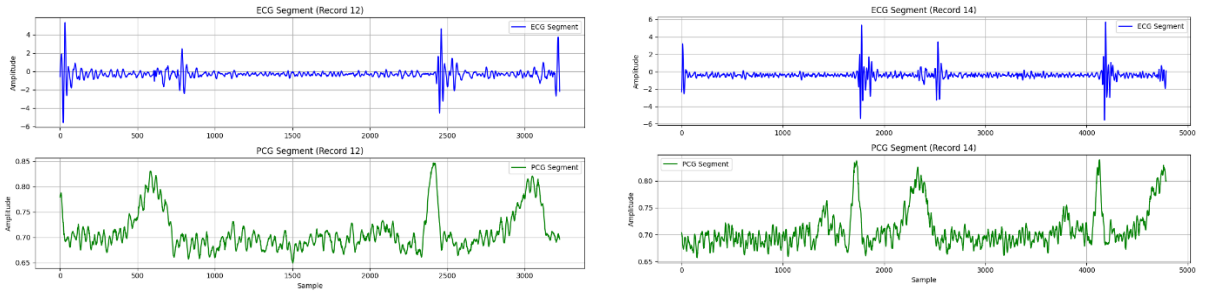


Figure 7. The distributions of ECG and PCG segment values before and after preprocessing

Figure 8-a shows the scatter plot of the segments ECG and PCG signals. The x-axis represents the normalized ECG values, and the y-axis represents the normalized PCG values. It gives the dispersion of the data points with the clusters in view, showing variability and correlation of the two physiological signals. Figure 8-b demonstrates simultaneous ECG and PCG data.



a) the scatter plot of the segments ECG and PCG signals



b) samples of simultaneous ECG and PCG data together in one record.

Figure 8. Relationship between ECG (Electrocardiogram) and PCG (Phonocardiogram) signal segments. The scatter plot shows the relation between the amplitude of ECG and PCG signals. Thus, this plot is capable of showing possible correlations or patterns between these two physiological signals. This analysis and figure b provides an understanding of the interaction between cardiac electrical activity (ECG) and acoustic heart sounds.

To standardize model inputs and to reduce the training time, sequences exceeding 1000 timestamps were truncated. Traditional methodologies for detecting R peaks often employ techniques such as wavelet transformations, frequency analysis, and digital filtering to pinpoint local maximum values. In our study, an R peak indicator algorithm exclusively focused on R peak indices was applied. It specifically calculates the intervals between consecutive R peaks, which are referred to as R-R intervals. Atrial fibrillation, as one of the common abnormal heart rhythms, is characterized by more than 4 R-R intervals, our algorithm segments the ECG signals by at least four R-R intervals. Unlike traditional approaches, this method intentionally avoids integrating additional indicators, enabling our model to directly extract features from unprocessed ECG signals for a more streamlined and precise analysis (see Appendix B). In contrast, the method described in [18] follows a different approach. Our method's streamlined approach distinguishes it by focusing solely on direct feature extraction from raw ECG signals. The PCG signals are recorded simultaneously with ECG signals, thus, we extract the same intervals of related PCG signals as a new input feature for our networks. This strategy reduces the amount of processing cost and provides more precise paired feature values. One of the advantages of employing the segmentation stage in our algorithm is that it effectively serves as a data augmentation approach. By segmenting the original database into smaller segments, we create a new, larger database with many more samples. This increase in sample size is beneficial, as it is well-known that having a larger number of samples can significantly improve the accuracy of deep learning algorithms. Figure 9A illustrates the segmentation of a sample ECG signal into episodes, regardless of their durations, while Figure 9B, gives the waveforms of how the beats have been segmented, where it clearly shows the positions of the P-wave, QRS-complex, and T-wave, which are all closely related to the position of the R-peak. t1 and t2 in Figure 9(B) are the start and end points of the segmented cardiac cycle, respectively. The selection of these points is done based on ECG R-peaks. This provides congruence in the alignment of ECG-PCG segments and hence effective multimodal analysis. The superiority of Figure 9B to Figure 9A lies in the fact that it details individual heartbeats. Figure 9B zeroed in on segmentation at the beat level, detailing the salient components: the P-wave, QRS-complex, and the T-wave. Such fine-grained segmentation is necessary for detecting specific cardiac events with a high degree of accuracy and hence enhances the reliability of the model in the pickup of abnormalities and consequently diagnostic accuracy. Whereas the coarser segmentation in Figure 9A may miss all this important information, it could reduce the efficacy of the model in extracting necessary features from the ECG signal.

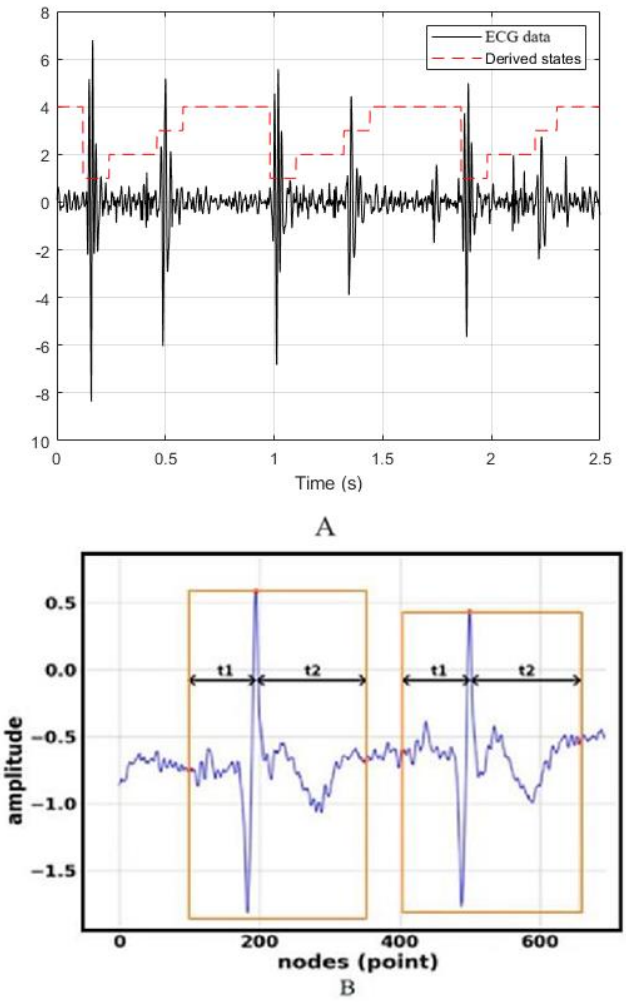


Figure 9. ECG signal segmentation (A) shows a coarse segmentation of 1600 nodes for rhythm features, useful for overall heart rate and rhythm analysis(B) focuses on finer segmentation at the beat level, with 252 nodes detailing individual components such as the P-wave, QRS-complex, and T-wave

2-3 Network Fundamentals

2-3-1 Bidirectional LSTM

Long Short-Term Memory (LSTM) networks have played a pivotal role in sequence modeling due to their memory-retaining capabilities [26]. To further enhance their performance, Bidirectional LSTM (BiLSTM) networks were introduced [27]. BiLSTMs, as the name suggests, process input sequences in both forward and backward directions simultaneously. This dual processing allows them to capture not only past but also future context, making them exceptionally adept at understanding the temporal dynamics of sequential data. The structure of LSTM is shown in Figure 10 which consists of three different gates: input, forget, and output gates. These gates play a vital role in ensuring that information can be selectively passed, allowing for the retention of essential information within the transmission unit. These block or allow the passing of information so that only the most discriminative details from ECG and PCG are allowed to pass. This selective

retention is achieved through a series of linear operations, which, in turn, ensures the preservation and invariance of critical information throughout transmission.

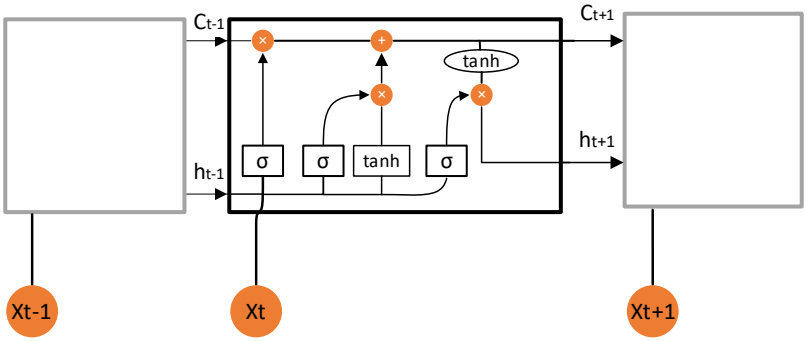


Figure10. the internal structure of an LSTM network, including input, forget, and output gates [26]

By strategically memorizing information while selectively discarding less significant data, the system maximizes the effective utilization of sample data, resulting in improved efficiency and accuracy. Here is equations for forget gate(f_t), input gate (I_t) and output gate (O_t):

$$I_t = \partial(X_t * U_t + H_{t-1} * W_t) \tag{1}$$

$$f_t = \partial(X_t * U_f + H_{t-1} * W_f) \tag{2}$$

$$O_t = \partial(X_t * U_o + H_{t-1} * W_o) \tag{3}$$

In above equations X refers to the input, U is the weight associated with the input, W is the weight matrix associated with the hidden state and H is the hidden state of the previous timestamp. To calculate the hidden state O_t and tanh of the updated cell state is used as follow:

$$H_t = O_t * \tanh(C_t) \tag{4}$$

Where the C_t represents the cell state which is defined as:

$$C_t = f_t * C_{t-1} + I_t * \tanh(X_t * U_c + H_{t-1} * W_c) \tag{5}$$

Which U_c and W_c are related to weight values. In the internal structure of a BiLSTM, similar to LSTM, the nodes in the hidden layers employ intricate activation mechanisms. They selectively retain or discard information to mitigate issues like gradient explosion and vanishing gradient. This advanced architecture has proven highly effective in various tasks such as natural language processing, speech recognition, and time series analysis, where capturing bidirectional dependencies is crucial for accurate modeling and predictions. The structure of BiLSTM is represented in Figure 11, which processes input sequences in both forward and backward directions. This eventually helps in capturing information from both past and future contexts, a very useful ability while analyzing sequential data such as ECG or PCG. This facilitates the ability of the model to incorporate information from both directions into diagnosing subtle irregularities within cardiac signals.

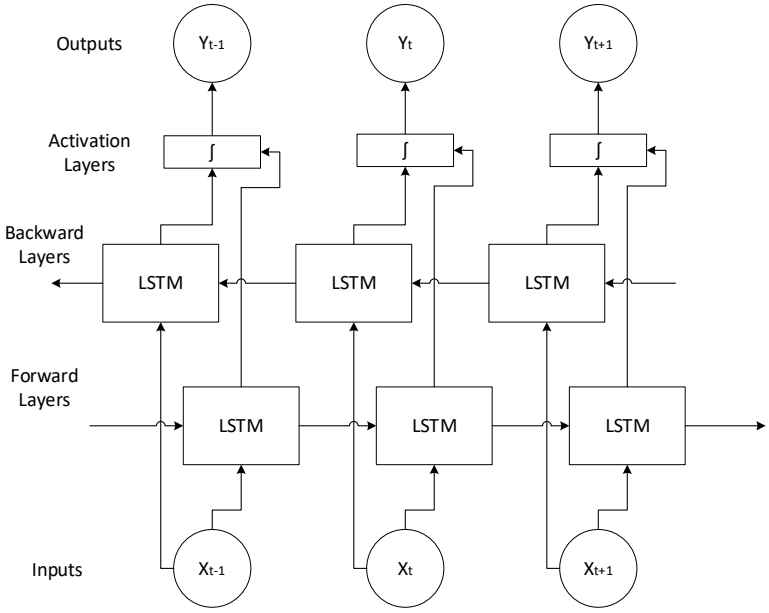


Figure 11. structure of BiLSTM network [27]

The input sample signal sequence has information from both the forward LSTM calculation and the reverse LSTM calculation. Both calculations (forward, reverse) would be considered to calculate the output y_t at time t as follows:

$$\vec{h}_t = \tanh(W_{\vec{h}t} x_t + W_{\vec{h}t\vec{h}t} h_{t-1} + b_{\vec{h}t}) \tag{6}$$

$$\overleftarrow{h}_t = \tanh(W_{\overleftarrow{h}t} x_t + W_{\overleftarrow{h}t\overleftarrow{h}t} h_{t-1} + b_{\overleftarrow{h}t}) \tag{7}$$

$$y_t = \tanh(W_{\vec{h}o} \vec{h}_t + W_{\overleftarrow{h}o} \overleftarrow{h}_t + b_{ho}) \tag{8}$$

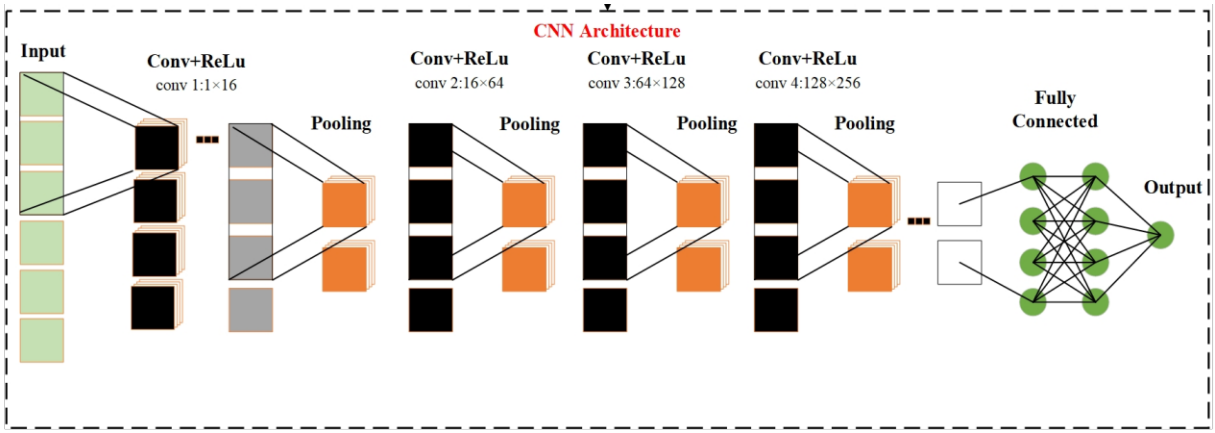


Figure 12. the proposed CNN architecture for extracting spatial features from the input ECG and PCG signals

2-3-2 CNN-BiLSTM structure

As shown in Figure 4, the proposed model employs two different structures for combining CNN and BiLSTM networks. In the first structure the output of the CNN network is fed into the LSTM network as known as CNN feed LSTM combination. Second structure, named CNN concat BiLSTM, is a parallel combination of two mentioned networks. The CNN captures implicit suitable features from ECG signals rather than hand-on feature extraction. On the other hand, the BiLSTM units perform temporal dimensionality reduction to avoid the overfitting problem introduced by the huge feature maps and therefore, the most useful features for the task of classification are selected. As a result, the combination of two networks makes an increase in the classification results. The CNN used here is a one-dimensional convolutional neural network that includes four convolutional layers. Each convolutional layer is followed by a normalization layer, a dropout layer, and Max Pool layer. Also, the first two layers of the network include normalization and dropout. The architecture of CNN which was described here is shown in Figure 12, As evidenced in this figure, input data is passed through several convolutional and ReLU layers. These CNN layers learn hierarchical representations of the signal progressively. The shapes of waves and their frequencies relevant to classifying cardiovascular diseases are then learned. The structure is parceled in such a way that only the most salient features are extracted that could enhance the performance of a hybrid CNN-BiLSTM model.

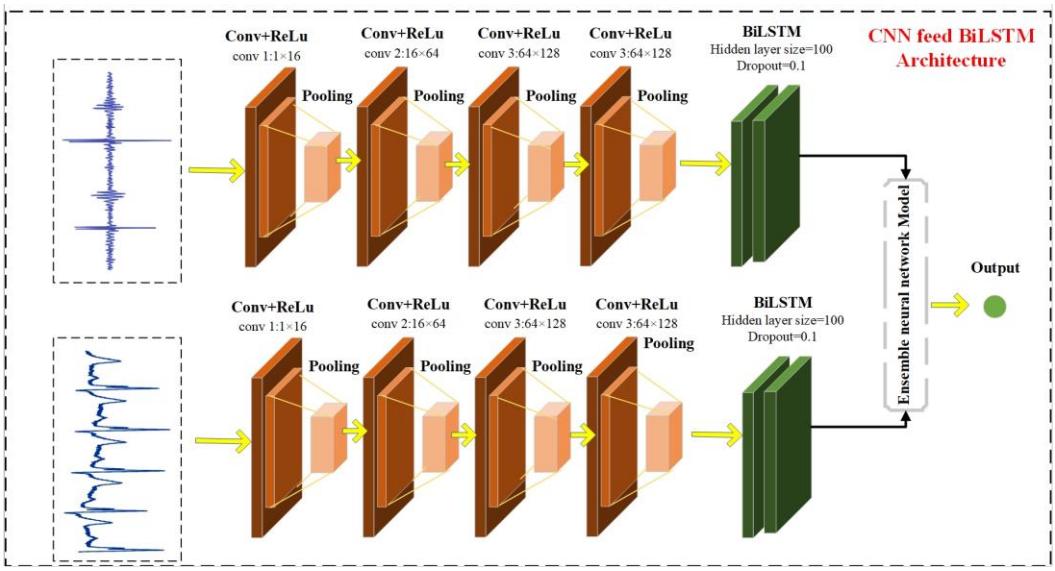


Figure 13. The key hyperparameters used in the first proposed CNN-BiLSTM architecture, detailing critical components such as input size, hidden layers, activation functions, and optimization techniques

In the proposed architecture, parameter tuning was guided by empirical evaluation and prior research on optimal network configurations for ECG and PCG signal analysis. Specifically, this network has varying filter sizes (1:1x16, 2:16x64, 3:64x128, 4:128x256), each followed by pooling layers to reduce dimensionality. The final feature maps are fully connected to the output layer. Consequently, an ECG signal of length 1000 is converted into a feature vector of size 256 after passing through this network. Two neural network models are designed for processing raw data through noise removal and segmentation (Figures 13 and 14). In the CNN feed BiLSTM architecture, the CNN extracts features from the time-series data and feeds them into a BiLSTM for classification, enabling the CNN to capture local patterns. Meanwhile, the BiLSTM learns

from the dependencies that exist in time before it can predict the outcome. This architecture sends the pre-processed input through several convolutional and ReLU layers before passing it to BiLSTM layers, akin to CNN. The output now goes through BiLSTM layers, with the hidden size of the layer being 100, after which it goes through a dropout of 0.1 to the ensemble neural network model for the final output prediction. In Figure 14, the CNN part extracts spatial features of time series data independently in the concat BiLSTM structure, while the BiLSTM part holds the temporal relationships in the features extracted.

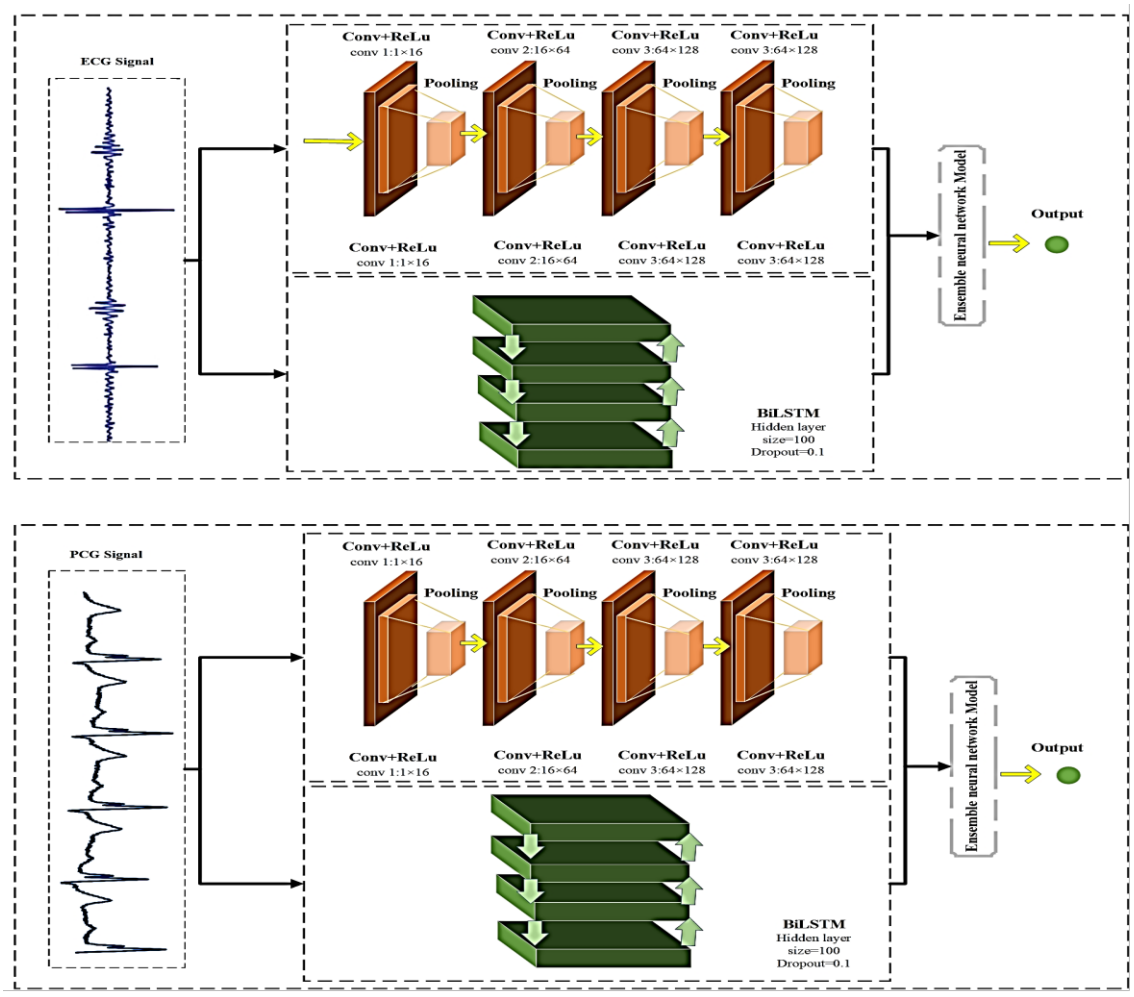


Figure 14. The hyperparameters for the second proposed CNN-BiLSTM architecture, which uses a parallel combination of CNN and BiLSTM networks. This figure provides a detailed breakdown of configurations like filter sizes, pooling layers, and BiLSTM hidden sizes

The features extracted from the BiLSTM and CNN are further extracted, combined end-to-end, and passed through fully connected layers for prediction within the input data. This combination method retains the potential of both models and therefore performs better than basically any individual model in tasks dealing with sequential data.

2-3-3 Ensemble modeling

As described earlier, the proposed model uses two different data series consisting of ECG and PCG signals to diagnose coronary disease. The combined CNN- BiLSTM architecture is applied

to each data series separately. Here, we propose an ensemble model based on a synthesized learning technique to fuse the outputs of the two branches as shown in Figure 4. First, the individual CNN for ECG and BiLSTM for PCG signals are trained separately. Their outputs are then fed to a bilinear layer, which learns the interactions of the features from both modalities. Then, the combined output is fed into a fully connected layer to make the final classification result. By doing this, the complementary strengths derived from each model contribute to enhancing robustness and final ensemble prediction accuracy. The proposed model enhances the classification accuracy by merging predictions from both models. In general context, the ensemble learning method combines the outputs of diverse models to improve the overall performance of the learning system. In ensemble learning, what is pursued is the mitigation of errors and enhancement of performance by leveraging the various diverse strengths of models, which not only encourages the robustness of predictions but also improves results on most tasks within machine learning and data analysis. There are different types of ensemble techniques such as simple but powerful ones, namely max voting, averaging, and weighted Averaging, or advanced techniques such as Stacking, Blending, Bagging, and Boosting. In this paper, we use the ensemble learning approach to create a deep learning-based multimodal ECG and PCG signal processing framework for cardiovascular disease classification. Here, the proposed model is inspired by the bagging ensemble learning method [28]. It involves fully connected layers at the end of a complex architectural model as an ensemble network. Therefore, in the ensemble section, we used a feedforward neural network as a fully connected layer. A fundamental part of the ensemble network is the bilinear layer which efficiently combines the output of two previous model types that were trained on two different datasets (ECG and PCG signals). Such ensemble networking can increase the accuracy of the prediction manyfold, as it would be treated with all the features that are extracted from both signals, therefore allowing more complex patterns to merge.

Table 1. Key factors in the model

Layer Type	Input Size	Output Size	Hidden Layers	MaxPooling	BiLSTM Hidden Size	Dropout	Activation Function
CNN	1x16	16x16	-	Yes	-	-	ReLU
CNN	16x64	64x64	-	Yes	-	-	ReLU
CNN	64x128	128x128	-	Yes	-	-	ReLU
BiLSTM	128x128	100	1	-	-	0.1	Tanh

2-4 Evaluation metrics

To evaluate the performance of the proposed algorithm, a set of essential performance metrics is utilized. These metrics encompass classifications, accuracy, sensitivity, specificity, and the F1 score. The division of signals into normal and abnormal categories dictates the treatment of abnormal signals as positive instances and normal signals as negative ones. Consequently, accurate classification of positive signals constitutes true positives (TP), while misclassification results in false negatives (FN). Similarly, precise classification of negative signals leads to true negatives (TN), while misclassification gives rise to false positives (FP) situations [24]. These diverse metrics collectively offer a comprehensive assessment of the algorithm's classification accuracy and errors, catering to a nuanced understanding of its performance across multiple dimensions. The number of instances correctly classified serves as a foundational indicator of the algorithm's overall performance, and accuracy provides a holistic measure of prediction correctness. Sensitivity becomes crucial in contexts where accurate identification of positive instances is

paramount, as in medical diagnoses. Conversely, specificity evaluates the algorithm's accuracy in identifying negative instances, especially crucial in applications like security systems. Moreover, the F1 score, a balanced amalgamation of precision and recall, provides a holistic view of the algorithm's ability to minimize both false positives and false negatives.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{9}$$

$$Recall = \frac{TP}{(TP + FN)} \times 100 \tag{10}$$

$$Specificity = \frac{TN}{FP + TN} \times 100\% \tag{11}$$

$$Precision = \frac{TP}{(TP + FP)} \times 100 \tag{12}$$

$$F1 - score = \frac{2 \times \mathrm{Re}\,call \times \mathrm{Pr}\,ecision}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision} \times 100\% \tag{13}$$

3- Experimental Results

3-1 Experimental setup

The experimentation was carried out utilizing an Intel (R) CORE-i5 CPU 13600KF3.5GHz processor with 32.00 GB of DDR5 RAM and a 3080 NVIDIA graphic card. The software employed for the experimental work was Python. Key parameter settings in the model encompass the size of the input vector, the number of hidden layers, as well as the pertinent training parameters, among others. The optimum configuration for parameters in the proposed network model is detailed in Table 1. These values are obtained empirically through our extensive experiments. All layers in the model use the Adam optimizer with a learning rate of 0.001, a batch size of 32, and are trained for 50 epochs.

3-2 Results Overview

To demonstrate the effectiveness of the proposed model, classification between healthy and unhealthy classes was performed on the dataset explained in Section 2.1. In this study, various combinations of deep neural networks were designed to identify coronary artery disease using PhysioNet datasets. Initially, a network was constructed by combining a CNN-LSTM, which utilized ECG and PCG data as input. In the subsequent phase, the ensemble technique was employed to combine the outcomes of two methods using the bilinear combination approach. This involved merging the outputs from numerous models generated by each method and then integrating them through the bilinear method. Also, to have a strong model evaluation, a k-fold cross-validation procedure was performed, with k set to 5 to obtain an appropriate tradeoff between model training and validation. Each iteration within a fold included a separate training and validation dataset to avoid leakage; thus, there was no overlap of samples used in training within the validation. The results reported hereafter are more reliable and robust. Moreover, this cross-

validation process prevents overfitting to the specific subsets of data that the model was shown, hence increasing the credibility of the results.

Table 2. the result of model

method	optimizer	Input signal	Accuracy	Sensitivity	Specificity	precision	F1 score
CNN	adam	ECG	0.795	0.794	0.786	0.795	0.723
		PCG	0.816	0.789	0.790	0.793	0.796
		ECG+PCG	**0.827**	**0.896**	**0.898**	**0.891**	**0.832**
BiLSTM	adam	ECG	0.915	0.924	0.926	0.915	0.905
		PCG	0.896	0.905	0.913	0.911	0.916
		ECG+PCG	**0.911**	**0.916**	**0.918**	**0.907**	**0.906**
CNN concat BiLSTM	sgd	ECG	0.902	0.904	0.896	0.915	0.915
		PCG	0.906	0.915	0.927	0.926	0.936
		ECG+PCG	**0.927**	**0.916**	**0.928**	**0.927**	**0.937**
CNN feed BiLSTM	sgd	ECG	0.914	0.903	0.909	0.915	0.921
		PCG	0.917	0.915	0.907	0.916	0.911
		ECG+PCG	**0.977**	**0.960**	**0.977**	**0.967**	**0.976**
CNN concat BiLSTM	adam	ECG	0.935	0.934	0.946	0.941	0.925
		PCG	0.946	0.941	0.947	0.951	0.956
		ECG+PCG	**0.927**	**0.926**	**0.931**	**0.933**	**0.936**
CNN feed BiLSTM	adam	ECG	0.925	0.944	0.926	0.945	0.944
		PCG	0.956	0.935	0.947	0.960	0.966
		ECG+PCG	**0.967**	**0.964**	**0.968**	**0.970**	**0.961**

By leveraging the ensemble technique, we incorporated predictions from diverse models, each capturing distinct aspects of the data. Subsequently, the bilinear combination method facilitated the fusion of these predictions, taking into account their interactions and boosting the overall predictive capacity of the ensemble. Thus, exploiting the ensemble technique, we combined those models looking after different features of the dataset. Whereas CNN layers capture a pattern of spatial signals, the BiLSTM layers model temporal dependencies, which improves the performance in prediction. This stage aimed to capitalize on the complementary strengths of different methods while mitigating potential drawbacks, ultimately resulting in more robust and accurate predictions. Through this iterative process, we aimed to optimize the ensemble model's performance and maximize its efficacy in addressing the target problem. The results were summarized, and the performance of the network was evaluated using metrics such as precision, accuracy, recall, and F1 score. Table 2 shows the best-performing models for both optimizers with different combination models. From Table 2, we can see, the proposed framework performs quite well on the PhysioNet dataset showing around 97.77% accuracy, 96.78% precision, 96.73% recall, and 97.68% F1 Score. The CNN exhibited a promising ability to discriminate between different

groups of heart diseases based on the provided ECG and PCG data. Nevertheless, further analysis and refinement of the model may be necessary to enhance performance and extend applicability to diverse data types and real-world scenarios. upon altering the optimization function from Stochastic Gradient Descent (SGD), an improvement in the results was noticed. The revised optimization function possibly introduced more efficient learning dynamics, promoting faster convergence and superior model performance. This modification may have facilitated the model in exploring the parameter space more effectively, allowing it to discover better solutions or escape from suboptimal local minima. Consequently, the optimization function adjustment enhanced the training process, leading to enhanced outcomes which can be seen in Table 2.

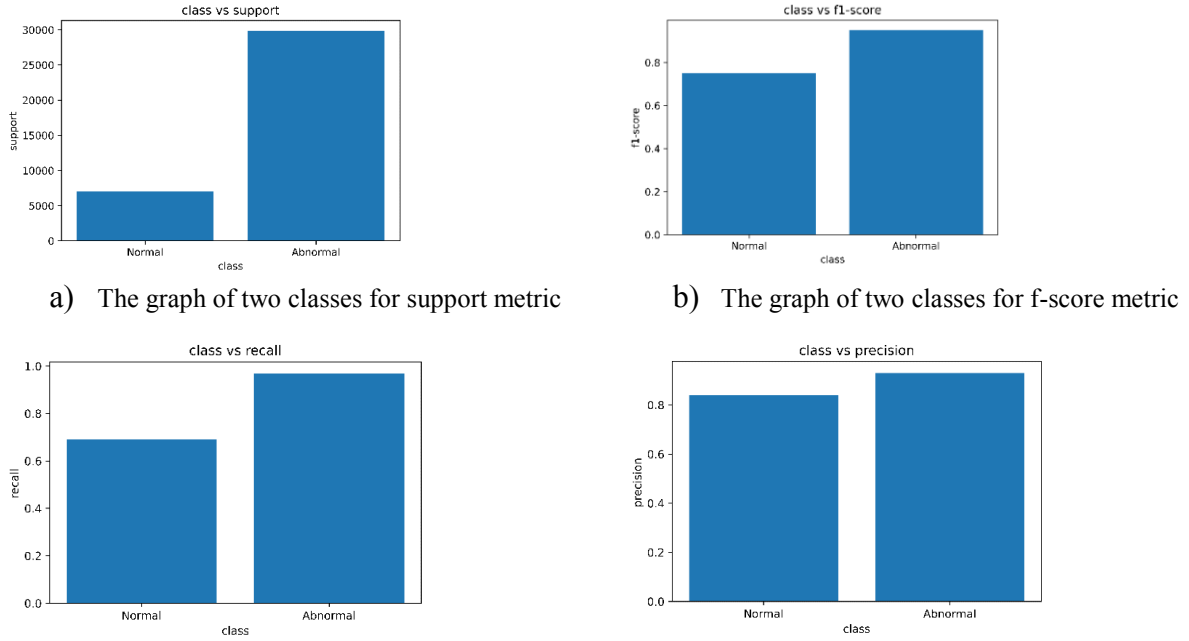
Table 3. The Comparison Between Proposed Method and Other Methods

Method	Sensitivity	Accuracy	F1- score	AUC
Support vector machine [29]	92.31%	92.5%	92.14%	92.23%
Dual-Input Neural Network[30]	98.5%	95.6%	96.58%	97.54%
1-D CNN[31]	90.8%	87.0%	90.85%	90.83%
BiLSTM-GoogLeNet-DS[32]	98.48%	96.13%	97.05%	97.77%
Proposed Method	**96.23%**	**97.77%**	**97.68%**	**96.98%**

3-2-1 Detailed Result Interpretation

In this paper, different methods drawn from the literature are compared. The research focuses more on deep learning methods and how efficient they are in the classification of signals. All the analyses are done on the ECG and PCG signals, mostly on the physionet Database. The results of this comparison are presented in Table 3. This table summarizes the comparison of various approaches in terms of sensitivity, accuracy, and F1-score. For this case, the Support Vector Machine will yield a sensitivity of 92.31%, an accuracy of 92.5%, and an F1-score of 92.14%. This seems to be doing fairly well but relatively lower compared to other techniques. Dual-Input Neural Network performs better on sensitivity with a value of 98.5% and an F1-score of 96.58%, though its accuracy is marginally inferior with a value of 95.6%. The 1-D CNN had the worst performance among all the methods listed, with a sensitivity of 90.8%, an accuracy of 87.0%, and an F1 score of 90.85%. One of the strong results is given by BiLSTM-GoogLeNet-DS: sensitivity comes out to be 98.48%, accuracy is 96.13%, and the F1-score is 97.77%. The proposed method is outstanding in terms of sensitivity—96.23%—the highest accuracy—97.05%—and the F1-score

of 97.68%, which thus indicates a higher overall performance and better robustness concerning the other methods under evaluation. As the results in Tables 2 & 3 demonstrate and as previously mentioned, the features examined are largely consistent, concentrating on time-domain, frequency-domain, and time-frequency domain attributes. Classification is carried out using 1D convolutional layers techniques. It is found that the method proposed in this paper outperforms other methods documented in the literature regarding classification accuracy. It means that the proposed algorithm has some advantages. In addition, it can be noticed that the discrimination ability of the classification approach with ECG and PCG signals is more potent than that with any single-modal signal. Among algorithms with ECG and PCG signals, the best classification performance belongs to that proposed in this paper. It is very similar to that of the approach in [33], although the proposed approach has slight advantages.



a) The graph of two classes for support metric

b) The graph of two classes for f-score metric

c) The graph of two classes for recall metric

d) The graph of two classes for precision metric

Figure 15. The performance of the proposed method is based on the classification activity between a normal and abnormal heart condition using the BIDMC database. These graphs compare the precision, recall, and F1-score for both classes. Precision refers to the ratio of the number of correct positive predictions, while recall refers to the proportion of actual positives correctly identified. The F1-score is indicative of the balance between precision and recall. Here, the proposed model showed higher accuracy in the "Normal" class. This proves the reliability of the model in different datasets for the detection of cardiac abnormalities.

Another experiment conducted was the evaluation of the proposed method on the BIDMC databases dataset, as shown in Figure 15. The provided diagram in Figure 15 consists of bar graphs that showcase the performance evaluation of the proposed model in a classification task, using precision, recall, and F1-score as metrics. The model's performance is compared for two classes, "normal" and "abnormal". The A graph demonstrates precision, which measures the ratio of correct positive predictions to all positive predictions. The "normal" class has a higher precision than the "abnormal" class. The c graph focuses on recall, representing the proportion of actual positive

instances accurately identified by the model. Here, the "abnormal" class has a higher recall than the "normal" class. The B graph presents the F1-score, a combined metric derived from precision and recall, offering an overall performance measure. In this scenario, the F1-score slightly favors the "normal" class. The loss function acts as a measure to assess the effectiveness of a prediction model in predicting anticipated outcomes. When employing the Mean Square Error (MSE) loss function, greater importance is given to outliers, particularly when errors demonstrate such values. In the process of collecting ECG data, the presence of uncontrollable noise interference can lead to isolated points and outliers. To address these errors during the training of ECG data, we opt for the Root Mean Square Error (RMSE) function. This function helps regulate excessively large or small error values, thereby mitigating the influence of outliers on the overall model. By adapting the model to minimize the impact of outlier data points, the proposed loss function enhances resilience against outliers. Figure 16 visually illustrates the curves of the training and test loss functions. Following the implementation of RMSE control, ECG data containing outliers and isolated points show reduced fluctuations in the loss function, resulting in a more consistent error trend. These results prove that the feature fusion of ECG and PCG signals outperforms single-signal methods, thus pointing out the advantage of multimodal analysis. The improved classification capability is due to the spatial feature extraction capability of the CNN component and the temporal dependency captured by BiLSTM. Besides, the proposed model was compared with other methods in the literature, as presented in Table 3, making it a competitive approach for the detection of CAD. These findings therefore suggest that the model of ensemble learning exploits the signals of both ECG and PCG very effectively, hence enhancing its diagnostic reliability.

3-2-2 Negative Testing Results

Toward this end, further negative testing experiments were performed, to validate the robustness of the proposed CNN-BiLSTM ensemble model. Scenarios have been tested that quantify performance under challenging situations where there are false positive and false negative cases. This kind of testing is rather important in establishing how well a model generalizes to cases quite far from typical patterns seen during training. Table 4 presents the negative testing results of the proposed model compared to other baseline models, such as CNN, BiLSTM, and Support vector machine [29] architecture. It calculates accuracy, sensitivity, specificity, and F1-score in conditions of increased noise, altered pattern of the signal, and data with artifacts. It is observed that the proposed model maintains higher accuracy and robustness compared to the comparative models and can be resilient in diverse and challenging scenarios.

Table 4. The negative testing results of the proposed model compared to other baseline models

Model	Optimizer	Input Signal	Scenario	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
Support Vector Machine	N/A	ECG+PCG	Without Noise	92.5	92.3	93.0	92.1
			Noise-added ECG & PCG	90.0	90.0	91.0	89.5
			Signal with Artifacts	87.5	88.0	89.0	87.0

			Randomized Signal Sequence	85.5	85.0	86.5	84.0
CNN	Adam	ECG	Without Noise	79.5	79.4	78.6	72.3
			Noise-added ECG & PCG	76.0	75.0	76.5	70.0
			Signal with Artifacts	72.5	71.0	73.0	68.0
			Randomized Signal Sequence	70.0	68.5	71.5	66.0
		PCG	Without Noise	81.6	78.9	79.0	79.6
			Noise-added ECG & PCG	78.5	76.5	77.0	75.2
			Signal with Artifacts	75.0	73.0	74.0	72.0
			Randomized Signal Sequence	73.0	71.5	73.5	70.0
		ECG+PCG	Without Noise	82.7	89.6	89.8	83.2
			Noise-added ECG & PCG	80.5	86.0	86.5	81.0
			Signal with Artifacts	77.5	83.0	83.5	78.0
			Randomized Signal Sequence	75.5	81.0	81.5	76.0
BiLSTM	Adam	ECG	Without Noise	91.5	92.4	92.6	90.5
			Noise-added ECG & PCG	89.5	91.0	91.5	88.5
			Signal with Artifacts	87.0	88.0	89.0	85.5
			Randomized Signal Sequence	85.0	86.0	87.0	84.0
		PCG	Without Noise	89.6	90.5	91.3	91.6
			Noise-added ECG & PCG	87.5	88.5	89.5	87.0
			Signal with Artifacts	85.0	86.0	87.0	84.0
			Randomized Signal Sequence	83.5	84.5	85.5	83.0
			Without Noise	97.06	96.2	97.50	97.7

Proposed Method	Adam	ECG+PCG	Noise-added ECG & PCG	95.5	94.8	96.2	94.5
			Signal with Artifacts	91.2	90.0	92.5	90.6
			Randomized Signal Sequence	89.0	88.0	90.0	87.5

These results indicate that the proposed CNN-BiLSTM ensemble model may outperform the rest of the models under several challenging scenarios by yielding higher accuracy and F1-scores in noise-added and artifact-prone environments. This points out the robustness and adaptability of the model; hence, it makes the model perfectly suitable for the identification of cardiovascular conditions even under suboptimal conditions.

3-2-3 Limitations

Although the proposed CNN-BiLSTM ensemble model had significant results, it still presents some limitations. The dataset on which this study is conducted had certain sample types, and the generalizability of the model may not be assured across a wide range of patients. The ensemble model, by its very nature, requires higher computational resources for training, which again may restrict its application in real-time or resource-constrained environments. Further exploration in optimizing the model for low computational power environments is needed.

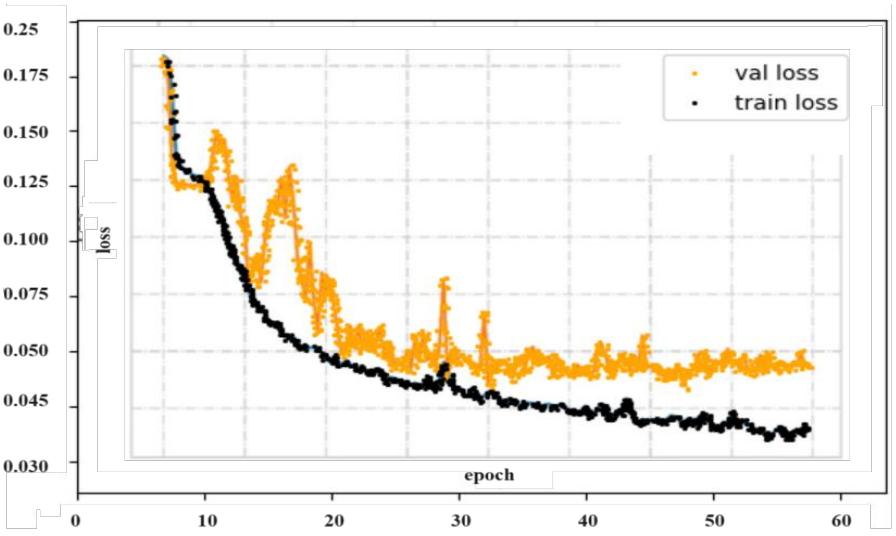


Figure 16. Loss function curves during the training and testing phases of the proposed CNN-BiLSTM model. This graph represents a decrease in error while the model is learning to better predict cardiovascular conditions by minimizing the root mean square error. The plot in this figure shows how the RMSE function makes the effect of outliers, especially in noisy ECG data, bound and hence can generate more stable and reliable model performance.

3-2-4 Future Work

Building on the results of this study, future research could be done on the following aspects:

- Model Optimization for Real-time Applications: This would make the model clinically more accessible at locations with resource constraints by developing lightweight versions.

- Including Additional Signal Modalities: The inclusion of complementary signals, such as PPG, would introduce a richer data pool to improve the diagnostic accuracy of the model.

- Extended Testing on Larger and More Diverse Datasets: It needs to be investigated further in larger datasets involving different demographics for its robustness and clinical application.

Conclusion

In this study, we applied CNNs and BiLSTM networks to extract time-varying and time-invariant features, respectively. This combination of representation and sequence residual learning achieved a remarkable accuracy of up to 97.77%, surpassing most results on the used benchmarks. We experimented with two structural designs: first, placing the BiLSTM layer after the CNN, and second, merging CNN with BiLSTM layers. In both designs, we preprocessed ECG and PCG signals through filtering and segmentation, followed by feeding the processed data into a classifier for analysis and prediction. The outputs from the classifiers were integrated into an ensemble model for final classification. Compared to other methodologies proposed in deep learning frameworks, our method demonstrates superior performance. Our unimodal-based multimodal ECG and PCG processing framework for cardiac analysis is more effective than classical machine learning approaches. The innovation of this work lies in the simultaneous combination of PCG and ECG signals within a CNN-BiLSTM network featuring two distinct architectures. This dual architecture better integrates both signals, enhancing the model's ability to represent temporal characteristics and time-invariant features. By processing PCG and ECG signals simultaneously, our model extracts complementary information from both sources, resulting in a more accurate and robust system for cardiovascular disease detection. This methodology sets a promising method with competitive performance for multimodal signal processing within deep learning frameworks.

Reference:

[1] Lv J, Dong B, Lei H, Shi G, Wang H, Zhu F, Wen C, Zhang Q, Fu L, Gu X, Yuan J, Guan Y, Xia Y, Zhao L, Chen H. Artificial intelligence-assisted auscultation in detecting congenital heart disease. Eur Heart J Digit Health. 2021 Jan 6;2(1):119-124. doi: 10.1093/ehjdh/ztaa017. PMID: 36711176; PMCID: PMC9708038.

[2] Kumar, M., Pachori, R. B., & Acharya, U. R. (2017). Characterization of coronary artery disease using flexible analytic wavelet transform applied on ECG signals. *Biomedical signal processing and control*, *31*, 301-308

[3] Makaryus, Amgad N., et al. "Utility of an advanced digital electronic stethoscope in the diagnosis of coronary artery disease compared with coronary computed tomographic angiography." *The American Journal of Cardiology* 111.6 (2013): 786-792.

[4] Gupta, V., Saxena, N.K., Kanungo, A., Kumar, P. and Diwania, S., 2022. PCA as an effective tool for the detection of R-peaks in an ECG signal processing. *International Journal of System Assurance Engineering and Management*, *13*(5), pp.2391-2403.

[5] Rekha, R., 2022. Medical Cyber-Physical Systems Security. In *Cyber-Physical Systems and Industry 4.0* (pp. 57-74). Apple Academic Press.

[6] Verrier, Richard L., Bruce D. Nearing, and Andre D'Avila. "Spectrum of clinical applications of interlead ECG heterogeneity assessment: From myocardial ischemia detection to sudden cardiac death risk stratification." *Annals of Noninvasive Electrocardiology* 26.6 (2021): e12894.

[7] Amal, S., Safarnejad, L., Omiye, J.A., Ghanzouri, I., Cabot, J.H. and Ross, E.G., 2022. Use of multi-modal data and machine learning to improve cardiovascular disease care. *Frontiers in cardiovascular medicine*, *9*, p.840262.

[8] Al-Hamdani, O., et al. "Multimodal biometrics based on identification and verification system." *Journal of Biometrics & Biostatistics* 4.2 (2013): 1-8.

[9] Phanphaisarn, W., et al. "Heart detection and diagnosis based on ECG and EPCG relationships." *Medical Devices: Evidence and Research* (2011): 133-144.

[10] Zhang, X., Lu, D., Hu, J., Banaei, A. and Abedi-Firouzjah, R., 2022. The role of ultrasound and mri in diagnosing of obstetrics cardiac disorders: A systematic review. *Journal of Radiation Research and Applied Sciences*, *15*(3), pp.261-269.

[11] Luo, H., Weerts, J., Bekkers, A., Achten, A., Lievens, S., Smeets, K., van Empel, V., Delhaas, T. and Prinzen, F.W., 2023. Association between phonocardiography and echocardiography in heart failure patients with preserved ejection fraction. *European Heart Journal-Digital Health*, *4*(1), pp.4-11.

[12] Zabihi, Morteza, et al. "Heart sound anomaly and quality detection using ensemble of neural networks without segmentation." *2016 computing in cardiology conference (CinC)*. IEEE, 2016.

[13] Kumar, Shruti Siva, et al. "Machine learning derived ECG risk score improves cardiovascular risk assessment in conjunction with coronary artery calcium scoring." *Frontiers in Cardiovascular Medicine* 9 (2022).

[14] Alkhodari, Mohanad, and Luay Fraiwan. "Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings." *Computer Methods and Programs in Biomedicine* 200 (2021): 105940.

[15] Li, Han, et al. "Integrating multi-domain deep features of electrocardiogram and phonocardiogram for coronary artery disease detection." *Computers in Biology and Medicine* 138 (2021): 104914.

[16] Hosseini, M.M., Mosahebeh, Z., Chakraborty, S. and Gharahbagh, A.A., 2024. Predicting the Early Detection of Breast Cancer Using Hybrid Machine Learning Systems and Thermographic Imaging. *International Journal of Imaging Systems and Technology*, *34*(6), p.e23211.

[17] Zhang, Huan, et al. "Detection of coronary artery disease using multi-modal feature fusion and hybrid feature selection." *Physiological Measurement* 41.11 (2020): 115007.

[18] Shuvo, Samiul Based et al. "CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification Using Heart Sound Recordings." *IEEE Access* 9 (2021): 36955-36967.

[19] Wang, Peng, et al. "A Wearable ECG Monitor for Deep Learning Based Real-Time Cardiovascular Disease Detection." (2022).

[20] Eleyan, Alaa and Ebrahim Alboghbaish. "Electrocardiogram Signals Classification Using Deep-Learning-Based Incorporated Convolutional Neural Network and Long Short-Term Memory Framework." *Comput.* 13 (2024): 55.

[21] Jamil, Sonain and Muhibur Rahman. "A Novel Deep-Learning-Based Framework for the Classification of Cardiac Arrhythmia." *Journal of Imaging* 8 (2022): n. pag.

[22] Pankaj et al. "Blood pressure estimation and classification using a reference signal-less photoplethysmography signal: a deep learning framework." *Physical and Engineering Sciences in Medicine* 46 (2023): 1589-1605.

[23] Arslan, Ö., 2022. Automated detection of heart valve disorders with time-frequency and deep features on PCG signals. *Biomedical Signal Processing and Control*, *78*, p.103929.

[24] Wang, Jiacheng, and Weiheng Li. "Atrial fibrillation detection and ECG classification based on CNN-BILSTM." *arXiv preprint arXiv:2011.06187* (2020).

[25] Zhang, Haobo, et al. "Co-learning-assisted progressive dense fusion network for cardiovascular disease detection using ECG and PCG signals." Expert Syst. Appl. 238 (2023): 122144

[26] Wang, Juliang, et al. "A pooling convolution model for multi-classification of ECG and PCG signals." Computer methods in biomechanics and biomedical engineering (2024): 1-14.

[27] Zhu, J., Liu, H., Liu, X., Chen, C. and Shu, M., 2025. Cardiovascular disease detection based on deep learning and multi-modal data fusion. Biomedical Signal Processing and Control, 99, p.106882.

[28] Wang, X., Li, J. and Wang, X., 2023, May. Multi-feature Fusion Network of ECG and VCG for coronary artery disease detection. In Proceedings of the 2023 4th International Conference on Computing, Networks, and Internet of Things (pp. 164-169).

[25] G. D. Clifford et al., "Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge 2016," 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 2016, pp. 609-612.

[26] Wang, S., Liu, L., Gan, L., Chen, H., Hou, X., Ding, Y., Ma, S., Zhang, D.W. and Zhou, P., 2021. Two-dimensional ferroelectric channel transistors integrating ultra-fast memory and neural computing. *Nature Communications*, *12*(1), p.53.

[27] Kim, J., & Moon, N. (2019). BiLSTM model based on multivariate time series data in multiple fields for forecasting trading area. *Journal of Ambient Intelligence and Humanized Computing*, 1-10.

[28] Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F.J.M., Ignatious, E., Shultana, S., Beeravolu, A.R. and De Boer, F., 2021. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access*, *9*, pp.19304-19326.

[29] Fatima, C.; Abdelilah, J.; Chafik, N.; Hammouch, A. Recognition of cardiac abnormalities from synchronized ECG and PCG Signals. Phys. Eng. Sci. Med. 2020, 43, 673–677.

[30] Li, H.; Wang, X.; Liu, C.; Wang, Y.; Li, P.; Tang, H.; Yao, L.; Zhang, H. Dual-Input Neural Network Integrating Feature Extraction and Deep Learning for Coronary Artery Disease Detection Using Electrocardiogram and Phonocardiogram. IEEE Access 2019,7, 146457–146469.

[31] Tschannen, M.; Kramer, T.; Marti, G.; Heinzmann, M.; Wiatowski, T. Heart sound classification using deep structured features. In Proceedings of the Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016; pp. 565–568.

[32] Li, J., Ke, L., Du, Q., Ding, X. and Chen, X., 2022. Research on the classification of ecg and PCG signals based on bilstm-googlenet-ds. *Applied Sciences*, *12*(22), p.11762.

[33] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, Castells F, Roig JM, Silva I, Johnson AEW, Syed Z, Schmidt SE, Papadaniil CD, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab MR, Sameni R, Mark RG, Clifford GD. An open access database for the evaluation of heart sound algorithms. Physiol Meas. 2016 Dec;37(12):2181-2213. doi: 10.1088/0967-3334/37/12/2181. Epub 2016 Nov 21. PMID: 27869105; PMCID: PMC7199391.

.

Appendix A:

Filter Design and Application

In this study, various Butterworth filters were utilized to denoise the ECG and PCG signals, remove baseline wander, and suppress unwanted noise, ensuring the signals were suitable for further processing. The filter designs and their application to the respective signals are described below:

1. ECG Signal Filtering

- A low-pass Butterworth filter was employed to preprocess the ECG signal. The filter was designed with a cutoff frequency of 20 Hz to remove high-frequency noise while retaining the primary frequency components of the ECG signal. The low-pass filter ensures that frequencies corresponding to the physiological ECG waveform are preserved, enabling accurate analysis and feature extraction. Additionally, a notch filter with a center frequency of 50 Hz and a quality factor (Q) of 30 was applied to eliminate power-line interference.

- The characteristics of the low-pass Butterworth filter are as follows:
- Filter Type: Low-pass Butterworth filter
- Cutoff Frequency: 20 Hz
- Impulse Response: Infinite Impulse Response (IIR)
- Stability: Conditionally stable due to its recursive nature; stability is ensured through proper design parameters.
- Difference Equation
- **Equation:**
    - **Difference Equation:**

      $y[p] = \sum_{k=0}^{M} b_k x[n-k] - \sum_{k=1}^{N} a_k y[n-k]$
    - Where b_k and a_k_ are filter coefficients, x[n] is the input signal, and y[n] is the output signal.
    - **Transfer Function:**
    - $H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1+\sum_{k=1}^{N} a_k z^{-k}}$
- **Filter Order:** For this study, the filter order (N) is 2, corresponding to a second-order Butterworth design.
- **MMM:** The number of feedforward coefficients, which in this filter equals the filter order (M=N=2). This design ensures a balance between computational efficiency and the accuracy required for medical signal processing.

The low-pass Butterworth filter effectively removes high-frequency noise while preserving the morphological characteristics of the ECG signal. This preprocessing step enhances the accuracy of subsequent feature extraction and classification algorithms

2. PCG Signal Filtering

To preprocess the PCG signal, a band-pass Butterworth filter was utilized. The filter was designed with a low cutoff frequency of 25 Hz and a high cutoff frequency of 400 Hz. This configuration allows for the removal of both low-frequency baseline wander and high-frequency noise, ensuring that the primary frequency components of the PCG signal, which are vital for cardiac analysis, are preserved.

The characteristics of the band-pass filter for the PCG signal are as follows:

- Filter Type: Band-pass Butterworth filter

- Cutoff Frequencies: 25 Hz (low cutoff), 400 Hz (high cutoff)
- Impulse Response: Infinite Impulse Response (IIR)
- Stability: Conditionally stable due to its recursive nature; stability is ensured through proper design parameters.
- Difference Equation:

$$y[p] = \sum_{k=0}^{M} b_k x[n-k] - \sum_{k=1}^{N} a_k y[n-k]$$

- Here, b_k_ and a_k_ represent the filter coefficients, x[n]x[n]x[n] is the input signal, and y[n]y[n]y[n] is the output signal.
- Transfer Function:
-

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 + \sum_{k=1}^{N} a_k z^{-k}}$$

- Filter coefficients:
  The coefficients b_k_ and _a_k are calculated based on the filter design parameters, including the cutoff frequencies, sampling rate, and filter order. These coefficients ensure the desired frequency response and stable filter operation, effectively isolating the specified band-pass range for the PCG signal.

Appendix B:

- R Peak Indicator Algorithm

Analysis and detection of R peaks in ECG signals are very vital in the assessment of heart rhythms along with other related cardiovascular diseases. Several traditional approaches to R peak detection in literature, in their essence, include a complex wavelet transform, frequency analysis, or digital filtering for locating a local maximum corresponding to an R-peak. Our R peak indicator algorithm introduces novelty, with focus on streamlining toward efficiency and specificity in the detection of R-R intervals actuated by the R peak indices.

- Algorithm Overview

The indicated R peak detection algorithm is developed to:

1. Direct identification of R peak indices with a minimum amount of preprocessing, factoring in the premise that the R peaks are the most prominent feature in the QRS complex of an ECG signal.

2. Computation of R-R intervals accurately and effectively, with each successive R peak being most crucial in the detection of certain arrhythmias, like atrial fibrillation.

Steps and Design of the Algorithm

1. Initial Signal Preprocessing:

The first step was to clean the ECG signal from high-frequency noise, masking the R peaks by a low-pass filter. In this filtered signal, all the main ECG features are preserved: P-wave, QRS complex, and T-wave; this algorithm uses only the R-peaks.

2. Differentiation and Squaring:

Differentiation, therefore, exaggerates the QRS complex in the filtered ECG by enhancing those abrupt alterations in the signal corresponding to the steep slopes of the QRS complex.

These characteristics are enhanced further by squaring the differentiated signal, a process that emphasizes large values and suppresses lower ones. Since the R peaks were already the largest features of the ECG signal, this increases their prominence even further.

3. Windowing Moving Average:

This will yield an envelope highlighting regions with R peaks, by taking the output of the moving average filter on a squared signal.

- The size of the moving average window is selected according to the heart rate range for the target population so that the peaks are well isolated and do not merge with the adjacent QRS complexes.

4. Adaptive Thresholding based R-Peak Detection:

Dynamic thresholding technique has been used for peak detection in the smoothed signal. The approach is quite different from the fixed threshold approach since it depends on the base of the signal and makes the algorithm robust irrespective of signal amplitudes.

Peaks in the signal are detected either by the zero-crossings in the differentiated signal or by maxima in the moving average-filtered signal.

5. Smoothening and the selection of R peaks It filters out the detected peaks that do not meet its pre-set criteria for R-peaks of minimum amplitude or width. This removes most false detections, especially from T-waves or other noise artifacts. The index position specifies every detected R-peak on which the R-R Intervals calculation is going to be based. 6. R-R Interval Calculation: Features of the different R-R intervals are computed from the indices of the R peaks detected as the time difference between successive R peaks. This feature, derived directly from the R peak

indices themselves, is considered critical in the determination of the heart rate variability and hence any abnormal heart rhythms.