A Machine Learning Approach for Integrating Phonocardiogram and Electrocardiogram Data for Heart Sound Detection

Thu Mains, Shruti Kshirsagar

Wichita State University, School of Computing, (Wichita, USA)

Abstract

Heart sound detection (HSD) is crucial for diagnosing cardiovascular diseases and monitoring cardiac health. While the traditional diagnostic methods often rely on either phonocardiogram (PCG) or electrocardiogram (ECG) data and often cause several performance degradations. In this work, we propose to combine the augmentation methodology with ECG and PCG fusion. Experiments are conducted with physioNet dataset used in CINC 2016 Challenges. Experimental results show the proposed method outperforming benchmark systems by providing complementary information, hence improving performance with modality fusion.

1. Introduction

Cardiovascular disease is among the biggest causes of death worldwide, emphasizing the necessity for fast diagnostics [1]. Machine learning (ML)-based methods improved heart disease screening accuracy [2]. Traditional methods observe heart functions using single-modality data like ECGs or PCGs [3], [4], [5]. ECGs show electrical issues, and PCGs detect heart noises; thus, combining them improves heart health insight [2].

Hettiarachchi [2] and Chakir [6] showed hybrid models can include ECG and PCG data. This foundation enables our study to build a multimodal ML model which employs ECG and PCG data to detect heart disease. Early cardiac condition detection and monitoring should be feasible with this method.

Following this introduction section, the paper is organized as follows: Section 2 describes the proposed methodology, including feature extraction, data augmentation, and building models. Section 3 presents experimental setup, dataset description, benchmark systems, and assessment techniques. Section 4 provides results and analysis, and Section 5 concludes the research work.

2. Proposed Methodology

This study proposes a fusion of ECG and PCG data to diagnose heart problems. Here, we explore the benefits of implementing standard feature extraction methods with data augmentation strategy. ECG and PCG provide a more complete picture of cardiac function, letting cardiologists make better judgments. In this section, we start by explaining feature extraction method for both ECG and PCG data, followed by data augmentation and proposed model.

2.1. Feature Extraction

2.1.1. Feature Extraction PCG

The signal's rhythmic pattern can be identified by extracting tempo and beat properties corresponding to heart sounds [7]. Next, we calculate ZCR and RMSE to obtain temporal features. The ZCR measures signal sign change, indicating high- and low-frequency components. Since the RMSE measures signal energy over time, it can provide heart sound intensity [8].The spectral centroid, bandwidth, and roll-off are also utilized to extract signal features [9]. To capture both temporal and spectral characteristics, Mel Frequency Cepstral Coefficients (MFCCs) [10] are used. By combining these features, our ML model can analyze both the temporal dynamics and overall spectral properties of PCG signals, emulating the detailed and multi-faceted approach used by human experts.

2.1.2. Feature Extraction ECG

The ECG signal is analyzed to extract the durations of the QRS, QT, PR, and RR intervals, and the heart rate [11]. Short-Time Fourier Transform (STFT) is employed to capture both overall and temporal ECG features [13]. STFT examines ECG signal temporal properties as well as frequency content. In addition, power spectral density (PSD) is utilized to examine the ECG signal's frequency energy distribution as well as frequency characteristics.

 Page 1 ISSN: 2325-887X DOI: 10.22489/CinC.2024.483

2.2. Data Augmentation

Data augmentation has been widely used and have shown to be useful strategies to improve system's performance in the past [14], [15] In order to expand the size and variety of the dataset, we implemented data augmentation methods on the training dataset, yielding a grand total of 2,592 files. Among these, 1,840 exhibit abnormalities, while the remaining 752 are within the normal class. In the final models, both the PCG and ECG data were subjected to augmentation methods in order to expand the dataset size. The applied augmentation techniques consist of lengthening the signals by 1.02% and compacting them by 0.98%, together with temporal shifts of 50 ms, 80 ms, 100 ms, 130 ms, and 180 ms.

2.3. Proposed Model

A convolutional neural network (CNN) and a bidirectional recurrent neural network (BiRNN) are combined in the model design [2]. The CNN retrieves local and hierarchical features, while the BiRNN collects signal temporal dependencies. Model components include:

2.3.1. Input Layers

Two separate input layers are designed for the PCG and ECG signals. Each input layer accepts a 1-dimensional signal array.

2.3.2. Convolutional Layers

PCG and ECG signals are convolutionally layered to capture local temporal patterns. Each block has a 1D convolutional layer, batch normalization, max-pooling, and dropout [2]. The first and second convolutional layers use 512 filters with 5 kernels and 1 stride. The third convolutional layer has 512 filters, kernel size 3, and stride 1. The fourth and fifth convolutional layers use 256 filters with 3 kernels and 1 stride [2]. Overfitting is prevented by adding 0.3-rate dropout layers following max-pooling layers.

2.3.3. Bidirectional LSTM Layers

Three bidirectional LSTM layers with 60, 50, and 50 units process the convolutional layer feature maps [2]. For training stability, batch normalization is used after each LSTM layer.

2.3.4. Concatenation and Dense Layers

The outputs from the LSTM layers processing the PCG and ECG signals are concatenated [2]. A dense layer with 128 units and ReLU activation is applied to the concatenated features. A dropout layer with a rate of 0.5 is included before the final output layer to prevent overfitting.

2.3.5. Output Layer

The output layer for binary classification is a dense layer with sigmoid activation for classifying normal vs abnormal heart sounds.

3. Experimental setups

3.1. Dataset description

We utilized the training-a datasets consisting of simultaneously recorded ECG and PCG signals from the PhysioNet Challenge 2016 [5]. The dataset comprises 288 anomalous and 117 normal recordings. To conduct our experiment, we partitioned the data into sets of 70%, 10%, and 20% for the purposes of training, validation, and testing.

3.2. Model Training

The model was compiled using Adam optimizer, binary cross-entropy loss, and accuracy. Experimental validation suggested 32 batches for optimal performance. A custom learning rate scheduler was used to change the learning rate during training, ranging from $1x10^{-5}$ to $1x10^{-3}$, ensuring a balanced learning rate. Overfitting was reduced by including an L2 regularization factor of 0.001 to the convolutional layers and Early Stopping strategies.

3.3. Benchmark systems

Several benchmark systems were explored in table 1. Benchmark systems are used for two reasons: (1) to find out how useful the suggested method is, and (2) to see how well it works with other methods that are already out there. Li et al. [3] used CNN with PCG data as the first method used as a baseline. Ren et al. [4] also investigated a second system that used PCG and the VGG model. Hybrid CNN was used by Hettiarachchi et al. [2] with PCG and ECG data. Along with these standards, we also look at state-of-the-art (SOTA) systems, such as the one by Chakir et al.[6] that uses SVM with PCG and ECG datasets.

3.4. Figure of merits

In order to evaluate the ML model's performance, we employed accuracy, sensitivity, and specificity [2]. Accuracy is a quantitative measure that evaluates the rate at which the ML model correctly predicts the output. Sensitivity and specificity measure how well ML classification

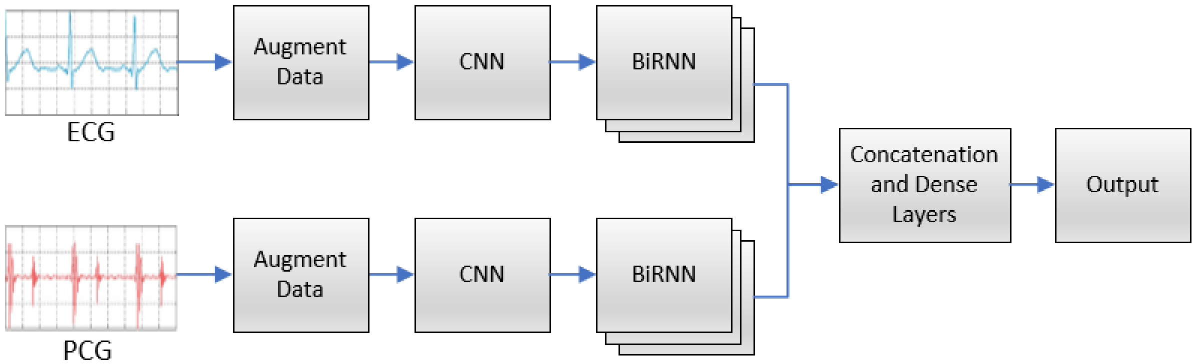

Figure 1. End-to-end pipeline for the proposed method using data augmentation and model fusion for HSD.

Table 1. Performance comparison of different state-of-the-art methods in terms of accuracy, sensitivity, and specificity

	Type	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ren et al. (VGG)	PCG	56.2	24.6	87.7
Hettiarachchi et al. (Hybrid CNN)	PCG	52.7	56.0	49.5
Hettiarachchi et al. (Hybrid CNN)	ECG	76.0	70.2	81.7
Chakir et al. (SVM)	PCG + ECG	92.5	92.3	92.9
Hettiarachchi et al. (Hybrid CNN)	PCG + ECG	90.4	94.7	75.0
Proposed approach	PCG + ECG	**98.0**	**96.4**	**98.3**

models detect positive and negative cases. A 100% sensitivity or true positive rate or recall means the model detected all positive cases without false negatives. A 100% specificity indicates that the model disregards all negative cases without false positives[2].

4. Results & Discussion

Data augmentation has shown to be extremely beneficial to improve accuracy with deep learning. Here, we explore the impact that data augmentation can have on the generalization capability of the BiRNN-CNN classifier. Overall, a total of 2595 ECG and PCG data files were used for training, hence a 7-fold increase relative to the case reported. To show the effectiveness of augmentation on HSD task, Table 2 shows the obtained results with and without data augmentation. As can be seen, the fusion of ECG and PCG features along with the augmentation method showed improvements relative to each modality alone (i.e., Table 2) by as much as 82% and 4% in terms of sensitivity and specificity. For the ECG dataset, 17, 132, and 1 % improvement has been observed with the proposed method (including fusion plus augmentation) vs. only the ECG modality in terms of accuracy, sensitivity, and specificity, respectively. Conversely, we observed similar results for PCG. Overall, we observed an average gain of 65, 33, and 28% in accuracy, sensitivity, and specificity scores with the proposed method compared to the fusion only. These findings corroborate those by Hettiarachchi [2] and Chakir [6], who showed that fusion modality achieved higher performance than single modalities. Furthermore, fusion plus augmentation improved sensitivity and specificity by 65 and 31%, respectively, compared to the fusion method alone. Overall, these findings suggest that the proposed method, combined with data augmentation during training, could also be a helpful resource for HSD tasks.

5. Conclusion

In this paper, we explore HSD strategies. Our model examines heart sound and electrical activity temporal dynamics and spectral characteristics that contribute to aid in cardiac disease detection and monitoring. We show the impact of fusion and data augmentation on heart sound performance and propose method that outperforms several benchmarks, thus highlighting the potential of the proposed system for HSD. Future work will be exploring an advanced ways to improve the model's interpretability which will help healthcare professionals grasp its decision-making process.

Table 2. Performance comparison of different modalities PhysioNet challenge dataset in terms of accuracy, sensitivity, and specificity, (Fusion) - ECG and PCG, (w aug) - with augmentation, and (w/o aug) -without augmentation.

Modality	Model	Accuracy (%)	Sensitivity (%)	Specificity (%)
PCG	CNN	89.0	71.1	95.4
PCG	LSTM	92.0	89.1	92.5
PCG	BiRNN	76.0	42.1	87.1
PCG	Hybrid	86.0	77.1	89.2
ECG	CNN	89.0	61.4	**98.3**
ECG	LSTM	82.0	39.8	97.1
ECG	BiRNN	78.0	28.9	95.3
ECG	Hybrid	80.0	28.9	97.5
Fusion(w/o aug)	CNN	73.0	69.2	75.0
Fusion (w aug)	CNN	97.0	91.6	**98.3**
Fusion(w/o aug)	LSTM	71.0	69.2	71.4
Fusion (w aug)	LSTM	93.0	87.8	94.6
Fusion(w/o aug)	BiRNN	61.0	38.5	71.4
Fusion (w aug)	BiRNN	97.0	94.0	97.5
Fusion(w/o aug)	Hybrid	68.0	46.2	78.6
Fusion (w aug)	Hybrid w/o dropout	94.0	95.4	88.0
Fusion (w aug)	Hybrid dropout 0.5	**98.0**	**99.2**	95.2
Fusion (w aug)	Hybrid dropout 0.3/0.5	**98.0**	96.4	**98.3**

References

[1] Roth, Gregory A., et al. ”Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update from the GBD 2019 Study.” Journal of the American College of Cardiology 76.25 (2020): 2982-3021.

[2] Hettiarachchi, et al. ”A Novel Transfer Learning-Based Approach for Screening Pre-existing Heart Diseases Using Synchronized Ecg Signals and Heart Sounds”. 2021 IEEE International Symposium on Circuits and Systems (ISCAS).

[3] F. Li et. al., “Classification of Heart Sounds using Convolutional Neural Network”, *Applied Sciences* (Switzerland), vol. 10, no. 11, 2020.

[4] Z. Ren, N. et.al., “Learning Image-based Representations for Heart Sound Classification”, *ACM International Conference Proceeding Series*, pp. 143–147, 2018.

[5] Clifford, et. al., “Classification of Normal/Abnormal Heart Sound Recordings: The PhysioNet/Computing in Cardiology Challenge 2016”, *Computing in Cardiology*, 37(9).

[6] Chakir, et al. ”Recognition of Cardiac Abnormalities from Synchronized ECG and PCG Signals”. Physical and Engineering Sciences in Medicine 43 (2020): 673-677.

[7] AMütze, et. al. ”The Effect of a Rhythmic Pulse on the Heart Rate: Little Evidence for Rhythmical ‘Entrainment’and ‘Synchronization’.” Musicae Scientiae 24.3 (2020): 377-400.

[8] Bachu, R. G., et al. ”Separation of Voiced and Unvoiced using Zero Crossing Rate and Energy of the Speech Signal.” American Society for Engineering Education (ASEE) Zone Conference Proceedings, 2008.

[9] Kshirsagar et al. ”Quality-aware Bag of Modulation Spectrum Features for Robust Speech Emotion Recognition.” IEEE Transactions on Affective Computing (2022): 1892-1905.

[10] M. Sahidullah,et al. ”Design, Analysis and Experimental Evaluation of Block Based Transformation in MFCC Computation for Speaker Recognition.” Speech Communication 54.4 (2012): 543-565.

[11] Sörnmo et. al. ”Electrocardiogram (ECG) Signal Processing.” Wiley Encyclopedia of Biomedical Engineering (2006).

[12] P. Bota et. al., ”BioSPPy: A Python Toolbox for Physiological Signal Processing,” SoftwareX, vol. 26, pp. 101712

[13] A. B. Jont. ”Short Time Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform”. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-25 (3): 235–238, June. 1977

[14] Kshirsagar et al. ”Cross-language Speech Emotion Recognition using Bag-of-word Representations, Domain Adaptation, and Data Augmentation.” Sensors 22.17 (2022): 6445.

[15] Kshirsagar et. al. ”Task-specific Speech Enhancement and Data Augmentation for Improved Multimodal Emotion Recognition Under Noisy Conditions.” Frontiers in Computer Science 5 (2023): 1039261.

Address for correspondence:

Thu Mains
Wichita State University, United States
tpmains@shockers.wichita.edu