Combining Deep Learning with Traditional Machine Learning to Improve Phonocardiography Classification Accuracy

M. Chowdhury[1], C. Li[2] and K. Poudel[3]

1. Computational Science, Middle Tennessee State University, Murfreesboro, Tennessee, 37132, USA
2. Department of Computer Science, Middle Tennessee State University, Murfreesboro, Tennessee, USA
3. Department of Mathematics & Computer Science, Embry-Riddle Aeronautical University, Florida, USA
mc7d@mtmail.mtsu.edu, Cen.Li@mtsu.edu, poudelk@erau.edu

Abstract— Phonocardiography (PCG) is a widely used technique to detect and diagnose cardiovascular diseases. We have combined the advantages of traditional machine learning (ML) and deep learning (DL) techniques to build deep hybrid PCG classification models. We have shown that, though DL models usually outperform ML models in classifying PCG signals, optimal classification can be achieved if we combine these two architectures to build a single PCG classification model. A Convolutional Neural Network (CNN) is used along with 7 traditional machine learning methods including Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), Decision Tree (DT), Naive Bayes (NB), Support Vector Machine (SVM), and AdaBoost (AB) to build hybrid PCG classification models. Our experimental results have shown that significant improvements in the classification accuracy can be achieved by using deep hybrid models compared to traditional machine learning models. We have also shown that some hybrid models performed better than the single deep learning model in classifying PCG signals. We have also compared the performance of the best hybrid model to 11 other PCG classification models and obtained better accuracy.

I. Introduction

Phonocardiogram (PCG) signal is the graphical representation of the activities of the human heart. Under normal conditions, the PCG signal is typically characterized by two fundamental heart sounds called the first heart sound (S1) and the second heart sound (S2). S1 and S2 result from the closure of the atrioventricular valves and semilunar valves, respectively. Systolic and diastolic intervals are also two very important parameters of PCG signal, which are used to detect cardiac changes related to high blood pressure. There are also two abnormal sound components called the third and fourth heart sounds (S3 and S4), which are the potential indicators of cardiac arrest. Table 1 shows different heart sounds and their properties. Aside from these four basic heart sounds, unusual sounds may be heard between heartbeats because of abnormal blood flow across the heart. These abnormal sounds are known as heart murmurs. Heart murmurs usually have higher frequencies compared to basic heart sounds [1].

Table 1. Properties of different heart sounds

Sound	S1	S2	S3	S4
Frequency	30-100 Hz	Above 100 Hz	20-25 Hz	Below 30 Hz

Heart murmurs may be a sign of cardiac diseases. Figure 1 shows an abnormal PCG signal. The main advantage of the PCG signal is that it can accurately identify these four basic heart sounds and murmurs. It can record heart sounds continuously and can extract cardiac information from heart sounds. This cardiac information can be analyzed to check the proper functionality of the heart [1]. Thus, it is possible to identify the presence of cardiovascular diseases in the primary stage by analyzing PCG signals. Moreover, PCG signal is used to complement ECG signal, as it is low-cost, non-invasive, and easy to measure. Hence, continuous monitoring of PCG signals is of great interest in remote health diagnostic systems to detect early-stage heart diseases and abnormal heartbeats [2][3].

In clinical environments, it is a very onerous task to analyze PCG signals by cardiologists. For this reason, automatic computer-aided diagnosis tools can be used as an alternative solution to examine PCG signals. The main objective of an automatic PCG signal analysis technique is to build a robust system that can detect abnormal PCG signals accurately and effectively. Various techniques have been reported in the literature for PCG signal classification including machine learning and deep learning algorithms. Different machine learning algorithms such as LR, RF, KNN, DT, NB, SVM, and AB have been used to classify PCG signals. However, ML techniques require manual feature engineering, which is extremely hard. The accuracy of these traditional classification methods heavily relies on features extracted from the original raw signal. Deep learning techniques do not need manual feature engineering techniques and they can extract meaningful features from the unstructured data using their own neural networks. Convolutional Neural Network (CNN) is a DL algorithm, which has shown impressive results in classifying PCG signals. We have proved that, though DL architectures are better than ML for learning highly descriptive features, they are not always optimal for the PCG classification.

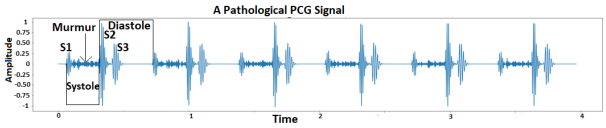

Figure 1. An abnormal PCG signal

On the other hand, while ML algorithms have limitations in learning important features on their own, they show significant improvements in PCG classification if informative learned features are fed to them. We have combined the advantages of these two architectures and developed deep hybrid models to classify PCG signals with great accuracy. First, the power spectrum of each PCG signal is calculated using Mel-scaled power spectrogram and Mel-frequency cepstral coefficients (MFCC). Further, DL architectures are used to learn distinctive features of PCG signals from their power spectrums. The output of the fully connected layer of the CNN is extracted as learned features which are sent as input to the classical supervised ML algorithms to predict the classification. Our proposed deep hybrid networks have shown significant accuracy improvements over the single DL and ML models.

II. PCG Dataset

In this paper, we have evaluated the deep hybrid classification models on the well-known 2016 PhysioNet Computing in Cardiology Challenge database [4]. This is the reference dataset that is most widely used to evaluate the performance of PCG classification methods. It contains 3240 unique PCG signals sampled at 2000 hz, with 5 to 120 seconds durations. Heart sound recordings from this database are divided into normal and abnormal classes. Many of the recordings were corrupted by different kinds of noise. The presence of noise and unnecessary information makes it difficult or even impossible to classify some PCG signals. We have used the 10-fold cross validation technique to test the performance of deep hybrid models.

III. Methods

In this section, we present the proposed methodology to derive deep hybrid models for the PCG classification. First, an extensive supervised fine-tuning step is carried out to derive the optimal hyperparameters from the raw signals. Then, these derived parameters are fed into an architecture that integrates a deep CNN architecture with the traditional machine learning methods.

III-A. Preprocessing

Raw PCG signals typically come with various noise and redundant information. Hence, preprocessing is an essential and crucial step to extract meaningful parameters from each PCG signal and it significantly improves the classification accuracy [2][3]. Mel-scaled power spectrogram and the MFCC have been proven to be very efficient to discriminate between two different acoustic signals as they mimic the functionality of human ears [1][5]. We have used these two methods to extract meaningful parameters from each PCG signal. The steps are as follows:

1) Dividing the PCG signal equally into several small frames using the hamming window.
2) Calculating the Fourier Transform of each frame to convert time domain signal to frequency domain.
3) Calculating the power spectrum of each frame and apply Mel-Scale filter banks to that.
4) Summing the energy in each filter and taking the logarithm of that to calculate the Mel-scaled power spectrogram.
5) MFCC of the signal is calculated by taking the Discrete Cosine Transform (DCT) of Mel-scaled power spectrogram.

Figure 2 shows the steps to obtain these two features from raw PCG signals. For each of the PCG signals, a total of 40 features are extracted using these two methods. This is the compact representation of original raw PCG signals. In signal classification, it is desirable to have a smaller data, since larger raw data often leads to lower classification accuracy. Figure 3 and Figure 4 show the MFCC and Mel-scaled power spectrogram of PCG signals, respectively.

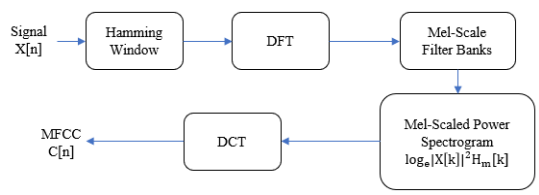

Figure 2. Steps to extract mel-scaled power spectrogram and MFCC from raw PCG signals

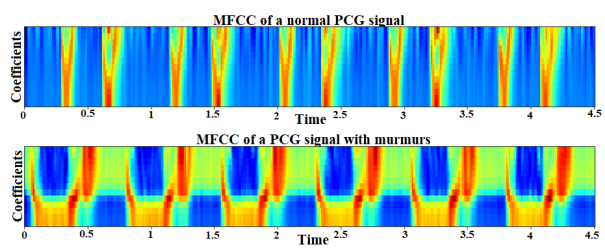

Figure 3. MFCC of a normal and an abnormal PCG signal

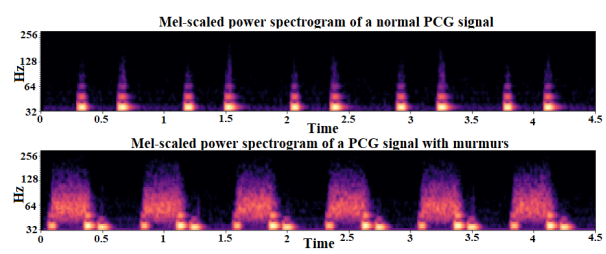

Figure 4. Mel-scaled power spectrogram of a normal and an abnormal PCG signal

III-B. Proposed CNN-ML Model

Figure 5 illustrates the design of the proposed deep hybrid architecture. It has four 1D convolution layers with 256, 512, 1,024, and 2048 filters, respectively. All the layers have the same kernel size of 2. Each convolution layer uses a Rectified Linear Unit (ReLU) activation function, and each max pooling layer is of size 2. A dropout layer with 10% dropping rate is connected with each convolution layer to handle overfitting. After the convolution and max pooling, the learned features are flattened to one long vector and are passed to a fully connected layer with 512 filters. The fully connected layer also used the ReLU activation function and 20% dropout rate to reduce overfitting problems. The fully connected layer works as a buffer between the learned features and the output. The cost function is minimized by using the Adam optimizer. The number of training epochs, batch size, and the learning rate are set to 100, 64, and 0.0001, respectively. After optimizing and training the CNN, the learned features from the fully connected layer are passed on to the ML classifiers for the final prediction task.

III-C. Performance Evaluation

We have compared the performance of the learning models by investigating their confusion matrix. We have used sensitivity, specificity, and accuracy for evaluating the classification models. The formulas to calculate these parameters are given below:

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (1)$$

$$Specificity = \frac{True\ Negative}{False\ Positive + True\ Negative} \quad (2)$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{All\ PCG\ Signals} \quad (3)$$

where True Positive (TP) is the number of abnormal PCG signal correctly identified as abnormal, True Negative (TN) is the number of normal PCG signal correctly identified as normal, False Positive (FP) is the number of normal PCG signal incorrectly identified as abnormal, and False Negative (FN) is the number of abnormal PCG signal incorrectly identified as normal signal.

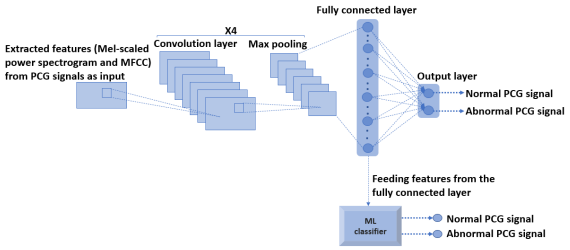

Figure 5. The proposed deep hybrid model architecture

IV. Results and Discussions

In this section, we have discussed the performance of the deep hybrid PCG classification approach. We have evaluated the effectiveness of the proposed deep hybrid models with other ML and DL algorithms in terms of sensitivity, specificity, and accuracy. First, Mel-scaled power spectrogram and MFCC are used to convert each of the PCG signals into a compact form of only 40 features, which are then fed into the classification models as input. Afterwards, DL and 7 traditional ML classifiers are used separately to classify PCG signals. Finally, our proposed hybrid models are trained to produce the classification models and to perform the classification task. The performance of the proposed hybrid models and other traditional ML and DL models with separate implementations are shown in Table 2.

As shown in Table 2, among all the ML classifiers, LR classifier performed worst with 67.50% accuracy, and best result is achieved by the RF classifier with 83.33% accuracy. The classifier learned with a single CNN model performed better than all other ML models with a classification accuracy of 92.00%. Our proposed deep hybrid CNN-ML methods were able to improve the accuracy obtained from a single CNN model and 7 ML models which were implemented separately.

The best classification accuracy of 94.30% is achieved by CNN-RF models followed by CNN-LR (92.70%), CNN-AB (92.62%), CNN-DT (92.37%), and CNN-SVM (92.10%) models. Though CNN-KNN and CNN-NB hybrid models showed around 92.00% accuracy, which is the same as the single CNN model, they showed significant improvement in the sensitivity of the model. The sensitivity of CNN-KNN and CNN-NB models are 95.33% and 97.00%, respectively. The sensitivity of both models is higher than that of the single CNN model (87.69%). The specificity of the single CNN model is 93.44%. For the specificity, there is a 3.00% reduction from the single CNN to the CNN-KNN and CNN-NB models, respectively.

Table 2. Comparison of the proposed deep hybrid models with ML and DL classification models implemented separately

Model	Sensitivity	Specificity	Accuracy
CNN	87.69%	93.44%	92.00%
LR	72.33%	66.25%	67.50%
CNN-LR	94.58%	92.27%	92.70%
RF	75.48%	85.35%	83.33%
CNN-RF	92.03%	94.83%	94.30%
KNN	75.19%	80.12%	79.10%
CNN-KNN	95.33%	90.79%	91.72%
DT	78.19%	76.42%	76.79%
CNN-DT	92.03%	92.47%	92.37%
NB	70.22%	67.30%	75.33%
CNN-NB	97.00%	90.17%	91.60%
SVM	77.29%	80.15%	79.56%
CNN-SVM	93.30%	91.80%	92.10%
AB	74.14%	81.13%	79.70%
CNN-AB	94.14%	92.23%	92.62%

In Table 3, the performance of the CNN-RF classification model is compared with 11 previous PCG classification models. As shown in Table 3, in past work, Sotaquirá et al. [6] achieved the highest accuracy of 92.60% to classify PCG signals. They used probability comparison and DNN algorithms to build their model. Langley et al. [7] used classification tree model and got the lowest classification accuracy of 79.00%. Potes et al. [8] proposed a hybrid model using AdaBoost-CNN algorithms with 86.02% accuracy and achieved 1st place in the 2016 PhysioNet Computing in Cardiology Challenge [4]. Nassralla et al. [9] also combined RF and DNN models for the PCG classification and got a good accuracy of 92.00%. Whitaker et al. [10] and Tang et al. [11] both used single implementation of ML algorithm (SVM) and got accuracies of 89.26% and 88.00%, respectively. Singh et al. [12] proposed two methods using KNN and ensemble of classifiers and obtained 90.00% and 92.47% accuracy, respectively. Krishnan et al. [13] divided each PCG signal into different segments and sent those segments to a feed forward DNN to accurately classify 85.65% PCG signals. Our proposed best hybrid model exceeded these models by achieving an accuracy of 94.30%.

Our proposed deep hybrid models are less computationally expensive and have less time complexity compared to traditional DL algorithms. The final classification layer of a DL model usually results in overfitting when the model is fed with unstructured or less data. This overfitting problem increases the time and computational complexity of traditional DL models, which is not present in traditional ML algorithms. In our proposed deep hybrid models, fully connected neural networks in the DL model are followed by the ML models. Thus, our proposed deep hybrid models are faster and do not require additional time for processing compared to traditional standalone DL models. In addition, our proposed deep hybrid models remove the need for feature engineering techniques on which all traditional ML algorithms are dependent. This automatic classification process can better help doctors and cardiologists to detect cardiac abnormalities and irregular heartbeats in the initial stage.

Table 3. Comparing the performance of the proposed CNN-RF deep hybrid model with previous models

Model	Sensitivity	Specificity	Accuracy
Potes et al., (2016) [8]	94.24%	77.81%	86.02%
Nassralla et al., (2017) [9]	78.00%	98.00%	92.00%
Whitaker et al., (2017) [10]	90.00%	88.45%	89.26%
Langley et al., (2017) [7]	77.00%	80.00%	79.00%
Han et al., (2018) [14]	98.33%	84.67%	91.50%
Tang et al., (2018) [11]	88.00%	87.00%	88.00%
Sotaquirá et al., (2018) [6]	91.30%	93.80%	92.60%
Singh et al., (2019) [12]	93.00%	90.00%	90.00%
Sing et al., (2020) [12]	94.08%	91.95%	92.47%
Nogueira et al., (2019) [15]	90.45%	85.25%	87.85%
Krishnan et al., (2020) [13]	86.73%	84.75%	85.65%
CNN-RF Model	**92.03%**	**94.83%**	**94.30%**

V. Conclusions

The main goal of this research is to develop algorithms to accurately detect cardiac abnormalities in the primary stage while continuous monitoring of the heart is conducted. We have combined the application of machine learning and deep learning approaches in PCG signals to detect abnormalities in the heart during the early-stage without the need for any doctor or cardiologist. Our proposed deep hybrid models took advantage of deep learning techniques to extract important features from unstructured data without the need for feature engineering. These models also overcome the overfitting problem of deep learning by applying machine learning techniques to classify PCG signals accurately with less computational complexity. Based on this result, it becomes realistic to develop a wearable sensor to detect heart diseases in their early-stage, which can cut the death rate due to cardiovascular diseases. Later, other advanced deep learning models may be implemented separately and combined with machine learning models to further increase the classification accuracy.

References

[1] T. H. Chowdhury, K. N. Poudel, and Y. Hu, " Time-Frequency Analysis, Denoising, Compression, Segmentation, and Classification of PCG Signals," *IEEE Access*, vol. 08, pp. 160 882–160 890, 2020.

[2] M. Chowdhury, K. Poudel, and Y. Hu, "Phonocardiography Data Compression Using Discrete Wavelet Transform," *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, 2018, pp. 1–4.

[3] M. Chowdhury, K. Poudel, and Y. Hu, "Automatic Phonocardiography Analysis using Discrete Wavelet Transform," *3rd International Conference on Vision, Image, and Signal Processing*, 2019, pp. 1–4.

[4] C. Liu, D. Springer, Q. Li, B. Moody, R. Juan, F. Chorro, F. Castells, J. Roig, I. Silva, A. Johnson, Z. Syed, S. Schmidt, C. Papadaniil, L. Hadjileontiadis, H. Naseri, A. Moukadem, A. Dieterlen, C. Brandt, H. Tang, M. Samieinasab, M. R. Samieinasab, R. Sameni, R. G. Mark, and G. D. Clifford, "An open access database for the evaluation of heart sound algorithms," *Physiological Measurement*, vol. 37, pp. 2181–2213, 2016.

[5] M. Chowdhury, K. Poudel, and Y. Hu, "Detecting Abnormal PCG Signals and Extracting Cardiac Information Employing Deep Learning and the Shannon Energy Envelope," *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, 2020, pp. 1–4.

[6] M. Sotaquirá, D. Alvear, and M. Mondragón, "Phonocardiogram classification using deep neural networks and weighted probability comparisons," *Journal of Medical Engineering Technology*, vol. 42, pp. 510–517, 2018.

[7] P. Langley and A. Murray, "Heart sound classification from unsegmented phonocardiograms," *Physiological Measurement*, vol. 38, 2017.

[8] C. Potes, S. Parvaneh, A. Rahman, and B. Conroye, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," *Proceedings of the Computing in Cardiology*, 2016, pp. 621–624.

[9] M. Nassralla, Z. E. Zein, and H. Hajj, "Classification of normal and abnormal heart sounds," *Fourth International Conference on Advances in Biomedical Engineering*, 2017, pp. 1–4.

[10] B. M. Whitaker, P. B. Suresha, C. Liu, G. D. Clifford, and D. V. Anderson, "Classification of normal and abnormal heart sounds," *Physiological Measurement*, vol. 38, 2017.

[11] H. Tang, Z. Dai, Y. Jiang, T. Li, and C. Liu, "PCG Classification Using Multidomain Features and SVM Classifier," *BioMed Research International*, pp. 1–4, 2018.

[12] S. A. Singh and S. Majumder, "Short unsegmented PCG classification based on ensemble classifier," *Turkish Journal of Electrical Engineering and Computer Sciences*, pp. 875–889, 2020.

[13] P. T. Krishnan, P. Balasubramanian, and S. umapathy, "Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network," *Physical and Engineering Sciences in Medicine*, vol. 43, 2020.

[14] W. Han, Z. Yang, J. Lu, and S. Xie, "Supervised threshold-based heart sound classification algorithm," *Physiological Measurement*, vol. 39, 2018.

[15] D. M. Nogueira, M. N. Zarmehri, C. A. Ferreira, A. M. Jorge, and L. Antunes, "Heart Sounds Classification Using Images from Wavelet Transformation," *Progress in Artificial Intelligence*, 2019.