Physical and Engineering Sciences in Medicine
https://doi.org/10.1007/s13246-020-00851-w

SCIENTIFIC PAPER

Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network

Palani Thanaraj Krishnan[1] · Parvathavarthini Balasubramanian[2] · Snekhalatha Umapathy[3]

Received: 29 November 2019 / Accepted: 7 February 2020

Abstract
Given the patient to doctor ratio of 50,000:1 in low income and middle-income countries, there is a need for automated heart sound classification system that can screen the Phonocardiogram (PCG) records in real-time. This paper proposes deep neural network architectures such as a one-dimensional convolutional neural network (1D-CNN) and Feed-forward Neural Network (F-NN) for the classification of unsegmented phonocardiogram (PCG) signal. The research paper aims to automate the feature engineering and feature selection process used in the analysis of the PCG signal. The original PCG signal is down-sampled at 500 Hz. Then they are divided into smaller time segments of 6 s epochs. Savitzky–Golay filter is used to suppress the high-frequency noises in the signal by data point smoothening. The processed data was then provided as an input to the proposed deep neural network (DNN) architectures. 1081 PCG records were used for training and validating the proposed DNN models. The Feed-forward Neural Network model with five hidden layers provided a better overall accuracy of 0.8565 with a sensitivity of 0.8673, and specificity of 0.8475. The balanced accuracy of the model was found to be 0.8574. The performance of the model was also studied using the Receiver Operating Characteristic (ROC) plot, which produced an Area Under the Curve (AUC) value of 0.857. The classification accuracy of the proposed models was compared to the related works on PCG signal analysis for cardiovascular disease detection. The DNN models studied in this study provided comparable performance in heart sound classification without the requirement of feature engineering and segmentation of heart sound signals.

Keywords Phonocardiogram · Heart sound · Convolutional neural network · Deep learning · Feature extraction · Time series classification · Feedforward neural network

Introduction

Heart disease or cardiovascular disease (CVD) is the leading cause of death in the world. Around 17.9 million people die prematurely because of heart disorders. Much of the deaths occur in low and middle-income countries [1]. CVD relates to a collection of heart or blood vessel malfunction such as coronary heart disease, congenital heart disease, and peripheral arterial disease. Diagnosis of these cardiac problems often involves listening to the heart sounds through the phenomenon of cardiac auscultation [2]. One of the common methods to perform cardiac auscultation electronically is through Phonocardiograph.

Phonocardiograph provides a way to record, store, and analyze the heart sounds for a thorough medical examination. Given the patient to doctor ratio of 50,000:1 in low income and middle-income countries, there is a need for automated heart sound classification system that can screen

✉ Snekhalatha Umapathy
sneha_samuma@yahoo.co.in

Palani Thanaraj Krishnan
palanithanaraj.k@gmail.com

Parvathavarthini Balasubramanian
parvathavarthini@gmail.com

1 Department of Electronics and Instrumentation Engineering, St. Joseph's College of Engineering, Anna University, Chennai, Tamil Nadu, India

2 Department of Computer Science and Engineering, St. Joseph's College of Engineering, Anna University, Chennai, Tamil Nadu, India

3 Department of Biomedical Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu 603203, India

Published online: 11 February 2020

the PCG records in real time [3]. These automated systems can be used as an aiding tool for the physician as it reduces the time to analyze the data. The automated heart sound classification system is one of the widely researched topics in computer-aided diagnosis of heart disease over many decades [4].

Many methods and techniques have been developed to study the PCG signal and diagnose heart problems. A plethora of these methods involves feature engineering that uses pre-defined features for the classification of heart sounds. Choi and Jiang proposed two spectral features such as the F_{max}, F_{width} to characterize the heart murmurs [5]. F_{max} is the maximum peak identified in the Power Spectral Density (PSD) function. F_{width} is the frequency bandwidth between the cross-over points in the PSD given by the user-defined threshold value. A set of support vector machines (SVM) were trained to identify normal and abnormal sounds. They reported accuracy between 81.2–99.6%, which was dependent on the threshold value set between 10–50%. The selection of an optimal threshold value is the key to attain high classification accuracy in this study. Redlarski et al. proposed a modified linear predictive coding (LPC) algorithm for the classification of heart sounds. They devised a variable window technique to extract the filter coefficients of LPC. A 24th order *filter coefficients* were taken as features to classify different heart sound classes. They reported an average accuracy of 94.48% and 91.26% for the SVM classifier using different fitness functions used in the Cuckoo search algorithm. The major limitation of their study is that the method was validated only with six records for each heart sound type [4].

Zheng et al. proposed a set of *energy fraction* and entropy features to characterize heart sounds and heart murmurs. They decomposed the PCG signal into wavelet packets with a scale factor of 5. Then the energy fraction and entropy features were computed for the fundamental heart sounds and heart murmurs. SVM classifier with different kernel functions was trained on the extracted features. They have reported a classification accuracy of 97.17% and a sensitivity of 93.48% with a proprietary dataset used in their study. High sensitivity reported in their work is also because of the high prevalence of abnormal sounds in the dataset, which is not a general case in clinical settings [6]. Zhang et al. proposed a *scaled spectrogram* feature to classify fundamental heart sound and murmurs. The most relevant time–frequency components of the spectrogram are used to train the SVM classifier. They have reported a precision of 0.60 and 0.76 for the normal sound of the two Pascal challenge datasets. Moreover, they have reported a heart murmur detection precision of 0.91and 0.65 for the two Pascal datasets. However, their method is a multi-step procedure that requires proper selection of parameters such as wavelet filter coefficients and a scaling factor for the spectrogram [7]. Hamidi et al. used *fractal dimension* and *Mel-frequency cepstrum coefficients* (MFCC) to identify abnormal heart sounds (murmurs) in the PCG signal. Fractal dimension and MFCC were stacked together for feature classification while a k-Nearest Neighbour (k-NN) classifier was used for classifying the heart sounds. They have used the Physionet challenge 2016 and Pascal dataset for validation. They have reported an overall classification accuracy for three datasets as 92%, 81%, and 98%, respectively [8].

Many of the studies discussed thus far made use of disparate PCG datasets for validation of heart sound classification methods. The inadequate number of open PCG datasets for widespread clinical validations is a major concern in computer-aided cardiac diagnosis. However, with the efforts of Liu and Springer et al. Physionet Challenge 2016 PCG open dataset was made available to the research community to clinical validations [3, 9]. Many research teams participated in the Physionet challenge 2016, and provided solutions for the classification of normal/abnormal heart sounds.

Whitaker et al. introduced matrix norm sparse coding and combined it with time-domain features to classify the PCG signals. They have combined the *sparse coding coefficients* with many other *time-domain features* and used it for training a group of SVM classifiers. They have reported a balanced accuracy (MAcc, Mean of sensitivity and specificity) of 0.8926 [10]. Plesinger et al. divided the PCG signal into different frequency bands, starting from low frequency (15–90 Hz) to ultra-high frequency (400–800 Hz). They extracted *amplitude envelopes* for these frequency bands, which summed to a total of 53 features. They introduced a probability assessment technique to perform feature selection and classification. They reported a MAcc test score of 0.853 with the challenge dataset [11]. Maknickas et al. proposed *Mel-frequency spectral coefficients* (MFSC) and used a deep convolutional neural network (CNN) for PCG signal classification. A three-dimensional image array was created with MFSC and used to train the CNN. They used 520 normal images and 567 abnormal images for training. They reported the best score of 86.02% for the hidden test dataset [12].

Langley and Murray used *spectral amplitude* and *wavelet entropy* to classify the unsegmented PCG signals. They have combined the two features and trained a decision tree classifier on the extracted features. They reported a MAcc of 80% for low noise PCG signals [13]. Kay and Agarwal proposed a drop-connected neural network (NN) for heart sound classification, in which *time–frequency* features were extracted from continuous wavelet transform (CWT), Mel-frequency cepstrum coefficients (*MFCC*), *inter-beat,* and complexity features such as *spectral entropy*. A total of 675 features were obtained to construct the feature vector. Their method produced a classification accuracy of 85.2% with the test dataset [14]. Homsi and Warrick proposed an ensemble of feature classifiers for classifying the heart sounds. They have created a feature space with 131 features obtained from *time, frequency* and *complexity* analysis. The two sets of ensemble classifiers were trained on the reduced feature set. The base classifier that was used in the work consisted of a cost-sensitive classifier (CSC), logit boost (LB), and random forest (RF). Their work reported an overall accuracy of 80.1% for the test dataset [15].

Thus far, it would be observed that the methods used in these studies are based on a set of handcrafted features. Although their results provided high accuracy, often the methods lead to subjective variations and bias due to feature engineering. To overcome this problem and to automate the feature extraction and feature selection processes, this study proposes a Deep Neural Network (DNN) based heart sound classification system in which a modified structure of CNN (1-dimensional CNN (1D-CNN)) and Feed-forward Neural Network (F-NN) would be used for heart sound classification.

Methods

Dataset

This study uses Physionet Challenge 2016 dataset for training and validation of the system for heart sound classification. This database consists of PCG records of both healthy and cardiovascular disease patients. The cardiac conditions in the dataset include valve defects such as mitral regurgitation, aortic stenosis, and valvular surgery. The PCG records containing these conditions are commonly grouped as abnormal heart sound. The available PCG records are named from 'a' to 'f'. The dataset is a collection of records obtained from different medical institutions with different recording settings. The duration of the records ranges from 5 to 120 s. PCG records that are less than 6 s were not considered in this study. Some of the PCG recording, namely dataset 'e' with high background noise, was also not considered for the analysis in this research effort. The total numbers of normal and abnormal heart sounds considered in this study are shown in Table 1. More details about the challenge 2016 dataset can be obtained from Physionet [3, 16].

Table 1 Description of the PCG data records used for the analysis of heart sound classification

Dataset name	Number of records	Normal heart sound	Abnormal heart sound
Training-a	409	117	292
Training-b	477	379	98
Training-c	31	7	24
Training-d	50	22	28
Training-f	114	80	34
Total	1081	605	476

Data pre-processing

The PCG records in the challenge 2016 dataset were recorded with a sampling frequency of 2000 Hz. The data records were down-sampled to a frequency of 500 Hz. To perform a time series classification of the signal, the data records are then divided into smaller time epochs of 6 s. This time duration is selected based on the methods used in recent studies, which suggests that at least 5 s data record is required for reliable detection of heart abnormality [3].

The data segment that is at the central portion of the data record is selected for this purpose. Each time segment with its corresponding class label was taken as a data sample to the proposed DNN models. The high-frequency noise component in the signal was reduced by using a digital filter. In this study, Savitzky–Golay filter was used, which performs a data point smoothening to remove the noise component in the signal. Figure 1 shows the 6 s time epoch of the filtered PCG signal of the patient records with normal and abnormal heart sounds.

Neural network architecture

In this study, the heart sound classification problem is viewed as a sequence classification of a time series data using DNN. It is a binary classification problem wherein PCG data records are classified as either normal or abnormal heart sounds. DNN can classify time series or any sequential data by many approaches. One of the standard methods is using Recurrent Neural Network (RNN). Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) are types of RNN that effectively handles the vanishing gradient and exploding gradient problem in RNN. RNN has been extensively used in natural language processing (NLP), such as speech processing and language translation. The main design constraint in using RNN is the time required to train the network, given its complex architecture [17]. An alternative approach to using RNN is converting the time series data to images and then using a convolution neural network (CNN) for classifying the signal. This approach uses the capability of CNN to identify and generalize the features from the images. The applications of CNN in computer vision problem is widely accepted and extensively used in various fields such as autonomous vehicles, medical imaging, and industrial robotics. However, this approach requires to use a certain type of transformation for converting to images such as time–frequency spectrum, spectrogram, MFCC, Gramian angular summation field, and Markov Transition Field [18]. The inherent problem in this approach of imaging time series

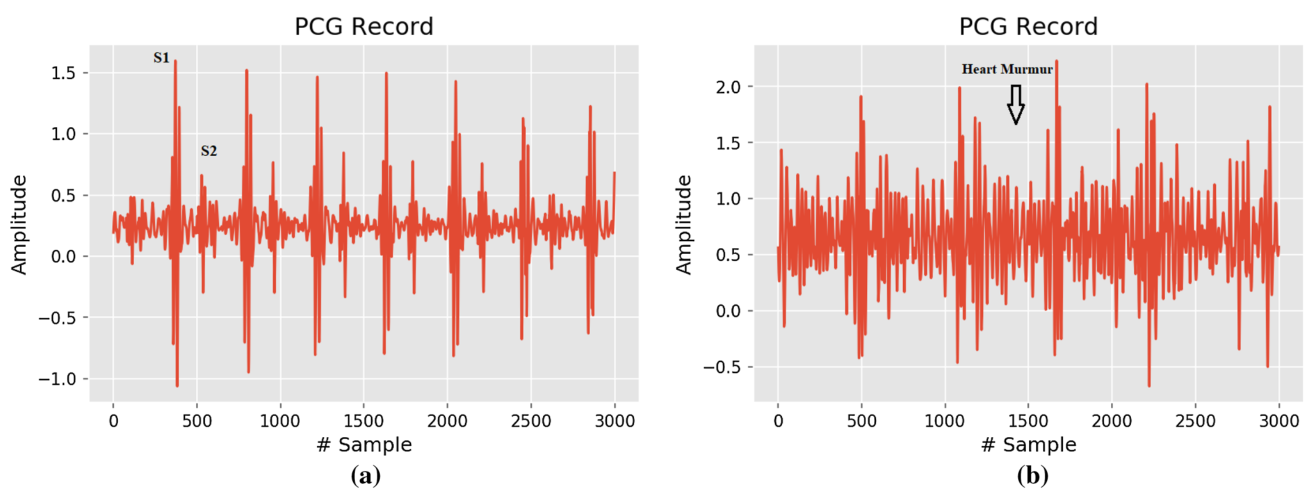

Fig. 1 PCG recording of: **a** a patient record with normal heart sounds ‘S1’ and ‘S2’; **b** patient record with abnormal heart sounds (Systolic Heart Murmurs)

is that the images depend on the underlying transformation and lead to subjective differences.

Recently, a one-dimensional convolution network (1D-CNN) was introduced for temporal data classification such that the approach employed does not require the pre-processing step of imaging time-series data. 1D-CNN is actively researched and also validated in biomedical signal processing. Acharya et al. used an 11 layer deep convolutional neural network based on 1D-CNN to detect coronary heart disease from the ECG signal and they have reported a top classification accuracy of 95.11% for the classification of normal/abnormal ECG segments [19]. Moreover, the same group has used DNN based on 1D-CNN to classify the cardiac heartbeats and validated their work with the Physiobank MIT-BIH arrhythmia dataset; they reported an accuracy of 94.03% in classifying the heart beats [20].

Given the advantages of using 1D-CNN for time series classification, this study evaluates different deep neural network architectures of heart sound classification. Table 2 describes the four network architectures designed in this study while the three networks Net-1, Net-2, and Net-3 use 1D-CNN; and Net-4, uses a Feedforward Neural network for abnormal heart sound detection.

Input layer

Temporal data classification of heart sound has been carried out in this work using different deep neural networks and the details of the proposed architecture are provided in Table 2. The filtered PCG signal, which is divided into smaller time segments of 6 s, is taken as the input layer of the networks. Thus, the time series data of heart sound

Table 2 Description of the network architecture evaluated for heart sound classification

Network architecture			
Net-1	Net-2	Net-3	Net-4
Input layer [1 × 3000] PCG time sequence			
Embedding layer [3000 × 200]	Embedding layer [3000 × 200]	Embedding layer [3000 × 200]	–
Convolution Layer-1D [1 × 100, 64, ReLU]	Convolution layer-1D [1 × 10, 32, ReLU] [1 × 10, 64, ReLU]	Convolution layer-1D [1 × 5, 32, ReLU]	–
MaxPooling layer flatten layer	GlobalAveragePooling layer	Flatten	–
Dense Layer [1 × 64, Sigmoid] [1 × 2, SoftMax]	Dense Layer [1 × 2, SoftMax]	Dense layer [1 × 128, ReLU] [1 × 64, ReLU] [1 × 2, SoftMax]	Dense layer [1 × 1024, ReLU] [1 × 512, ReLU] [1 × 128, ReLU] [1 × 64, Sigmoid] [1 × 2, SoftMax]
Output layer [1 × 2] [normal, abnormal]			

Specification of convolution layer: [Dimension of the Kernel, # Filter, Activation function]

Specification of dense layer: [Dimension of the Feedforward Neural Net, Activation function] can be represented as, $X = \{x_0, x_1 \cdots x_{n-1}\}$, here '$n$' is the data length for the sample input ($n = 3000$).

Embedding layer

One key point we have to notice is that the dataset contains PCG data, which are recorded in different clinical settings such that the amplitude values are not scaled to a common reference range. Thus, the dataset is normalized by dividing the input samples by the standard deviation of the data samples. Moreover, the proposed architecture uses an Embedding Layer to transform the input samples into a dense vector which is optimized with the model architecture. This allows the data sample to be referenced in a common vector space. This is similar to a word embedding commonly used in text processing. The Embedding layer used in the architecture can be represented as, $(e_{0,0}, e_{0,1} \cdots e_{0,p-1}), (e_{1,0}, e_{1,1} \cdots e_{1,p-1}) \cdots, (e_{n-1,0}, e_{n-1,1} \cdots e_{n-1,p-1})$ where 'p' is the dimension of the dense vector ($p = 200$). The embedded layer creates an index table of the data points and maps the integers to the dense vector, thus creating a common reference space of the data samples.

1D-convolutional neural network

The feature extraction from the heart sound time series data is obtained by 1D-convolutional layers. 1D-CNN consists of different filters that are randomly initialized for extracting the features from the input time sequence. The parameters of this layer consist of the number of filters, filter length, and stride (equals 1). Three networks are designed with different configurations of the 1D-CNN layers. Net-1 uses one convolutional layer with a filter length of 100 and with 64 convolutional filters. Net-2 uses two convolutional layers with a filter length of 10 and uses 32 and 64 filters, respectively. Net-3 uses one convolutional layer with a filter length of 5 and 32 filters. All the convolutional layers use the rectified linear unit (ReLU) as the activation function and the entire process of feature extraction can be given as:

$$c_{j,i} = \begin{cases} \overbrace{b_i + \sum_i k_i * X_i}^{z}, & z > 0 \\ 0 & z < 0 \end{cases} \tag{1}$$

where, 'c' represents the feature maps of the convolutional layers, 'b' represents the bias, 'k' represents the kernel or filter weight, and 'X' represents the data sample.

Max-pooling layer (MPL)

The feature maps obtained from the convolutional layers are down-sampled based on the maximum or average value of the feature maps. Here the pool length of the kernel function is taken to be two, and the maximum of the feature values is taken as the output of this layer. The Max-Pooling layer is used in Net-1 architecture.

Global average pooling layer (GAP)

Global Average Pooling is similar to the Max pooling layer as both involved in down-sampling of the feature maps obtained from the previous convolutional layers. However, GAP determines the average of the feature maps while MPL determines maximum of the feature maps. The spatial dimension of the feature maps with the dimension $h \times d$ is reduced to $1 \times d$ using the Global Average Pooling layer, and the average of the feature maps is computed. Both Max Pooling and Global Pooling layers minimize the over-fitting of the model by reducing the number of parameters through the down-sampling procedure. Net-2 uses the Global Average pooling layer.

Flatten layer

Flatten Layer converts the feature maps of dimension $h \times d$ into the vector dimension of $1 \times h \times d$. It is usually done when connecting the convolutional layer or pooling layer output to a fully connected multilayer perceptron layer. Flatten Layer has been employed in Net-1 and Net-3 architectures.

Feedforward neural network

The high-level feature inference is obtained by using the dense layers. Dense layers consist of a multilayer perceptron neural network that involves extracting the high-level features obtained from the convolutional neural network. Net-1 uses two dense layers with 64 neurons and 2 neurons with sigmoid and softmax activations, respectively. Net-2 uses one dense layer of 2 neurons with softmax activation. Net-3 uses three dense layers of 128, 64, and 2 neurons with Relu, and softmax activations, respectively. Net-4 consists only of MLP layer with 5 dense layers with the first layer with 1024 neurons with Relu activation, second dense layer with 512 neurons and Relu activation function, third layer with 128 neurons with Relu activation, fourth layer with 64 neurons with sigmoid activation and fifth layer with 2 neurons and softmax activation layer.

Output layer

Feature Classification is done in the output layer such that the target output is represented as $Y=\{y_0\}$, where $y_{0=-1}$ for normal heart sound and $y_0=1$ for abnormal heart sound. This layer provides the classification output of the input data sample with the prediction probability. This can be represented as $P_{Y|X}(y_j) = \frac{e^{y_j}}{\sum_j e^{y_j}}$, where *'j'* is two as it is the binary classification of heart sounds. The output neuron with the highest probability score is taken as the classification output of the given data sample. If the softmax probability score is high for class '-1', then the time series data is classified as normal heart sound else the time series data is labeled as abnormal heart sound. Figure 2 provides a pictorial representation of the deep neural network layers with the architecture details of the models as developed in this study.

Training and validation

The heart sound data used in this study is divided into two non-overlapping datasets, namely the training set (#765), and the validation set (#216). The training of the proposed architecture for heart sound classification is based on supervised learning. The training pair of the *'m'* samples in the dataset is given as $(X_0, Y_0),\ (X_1, Y_1) \cdots (X_m, Y_m)$. Each input data sample has a target output *'Y_i'* and the kernel weights are updated by minimization of the binary cross-entropy or log-loss function. As the heart sound classification is stated here as a binary classification problem, the number of classes equals 2, and the binary cross-entropy can be calculated as:

$$J = -\frac{1}{N}\sum_i y_i \cdot \log\left(P(y_i)\right) + (1-y_i)\cdot \log\left(1-P(y_i)\right) \quad (2)$$

Where, *'J'* is the loss function, which is reduced during each training epoch. The number of training epochs is set as 20, and the batch size is set as 16. Adam optimizer is used to update the kernel weights by reducing the loss function. The parameters of the Adam optimizer are as follows: (a) learning rate = 0.001, (b) exponential decay rate for the first moment, $\beta_1 = 0.9$, (c) exponential decay rate for the second moment, $\beta_2 = 0.999$. The deep learning architecture is run on a Google Cloud infrastructure with a virtual machine (VM) instance of type 'n1-standard-4' with 4 vCPUs and RAM size of 15 GB. Moreover, to accelerate the deep learning process, a dedicated GPU of type 'NVIDIA TESLA P100' with 16 GB RAM is initialized in the VM instance. The proposed architecture is designed and validated using an open-source software tool called 'DEEPCOGNITION.AI'. It supports the visual programming of a deep learning library known as 'MXNet' [21]. The input data samples are trained as batches, and the trained weights are validated at the end of each epoch with the validation dataset. This process of updating the kernel weights using the validation dataset provides reliable training of the model. This step resembles the cross-validation technique often used in machine learning approaches. The best weights for which the validation accuracy and loss are optimal are stored during the training process. The stored weights are then used for testing the proposed architecture by estimating the performance metrics.

Results

The trained model is tested, and its performance was assessed using the test dataset. The test data set is made up of 100 PCG records that were not part of the training and validation dataset. The performance measures such as sensitivity (recall), specificity, balanced accuracy (MAcc), overall accuracy, and precision, which is also known as positive predictive value (PPV), are tabulated in Table 3. Moreover, the average of precision and recall, which is also known as F1 score, is also computed for the models.

Net-4, which consists of only MLP layers, produced the highest sensitivity of 0.8673. Net-3 produced the highest detection of true negative cases with a specificity of 0.9298. Net-4 provided the best overall performance compared to other DNN models with respect to MAcc of 0.8574. Besides, Net-4 resulted in the highest F1 score of 0.8458. The Receiver operating characteristic (ROC) curve is an alternative representation of the classification performance of the classifiers. The ROC plot of the four network architectures used in the detection of abnormal heart sounds are shown in Fig. 3. Net-4 provided the highest Area under the Curve (AUC) value of 0.857, signifying the better classification accuracy of the Net-4 compared to other DNN models considered in this study. Net-2 and Net-3 provide an acceptable AUC value of 0.720 and 0.857 for the heart sound classification.

The training of the deep neural networks involved multiple trial runs (trials = 10) in which the weights of each DNN are monitored. The weights for which the maximum validation accuracy is stored for testing the DNN. Net-1 took on an average time of 374 s, Net-2 took 97 s, Net-3 took 104 s, and Net-4 took 197 s, respectively for the training process.

Discussion

Automated classification of the cardiovascular malfunction from the PCG signal is widely detected through the feature engineering process. In literature, handcrafted features have been extracted from the phonocardiogram signal and the features were used to train classifiers such

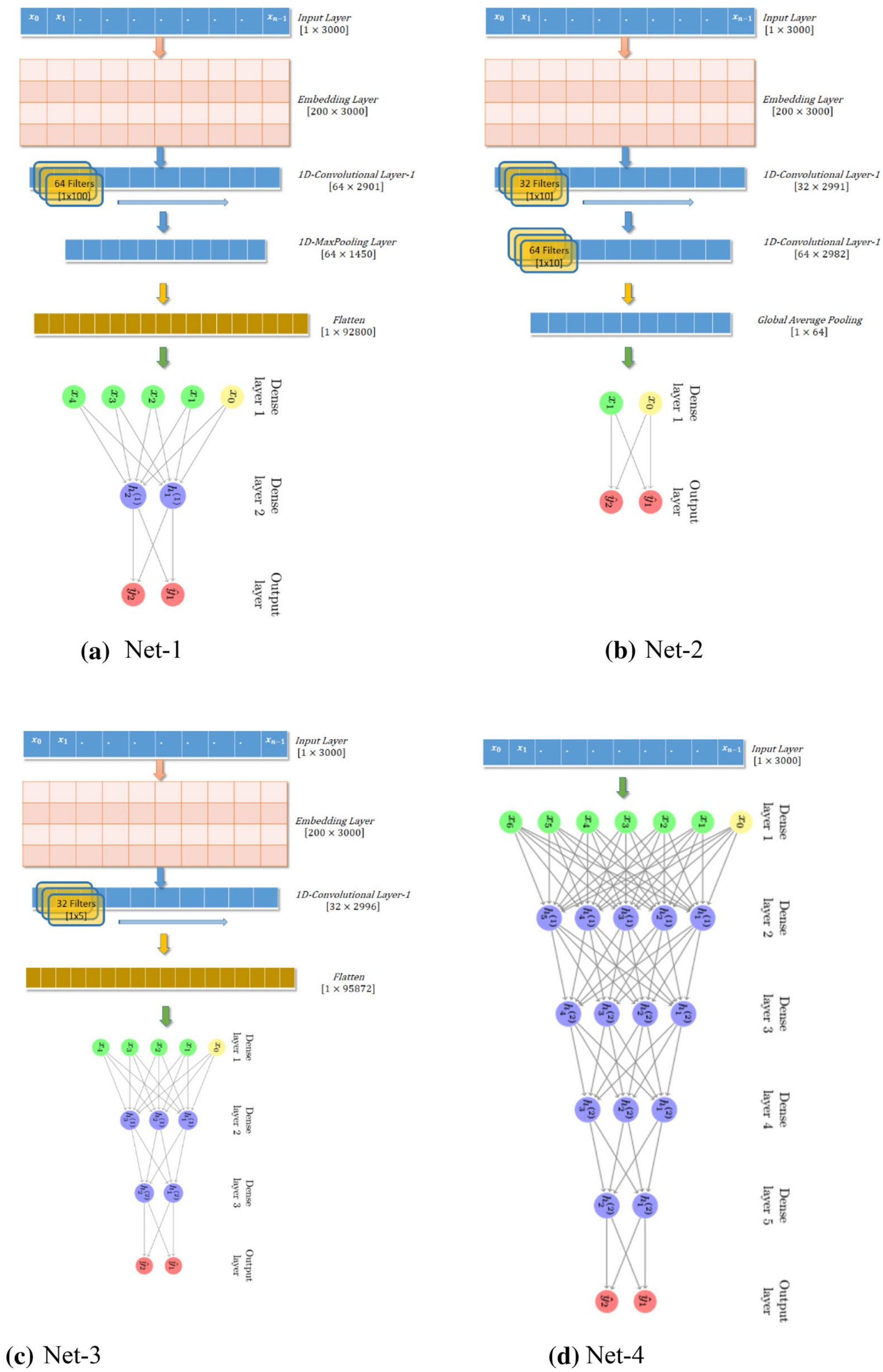

Fig. 2 Proposed DNN architectures for heart sound classification

Table 3 Performance measures of the DNN models for abnormal heart sound detection from PCG data

Network Architecture	Sensitivity	Specificity	MAcc	Accuracy	Precision	F1 Score
Net-1	0.5111	0.6608	0.5860	0.6296	0.2840	0.3651
Net-2	0.5556	0.7953	0.6754	0.7454	0.4167	0.4762
Net-3	0.5778	0.9298	0.7538	0.8565	0.6842	0.6265
Net-4	0.8673	0.8475	0.8574	0.8565	0.8252	0.8458

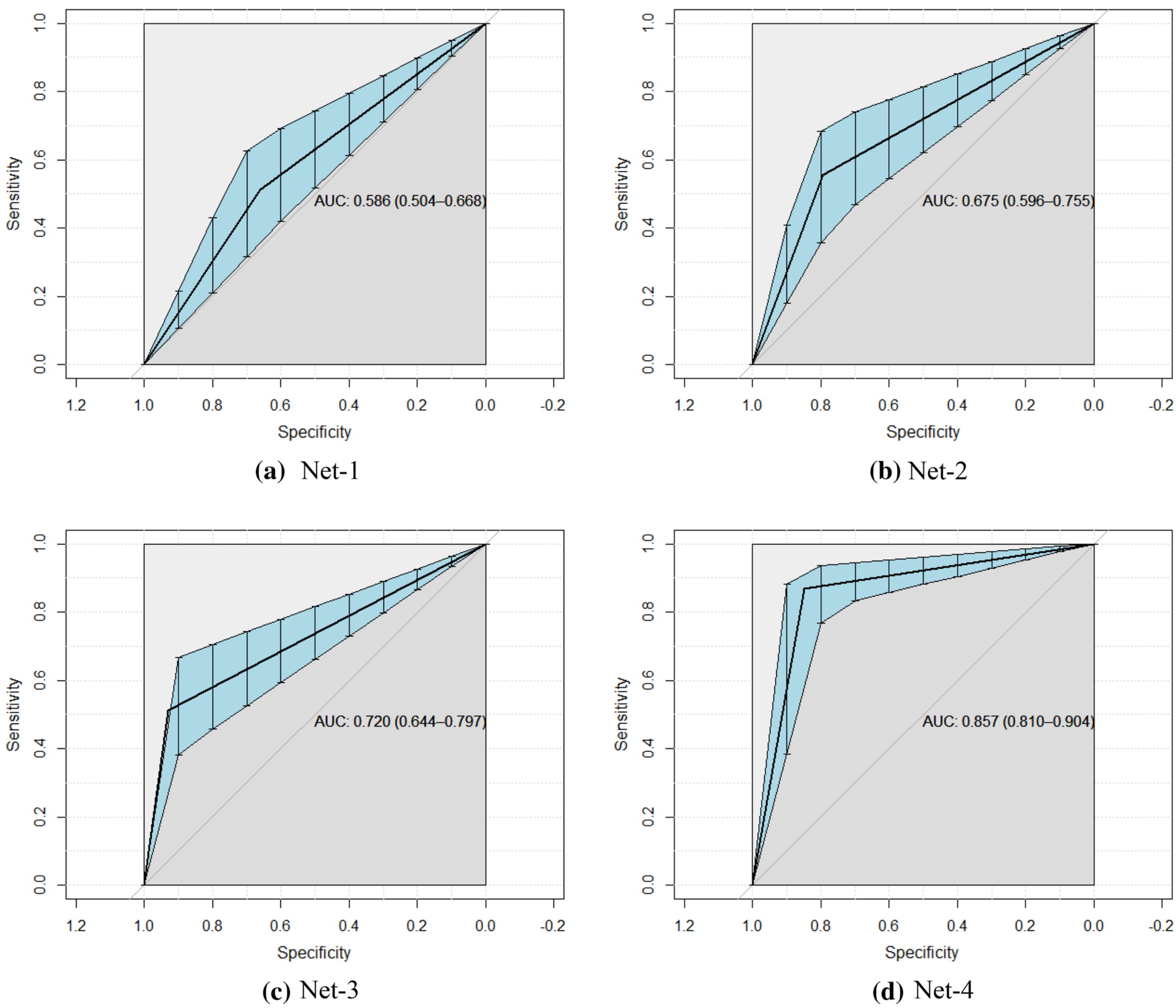

Fig. 3 Receiver operating characteristic (ROC) plots of the DNN models with area under the curve (AUC) values

as the support vector machine, random forest, or k-NN for classification [15]. Though methods on feature engineering have been reported with good classification accuracy for the detection of abnormal sounds, they inherently have a problem of subjective analysis. Deep neural networks proposed in this study can perform feature engineering in an automated manner. The filters of the 1D-CNN are randomly initialized and are trained on the input data samples. The kernel weights are updated iteratively such that it allows the network to select the appropriate filters that best suits feature classification. Moreover, DNN architectures proposed in this study can fully take advantage of the latest GPU hardware for accelerated training.

Table 4 Comparison of related work in heart sound classification based on physionet challenge 2016 dataset

Related work	Method	Performance metrics
Kay & Agarwal [14]	Dropconnected neural networks trained on time–frequency	Sensitivity:87% Specificity:82% MAcc:85%
Homsi et al. [15]	Random forest + LogitBoost + Cost-sensitive classifier	Sensitivity: 80% Specificity:81% MAcc:80%
Whitaker et al. [10]	Sparse coding and time-domain features	Sensitivity: 90% Specificity: 88% MAcc: 89%
Plesinger et al. [11]	Probability assessment	Sensitivity: 87% Specificity: 94% MAcc: 90%
Maknickas et al. [12]	Deep convolutional neural networks and Mel-frequency spectral coefficients	MAcc: 86.02%
Langley et al. [13]	Wavelet entropy from unsegmented phonocardiograms	Sensitivity: 94% Specificity: 65% MAcc: 80%
This study	Deep neural network 1. 1D-CNN (Net-3)	Sensitivity: 58% Specificity: 93% MAcc: 75%
	2. Feedforward NN (Net-4)	Sensitivity: 87% Specificity: 85% MAcc: 86%

Table 4 shows how the results obtained in this study compare with related studies on the detection of abnormal heart sound from PCG. The methods employed in this study involve the same challenge 2016 dataset of Physionet. However, it has to be noted that the proposed method is tested only with the training dataset available from the PhysioNet Challenge 2016 database. Kay and Agarwal reported an overall accuracy of 85% for a combination of time–frequency and inter-beat features. The authors have also reported a better estimate of 74.8% from their algorithm by balancing the dataset. However, their method requires the segmentation of heart sounds and noiseless signal data for better classification performance [14]. Homsi et al. also proposed similar features obtained from time–frequency, wavelet, and statistical analysis. However, they have used an ensemble of the classifier to improve the performance of the heart sound detector and reported an overall accuracy of 80%, which also requires prior segmented PCG records [15]. Whitaker et al. used a sparse coding technique to extract features from the data records. They have proposed matrix norm sparse coding and time-domain features. They have reported a cross-validated overall score of 89% for the test dataset [10]. The mean score for their algorithm on a smaller dataset was only 56%. This suggests that their method requires a large training dataset for improved classification performance. Moreover, the sparse coding scheme is computationally intensive as it has been stated in their work that training requires many hours.

Plesinger et al. proposed a semi-automated method that requires a prior segmentation of the PCG signal into fundamental heart sounds. Moreover, their method requires a feature selection stage based on a probabilistic approach to select discriminant features from a set of 228 statistical features. They have reported a balanced accuracy of 90%. However, their method required prior segmentation of PCG signal and probability assessment of features were time-consuming [11]. Maknickas et al. proposed a deep neural network based on 2D-CNN for heart sound classification. It required converting the time frames into image frames through MFSC. Their work involved a smaller training dataset of only 520 normal frames and 567 abnormal frames. Often training of deep learning architectures requires more data samples to improve generalization capacity [12]. Langley et al. proposed two features namely, normalized spectral amplitude and wavelet entropy for classification of unsegmented PCG signals. They have reported that the wavelet entropy showed improved results compared to spectral amplitude with an accuracy of 80% for low noise datasets [13]. However, their method produced an accuracy of 76% with high noise dataset. It can be observed that the majority of research works related to this study have used extensive feature engineering for heart sound classification while most of the methods except the work of Langley et al. used segmented heart sound for feature processing. The classification accuracy of abnormal sound detection is related to the performance of the segmentation of heart sounds. Most feature engineering methods proposed high dimensional feature set, which in turn requires feature selection for selecting relevant features for training the classifier. Many features do not correlate with an increase in accuracy, as demonstrated by Langley et al., which used only wavelet entropy to produce comparable classification accuracy.

The main contribution of this study is that a fully automated heart sound classification system was designed and validated using a set of DNN models such that proposed models do not involve extensive feature engineering or feature selection process. However, the experimental analysis suggests that not all the proposed models performed well in the detection of the abnormal heart sound. Net-1 and Net-2 produced low classification performance with the MAcc value of 0.5860 and 0.6754, respectively. It is due to the large value of the filter length of 100 and 10 for the 1D-CNN layers. It had failed to capture the temporal features of the PCG signal. However, reducing the filter size of the 1D-CNN to 5 produced a moderate performance of Net-3 with the MAcc value of 0.7538. Moreover, the performance is also related to the fully connected layers of each network. It can be seen that Net-4, which consisted of only MLP layers, produced the highest classification accuracy with the MAcc value of 0.8574. The simple Feedforward neural network used in the proposed work (Net-4) was able to capture the temporal relationship of normal and abnormal PCG time series data. The comparative analysis provided in Table 4 suggests that the deep Feedforward neural net (Net-4) proposed in this study can detect abnormal heart sounds without much of pre-processing of the PCG signal. It can reduce the analysis time of screening the PCG records for heart disease identification, thus assisting the cardiologist in providing a faster treatment plan to the patients.

Conclusion

In this study, we have demonstrated the application of deep neural networks for the classification of the PCG signal without feature engineering and feature selection stages. The method directly takes the sequential signal data for classifying into normal or abnormal heart sounds. The Feed-forward Neural Network (Net-4) produced the highest balanced accuracy of 0.8574 compared to other DNN models. Moreover, the proposed method does not require prior segmentation of PCG sounds into the fundamental heart sounds. The Net-4 model produced better classification accuracy when compared to the work of Langley et al. for the unsegmented PCG records with a low noise component. Thus, the deep neural network based on Feed forward neural network as proposed in this study can be a viable tool in detecting abnormal heart sounds from an unsegmented PCG signal without any feature engineering process.

Funding This article does not receive funding from any agencies. All expenditures for this study have been managed by the first author.

Data availability The patient medical data used in this study were obtained from a publicly available source: Physionet.org.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. Cardiovascular diseases (CVDs): WHO Fact Sheet. WHO (2017) https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed 1 May 2019
2. Liu C, Springer D, Clifford GD (2017) Performance of an open-source heart sound segmentation algorithm on eight independent databases. Physiol Meas 38:1730–1745. https://doi.org/10.1088/1361-6579/aa6e9f
3. Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ et al (2016) An open access database for the evaluation of heart sound algorithms. Physiol Meas 37:2181–2213. https://doi.org/10.1088/0967-3334/37/12/2181
4. Redlarski G, Gradolewski D, Palkowski A (2014) A system for heart sounds classification. PLoS ONE 9:e112673. https://doi.org/10.1371/journal.pone.0112673
5. Choi S, Jiang Z (2010) Cardiac sound murmurs classification with autoregressive spectral analysis and multi-support vector machine technique. Comput Biol Med 40:8–20. https://doi.org/10.1016/J.COMPBIOMED.2009.10.003
6. Zheng Y, Guo X, Ding X (2015) A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification. Expert Syst Appl 42:2710–2721. https://doi.org/10.1016/J.ESWA.2014.10.051
7. Zhang W, Han J, Deng S (2017) Heart sound classification based on scaled spectrogram and partial least squares regression. Biomed Signal Process Control 32:20–28. https://doi.org/10.1016/J.BSPC.2016.10.004
8. Hamidi M, Ghassemian H, Imani M (2018) Classification of heart sound signal using curve fitting and fractal dimension. Biomed Signal Process Control 39:351–359. https://doi.org/10.1016/J.BSPC.2017.08.002
9. Springer D, Tarassenko L, Clifford G (2015) Logistic Regression-HSMM-based heart sound segmentation. IEEE Trans Biomed Eng 63:1–1. https://doi.org/10.1109/TBME.2015.2475278
10. Whitaker BM, Suresha PB, Liu C, Clifford GD, Anderson DV (2017) Combining sparse coding and time-domain features for heart sound classification. Physiol Meas 38:1701–1713. https://doi.org/10.1088/1361-6579/aa7623
11. Plesinger F, Viscor I, Halamek J, Jurco J, Jurak P (2017) Heart sounds analysis using probability assessment. Physiol Meas 38:1685–1700. https://doi.org/10.1088/1361-6579/aa7620
12. Maknickas V, Maknickas A (2017) Recognition of normal–abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients. Physiol Meas 38:1671–1684. https://doi.org/10.1088/1361-6579/aa7841
13. Langley P, Murray A (2017) Heart sound classification from unsegmented phonocardiograms. Physiol Meas 38:1658–1670. https://doi.org/10.1088/1361-6579/aa724c
14. Kay E, Agarwal A (2017) DropConnected neural networks trained on time-frequency and inter-beat features for

classifying heart sounds. Physiol Meas 38:1645–1657. https://doi.org/10.1088/1361-6579/aa6a3d

15. Nabhan Homsi M, Warrick P (2017) Ensemble methods with outliers for phonocardiogram classification. Physiol Meas 38:1631–1644. https://doi.org/10.1088/1361-6579/aa7982
16. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG et al (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101:e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215
17. Karim F, Majumdar S, Darabi H, Chen S (2018) LSTM Fully convolutional networks for time series classification. IEEE Access 6:1662–1669. https://doi.org/10.1109/ACCESS.2017.2779939
18. Sadouk L. CNN Approaches for time series classification. Time Series Analysis, IntechOpen; 2019. doi:10.5772/intechopen.81170.
19. Acharya UR, Fujita H, Lih OS, Adam M, Tan JH, Chua CK (2017) Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. Knowl-Based Syst 132:62–71. https://doi.org/10.1016/j.knosys.2017.06.003
20. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adam M, Gertych A et al (2017) A deep convolutional neural network model to classify heartbeats. Comput Biol Med 89:389–396. https://doi.org/10.1016/j.compbiomed.2017.08.022
21. Deep Cognition—DeepCognition.ai n.d. https://deepcognition.ai/. Accessed 23 Nov 2018