Classification of Heart Sound Recordings using Convolution Neural Network

Heechang Ryu, Jinkyoo Park, Hayong Shin

Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

Abstract

Aims: This study proposes a cardiac diagnostic model using convolution neural network (CNN). This model can predict whether a heart sound recording is normal or not by classifying phonocardiograms (PCGs) from both clinical and nonclinical environments – in accordance with the "2016 Physionet/CinC Challenge".

Methods: Heart sound recordings in the training data set are filtered by using Windowed-sinc Hamming filter algorithm to remove signals regarded as noise. The filtered recordings are then scaled and segmented. Using the filtered and segmented recordings, CNN is trained to extract features and construct a classification function. The CNN is trained by back propagation algorithm with stochastic gradient descent and mini-batch learning. To classify one sound recording, the signal should be filtered and segmented. Each segment of the signal is then classified by the trained CNN model. The model assigns each segment signal a relative probability between normal and abnormal labels. By accumulating these relative probability values for all the segmented signals, one can reliably and robustly determine whether the target signal is normal or abnormal.

Results: The proposed model achieved an overall score of 79.5 with a sensitivity of 70.8 and a specificity of 88.2.

1. Introduction

Cardiac abnormalities are known as one of the life-threatening abnormalities. Heart sounds and murmurs are usually caused by the blood stream in the chambers of a heart and opening and closing mechanism of the heart valves [1]. The heart sounds and murmurs have been used for diagnosing cardiac abnormalities through auscultation system by medical experts. Only highly skilled physicians can interpret the heart sounds and diagnose potential diseases through the auscultation process [2]. With advancements in many computing algorithms and classification methods, automated classification methods, such as threshold-based method [3], hidden Markov model [4] and artificial neural network [2,5] have been employed in many disciplines for over 50 years. However, such models have flaws: (1) they require carefully designed input features to obtain a high classification accuracy and (2) the performance, i.e., test error, of such methods is evaluated using data sets collected from well-controlled environments.

With an aid of graphic processing units that is capable of computing massive operations efficiently, deep learning including neural network has drawn a lot of interest from researchers again after 2000. Especially, Convolution Neural Network (CNN) has been applied with great success to detecting, segmenting and recognizing visual images and time series. Since it extracts features autonomously, CNN has been used in various disciplines, such as speech recognition and pattern recognition.

The aim of this study is to develop a CNN-based classification algorithm that can identify abnormal heart sound recordings from both clinical and nonclinical environments, even with very poor signal quality [6].

2. Methods

As shown in Figure 1, the proposed model is divided into three parts – filtering, segment classification and recording classification.

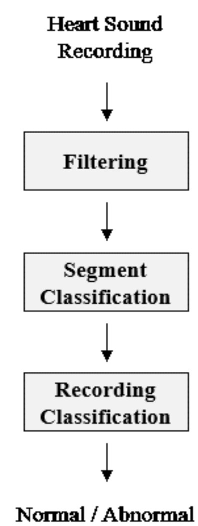

Figure 1. Proposed model.

 ISSN: 2325-887X DOI:10.22489/CinC.2016.329-134

Through the model, it states whether the heart sound recording is normal or abnormal as cardiac diagnosis. Filtering process removes signals regarded as noise from the sound recordings by using Windowed-sinc Hamming filter algorithm [2]. In addition, the filtered sound recordings are scaled. The scaled sound recordings are then divided into multiple segments. Segment classification process first extracts the features from each segment. The extracted features are then used to determine whether each signal segment is normal or not using CNN. Lastly, recording classification process determines whether the whole sound recording is normal or not based on the classification results of segmented recordings. Specifically, it classifies the heart sound signal based on the majority voting.

2.1. Filtering

The heart sound recordings containing only PCGs have been resampled with 2,000 Hz. The sound recordings contain not only heart murmur caused by abnormal activities of a heart but also noise from environments. It is desirable to remove noise from the signal because noise can lower the classification accuracy. However, in general, it is difficult to know which components of the signal are noises that are not informative for the classification. Assuming that the high frequency components of the signal are not directly relevant to cardiac activities, we remove the signal over a certain frequency threshold. The Windowed–sinc Hamming filter algorithm with a filter kernel using 101 points is used for hi-pass filtering. The filtering procedure is illustrated in Algorithm 1.

Algorithm 1 Windowed-sinc Hamming filter algorithm

```
X ← amplitude of heart sound recording
for i ← 1, … ,101 do
  Hamming window ← (0.54-0.46*cos(2*pi*(i/m)))
  H(i) ← sin(2*pi*fc*(i-m/2))/(i-m/2)*Hamming window
  if i = 50 then
    H(i) ← sin(2*pi*fc)*Hamming window
  m ← 100
  fc ← 400/2000 (cut off frequency / sampling frequency)
H ← H / sum(H)
for all j > 102 in X's points do
  for k ← 1, … ,101 do
    Y(j) ← Y(j) + X(j-k) * H(k)
```

The filter kernel is normalized by its summation and convolved with heart sound record signal to get filtered signal. Figure 3 shows a typical example of filtering with the 400 Hz cut off frequency. The signals before filtering and after filtering are compared in Figure 3.

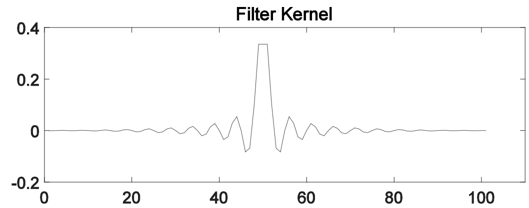

Figure 2. Normalized filter kernel with 101 points.

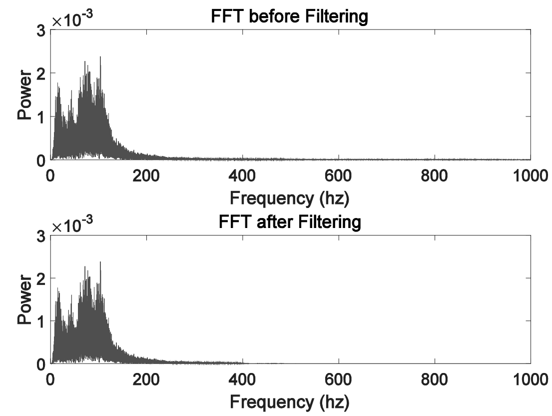

Figure 3. Before and after Windowed–sinc Hamming filtering on training data recording “a0011” are expressed through Fast Fourier transform (FFT).

The filtered signal is scaled by using maximum and minimum of the signal and segmented by 5000 points corresponding to about 2.5 seconds, as shown in Figure 4.

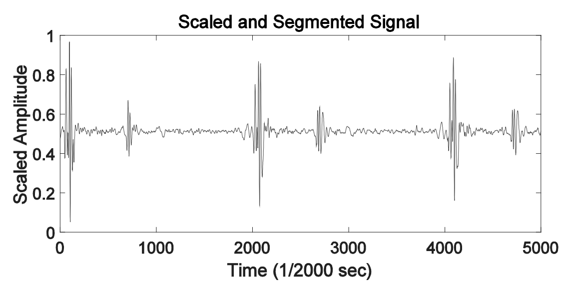

Figure 4. One of scaled and segmented signals after Windowed–sinc Hamming filtering on training data recording “a0011”.

2.2. Segment classification

Using the filtered and segmented sound recordings, CNN is trained to extract features and construct a classification function. As shown in Figure 5, CNN consists of two layers, namely convolution layers and fully connected layers.

First, convolution layers extract high level (abstracted) features by sequentially transforming raw input data into high-level abstract features. The convolution layers contain non-linear activation function, batch normalization and max pooling layers. The non-linear activation function is used to increase the representability of a network (i.e.,

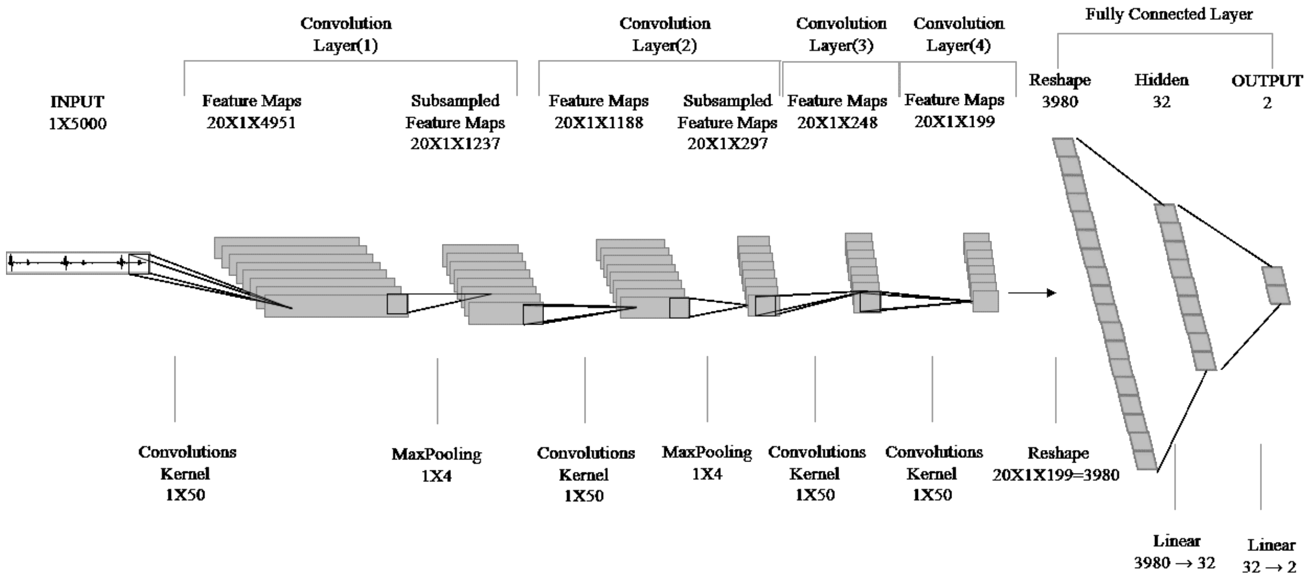

Figure 5. Structure of convolution neural network for segment classification.

increase the fitness of a model to the training data). Batch normalization is used to achieve a higher learning rate and prevents overfitting [8]. The max pooling layers select maximum value among its neighbors to reduce noise and extract more abstract features. After having gone through the convolution layers, the features are autonomously extracted from the segmented signals.

Fully connected layers output values for classification predictions by linearly combining the features extracted from convolutions layers and having the combined values go through non-linear activation functions. As one segmented signal goes through the convolution and fully-connected layers, the features are extracted and used to classify the signal, respectively.

After an architecture of CNN is designed, the parameters of CNN have been trained by back propagation algorithm with stochastic gradient descent and mini-batch learning using training data. We use a deep learning framework, Torch [7], to construct and train our model.

2.3. Recording classification

To classify one sound recording, the signal should be filtered and segmented into N segmented. Each segment of the signal recording is then classified by the trained CNN model. The model assigns each segment signal a relative probability between normal and abnormal. By accumulating these relative probability values for all the segmented signals, one can reliably and robustly determine whether the target signal is normal or abnormal.

3. Results

The proposed model was trained with the training data set (3,126 recordings), and its performances were evaluated using the validation data set (300 recordings) and hidden test data set. Table 1 summarizes performances of the proposed model with the cut off frequencies used to filter out noise signal.

Table 1. Results. Validation - testing on validation data set, Score - testing on hidden test data set in Physionet test environment.

Cut off frequency	Validation	Score
100Hz	72.33	-
150Hz	75	78.3
200Hz	72.67	-
250Hz	73	-
300Hz	74	79.5
350Hz	73.67	76.9
400Hz	74	79.5
450Hz	71	-
500Hz	71.33	-

Using the proposed model, we have achieved an overall score of 79.5 with a sensitivity of 70.8 and a specificity of 88.2 in “CinC Challenge 2016” official phase.

Acknowledgements

This research was supported by 2016 College of Engineering Graduate Exploratory Research fund by Korean Advanced Institute of Science and Technology (project number: N11160144).

References

[1] Leatham, A. Auscultation of the heart and phonocardiography. Churchill Livingstone: 1975.
[2] Shamsuddin N, Mustafa MN, Husin S, Taib MN: Classification of heart sounds using a multilayer feed-forward neural network. In: Asian Conference on Sensors and the International Conference on New Techniques in Pharmaceutical and Biomedical Research; 2005;p:87–90.
[3] Gerbarg, D.S.; Taranta, A.; Spagnuolo, M.; Hofler, J.J. Computer analysis of phonocardiograms. Progress in Cardiovascular Diseases 1963;5:393-405.
[4] Wang, P.; Lim, C.S.; Chauhan, S.; Foo, J.Y.; Anantharaman, V. Phonocardiographic signal analysis method using a modified hidden markov model. Ann Biomed Eng 2007;35:367-374.
[5] Gupta, C. N., R. Palaniappan, S. Swaminathan, and S. M. Krishnan. Neural network classification of homomorphic segmented heart sounds. Appl. Soft Comp. In Press, 2007;7:286-297.
[6] Liu et al, "An Open Access Database for the Evaluation of Heart Sound Algorithms", Physiol. Meas, 2016;37:9.
[7] R. Collobert, K. Kavukcuoglu, and C. Farabet. Torch7: A MATLAB-like environment for machine learning. In BigLearn, NIPS Workshop, 2011.
[8] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In arXiv:1502.03167, 2015.

Heechang Ryu
Industrial and Systems Engineering
Korea Advanced Institute of Science and Technology
291 Daehak-ro
Yuseong-gu
Daejeon 34141
Republic of Korea
rhc93@kaist.ac.kr