

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Li, Fan; Tang, Hong; Shang, Shang; Mathiak, Klaus; Cong, Fengyu

Title: Classification of Heart Sounds Using Convolutional Neural Network

Year: 2020

Version: Published version

Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland.

Rights: CC BY 4.0

Rights url: <https://creativecommons.org/licenses/by/4.0/>

Please cite the original version:

Li, F., Tang, H., Shang, S., Mathiak, K., & Cong, F. (2020). Classification of Heart Sounds Using Convolutional Neural Network. *Applied Sciences*, 10(11), Article 3956.

<https://doi.org/10.3390/app10113956>

Article

Classification of Heart Sounds Using Convolutional Neural Network

Fan Li ¹, Hong Tang ¹ , Shang Shang ¹, Klaus Mathiak ²  and Fengyu Cong ^{1,3,4,5,*}

¹ School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, No. 2 Linggong Street, Dalian 116024, China; lifandlpu@foxmail.com (F.L.); tanghong@dlut.edu.cn (H.T.); shangshang@mail.dlut.edu.cn (S.S.)

² Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Uniklinik RWTH Aachen, Pauwelsstr. 30, 52074 Aachen, Germany; KMathiak@UKAachen.de

³ School of Artificial Intelligence, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, No. 2 Linggong Street, Dalian 116024, China

⁴ Key Laboratory of Integrated Circuit and Biomedical Electronic System, Liaoning Province, Dalian University of Technology, No. 2 Linggong Street, Dalian 116024, China

⁵ Faculty of Information Technology, University of Jyväskylä, 40014 Jyväskylä, Finland

* Correspondence: cong@dlut.edu.cn; Tel.: +86-0411-8470-8704

Received: 11 April 2020; Accepted: 4 June 2020; Published: 6 June 2020



Featured Application: Combining of multi-features extracted manually and convolutional neural network classifier for automatic heart sounds classification.

Abstract: Heart sounds play an important role in the diagnosis of cardiac conditions. Due to the low signal-to-noise ratio (SNR), it is problematic and time-consuming for experts to discriminate different kinds of heart sounds. Thus, objective classification of heart sounds is essential. In this study, we combined a conventional feature engineering method with deep learning algorithms to automatically classify normal and abnormal heart sounds. First, 497 features were extracted from eight domains. Then, we fed these features into the designed convolutional neural network (CNN), in which the fully connected layers that are usually used before the classification layer were replaced with a global average pooling layer to obtain global information about the feature maps and avoid overfitting. Considering the class imbalance, the class weights were set in the loss function during the training process to improve the classification algorithm's performance. Stratified five-fold cross-validation was used to evaluate the performance of the proposed method. The mean accuracy, sensitivity, specificity and Matthews correlation coefficient observed on the PhysioNet/CinC Challenge 2016 dataset were 86.8%, 87%, 86.6% and 72.1% respectively. The proposed algorithm's performance achieves an appropriate trade-off between sensitivity and specificity.

Keywords: automatic heart sound classification; feature engineering; convolutional neural network

1. Introduction

Statistics show that cardiovascular disease (CVD) is one of the main reasons for mortality in the world [1,2]. Heart sounds are a kind of mechanical vibrations caused by the movement of blood in the cardiovascular system [3] and they are considered to be an important indicator in the diagnosis of several CVDs. Compared with electrocardiograph (ECG) signals, a phonocardiogram (PCG), which is a graphical recording of heart sounds, can clearly express changes in the state of the heart. So PCG can be used to diagnose deformation in heart organs and damage to heart valves [4]. Auscultation is the major means by which clinicians listen to heart sounds. The results of an auscultation are then used by doctors to detect cardiac diseases. This process is subjective, time-consuming and requires

extensive training and experience in auscultation [5]. Therefore, it is necessary to develop a method for the automatic classification of heart sounds.

The common approaches to heart sound or PCG classification include three steps—(1) feature extraction; (2) feature selection; and (3) classification by a classifier. The performance of PCG classification is influenced by all three of these processes. Gerbarg et al. [6] were the first researchers to attempt the automatic classification of a pathology in a PCG recording by the threshold-based method. In our previous work [7,8], we, in Reference [7], extracted 324 features and used a back propagation neural network (BP, NN) for PCG classification. Furthermore, in Reference [8], we extracted 515 features from nine feature domains on the same dataset as used in the present studying and trained a support vector machine (SVM) with a radial basis kernel function for signals classification. Yaseen et al. [9] applied an SVM and a deep neural network (DNN) to extract Mel Frequency Cepstral Coefficients (MFCCs) and Discrete Wavelets Transform (DWT) features from their own database and evaluated the contribution of each feature. Other researchers have worked on extracting features from heart sound signals; however, to the best of our knowledge, the features extracted in Reference [8] include almost all of the features that are referred to in other related research.

Deep learning [10], especially for convolutional neural network (CNN), has been used to achieve great success in such fields, image detection [11,12], speech recognition [13] and medical image and signal analysis [14]. Recently, many researchers have used deep learning methods to automatically classify heart sounds. A CNN was used as a feature extractor in Reference [15] and the features extracted by the CNN were fed into an SVM for heart sound classification. Time-frequency features of raw heart sound signals were fed into a CNN model to classify normal and abnormal heart sounds in References [16,17]. In Reference [18], the authors used Mel-frequency spectral coefficients (MFSCs) features as inputs into a CNN model. Rubin et al. in Reference [19] proposed an ensemble of feature-based and CNN -based classifiers for heart sound classification tasks. Andries et al. compared the classification performance between a CNN and two other classifiers (SVM and K-nearest Neighbor, (KNN)) based on the features extracted from the CNN. These authors also compared the results obtained from the same classifiers using different features extracted from a linear binary pattern (LBP) and the CNN [20].

All of these studies have made a great contribution to the field of automatic heart sound classification. However, some challenges remain in the field of heart sound classification:

- (1) The extraction of new feature types from heart sounds.
- (2) New methods for heart sound classification.
- (3) imbalanced numbers of normal and abnormal heart sounds.

Deep learning is a big data algorithm. However, we used a small number of dataset in this study, which would likely have led to overfitting if we employed a deep neural network [21]. Deep learning methods have been combined with the traditional feature engineering method in some fields recently [22]. A CNN has been used to extract features automatically in many recent studies. However, the features that are extracted by a CNN are abstract and have low explanatory power. Therefore, in this study, we classified normal and abnormal heart sounds using a combination of the feature engineering method and a compact CNN model. We also aimed to explore the feature selection and classification capabilities of the CNN. First, 497 features were extracted from eight domains in accordance with our previous work. Then, the designed CNN was used to select features and distinguishing abnormal heart sounds from normal heart sounds. We introduced dropout and global average pooling strategies into the proposed model to avoid overfitting and class weights were set in order to solve the class imbalance problem. Finally, we evaluated the performance of the proposed method using mean accuracy, sensitivity, specificity and Matthews correlation coefficient (MCC) based on stratified five-fold cross-validation. To demonstrate the validity of the designed CNN model, we further tested several machine learning algorithms (support vector machine [23], random forest [24] and adaboost [25]). The performances of the proposed method were more remarkable than those

machine learning classifiers. Compared with recurrent neural network (RNN) [26], which is said to be better for sequential tasks, CNN can reduce the computation overhead [27]. The results indicate that the proposed method efficiently classifies normal and abnormal heart sounds and demonstrate the feasibility of the CNN as a feature selector and classifier.

2. Materials and Methods

2.1. Dataset

In this study, we used a dataset from PhysioNet/Computing in Cardiology Challenge2016 [28], which aims to facilitate research on novel methods for heart sound classification. The dataset, which contains a total of 3153 PCG recordings, is composed of six subsets that were provided by different groups around the world. All of these PCGs recorded the heart sounds of adults and children using clinical and nonclinical methods. Their sampling frequency is 2000 Hz. The length of the recordings varies from 5 s to over 120 s. They are all in the wav format. The dataset includes 2488 normal samples and 665 abnormal samples manually labelled with -1 and 1 , respectively [29]. For convenience, the labels -1 and 1 were replaced with 1 and 0 , respectively, in this study. The normal recordings are from healthy subjects and the abnormal recording are from patients with a confirmed cardiac diagnosis, which is not specified [30]. For more details, please refer to Reference [29].

Normal heart sounds usually include a first (S1) and a second (S2) heart sound. S1 occurs at the beginning of the isovolumetric ventricular contraction, when the mitral and tricuspid valves close due to the rapid increase in pressure within the ventricles. S2 occurs at the beginning of diastole with the closure of the aortic and pulmonic valves. Between S1 and S2 or starting at or after S2 and ending before or at S1 separately, the cardiac cycle also contains systole (Sys) and diastole (Dia). So, one PCG recording can be divided into S1, systole, S2 and diastole [31].

The proposed method includes two major steps—feature extraction and feature selection and classification based on the designed CNN. The workflow of this approach is shown in Figure 1. We describe each step in the following subsections.

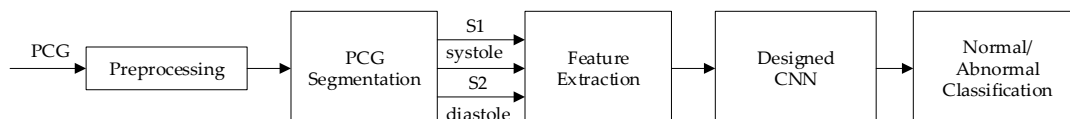


Figure 1. Block diagram of the proposed method. First, the phonocardiogram (PCG) is preprocessed and segmented. After that, the four states of each PCG recording are used to extract multiple features. Finally, the extracted features are fed into the designed convolutional neural network (CNN) model to classify normal and abnormal heart sounds.

2.2. Preprocessing

According to the characteristics of the PCG recordings described in Section 2.1, we performed filtering, regularization and segmentation of the heart sounds before we designed the classification algorithm.

Firstly, each PCG recording was filtered by a high-pass filter with a cut-off frequency of 10 Hz to remove baseline drift. Secondly, the spike removal algorithm was applied to the filtered recordings [32]. Thirdly, the recordings were normalized to zero mean and unit standard deviation. Finally, the PCG recordings were segmented using the method of the hidden semi-Markov model (HSMM) segmentation method proposed by Springer et al. [33].

2.3. Multi-Feature Extraction

A total of 497 features were extracted from eight domains. All these procedures were processed in Matlab2016b. The number of features extracted from each domain is listed in Table 1. Details on how

these features were extracted are given in the following subsections. The features of each domain are summarized in the Appendix A (Tables A1–A6).

Table 1. Summary of the numbers of features extracted from eight domains.

Index	Domain	Num. of Features
1	Time	20
2	Amplitude	12
3	Energy	62
4	High-order statistics	16
5	Cepstrum	65
6	Frequency	308
7	Cyclostationary	4
8	Entropy	10
Total	–	497

2.3.1. Features in the Time Domain

The features from this domain are listed in the Appendix A (Table A1). The methods used to extract the features in index 1–16 have been described in References [7,29], which specify the physiological meanings from the point of view of heart physiology. We added another four features (index 17–20). So, 20 features in total were extracted from the time domain.

2.3.2. Features in the State Amplitude Domain

Previous studies have shown that the amplitude of a heart sound is related to the heart’s hemodynamics [3,34,35]. So, we considered the amplitudes of heart sounds as features. First, the absolute values of the amplitude were normalized. Then, the ratios of the absolute amplitude between different states were calculated. Furthermore, the mean and standard deviation of the ratio of the absolute amplitude between any two different states in a PCG recording were taken to be the features of this domain. Hence, a total of 12 amplitude features were extracted from four states. They are listed in the Appendix A (Table A2).

2.3.3. Features in the Energy Domain

There are two kinds of features in the energy domain—(1) the energy ratio of the band-passed signal to the original signal; and (2) the energy ratio of one state to another.

For (1), we considered the frequency bands with an initial bandwidth of 10 Hz and an incremental bandwidth of 10 Hz. We obtained 42 frequency bands from 10 Hz to 430 Hz ([10 20] Hz, [20 30] Hz, [30 40] Hz, . . . , and [420 430] Hz). So, there were 42 features for each of these bands. First, we applied the Butterworth filter to the raw signals. The output y_i of each fifth-order Butterworth filter applied to each band signal was calculated using Equation (1).

$$y_i = \text{filter}(\mathbf{b}_i, \mathbf{a}_i, \mathbf{x}), i = 1, 2, \dots, 42, \tag{1}$$

where \mathbf{b}_i (the numerator) and \mathbf{a}_i (the denominator) denote the coefficients of the Butter-worth IIR filter. Then, the energy ratio of each band was calculated using Equation (2).

$$\text{Ratio_energy_band}_i = \frac{\sum |y_i|^2}{\sum |x|^2}, i = 1, 2, \dots, 42. \tag{2}$$

We also needed to consider the energy ratio between any two states. Taking as an example the energy ratio of S1 to the cycle period, if we have N cycles in a PCG recording and each cycle contains n discrete time indices, then the “Ratio_energy_state” can be defined as in Equation (3).

$$\text{Ratio_energy_S1_cycle} = \frac{\sum_n |S_{1i}(n)|^2}{\sum_n |c_i(n)|^2}, \quad i = 1, 2, \dots, N. \quad (3)$$

In this study, the mean and the standard deviation (SD) of $\text{Ratio_energy_state}_{S1_cycle}$ in the i th cycle were calculated as two features, namely, $m_Ratio_energy_S1_cycle$ and $sd_Ratio_energy_S1_cycle$. We also obtained the energy ratios of state S1 to the S2, systole and diastole states, respectively. So, for state S1, eight features were extracted. Similarly, we extracted six, four and two features for the S2, systole and diastole states, respectively. All of the features from this domain are listed in the Appendix A (Table A3).

2.3.4. Features in the Higher-Order Statistics Domain

Skewness is a measure of the deviation of the direction and degree in a data distribution. It is defined as a third-order normalized moment of a sample. Kurtosis is used to represent the peak value of a probability density distribution curve at the average value. The skewness and kurtosis of each state (s_i) in a cycle are defined in Equations (4) and (5), respectively. As shown in Equation (5), kurtosis is a fourth-order statistic.

$$\text{skewness} = E\left[\left(\frac{S_i - \mu}{\sigma}\right)^3\right], \quad i = 1, 2, \dots, N, \quad (4)$$

$$\text{kurtosis} = E\left[\left(\frac{S_i - \mu}{\sigma}\right)^4\right], \quad i = 1, 2, \dots, N, \quad (5)$$

where N is the number of cycles in a PCG recording, μ and σ denote the mean and variance, respectively, of one state (S) in the i th cycle and $E[\cdot]$ is the mathematical expectation. In this study, we considered the mean and standard deviation of the *skewness* and *kurtosis* of each state to be two separate features. Therefore, there were 16 features in this domain. They are listed in the Appendix A (Table A4).

2.3.5. Features in the Cepstrum Domain

In this domain, we first calculated the cepstrum of a PCG recording $p(n)$ using Equations (6)–(8). As described in Reference [36], the cepstrum coefficient decays quickly, so we only considered the first 13 cepstral coefficients. Furthermore, a new discrete sequence was generated by combining together all of the S1 states of a PCG recording. The first 13 cepstral coefficients were extracted from the cepstrum of the new discrete sequence. The same operation was performed on the new discrete sequences generated by all of the S2, systole and diastole states of a PCG recording. Finally, a total of 65 ($13 + 13 \times 4$) features was obtained.

$$P(k) = \text{DFT}[p(n)] \quad (6)$$

$$P(\hat{k}) = \log[P(k)] \quad (7)$$

$$P(\hat{n}) = \left| \text{IDFT}[P(\hat{k})] \right|. \quad (8)$$

In Equation (6), we first applied the discrete Fourier transform (DFT) to the signal $p(n)$. Then, we obtained the natural logarithm (\log) $P(\hat{k})$ of $P(k)$. Finally, by applying the inverse discrete Fourier transform (IDFT) and the absolute operation ($|\cdot|$), we obtained the cepstrum $P(\hat{n})$ of the signal $p(n)$.

2.3.6. Features in Frequency Domain

In this domain, we estimated the frequency spectrum of each state in each cardiac cycle using a rectangular window and the DFT. Then, we calculated the mean frequency spectrum of each state over

all cycles in a PCG recording. The spectrum values from 20 Hz to 400 Hz with a 5 Hz interval were extracted as features. Therefore, we could obtain 77 features from each state. A total of 308 (77×4) features were obtained from the frequency domain.

2.3.7. Features in the Cyclostationarity Domain

Ting et al. in Reference [37] have discussed the degree of cyclostationarity, which indicates the level of signal repetition in a PCG recording. We set

$\gamma(\alpha)$ as the cycle frequency spectral density (CFSD) of a PCG recording at cycle frequency α and $\gamma(\eta)$ as the CFSD of the PCG recording at the cycle frequency η , which is defined by the main peak location of $\gamma(\alpha)$. The degree of cyclostationarity is defined as in Equation (9).

$$d(\eta) = \frac{\gamma(\eta)}{\int_0^\beta \gamma(\alpha) d\alpha}, \quad (9)$$

where β denotes the maximum cycle frequency. It is obvious that $d(\eta)$ will be infinity with cyclical events and that random signals will have a small value. We first divided each PCG recording into several equal subsequences. Each subsequence had its own degree of cyclostationarity. Then, the mean and standard deviation, based on the degree of cyclostationarity of all subsequences in a PCG recording, were calculated as two separate features.

We also considered the sharpness of the peak of the CFSD in this domain. It is defined as the ratio of the maximum CFSD to the median CFSD:

$$peak_sharpness = \frac{\max(\gamma(\alpha))}{\text{median}(\gamma(\alpha))}. \quad (10)$$

Similarly, the mean and standard deviation of the *peak_sharpness* of all of the subsequences in a PCG recording can be calculated as another two separate features. Therefore, four features were obtained from this domain. They are listed in the Appendix A (Table A5).

2.3.8. Features in the Entropy Domain

Entropy is a commonly used feature in signal processing. Sample entropy and fuzzy measure entropy indicate the complexity of a PCG recording. Details on the calculation of sample entropy and fuzzy measure entropy can be found in References [7,36]. In this domain, we first calculated the sample entropy and fuzzy measure entropy of one PCG recording. As with the features in the cepstrum domain, we then extracted these two entropy features from four newly generated sequences. In total, 10 features were extracted from the entropy domain. They are listed in the Appendix A (Table A6).

2.4. Heart Sound Classification Based on the Designed CNN

2.4.1. Model Structure

The conventional CNN [38] typically alternates between two different types of layers—convolution and pooling. The last stage usually consists of at least one fully connected layer and a classification layer with the sigmoid function (for binary classification) [39] or the softmax function (for multi-class classification). For images or temporal series, convolution layers preserve the spatial information by using local receptive fields. The strategy of sharing the same weights can be used to learn a set of position-independent latent features. Pooling layers are used to reduce the dimensions of the feature maps generated by the convolution layers.

As shown in Figure 2, the proposed model is composed of three Conv-blocks, a global average pooling (GAP) layer and a classification layer with the sigmoid function. This classification step was completed by Keras one of deep learning architectures with python 3.6. Each Conv-block includes a convolutional1D (Conv-1D) layer and a maxpooling1D layer, followed by the strategy of dropout

to prevent overfitting [40]. The numbers of filters for the Conv1D layers were set to 32, 64 and 128, respectively and each filter had a kernel size of 3 and strides 3. The pooling size was 2 with two strides 2. Relu was adopted as the activation function in each convolution layer. We tested the dropout rate from 0.2 to 0.5 and found that the model achieved the best performance when the dropout rate was 0.25. Adam, with a learning rate of 0.001, $\beta_1(\text{beta}_1) = 0.9$ and $\beta_2(\text{beta}_2) = 0.999$ and binary cross-entropy were used as the optimizer function and the objective function, respectively.

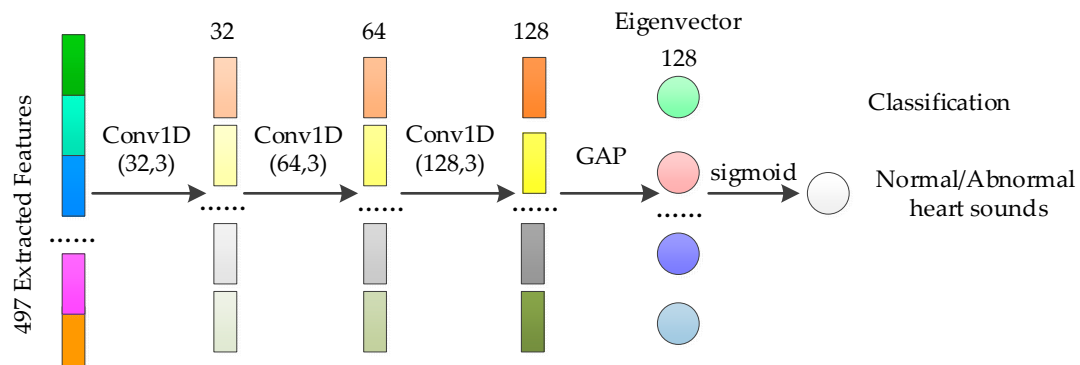


Figure 2. Structure of the designed convolutional neural network (CNN) model for normal and abnormal PCG classification. Conv1D denotes a one-dimensional convolution layer. GAP denotes the global average pooling layer.

In the proposed classification method depicted in Figure 2, as the input of the model comprises all of the extracted features, we assume that the designed model is used only to select features and classify normal and abnormal heart sounds. Therefore, the designed model is a shallow or compact model. We performed valid convolution operations on the convolution layers to explore the relationship between the different features described in Section 2.3. We also adopted pooling layers to reduce the feature dimensions. So, the dimensions of the input features gradually decreased from the bottom layers to the top layers of the Conv-blocks. This procedure reorganizes and reduces the dimensionality of the input (497 features in total), which can also be viewed as feature selection in signal processing.

The feature maps generated from the last Conv1D layer are fed into the global average pooling (GAP) layer, the principle of which is introduced in the following subsection. Each feature map outputs a value named an eigenvalue. Then, all of these eigenvalues make up the eigenvector shown in Figure 2. Finally, the eigenvector is fed into the classification layer with one neural and one sigmoid activation function to output a predicted probability.

2.4.2. Global Average Pooling

Compared with the traditional CNN, in the proposed CNN model, there is a GAP layer, which was first introduced in Reference [41], rather than fully connected layers before the classification layer. The differences between the fully connected layers and the GAP layer are shown in Figure 3.

As shown in Figure 3a, the feature points of each feature map generated by the last Conv1D layer are first flattened into a vector. Then, the vector is connected to all of the neural nodes of the fully connected layer, which produce various permutations and combinations of the feature nodes. Given m feature points and n neural nodes, the output of the fully connected layer is calculated as shown in Equation (2). There will be $mn + n$ parameters to be trained. If we increment the parameters, the model

will be prone to overfitting. On the other hand, it acts as a black hole from the fully connected layer to the classification layer.

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = [x_1 \ x_2 \ \dots \ x_m] \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{m1} & w_{m2} & \dots & w_{mn} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix}, \tag{11}$$

where, w_{ij} ($i = 1, 2, \dots, m, j = 1, 2, \dots, n$) and b_n ($n = 1, 2, \dots, n$) are the connected weights and bias, respectively, x_n is the number of feature nodes and y_n is the output.

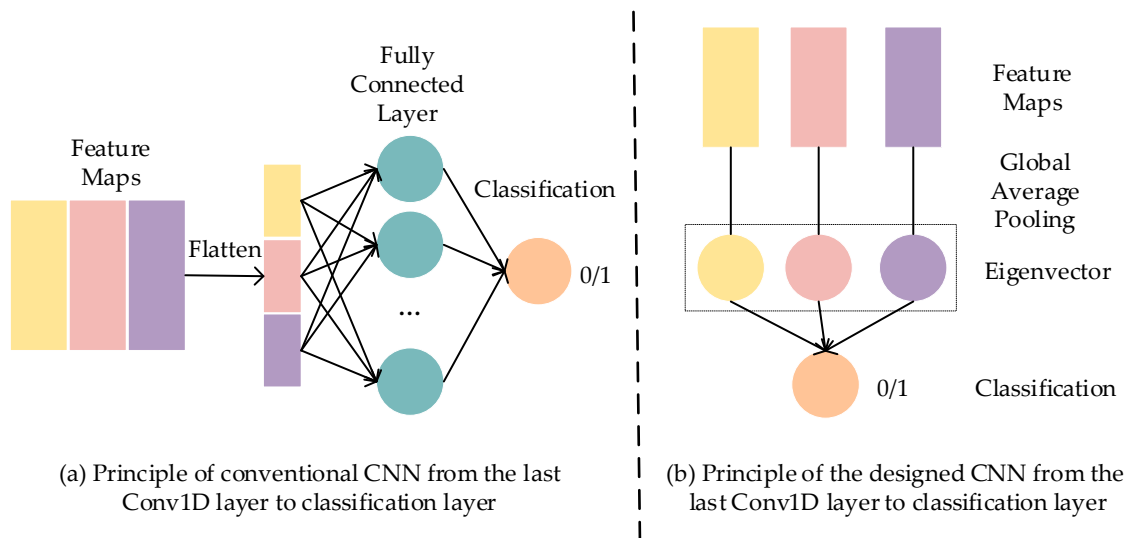


Figure 3. Comparison of the working principle between a fully connected layer and the global average pooling layer. “0” and “1” denote abnormal and normal heart sounds, respectively. The feature maps are generated by the convolution layers of the last Conv-block. (a) Description of the workflow from feature maps to the classification layer. The feature maps should first be flattened in this process. (b) Description of the principle of the global average pooling layer.

The GAP layer carries out average pooling on the global feature map, so it is able to learn global information about each feature map. Each feature map generates a feature node and all of these feature maps are mapped into an eigenvector. As shown in Figure 3b, three feature maps are obtained from the last Conv1D layer. So, we obtain an eigenvector with a size of 3 that is then fed into the classification layer with the sigmoid activation function to calculate the classification probabilities. This procedure can be defined in Equations (12) and (13).

$$x_i = \frac{1}{m} \sum_{j=1}^m a_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, m, \tag{12}$$

$$\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n], \tag{13}$$

where a_{ij} is the j th feature point of the i th feature map, n and m are the number of feature maps and the size of each feature map, respectively, x_i denotes the global average of the i th feature map and \mathbf{x} is the eigenvector generated by the GAP layer.

The GAP layer plays the role of a bridge between the Conv-block and a classification task without requiring extra parameters to be trained. Therefore, it regularizes the network in order to prevent overfitting. Furthermore, one of the authors of the present study has proven that the strategy of replacing the fully connected layer with a GAP layer is efficient [12]. In this study, we introduced a GAP layer into the CNN model to process medical signals.

2.4.3. Addressing the Problem of Class Imbalance

For binary classification tasks, in the conventional CNN, the feature nodes generated by the fully connected layer are fed into the classification layer to output a probability. When the model calculates the loss value, each category has an equal loss weight, which may lead the model to predict the input as the class with the larger number of samples when performing a classification task with a class imbalance. However, in clinical practice, it is common for the number of normal samples to be much larger than that of abnormal samples. So, in clinical applications of deep learning algorithms, it is essential to deal with the problem of class imbalance.

To solve the problem of class imbalance for normal and abnormal heart sound samples, in this study, we set different class weights in the loss function, that is, the model applies different penalties to classes with mispredictions. The new weight for each class is defined in Equation (14). It is clear that abnormal samples will acquire a larger weight when they are incorrectly predicted.

$$class_i_weight = \frac{1}{2} \times \frac{Number\ of\ training\ samples}{Number\ of\ class_i}, i = 0, 1. \quad (14)$$

3. Results

3.1. Performance of the Normal and Abnormal Heart Sound Classification

We applied stratified five-fold cross-validation to the heart sound dataset described in Section 2.1 in order to train and test the proposed model. The dataset was split into five equal-sized subsets and the proportion of each part was the same as that of the original dataset. Four subsets were used to train and validate the model, while the members of the remaining subset were retained as test samples. This process was repeated five times.

The sensitivity, specificity, MCC [42] and mean accuracy (Macc) for each fold of the cross-validation were used to evaluate the performance of the proposed model. In clinical applications, sensitivity measures the proportion of sick people who are correctly identified as positive. Specificity measures the percentage of healthy people who are correctly identified as negative. MCC can produce a truthful score for binary classification tasks or multi-class prediction [43]. Macc is not sensitive to the numbers of classes, so it can be used to indicate the performance of the classifier. Because the classification task involved a class imbalance, we applied the arithmetic mean of sensitivity and specificity, that is, the mean accuracy (Macc), to eliminate any reported high virtual accuracy. The classification performances of the proposed method are presented in Table 2. It can be observed that the best outcomes obtained from fourth-fold are 94%.

Table 2. Classification performance of the proposed method based on five-fold cross-validation.

Criteria	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean
Sensitivity	0.919	0.838	0.741	0.943	0.909	0.87
Specificity	0.735	0.92	0.944	0.937	0.795	0.866
MCC	0.593	0.797	0.703	0.906	0.606	0.721
Macc	0.827	0.879	0.8425	0.94	0.852	0.868

3.2. Performance Comparison between the Fully Connected Layer and the Global Average Pooling Layer

In the proposed method, the fully connected layers of the conventional CNN were replaced by a GAP layer to avoid overfitting and to improve the model's performance. To demonstrate the advantages of using a GAP layer, we performed another experiment on the model with one fully connected layer followed by the dropout strategy with a dropout rate of 0.25. The comparison of the average performances, in terms of five-fold cross-validation, between the fully connected layer and the global average pooling layer is depicted in Figure 4.

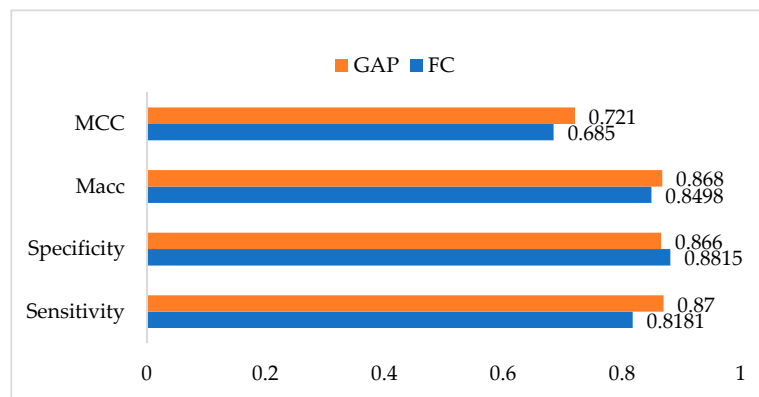


Figure 4. Performance comparison between the fully connected layer and the global average pooling layer. FC, fully connected layer.

Figure 4 shows that the proposed model with a GAP layer is more powerful for normal and abnormal heart sound classification than the model with a fully connected layer. The mean accuracy and sensitivity of the former are approximately 2 percentage points and 6 percentage points higher than the latter, respectively. The MCC observed from the proposed model is about 4 percentage points higher than the model with fully connected layer. However, the specificity of the fully connected layer is better than that of the GAP layer. From a clinical standpoint, sensitivity is more important than specificity, so the classification results from the model with a GAP layer are promising.

3.3. Performance Comparison between the Default Loss Function and the Loss Function with Class Weights

We set class weights in the loss function during the training process to tackle the problem of class imbalance and to improve the classifier’s Macc, especially for abnormal heart sounds. The same stratified five-fold cross-validation was applied to evaluate the performance of the training model with class weights and without class weights. The values of class weights for the normal and abnormal classes were calculated using Equation (14). For each fold of the cross-validation, the total number of training samples was 2523 with 1991 normal samples and 532 abnormal samples. Therefore, the class weights for the normal and abnormal samples were 0.63 and 2.37, respectively. The average performances are shown in Figure 5.



Figure 5. Performance comparison between the default loss function and loss function with class weights. The default loss function is with the equal class weights for normal and abnormal samples, while the class weights-based method sets 2.73 for abnormal samples and a small value 0.63 to the normal samples.

As shown in Figure 5, the classification mean accuracy obtained by setting the class weights in the loss function during the training process is 5 percentage points higher than that of the default equal weights loss function. The sensitivity of the model based on the class weights loss function is much higher than that of the default loss function (up to 16 percentage points). The default equal weights loss function obtained a specificity of about 93%, which is higher than that of class weights-based loss function. Besides, the proposed model with different class weights obtained MCC of 72.1% which is higher than the default equal class weights. With the improvement of the sensitivity, the specificity of the model is still above 85%, which indicates the validity of this method.

4. Discussion

In this study, we presented a method for recognizing normal and abnormal heart sounds that uses an ensemble of a feature engineering method and a deep learning algorithm. The outcomes from the proposed method outperform the reported best accuracy (0.8602) released by PhysioNet/Computing in Cardiology Challenge2016. In this study, we have also explored the challenges that were mentioned in the introduction.

Regarding the challenge of feature types, we extracted 497 features from eight domains based on our previous work. To the best of our knowledge, this feature set contains almost all of heart sounds features that have been referred to in the existing literature. We hope that these features will be helpful for the research of heart sounds or heart diseases. Regarding the method challenge, we explored a new method that combines a multi-features and deep learning method. For deep learning applications on temporal series, such as series of electroencephalogram (EEG), electrocardiograph (ECG) or Electromyography (EMG) signals, which have a low signal-to-noise (SNR), our proposed method may help with event detection or signal classification involving a combination of manually extracted features and deep learning algorithms. To some extent, the manual extraction of features is subjective. Combining automatically extracted features with deep learning algorithms, may make up the flaw of manually extracted features in some classification tasks. In addition, we introduced a GAP layer into the CNN model to improve the classifier's performance as reflected in Figure 4. The use of a GAP layer rather than a fully connected layer allows us to avoid overfitting when the number of PCG recordings is insufficient for a deep learning algorithm. On the class imbalance issue, which commonly arises in clinical practice, we adopted the strategy of setting class weights during the training process to solve the problems with class imbalance. The results shown in Figure 5 prove the effectiveness of this strategy.

Besides, as the different metrics used in other research, it is confusion whether we should adopt the usually used accuracy (Acc) or the Macc as the final metric. In this study we compare the Acc and Macc in the same model as presented in Figure 6. We can get that—(1) when the specificity of the model is large, the Acc of the model is higher than the Macc and (2) the Macc is bigger than Acc when the sensitivity of the model is large. We believe Acc is not enough for the evaluation of the model, especially for the task of class imbalance. While, Macc is defined on the sensitivity and specificity, so it is suitable as an important metric.

Although we have used the proposed method to achieve good performance on the heart sound dataset, there are some limitations to this research. The designed CNN was applied to manually extracted features in order to select features and classify normal and abnormal heart sounds. It has been the trend to perform classification or detection tasks that involve medical images or signals using end-to-end deep learning methods. However, in our experiments, the CNN model's performance was unsatisfactory when we used raw heart sounds as the input. On the other hand, the proposed method was able to classify normal and abnormal heart sounds, without the ability to recognize the categories of the abnormal heart sounds, which is crucial for the clinical diagnosis of related heart diseases. In further work, we will focus on the use of an end-to-end deep learning method for the classification of heart sounds and explore methods of mixing manual features and automatically extracted features in EEG signals for sleep staging.

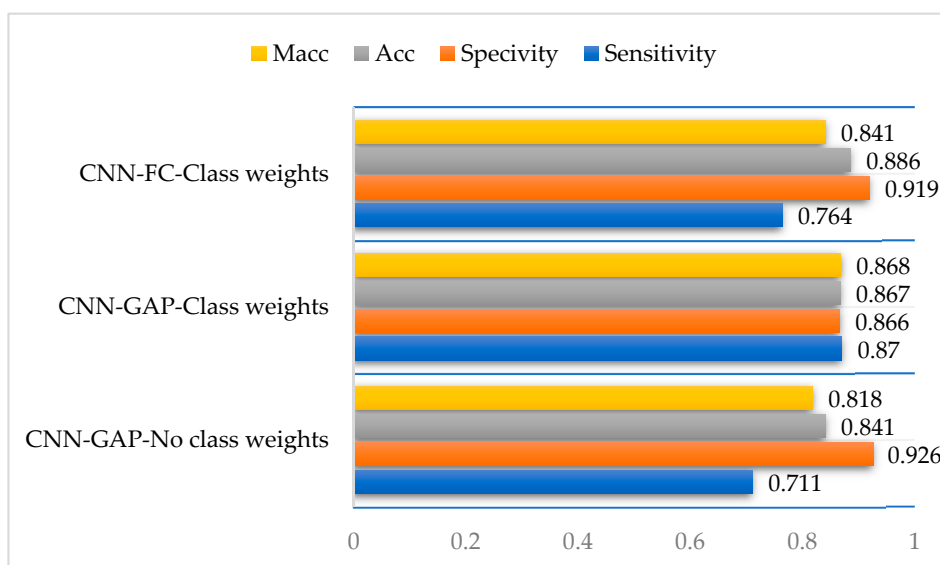


Figure 6. The comparison between the Acc and Macc in three different models, respectively.

5. Conclusions

In this study, we proposed a method for classifying normal and abnormal heart sounds. Multiple features extracted from eight domains were fed into the designed CNN model, which was used to select features and classify heart sounds. A GAP layer was introduced into the designed CNN in order to obtain global information about the feature maps to improve the classification performance. This was also shown to reduce the model's complexity and avoid overfitting. Class weights were set during the training process to solve the problem of class imbalance, which is an issue that commonly arises during the processing of medical images and signals. The results obtained in this study have demonstrated the efficacy of the proposed method.

Author Contributions: Conceptualization, F.C. and H.T.; methodology, F.L. and H.T.; software, F.L.; validation, F.L. and S.S.; formal analysis, F.L.; investigation, F.L. and H.T.; resources, H.T.; data curation, H.T.; writing—original draft preparation, F.L.; writing—review and editing, F.C. and K.M.; visualization, F.L.; supervision, F.C.; project administration, F.C. and H.T. and K.M.; funding acquisition, F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 91748105&81471742; the Fundamental Research Funds for the Central Universities [DUT2019] in Dalian University of Technology in China.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Note: Because there are a lot of operations of mean and standard deviation, for simplicity and clarity, in the second column of all the tables, “m” and “sd” are used to represent mean and standard deviation respectively. Furthermore, “Sys” and “Dia” denote systole and diastole states separately.

Table A1. The summary of the features extracted from time domain.

Index	Features	Concepts
1–2	m_RR sd_RR	mean and standard deviation of RR intervals
3–4	m_IntS1 sd_IntS1	mean and standard deviation of S1 intervals
5–6	m_IntS2 sd_IntS2	mean and standard deviation of S2 intervals
7–8	m_IntSys sd_IntSys	mean and standard deviation of systolic intervals
9–10	m_IntDia sd_IntDia	mean and standard deviation of diastolic intervals
11–12	m_Ratio_SysRR sd_Ratio_SysRR	mean and standard deviation of the ratio of systolic interval to RR interval of each heartbeat
13–14	m_Ratio_DiaRR sd_Ratio_DiaRR	mean and standard deviation of ratio of diastolic interval to RR interval of each heartbeat
15–16	m_Ratio_SysDia sd_Ratio_SysDia	mean and standard deviation of the ratio of systolic to diastolic interval of each heartbeat
17–18 ¹	m_Ratio_S1RR sd_Ratio_S1RR	mean and standard deviation of the ratio of S1 interval to RR interval of each heartbeat
19–20 ¹	m_Ratio_S2RR sd_Ratio_S2RR	mean and standard deviation of the ratio of S2 interval to RR interval of each heartbeat

¹ The authors add another four features in time domain.

Table A2. The summary of the amplitude features extracted from four states.

Index	Features	Concepts
1–2	m_Amp_SysS1 sd_Amp_SysS1	mean and standard deviation of the ratio of the mean absolute amplitude during systole to that during S1 in each heartbeat
3–4	m_Amp_DiaS2 sd_Amp_DiaS2	mean and standard deviation of the ratio of the mean absolute amplitude during diastole to that during S2 in each heartbeat
5–6	m_Amp_S1S2 sd_Amp_S1S2	mean and standard deviation of the ratio of the mean absolute amplitude during S1 to that during the S2 in each heartbeat
7–8	m_Amp_S1Dia sd_Amp_S1Dia	mean and standard deviation of the ratio of the mean absolute amplitude during S1 to that during diastole in each heartbeat
9–10	m_Amp_SysDia sd_Amp_SysDia	mean and standard deviation of the ratio of the mean absolute amplitude during systole to that during diastole in each heartbeat
11–12	m_Amp_S2Sys sd_Amp_S2Sys	mean and standard deviation of the ratio of the mean absolute amplitude during S2 to that during systole in each heart beat

Table A3. The summary of the energy ration features extracted between any two states.

Index	Features	Concepts
1–42	Ratio_energy_band _i	the energy ratios of the band-passed signal to the original one
43–44	m_Ratio_energy_S1_cycle sd_Ratio_energy_S1_cycle	mean and standard deviation of Ratio_energy_S1_cycle
45–46	m_Ratio_energy_S2_cycle sd_Ratio_energy_S2_cycle	mean and standard deviation of Ratio_energy_S2_cycle
47–48	m_Ratio_energy_Sys_cycle sd_Ratio_energy_Sys_cycle	mean and standard deviation of Ratio_energy_Sys_cycle
49–50	m_Ratio_energy_Dia_cycle sd_Ratio_energy_Dia_cycle	mean and standard deviation of Ratio_energy_Dia_cycle
51–52	m_Ratio_energy_S1_S2 sd_Ratio_energy_S1_S2	mean and standard deviation of Ratio_energy_S1_S2
53–54	m_Ratio_energy_S1_Sys sd_Ratio_energy_S1_Sys	mean and standard deviation of Ratio_energy_S1_Sys
55–56	m_Ratio_energy_S1_Dia sd_Ratio_energy_S1_Dia	mean and standard deviation of Ratio_energy_S1_Dia
57–58	m_Ratio_energy_S2_Sys sd_Ratio_energy_S2_Sys	mean and standard deviation of Ratio_energy_S2_Sys
59–60	m_Ratio_energy_S2_Dia sd_Ratio_energy_S2_Dia	mean and standard deviation of Ratio_energy_S2_Dia
61–62	m_Ratio_energy_Dia_Sys sd_Ratio_energy_Dia_Sys	mean and standard deviation of Ratio_energy_Dia_Sys

Table A4. The summary of the skewness and kurtosis features extracted from four states.

Index	Skewness Features	Concepts
1–2	m_skewness_S1 sd_skewness_S1	mean and standard deviation of the skewness of S1 state
3–4	m_kurtosis_S1 sd_kurtosis_S1	mean and standard deviation of the kurtosis of S1 state
5–6	m_skewness_S2 sd_skewness_S2	mean and standard deviation of the skewness of S2 state
7–8	m_kurtosis_S2 sd_kurtosis_S2	mean and standard deviation of the kurtosis of S2 state
9–10	m_skewness_Sys sd_skewness_Sys	mean and standard deviation of the skewness of systole state
11–12	m_kurtosis_Sys sd_kurtosis_Sys	mean and standard deviation of the kurtosis of systole state
13–14	m_skewness_Dia sd_skewness_Dia	mean and standard deviation of the skewness of diastole state
15–16	m_kurtosis_Dia sd_kurtosis_Dia	mean and standard deviation of the kurtosis of diastole state

Table A5. The summary of the features in Cyclostationary domain.

Index	Features	Concepts
1–2	m_cyclostationarity sd_cyclostationarity	mean and standard deviation of the degree of cyclostationarity
3–4	m_peak_sharpness sd_peak_sharpness	mean and standard deviation of the sharpness measure

Table A6. The summary of the features in entropy domain.

Index	Features	Concepts
1	sample_entropy_one	sample entropy of a PCG recording
2	fuzzy_measure_entropy_one	fuzzy measure entropy of a PCG recording
3	sample_entropy_S1	sample entropy of jointed S1 state
4	fuzzy_measure_entropy_S1	fuzzy measure entropy of jointed S1 state
5	sample_entropy_S2	sample entropy of jointed S2 state
6	fuzzy_measure_entropy_S2	fuzzy measure entropy of jointed S2 state
7	sample_entropy_Sys	sample entropy of jointed systole state
8	fuzzy_measure_entropy_Sys	fuzzy measure entropy of jointed systole state
9	sample_entropy_Dia	sample entropy of jointed diastole state
10	fuzzy_measure_entropy_Dia	fuzzy measure entropy of jointed diastole state

References

1. Medina-Leyte, D.J.; Domínguez-Pérez, M.; Mercado, I.; Villarreal-Molina, M.T.; Jacobo-Albavera, L. Use of Human Umbilical Vein Endothelial Cells (HUVEC) as a Model to Study Cardiovascular Disease: A Review. *Appl. Sci.* **2020**, *10*, 938. [\[CrossRef\]](#)
2. Perpetuini, D.; Chiarelli, A.M.; Cardone, D.; Rinella, S.; Massimino, S.; Bianco, F.; Bucciarelli, V.; Vinciguerra, V.; Fallica, G.; Percivalle, V.; et al. Photoplethysmographic Prediction of the Ankle-Brachial Pressure Index through a Machine Learning Approach. *Appl. Sci.* **2020**, *10*, 2137. [\[CrossRef\]](#)
3. Luisada, A.A.; Liu, C.K.; Aravanis, C.; Testelli, M.; Morris, J. On the mechanism of production of the heart sounds. *Am. Heart J.* **1958**, *55*, 383–399. [\[CrossRef\]](#)
4. Wang, F.; Syeda-Mahmood, T.; Beymer, D. Finding disease similarity by combining ECG with heart auscultation sound. *Comput. Cardiol.* **2007**, *34*, 261–264.
5. Roy, D.; Sargeant, J.; Gray, J.; Hoyt, B.; Allen, M.; Fleming, M. Helping family physicians improve their cardiac auscultation skills with an interactive CD-ROM. *J. Contin. Educ. Health Prof.* **2002**, *22*, 152–159. [\[CrossRef\]](#)
6. Gerbarg, D.S.; Taranta, A.; Spagnuolo, M.; Hofler, J.J. Computer analysis of phonocardiograms. *Prog. Cardiovasc. Dis.* **1963**, *5*, 393–405. [\[CrossRef\]](#)
7. Tang, H.; Chen, H.; Li, T.; Zhong, M. Classification of Normal/Abnormal Heart Sound Recordings Based on Multi-Domain Features and Back Propagation Neural Network. In Proceedings of the 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016; Volume 43, pp. 593–596.
8. Tang, H.; Dai, Z.; Jiang, Y.; Li, T.; Liu, C. PCG Classification Using Multidomain Features and SVM Classifier. *Biomed Res. Int.* **2018**, *2018*, 1–14. [\[CrossRef\]](#)
9. Son, G.Y.; Kwon, S. Classification of heart sound signal using multiple features. *Appl. Sci.* **2018**, *8*, 2344. [\[CrossRef\]](#)
10. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 7553. [\[CrossRef\]](#)
11. Pan, H.; Wang, B.; Jiang, H. Deep Learning for Object Saliency Detection and Image Segmentation. *arXiv* **2015**, arXiv:1505.01173.
12. Shang, S.; Long, L.; Lin, S. Automatic Zebrafish Egg Phenotype Recognition from Bright-Field Microscopic Images Using Deep Convolutional Neural Network. *Appl. Sci.* **2019**, *9*, 3362. [\[CrossRef\]](#)

13. Abdel-Hamid, O.; Mohamed, A.R.; Jiang, H.; Deng, L.; Penn, G.; Yu, D. Convolutional neural networks for speech recognition. *IEEE Trans. Audio Speech Lang. Process.* **2014**, *22*, 1533–1545.
14. Yildirim, Ö. A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification. *Comput. Biol. Med.* **2018**, *96*, 189–202. [[CrossRef](#)] [[PubMed](#)]
15. Tschannen, M.; Kramer, T.; Marti, G.; Heinzmann, M.; Wiatowski, T. Heart Sound Classification Using Deep Structured Features. In Proceedings of the 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016; Volume 43, pp. 565–568.
16. Nilanon, T.; Yao, J.; Hao, J.; Purushotham, S.; Liu, Y. Normal/Abnormal Heart Sound Recordings Classification Using Convolutional Neural Network. In Proceedings of the 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016; Volume 43, pp. 585–588.
17. Rubin, J.; Abreu, R.; Ganguli, A.; Nelaturi, S.; Matei, I.; Sricharan, K. Recognizing abnormal heart sounds using deep learning. *CEUR Workshop Proc.* **2017**, *1891*, 13–19.
18. Maknickas, V.; Maknickas, A. Recognition of normal-abnormal phonocardiographic signals using deep convolutional neural networks and mel-frequency spectral coefficients. *Physiol. Meas.* **2017**, *38*, 1671–1684. [[CrossRef](#)] [[PubMed](#)]
19. Potes, C.; Parvaneh, S.; Rahman, A.; Conroy, B. Ensemble of Feature-Based and Deep Learning-Based Classifiers for Detection of Abnormal Heart Sounds. In Proceedings of the 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016; Volume 43, pp. 621–624.
20. Meintjes, A.; Lowe, A.; Legget, M. Fundamental Heart Sound Classification Using the Continuous Wavelet Transform and Convolutional Neural Networks. In Proceedings of the IEEE Engineering in Medicine and Biology Society, EMBS, Honolulu, HI, USA, 18–21 July 2018; pp. 409–412.
21. Poggio, T.; Kawaguchi, K.; Liao, Q.; Miranda, B.; Rosasco, L.; Boix, X.; Hidary, J.; Mhaskar, H. Theory of Deep Learning III: Explaining the non-overfitting puzzle. *arXiv* **2017**, arXiv:1801.00173.
22. Bizzego, A.; Bussola, N.; Salvalai, D.; Chierici, M.; Maggio, V.; Jurman, G.; Furlanello, C. Integrating Deep and Radiomics Features in Cancer Bioimaging. In Proceedings of the 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Siena, Italy, 9–11 July 2019.
23. Huang, X.; Maier, A.; Hornegger, J.; Suykens, J.A.K. Indefinite kernels in least squares support vector machines and principal component analysis. *Appl. Comput. Harmon. Anal.* **2017**, *43*, 162–172. [[CrossRef](#)]
24. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2003**, *3*, 18–22.
25. Freund, Y.; Schapire, R.E. Experiments with a New Boosting Algorithm. In Proceedings of the 13th International Conference on Machine Learning, Bari, Italy, 3–6 June 1996; Volume 96, pp. 148–156.
26. Latif, S.; Usman, M.; Rana, R.; Qadir, J. Phonocardiographic Sensing Using Deep Learning for Abnormal Heartbeat Detection. *IEEE Sens. J.* **2018**, *18*, 9393–9400. [[CrossRef](#)]
27. Kim, H.; Jeong, Y.S. Sentiment classification using Convolutional Neural Networks. *Appl. Sci.* **2019**, *9*, 1–14.
28. Classification of Heart Sound Recordings—The PhysioNet Computing in Cardiology Challenge 2016. Available online: <https://physionet.org/challenge/2016/> (accessed on 4 March 2016).
29. Liu, C.; Springer, D.; Li, Q.; Moody, B.; Juan, R.A.; Chorro, F.J.; Castells, F.; Roig, J.M.; Silva, I.; Johnson, A.E.W.; et al. An open access database for the evaluation of heart sound algorithms. *Physiol. Meas.* **2016**, *37*, 2181–2213. [[CrossRef](#)] [[PubMed](#)]
30. Dominguez-Morales, J.P.; Jimenez-Fernandez, A.F.; Dominguez-Morales, M.J.; Jimenez-Moreno, G. Deep Neural Networks for the Recognition and Classification of Heart Murmurs Using Neuromorphic Auditory Sensors. *IEEE Trans. Biomed. Circuits Syst.* **2018**, *12*, 24–34. [[CrossRef](#)] [[PubMed](#)]
31. Tang, H.; Chen, H.; Li, T. Discrimination of aortic and pulmonary components from the second heart sound using respiratory modulation and measurement of respiratory split. *Appl. Sci.* **2017**, *7*, 690. [[CrossRef](#)]
32. Schmidt, S.E.; Holst-Hansen, C.; Graff, C.; Toft, E.; Struijk, J.J. Segmentation of heart sound recordings by a duration dependent Hidden Markov Model. *Physiol. Meas.* **2010**, *31*, 513–529. [[CrossRef](#)] [[PubMed](#)]
33. Springer, D.B.; Tarassenko, L.; Clifford, G.D. Logistic regression-HSMM-based heart sound segmentation. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 822–832. [[CrossRef](#)]
34. Sakamoto, T.; Kusukawa, R.; MacCanon, D.M.; Luisada, A.A. Hemodynamic Determinants of the Amplitude of the First Heart Sound. *Circ. Res.* **1965**, *16*, 45–57. [[CrossRef](#)]
35. Sakamoto, T.; Kusukawa, R.; MacCanon, D.M.; Luisada, A.A. First heart sound amplitude in experimentally induced alternans. *Dis. Chest* **1966**, *50*, 470–475. [[CrossRef](#)]

36. Boucheron, L.E.; De Leon, P.L.; Sandoval, S. Low bit-rate speech coding through quantization of mel-frequency cepstral coefficients. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 610–619. [[CrossRef](#)]
37. Li, T.; Tang, H.; Qiu, T.; Park, Y. Best subsequence selection of heart sound recording based on degree of sound periodicity. *Electron. Lett.* **2011**, *47*, 841–843. [[CrossRef](#)]
38. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
39. Han, J.; Moraga, C. The influence of the sigmoid function parameters on the speed of backpropagation learning. In *International Workshop on Artificial Neural Networks*; Mira, J., Sandoval, F., Eds.; Springer: Berlin/Heidelberg, Germany, 1995; pp. 195–201.
40. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
41. Lin, M.; Chen, Q.; Yan, S. Network In Network. *arXiv* **2013**, arXiv:1312.4400.
42. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 1–13. [[CrossRef](#)] [[PubMed](#)]
43. Jurman, G.; Riccadonna, S.; Furlanello, C. A comparison of MCC and CEN error measures in multi-class prediction. *PLoS ONE* **2012**, *7*, e41882. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).