

Article

Automatic Classification of Normal–Abnormal Heart Sounds Using Convolution Neural Network and Long-Short Term Memory

Ding Chen ¹, Weipeng Xuan ^{1,*}, Yexing Gu ¹, Fuhai Liu ^{1,2}, Jinkai Chen ¹, Shudong Xia ³, Hao Jin ⁴, Shurong Dong ⁴ and Jikui Luo ^{1,4,*}

- ¹ Ministry of Education Key Laboratory of RF Circuits and Systems, College of Electronics & Information, Hangzhou Dianzi University, Hangzhou 310018, China; 192040216@hdu.edu.cn (D.C.); 18362057055@163.com (Y.G.); techliu1123@hdu.edu.cn (F.L.); chenjk09@hdu.edu.cn (J.C.)
- ² Special Equipment Institute, Hangzhou Vocational & Technical College, Hangzhou 310018, China
- ³ The Fourth Affiliated Hospital Zhejiang University School of Medicine, Yiwu 322000, China; shystone@zju.edu.cn
- ⁴ Key Laboratory of Advanced Micro/Nano Electronic Devices & Smart Systems of Zhejiang, College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou 310058, China; hjin@zju.edu.cn (H.J.); dongshurong@zju.edu.cn (S.D.)
- * Correspondence: xuanweipeng@hdu.edu.cn (W.X.); jackluo@zju.edu.cn (J.L.)

Abstract: The phonocardiogram (PCG) is an important analysis method for the diagnosis of cardiovascular disease, which is usually performed by experienced medical experts. Due to the high ratio of patients to doctors, there is a pressing need for a real-time automated phonocardiogram classification system for the diagnosis of cardiovascular disease. This paper proposes a deep neural-network structure based on a one-dimensional convolutional neural network (1D-CNN) and a long short-term memory network (LSTM), which can directly classify unsegmented PCG to identify abnormal signal. The PCG data were filtered and put into the model for analysis. A total of 3099 pieces of heart-sound recordings were used, while another 100 patients' heart-sound data collected by our group and diagnosed by doctors were used to test and verify the model. Results show that the CNN-LSTM model provided a good overall balanced accuracy of 0.86 ± 0.01 with a sensitivity of 0.87 ± 0.02 , and specificity of 0.89 ± 0.02 . The F1-score was 0.91 ± 0.01 , and the receiver-operating characteristic (ROC) plot produced an area under the curve (AUC) value of 0.92 ± 0.01 . The sensitivity, specificity and accuracy of the 100 patients' data were 0.83 ± 0.02 , 0.80 ± 0.02 and 0.85 ± 0.03 , respectively. The proposed model does not require feature engineering and heart-sound segmentation, which possesses reliable performance in classification of abnormal PCG; and is fast and suitable for real-time diagnosis application.

Keywords: heart sound; phonocardiogram; deep neural network; machine learning



Citation: Chen, D.; Xuan, W.; Gu, Y.; Liu, F.; Chen, J.; Xia, S.; Jin, H.; Dong, S.; Luo, J. Automatic Classification of Normal–Abnormal Heart Sounds Using Convolution Neural Network and Long-Short Term Memory. *Electronics* **2022**, *11*, 1246. <https://doi.org/10.3390/electronics11081246>

Academic Editors: Andrea Prati, Luis Javier García Villalba and Vincent A. Cicirello

Received: 24 March 2022

Accepted: 11 April 2022

Published: 14 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cardiovascular disease (CVD) is one of the main causes of death worldwide. According to the statistics provided by WHO in 2019, approximately 17.9 million people die from CVD every year worldwide [1]. Early diagnosis of CVD allows prevention measures and medications to be taken and saves lives. Many technologies have been developed for the diagnosis of CVD, such as Coronary Computed Tomography (CT) and Echocardiography. Another method of CVD diagnosis is through the analysis of heart sounds recorded. Heart sound is a kind of heart-structure vibration caused by blood flow in the cardiovascular system [2], and specific heart sounds can reflect the health status of the heart. The phonocardiogram (PCG) is a graphical representation of heart-sound recording, and has been widely used for CVD diagnosis. Clinicians always use a stethoscope to listen to heart sounds of the patient, and then diagnose whether the patient has a cardiovascular disease or not based

on the PCG information obtained. This process is time-consuming and requires the doctor to have a wealth of experience.

An automated system can be used as an auxiliary tool for doctors to perform diagnoses of heart problems and reduce their burden. Numerous methods have been proposed to study PCG signals for the diagnosis of heart problems based on heart-sound segmentation and predefined manual features. Heart-sound segmentation can determine the systolic and diastolic regions of the PCG signal, which is convenient for subsequent artificial extraction of relevant features to classify the PCG signal. The feature domains used for classification generally include time, frequency, wavelet, energy, high-order statistics and entropy [3–9]. The methods used for classification include an artificial neural network (ANN) [10–13], a support vector machine (SVM) [14], clustering [15,16] and so on. Gerbarg et al. were the first researchers to attempt the classification of PCG using a threshold-based method [17]. Springer et al. proposed an improved hidden semi-Markov model (HSMM) and improved PCG segmentation performance [18]. Tang et al. extracted up to 515 features from nine different dimensions and used SVM to identify the abnormal signal (murmurs) in the PCG signal [19].

So far, it can be found that most of the automatic heart-sound-classification methods are based on a complex set of features, such as time interval, frequency spectrum of states, state amplitude, energy, frequency spectrum of records, cepstrum, cyclostationarity, high-order statistics and entropy. These features are usually calculated manually. Although the accuracy of the results is high, the extracted features are too complicated, which may lead to subjective bias and variations. Krishnan et al. proposed a deep neural-network (DNN) model without using feature engineering. They divided the original PCG signal into shorter time segments of 6 s epochs. The processed data were then inputted to the proposed DNN architecture for analysis, resulting in an accuracy of 0.8565 with a sensitivity of 0.8673, and a specificity of 0.8475 in detection of abnormal heart sounds [20].

In response to these situations, in this study, we develop an end-to-end DNN for PCG analysis and classification. The research aims to simplify the PCG signal feature engineering and speed up the analysis of PCG recordings, thereby helping cardiologists provide faster treatment plans for patients.

2. Materials and Methods

The workflow of this study is shown in Figure 1. After the PCG signal is preprocessed, it is directly input into classifier for analysis, and then it is identified as to whether it is abnormal or normal.

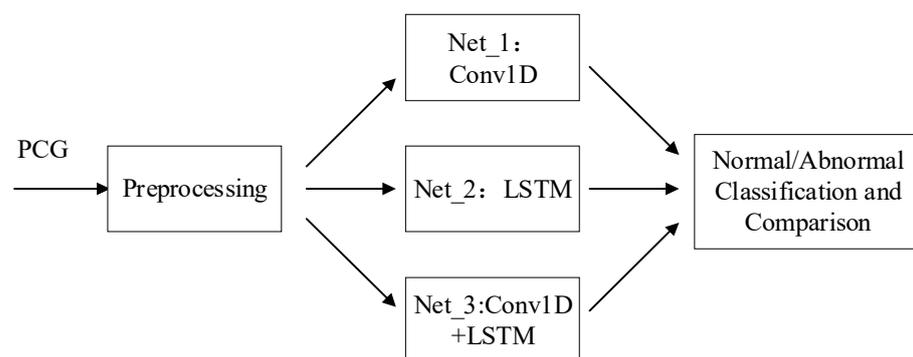


Figure 1. Block diagram of the proposed method. First, the phonocardiogram (PCG) is preprocessed; then, the PCG is fed into the designed model to classify normal and abnormal heart sounds.

2.1. Dataset

This study used the Physionet Challenge 2016 dataset (CinC) to train and validate the PCG classification model. The database was sourced from several contributors around the world and collected at either a clinical or nonclinical environment from both healthy subjects and pathological patients recorded at a sampling frequency of 2000 Hz. It consists of six databases (A through F). The databases contain a total of 3240 PCG recordings in

mV, lasting from 5 s to over 120 s, and the dataset ‘E’ has high background noise. The PCG recordings were collected from different locations on a human body. The PCG signals are divided into two types: normal and abnormal records. Normal records are from healthy people, while abnormal records are from patients being diagnosed with heart disease. Figure 2 shows typical normal and abnormal PCG, respectively. The normal PCG have regular beats with one relatively weak beat between the two strong beats, while there is random distribution of weak sounds between the two strong beats with irregular signal amplitudes for the abnormal heart sounds. More details about the CinC can be obtained from Physionet [21,22]. All PCG signals used in this study are shown in Table 1.

To demonstrate the robustness and generalizability of our method, we also used the ZJU4H PCG dataset from the Fourth Affiliated Hospital of Zhejiang University, School of Medicine as the test dataset. The dataset contains a total of 1075 heart-sound records from 100 patients diagnosed by cardiovascular experts.

Table 1. Summary of the PCG records used for the training and verification of the classification.

| Dataset | Number of PCG | Normal PCG | Abnormal PCG | Training Datasets | Validation Datasets | Test Datasets |
|------------|---------------|------------|--------------|-------------------|---------------------|---------------|
| Dataset-a | 409 | 117 | 292 | 245 | 82 | 82 |
| Dataset-b | 490 | 370 | 102 | 294 | 98 | 98 |
| Dataset-c | 31 | 7 | 24 | 19 | 6 | 6 |
| Dataset-d | 55 | 21 | 28 | 33 | 11 | 11 |
| Dataset-e | 2000 | 1876 | 124 | 1200 | 400 | 400 |
| Dataset-f | 114 | 76 | 34 | 68 | 23 | 23 |
| Total Cinc | 3099 | 2483 | 616 | 1859 | 620 | 620 |

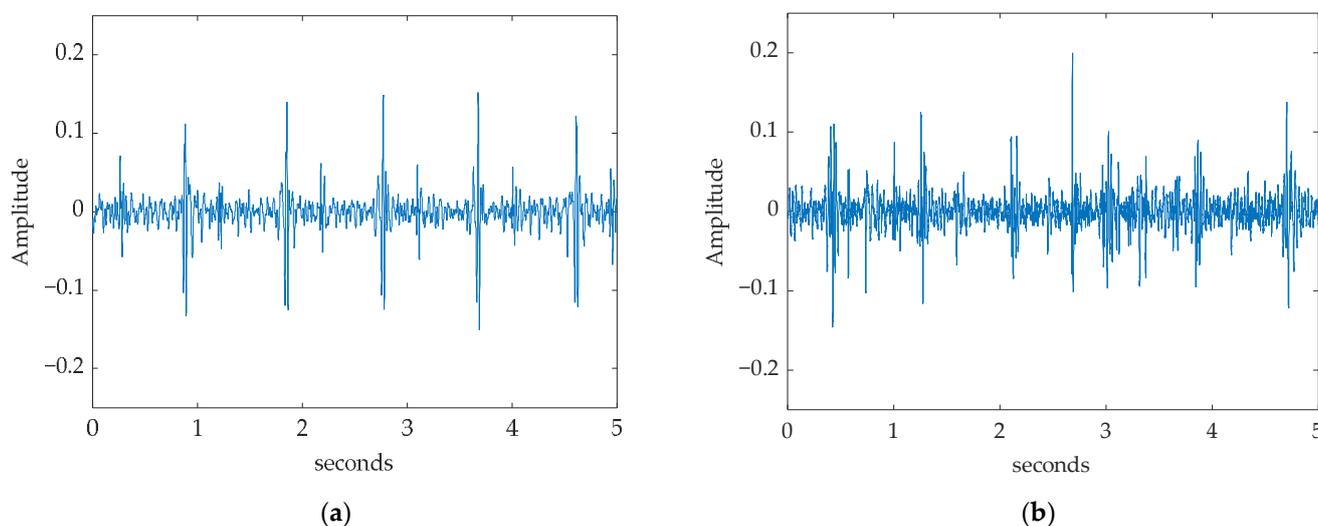


Figure 2. PCG recording of: (a) normal signal; (b) abnormal signal.

2.2. Preprocessing

Since the PCG data used are from different medical institutions, the amplitude values of the sound signals vary significantly, which is difficult for analysis and model establishment. The data were normalized using the min–max normalization. The frequency of PCG was between 1 and 800 Hz, and the frequency for the important signal component was above 20 Hz [21]. Therefore, the Butterworth bandpass filter with a frequency range of 20–800 Hz was used to remove the low- and high-frequency noises. In order to facilitate the signal processing, the signals of different lengths were segmented to the same length of 5 s. This number was selected based on reference [21], which pointed out that it takes at least 5 s of data to detect cardiac abnormalities. The processed signal can be seen as a vector with a dimension of $1 \times 10,000$.

2.3. Model

In this study, the automatic classification of heart sounds can be seen as a binary-classification problem that uses deep neural networks to classify recorded PCG signals into normal and abnormal categories. In this study, we first thought of using CNN to process heart-sound signals. CNN shares convolution kernels, processing high-dimensional data without pressure, and can automatically perform feature extraction. However, it is easy to ignore the correlation between the part and the whole. LSTM is a kind of RNN, which emphasizes the relevance of data, which can solve this problem. Therefore, we combined one-dimensional convolutional neural networks (1D-CNN) and long short-term memory networks (LSTM) to classify heart sounds. We designed three network structures to evaluate the heart-sound classifiers. Net_1 uses 1D-CNN, Net_2 uses LSTM, while Net_3 uses 1D-CNN and LSTM. Table 2 is the summary of the structure and parameters of the three structure models.

Table 2. Summary of the Net_1 architecture evaluated for heart-sound classification.

| Layer (Type) | Params |
|--------------|---|
| Input | [1 × 10,000] PCG time sequence |
| Conv1D | Filters = 64, 1 × 100, ReLU Filters = 32, 1 × 100, ReLU Filters = 32, 1 × 100, ReLU |
| Dropout | dropout rate = 0.5 |
| MaxPooling1D | Pool size = 2, strides = 2 |
| Flatten | |
| Dropout | dropout rate = 0.3 |
| Dense | 1 × 512, ReLU 1 × 128, ReLU 1 × 64, ReLU |
| Output | 1 × 2, Softmax |

2.3.1. Input Layer

We used three different network structures for comparative analysis. The input layer of each network structure was different. The processed PCG signal was 5 s long data, and can be expressed as a vector with a dimension of 1 × 10,000. In Net_1 and Net_2, the input was a PCG time sequence with a dimension of 1 × 10,000. In Net_3, a processed PCG signal was divided into 100 parts. The input was a vector with a dimension of 1 × 100 × 100.

2.3.2. 1D-Convolutional Neural Network

1D-CNN was used in this study for two network structures: Net_1 and Net_3. In Net_1, we used three one-dimensional convolutional layers. The first convolutional layer had 64 convolutional filters, the second convolutional layer had 32 convolutional filters, while the third convolutional layer had 16 convolutional filters. The input sample was 10,000 points. We set the total length of all convolutional filters in Net_1 to 100.

In Net_3, 1D-CNN was used to extract the features of the heart-sound sequence. This part consisted of three one-dimensional convolutional layers. The length of the convolutional filter was 5. The number of filters owned by the three convolutional layers was 8, 4 and 2, respectively. Batch normalization was introduced between each layer to renormalize the output value of the previous layer. We used TimeDistributed to apply 1D-CNN to 100 parts of a PCG signal at the same time.

2.3.3. Pooling Layers

After the features are obtained through 1D-CNN, a pooling layer needs to be used to reduce the dimension of the obtained features to reduce the amount of computation. In

this study, a one-dimensional max-pooling layer was used in Net_1 and Net_3, with both pool size and stride of 2.

2.3.4. Long Short-Term Memory Network

Time-series data at different points in time are interrelated, such as heart sounds; so how do we make the association between the data also being analyzed by neural network? Think about how we humans analyze the associations of various things. The most basic way is to remember what happened before. Recurrent neural network (RNN) will store the analysis results at different time points, and finally will accumulate all the previous results and analyze them together. RNN learns on sequential data; in order to remember these data, RNN will generate memories of previous events like a human. Similarly, they will forget like humans. This will cause gradient-disappearance and gradient-explosion problems.

The long short-term memory network (LSTM) is a kind of RNN that is specially designed to solve the problems in general RNNs. Compared with ordinary RNN, LSTM has three more controllers: input control, output control, and forget control [23]. The three controllers are all based on the original RNN system. If the information at this time is very important to the result, the input control will store this information according to the degree of importance of the information for analysis. If the information at this time changes our thoughts on the previous information, then the control units will forget some of the previous information and replace the current information proportionally. The output control will output the correct information based on these.

In Net_2, we directly used the three LSTM layers to process the data. Detailed parameters can be seen in Table 3.

In Net_3, we divided the heart-sound data records into one hundred segments, and applied TimeDistributed to 100 parts of a PCG signal at the same time. The features extracted from the 1D-CNN will be input to the LSTM for processing, and the processed data will be used as the input of the two dense layers. Table 4 shows the parameters of LSTM.

Table 3. Summary of the Net_2 architectures evaluated for heart-sound classification.

| Layer (Type) | Params |
|--------------|---|
| Input | [1 × 10,000] PCG time sequence |
| LSTM | Units = 512 Units = 256 Units = 128 |
| Dropout | dropout rate = 0.5 |
| Dense | 1 × 64, ReLU |
| Output | 1 × 2, Softmax |

2.3.5. Dense Layers

The network structures used in this study contain dense layers using the ReLU activation function. The parameters of dense layers are shown in Tables 2–4. The output layer of three network structures is two neurons with softmax activation. The result is represented as \hat{y} . \hat{y} is a vector with a 1×2 dimension; \hat{y}_0 represents the probability of abnormal heart sounds; and \hat{y}_1 represents the probability of normal heart sounds. Abnormal heart sounds are represented by 0 (class 0), and normal heart sounds are represented by 1 (class 1). The sum of \hat{y}_0 and \hat{y}_1 is 1. If the probability score of \hat{y}_1 is greater than that of \hat{y}_0 , then the PCG data are classified as normal signals; otherwise, the PCG data are classified as abnormal signals.

Table 4. Summary of the Net_3 architectures evaluated for heart-sound classification.

| Layer (Type) | Params |
|-------------------------------|--|
| Input | [1 × 100 × 100] PCG time sequence |
| Time Distributed Conv1D | Filters = 8, 1 × 5, ReLU Filters = 4, 1 × 5, ReLU Filters = 2, 1 × 1, ReLU |
| Dropout | dropout rate = 0.5 |
| Time Distributed MaxPooling1D | Pool size = 2, strides = 2 |
| Time Distributed Flatten | |
| LSTM | Units = 256 |
| Dense | 1 × 128, ReLU |
| Dropout | dropout rate = 0.3 |
| Output | 1 × 2, Softmax |

2.3.6. Dropout Layers

In order to prevent overfitting, we need dropout layers to randomly discard some neurons. These neurons are not really discarded, but temporarily disabled, which will discard some features and increase the robustness of the model. Dropout layers force a neuron unit to work with other neuron units selected at random, weakening the joint adaptability between neuron nodes and enhancing the generalization ability. The detailed dropout rates are shown in Tables 2–4.

2.4. Class Weight

In the classification model, we often encounter two types of problems. The first one is that misclassification is expensive. The second is that the sample is highly imbalanced. Therefore, in the clinical application of deep-learning algorithms, it is very important to deal with the problem of class imbalance. For example, imagine you have two classes: A and B. Class A is 90% of the dataset and Class B is 10%, but you are most interested in identifying instances of Class B. You could predict Class A every time, which easily achieves 90% accuracy, but it is a useless classifier for your intended use case. This is a common scenario when performing detection, such as detecting malicious content online or disease markers in medical data. The prediction accuracy can be improved by adjusting the degree of influence on the loss function with few categories.

In this study, to solve the problem of class imbalance for normal and abnormal heart-sound samples, we set different class weights in the loss function. Abnormal samples acquire a larger weight when they are incorrectly predicted; that is, the model applies different penalties to classes with mispredictions. The new weight for each class is defined as Equation (1). The number of the class represents the number of categories. In this study, this value is 2, because of a binary classification. Class i means a certain category. Here, the value of i is 0 and 1. 0 represents abnormal heart sounds; 1 represents normal heart sounds. Class weights weight the loss function during training so that the model focuses on samples from underrepresented classes. The class weights will be dynamically adjusted based on the data of each round of training.

$$class_weight_i = \frac{Number\ of\ training\ samples}{Number\ of\ class \times Number\ of\ class_i} \quad (1)$$

2.5. Training and Validation

The proposed classification model is based on supervised learning; each heart sound data has a label, and the weight is updated by minimizing the binary cross entropy to achieve the optimal solution. The learning rate of Adam optimizer [24] is 0.0006. The model

is based on the keras framework and runs on Google Colaboratory, which is a research tool developed by Google, and is mainly used for machine-learning-based R&D. The advantage of Google Colaboratory is that it provides free GPU usage to most of the AI developers. The GPU is NVIDIA TESLA T4, 16GB. We use sensitivity (recall), specificity, balanced accuracy (MAcc), F1-score and receiver-operating characteristic curve (ROC) to evaluate the performance.

3. Results

These performance evaluations of the three network structures are shown in Figures 3 and 4. The performance of Net_1 was not so good; Net_2 produced the highest detection of true negative cases, with a specificity of 0.88; Net_3 provided the best overall balanced performance compared to the other two network architectures. In the training process of Net_3, we set class weights in the loss function to solve the problem of category imbalance.

Figure 4 shows the receiver-operating characteristic curve (ROC). Net_3 with the setting of class weights provided the highest area-under-the-curve (AUC) value of 0.92 ± 0.01 ; Net_1 and Net_2 provided an acceptable AUC value of 0.87 ± 0.02 and 0.74 ± 0.03 for the PCG classification. From the combined view of MAcc, F1-score and ROC, Net_3 with class weight is the best of the four models.

With the setting of the class weights, Net_3 provided the best overall performance compared to the default equal-weight loss function with respect to MAcc of 0.87, F1-score of 0.91 and AUC value of 0.92. Figure 5 shows the training and validation loss and accuracy vs. epoch curve of Net_3. As the number of iterations increases, the loss-function curve will gradually converge. In order to prevent overfitting, we set the number of iterations to 500. The training of the neural networks involved 10 runs. The final result is the average of these 10 runs.

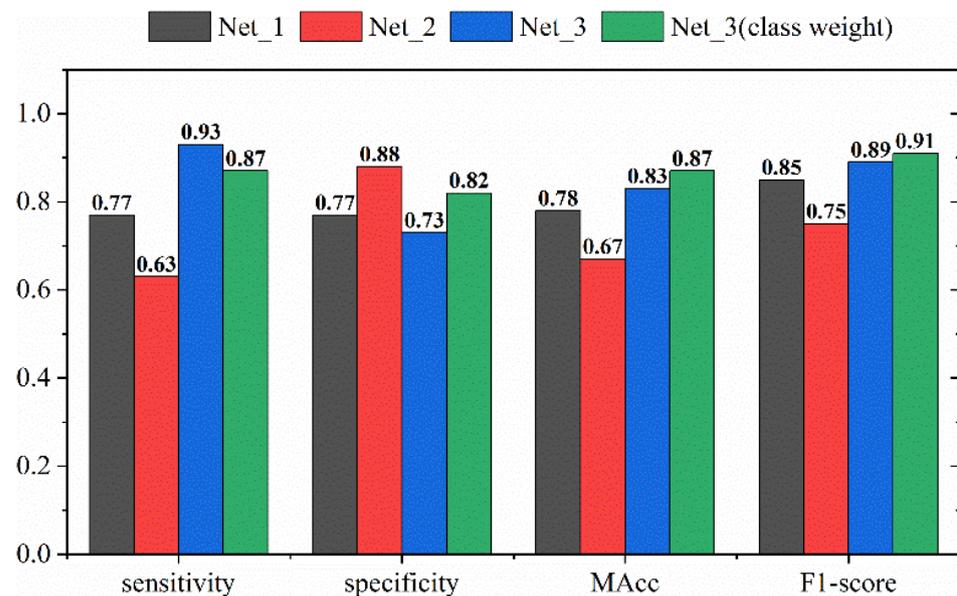


Figure 3. Performance measures of the network architectures for heart-sound classification.

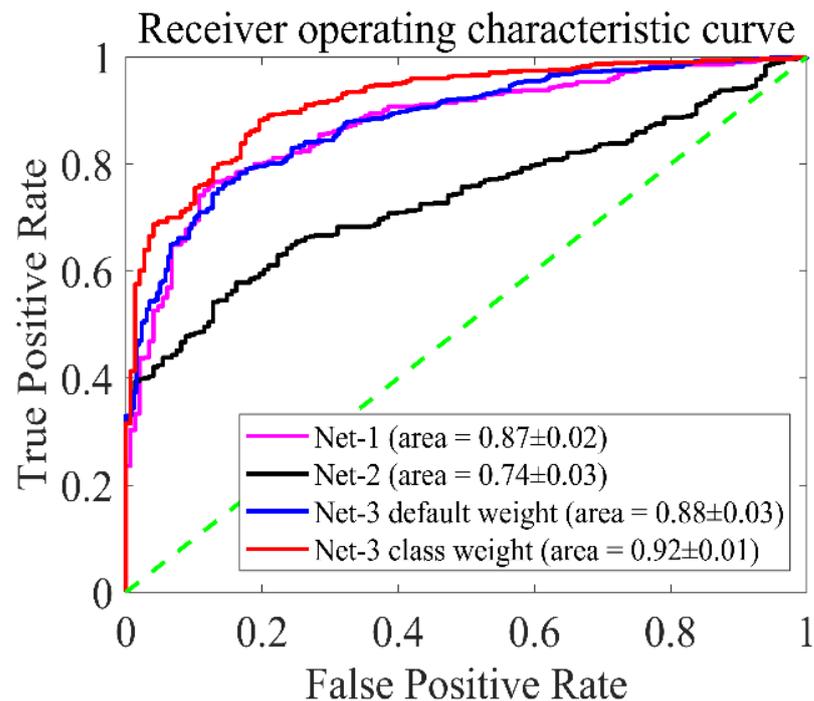


Figure 4. The ROC curve shows the classification performance of the heart-sound classifiers. Area under curve of Net_1 is 0.87 ± 0.02 . Area under curve of Net_2 is 0.74 ± 0.03 . Area under curve of Net_3 with default weight is 0.88 ± 0.03 . Area under curve of Net_3 with the setting of the class weights is 0.92 ± 0.01 .

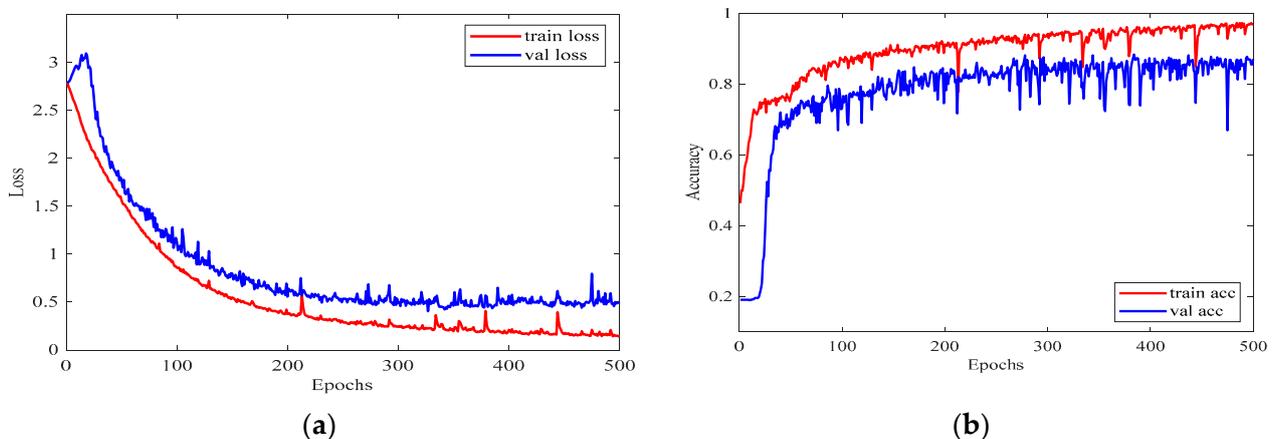


Figure 5. The training and validation loss vs. epoch curve of Net_3 with setting class weight. (a) is training and validation loss vs. epoch; (b) is training and validation accuracy vs. epoch.

4. Discussion

We can see that the three models perform differently for the same data. Net_1 uses CNN. Net_2 uses RNN. Net_3 combines the advantages of these two DNNs and has better performance than Net_1 and Net_2.

So far, most of the classification methods for heart sounds have been manually based on the characteristic features of the projects, and then a variety of machine-learning methods were used for analysis, such as SVM, random forests and k-nearest neighbor. Although these methods have good accuracy for detecting abnormal heart sounds, subjective deviations will inevitably occur. Table 5 compares the method proposed in this study with the existing methods for heart-sound classification. All these methods used the same CinC as a dataset in this study. Masun et al. extracted features from time, frequency and complexity

analysis and achieved 80% accuracy [25]. Li et al. first segmented the heart-sound data, and then extracted as many as 497 features from eight categories, including time, amplitude, energy, frequency, cepstrum, cyclostationary, high-order statistical and cross-entropy features. They then fed these features into CNN and obtained an accuracy of 86.8% [26]. As can be seen from Figure 5, the way of using DNN is better than the method of heart-sound segmentation and manual feature recognition. Because the features are calculated and selected manually, there may be problems of calculation bias and selected features that are not comprehensive enough. The end-to-end DNN is that the model recognizes the features by itself, and does not need to be constructed manually, which greatly simplifies the complexity of the model. By reducing manual preprocessing and subsequent processing, the model can go from the original input to the final output as much as possible, giving the model more space for automatic adjustment according to the data, and increasing the overall fit of the model. Krishnan et al. proposed a deep neural-network model that can directly classify heart-sound data with a classification accuracy of 85.74% [20]. Considering the correlation before and after the heart-sound signal sequence, we used a combination of CNN and LSTM to process the heart sound and obtained a higher accuracy of 86%, with a sensitivity of 87% and specificity of 82%. The F1-score was 91%.

Table 5. Comparison of related work in heart-sound classification based on Physionet Challenge 2016 dataset.

| Related Work | Method | Performance | |
|---|--|--|---------------------------------|
| Masun et al. [25] | Random forest + LogitBoost + Cost-sensitive classifier | Sensitivity: Specificity: MAcc: | 80% 81% 80% |
| Li et al. [26] | CNN | Sensitivity: Specificity: MAcc: | 87% 86% 86% |
| Krishnan et al. [20] | DNN | Sensitivity: Specificity: MAcc: F1-score: AUC: | 87% 85% 86% 85% 86% |
| This study (Net_3 with class weight) | 1D-CNN + LSTM | Sensitivity: Specificity: MAcc: F1-score: AUC: | 87% 82% 86% 91% 92% |

Figure 6 shows the performance of Net_3 with the setting of the class weight under two test sets, CinC 2016 and ZJU4H. The model provided a MAcc of 0.85, with a sensitivity of 0.83 and specificity of 0.80. The F1-score of the model was found to be 0.90. The performance of the two test sets was similar to Net_3 with the setting of the class weight. This proves that Net_3 with the setting of the class weight has good versatility.

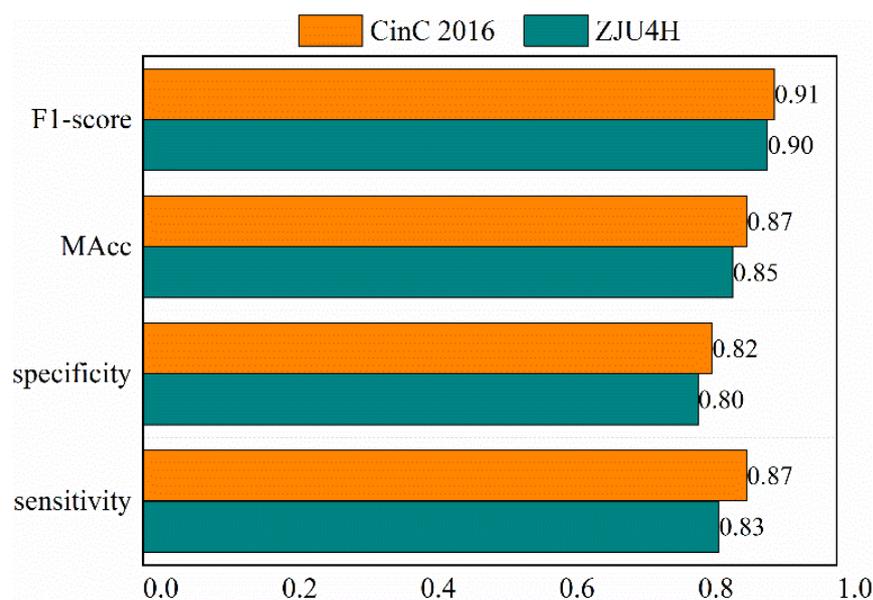


Figure 6. Summary of performance of Net_3 with setting of class weight evaluated using the dataset (ZJU4H) from the Fourth Affiliated Hospital of Zhejiang University, School of Medicine and Physionet Challenge 2016 (CinC 2016).

5. Conclusions

In this study, we propose an end-to-end 1D-CNN-LSTM without PCG segmentation and feature engineering for PCG analysis and classification, and aim to automate the feature-engineering and feature-selection process used in the analysis of the PCG signal and to reduce the analysis time of PCG records for heart-disease identification, thus assisting the cardiologist in providing a faster treatment plan to the patients. This method directly obtains PCG data and classifies them as normal or abnormal heart sounds. In addition, this method does not require presegmentation of PCG sounds into basic heart sounds, and also has good performance for heart-sound data with high-noise components. Moreover, the model has been also verified by the data obtained from the Fourth Affiliated Hospital of Zhejiang University, School of Medicine, which show a good performance. Therefore, the neural-network architecture based on the convolutional neural network and long short-term memory network proposed in this study can be used as a feasible tool to detect abnormal PCG from unsegmented heart-sound signals without any feature-engineering processing.

Of course, there are still some limitations in our work. Compared with the methods based on feature engineering, the end-to-end method proposed by us lacks some interpretabilities. In addition, we achieved the classification of normal and abnormal heart sounds but could not identify a specific heart disease. In the future, we will optimize the method and be able to diagnose specific heart diseases according to the heart-sound signals.

Author Contributions: Conceptualization, D.C. and Y.G.; methodology, D.C. and Y.G.; software, D.C.; validation, D.C., F.L. and Y.G.; resources, S.X.; writing—original draft preparation, D.C.; writing—review and editing, H.J., S.D. and J.L.; project administration, J.C.; funding acquisition, W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key R&D Program of China, grant number 2021YFB3602200; Zhejiang Province Key R & D programs, grant number 2021C05004; National Natural Science Foundation of China, grant number 61827806, 61974037, 61904042; NSFC-Zhejiang Joint Fund for the Integration of Industrialization and information, grant number U1909212 and The APC was funded by Zhejiang University.

Data Availability Statement: Data sharing not applicable.

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

References

1. Cardiovascular Diseases (CVDs). Available online: [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) (accessed on 3 August 2020).
2. Luisada, A.A.; Liu, C.K.; Aravanis, C.; Testelli, M.; Morris, J. On the mechanism of production of heart sounds. *Am. Heart J.* **1958**, *55*, 383–399. [[CrossRef](#)]
3. Goda, M.A.; Hajas, P. Morphological Determination of Pathological PCG Signals by Time and Frequency Domain Analysis. In Proceedings of the 2016 Computing in Cardiology Conference, Vancouver, BC, Canada, 11–14 September 2016.
4. Langley, P.; Murray, A. Abnormal Heart Sounds Detected from Short Duration Unsegmented Phonocardiograms by Wavelet Entropy. In Proceedings of the 2016 Computing in Cardiology Conference, Vancouver, BC, Canada, 11–14 September 2016.
5. Rubin, J.; Abreu, R.; Ganguli, A.; Nelaturi, S.; Matei, I.; Sricharan, K. Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients. In Proceedings of the 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016.
6. Singh-Miller, N.; Singh-Miller, N. Using Spectral Acoustic Features to Identify Abnormal Heart Sounds. In Proceedings of the 2016 Computing in Cardiology Conference, Vancouver, BC, Canada, 11–14 September 2016.
7. Zhang, X.; Liang, Y.; Zhou, J.; Zang, Y. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement* **2015**, *69*, 164–179. [[CrossRef](#)]
8. Plesinger, F.; Viscor, I.; Halamek, J.; Jurco, J.; Jurak, P. Heart sounds analysis using probability assessment. *Physiol. Meas.* **2017**, *38*, 1685–1700. [[CrossRef](#)]
9. Zhang, W.; Han, J.; Deng, S.-W. Heart sound classification based on scaled spectrogram and partial least squares regression. *Biomed. Signal Process. Control* **2017**, *32*, 20–28. [[CrossRef](#)]
10. Durand, L.G.; Blanchard, M. Comparison of pattern recognition methods for computer-assisted classification of spectra of heart sounds in patients with a porcine bioprosthetic valve implanted in the mitral position. *IEEE Trans. Biomed. Eng.* **1990**, *37*, 1121–1129. [[CrossRef](#)]
11. Uğuz, H. A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases. *J. Med. Syst.* **2012**, *36*, 61–72. [[CrossRef](#)]
12. Oelmez, T.; Dokur, Z. Classification of Heart Sounds Using Artificial Neural Network. *Pattern Recognit. Lett.* **2003**, *24*, 617–629. [[CrossRef](#)]
13. Dokur, Z.; Ölmez, T. Heart sound classification using wavelet transform and incremental self-organizing map. *Digit. Signal Process.* **2008**, *18*, 951–959. [[CrossRef](#)]
14. Ari, S.; Hembram, K.; Saha, G. Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier. *Expert Syst. Appl.* **2010**, *37*, 8019–8026. [[CrossRef](#)]
15. Avendao-Valencia, L.D.; Godino-Llorente, J.I.; Blanco-Velasco, M.; Castellanos-Dominguez, G. Feature extraction from parametric time-frequency representations for heart murmur detection. *Ann. Biomed. Eng.* **2010**, *38*, 2716–2732. [[CrossRef](#)]
16. Quiceno-Manrique, A.F.; Godino-Llorente, J.I.; Blanco-Velasco, M.; Castellanos-Dominguez, G. Selection of Dynamic Features Based on Time–Frequency Representations for Heart Murmur Detection from Phonocardiographic Signals. *Ann. Biomed. Eng.* **2010**, *38*, 118–137. [[CrossRef](#)] [[PubMed](#)]
17. Gerbarg, D.S.; Taranta, A.; Spagnuolo, M.; Hofler, J. Computer analysis of phonocardiograms. *Prog. Cardiovasc. Dis.* **1963**, *5*, 393–405. [[CrossRef](#)]
18. Springer, D.B.; Tarassenko, L.; Clifford, G.D. Logistic Regression-HSMM-Based Heart Sound Segmentation. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 822–832. [[CrossRef](#)] [[PubMed](#)]
19. Hong, T.; Ziyin, D.; Yuanlin, J.; Ting, L.; Chengyu, L. PCG Classification Using Multidomain Features and SVM Classifier. *BioMed Res. Int.* **2018**, *2018*, 4205027.
20. Krishnan, P.; Balasubramanian, P.; Umapathy, S. Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network. *Phys. Eng. Sci. Med.* **2020**, *43*, 505–515. [[CrossRef](#)]
21. 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. An open access database for the evaluation of heart sound algorithms. *Physiol. Meas.* **2016**, *37*, 2181–2213. [[CrossRef](#)]
22. Goldberger, A.; Luis, M.; Amaral, N.; Glass, P.; Jeffrey, P.; Hausdorff, J.; Peng, C.-K.; Stanley, H. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation* **2000**, *101*, E215–E220. [[CrossRef](#)]
23. Karim, F.; Majumdar, S.; Darabi, H.; Chen, S. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access* **2018**, *6*, 1662–1669. [[CrossRef](#)]
24. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2016**, arXiv:1609.04747.
25. Homsy, M.N.; Warrick, P. Ensemble methods with outliers for phonocardiogram classification. *Physiol. Meas.* **2017**, *38*, 1631–1644. [[CrossRef](#)]
26. Li, F.; Tang, H.; Shang, S.; Mathiak, K.; Cong, F. Classification of Heart Sounds Using Convolutional Neural Network. *Appl. Sci.* **2020**, *10*, 3956. [[CrossRef](#)]