Article

Application of Heartbeat-Attention Mechanism for Detection of Myocardial Infarction Using 12-Lead ECG Records

Yue Zhang * and Jie Li

The Division of Information Science and Technology, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

* Correspondence: zhangyue@mail.tsinghua.edu.cn

Received: 12 June 2019; Accepted: 6 August 2019; Published: 13 August 2019

Abstract: Early detection and effective treatment of myocardial infarction can prevent the deterioration of ischemic heart disease and greatly reduce the possibility of sudden death. On the basis of standard 12-lead electrocardiogram (ECG) records, this paper proposes a bidirectional, long short-term memory (Bi-LSTM) network with a heartbeat-attention mechanism to effectively and automatically detect myocardial infarction (MI). First, we divide the standard 12-lead ECG records into sliding windows with the same number of heartbeats. Subsequently, we do not use any labels of heartbeats to train the Bi-LSTM network and the heartbeat-attention mechanism is applied to automatically weight the difference between unlabeled heartbeats. Finally, our method is validated by patients' complete ECG records and real labels in the Physikalisch-Technische Bundesanstalt (PTB) diagnostic ECG database. When compared with the same network without the heartbeat-attention mechanism or other existing methods, our method achieves comparable or better performance. The accuracy, sensitivity, and specificity reach 94.77%, 95.58%, and 90.48%, respectively.

Keywords: myocardial infarction; heartbeat-attention mechanism; 12-lead ECG records

1. Introduction

Myocardial infarction (MI) is also called myocardial avascular necrosis. It refers to the necrosis of heart tissue when it is in an ischemic state for a long time, which leads to coronary artery occlusion and significantly reduced ability of the heart to transport blood. This can lead to more severe cases of syncope malignant arrhythmia or sudden death. Clinical manifestations include persistent chest pain and systemic fever. According to electrocardiogram (ECG) signals, there are significant, progressive changes. Myocardial infarction mostly occurs in high-risk groups with coronary heart disease, high blood pressure, or a history of myocardial infarction; it may also occur in healthy people with emotional excitement, an infection, or who are overworked.

In addition, some episodes of myocardial infarction have no specific symptoms, and patients eventually deteriorate to ischemic heart disease without knowing it. As the clinical symptoms are not clear, these patients often do not go to hospital for treatment until the condition worsens. At this stage, the myocardium is extensively necrotic and difficult to repair, which contributes to a significant increase in the risk of recurrent myocardial infarction and sudden death. According to statistics from American Health Association, nearly 750,000 people in the United States have heart attacks each year; only 210,000 of them have recurrent heart attacks [1], suggesting that about 72% of heart attacks are silent. If these patients do not find the episode in time and take appropriate treatment measures, their life will be greatly threatened.

According to World Health Organization standards, the diagnosis of myocardial infarction includes three components: an abnormal electrocardiogram (ECG), such as significant elevation of

ST segment and abnormal Q wave, as shown in Figure 1; abnormal serum enzyme content; and chest pain symptoms [2]. As long as two of the criteria are met, the patient can be diagnosed with myocardial infarction. The location of infarction depends on the coronary artery that has been occluded. This evidence can appear in different ECG leads, which requires simultaneous investigation of all the 12 ECG leads [3].

Figure 1. Electrocardiogram (ECG) signal of myocardial infarction.

With the advancement of electronic technology and big data cloud platforms, portable medical wearable devices have also undergone significant development, making real-time ECG monitoring and automatic detection of myocardial infarction possible. On the basis of the standard 12-lead ECG signals collected by portable medical wearable devices, after amplification and filtering, the automatic detection algorithm gives the diagnosis result of myocardial infarction. If an abnormality occurs, the devices will alert the patient or directly notify the hospital to take appropriate action. In our work, we are committed to developing a practical, medical-grade detection algorithm of myocardial infarction. With the increasing popularity of machine learning, many researchers [3–15] have applied machine learning techniques to the detection of myocardial infarction and achieved promising results.

Reddy M et al. [4] extracted 15-dimensional morphology feature vector from QRS measurement and introduced neural networks into the classification of myocardial infarction, achieving 79% accuracy. Pei Chann C et al. [5] converted ECG signals into a density model and extracted the feature vector using hidden Markov models (HMMs). The myocardial infarction was classified by Gaussian mixture models (GMMs) and their accuracy, sensitivity and specificity achieved 82.50%, 85.71%, and 79.82%, respectively. Safdarian N et al. [6] used two new features, T-wave integral and total integral, in the detection of myocardial infarction; accuracy on the test set reached 94.74%. In a study by Sharma L Net al. [3], a novel feature extraction method, based on multi-scale wavelet energy and multi-scale covariance matrix, was proposed. Their accuracy, sensitivity, and specificity were 96%, 93%, and 99%, respectively. Dohare A et al. [7] synthesized the standard 12-lead ECG records and performed a morphological examination including the P-wave duration, the ST–T interval, and the peak of the QRS complex. The principal component analysis (PCA) was used to reduce the dimension of feature vectors. Their sensitivity, specificity, and accuracy were all 96.66% in 60 samples. All of the above studies used traditional classifiers, such as a support vector machine (SVM), k-Nearest Neighbor (KNN), and neural networks (NN). However, these methods [3–7] require complex feature extraction and selection process and may be only able to achieve good results for specific tasks or data sets.

The automatic detection algorithm for myocardial infarction has also taken advantage of the progress of deep learning. Acharya U et al. [8] introduced the convolutional neural network to analyze ECG heartbeats for myocardial infarction. The accuracies of their method for heartbeats without noise and with noise were 95.22% and 93.53%, respectively. However, they only used single-lead heartbeats to verify their method, which was not present in a rigorous and practical manner. In clinical myocardial infarction diagnosis, the standard 12-lead ECG record is necessary [3] and the beat-wise annotations are expensive. Accordingly, we need to measure the difference between unlabeled heartbeats based

on standard 12-lead ECG records. Sun L et al. [9] first applied Multiple Instance Learning (MIL) to the classification of myocardial infarction. They proposed a new MIL strategy, called latent topic MIL, to map polynomial-fitted heartbeats to subject spaces. It measured the difference between each unlabeled heartbeat based on 12-lead ECG records. Their best performance was obtained using the SVM classifier, with sensitivity and specificity reaching 92.6% and 82.4%, respectively.

In summary, existing studies have the following three issues. First, deep learning networks tailored for the detection of myocardial infarction have not been well studied. In addition, only one or some specific leads ECG signal are unreliable for the detection of myocardial infarction [3] and the beat-wise annotations are of great cost. Finally, few studies have weighted the difference between unlabeled heartbeats based on standard 12-lead ECG records.

In this paper, we propose a bidirectional, long short-term memory (Bi-LSTM) network with heartbeat-attention mechanism. First, Bi-LSTM networks can automatically extract and select features similar to convolutional neural networks. Furthermore, we propose a heartbeat-attention mechanism to automatically weigh the difference between heartbeats based on 12-lead ECG records. Finally, our method can automatically detect myocardial infarction without any labels of heartbeats.

The structure of this paper is as follows: The second part introduces the methods and related theories proposed in this paper, including the system architecture, preprocessing and network framework of myocardial infarction detection algorithm. The third part presents the experimental condition and results of our proposed method. Finally, we conduct related discussion and present the conclusion in the fourth part and fifth part of this paper.

2. Materials and Methods

2.1. Data Description

In this study, we used an available database: Physikalisch-Technische Bundesanstalt (PTB) diagnostic ECG database [16]. As shown in Table 1, the database contains a total of 549 ECG records from 290 patients. All patients are labeled as one of different 10 classes; the two classes with the largest number of samples are myocardial infarction (MI) and healthy control (HC). Each ECG record without any labels of heartbeats contains 15 simultaneously collected ECG records: the standard 12 leads ECG records (I, II, III, avr, avl, avf, V1, V2, V3, V4, V5, V6) and three Frank leads ECG records (V_X , V_Y , V_Z). The database uses a sampling frequency of 1000 Hz and resolution of 16 bits to ensure the quality of ECG signal.

In our study, we extracted standard 12-lead ECG records focused on MI (369 records) and HC (79 records) to analyze and validate our proposed method.

Table 1. The summary of PTB diagnostic ECG Database.

No	Class	No. of Subjects	No. of Records
1	Myocardial infarction (MI)	148	369
2	Cardiomyopathy/Heart failure	18	20
3	Bundle branch block	15	17
4	Dysrhythmia	14	16
5	Myocardial hypertrophy	7	7
6	Valvular heart disease	6	6
7	Myocarditis	4	4
8	Miscellaneous	4	4
9	Healthy controls (HC)	52	79
10	n/a	22	27
Total	10	290	549

2.2. Proposed Method

As shown in Figure 2, the proposed method consists of three parts: preprocessing, model, and output. First, the standard 12-lead ECG records are processed to filter out baseline wander and high-frequency noise. In addition, R-wave detection is performed on the pure ECG records, and sliding windows with the same number of heartbeats are generated based on the position of the R-wave. These sliding windows are an input to a bidirectional long short-term memory (Bi-LSTM) network with heartbeat-attention mechanism. Finally, real output is comprehensively determined by the prediction results of sliding windows.

Figure 2. System Framework.

2.2.1. Preprocessing

Because ECG signals contain noise such as baseline wander, powerline interference, and electromyography, our method uses a wavelet transform-based filter proposed by Donoho, DL et al. [17] to process all ECG records. In addition, the accuracy of the R-wave detection algorithm, based on empirical mode decomposition [18], has reached more than 98% in the PTB database. To obtain complete heartbeats, we apply this algorithm [18] to detect the R-wave position.

As each ECG record in the database has a duration of about 110 records, this is a long-term ECG record for physicians and Bi-LSTM networks. As shown in Figure 3, we sequentially took sliding windows with the same number of heartbeats from a complete ECG record. The hypothetical labels of sliding windows are consistent with the real labels of complete ECG records. In this way, the role of each unlabeled heartbeat in the ECG record can be fully utilized, and capacity of data is invisibly increased to better train the Bi-LSTM network. Although Bi-LSTM networks are robust to long-term dependencies,

the duration of 110 records is daunting. Therefore, the size of sliding windows (the number of heartbeats contained in a sliding window) should be determined by the ability of Bi-LSTM networks.

Figure 3. Sliding Window.

2.2.2. Model

As shown in Figure 4, we propose a recurrent neural network based on the Bi-LSTM network and heartbeat-attention mechanism to classify sliding windows. It consists of four parts: Input Layer, LSTM Layer, Attention Layer, and Output Layer.

Figure 4. Model Framework.

- Input Layer

A sliding window with some unlabeled heartbeats is an input to the model. As shown in Figure 3, we assumed that the size of sliding windows was set as 10. At this time, $T = 10$ in the input $X[T]$ means that the time step of Bi-LSTM network is 10, as shown in Figure 4. To ensure the dimension of input $X[i]$ ($i \in [1, T]$) was constant, we intercepted each heartbeat from one-third of the previous R–R interval and two-thirds of the next R–R interval, and resampled each heartbeat to 100 sample points. As the use of 12-lead ECG, the input $X[i]$ was converted to a heartbeat vector with a dimension of $12 * 100$.

- LSTM Layer

The long short-term memory (LSTM) units were first proposed by Hochreiter S et al. [19] based on recurrent neural networks. The gating mechanism is used to solve the problem of gradient vanishing. Subsequently, due to its good capability for sequence memory, LSTM units are widely used in the field of natural language processing. Some improved versions based on the LSTM units are also emerging.

The basic LSTM units consist of four parts:

Forget gate $\mathcal{F}_t$: its expression is as (1),

$$\mathcal{F}_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

where σ is a sigmoid function, a forget gate controls how much memory state of previous step is forgotten.

Input gate $\mathcal{I}_t$: its expression is shown as (2) and (3),

$$\mathcal{I}_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

where $\check{C}_t$ is the state transition and $\tanh$ is a hyperbolic tangent function, a input gate determines how much input is retained.

Memory state C_t : it is determined by the state transition $\check{C}_t$, input gate $\mathcal{I}_t$ and forget gate $\mathcal{F}_t$, as shown in (4).

$$C_t = \mathcal{F}_t \cdot C_{t-1} + \mathcal{I}_t \cdot \tilde{C}_t \quad (4)$$

Output gate $\mathcal{O}_t$: it controls how much the current unit output h_t (hf_t and hb_t) depends on the current memory state, and its expression is shown as (5) and (6).

$$\mathcal{O}_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = \mathcal{O}_t \cdot \tanh(C_t) \quad (6)$$

Bi-LSTM networks [20] are introduced based on the basic LSTM units. As shown in Figure 4, the model includes forward ($hf[i]$) and backward ($hb[i]$) LSTM channels, which simultaneously capture past and future information. It can extract deep and robust features regarding classification tasks of time series. The output $\mathcal{H}[i]$ of the i^{th} heartbeat is obtained by combining $hf[i]$ and $hb[i]$, as shown in (7).

$$\mathcal{H}[i] = [hf[i] \oplus hb[i]] \quad (7)$$

- Attention Layer

The attention mechanism was first proposed in the field of visual image recognition, and is widely used in natural language processing. For example, Bahdanau D et al. [21] introduced the attention mechanism to the machine translation for the first time and achieved excellent results. The attention mechanism, as the name suggests, was inspired by the performance of the human brain when it focuses on things, giving different weights to different parts of things.

In the field of ECG diagnosis, doctors may pay more attention to abnormal heartbeats when analyzing patients' ECG records. Inspired by this trend, we propose a Bi-LSTM network based on the heartbeat-attention mechanism to detect myocardial infarction. The attention-based network is a measure of similarity. The more similar the current input ($\mathcal{H}_i$) is to the state of real labels, the greater the weights of the current input, indicating that the current output (r) is more dependent on the current input ($\mathcal{H}_i$). Specifically, the current input ($\mathcal{H}_i$) reflects the characteristics of the corresponding heartbeat, and the weight (a_i) reflects differences between heartbeats in the detection of myocardial infarction. The weight (a) is determined by the parameter (ω) automatically learned during training and the input ($\mathcal{H}$) as shown in (8), (9), and (10).

$$\mathcal{M} = \tanh(\mathcal{H}) \quad (8)$$

$$a = \text{softmax}(\omega^T \mathcal{M}) \quad (9)$$

$$r = \mathcal{H} a^T \quad (10)$$

Then the output of attention layer $\mathcal{H}^*$ can be expressed as (11).

$$\mathcal{H}^* = \tanh(r) \quad (11)$$

- Output Layer

We use *softmax* function and conditional probability maximum criterion to determine the prediction, as shown in (12).

$$Y = \arg\max(\text{softmax}(W_B \cdot \mathcal{H}^* + b_B)) \quad (12)$$

As shown in (13), a cross entropy function is employed as loss function. And we add L2 regularization to reduce over-fitting, where λ is a hyper-parameter.

$$\text{Loss}_{\theta} = -\frac{1}{n} \sum_{i=1}^n y_i \log(Y_i) + \lambda \|\theta\|_F^2 \quad (13)$$

2.2.3. Output

Because there are no any labels of heartbeats in the PTB diagnostic database, the predictions for sliding windows are not the final indicator for evaluating the performance. To obtain the final predictions for complete ECG records, we set the decision function. According to the counts of sliding windows, the predictions for complete ECG records are determined as the predictions for the majority of their sliding windows. Finally, we compared final predictions with the annotations in PTB diagnostic database to evaluate the proposed method.

3. Results

To illustrate the performance of the Bi-LSTM network with heartbeat-attention mechanism, we performed multiple sets of experiments in the PTB diagnostic ECG database. The experiments include two aspects. First, because deep learning networks are involved, optimal parameters are essential to improve the experimental results. The key parameters in our method include the Window Size: the number of heartbeats contained in a sliding window (corresponding to the time step in Bi-LSTM networks), the Beat Length: dimension of heartbeats (set as 12 * 100 described above) and important hyper-parameters in Bi-LSTM networks: dimension of hidden unit (Hidden Size), dropout ratio (keep prob). Second, to explain the effect of heartbeat-attention mechanism, we only removed the Attention Layer in our model and compared their experimental results in the same conditions.

3.1. Performance Evaluation

In the field of ECG diagnosis, the most common performance evaluation is sensitivity (Sen), specificity (Spec), and accuracy (Acc) respectively. To facilitate comparison with related studies, we also use them to evaluate our method. The expression is shown as (14), (15) and (16):

$$\text{Sensitivity} = TP / (TP + FN) \quad (14)$$

$$\text{Specificity} = TN / (TN + FP) \quad (15)$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (16)$$

The positive sample is MI (label as 1) and the negative sample is HC (label as 0). Among them, TP, FP, TN, and FN, respectively, represent the number of sick samples diagnosed correctly as sick. Healthy samples are wrongly diagnosed as sick, healthy samples are correctly diagnosed as healthy, and sick samples are wrongly diagnosed as healthy.

3.2. Parameter Setting

In our experiments, 448 12-lead ECG records in the PTB diagnostic ECG database are randomly divided, 70% as the training set and the other 30% as the testing set. To obtain reliable experimental results, we repeat each set of experiments 20 times independently based on a Bi-LSTM network with heartbeat-attention mechanism. And the testing set are also randomly divided for each repeated experiment. Our model was trained on a processor for the Intel Core i5-8400 and a GPU for the Nvidia GeForce GTX 1060 6GB. Its convergence usually occurs after 1000 batches (batch size is 12) and takes about 5090.1 s.

To find optimal experimental parameters, we studied the impact of three parameters on experimental performance by control variable method, including the number of heartbeats in a sliding window (Window Size), dimension of hidden unit (Hidden Size), and dropout ratio (keep prob). First, only the Window Size was adjusted, and the values of the Hidden Size and the keep prob were set as 80 and 0.4, respectively. As shown in Figure 5, we can see that the accuracy, sensitivity, and specificity all have similar trends. When the Window Size is set to 24, the experimental performance is optimal, and the accuracy can reach close to 94%.

Figure 5. Impact of window size.

With regard to the hyper-parameters of Bi-LSTM networks: Hidden Size and keep prob, we can observe from Figure 6 that the red curve is in a relatively high position. Accordingly, the performance is

relatively good when the keep prob is 0.3. As the green curve is in a relatively low position, we believe that the network may be over-fitting when the keep prob is 0.5. Because the peak point of the red curve is obtained when Hidden Size is 90, we conclude that the relatively optimal Hidden Size and keep prob are 90 and 0.3, respectively, and the accuracy can reach nearly 95%.

Figure 6. Impact of hyper-parameters.

3.3. Experimental Results

On the basis of optimal experimental parameters, we respectively trained and tested the model without heartbeat-attention mechanism (Experiment 1) and with heartbeat-attention mechanism (Experiment 2) in the same conditions. The performance evaluation was calculated on the testing set containing 134 samples, and the average results were obtained after 20 independent repetitions.

As shown in Tables 2 and 3, the accuracy of Experiment 2 is about 2% higher than Experiment 1, and the better accuracy can reach 94.77%. In addition, it can be seen from the confusion matrix that many healthy samples (HC) are wrongly identified as sick (MI) in Experiment 1, and its specificity only reaches 85.00%. The results of Experiment 2 clearly show that the specificity is significantly improved, reaching 90.48%. Simultaneously, the sensitivity in Experiment 2 is also maintained at a high level, reaching 95.58%.

Table 2. The performance of Bi-LSTM without the heartbeat-attention mechanism (Experiment 1).

		Predicted				
		HC	MI	Accuracy	Sensitivity	Specificity
Real	HC	17	3	92.54%	93.86%	85.00%
	MI	7	107			

Table 3. The performance of Bi-LSTM with the heartbeat-attention mechanism (Experiment 2).

		Predicted				
		HC	MI	Accuracy	Sensitivity	Specificity
Real	HC	19	2	94.77%	95.58%	90.48%
	MI	5	108			

The model with heartbeat-attention mechanism achieves better performance when the classifier and experimental condition are identical. From the methods section above, the principle of the heartbeat-attention mechanism is to make the model pay more attention to the key heartbeats

by giving different weights to different heartbeats. As shown in Figure 7, we plotted a 12-lead record that was correctly predicted as myocardial infarction. The weights given to heartbeats by the heartbeat-attention mechanism in this record are represented by the green histogram. The thirteenth heartbeat is an obviously abnormal heartbeat, and it is given a larger weight. In addition, other general heartbeats are relatively evenly assigned weights. The model with heartbeat-attention mechanism can weigh different heartbeats differently for samples of myocardial infarction. Further, the difference in weights will make feature vectors more representative and result in better experimental performance, as shown in Table 3.

Figure 7. Weights assigned for different heartbeats in ECG records.

4. Discussion

As shown in Table 4, we listed some classical algorithms in the field of detection of myocardial infarction. Most [3–7] applied machine learning techniques to the detection of myocardial infarction and achieved promising results. However, some of them [3–7] achieved good results only for some specific ECG leads or ECG records. In our work, the Bi-LSTM network does not require complex feature extraction techniques and we achieved comparable or better performance in a relatively complete PTB diagnostic ECG database.

Acharya et al. [8] applied a convolutional neural network to classify heartbeats of lead II ECG records, also eliminating the need for manual extraction of features. Simultaneously, it achieved an accuracy of 95.22%. In our work, we do not use any labels of heartbeats to train our model based on the standard 12-lead ECG records. Further, our model is validated by the patients' complete records and their real labels from the PTB diagnostic ECG database. Only a 1-lead ECG is unreliable for detection of myocardial infarction [3] and the beat-wise annotations are of great cost. Accordingly, our method is more effective and practical.

The experimental conditions of Sun et al. [9] were the same as for our experiments. They applied multiple instance learning to measure the difference between unlabeled heartbeats based on 12-lead ECG records. Our method can automatically complete this process by a Bi-LSTM network with heartbeat-attention mechanism. As analyzed in the previous section, the heartbeat-attention mechanism

can weight different unlabeled heartbeats based on 12-lead ECG records to improve performance. Additionally, our method has been significantly and comprehensively improved in experimental performance in comparison with theirs [9].

In summary, our method has the following advantages:

1. It does not require any labels of heartbeats and is based on the standard 12-lead ECG records.
2. The heartbeat-attention mechanism is implemented to weight different unlabeled heartbeats.

Table 4. Summary of selected studies conducted for the detection of myocardial infarction (MI).

Author, Year	Leads Used	Data Used	Method	Performance
Reddy et al. [4] 1992	V2–V4 leads	Normal:320 records AMI:182 records	Morphology feature ANN	Acc = NA Sen = 79.00% Spec = 97.00%
Pei et al. [5] 2012	V1–V4 leads	Normal: 582 MI: 547 (heartbeats)	HMMs with GMMs	Acc = 82.50% Sen = 85.71% Spec = 79.82%
Safdarian et al. [6] 2014	Lead II	549 records (PTB)	T-wave integral Naive Bayes	Acc = 94.74% Sen = NA Spec = NA
Sharma et al. [3] 2015	12 leads	2148 instances (a frame of four heartbeats from PTB)	Multiscale Energy SVM	Acc = 96.00% Sen = 93.00% Spec = 99.00%
Dohare et al. [7] 2018	12 leads	Normal: 60 records MI: 60 records (selected from PTB)	Morphological examination PCA SVM	Acc = 96.96% Sen = 96.96% Spec = 96.96%
Acharya et al. [8] 2017	Lead II	Normal: 10546 MI: 40182 (heartbeats from PTB)	Convolutional neural network	Acc = 95.22% Sen = 95.49% Spec = 94.19%
Sun L et al. [9] 2012	12 leads	Normal: 79 records MI: 369 records (PTB)	multiple instance learning SVM	Acc = NA Sen = 92.60% Spec = 82.40%
Our method	12 leads	Normal: 79 records MI: 369 records (PTB)	Bi-LSTM Heartbeat-attention	Acc = 94.77% Sen = 95.58% Spec = 90.48%

Although our method achieves good performance, we also discovered drawbacks from misclassified samples. As shown in Figure 8, this is a false negative sample with only four abnormal heartbeats. Although the predicted value of this sliding window is correct, the prediction of the ECG record containing this sliding window is wrong. The reason is that the complete ECG record of ‘Patient049/s0173’ only contains four abnormal heartbeats, which results in many sliding windows to be predicted as HC. Eventually, the output of the decision function is wrong. This problem may be solved in two ways. On the one hand, a larger number of ECG records are needed to optimize the parameters in the decision function. On the other hand, we should add other symptoms of patients into the decision function.

For the false positive sample, we can see from Figure 9 that most of the heartbeats in the sliding window are distorted, and patient’s motion state or the stable type of acquisition devices has a significant impact on the quality of ECG signals. Poor quality will lead to degraded performance. We believe that the algorithms for evaluating quality of ECG signals [22] will improve this problem.

In summary, our method has following drawbacks:

1. It requires more data to further improve the performance.
2. Its performance is greatly affected by the quality of ECG signals.

Figure 8. The ECG waveform of a false negative sample.

Figure 9. The ECG waveform of a false positive sample.

5. Conclusions

With the tremendous development of medical big data and mobile monitoring, detection algorithms of myocardial infarction are being applied to real-time ECG monitoring equipment. We introduce a heartbeat-attention mechanism for the detection of myocardial infarction using a standard 12-lead ECG records; the mechanism achieved good performance, with the accuracy of 94.77%, sensitivity of 95.58%, and specificity of 90.48%. Our method does not use any labels of heartbeats to train the Bi-LSTM network and automatically weights the difference between unlabeled heartbeats to improve performance. Two drawbacks can be improved in our study. In future work, we will deploy our method to medical wearable devices [23] to obtain more real 12-lead ECG records to improve our method for greater accuracy and robustness. Eventually, the algorithm we are studying can be applied to the actual product and will serve patients well.

Author Contributions: Conceptualization, L.J. and Z.Y.; methodology, L.J. and Z.Y.; software, L.J.; validation, L.J.; formal analysis, L.J.; investigation, L.J.; resources, L.J.; data curation, L.J.; writing—original draft preparation, L.J. and Z.Y.; writing—review and editing, L.J. And Z.Y.; visualization, L.J.; supervision, L.J.; project administration, Z.Y.; funding acquisition, Z.Y.

Funding: This research was funded by National Natural Science Foundation of China, No. 61571268.

Acknowledgments: The authors are grateful to Yuan Bo for improving our paper in English language and style.

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

References

1. Mozaffarian, E.; Benjamin, E.J.; Go, A.S.; Arnett, D.K.; Blaha, M.J.; Cushman, M.; Das, S.R.; de Ferranti, S.; Després, J.P. Heart disease and stroke statistics—2016 update. *Circulation* **2016**, *133*, e38–e48. [[CrossRef](#)] [[PubMed](#)]
2. Antman, E.; Bassand, J.P.; Klein, W.; Ohman, M.; Sendon, J.L.L.; Rydén, L.; Simoons, M.; Tendera, M. Myocardial infarction redefined—A consensus document of The Joint European Society of Cardiology/American College of Cardiology committee for the redefinition of myocardial infarction: The Joint European Society of Cardiology/American College of Cardiology Committee. *J. Am. Coll. Cardiol.* **2000**, *36*, 959–969.
3. Sharma, L.N.; Tripathy, R.K.; Dandapat, S. Multiscale Energy and Eigenspace Approach to Detection and Localization of Myocardial Infarction. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1827–1837. [[CrossRef](#)] [[PubMed](#)]
4. Reddy, M.R.S.; Edenbrandt, L.; Svensson, J.; Haisty, W.K.; Pahlm, O. Neural network versus electrocardiographer and conventional computer criteria in diagnosing anterior infarct from the ECG. In Proceedings of the Computers in Cardiology, Durham, NC, USA, 11–14 October 1992.
5. Chang, P.C.; Lin, J.J.; Hsieh, J.C.; Weng, J. Myocardial Infarction Classification with Multi-lead ECG Using Hidden Markov Models and Gaussian Mixture Models. *Appl. Soft Comput.* **2012**, *12*, 3165–3175. [[CrossRef](#)]
6. Safdarian, N.; Dabanloo, N.J.; Attarodi, G.A. New Pattern Recognition Method for Detection and Localization of Myocardial Infarction Using T-Wave Integral and Total Integral as Extracted Features from One Cycle of ECG Signal. *J. Biomed. Sci. Eng.* **2014**, *5*, 818–824. [[CrossRef](#)]
7. Dohare, A.K.; Kumar, V.; Kumar, R. Detection of Myocardial Infarction in 12 lead ECG using Support Vector Machine. *Appl. Soft Comput.* **2018**, *64*, 138–147. [[CrossRef](#)]
8. Acharya, U.R.; Fujita, H.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adam, M. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Inf. Sci.* **2017**, *415*, 190–198. [[CrossRef](#)]
9. Sun, L.; Lu, Y.; Yang, K.; Li, S. ECG Analysis Using Multiple Instance Learning for Myocardial Infarction Detection. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 3348–3356. [[CrossRef](#)] [[PubMed](#)]
10. Gallino, A.; Chierchia, S.; Smith, G.; Croom, M.; Morgan, M.; Marchesi, C.; Maseri, A. Computer system for analysis of ST segment changes on 24-hour Holter monitor tapes: Comparison with other available systems. *J. Am. Coll. Cardiol.* **1984**, *4*, 245–252. [[CrossRef](#)]
11. Sun, G.; Thomas, C.W.; Liebman, J.; Rudy, Y.; Reich, Y.; Stilli, D.; Macchi, E. Classification of normal and ischemia from BSPM by neural network approach. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, New Orleans, LA, USA, 4–7 November 1988; pp. 1504–1505.
12. Passariello, G.; Mora, F.; De La Cruz, E.; Gotoc, J.; Cerreult, B. Realtime detection and quantification of ischemic ECG changes. In Proceedings of the Twelfth Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Philadelphia, PA, USA, 1–4 November 1990; pp. 809–810.
13. Krucoff, M.W.; Croll, M.A.; Pendley, L.P.; Burdette, D.L.; Pope, J.E.; Hutchinson, D.D.; Stone, J.S.; Weber, R.A.; Califf, R.M. Continuous computer-assisted electrocardio graphic monitoring in patients with acute myocardial infarction: Early experience. In Proceedings of the Computers in Cardiology, Jerusalem, Israel, 19–22 September 1989; pp. 197–200.
14. Thakor, N.; Gramatikov, B.; Mita, M. Multiresolution wavelet analysis of ECG during ischemia and reperfusion. In Proceedings of the Computers in Cardiology Conference, London, UK, 5–8 September 1993.
15. Lemire, D.; Pharand, C.; Rajaonah, J.C.; Dubé, B.; LeBlanc, A.R. Wavelet time entropy, T wave morphology and myocardial ischemia. *IEEE Trans. Biomed. Eng.* **2000**, *47*, 967–970. [[CrossRef](#)] [[PubMed](#)]
16. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [[CrossRef](#)] [[PubMed](#)]
17. Donoho, D.L. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **1995**, *41*, 613–627. [[CrossRef](#)]
18. Pal, S.; Mitra, M. QRS Complex detection using Empirical Mode Decomposition based windowing technique. In Proceedings of the 2010 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 18–21 July 2010.

19. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
20. Zhang, D.; Wang, D. Relation Classification via Recurrent Neural Network. *arXiv* **2015**, arXiv:1508.01006.
21. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* **2014**, arXiv:1508.01006.
22. Wang, J.Y. A new method for evaluating ECG signal quality for multi-lead arrhythmia analysis. In Proceedings of the Computers in Cardiology, Memphis, TN, USA, 22–25 September 2002.
23. Sun, X.; Zhang, Y. Design and Implementation of Portable ECG and Body Temperature Monitor. In Proceedings of the 2014 International Symposium on Computer, Taichung, Taiwan, 10–12 June 2014.

© 2019 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/>).