ISSN 1674-1056
http://cpb.iphy.ac.cn
http://iopscience.iop.org/cpb
中国物理 B
Chinese Physics
中国物理
Chinese Physics B

Preliminary abnormal electrocardiogram segment screening method for Holter data based on long short-term memory networks

Siying Chen(陈偲颖), Hongxing Liu(刘红星)

Citation: Chin. Phys. B . 2020, 29(4): 040701 . **doi:** 10.1088/1674-1056/ab6d52
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Multiple Lagrange stability and Lyapunov asymptotical stability of delayed fractional-order Cohen-Grossberg neural networks

Yu-Jiao Huang(黄玉娇), Xiao-Yan Yuan(袁孝焰), Xu-Hua Yang(杨旭华), Hai-Xia Long(龙海霞), Jie Xiao(肖杰)
Chin. Phys. B . 2020, 29(2): 020703 . **doi:** 10.1088/1674-1056/ab6716

Coexistence and local Mittag-Leffler stability of fractional-order recurrent neural networks with discontinuous activation functions

Yu-Jiao Huang(黄玉娇), Shi-Jun Chen(陈时俊), Xu-Hua Yang(杨旭华), Jie Xiao(肖杰)
Chin. Phys. B . 2019, 28(4): 040701 . **doi:** 10.1088/1674-1056/28/4/040701

Synthesization of high-capacity auto-associative memories using complex-valued neural networks

Yu-Jiao Huang(黄玉娇), Xiao-Yan Wang(汪晓妍), Hai-Xia Long(龙海霞), Xu-Hua Yang(杨旭华)
Chin. Phys. B . 2016, 25(12): 120701 . **doi:** 10.1088/1674-1056/25/12/120701

Multistability of delayed complex-valued recurrent neural networks with discontinuous real-imaginary-type activation functions

Huang Yu-Jiao, Hu Hai-Gen
Chin. Phys. B . 2015, 24(12): 120701 . **doi:** 10.1088/1674-1056/24/12/120701

Stability analysis of Markovian jumping stochastic Cohen–Grossberg neural networks with discrete and distributed time varying delays

M. Syed Ali
Chin. Phys. B . 2014, 23(6): 060702 . **doi:** 10.1088/1674-1056/23/6/060702

Preliminary abnormal electrocardiogram segment screening method for Holter data based on long short-term memory networks

Siying Chen(陈偲颖)[‡] and Hongxing Liu(刘红星)[†]

School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, *China*

(Received 23 December 2019; revised manuscript received 13 January 2020; accepted manuscript online 20 January 2020)

Holter usually monitors electrocardiogram (ECG) signals for more than 24 hours to capture short-lived cardiac abnormalities. In view of the large amount of Holter data and the fact that the normal part accounts for the majority, it is reasonable to design an algorithm that can automatically eliminate normal data segments as much as possible without missing any abnormal data segments, and then take the left segments to the doctors or the computer programs for further diagnosis. In this paper, we propose a preliminary abnormal segment screening method for Holter data. Based on long short-term memory (LSTM) networks, the prediction model is established and trained with the normal data of a monitored object. Then, on the basis of kernel density estimation, we learn the distribution law of prediction errors after applying the trained LSTM model to the regular data. Based on these, the preliminary abnormal ECG segment screening analysis is carried out without R wave detection. Experiments on the MIT-BIH arrhythmia database show that, under the condition of ensuring that no abnormal point is missed, 53.89% of normal segments can be effectively obviated. This work can greatly reduce the workload of subsequent further processing.

Keywords: electrocardiogram, long short-term memory network, kernel density estimation, MIT-BIH arrhythmia database

PACS: 07.05.Mh, 02.50.Cw, 07.05.Kf **DOI:** 10.1088/1674-1056/ab6d52

1. Introduction

Arrhythmia manifests as irregularity in heart rate or rhythm, and the exact moment of occurrence is uncertain, usually needing to wear Holter for 24–72 hours to capture. It is time consuming and labor intensive to manually interpret long Holter data by doctors. Therefore, automated analysis is required.

Among the automatic analysis techniques for Holter data, there is a kind of methods which need R wave detection.[1–5] For example, the work[5] separates each heartbeat waveform one by one based on R wave detection, and then sends them to the deep convolutional neural network (CNN) for heart beat classification. The work divides heart beats into one of five categories. This kind of Holter data analysis method starts with the detailed R wave detection that is somewhat hard to accomplish well, which does not conform to the principle from coarse to fine.

A reasonable approach is not to do the R wave detection first, but to make judgments on the electrocardiogram (ECG) data segments, in order that it can screen out normal data segments as many as possible, i.e., reduce the data quantity first on the premise of ensuring the retention of abnormal segments.

Mhuah and Fu[6] divided segments according to whether the distance between ECG segment and nearest non-self match exceeds the threshold. In the method, determining the threshold needs to learn from the abnormal data segment. Acharya *et al.*[7] proposed to use the deep CNN network to identify four types of ECG segments: atrial flutter, atrial fibrillation, ventricular fibrillation, and normal. Although the shown performance is good, the given classes are too many, needing much more labeled data for training, which means that it is of weak operability and not suitable for preliminary screening.

Chauhan and Vig[8] used the error generated by the LSTM prediction model to classify the segments. If the log(PDF) value of an error exceeds a threshold, the corresponding segment is divided into the anomaly. The threshold is determined as follows: a validation set containing abnormal signals is used, to maximize the F1-score on the validation set to get a fixed threshold, i.e., the target function pursues the largest proportion of true positive samples. However, we think the most important thing in the screening of data is that the abnormal segments cannot be missed. Avoiding false negatives should be the constraint.

The shortcomings of the above screening methods by segment can be summarized as follows: (1) Abnormal segments are needed for learning, which is inconvenient in practice. (2) The design of the objective function is not reasonable, and avoiding false negatives should be the premise to screen away the normal data segments. This paper will propose an improved Holter data anomaly segment screening method based on the LSTM prediction model. It does not need abnormal data for learning. It only needs to monitor the normal ECG signal points for a few minutes to train the prediction model. The determination of the threshold is based on the maximum

[†]Corresponding author. E-mail: njhxliu@nju.edu.cn

[‡]Corresponding author. E-mail: siying_chen22@163.com; 495777679@qq.com

© 2020 Chinese Physical Society and IOP Publishing Ltd

probability of avoiding missing abnormal data segments.

2. Materials

2.1. Long short-term memory networks

As one kind of RNN networks, the LSTM as shown in Fig. 1 is good at processing sequences. It is not prone to gradient disappearance or gradient explosion while learning.[9] It is characterized by the use of the LSTM unit, as shown in Fig. 2, which consists of four main elements: an input gate [I_G], a neuron with a self-circulating connection, a forgetting gate [F_G] and an output gate [O_G].

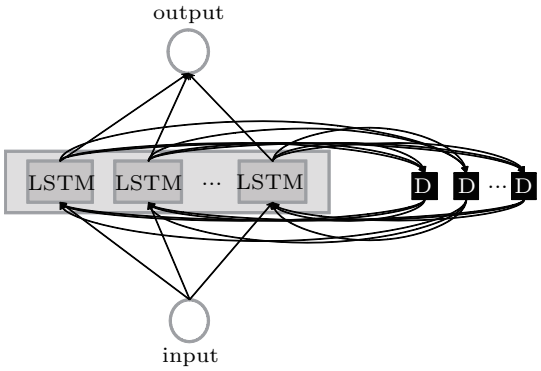


Fig. 1. Basic representation of the LSTM model.

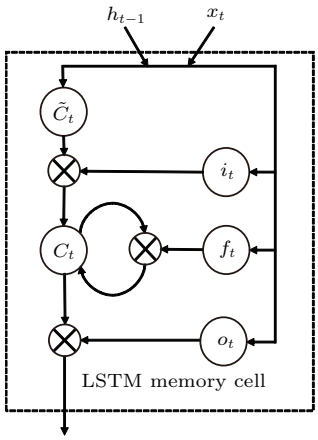


Fig. 2. LSTM memory cell.

The input of the input gate [I_G] of an LSTM unit is the input vector x_t from the previous layer at the current moment and the output vector h_{t-1} of all LSTM units of the current layer at the previous moment. After the sigmoid unit ($\sigma(\cdot)$), a real number between 0 and 1 is obtained. W_i and U_i are two weight vectors, and b_i is the offset. The forgetting gate [F_G] and the output gate [O_G] are calculated in a similar manner to the input gate. The unit function for $\tilde{C}_t$, which is with the self-loop connected neuron portion, is tanh and the output is a value between -1 and 1. The unit $\tilde{C}_t$, the memory state C_{t-1} at the previous moment, the current input gate and the forgetting gate together determine the current state C_t. The specific calculation formulas are as follows:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right), \tag{1}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right), \tag{2}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + V_o C_t + b_o\right), \tag{3}$$

$$\tilde{C}_t = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right), \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{5}$$

$$h_t = o_t * \tanh\left(C_t\right), \tag{6}$$

where x_t is the input of the LSTM unit at a given time t, h_{t-1} is the output vector of the entire LSTM layer at the time $t-1$. W_i, W_c, W_f, W_o, U_i, U_c, U_f, U_o, and V_o are weight matrices, b_i, b_c, b_f, b_o are offset vectors, h_t is the output of the LSTM unit, C_{t-1} is the output of the operator C_t, the symbol $*$ refers to the product of the vector elements.

2.2. Database

The data used in the following study is from the MIT-BIH arrhythmia database.[4] The data in the database was independently labeled by two cardiologists who marked all abnormal beats and added rhythm and signal quality labels. The data for each patient is approximately for 30 min and the sampling rate is 360 Hz, including 2 channels.

In the following study, only 16 patients are selected for modeling research. They are 100, 101, 105, 108, 114, 121, 123, 202, 205, 209, 210, 213, 219, 223, 230, 234. These data have normal sequences which are longer than one minute and can be used for the modeling of this method, while other data does not have long enough normal sequences for this model.

In addition, this study only selects the MLII channel of the 16 data for research. Compared with the other channel, it can better meet the requirements of normal sequences with longer than 1 min.

3. Proposed method

3.1. General idea

The whole method consists of two phases: the learn phase and the work phase.

Due to the individual differences between the ECG signals of different patients, each patient data is arranged for their own model learning. In the learn phase, a normal ECG sequence of the patient is required. First pre-process the sequence to reduce noise, and the noise-reduced data is used to train the LSTM network, which is essentially an autoregressive prediction model. Using the trained model, the prediction errors of the input normal data are calculated. On the basis of kernel density estimation (KDE), the probability density function of the prediction errors can be obtained, and subsequently we can determine the prediction error threshold interval for accepting the assumption that the data is normal, as shown in Fig. 3(a).

In the work phase, as a judging unit, each ECG segment of a sequence will be classified as a normal one or an abnormal one. The sequence to be tested is sent to the trained LSTM prediction model, and the prediction error of each point in the segment is calculated to see whether any prediction error exceeds the set threshold interval. If yes, it will be an abnormal segment considered. If no, a normal segment considered, which can be screened out. See Fig. 3(b).

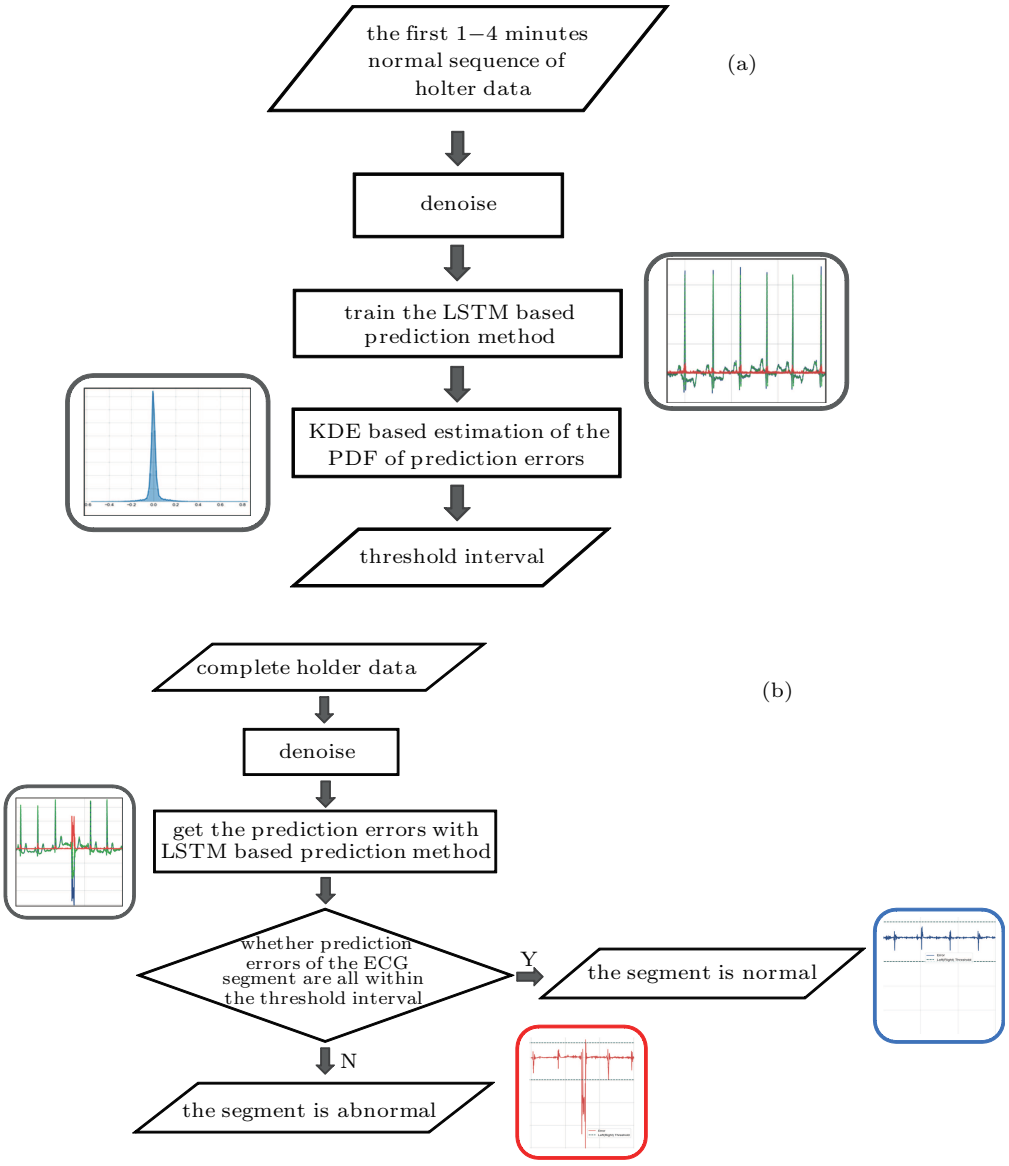


Fig. 3. Training and working flowchart of the method: (a) the training LSTM model and threshold interval, (b) determining if ECG segment is normal.

3.2. LSTM based prediction

The LSTM network used is called the stacked LSTM network. As shown in Fig. 4, it consists of 2 LSTM layers, a fully connected layer, and a linear layer. The first LSTM layer owns 60 LSTM units and the second LSTM layer owns 40 LSTM units. LSTM units in the same layer are not connected to each other, but their outputs will be used as the input of each LSTM unit in the same layer at the next moment and the input of each LSTM unit in the next layer. There is only one neuron in both the fully connected layer and the linear layer. The input of the entire network model is unidimensional, that is, the ECG signal point at the current time, and the output of the model is also unidimensional, that is, only the ECG signal point at the next moment is predicted.

The cost function L of network training is defined as the mean squared error function (MSE)

$$L = \mathrm{MSE}(\hat{x}) = E(\hat{x} - x)^2, \tag{7}$$

where $\hat{x}$ is the predicted value and x is the true value. The weight and bias parameters are updated using the adaptive learning rate algorithm.[10] The dropout[11,12] mechanism is used to prevent overfitting. There are training rounds, i.e., epoch, batch size, timestep and other parameters, which need to be configured. After the experiment, epoch is set to 200

times, and batch size is 360.

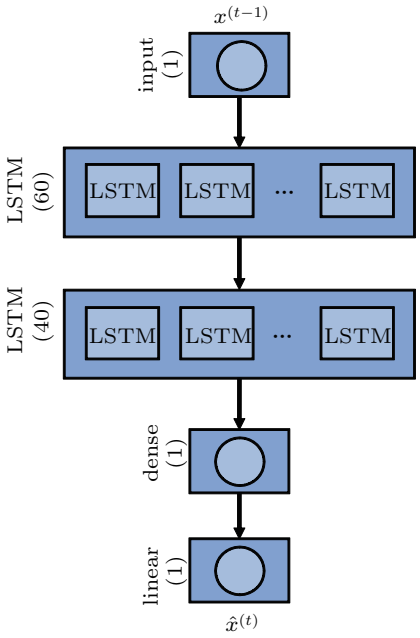


Fig. 4. The determined stacked LSTM network architecture.

Timestep is a parameter that controls the complexity of the training process, and it limits the number of steps where the gradient can expand in the backpropagation. Essentially, it is assumed that the model relies on the current ECG samples and the latest timestep ECG samples.

3.3. KDE based anomaly screening

We define the prediction error $e^{(t)}$ as follows:

$$e^{(t)} = \hat{x} - x, \tag{8}$$

estimating its probability density function $p(e^{(t)})$ with all the prediction errors of a normal sequence is the key to learn the threshold which can distinguish between the abnormal and normal data points.

After observing, the histogram of the prediction errors in the experiment, as shown in Fig. 5(a), cannot be described as a known form of probability density function. Considering the large number of error samples, kernel density estimation can be used to characterize the probability density function of the prediction error. In the method, the probability $p(e^{(t)})$ of any prediction error $e^{(t)}$ is estimated as the probability density $\hat{p}(e^{(t)})$ as follows:

$$\hat{p}(e^{(t)}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{e^{(t)} - e^{(i)}}{h}\right), \tag{9}$$

where n is the number of samples, $e^{(i)}$ is the prediction error of the observed sample i, and h is the scale parameter of the Gaussian kernel function $K(x)$:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right). \tag{10}$$

In Khosrow's work,[13] the optimal scale parameter h_{opt} in the kernel density estimation can be calculated by the following formula:

$$\begin{aligned} h_{\text{opt}} &= (4\pi)^{-1/10}\left(\frac{3}{8}\pi^{-1/2}\right)^{-1/5}\sigma n^{-1/5} \\ &= (4/3)^{1/5}\sigma n^{-1/5} = 1.06\sigma n^{-1/5}, \end{aligned} \tag{11}$$

where σ can be the standard deviation. At this point, the probability density $\hat{p}(e^{(t)})$ of $e^{(t)}$ can be calculated. In the following experiment, the calculated error probability density function curve of the case in the Fig. 5(a) is as shown in Fig. 5(b), which is consistent with the shape of Fig. 5(a).

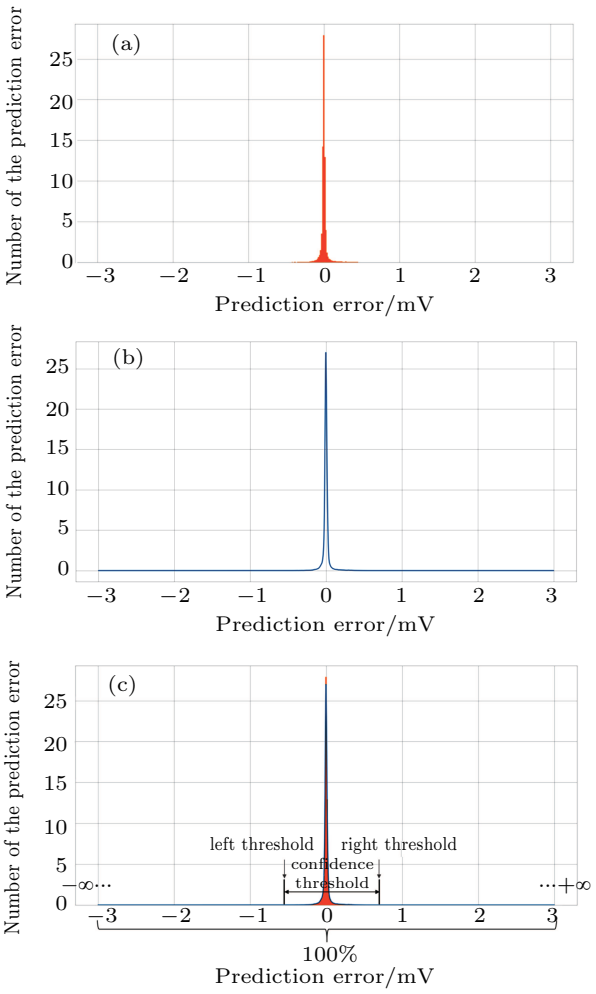


Fig. 5. Kernel density estimation and confidence interval determination: (a) histogram of the prediction error produced by the prediction model on the normal signal points, (b) kernel density estimation based probability density function curve of normal ECG signal points, (c) confidence threshold and determined corresponding left and right thresholds through integration.

Set the confidence value α. If a prediction error $e^{(t)}$ in an ECG segment is outside the corresponding threshold interval, the ECG segment can be judged to be abnormal. Since the prediction errors do not obey any classical distribution, it is necessary to calculate the threshold interval by the integration

method. According to the known confidence value, the way to obtain the left and right thresholds is as follows: accumulate the probability from the negative infinity to the left threshold γ_1 until the sum reaches $1/2(1-\alpha)$ and accumulate the probability from the positive infinity to the right threshold γ_2 until the sum of the process also reaches $1/2(1-\alpha)$, as shown in Fig. 5(c). Hence, $[\gamma_1, \gamma_2]$ is the corresponding threshold interval when the confidence value is α.

4. Experiments

4.1. Data annotation for normal segments and abnormal segments

For the data in the selected database, we divide each long sequence into several connected time segments. If there is an abnormal signal point in the time range of a segment, the segment will be marked as an abnormal one; otherwise, it will be marked as the normal. Since a supraventricular ectopic activity[14] has small QRS wave morphological change, the segment with supraventricular ectopic activities is also marked as the normal one in our method. The determination of the length of each segment should make sure the convenience of the doctors observing ECG waveform, and 10 s duration is selected as the size of one segment in this paper. Each patient data is divided into 181 labeled segments.

4.2. Data pre-processing

Each patient's data is divided into two parts: training set, test set. As shown in Fig. 6, the training set is a normal sequence located in the first 1–9 minutes, and the test set is a complete 30-min ECG data of the patient, that is, a mixed sequence of normal and abnormal signal points. The training set is used for (1) training of the LSTM model, (2) obtaining a probability density function of the normal signal points, and (3) determining the threshold interval according to the confidence value. The test set is used to verify the effect of the algorithm.



Fig. 6. Training set and test set.

The ECG data is subjected to discrete wavelet transform (Daubechies 6) for noise and baseline wander removal.[15] Subsequently, to eliminate the problem of amplitude scaling,[16] the sequences are processed via Z-score normalization ($z = (X - \bar{X})/S$), where $\bar{X}$ is the mean of the input samples, and S is the variance of the input samples.

4.3. Results

The main objective of the abnormal data screening method proposed in this paper is to eliminate more normal segments to reduce the workload of the subsequent further fine classification, while ensuring that no disease segments are missed. To describe the screening ability for anomalies, we define the evaluation indicator: effective screening rate (ESR).

$$\mathrm{ESR} = \frac{\mathrm{TN}}{\mathrm{FP}+\mathrm{TN}}, \quad \text{st. } \mathrm{FN} = 0, \tag{12}$$

where FP represents the number of normal segments that are judged to be abnormal, TN represents the number of normal segments that are judged to be normal, and FN represents the number of abnormal segments that are judged to be normal (i.e., missed detection).

The condition that FN is 0, means that no abnormal segment is misjudged as the normal. The higher the ESR, the larger the correct rate of the classification result on the normal segments, and the greater the normal segments can be excluded, whereas neither of the abnormal segments can be judged as the normal ones.

Table 1. Results of all test sets.

Patient Id	Threshold interval	TP	FP	TN	FN	ESR
100	[−0.627, 1.116]	1	67	113	0	62.78%
101	[−0.615, 0.674]	2	74	105	0	58.66%
105	[−0.415, 0.700]	43	95	43	0	31.16%
108	[−0.478, 0.557]	17	77	87	0	53.05%
114	[−0.657, 1.065]	36	22	123	0	84.83%
121	[−0.370, 0.497]	1	98	82	0	45.56%
123	[−0.981, 0.849]	3	57	121	0	67.98%
202	[−0.811, 0.436]	19	55	107	0	66.05%
205	[−0.618, 0.631]	30	94	57	0	37.75%
209	[−0.907, 1.047]	1	83	97	0	53.89%
210	[−0.362, 0.432]	102	41	38	0	48.10%
213	[−0.391, 0.424]	117	57	7	0	10.94%
219	[−0.486, 0.423]	50	31	100	0	76.34%
223	[−0.441, 0.334]	89	87	5	0	5.43%
230	[−0.930, 0.815]	1	22	158	0	87.78%
234	[−0.348, 0.439]	3	50	128	0	71.91%
Average ESR: 53.89%						

The selection of the confidence value needs to be optimized. If the confidence value is set too small, the threshold interval will be too small. This may cause the prediction errors of the normal points to exceed the threshold limit and be misjudged as abnormal. What we need is the maximization of the confidence value under $= 0$ to maximize ESR. After experiments, we find that when the confidence threshold is set to 99.993%, ESR is highest at 53.89%. In other words, 53.89 per cent normal data can be excluded and there is no missed anomaly detection, as shown in Table 1.

Figures 7 and 8 show the intercepted 1-min sequence test result for the patients 100 and 114, respectively. The prediction errors of the normal points are within the threshold interval, while there appears a distinct bipolar camp with the prediction errors generated for the anomalous points.

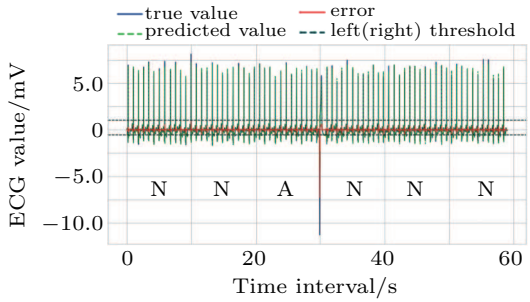


Fig. 7. The 1-min sequence test result for No. 100 patient (N: normal, A: abnormal).

As shown in Fig. 9, the test set contains a variety of arrhythmia diseases. Although the model only needs the knowledge of normal ECG data without learning the characteristic patterns of various diseases, it can still successfully detect a variety of diseases.

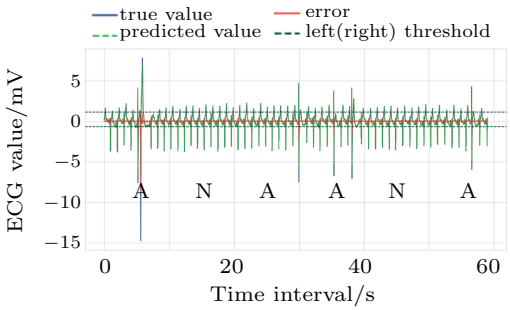


Fig. 8. The 1-min sequence test result for No. 114 patient (N: normal, A: abnormal).

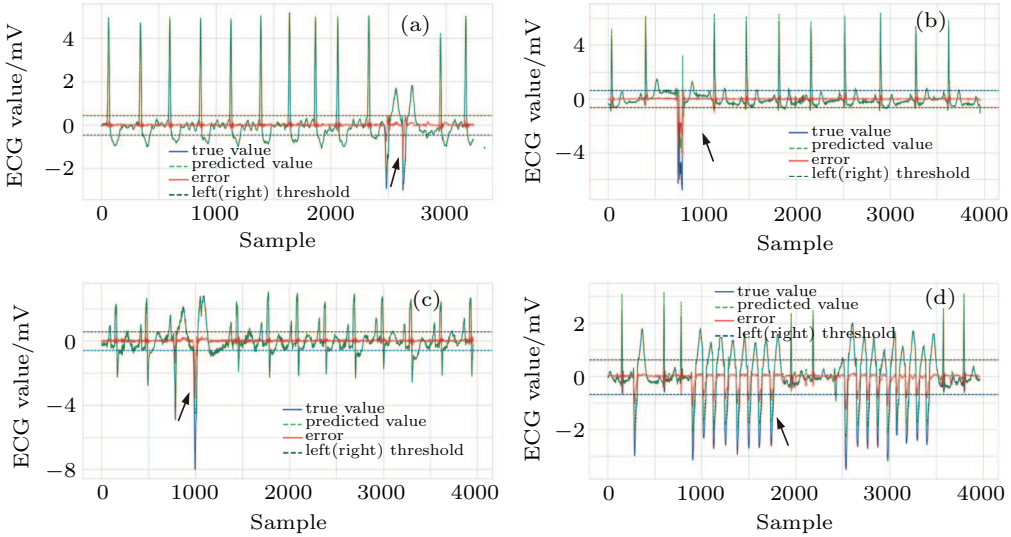


Fig. 9. Successfully detected different kinds of anomalies in the test set, (a) ventricular couplets, (b) unclassifiable beat, (c) fusion of ventricular and normal beat, (d) ventricular tachycardia.

The experiment is carried out with Intel Core i7-8700 K CPU @3.70 GHz, 32GB RAM, GeForce GTX 1080 Ti GPU, and our network is built based on TensorFlow. It takes an average of about 10 min to train the model, and about 1.5 min to test the 30-min data offline.

4.4. Discussion

Among the methods for solving two-class classification with an ECG segment as a unit, our proposed method seems to be conservative, and the obtained ESR 53.89% in the test is not very high. This is due to our strict method constraint that FN must be extremely close to zero. Under such a strict condition, the error threshold interval is narrower and the prediction errors of the normal points are easier to exceed the threshold interval and be misjudged to abnormal, as shown in Fig. 10. However, the authors think it is reasonable to have the strict constraint in the preliminary seeming stage to make sure the avoiding of the abnormal segments dropping.

Moreover, in terms of test set size, our method is tested on 16 complete 30-min sequences from MIT-BIH arrhythmia database, and the test results corresponding to each patient were given, while Chauhan and Vig[8] only mentioned a 20-min sequence consisting of several positive and negative ECG segments for test. Thus it is unfair to compare two methods straightly.

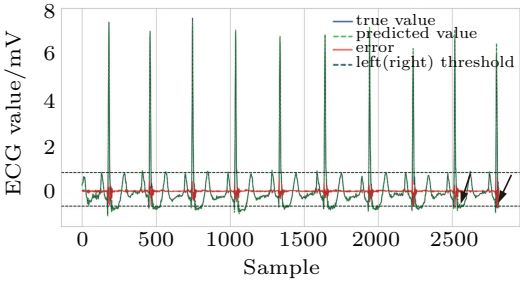


Fig. 10. The normal points mistaken to abnormal.

On the other hand, our test is limited by the data length of MIT-BIH database. For some test data we can find long continuous normal sequence in them, but for others such as data 105, 210, 213, we can only pick out continuous normal sequences of about 1 min. This is why we take different lengths of nor-

mal sequences in the above training and the length of the training set cannot be uniformly fixed for all patients. However, in practical wearable applications, it is believed that, continuous normal sequences usually occupy a large certain proportion in Holter data.

We do further analysis to the influence on the screening result of our problem. According to the length size of the training sequences in the experiment, samples are divided into five class, letting the number of training sequences in each class is 3–4 (each class approximately has equal samples). Then, the mean values of ESR in each class are respectively counted, as shown in Fig. 11. With the length of the training set increasing, the mean value of ESR also increases significantly. Comparing with the 1–1.5 min training set, the 6–9 min training set gives the method nearly 3 times boost in performance.

In the above experiments, such as Nos. 223 and 213, the length of the training set is below 2.5 min, so that the corresponding effective screening rate (ESR) is negatively affected. However, it is also observed that a small training set does not necessarily results in a poor experiment result, such as No. 219.

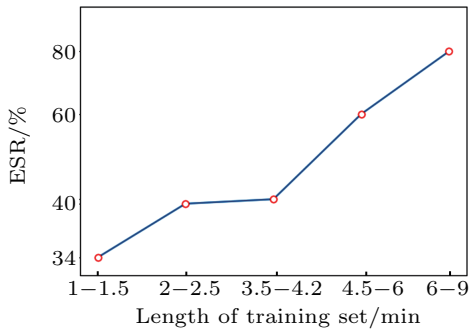


Fig. 11. Length of training sets and the corresponding ESRs.

5. Conclusion

Among the existing analysis methods of ECG segments, the knowledge of abnormal data is usually needed to determine the judgment threshold, but the actual disease data is rare. Also, for clinical monitoring requirements, it should be ensured that abnormal data will not be misjudged as normal data. We have innovatively proposed a preliminary anomaly screening method for Holter data based on the LSTM model.

In our proposed method, the LSTM model is utilized to learn the pseudo-periodic mode of the normal ECG signal points, and the kernel density estimation method is utilized to calculate the probability density of prediction errors of normal data. Based only on the normal signal points, we can achieve the threshold interval that distinguishes the abnormal segments from the normal ones.

The experiments show that the proposed method can filter out a large number of normal segments that do not require further analysis, reducing the workload of subsequent multiple disease classifications. It achieves a certain effect of distinguishing normal and abnormal conditions. The stability of the model is good. The average effective screening rate obtained by repeated tests is 53.89%. The normal segments are excluded to a large extent while ensuring that no abnormal signal point is missed.

For the multi-channel ECG data, our model only needs to change the input dimension on the basis of the number of channels, and it can work.

References

[1] de Albuquerque V H C, Nunes T M, Pereira D R, Luz E J D S, Menotti D, Papa J P and Tavares J M R 2018 *Neural Comput. Appl.* **29** 679
[2] Lin C C and Yang C M 2014 *International Symposium on Computer, Consumer and Control (IS3C)*, June 10–12, 2014, Taichung, Taiwan, China, pp. 650–653
[3] Ye C, Coimbra M T and Kumar B V K V 2010 *Annual International Conference of the IEEE Engineering in Medicine and Biology*, August 31–September 4, 2010, Buenos Aires, Argentina, p. 1918
[4] Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 *Circulation* **101** e215
[5] Acharya U R, Oh S L, Hagiwara Y, Tan J H, Adam M, Gertych A and Tan R S 2017 *Comput. Biol. & Med.* **89** 389
[6] Chuah M C and Fu F 2007 *Frontiers of High Performance Computing and Networking ISPA, 2007 Workshops*, August 28–September 1, 2007, Niagara Falls, Canada, pp. 123–135
[7] Acharya U R, Fujita H, Lih O S, Hagiwara Y, Tan J H and Adam M 2017 *Inf. Sci.* **405** 81
[8] Chauhan S and Vig L 2015 *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, October 9–21, 2015, Paris, France, p. 1
[9] Hochreiter S and Schmidhuber J 1997 *Neural Comput.* **9** 1735
[10] Krizhevsky A, Sutskever I and Hinton G E 2012 *Proceedings of the 25th International Conference on Neu ral Information Processing Systems*, December 3–6, 2012, Lake Tahoe, Nevada, p. 1097
[11] Hinton G E, Srivastava N, Krizhevsky A, Sutskever I and Salakhutdinov R R 2012 arXiv preprint arXiv:1207.0580.
[12] Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 *J. Mach. Learn. Res.* **15** 1929
[13] Khosrow D 1986 *Technometrics* **29** 495
[14] EC57 A A 1998 *ANSI/AAMI EC57: 2012 Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms*
[15] Singh B N and Tiwari 2006 *Digital Signal Process.* **16** 275
[16] Nair V and Hinton G E 2010 *Proceedings of the 27th international conference on machine learning*, June 21–24, 2010, Haifa, Israel, pp. 807–814