



Automated detection of cardiovascular disease by electrocardiogram signal analysis: a deep learning system

Xin Zhang^{1,2,3,4}, Kai Gu⁵, Shumei Miao^{1,2,3}, Xiaoliang Zhang^{1,2,3}, Yuechuchu Yin^{1,2,3}, Cheng Wan^{2,3}, Yun Yu^{2,3}, Jie Hu^{2,3}, Zhongmin Wang^{1,2,3}, Tao Shan^{1,2,3}, Shenqi Jing^{1,2,3}, Wenming Wang^{1,2,3}, Yun Ge⁴, Yin Chen⁴, Jianjun Guo^{1,2,3}, Yun Liu^{1,2,3}

¹Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China; ²Department of Information, The First Affiliated Hospital, Nanjing Medical University, Nanjing 210029, China; ³Institute of Medical Informatics and Management, Nanjing Medical University, Nanjing 210029, China; ⁴School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China; ⁵Division of Cardiology, The First Affiliated Hospital, Nanjing Medical University, Nanjing 210029, China

Correspondence to: Yun Liu. Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China. Email: liyun@njmu.edu.cn.

Abstract: Automated electrocardiogram (ECG) diagnosis could be a useful aid for clinical use. We applied a deep learning method to build a system for automated detection and classification of ECG signals. We first trained a convolutional neural network (CNN) to detect cardiovascular disease in ECG signals using a training data set of 259,789 ECG signals collected from the cardiac function rooms of a tertiary care hospital. The CNN classification was validated using an independent test data set of 18,018 ECG signals. The labels used covered >90% of clinical diagnoses. The system grouped ECGs into 18 classifications—17 different types of abnormalities and normal ECG. The overall accuracy of the model was tested and found to be close to 95%; the accuracy for diagnosis of normal rhythm/atrial fibrillation was 99.15%. The proposed CNN model could help reduce misdiagnosis and missed diagnosis in primary care settings and also improve efficiency and save manpower cost for large general hospitals.

Keywords: Deep learning; electrocardiogram (ECG); neural network; algorithm

Submitted Sep 27, 2019. Accepted for publication Dec 09, 2019.

doi: 10.21037/cdt.2019.12.10

View this article at: <http://dx.doi.org/10.21037/cdt.2019.12.10>

Introduction

Automated analysis of electrocardiogram (ECG) patterns could help in prompt detection of life-threatening arrhythmias such as atrioventricular block, ventricular tachycardia, and atrial fibrillation and be of great help to clinicians (1-4). Such systems will have to use algorithms to identify different waveform types in an ECG and recognize complex relationships between them over time. However, wide variability in wave morphology between patients and the presence of noise are major challenges (3).

Computerized recognition of ECG abnormalities is routinely used by cardiologists classifying long-term ECG records. Feature extraction methods include wave shape functions (5,6), Hermite functions (7), wavelet-based

features (8-10), and statistical features (11). Methodologies to classify these extracted features include support vector machines (12), k-th nearest-neighbor rules (13,14), decision trees (12), artificial neural networks (10,15-21), and linear discriminants (5). State of the art automated ECG recognition systems often rely on a pattern-matching framework that represents the ECG signal as a sequence of stochastic patterns. They require complex feature extraction methods and high sampling rates and are therefore time taking (1). For real-time implementation in the clinic at reasonable cost these systems must use a simple set of features and a lower sampling rate.

A limitation of several algorithms that are used for automatic classification of ECG is the inability to handle large intraclass variations. They are highly dependent on

supervised training datasets and perform poorly when processing large numbers of new ECG records. In addition, the application of dimensionality reduction algorithm to extract complex features in the transform domain significantly improves the computational complexity of the whole process. Moreover, classifier algorithms do not perform when there are wide interpatient variations in ECG signals. Thus, inconsistent performance makes classifier algorithms unreliable in the clinical setting.

Deep learning is a new machine learning technique that is becoming the mainstream for pattern recognition (22,23). It has been successfully used for object recognition, image verification, classification, and speech recognition. Deep learning approaches have greatly improved the accuracy of recognition tools. They have been used to create a deep, multistage architecture for unsupervised learning and recognition systems. We drew on previous work in convolutional neural networks (CNNs) (3) to build a more accurate and robust approach for automated ECG diagnosis. In this paper we describe our algorithm-based system, which we call the Cardiovascular Disease Whole Process Management Platform.

Materials and methods

Data sets and reference standards

To develop the CNN, we constructed a data set from the ECG management system of the First Affiliated Hospital of Nanjing Medical University. A total of 277,807 12-lead static ECG recorded in the cardiac function rooms of the institute between August 1, 2018, and May 31, 2019, were included in the database. The ECGs lasted for 10–60 seconds, with most being in the range of 24 to 30 seconds. After cleaning, the ECGs were labeled according to clinical diagnosis by two experienced electrocardiologists. In rare cases, disagreements were settled by consultation with a senior cardiologist (a chief physician or an associate chief physician). The data set was randomly separated into training data set (n=259,789) and a testing data set (n=18,018). Each data set contained 18 classes of abnormal and sinus ECG signals (Table 1). Figure 1 shows the data processing flow.

CNN architecture and training

Our deep learning system takes as input an ECG waveform

between 10 and 60 seconds long and outputs a label prediction of one of the 18 rhythm classes, along with a probability distribution over the 18 classes. Figure 2 shows the CNN architecture that was used.

Implementation and optimization

Python 3.5 on the Keras library (TensorFlow background) was used to implement the proposed deep CNN model, which was trained and evaluated using graphics processing unit (NVIDIA Tesla P100) computing in an Ubuntu 16.04 environment. The training for cardiovascular disease detection was fully supervised. It back-propagated the gradients from the fully-connected layer through to the convolutional layers. As a loss function, we minimized the binary cross-entropy to optimize the model parameters. The gradient descent with the Adam update rule was utilized.

Results

Performance evaluation

The diagnostic capability of the proposed system was evaluated in terms of accuracy, precision, and specificity. The basic definitions used were as follows:

Patient: positive for the disease;

Healthy: negative for the disease;

True positive (TP) = the number of cases where the patient was correctly defined;

False positive (FP) = the number of cases where the patient was incorrectly defined;

True negative (TN) = the number of cases where a healthy individual was correctly defined;

False negative (FN) = the number of cases where a healthy individual was incorrectly defined.

The definitions of accuracy (ACC), precision (P), specificity (S) and f1-score are as follows (Eq. [1]–Eq. [4]):

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \quad [1]$$

$$P = \frac{TP}{TP + FN} \quad [2]$$

$$S = \frac{TN}{TN + FP} \quad [3]$$

$$\text{f1-score} = \frac{2TP}{2TP + FP + FN} \quad [4]$$

Table 1 Summary of the ECG rhythm data set

ECG rhythm diagnosis	Training data set (n=259,789)		Testing data set (n=18,018)	
	n	%	n	%
Normal	160,115	61.63	10,801	59.95
Premature atrial beats	6,875	2.65	504	2.80
Atrial fibrillation	6,543	2.52	482	2.68
Atrial flutter	881	0.34	71	0.39
Ventricular premature complex	6,688	2.57	505	2.80
No cardiac electrical activity	854	0.33	83	0.46
Pervious myocardial infarction	330	0.13	15	0.08
Acute myocardial infarction/ST segment elevation	3,996	1.54	276	1.53
Left ventricular high voltage	9,212	3.55	640	3.55
Post-ischemic T-wave changes/ST segment depression	30,535	11.75	2,115	11.74
Hyperkalemia pattern/tall peaked T-wave	1,177	0.45	117	0.65
T-wave abnormalities (peaked, symmetrical, biphasic, flat, inverted)	19,246	7.41	1,548	8.59
Left ventricular hypertrophy	1,314	0.51	85	0.47
First-degree atrioventricular block	732	0.28	37	0.21
Second-degree atrioventricular block	81	0.03	3	0.02
Left bundle branch block	1,028	0.40	84	0.47
Right bundle branch block	11,361	4.37	764	4.24
Ventricular pre-excitation syndrome	431	0.17	25	0.14

ECG, electrocardiogram.

Experimental results

The model was tested on a random sample of 18,018 ECGs. *Table 2* shows the accuracy, precision, specificity and f1-score of every classification. The labels used covers more than 90% of clinical diagnoses. The overall accuracy of the model was nearly 95%; the accuracy of the model for diagnosis of normal rhythm/atrial fibrillation was 99.15%. For atrial fibrillation, the most frequently identified disorder, the accuracy was 98.27%. And in all labels, the highest accuracy is up to 99.75%.

The cardiovascular disease whole process management platform

We established the Cardiovascular Disease Whole Process Management Platform shown in *Figure 3*. The system provides a labeling tool (*Figure 4*). After training the CNN model, the system also offers the result of evaluation (*Figure 5*).

Discussion

In this paper we present a novel application of deep learning for classification of ECGs. Since existing deep learning networks do not have a suitable structure to handle the 12 channels of the ECG recording, we applied the structure of channel convolution.

As *Table 3* shows, we achieved accuracy of 98.27% for recognition of 18-classes of heart rhythms. Our CNN network has achieved good performance under the condition of more classification. Different from other ECG analysis algorithms reported earlier, our system considers 18 classifications. A single ECG tracing might contain multiple main categories and subcategories of the label. The main categories included sinus rhythm, atrial fibrillation, atrial flutter, ventricular premature beat, atrial premature beat, low and flat T-wave, and so on. The main category of "sinus rhythm", for example, could include subcategories such as "sinus arrhythmia" or "sinus tachycardia".

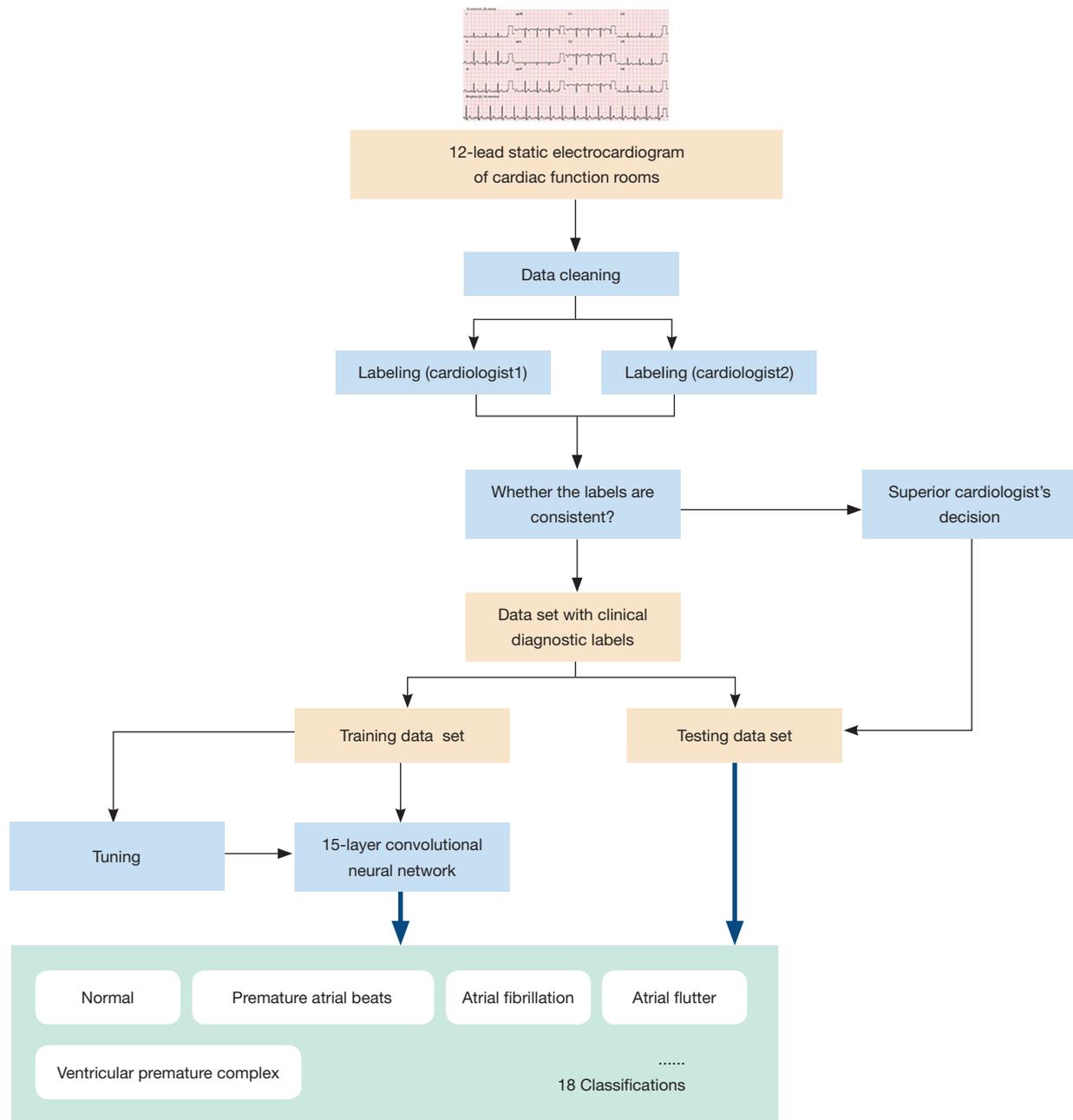


Figure 1 The data processing flow.

Unequal lengths of signals and unbalanced data in ECG signals posed a problem. To solve the problem of unequal lengths of signals, we adopted the method of frame division. To address the issue of unbalanced distribution of abnormal data and normal data, a data amplification method was

introduced to enhance the data.

Some of the published work is based on open datasets. We built our own datasets, and these data sets continue to grow. At present, because some individual labels have not enough data to adjust the parameters of the model,

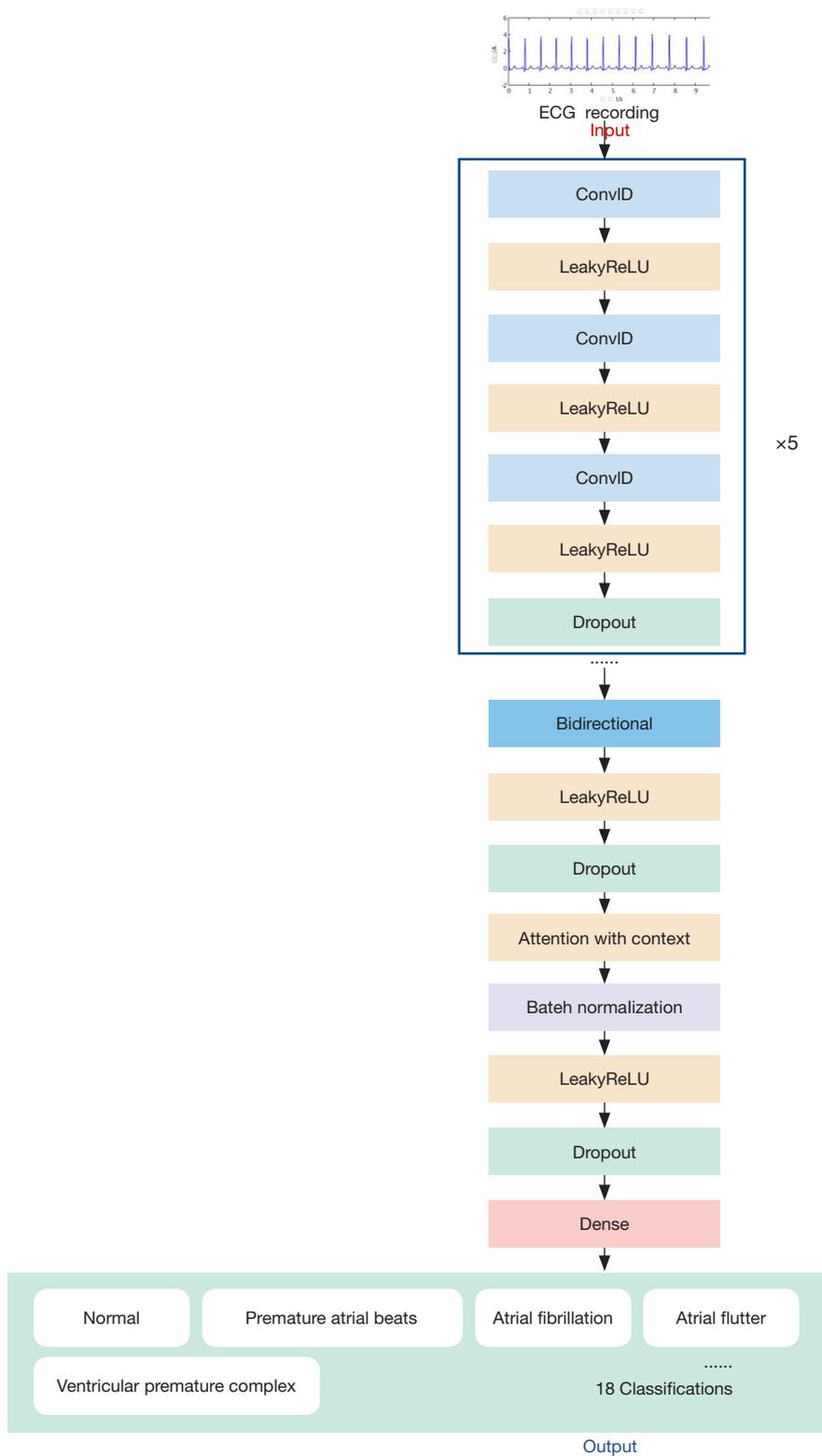


Figure 2 The architecture of the CNN. CNN, convolutional neural network; ECG, electrocardiogram.

Table 2 Accuracy of the proposed automated diagnostic system for different ECG features

Class	ACC, %	P, %	S, %	f1-score, %
Normal	85.49	81.14	88.52	82.11
Premature atrial beats	95.69	38.77	99.80	54.80
Atrial fibrillation	98.27	60.93	99.95	75.24
Atrial flutter	95.42	7.25	99.96	13.42
Ventricular premature complex	98.01	59.29	99.77	72.19
No cardiac electrical activity	99.75	64.84	100.00	78.67
Pervious myocardial infarction	87.28	0.61	99.99	1.21
Acute myocardial infarction/ST segment elevation	90.94	12.78	99.73	22.20
Left ventricular high voltage	96.64	51.46	99.86	67.07
Post-ischemic T-wave changes/ST segment depression	91.10	58.85	97.27	67.98
Hyperkalemia pattern/tall peaked T-wave	95.22	11.16	99.94	19.89
T-wave abnormalities (peaked, symmetrical, biphasic, flat, inverted)	89.39	43.77	98.19	57.14
Left ventricular hypertrophy	95.54	8.83	99.95	16.09
First-degree atrioventricular block	94.75	3.00	99.95	5.78
Second-degree atrioventricular block	92.16	0.21	100.00	0.42
Left bundle branch block	98.77	26.85	99.98	41.88
Right bundle branch block	96.41	54.34	99.85	69.59
Ventricular pre-excitation syndrome	96.49	3.23	99.98	6.22

ECG, electrocardiogram; ACC, accuracy; P, precision; S, sensitivity.

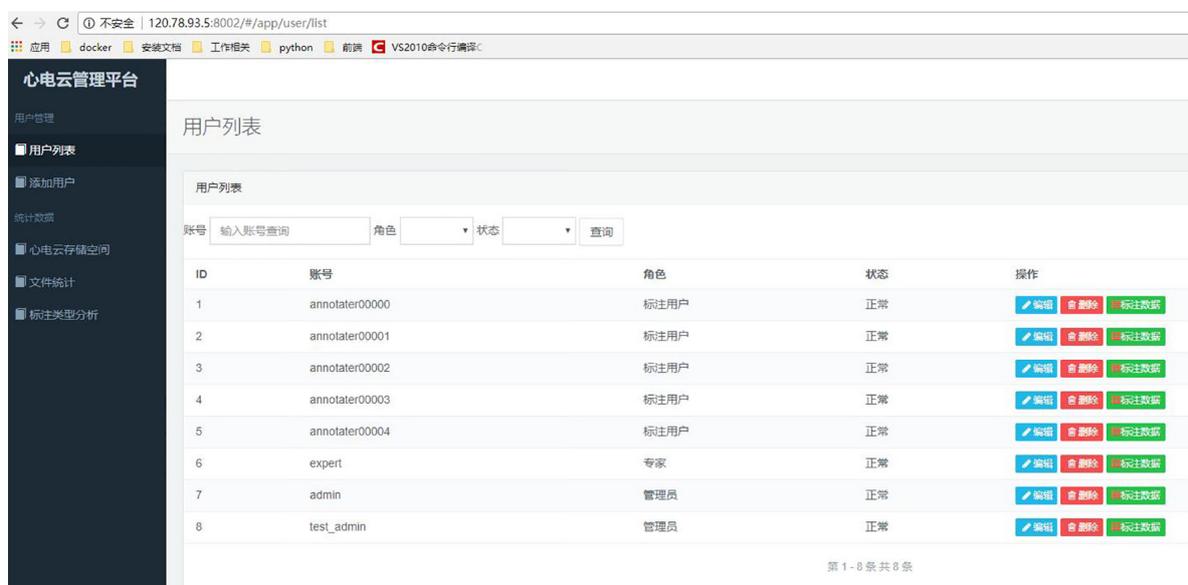
**Figure 3** The interface of the cardiovascular disease whole process management platform.



Figure 4 The interface of the labeling tool.

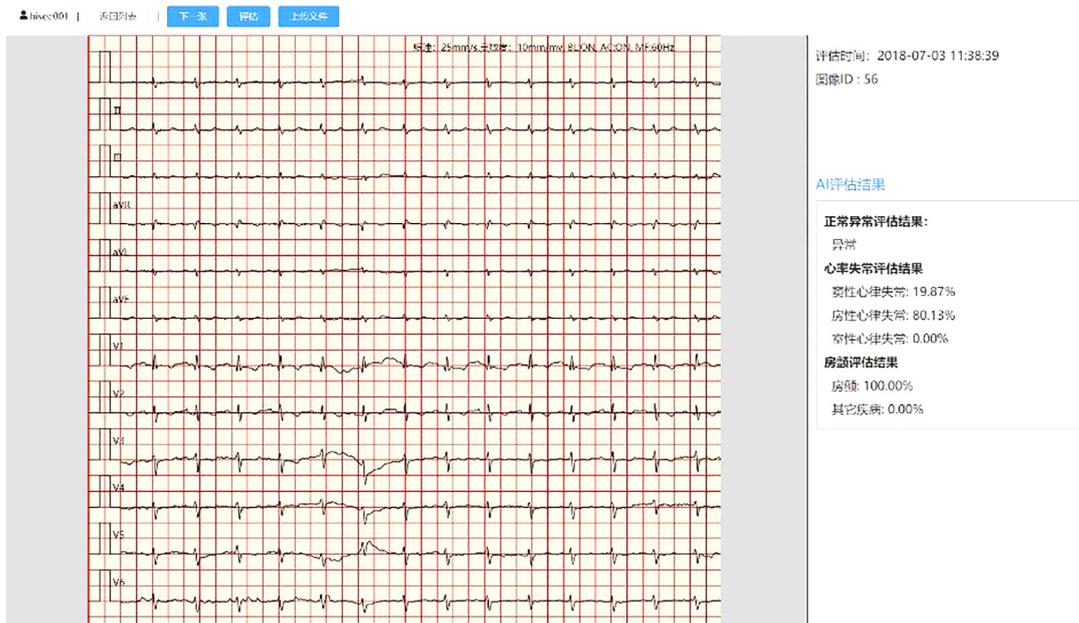


Figure 5 The interface of the evaluate result in the platform.

Table 3 Comparison between the related work and the method proposed in this work

Works	Year	Classes	Methods	ACC, %	P, %	S, %
Jung and Lee (14)	2017	4 beat types	WKNN	96.12	96.12	99.97
Li <i>et al.</i> (10)	2017	6 beat types	GA-BPNN	97.78	97.86	99.54
Kachuee <i>et al.</i> (20)	2018	5 beat types	Deep CNN	93.4	95.1	95.2
Yildirim (17)	2018	5 beat types	DULSTM-WS2	99.25	–	–
Oh <i>et al.</i> (18)	2018	5 beat types	CNN-LSTM	98.10	97.50	98.70
Pandey <i>et al.</i> (21)	2019	5 beat types	CNN	98.3	95.51	86.06
Yildirim <i>et al.</i> (19)	2019	5 beat types	LSTM	99.23	99.00	99.00
Gao <i>et al.</i> (16)	2019	8 beat types	LSTM, FL	99.26	99.26	99.14
Our work	2019	18 beat types	CNN	98.27	60.93	99.95

ACC, accuracy; P, precision; S, sensitivity; WKNN, weighted k-nearest neighbor; GA-BPNN, genetic algorithm-backpropagation neural network; CNN, convolutional neural network; LSTM, long short-term memory; DULSTM-WS2, deep unidirectional LSTM network-based wavelet sequences 2; FL, focal loss.

the individual training effect is not ideal. We are gradually accumulating data and learning.

Conclusions

With the development of optimization methods for processing of the large amounts of data being accumulated, the sensitivity and specificity of automated ECG diagnosis will improve. The AI-aided ECG diagnosis system that we developed appears to be sufficiently reliable for clinical use. It could help reduce misdiagnosis and missed diagnosis in the primary care setting and also save manpower costs for large general hospitals.

Future research should attempt to improve the sensitivity and specificity in the individual classifications by adjusting the different parameters. Machine learning could also be combined with other techniques such as computational modeling and simulation to explain the results of machine learning. That will make the clinical application of the proposed system more interpretable and more credible.

Acknowledgments

Funding: This work was supported by grants from the National key Research & Development plan of the Ministry of Science and Technology of the People's Republic of China (grant no. 2018YFC1314900, 2018YFC1314901), the 2018 provincial industrial and information industry transformation and upgrading project [grant no. (2018)0419, (2017)79], and the 2016 projects of Nanjing Science Bureau

(grant no. 201608003). Yun Liu is the guarantor of this paper.

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/cdt.2019.12.10>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The protocol was approved by the Ethics Committee of Nanjing Medical University [2019(373)].

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Mathews SM, Kambhamettu C, Barner KE. A novel

- application of deep learning for single-lead ECG classification. *Comput Biol Med* 2018;99:53-62.
2. Glass L. Cardiac oscillations and arrhythmia analysis. In: Deisboeck TS, Kresh JY. editors. *Complex systems science in biomedicine*. Boston: Springer, 2006:409-22.
 3. Rajpurkar P, Hannun AY, Haghpanahi M, et al. Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*, 2017.
 4. Mincholé A, Camps J, Lyon A, et al. Machine learning in the electrocardiogram. *J Electrocardiol* 2019;57S:S61-4.
 5. de Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans Biomed Eng* 2004;51:1196-206.
 6. Ye C, Kumar BV, Coimbra MT. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Trans Biomed Eng* 2012;59:2930-41.
 7. Lagerholm M, Peterson C, Braccini G, et al. Clustering ECG complexes using hermite functions and self-organizing maps. *IEEE Trans Biomed Eng* 2000;47:838-48.
 8. Ince T, Kiranyaz S, Gabbouj M. A generic and robust system for automated patient-specific classification of ECG signals. *IEEE Trans Biomed Eng* 2009;56:1415-26.
 9. Senhadji L, Carrault G, Bellanger JJ, et al. Comparing wavelet transforms for recognizing cardiac patterns. *IEEE Engineering in Medicine and Biology Magazine* 1995;14:167-73.
 10. Li H, Yuan D, Ma X, et al. Genetic algorithm for the optimization of features and neural networks in ECG signals classification. *Sci Rep* 2017;7:41011.
 11. de Lannoy G, Francois D, Delbeke J, et al. Weighted conditional random fields for supervised interpatient heartbeat classification. *IEEE Trans Biomed Eng* 2012;59:241-7.
 12. Rodríguez J, Goñi A, Illarramendi A. Real-time classification of ECGs on a PDA. *IEEE Trans Inf Technol Biomed* 2005;9:23-34.
 13. Christov I, Jekova I, Bortolan G. Premature ventricular contraction classification by the Kth nearest-neighbours rule. *Physiol Meas* 2005;26:123-30.
 14. Jung WH, Lee SG. An arrhythmia classification method in utilizing the weighted KNN and the fitness rule. *IRBM* 2017;38:138-48.
 15. Jiang W, Kong SG. Block-based neural networks for personalized ECG signal classification. *IEEE Trans Neural Netw* 2007;18:1750-61.
 16. Gao J, Zhang H, Lu P, et al. An effective LSTM recurrent network to detect arrhythmia on imbalanced ECG dataset. *J Healthc Eng* 2019;2019:6320651.
 17. Yildirim Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Comput Biol Med* 2018;96:189-202.
 18. Oh SL, Ng EYK, Tan RS, et al. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput Biol Med* 2018;102:278-87.
 19. Yildirim O, Baloglu UB, Tan RS, et al. A new approach for arrhythmia classification using deep coded features and LSTM networks. *Comput Methods Programs Biomed* 2019;176:121-33.
 20. Kachuee M, Fazeli S, Sarrafzadeh M. Ecg heartbeat classification: a deep transferable representation. 2018 IEEE International Conference on Healthcare Informatics (ICHI). *IEEE* 2018:443-4.
 21. Pandey SK, Janghel RR. Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE. *Australas Phys Eng Sci Med* 2019;42:1129-39.
 22. Sharif Razavian A, Azizpour H, Sullivan J, et al. CNN features off-the-shelf: an astounding baseline for recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2014:806-13.
 23. Tompson J, Stein M, Lecun Y, et al. Real-time continuous pose recovery of human hands using convolutional networks. *ACM Transactions on Graphics (ToG)* 2014;33:169.

Cite this article as: Zhang X, Gu K, Miao S, Zhang X, Yin Y, Wan C, Yu Y, Hu J, Wang Z, Shan T, Jing S, Wang W, Ge Y, Chen Y, Guo J, Liu Y. Automated detection of cardiovascular disease by electrocardiogram signal analysis: a deep learning system. *Cardiovasc Diagn Ther* 2020;10(2):227-235. doi: 10.21037/cdt.2019.12.10