IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 30, 2019, accepted January 1, 2020, date of publication January 8, 2020, date of current version January 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964749

ECG Heartbeat Classification Using Convolutional Neural Networks

XUEXIANG XU AND HONGXING LIU
School of Electronic Science and Engineering, Nanjing University, Nanjing 210008, China
Corresponding author: Hongxing Liu (njhxliu@nju.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61271079. The URLs of the funders' Website is http://www.nsfc.gov.cn/

ABSTRACT Electrocardiogram (ECG) data recorded by Holter monitors are extremely hard to analyze manually. Therefore, it is necessary to automatically analyze and categorize each heartbeat using a computer-aid method. Because convolutional neural networks (CNNs) can classify ECG signals automatically without trivial manual feature extractions, they have received extensive attention. However, it is anticipated that improving the existing CNN classifiers might provide better overall accuracy, sensitivity, positive predictivity, etc. In this study, we proposed a CNN based ECG heartbeat classification method. Based on the MIT-BIH arrhythmia database, our proposed method achieved a sensitivity of 99.2% and positive predictivity of 99.4% in VEB detection; a sensitivity of 97.5% and positive predictivity of 99.1% in SVEB detection; and an overall accuracy of 99.43%. Our proposed system can be directly implemented on wearable devices to monitor long-term ECG data.

INDEX TERMS Heartbeats, Holter, convolutional neural networks, MIT-BIH arrhythmia database, electrocardiogram signals.

I. INTRODUCTION

In cases with suspected arrhythmias, doctors often ask the subjects to wear a Holter to continuously record ECG data for 24 hours or longer. Because the amount of ECG data recorded by the Holter is extremely large, it is necessary to analyze the recordings using a computer and categorize the type of each heartbeat automatically [1].

According to the Association for the Advancement of Medical Instrumentation (AAMI) [2], non-life-threatening arrhythmia signals can be divided into five categories: non-ectopic beat (N), supra ventricular ectopic beat (SVEB, S), ventricular ectopic beat (VEB, V), fusion (F), and unknown (Q). In this paper, we focus on classifying these 5 types of arrhythmia heartbeat.

With the wide use of convolutional neural networks (CNN) in many fields [8]–[10], CNN have also become a popular option to automatically classify ECG signals recorded by the Holter [3]–[7]. Compared with traditional methods, the CNN classifier can directly input heartbeats without additional feature extraction and selection; it also demonstrates competitiveness in classification accuracy [3], [11], [16].

Many approaches on arrhythmia heartbeat classification with CNN have been proposed. Kiranyaz *et al.* [3] improved the two-dimensional (2-D) CNN used for images, and first proposed a 1-D CNN for the automatic classification of heartbeats. Acharya *et al.* [4] balanced the five types of heartbeats by oversampling, and proposed a nine-layer CNN structure. Zhai *et al.* [5] coupled two adjacent sets of dual ECG heartbeat data to generate a dual heartbeat coupling matrix, which was used as an input sample and subsequently sent to a seven-layer 2-D CNN for classification. Xiang *et al.* [6] added a manual extraction of RR interval features to facilitate training and proposed a two-level CNN structure. Sellami and Hwang [7] proposed a robust deep convolutional neural network with batch-weighted loss.

The results of the researches mentioned above were remarkable. However, none of them fully addressed high accuracy, high sensitivity, high specificity and high positive predictive value of classification at the same time. Consequently, we intend to further optimize the CNN heartbeat classifier in this study.

II. MATERIALS

A. DATABASE

Data used in this study were obtained from the MIT-BIH Arrhythmia Database [14], [15]. The database contains 48 sets of dual-channel ECG data obtained from 47 subjects, and the data have a sampling rate of 360 Hz. Each set of data has a length of approximately 30 min and contains recordings from two leads: one is MLII and the other is one of V1, V2, V4, or V5. The position of each R-peak was marked and the

The associate editor coordinating the review of this manuscript and approving it for publication was Dusmanta Kumar Mohanta.

TABLE 1. Summary of the 5 types of heartbeats.

Heartbeat Type	Number
Non-ectopic (N)	90265
Supra ventricular ectopic (SVEB)	2503
Ventricular ectopic (VEB)	7106
Fusion (F)	789
Unknown (Q)	7016
Total	107679

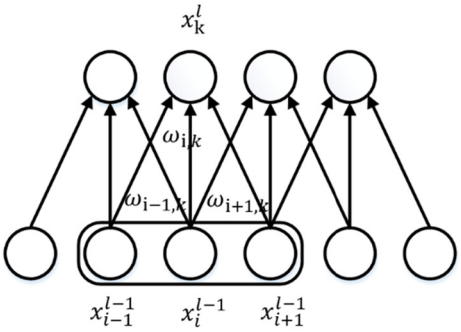


FIGURE 1. Schematic diagram of the convolution layer connection of a CNN.

corresponding category was manually annotated by two or more cardiovascular disease experts. In our experiments, only data from the MLII lead were used.

According to the R-peaks information available in the database, we segmented the signal into heartbeats centered about each R-peak and noted their given corresponding types. Each heartbeat consists of 300 sample points (150 samples before the R-peak and 149 samples after the peak). In light of the AAMI recommendations, we extracted 107679 ECG signal segments of five types: non-ectopic, supra ventricular ectopic, ventricular ectopic, fusion, and unknown, as shown in Table 1.

B. CONVOLUTIONAL NEURAL NETWORKS

Traditional neural networks use matrix multiplication to establish a connection between the input and output, meaning that each output unit interacts with each input unit, i.e., the so-called fully connected network. As the number of neurons increase, there will be more calculations involved.

An important feature of the CNNs is the introduction of convolutional layers [10], [17]. The convolutional layer replaces the general matrix multiplication operation by convoluting with the input using the convolution kernel. Figure 1 shows a convolutional layer connection with a convolution kernel of size three. As shown in the figure, each neuron in convolutional layer l is only connected to three neurons of layer l-1 with the same set of three weight values, that is, the so-called parameter sharing. Because the size of the convolution kernel is significantly smaller than the input size, it is used to perform sparse interactions. The characteristics of sparse interaction and parameter sharing reduce the number

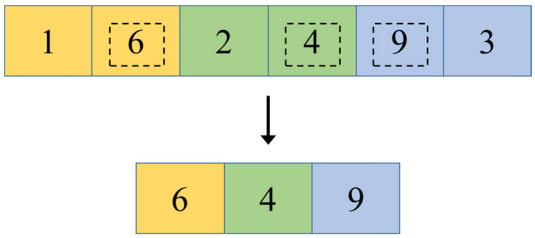


FIGURE 2. Schematic diagram of pooling layer connection of a convolutional neural network.

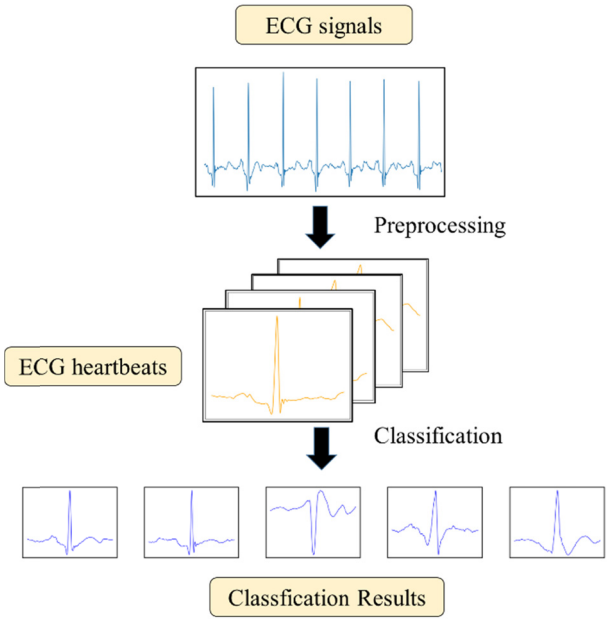


FIGURE 3. Main procedures of ECG heartbeats classification.

of parameters of the CNN and computational complexity of the neural network considerably.

In addition, the CNN typically introduces a pooling layer, also called a subsampling layer, to further reduce the number of computations. The pooling layer uses the overall characteristics of the adjacent area of an input location to represent the characteristics of the area. For instance, the maximum pooling uses the maximum value in the adjacent area as the output, and the average pooling uses the average value in the adjacent area as the output. Figure 2 shows a pooled layer connection with a maximum pooling of two steps. Because pooling combines the feedbacks of all adjacent units, the number of units after pooling are reduced significantly. Therefore, the calculation efficiency is increased.

III. PROPOSED METHOD

In this paper, all the ECG signals are retrieved from MIT-BIH arrhythmia database as we mentioned before. Our ECG heartbeat classifier is composed of two main steps: preprocessing and classification. Overall steps are shown in figure 3:

A. PREPROCESSING

The raw signal from MIT-BIH is corrupted by noise such as myoelectric interference, power line interference and baseline drift. To remove those noise, the raw ECG signal is

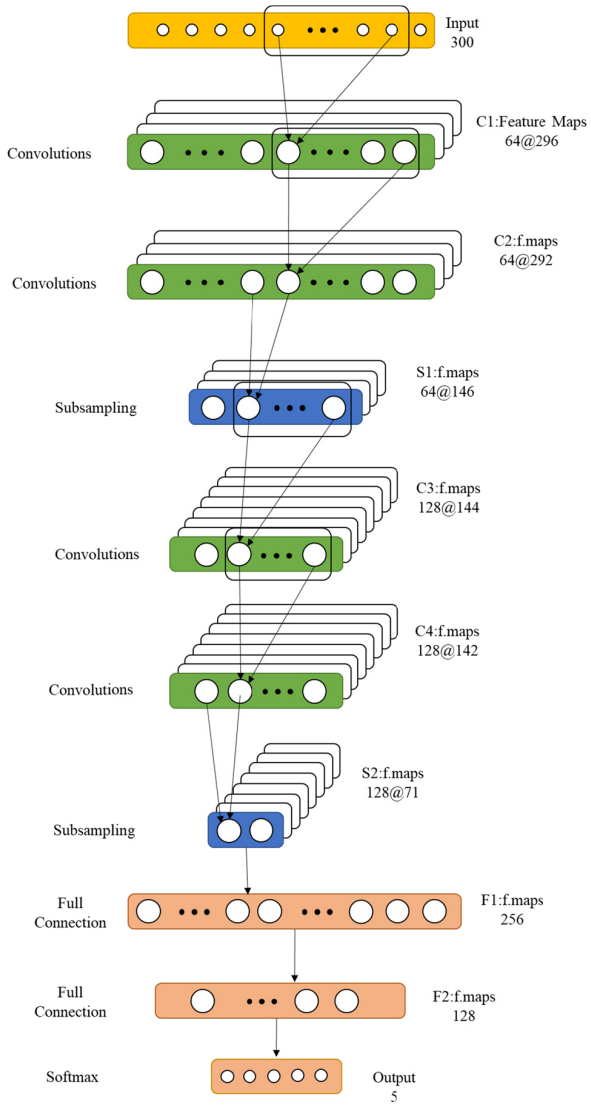


FIGURE 4. Architecture of the proposed CNN model.

denoising by wavelet filters. Raw signal is decomposed by Daubechies wavelet 6 (db6) to six levels and the wavelet coefficients from 3rd level to 6th level were retained and used for reconstructing the signal [12].

After denoising, we segmented the signal to heartbeats by leveraging the position information of R-peaks annotated by the MIT-BIH arrhythmia database. Each heartbeat consists of 300 samples (149 before R-peak position and 150 after R-peak position).

B. CLASSIFIER ARCHITECTURE

Figure 4 shows a schematic of our CNN classifier. The network consists of nine layers, including four convolutional layers, two subsampling layers, two fully connected layers, and one Softmax layer.

Convolutional layers C1, C2, C3, and C4 perform the convolution operations as expressed in equation (1) on the output of a previous layer using the current convolution kernel, and their convolution kernel sizes are 5, 5, 3, and 3, respectively.

$$x_k^l = f(\sum_{i \in M_k} x_i^{l-1} * \omega_{ik} + b_k) \quad (1)$$

where M_k is the effective range of the convolution kernel, x_k^l represents the output of the kth neuron in layer l, b_k is the bias of the kth neuron in layer l, ω_{ik} is the weight kernel between the ith neurons in layer *l-1* and the kth neuron in layer l, and f(·) represents the Rectified Linear Unit (ReLU) activation function [18].

Subsampling layers (S1 and S2) were used to reduce the input size of the next layer, compress the dimension of the ECG data, reduce the amount of calculation, and further extract useful features. We used the max-pooling function to obtain the maximum value among the adjacent region to replace it. The output of the kth neuron of the subsampling layer l is calculated by equation (2).

$$x_k^l = subsample\left(x_{k_{cluster}}^{l-1}\right) \quad (2)$$

where x_k^l represents the output of the kth neuron of layer l, down represents the subsampling operation, and $x_{k\_cluster}^{l-1}$ represents the kth output cluster of layer *l-1*.

Fully connected layers (F1, F2, and F3) were used to further increase the number of nonlinear operations. The output of the kth neuron in fully connected layer l was calculated using equation (3).

$$x_k^l = f(\sum_{i=1}^{N} x_i^{l-1} * \omega_{ik} + b_k) \quad (3)$$

where x_k^l represents the output of the kth neuron in layer l, b_k is the bias of the kth neuron in layer l, ω_{ik} is the weight vector between the kth neuron in layer l and the ith neuron in layer *l-1*, and N is the total number of neurons in layer *l-1*.

In this study, we used a rectified linear unit as the activation function in both the convolutional layers and fully connected layers [18]–[20]. In the output layer, we used the Softmax activation function to obtain five categories of heartbeats.

C. TRAINING METHOD

The purpose of the training is to reduce the value of the loss function L by adjusting the weights and biases. We used the cross entropy function as the loss function (4):

$$\mathrm{L} = -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log\left(\hat{y}_n\right) + (1 - y_n)\log\left(1 - \hat{y}_n\right)\right] \quad (4)$$

where N represents the total number of neurons in the output layer, y_n represents the actual category, and $\hat{y}_n$ represents the predicted category. We update the weights and offsets using the Adam algorithm [21], which is as follows:

First, a batch of samples was sent to calculate the gradient equation (5), and we set the batch size as 256:

$$g = \frac{1}{m}\nabla_\theta \sum_i L\left(f\left(x^{(i)}; \theta\right), y^{(i)}\right) \quad (5)$$

where g is the gradient, m is the batch size, θ is the parameter to be updated, $f\left(x^{(i)}; \theta\right)$ is the heartbeat type predicted by the

FIGURE 5. Distribution of ECG segments used for training and testing.

ith sample, $y^{(i)}$ is the actual type of the ith sample, and L is the loss function.

Subsequently, we update the biased first moment estimate and the biased second moment. Further, we correct the biases in the first and second moments.

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \tag{6}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \tag{7}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{8}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{9}$$

In the formula above, m_t and v_t are the first and second moment estimates of the gradient, respectively. $\hat{m}_t$ and $\hat{v}_t$ are the corresponding corrections of the biases. β_1 and β_2 are the decay rates for the moment estimates, set as 0.9 and 0.999, respectively.

$$\theta\,(t) = \theta\,(\mathrm{t}-1) - \frac{1}{\sqrt{\hat{v}_t}+\epsilon}\hat{m}_t \tag{10}$$

Finally, the parameter θ is updated by modifying the corrected bias in the first moment $\hat{m}_t$, the corrected bias in the second moment $\hat{v}_t$, and the step ϵ (set to 0.001):

After replacing θ with ω_{ik} and b_k, we can obtain the update of the weight and bias as follows:

$$\omega_{ik}\,(t) = \omega_{ik}\,(\mathrm{t}-1) - \frac{1}{\sqrt{\hat{v}_t}+\epsilon}\hat{m}_t \tag{11}$$

$$b_k\,(t) = b_k\,(\mathrm{t}-1) - \frac{1}{\sqrt{\hat{v}_t}+\epsilon}\hat{m}_t \tag{12}$$

Dropout [22], [23] was used to prevent overfitting on convolutional layers C2 and C4; and fully connected layers F1 and F2. Dropout allows for the weights of the hidden layer neurons to be set randomly to zero during training, which invalidates these nodes. A large number of experiments prove that the introduction of the dropout layer can reduce the possibility of model overfitting considerably.

IV. EXPERIMENTS

We used stratified sampling to divide 90% of the samples into the training set and 10% into the test set. The training and test sets are mutually exclusive. Subsequently, in the training set, 90% of the samples are divided into the training set, and 10% into the validation set by stratified sampling. Similarly, the training and validation sets are mutually exclusive. The division is shown in Figure 5.

The CNN was trained using an Intel Core i7-8700K CPU @3.70 GHz, 32G B RAM, GeForce GTX 1080 Ti GPU.

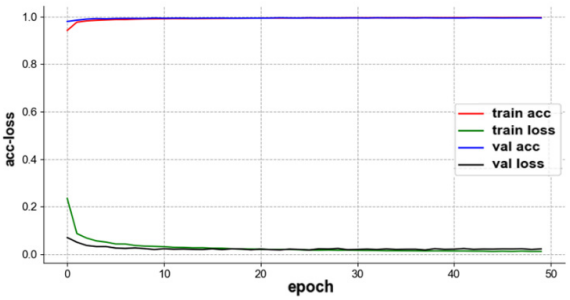


FIGURE 6. Learning curve of model training process.

TABLE 2. Confusion matrix of ECG heartbeats on all samples.

Ground Truth	Classification Result				
	N	S	V	F	Q
N	90213	21	12	18	1
S	62	2440	1	0	0
V	19	0	7054	33	0
F	43	0	26	720	0
Q	1	0	0	0	7015

TABLE 3. Confusion matrix of ECG heartbeats on test samples.

Ground Truth	Classification Result				
	N	S	V	F	Q
N	9041	6	5	1	0
S	22	228	0	0	0
V	10	0	698	3	0
F	9	0	5	65	0
Q	0	0	0	0	702

Our network was based on the TensorFlow framework. After the training was completed, the network parameters were stored in the HDF5 file. The training time of each epoch was approximately 5 s, and the maximum epoch number was set to 50. The total training time was between 260.65 s and 268.46 s. The training process curve is shown in Figure 6. As shown in the figure, the red and blue curves represent the accuracies of the training set and validation set, respectively. The green and black curves represent the losses in the training set and validation set, respectively. As the training progresses, the validation set and training set loss curves coincide. Thus, overfitting does not occur.

Table 2 shows the confusion matrix of the training results on all samples, and Table 3 shows the confusion matrix of the training results on the test set. To prove the stability of the model learning, we performed 20 experiments; the training time and overall accuracy are shown in Table 4. It can be observed that the overall accuracy of our model is between 99.41%–99.49%, and training time is between 260.65–268.46 s. It proves that our model exhibits good stability.

V. DISCUSSION

The result of this study is compared with former studies shown in Table 5. Our improved CNN model is superior

TABLE 4. Overall accuracy and consuming time of 20 repeat training.

Serial Number	Time consumed (s)	Overall accuracy (%)	Serial Number	Time consumed (s)	Serial Number
1	260.65	99.41	11	268.08	99.42
2	261.60	99.41	12	268.46	99.44
3	263.74	99.49	13	268.42	99.43
4	265.66	99.41	14	266.80	99.42
5	267.06	99.48	15	266.50	99.41
6	263.65	99.41	16	267.37	99.41
7	264.60	99.45	17	268.06	99.41
8	266.49	99.43	18	266.49	99.42
9	267.19	99.41	19	267.23	99.46
10	267.99	99.45	20	264.05	99.43

TABLE 5. VEB and SVEB classification performance of previous studies.

Methods	VEB				SVEB			
	Acc	*Sen*	*Spe*	*Ppv*	*Acc*	*Sen*	*Spe*	*Ppr*
Kiranyaz S et al.[3]	99.0	93.9	98.9	90.6	97.6	60.3	99.2	63.5
Acharya et al.[4]	97.9	94.2	98.8	95.3	97.0	90.6	98.6	94.3
Zhai X et al.[5]	99.1	96.4	99.5	96.4	97.3	85.3	98.0	71.8
Yangde X et al.[6]	99.2	93.7	99.6	94.8	98.9	85.0	99.3	78.2
Ali Sellami et al[7]	99.8	99.0	99.9	98.9	99.7	90.8	99.9	97.3
Proposed model	**99.9**	**99.2**	**99.9**	**99.4**	**99.9**	**97.5**	**99.9**	**99.1**

to those in existing literature in terms of VEB or SVEB classification performance.

In Table 5, the parameters Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), and Positive predictive value (Ppv) are defined as follows: $Acc = (TP + TN)/(TP + TN + FP + FN)$, $Sen = TP/(TP + FN)$, $Spe = TN/(TN + FP)$, $Ppv = TP/(TP + FP)$, where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative [13].

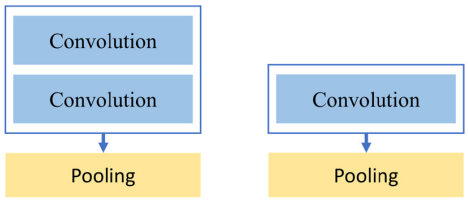
We obtained such good results because of the positive effects of the following two aspects:

A. USING COUPLED-CONVOLUTION LAYER STRUCTURE

In common situation, each convolution layer is followed by a subsampling layer. However, we believe that the mapping relationship between the heartbeat category and its waveform is extremely complicated. Single convolution layer is not complicated enough. Therefore, we use a coupled-convolution structure (shown in Figure 7) to achieve a more powerful fitting capability.

B. BUSING DROPOUT

Many experiments have indicated that the dropout can prevent overfitting effectively [23]. Dropout sets a probability parameter, p. In each training epoch, every node exhibits the probability of p to be maintained. If they are not maintained, their weights will be set to 0. In fact, several subnets are trained during each epoch. At the test or working period, all the weights are multiplied by the probability p. Thus, we can obtain the average result of each subnet.



(a) Coupled-convolution structure (b) Common structure

FIGURE 7. Comparison between common structure and coupled-convolution structure.

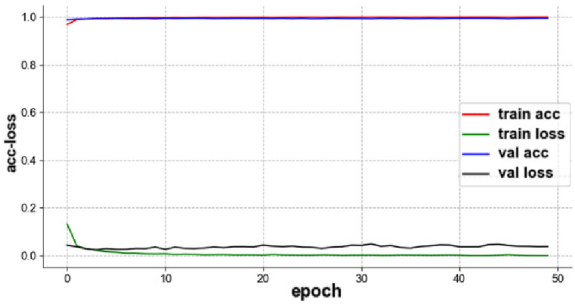


FIGURE 8. Learning curve of model training process without dropout.

In this study, when the dropout is not used, the network training process curve is as shown in Figure 8. In the figure, the red and blue curves represent the accuracies of the training set and validation set, respectively. The green and black curves represent the losses in the training set and validation set, respectively. It is clear that as the training progresses, starting from the third epoch, the validation set loss curve exhibits an upward trend, and intersects the training set loss curve, indicating that the network has been overfitted. In contrast, the CNN network using dropout training in all our previous experiments—conducted 20 times—did not exhibit any overfitting, as shown in Figure 6.

It is worth noting that when applying our trained model in real-life scenarios, we may encounter some other problems. First, the sampling frequency of the training data from the MIT-BIH arrhythmia database is 360 Hz, but some Holters may acquire data with different sampling rates; subsequently, we must resample the acquired data to 360 Hz first before using our model. Next, we should continue to further retrain our trained model with new data to render it more suitable for the new situation while using it in real-life applications. Subsequently, we used only the MLII lead data for our model training; however, in an actual situation, the Holter may collect data from more than 12 leads. Fortunately, only a slight adjustment is required for our model to fit this situation—simply changing the input channels as multichannel ECG signals and setting the initial corresponding input weights to zeros, which will be updated when performing new training.

The model proposed in this paper relies on the correct detection of R waves in practical applications. Hitherto, many studies on R-peak detection have been conducted, and some of them have achieved high accuracies. However, in the actual Holter monitoring, much interference may occur that will affect the correct detection of R waves. Therefore, in the future, we are interested in investigating a robust and high accuracy R-peak detection algorithm.

VI. CONCLUSION

In this study, we designed a Holter data CNN heartbeat classifier based on the MLII lead by using coupled-convolution layer structure and adopting the dropout mechanism. It was proved experimentally that the overall accuracy of the classification could reach 99.43%, while the sensitivity and positive predictivity of VEB could reach 99.2% and 99.4%, respectively. Further, the sensitivity and positive predictivity of VEB could reach 97.5% and 99.1%, respectively. Compared with the methods in previous literature, our model performed better in terms of VEB, SVEB classification accuracy, and overall accuracy. Moreover, our model could achieve an overall accuracy of approximately 99.4% in all 20 repeated experiments, indicating its good learning stability and facilitating the promotion and renewal of the model. For the 107679 samples extracted from the MIT-BIH arrhythmia database, the training time was approximately 4 minutes.

We believed that our trained CNN heartbeat classifier model can be used for real-life and real-time Holter applications [24] with a slight adaptive adjustment before use.

REFERENCES

[1] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 437–448, Sep. 2013.

[2] *Testing and Reporting Performance Results of Cardiac Rhythm and St Segment Measurement Algorithms*, document ANSI/AAMI EC38 1998, Association for the Advancement of Medical Instrumentation, 1998.

[3] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.

[4] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.

[5] X. Zhai and C. Tin, "Automated ECG classification using dual heartbeat coupling based on convolutional neural network," *IEEE Access*, vol. 6, pp. 27465–27472, 2018.

[6] Y. Xiang, J. Luo, T. Zhu, S. Wang, X. Xiang, and J. Meng, "ECG-based heartbeat classification using two-level convolutional neural network and RR interval difference," *IEICE Trans. Inf. Syst.*, vol. E101-D, no. 4, pp. 1189–1198, 2018.

[7] A. Sellami and H. Hwang, "A robust deep convolutional neural network with batch-weighted loss for heartbeat classification," *Expert Syst. Appl.*, vol. 122, pp. 75–84, May 2019.

[8] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Comput. Methods Programs Biomed.*, vol. 161, pp. 1–13, Jul. 2018.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[11] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009.

[12] H. Shi, H. Wang, F. Zhang, Y. Huang, L. Zhao, and C. Liu, "Inter-patient heartbeat classification based on region feature extraction and ensemble classifier," *Biomed. Signal Process. Control*, vol. 51, pp. 97–105, May 2019.

[13] P. Dechazal, M. O'Dwyer, and R. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[14] R. Mark and G. Moody, *MIT-BIH Arrhythmia Database Directory*. Cambridge, MA, USA: Massachusetts Institute of Technology, 1988.

[15] R. Acharya, Eds. *Advances in Cardiac Signal Processing*. Berlin, Germany: Springer, 2007.

[16] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Inf. Sci.*, vol. 405, pp. 81–90, Sep. 2017.

[17] Y. LeCun and Yoshua Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10. Cambridge, MA, USA: MIT Press, 1995, p. 1995.

[18] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012.

[20] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural Comput.*, vol. 7, no. 2, pp. 219–269, Mar. 1995.

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[22] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*. [Online]. Available: https://arxiv.org/abs/1207.0580

[23] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[24] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

XUEXIANG XU received the B.S. degree in microelectronics science and engineering from the Hefei University of Technology, China, in 2017. He is currently pursuing the M.S. degree in electronic information engineering with Nanjing University, China.

HONGXING LIU received the Ph.D. degree from Xi'an Jiaotong University, China, in 1997. He is currently a Professor of information electronics with the School of Electronic Science and Engineering, Nanjing University, China, and also a Master's Tutor for signal and information processing. He is a member of the Biomedical Electronics Branch, Chinese Institute of Electronics. In recent years, he has mainly carried out research in the field of intelligent information processing and biomedical electronics. He has focused on fetal electrocardiogram technology and achieved remarkable results. He has presided over three projects of the National Natural Science Foundation, one social development project of Jiangsu Science and Technology Support Program, more than ten horizontal cooperation projects, and published nearly 100 articles by the first author or correspondent author, including more than 20 articles in SCI. The first inventor applied for 33 invention patents, of which 17 were authorized.