ResNet-TCN: A Joint Model for ECG Heartbeat Classification with High Accuracy

Shaogang Hu
State Key Laboratory of Electronic Thin Films and Integrated Devices
University of Electronic Science and Technology of China
Chengdu, China
sghu@uestc.edu.cn

Renwei Li
School of Software Engineering
Chongqing University of Posts And Telecommunications
Chongqing, China
1005411762@qq.com

Qiang Hu
School of Software Engineering
Chongqing University of Posts And Telecommunications
Chongqing, China
qianghu98@163.com

Guanchao Qiao*
State Key Laboratory of Electronic Thin Films and Integrated Devices
University of Electronic Science and Technology of China
Chengdu, China
gcqiao@uestc.edu.cn

2023 International Conference on Information Networking (ICOIN) | 978-1-6654-6268-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICOIN56518.2023.10048913

***Abstract*—The electrocardiogram (ECG) is a reliable indicator of heart health and is widely used to diagnose arrhythmias. In this work, we propose ResNet-TCN, a joint model based on residual network (ResNet) and temporal convolutional network (TCN) for more accurate ECG classification. Specifically, the ResNet model extracts the spatial features of the signal, the TCN model extracts the temporal features of the signal, and the linear layer combines the features for classification. On the MIT-BIH public dataset, the proposed ResNet-TCN reaches a high accuracy (Acc) of 98.84%, a positive predictive value (P^+) of 93.34%, and a sensitivity (Se) of 91.69%. Ablation studies show that compared with single ResNet or TCN, the Acc and P^+ of ResNet-TCN improved by 0.72% and 5.13%, 0.48% and 5.59%, respectively, and the Se is comparable. Moreover, compared with previous works, the proposed model achieves a 1%~4.8% accuracy improvement. The proposed model is pretty suitable for clinical ECG detection that requires high accuracy.**

Keywords*—*electrocardiogram (ECG), joint model, residual network (ResNet), temporal convolutional network (TCN)

I. Introduction

In recent years, cardiovascular disease has become a major threat to human health. According to official data, the number of people currently suffering from cardiovascular disease in China is about 330 million, and the ensuing death is the first leading cause of death [1]. Clinical studies have shown that arrhythmias frequently precede other cardiovascular diseases and are considered precursors to serious disease [2]. Therefore, timely diagnosis of arrhythmias is of tremendous value for avoiding cardiovascular diseases.

In clinical practice, clinicians diagnose arrhythmias by monitoring the electrocardiogram (ECG). With the development of electronic information computing, automatic classification methods for arrhythmia emerge in an endless stream. The processing steps are generally divided into signal preprocessing, feature wave localization, heartbeat interception, feature extraction, and classification. Feature extraction is a crucial step in automatic classification and directly affects the performance of algorithm. Traditional machine learning involves manual feature extraction to extract characteristics of ECG in the time, frequency, or wavelet domains. The classification effect entirely depends on the manually extracted feature information and requires the support of medical domain knowledge. The method is highly dependent on individuals and lacks generality. Compared with traditional algorithms, end-to-end deep learning algorithms are increasingly being employed by academics for ECG signal classification. Acharya et al. [3] used an 11-layer 1D-CNN model to obtain 92% classification accuracy in the automated categorization of ECG data; Yildirim [4] proposed a deep bidirectional LSTM wavelet sequence model to extract the time characteristics of ECG data, which significantly improved the recognition performance of the network; Brito [5] et al. constructed a deep ResNet model to alleviate the gradient disappearance problem and achieve better classification results; Izci [6] et al. present a simple 2D-CNN model that classifies five diseases of cardiac arrhythmias using ECG grayscale images as input data.

However, it is difficult for a single network model to extract the multi-dimensional characteristics of heartbeat signals. In addition, some methods do not classify according to the AAMI standard, resulting in low comparability of similar studies. Thus, we propose a joint classification model based on ResNet and TCN, termed ResNet-TCN, which can combine features from different dimensions of ECG. The proposed model extracts the heartbeat's spatial characteristics through the ResNet model and the heartbeat's temporal characteristics through the TCN model. The features extracted from the two modules are fused in the linear layer to classify the four arrhythmia categories of the AAMI criteria.

II. Background

A. Data Source

In this work, we use the MIT-BIH [7] dataset, which is jointly by MIT and Beth Israel Hospital in Boston and is the most popularly used ECG signal dataset in ECG signal classification studies. The dataset contains 48 sets of half-hour-long records acquired from the chest locations of 47 patients. The records are in binary format with a 360 Hz sampling frequency, and the type of heartbeat is indicated in the '.atr' file.

In each record, the first channel is the MLII lead, whereas the second channel is typically used in lead V1 and, in a few records, in V2, V5, or V4. To avoid the influence of data origin, we use the first channel data which all recordings are of the same lead and remove the four ECG recordings collected by the pacemaker (102, 104, 107 and 217).

B. Temporal Convolutional Networks

The temporal convolutional network is a kind of network presented by Bai et al. [8] for solving time series problems, which can effectively extract correlations between data and make predictions for subsequent data. The model currently performs well on tasks such as machine translation, audio synthesis, and natural language processing. The main structure of TCN is the dilated causal convolution, which acts as the convolution kernel on a more extensive region by skipping some of the inputs. Dilated convolution adjusts the receptive field's size by changing the expansion coefficient, allowing the network to flexibly adjust the amount of historical information received by the output. For a one-dimensional sequence of inputs x and filter f, the receptive field can be expanded by the filter coefficient k and the expansion coefficient d. The expansion convolution equation is as follows:

$$F(x) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d\cdot i} \tag{1}$$

In the formula, d is the expansion coefficient; s-d·i is the historical data in the input sequence; k is the filtering window size. Fig. 1 shows the dilated causal convolution structure, from which it can be seen that the perceptual field size of the y_T point in the output sequence is adjusted by k and d, and the output of this point is only affected by the historical data.

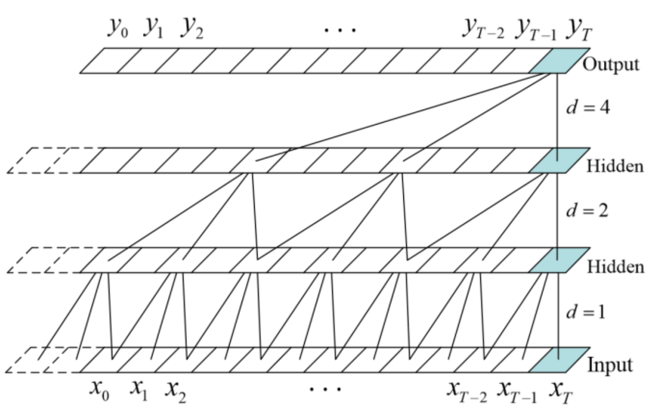

Fig. 1. Dilated causal convolution structure.

C. Residual Network

The Residual network model is a novel deep learning structure proposed by Kaiming He [9] and others, which achieved first place in the ImageNet classification competition in 2016. Previously, the training accuracy was usually improved by adding layers to the network. But experiments proved that the more layers of the network are designed, the more distinct the problems of network degradation and gradient disappearance, leading to a decrease in network accuracy. Compared with other networks, the residual neural network avoids these problems by adding "shortcut connections" to the network structure and achieves direct information transfer from the high level to the low level.

III. Model Design

A. ResNet-TCN model

The proposed classification network consists of two modules, ResNet and TCN. The ResNet uses convolutional kernel downsampling with a step size of 2 to reduce feature dimensionality without losing feature information. Moreover, ResNet uses residual edges to pass the information of each layer to avoid overfitting and gradient loss due to the cascading convolutional layers so that the underlying features of the input data can be well extracted. The TCN uses causal convolution to ensure that the model does not reverse the sequence order. The output at a given moment can only be convolved with the corresponding input in the previous layer and earlier moments, independent of future moments. Moreover, TCN uses dilation convolution, which increases the convolution kernel's size and the dilation coefficient value to increase the receptive field of the data and form a longer-time convolutional memory. As a result, TCN is able to detect the intricate, temporal context connections between the input and the target [10]. It is appropriate for tasks such as ECG signal classification. The method steps and the overall structure of the classification network can be seen as shown in Fig. 2.

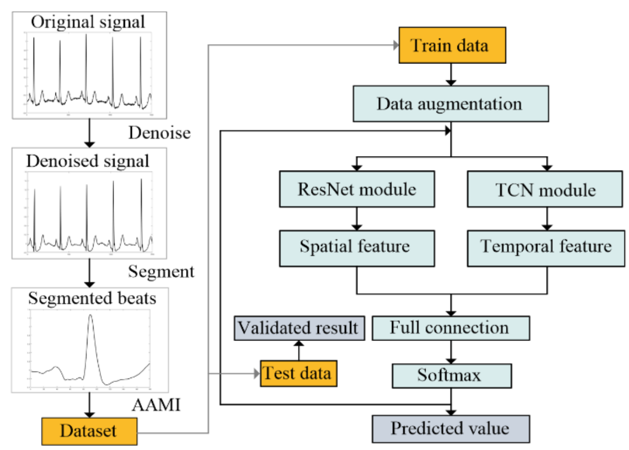

Fig. 2. Algorithmic framework diagram.

The ResNet-TCN heartbeat classification model proposed in this study consists of four steps: data acquisition, data preprocessing, classifier feature extraction, and fully connected heartbeat classification. The optimizer of the classification network uses Adam; the loss function uses cross-entropy; the learning rate is 0.0001; the number of iterations is 50. Detailed descriptions of the ResNet and TCN models are in Sections B and C. The most commonly used 1D convolution and ReLU activation function formulas in the model are defined as follows:

$$ReLU = \begin{cases} x, x > 0 \\ x, x \leq 0 \end{cases} \tag{2}$$

$$c^l = \sum_{i=0}^{m} \left(w_i^l x_i^l \right) + b^l \tag{3}$$

In the convolution formula, w^l and b^l are the weights and biases of the lth layer, and m is the convolution kernel size.

B. TCN Model Design

The fundamental component of the TCN, the TCN residual block, is composed of the Dilated Causal Conv layer, the WeightNorm layer, the ReLU layer, and the Dropout layer. This module is called the Temporal Block in this paper. It contains an optional convolutional (1×1) residual edge that allows the network to better transfer information across layers and mitigate gradient disappearance. The dilation of two convolutional layers in the same Temporal Block is the same, with a 2-fold relationship between different Temporal Blocks.

As shown in Fig. 3, the TCN model uses four Temporal Block modules. The convolutional kernel size is 7; the expansion coefficients of the convolutional kernels in each module is 1, 2, 4, and 8; the convolutional kernel number is 32, 64, 128 and 256, and the dropout is 0.3. Based on the properties of TCN, the last node of each channel is the feature that the network extracts. We add the channel attention mechanism, Squeeze-and-Excitation (SE) block [11], after the last layer of the TCN model. It makes the model learn to assign different feature weights to each channel according to the loss, so the network pays attention to the important features.

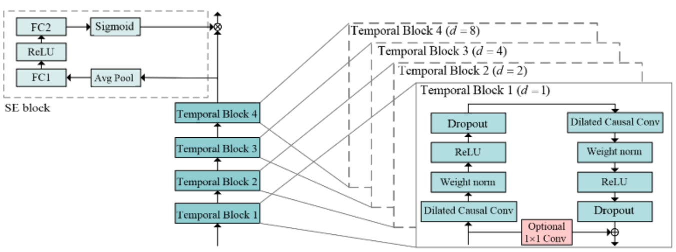

Fig. 3. TCN model structure.

C. ResNet Model Design

Fig. 4, shows the structure of the ResNet model. Since the heartbeat signal is simple one-dimensional data, we replace the convolutional kernels with one-dimensional structures. Then we remove the final global average pooling layer to avoid too much compression of the signal dimension, resulting in information loss. Meanwhile, in order to extract features at multiple scales and enhance the learning performance of the network, we replace the initial convolutional layer with a one-dimensional Inception [12] module. The Inception module contains three convolutional kernels (1×1, 1×3, and 1×5) and a maximum pooling layer (1×3). The processed ECG signal first passes through a one-dimensional Inception module to extract features at different scales. Then the features are extracted by the residual structure part consisting of a cascade of six residual modules. The residual structure part can be divided into 3 layers with the number of convolution kernels of 32, 48 and 64 for each layer.

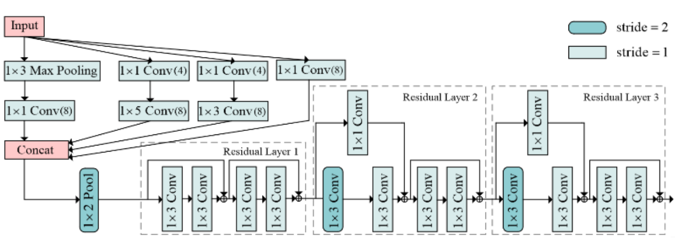

Fig. 4. ResNet model structure.

IV. Experiments and Analysis

A. Data Processing

ECG is susceptible to various noises in the process of acquisition. To minimize the effect of noise on the classification results, we use discrete wavelet transformation (DWT) to eliminate the ECG's baseline drift and high-frequency noise. We use the Matlab software tool to perform the db6 wavelet decomposition of ECG and zero out the coefficients at certain scales before reconstructing the signal. In the denoised ECG, we perform heartbeat interception by the position of the R-peaks annotated in the signal, and the interception is performed by 90 each of the left and right R-peaks, and the same method is used in other works [13][14]. In the AAMI standards, there are five categories of heartbeats: N, SVEB, VEB, F, and Q. The total number of each heartbeat category is 89841 N, 3026 SVEB, 7008 VEB, 802 F, and 15 Q. Because the number of Q is too small in this work, we remove the Q category and use the remaining four categories for classification. The classified data are randomly separated by 50% into training and test sets. Due to the training data's class imbalance (89.2% of heartbeats are N, 7% are V, 3% are S, and 0.79% are F), we extend the number of S, V, and F signals to be consistent with N by a proposed method that is suitable for ECG signal expansion. The approach expands the data by randomly intercepting the heartbeats and then interpolating the intercepted data back to the original size without loss of heartbeat characteristics.

B. Evaluation Methods

To assess the model's capability for classifying ECG signals, as in most methods, the following three performance indicators: accuracy (Acc), positive predictive value (P^+), and sensitivity (Se), are used in this paper, with the following equations.

$$Se = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

$$P^+ = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (6)$$

In the formula, TP is the true positive, FP is the false positive, FN is the false negative, and TN is the true negative.

C. Experiment Results

The experimental platform Spyder 4.0 is built on the Windows 10 operating system. We use Python as the programming language and the Numpy library and Pytorch library as data processing and classification tools. The hardware environment consists of a 3.6 GHz Intel Core i7 CPU, 8GB of memory, and a GeForce GTX 1650 GPU.

In this paper, we first validated the effectiveness of Inception module in the ResNet and SE block module in the TCN. The accuracy comparison is shown in Fig. 5. As seen from the comparison chart, the addition of the Inception and SE block modules effectively improves the overall accuracy of the ResNet model and the TCN model.

TABLE I. PERFORMANCE COMPARISON WITH RELATED WORKS

Model	Acc	F1-score	N		S		V		F	
			Se	P^+	*Se*	P^+	*Se*	P^+	*Se*	P^+
Acharya et al.[15]	0.940	0.924	0.916	0.852	0.890	0.948	0.941	0.951	0.952	0.947
Chen et al.[16]	0.978	0.911	0.992	0.985	0.875	0.953	0.947	0.952	0.739	0.861
Hua et al.[17]	0.974	0.931	0.965	0.971*	0.928	0.957*	0.928	0.964*	0.836	0.904*
Gai [18]	0.986	0.912	0.996	0.990	0.797	0.906	0.959	0.971	0.846	0.840
ResNet-TCN (This work)	**0.988**	**0.925**	**0.994**	**0.994**	**0.917**	**0.890**	**0.968**	**0.969**	**0.783**	**0.882**

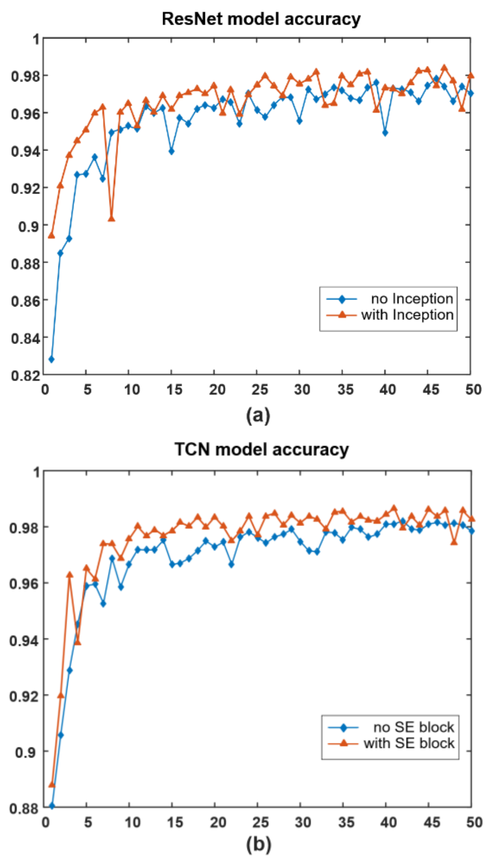

Fig. 5. Comparision of accuracy on (a) ResNet model and (b) TCN model.

TABLE II. ABLATION STUDY RESULTS

Classifier	Acc (%)	Se (%)	P^+ (%)	F1-score (%)
ResNet	98.12	89.62	88.21	88.91
TCN	98.36	91.76	87.75	89.71
ResNet-TCN (This work)	**98.84**	**91.69**	**93.34**	**92.51**

We verified the presented joint network model with papers that had recently been published to further confirm the efficacy of the proposed classification algorithm. Table 1 displays the method performance comparison. It can be seen that the classification accuracy of the proposed model is superior to that of the other works, the sensitivity of each category is also greater than most works, and the positive predictive value of each category is comparable to other works.

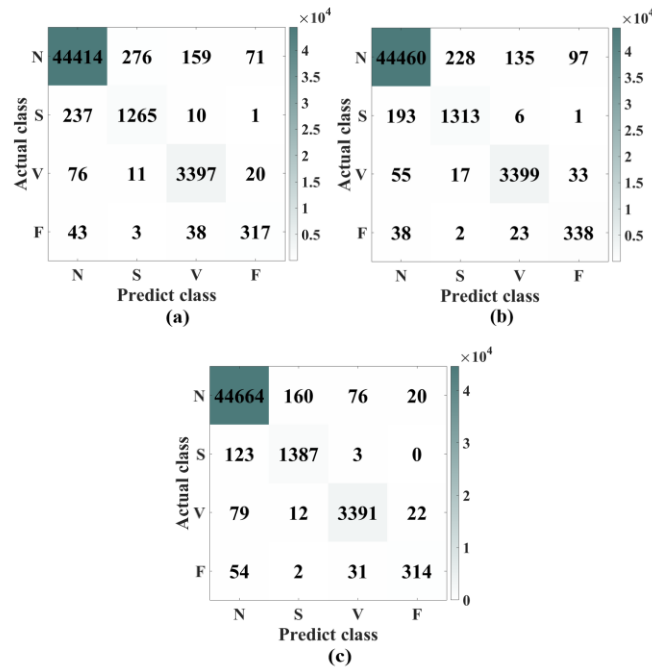

Fig. 6. Confusion matrix for (a) ResNet, (b) TCN, and (c) ResNet-TCN.

We next performed ablation experiments on the MIT-BIH for the single ResNet model, TCN model, and combined ResNet-TCN model to confirm the efficacy of the proposed joint ResNet-TCN model for ECG signal categorization. Table 2 and Fig. 6 indicates that the joint ResNet-TCN model performs superior to the single ResNet and TCN models on the MIT-BIH dataset.

V. CONCLUSION

In this work, we propose a joint model of ECG heartbeat four-classification based on ResNet and TCN. We first add Inception and SE block modules to ResNet and TCN, respectively, in the independent module design, which enhances the single model performance of ResNet and TCN. Then the joint model extracts the spatial information of data by the ResNet model and the temporal information of data by the TCN model and fuses the two to enhance model classification accuracy. The proposed model is validated with the MIT-BIH public dataset that uses the AAMI classification criteria, and the final experimental results demonstrate the validity and feasibility of the model. Meanwhile, the proposed data extension method provides a reference for ECG signals to handle the data imbalance problem.

REFERENCES

[1] The Writing Committee of the Report on Cardiovascular Health and Diseases in China. 2022. Report on Cardiovascular Health and Diseases in China 2021: an Updated Summary. Chin Circul J 37, 6 (June 2022), 553-578. https://doi.org/10.3969/j.issn.1000-3614.2022.06.001

[2] Hongqi Xia and Kaibo Deng. 2002. Clinical analysis and decision making of arrhythmia. Pecking Union Medical College Press. Beijing, CHN.

[3] U. Rajendra Acharya, Hamido Fujita, Oh Shu Lih, Yuki Hagiwara, Jen Hong Tan and Muhammad Adam. 2017. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. Inf. Sci. 405 (September 2017), 81-90. https://doi.org/10.1016/j.ins.2017.04.012

[4] Özal Yildirim. 2018. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Comput. Biol. Med. 96 (May 2018), 189-202. https://doi.org/10.1016/j.compbiomed.2018.03.016

[5] Cláudia Brito, Ana Machado, and António Sousa. 2019. Electrocardiogram Beat-Classification Based on a ResNet Network. Stud Health Technol Inform 264 (August 2019), 55-59. https://doi.org/10.3233/shti190182

[6] Elif Izci, Mehmet Akif Ozdemir, Murside Degirmenci, and Aydin Akan. 2019. Cardiac Arrhythmia Detection from 2D ECG Images by Using Deep Learning Technique. In 2019 Medical Technologies Congress (TIPTEKNO) . IEEE, Izmir, TR, 1-4. https://doi.org/10.1109/TIPTEKNO.2019.8895011

[7] George B Moody and Roger G Mark. 2001. The impact of the MIT-BIH arrhythmia database. IEEE eng. med. biol. mag. 20, 3 (May-June 2001), 45-50. https://doi.org/10.1109/51.932724

[8] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. ArXiv.org 2 (March 2018), arXiv:1803.01271. https://doi.org/10.48550/arXiv.1803.01271

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . IEEE, Las Vegas, NV, 770-778. https://doi.org/10.1109/cvpr.2016.90

[10] Raul Fernandez, Asaf Rendel, Bhuvana Ramabhadran, and Ron Hoory.2015. Using deep bidirectional recurrent neural networks for prosodic-target prediction in a unit-selection text-to-speech system. In Sixteenth Annual Conference of the International Speech Communication Association (INTERSPEECH) . ISCA, Dresden, GER, 1606-1610. https://doi.org/10.21437/interspeech.2015-372

[11] Jie Hu, Li Shen, and Gang Sun. 2018. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . IEEE, Salt Lake City, UT, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745

[12] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going Deeper With Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . IEEE, Boston, MA, 1-9. https://doi.org/10.1109/cvpr.2015.7298594

[13] V Mondéjar-Guerra, Jorge Novo, José Rouco, Manuel G Penedo and Marcos Ortega. 2019. Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. Biomed Signal Process Control 47 (January 2019), 41-48. https://doi.org/10.1016/j.bspc.2018.08.007

[14] Zhanglu Yan, Jun Zhou and Wengfai Wong. 2021. Energy efficient ECG classification with spiking neural network. Biomed Signal Process Control 63 (January 2021), 102170. https://doi.org/10.1016/j.bspc.2020.102170

[15] U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Arkadiusz Gertych and Ru San Tan. 2017. A deep convolutional neural network model to classify heartbeats. Comput. Biol. Med. 89 (October 2017), 389-396. https://doi.org/10.1016/j.compbiomed.2017.08.022

[16] Xiaohe Chen, Yan Wang and Lirong Wang. 2018. Arrhythmia recognition and classification using ECG morphology and segment feature analysis. IEEE/ACM Trans Comput Biol Bioinform 16, 1 (June 2018), 131–138. https://doi.org/10.1109/tcbb.2018.2846611

[17] Xuan Hua, Jungang Han, Chen Zhao, Haipeng Tang, Zhuo He, Qinghui Chen, Shaojie Tang, Jinshan Tang and Weihua Zhou. 2020. A novel method for ECG signal classification via one-dimensional convolutional neural network. Multimedia Systems (November 2020), 1-13. https://doi.org/10.1007/s00530-020-00713-1

[18] Gai N D. 2022. ECG beat classification using machine learning and pre-trained convolutional neural networks. ArXiv.org (Jun 2022), arXiv:2207.06408. https://doi.org/10.48550/arXiv.2207.06408