



# SEVGGNet-LSTM: A Fused Deep Learning Model for ECG Classification

Tongyue He<sup>1</sup> , Yiming Chen<sup>1</sup>, Bo Fang<sup>1</sup> , and Junxin Chen<sup>2</sup>  

<sup>1</sup> College of Medicine and Biological Information Engineering,  
Northeastern University, Shenyang 110167, China

<sup>2</sup> School of Software, Dalian University of Technology, Dalian 116620, China  
[junxinchen@ieee.org](mailto:junxinchen@ieee.org)

**Abstract.** With the dramatic progress of smart sensing and wearable device, continuous and real-time acquisition of electrocardiograph (ECG) tends to be realized in a convenient way. Data mining of ECG signals has therefore been extensively researched, among which ECG classification is a hot topic. This paper presents a fused deep learning algorithm for ECG classification. It takes advantages of the combined convolutional and recurrent neural network for ECG classification, and the weight allocation capability of attention mechanism. The input ECG signals are firstly segmented and normalized, and then fed into the combined VGG and LSTM network for feature extraction and classification. An attention mechanism (SE block) is embedded into the core network for increasing the weight of important features. Two databases from different sources and devices are employed for performance validation, and the results well demonstrate the effectiveness and robustness of the proposed algorithm for classifying ECG signals obtained from wearable ECG devices and professional medical equipment.

**Keywords:** ECG classification · Deep learning · Arrhythmia · Attention mechanism

## 1 Introduction

The electrocardiogram (ECG) analysis is an important non-invasive means for diagnosing and evaluating cardiac diseases [13]. However, the inherent complexity of arrhythmia often brings difficulties to medical workers in ECG classification, and may lead to mis-diagnosis. With the popularity of artificial intelligence (AI), developing computer-aided ECG classification is in a high demand.

Deep learning based solution for ECG classification has drawn world-wide concerns in recent years. It is able to automatically extract features, and hence get rid of the dependence of manual feature extraction in traditional machine learning methods. About features extraction, as reported in [7], attention mechanism is able to increase the weight of important features and further promote

---

Supported by the National Natural Science Foundation of China (No. 62171114).

the classification performance. This advantage has been widely exploited in the fields of neural machine translation and computer vision [4, 15]. It is therefore plausible to infer that adding attention mechanisms to the ECG classification model is likely to achieve performance improvement. [16] presented a CNN model with a non-local convolutional block attention module capable of distinguishing relationships between local and global segments. In addition, previous works [1, 10] have demonstrated the great potentials of the combined structure of convolutional neural network (CNN) and recurrent neural network (RNN) in the ECG classification problems. As a popular CNN, the visual geometry group network (VGGNet) [14] has more nonlinear transformation and enhanced ability to learn features. On the other hand, long short-term memory neural network (LSTM) is an improved RNN, and it is able to avoid the problems of gradient explosion or gradient disappearance.

This paper proposes the SEVGGNet-LSTM model for ECG classification. Our proposal takes advantages of the combined structure of convolutional and recurrent neural network for ECG classification, and also makes full use of the attention mechanism's weight allocation capability. First, ECG data records are split into 10s segments. After that, the amplitudes of the divided segments are normalized. The preprocessed ECG signals are then fed into the SEVGGNet-LSTM, which is a sequential combination of VGG and LSTM, with an attention mechanism (squeeze-and-excitation block, SE block) to increase the weight of important features. Finally, ECG classification is completed by the fully connected layers. Two databases from different sources and devices are employed for performance validation, and the experimental results well demonstrate the effectiveness and advantages of the proposed algorithm.

Our main contributions are as follows. 1) A fused deep convolutional neural network is constructed for ECG classification. 2) The combined convolutional and recurrent neural network and the weight allocation capability of the attention mechanism are exploited for performance promotion. 3) Two databases with ECG records collected by different devices are employed for performance evaluation.

## 2 The Proposed Method

### 2.1 Architecture

The proposed algorithm mainly consists of the preprocessing module, feature extraction module, and classification module, as illustrated in Fig. 1. The preprocessing module includes signal segmentation and normalization. The ECG signals are split into 10s segments, after that a normalization technique is used to normalize their amplitudes. In the feature extraction module, SEVGGNet and LSTM constitute the deep neural network, and the preprocessed ECG signals are fed into the deep neural network for feature extraction. Finally, a three-fully connected layer and a softmax layer constitute the classification module to realize the multi-classification problem.

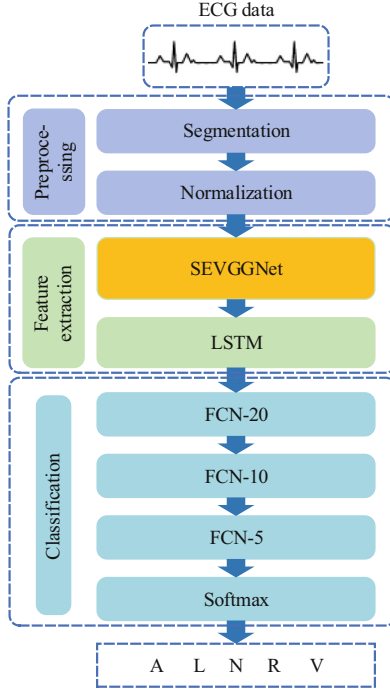


Fig. 1. Block diagram of the proposed model.

### 2.2 Data Preprocessing

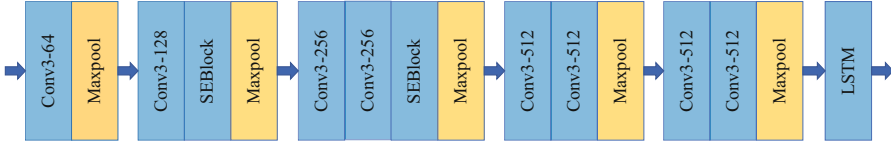
The data preprocessing module primarily concentrates on two key tasks: signal segmentation and normalization. To alleviate computational overhead, ECG signals are segmented into multiple 10-s segments. Since ECG segment amplitudes can significantly vary owing to individual differences and lead positions, normalization becomes crucial. This normalization process involves adjusting the amplitudes, ensuring uniformity, and improving data consistency for further analysis and model training. The normalization is conducted by

$$Normalized(X) = \frac{X - \bar{X}}{S}, \tag{1}$$

where  $X$  is value of each record,  $\bar{X}$  and  $S$  refer to the average and standard deviation of all the records, respectively.

### 2.3 Feature Extraction

The core network of the proposed algorithm is illustrated in Fig. 2. It includes eight convolutional layers, one LSTM layer, five maximum pooling layers, and two SE blocks. The convolutional layers are split into five parts by maximum



**Fig. 2.** Core network of the proposed model.

pooling layers. The numbers of the convolutional layers in each part are 1, 1, 2, 2, and 2, and the sizes of convolutional kernels are 64, 128, 256, 512, and 512, respectively. Each part of convolutional layer is followed by a maximum pooling layer to reduce the data length and computational burden of the model. A LSTM layer is connected after the convolutional layers and maximum pooling layers of VGGNet to avoid the gradient disappearance and gradient explosion. In addition, two SE blocks are added to the fusion model of VGGNet and LSTM, which strengthen the features of R peak. After that, the output of LSTM is given to the fully connected layer for ECG classification. During the model training process, we selected the following optimal parameters: optimizer = “Adam”, learning rate = 0.001, epoch = 50, and batch size = 32 for improved performance.

The SENet, a kind of attention mechanism, is presented in the form of the SE block. It contains the squeeze and excitation modules. Squeeze operation is carried out after the traditional convolution module, that is, all spatial features in a channel are encoded into a global feature. It is defined as

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j), z_c \in R^c, \tag{2}$$

where  $Z_c$  refers to the  $c$ -th element of the squeezed channels,  $u_c$  represents the  $c$ -th channel of the input,  $F_{sq}$  is the squeeze function, and  $H$  and  $W$  are the height and width of the input, respectively. After that, two fully connected layers are employed for better generalization, an activation function is used to obtain the channel-wise dependencies. The process is described by

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)), \tag{3}$$

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c u_c, \tag{4}$$

where  $F_{ex}$  denotes the excitation function,  $W_1$  and  $W_2$  represent the widths of the inputs in dimensionality reduction and increasing layers,  $\delta$  denotes the ReLU activation function,  $\sigma$  is the sigmoid function, and  $F_{scale}$  represents channel-wise multiplication.

### 2.4 Classification Module

At the end of our model, a 3-layers fully connected layer of the VGGNet is selected, and the softmax function is used to realize multi-classification problem.

The fully connected layers reassembles the local features into a complete graph through the weight matrix. The representative features are integrated into a value, which has the advantage of reducing the influence of feature locations on classification results, and improving the robustness of the whole network. In addition, multiple fully connected layers are connected, so that the ability of nonlinear expression is improved.

### 3 Experiment Configuration

#### 3.1 Database and Performance Metrics

Two different databases are employed for performance evaluation, the MIT-BIH arrhythmia database and the 2017 PhysioNet/CinC Challenge database (2017 PCCD). The MIT-BIH arrhythmia database includes 48 recordings, each of which is sampled at 360 Hz and contains a series of two-leads ECG data. Based on the proportion of various arrhythmias in the clinic, normal rhythm (N), left bundle branch block (R), right bundle branch block (L), ventricular precontraction (V) and atrial premature contractions (A) are selected for classification. On the other hand, there are 8528 single-lead ECG records in 2017 PCCD. All of them are collected by wearable devices at a sampling frequency of 300 Hz. The types of normal (N) and atrial fibrillation (AF) samples are selected to validate the performance of our model. Because of the data imbalance problem, samples in the training set are balanced by oversampling, as detailed in Sect. 3.2.

In this paper, accuracy (*Acc*), sensitivity (*Sen*), precision (*Pre*), and F1 score are employed for performance evaluation. The *Acc* is the proportion of correctly classified samples, *Sen* denotes the recognition ability of positive examples, *Pre* represents the proportion of correct predictions in the samples with positive predictions, and F1 score is the weighted average of *Sen* and *Pre*, respectively.

All of the models are implemented on the framework Tensorflow-gpu 2.6.2, using the Windows-11 operation system. The algorithm is deployed on a workstation with Intel(R) Core(TM) i7-12700H at 2.7 GHz, and an RTX 3060 GPU with a 14 GB memory.

#### 3.2 Oversampling

Oversampling is adopted to solve the data imbalance problem, by increasing the number of certain arrhythmia samples. In this paper, the random oversampling method is used. It works as follows. 1) Taking the class with the largest number of samples as the benchmark, calculate the multiple of the benchmark class and the minority classes. 2) If the multiple is greater than 2, the minority classes are copied by corresponding multiples. 3) If the multiple is less than 2, a certain proportion (multiple minus one) of samples from the minority classes are randomly selected for duplication. Taking MIT-BIH arrhythmia database as an example, and samples counts before and after oversampling is listed in Table 1.

**Table 1.** Oversampling on MIT-BIH arrhythmia database

Type	Before oversampling	After oversampling
N	6735	6735
V	3005	6010
L	1202	7212
R	1179	7074
A	771	6939
Total	12892	33970

### 3.3 Cross Validation

Under the condition of limited size of dataset, 10-fold cross validation is able to achieve multiple random partitioning of the training set and test set. The original dataset is divided into 10 equal sized parts, of which nine parts are considered as training dataset and the other one is used for testing. 10-fold cross validation avoids overlap between the training and testing datasets. In each iteration, the balanced training data set is used to get the optimized parameters of the model, and then the corresponding test set is employed for performance evaluation. After 10 iterations, the results from each iteration are combined to yield the average performance of the model.

## 4 Results and Discussion

### 4.1 Overall Performance

Table 2 lists the results when using the proposed algorithm to classify various heart rhythms in the MIT-BIH arrhythmia database. The overall *Acc*, *Sen*, *Pre* and F1 values are 0.996, 0.984, 0.988, and 0.986, respectively. By considering both *Sen* and *Pre*, the F1 score better demonstrate the overall performance of the proposed algorithm. For A, L, N, R, and V types of rhythms, the achieved F1 scores of our method are 0.955, 0.995, 0.993, 1.000, and 0.987, respectively. For R rhythm, all of the performance metrics are optimal, that is, 1, which indicates the good performance in identifying R rhythms. In addition, all of the *Acc* values are higher than 0.992, demonstrating that the classification performance of our algorithm is very satisfactory.

In order to verify the universality of the proposed algorithm, it is further applied to 2017 PCCD, and the performance records are listed in Table 3. As can be observed, the *Acc*, *Sen*, *Pre* and overall F1 records are 0.962, 0.931, 0.914, and 0.922, respectively. As concluded above, the proposed algorithm shows satisfactory performance on both databases.

**Table 2.** Performance on MIT-BIH arrhythmia database

Types	<i>Acc</i>	<i>Sen</i>	<i>Pre</i>	F1
A	0.994	0.955	0.955	0.955
L	0.999	0.991	1.000	0.995
N	0.992	0.997	0.989	0.993
R	1.000	1.000	1.000	1.000
V	0.994	0.978	0.997	0.987
Overall	0.996	0.984	0.988	0.986

**Table 3.** Performance on 2017 PCCD

Types	<i>Acc</i>	<i>Sen</i>	<i>Pre</i>	F1
N	0.962	0.974	0.982	0.978
A	0.962	0.888	0.845	0.866
Overall	0.962	0.931	0.914	0.922

## 4.2 Comparison Results

In order to further validate the performance of our proposal, some state-of-the-art algorithms are employed for comparison. For fair comparison, only the algorithms developed for the same classification problems and tested on the same database are introduced. Regarding the MIT-BIH arrhythmia database, the selected models for comparison are as follows. Liu *et al.* [8] proposed a model based on LSTM to obtain time-series features of ECG. In [10], the CNN layer is used to extract feature maps, and the LSTM layer captures the temporal dynamics. The model [5] is composed of a eight-layers CNN, a eight-layers LSTM and a fully connected layer. Detailed performance records of the compared models are listed in Table 4. The F1 score, *Acc*, *Sen* and *Pre* of the proposed model are higher than those of the compared models. The advantages of our proposal are therefore validated.

In addition, the two-class classification performance of wearable ECG is also compared on the 2017 PCCD. The results are listed in Table 5. Compared with the peer algorithms that have reported their F1 performance, the proposed model has the highest F1 score of 0.922. The comparison results further verify the good performance of the proposed model.

**Table 4.** Comparison on MIT-BIH arrhythmia database

Methods	<i>Acc</i>	<i>Sen</i>	<i>Pre</i>	F1
LSTM [8]	0.986	0.980	0.976	0.978
CNN+LSTM [10]	0.981	0.975	–	–
CNN+LSTM+HOS [5]	0.989	0.965	0.969	0.967
U-Net [11]	0.973	–	–	–
RBF-BA [2]	0.952	0.956	0.906	0.930
LDA [17]	0.962	0.925	0.947	0.936
KNN [6]	–	0.809	0.769	0.788
<b>Proposed method</b>	<b>0.996</b>	<b>0.984</b>	<b>0.988</b>	<b>0.986</b>

**Table 5.** Performance comparison on 2017 PCCD

Methods	<i>Acc</i>	<i>Sen</i>	<i>Pre</i>	F1
CNN [12]	0.899	0.671	0.590	0.628
VGGNet [14]	0.976	0.909	0.903	0.906
MS-CNN [3]	0.977	0.943	0.886	0.914
BT [9]	0.966	0.832	–	–
<b>Proposed</b>	<b>0.962</b>	<b>0.931</b>	<b>0.914</b>	<b>0.922</b>

**Table 6.** Results of ablation experiments on the MIT-BIH arrhythmia database

Methods	<i>Acc</i>	<i>Sen</i>	<i>Pre</i>	F1
VGG11	0.972	0.901	0.898	0.898
VGG13	0.979	0.873	0.943	0.895
VGG16	0.963	0.810	0.858	0.824
VGG11-LSTM	0.980	0.911	0.911	0.911
VGG13-LSTM	0.980	0.894	0.919	0.906
VGG16-LSTM	0.977	0.867	0.911	0.888
SEVGG11-LSTM	0.980	0.933	0.909	0.919
<b>Proposed (SEVGG11-LSTM-O)</b>	<b>0.996</b>	<b>0.984</b>	<b>0.988</b>	<b>0.986</b>

**Table 7.** Results of ablation experiments on the 2017 PCCD

Methods	<i>Acc</i>	<i>Sen</i>	<i>Pre</i>	F1
VGG11	0.914	0.647	0.800	0.689
SEVGG11	0.972	0.901	0.898	0.898
VGG11-LSTM	0.952	0.841	0.950	0.885
VGG13-LSTM	0.945	0.868	0.893	0.880
VGG16-LSTM	0.941	0.798	0.950	0.853
SEVGG11-LSTM	0.952	0.877	0.913	0.894
<b>Proposed (SEVGG11-LSTM-O)</b>	<b>0.962</b>	<b>0.931</b>	<b>0.914</b>	<b>0.922</b>

### 4.3 Discussions

In order to determine the optimized network, convolutional networks of various depths are created. In specific, networks with 11, 13 and 16 convolutional layers are compared on the MIT-BIH arrhythmia database. As listed in Table 6, with the increasing counts of convolutional layers, the performance metrics decrease. Compared with the basic models, VGG11-LSTM is finalized for further improvement. When enhanced by the attention mechanism, the overall F1 score increases, indicating that the SEVGG11-LSTM has better classification performance. After oversampling is implemented, the classification performance is further improved, and the strong competitiveness on the MIT-BIH arrhythmia database is advantageous to that of the state-of-the-art algorithms.

Similar ablation experiments are also performed on the 2017 PCCD, and the performance records are comparatively listed in Table 7. Similarly, the F1 score decreases with the increasing counts of the convolutional layers, and is increased after introducing the attention mechanism. When oversampling is further used, the F1 score reaches the optimum value. Tables 6 and 7 well validate the great potentials of attention mechanism and oversampling for ECG classification.

Compared with the validation on a single database, two databases are employed in our paper in a “dual-centers” fashion. More participants are involved to avoid the limitation of a single database, and the conclusion is therefore more reliable. In addition, researchers can draw on the wisdom of the masses in the two databases to improve clinical trials.

## 5 Conclusion

This paper demonstrates a novel deep learning algorithm for ECG classification. It makes use of the combined convolutional and recurrent neural network for classifying ECG as well as the attention mechanism to assign weights. The input ECG signals are sequentially segmented and normalized. After that, the pre-processed signals are fed into the combined VGG and LSTM network for feature extraction and classification. The core network contains an attention mechanism that increases the weight of significant features. Two databases from different

sources and devices are employed for performance validation, and the results well demonstrate the effectiveness and advantages of the algorithm.

## References

1. Chen, C.Y.: Automated ECG classification based on 1D deep learning network. *Methods* **202**, 127–135 (2022)
2. Ebrahimzadeh, A., Shakiba, B., Khazaei, A.: Detection of electrocardiogram signals using an efficient method. *Appl. Soft Comput.* **22**, 108–117 (2014)
3. Fan, X., Yao, Q., Cai, Y., Miao, F., Sun, F., Li, Y.: Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings. *IEEE J. Biomed. Health Inf.* **22**(6), 1744–1753 (2018)
4. Guo, W., Zhang, Y., Yang, J., Yuan, X.: Re-attention for visual question answering. *IEEE Trans. Image Process.* **30**, 6730–6743 (2021)
5. Huang, Y., Li, H., Yu, X.: A multiview feature fusion model for heartbeat classification. *Physiol. Meas.* **42**(6), 065003 (2021)
6. Jekova, I., Bortolan, G., Christov, I.: Assessment and comparison of different methods for heartbeat classification. *Med. Eng. Phys.* **30**(2), 248–257 (2008)
7. Lai, Q., Khan, S., Nie, Y., Sun, H., Shen, J., Shao, L.: Understanding more about human and machine attention in deep neural networks. *IEEE Trans. Multimedia* **23**, 2086–2099 (2020)
8. Liu, P., Sun, X., Han, Y., He, Z., Zhang, W., Wu, C.: Arrhythmia classification of LSTM autoencoder based on time series anomaly detection. *Biomed. Signal Process. Control* **71**, 103228 (2022)
9. Mei, Z., Gu, X., Chen, H., Chen, W.: Automatic atrial fibrillation detection based on heart rate variability and spectral features. *IEEE Access* **6**, 53566–53575 (2018)
10. Oh, S.L., Ng, E.Y., San Tan, R., Acharya, U.R.: Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. *Comput. Biol. Med.* **102**, 278–287 (2018)
11. Oh, S.L., Ng, E.Y., San Tan, R., Acharya, U.R.: Automated beat-wise arrhythmia diagnosis using modified U-Net on extended electrocardiographic recordings with heterogeneous arrhythmia types. *Comput. Biol. Med.* **105**, 92–101 (2019)
12. Pourbabaee, B., Roshtkhari, M.J., Khorasani, K.: Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. *IEEE Trans. Syst. Man Cybern. Syst.* **48**(12), 2095–2104 (2018)
13. Mukhopadhyay, S.K., Mitra, S., Mitra, M.: An ECG signal compression technique using ASCII character encoding - sciencedirect. *Measurement* **45**(6), 1651–1660 (2012)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)* (2014)
15. Vaswani, A., et al.: Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30** (2017)
16. Wang, J., et al.: Automated ECG classification using a non-local convolutional block attention module. *Comput. Methods Programs Biomed.* **203**, 106006 (2021)
17. Yeh, Y.C., Wang, W.J., Chiou, C.W.: Cardiac arrhythmia diagnosis method using linear discriminant analysis on ECG signals. *Measurement* **42**(5), 778–789 (2009)