Future Generation Computer Systems 102 (2020) 670–679

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

# A deep learning approach for atrial fibrillation signals classification based on convolutional and modified Elman neural network

Jibin Wang *

*School of Mathematics, Tianjin University, Tianjin 300354, China*
*Visual Pattern Analysis Research Lab, Tianjin University, Tianjin 300072, China*

ARTICLE INFO

*Article history:*
Received 11 July 2019
Received in revised form 22 August 2019
Accepted 6 September 2019
Available online 18 September 2019

*MSC:*
68Uxx
62P10

*Keywords:*
Electrocardiogram
Atrial fibrillation
Deep learning
Convolutional neural network
Elman neural network

ABSTRACT

Atrial fibrillation (AF) is one of the main causes of life-threatening heart disease. Its detection and diagnosis have been highly concerned by physicians in recent years. However, the conventional AF detection through visual inspection of electrocardiogram (ECG) data by expert cardiologists is low-efficient and time-consuming. In this work, we develop a novel approach for the automated AF detection based on an 11-layers neural network. The network structure is primarily stacked by convolutional neural network (CNN) and the modified Elman neural network (MENN), while automatically performing end-to-end signals classification. To verify the superiority of the proposed model, two relevant deep network models were specially constructed for comparison. Moreover, ten-fold cross-validation was employed to evaluate the classification performance of the model on the MIT-BIH AF database. Compared with the two relevant models and several state-of-the-art methods, the model yielded excellent classification performance with the accuracy, sensitivity and specificity of 97.4%, 97.9%, and 97.1%, respectively. To the best of our knowledge, this is also the first time to stack CNN and MENN for ECG signals analysis and AF detection in particular. Its outstanding performance demonstrates that the proposed model has great potential as an efficient and robust identification system to assist physicians and reduce mortality.



## 1. Introduction

Atrial fibrillation (AF) is one of the most common arrhythmias in recent years. The disease is closely associated with the occurrence of thrombus, cerebral apoplexy, even death, which creates an imminent threat to human health [1,2]. According to the World Health Organization (WHO), it was also reported that over 6 million people have been threatened in the United States and about 90 million people have AF worldwide [3,4]. At present, the morbidity has reached 5% in people aged 50 to 59, even about 10% in people aged 80 to 89. This trend has increased significantly as age [3]. Notably, the symptoms of AF are usually not so obvious in patients that they are hard to perceive, which also undoubtedly aggravates the deterioration of the condition. Therefore, the indispensable combat for this prevalence and damage to maintain public health is of fundamental significance.

However, the conventional AF detection is often diagnosed through visual inspection of electrocardiogram (ECG) data by cardiologists. It is undeniable that the diagnostic results of the disease are often influenced by the doctor's subjective assessment. Also, the large amount of ECG data are increasing the workload of physicians [5]. These factors make AF detection time-consuming and low-efficient during the manual diagnosis. Hence, it is vital to develop the automated AF detection strategy for accurate and efficient disease diagnosis.

The ECG is able to serve as a widely used and non-invasive diagnostic tool in medical research, which mainly illustrates the electrical activity of heart recorded information by installing electrodes on the skin for a period of time [1–5]. Specifically, this method is well available for providing the vital clinical characteristics about heart condition of patients and interpreting cardiac pathological mechanisms [1,6]. A variety of classical methods based on the ECG features such as RR intervals (adjacent heartbeat intervals) have been implemented to automatically detect AF segments from ECG data over the past decade.

More specifically, most of them are the traditional machine learning algorithms for ECG signals analysis [6–8]. These algorithm structures are mainly consist of three parts: feature extraction, feature selection and classification. The first two parts are to build an analysis system that can extract essential features from the raw data. Classification is ultimately accomplished by combining the extracted features with the classifier. For instance, Thomas et al. [9] employed the coefficients of discrete wavelet transform (DWT) and dual tree complex wavelet transform (DTCWT) with four morphological features for the analysis

* Correspondence to: Visual Pattern Analysis Research Lab, Tianjin University, Tianjin 300072, China.
*E-mail address:* wangjibin0429@163.com.

https://doi.org/10.1016/j.future.2019.09.012
0167-739X/

of AF episodes. A shallow artificial neural network (ANN) was implemented for the classification. Henzel et al. [10] used four statistical features based on RR intervals and the generalized linear classifier to automatically classify AF and normal sinus rhythm (NSR) segments. Acharya et al. [11] combined the coefficients of discrete cosine transform (DCT), empirical mode decomposition (EMD) and DWT with the K-nearest neighbor (KNN) classifier for ECG signals classification. Kumar et al. [12] developed a novel decision-making system based on the flexible analytic wavelet transform (FAWT). The log-energy entropy (LEE) from the subbands of ECG signals was incorporated with the random forest (RF) classifier for the analysis of AF segments. Boon et al. [13] proposed an optimization algorithm using heart rate variability (HRV) series and non-dominated sorting genetic algorithm to distinguish AF signals. Hochstadt et al. [14] developed a portable system based on RR intervals and photoplethysmograph for AF episodes classification. Although these methods were demonstrated to be helpful for ECG signals analysis, they have many obvious shortcomings. On the one hand, feature extraction and feature selection need the manual design. So far, however, there is no foolproof recipe to demonstrate which artificial design strategy is suitable for the given problem; on the other hand, these hand-designed strategies fail to guarantee the reliable robustness and easily cause the over-fitting problems, especially validated against different types of data sets. Therefore, it is unreliable to ensure the generalization ability of method system by traditional hand-crafted feature extraction algorithms.

To overcome the above problems, deep learning approach has been developed into many fields such as face identification, behavior classification, image processing, and some remarkable results have been presented [6,8,15]. Unlike traditional learning algorithms, deep learning methods have integrated feature extraction and feature selection into the model, thus the hand-crafted features are not needed. Also, these methods are able to mine well different types of data sources and have well generalization ability, allowing the computer to automatically learn and extract related features for any given issues. A great superiority of deep learning is that with the accumulation of data and results or corresponding accurate explanations, the model performance of the system can be constantly heightened. Their typical frameworks mainly include convolutional neural network (CNN), stacked autoencoder (SAE), deep belief network (DBN) and recurrent neural network (RNN), and so forth [8,16]. Presently, a great quantity of literatures have reported that deep learning models are more effective and accurate than traditional machine learning algorithms, especially for ECG signals analysis and classification [17]. Rahhal et al. [18] performed a novel system based on the denoising SAE with sparsity constraint for the active classification of ECG signals. Acharya et al. [19] developed an 11-layers CNN structure to distinguish four types of heartbeats. Subsequently, they also constructed an 11-layers CNN model for the classification of shockable and non-shockable ventricular arrhythmias [15]. Xiong et al. [20] proposed a 16-layers CNN structure with skip connection for ECG signals classification. Singh et al. [21] implemented a 3-layers long short-term memory (LSTM) network with 128, 256 and 100 number of neurons in each layer for the analysis of ECG data. With the max-pooling layer, Faust et al. [22] developed a deep learning architecture based on RNN with LSTM to classify AF segments. Besides, Andersen et al. [17] extracted the high-level features from RR intervals and combined CNN with RNN for the analysis of ECG signals.

Since RNN designed with the feedback loop is dynamic and thus has short-term memory ability compared to the traditional static network such as the multi-layer perceptron (MLP). This network structure has been paid considerable attention by scholars for ECG signals analysis. As one of the typical RNNs, Elman neural network (ENN) is trained through a supervised way and optimized based on the back propagation (BP) algorithm [23,24]. In an ENN system, the context unit is especially installed as a delay operator such that the output information of the previous moment from the hidden layer can be stored by them [25]. Due to the existence of this structure, ENN has been demonstrated to have well dynamic information processing ability and fast convergence rate, which has been successfully performed in prediction and classification tasks [26–28].

However, there are few research on the application of ENN architecture for ECG signals classification. Also, due to the variations in ECG signals among different individuals over time, the analysis of ECG data should be considered dynamically. Thus, it is promising to develop the identification of ECG signals by means of the ENN structure. More importantly, the model performance of the standard ENN can be further enhanced by considering various improvement strategies. It can be expected that the combination of CNN and the modified ENN (CNN–MENN) will be effective to distinguish AF and NSR segments. Therefore, in this work, a novel deep network structure based on an 11-layers CNN–MENN model was designed for AF detection. Ten-fold cross-validation was also employed to evaluate the model performance on the MIT-BIH AF database. Especially compared with the two relevant CNN-MLP and CNN–ENN models, more superior classification performance and faster convergence rate of the proposed model were demonstrated in ECG signals analysis for the automatic diagnosis of AF.

The structure of the paper is summarized as follows. Section 2 describes the ECG data set, pre-processing, CNN, and the CNN–MENN model. The experimental results and related discussions are illustrated in Section 3 and Section 4, respectively. Section 5 summarizes the paper and gives the future work.

## 2. Methods and materials

### 2.1. Data set

To evaluate the classification performance of the proposed model, ECG signals from the MIT-BIH database were analyzed. The database is publicly available for medical research and analysis, which mainly includes NSR data set, arrhythmia data set, AF data set, long-term ECG data set and so on [29]. In this study, we used the AF data set to conduct the numerical experiments. The data set mainly includes 23 records from the 23 different patients and each record is about 10 h. The sampling rate is 250 Hz and the resolution is 12 bit with an interval of $\pm 10$ mV. All the records were labeled by expert physicians. The ECG episodes including NSR and AF derived from all 23 AF records were employed. And an example of the AF signal and the normal ECG signal from the record #04015 are illustrated in Fig. 1.

### 2.2. Pre-processing

Before the ECG signals are fed into the proposed model for training and testing, it is necessary to remove the unexpected artifacts in ECG data. The ECG signals are often polluted by baseline drift, power-line interference, muscle noise [30], so the bandpass filtering with the frequency range of 0.5–30 Hz and the 50 Hz notch filtering were assembled to eliminate these artifacts for the further ECG data analysis. Subsequently, the filtered ECG episodes were divided every 4 s into segments, and then these raw ECG segments were Z-score normalized to address the problem of amplitude scaling and offset effect before fed into the proposed model.

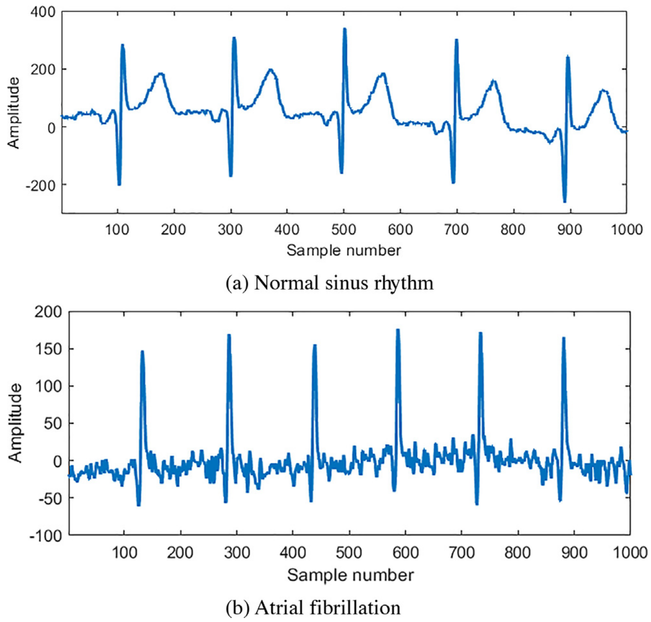


**Fig. 1.** An example of ECG segments from the record #04015.

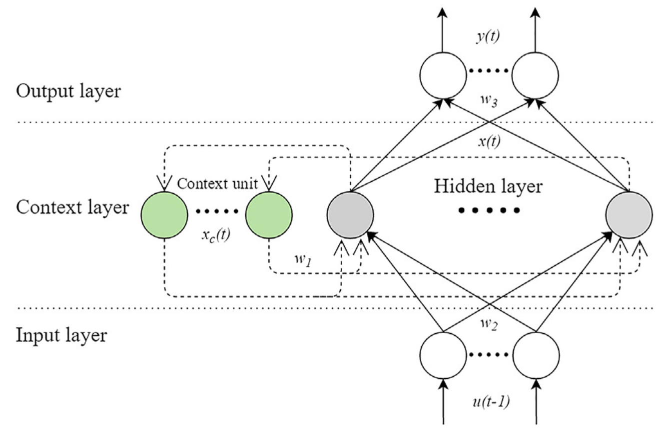


**Fig. 2.** The network structure of ENN.

### 2.3. Convolutional neural network

CNN was initially proposed by Fukushima in 1980, and then greatly boosted by Lecun et al. [31,32]. The network is actually a kind of typical MLP. Unlike the traditional neural network that cannot make full use of any spatial or temporal information in the data, CNN opens up a new approach based on the local connectivity and weight sharing pattern to incorporate these information while greatly reducing the complexity of network [32]. Its structure is mainly composed of input layer, convolutional layer, pooling layer, fully connected layer. The depicts of different layers in the model are summarized as follows:

- **Input layer:** The input data are initially input through it to maintain the values.
- **Convolutional layer:** This layer is the essential core structure layer of CNN. The convolution kernels (filters) that slide across the input data in this layer are convolved with them, and the stride regulates how much the kernels convolve with the input data. The corresponding output is also called the feature map. This operation not only reduces the dimension of the input data, but also automatically extracts useful features from the original data.
- **Pooling layer:** This layer, also known as sub-sampling layer, is another important structural layer of CNN. It mainly includes max-pooling and average pooling. The aim of input data passing through this layer is to compress the number of parameters and reduce the computational cost of network, while retaining the essential features. Also, the structure alleviates the over-fitting problem effectively.
- **Fully connected layer:** This layer is equivalent to the traditional MLP. The final features originated from the processing of multiple convolutional and pooling layers can be well classified in this layer.

Dependent on the structure characteristics of CNN, it eliminates the requirement of feature extraction and feature selection process [33,34]. Furthermore, CNN can greatly alleviate the workload of training and reduce the computational complexity, while providing the optimal feature extraction strategy for classification. It is apparent that if the suitable structures or strategies are combined with CNN, there is a great possibility of yielding the promising results. Therefore, we employed CNN in the proposed model for these reasons.

### 2.4. The modified elman neural network

Originally developed by Elman in 1990 on the basis of Jordan network [35], the network structure of ENN is divided into four parts: input layer, hidden layer, context unit and output layer as illustrated in Fig. 2. The input layer can be used to transmit the raw data; the imported weighted data are mapped linearly or non-linearly through the transfer function of the hidden layer; finally the processed data are performed by the linear weighted method in the output layer. However, unlike BP feedforward network, ENN adds extra the context unit to store the output information of the previous moment of the hidden layer and feedback it to the next moment of the hidden layer [36]. Hence, this network not only enhances the sensitivity to historical data, but also has dynamic information memory ability than the traditional static network [37]. The mathematical model of standard ENN structure can be summarized as follows:

$$y(t) = g(w_3 x(t)) \tag{1}$$

$$x(t) = f(w_1 x_c(t) + w_2(u(t-1))) \tag{2}$$

$$x_c(t) = x(t-1) \tag{3}$$

where $x(t), y(t), x_c(t)$ represent the $t$th output of the hidden layer, the output layer and the context unit, respectively. $u$ is the input vector of the input layer. $w_1$, $w_2$, $w_3$ denote the connection weight of the context unit to the hidden layer, the input layer to the hidden layer, and the hidden layer to the output layer, respectively. $f(\cdot)$ and $g(\cdot)$ represent the transfer function of the hidden layer and the output layer, respectively.

However, in a standard ENN, it is easy to find that the output of the context unit is only the information feedback of the hidden layer at the previous moment, which may not fully take advantage of the spatio-temporal information in the network. So in order to further improve the model performance of standard ENN, in this work, the MENN structure was constructed. It not only accounts for the full information feedback from the hidden layer, but also the information feedback from the context unit itself. The more relevant information are fed back into the network such that the sample data can be analyzed more comprehensively, which is more conducive for the final classification. The mathematical model of the MENN structure can be modified in the context unit as follows:

$$x_c(t) = \alpha \cdot x_c(t-1) + x(t-1) \tag{4}$$

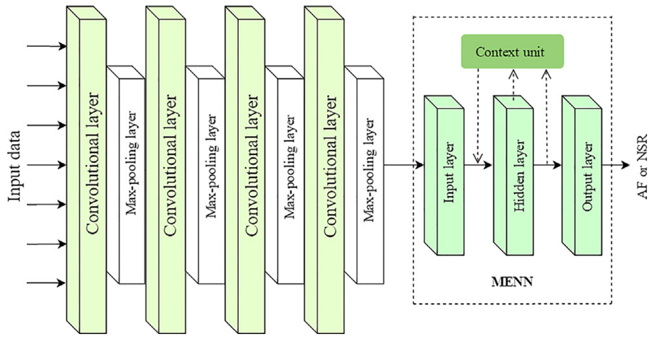


**Fig. 3.** Architecture of the CNN–MENN model for AF detection.

**Table 1**
The details of the CNN–MENN model.

| Layers | Type | Number of neurons (output layer) | Kernel size for each output feature map | Stride |
|---|---|---|---|---|
| 0–1 | Convolution | 978 × 3 | 23 | 1 |
| 1–2 | Max-pooling | 489 × 3 | 2 | 2 |
| 2–3 | Convolution | 476 × 5 | 14 | 1 |
| 3–4 | Max-pooling | 238 × 5 | 2 | 2 |
| 4–5 | Convolution | 232 × 10 | 7 | 1 |
| 5–6 | Max-pooling | 116 × 10 | 2 | 2 |
| 6–7 | Convolution | 112 × 10 | 5 | 1 |
| 7–8 | Max-pooling | 56 × 10 | 2 | 2 |
| 9–11 | MENN | 30, 14, 2 | – | – |

**Table 2**
The details of the CNN-MLP model.

| Layers | Type | Number of neurons (output layer) | Kernel size for each output feature map | Stride |
|---|---|---|---|---|
| 0–1 | Convolution | 978 × 3 | 23 | 1 |
| 1–2 | Max-pooling | 489 × 3 | 2 | 2 |
| 2–3 | Convolution | 476 × 5 | 14 | 1 |
| 3–4 | Max-pooling | 238 × 5 | 2 | 2 |
| 4–5 | Convolution | 232 × 10 | 7 | 1 |
| 5–6 | Max-pooling | 116 × 10 | 2 | 2 |
| 6–7 | Convolution | 112 × 10 | 5 | 1 |
| 7–8 | Max-pooling | 56 × 10 | 2 | 2 |
| 8–9 | Fully connected | 30 | – | - |
| 9–10 | Fully connected | 10 | – | - |
| 10–11 | Fully connected | 2 | – | – |

**Table 3**
The details of the CNN–ENN model.

| Layers | Type | Number of neurons (output layer) | Kernel size for each output feature map | Stride |
|---|---|---|---|---|
| 0–1 | Convolution | 978 × 3 | 23 | 1 |
| 1–2 | Max-pooling | 489 × 3 | 2 | 2 |
| 2–3 | Convolution | 476 × 5 | 14 | 1 |
| 3–4 | Max-pooling | 238 × 5 | 2 | 2 |
| 4–5 | Convolution | 232 × 10 | 7 | 1 |
| 5–6 | Max-pooling | 116 × 10 | 2 | 2 |
| 6–7 | Convolution | 112 × 10 | 5 | 1 |
| 7–8 | Max-pooling | 56 × 10 | 2 | 2 |
| 9–11 | ENN | 30, 12, 2 | – | – |

where $\alpha$ is a scalar constant that modulates the weight of information feedback derived from the context unit at the previous moment. For $\alpha = 0$, the MENN structure is actually just a standard ENN model and has no memory of the context unit.

In addition, the cross entropy is used as loss function and defined as follows:

$$L(d, y) = -\sum_{j=1}^{N}\Big[d_j(t)\log y_j(t) + \big(1 - d_j(t)\big)\log\big(1 - y_j(t)\big)\Big] \tag{5}$$

where $d_j(t)$ is the target value of the raw data and $y_j(t)$ is the obtained output value.

### 2.5. Architecture of the proposed model

Fig. 3 illustrates the architecture of the proposed model for AF detection and Table 1 summarizes the detailed parameters and compositions of the structure. Firstly, layers 1 to 8 are alternately composed of convolutional layers and max-pooling layers. Subsequently, layers 9 to 11 are installed as the MENN structure for the final classification, where the number of neurons in each layer were set to 30, 14 and 2, respectively. Leaky rectifier linear unit (LReLU) function was employed as the activation function for layer 1, 3, 5, 7, 9 and 10, respectively. In the last layer, the softmax function was used as the activation function, which can predict the class as being NSR or AF through determining the category probability of each ECG segment. The strides were set to 1 and 2 for convolution layer and max-pooling layer, respectively.

In this model, Xavier algorithm was employed for the weight initialization of the network [38]. The goal of this operation is to overcome the vanishing gradient problem in the training stage. Adam algorithm was used for the weight optimization [39], and the values of all biases in the network were originated from the random generation of Gauss distribution. In the context unit of the MENN structure, the parameter $\alpha$ was tuned to 0.21 since it provided the best classification results. The network was trained by setting the value of loss function no more than 0.002 with a batch size of 64 and the maximum epoch limit of 40.

Meanwhile, to demonstrate the classification performance of the proposed model, both the CNN-MLP and CNN–ENN models are stacked by means of the same convolutional layers and pooling layers as the CNN–MENN model for a fair comparison. It should be noted that the only difference among the three network frameworks is the final classification structure, and they are MLP, ENN as well as MENN, respectively. Tables 2 and 3 illustrate the detailed parameters and compositions of CNN-MLP and CNN–ENN, respectively. These parameters of the three models were obtained by the trial-and-error method through the optimization of the accuracy in the network. Meanwhile, an example of layer 7 of a sample normal and AF ECG segment with output shape of 112 × 10 are illustrated as shown in Figs. 4 and 5, respectively.

### 2.6. Performance evaluation

In this work, ten-fold cross-validation was performed to assure the generalization ability of the proposed model for AF detection. The data set is divided into ten equal parts, nine parts are selected for training the model and rest one for testing, and this process is repeated 10 times by shifting the testing data portion [40]. In this case, the mean values of all ten-folds are used for the final evaluation performance, and three standard performance metrics composed of accuracy (ACC), sensitivity (SEN) and specificity (SPF) are obtained as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$SEN = \frac{TP}{TP + FN} \tag{7}$$

$$SPF = \frac{TN}{TN + FP} \tag{8}$$

where:

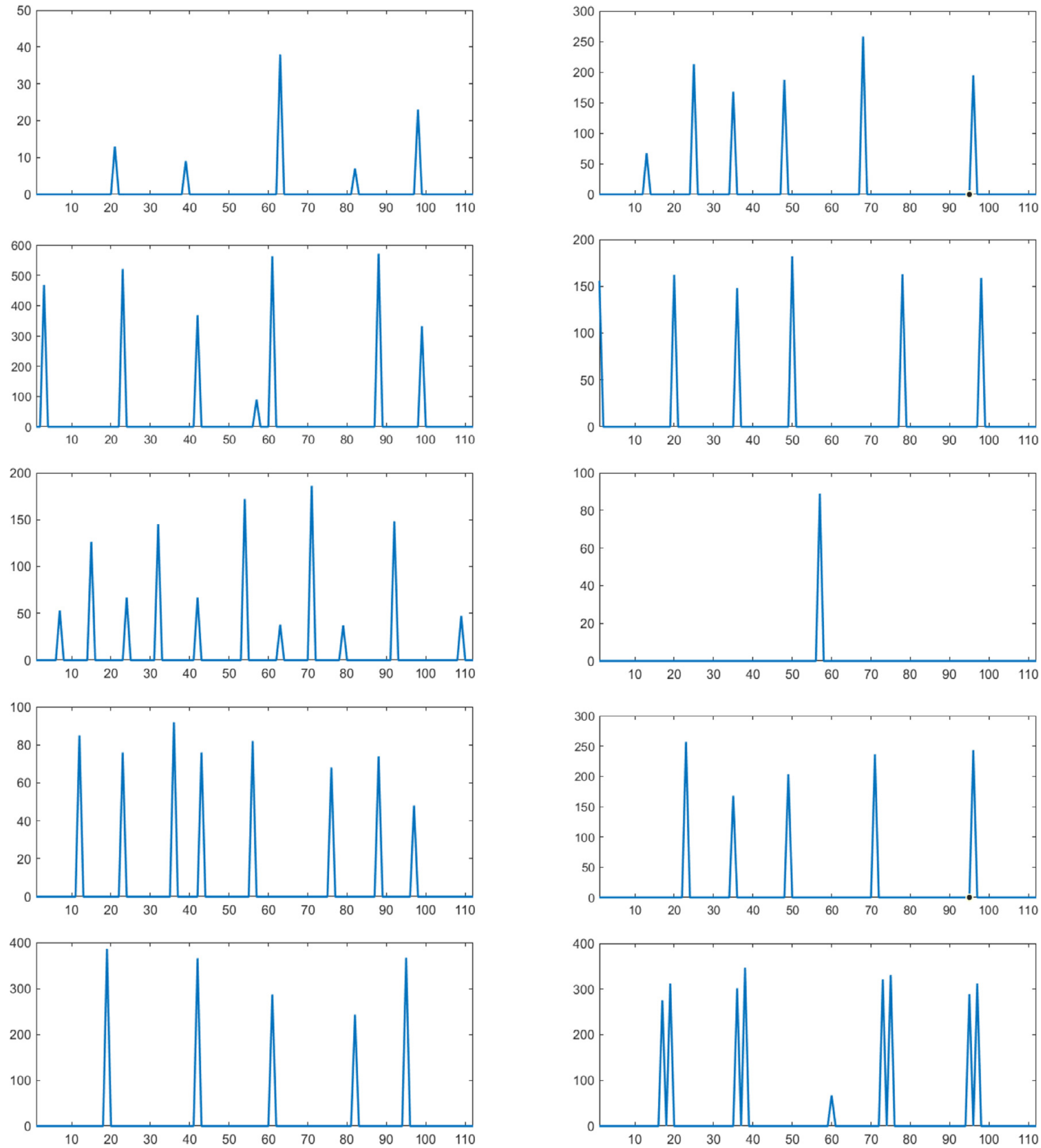

**Fig. 4.** An example of layer 7 of a sample normal ECG segment with output shape of 112 × 10.

- True Positive (TP) indicates the number of AF signals correctly classified.
- True Negative (TN) indicates the number of NSR signals correctly classified.
- False Positive (FP) indicates the number of NSR signals mistaken as AF signals.
- False Negative (FN) indicates the number of AF signals mistaken as NSR signals.

## 3. Results

In this study, a total of 38,312 AF segments and 74,038 NSR segments fed into the proposed model were analyzed across ten-fold cross-validation. To further verify the superiority of the proposed CNN–MENN model, the CNN-MLP and CNN–ENN models were constructed and their classification results were also compared. All the experiments were carried out on a computer of Inter-core i7-8565U CPU@1.80 GHz with 16 GB of RAM and MATLAB software R2018b.

Table 4 shows the confusion matrix for the three proposed models on the MIT-BIH AF data set across ten-fold cross-validation. It can be shown in Table 4 that 97.9% of ECG segments are correctly classified as AF segments by the proposed CNN–MENN model. Also, 97.1% of ECG segments are correctly classified as NSR segments. As for the CNN–ENN and CNN-MLP models, 96.7% and 96.2% of ECG segments are correctly classified as AF segments, respectively, then 96.8% and 95.2% of ECG segments are correctly classified as NSR segments, respectively. In addition, the experimental results demonstrate that both the ENN models improve the classification performance of the model in varying degrees as compared to CNN-MLP. Especially in CNN–MENN, ACC, SEN and SPF are increased by 1.5 to 2 percentage points, which indicates that the proposed model is proven to have a higher precision and discrimination ability for ECG signals,

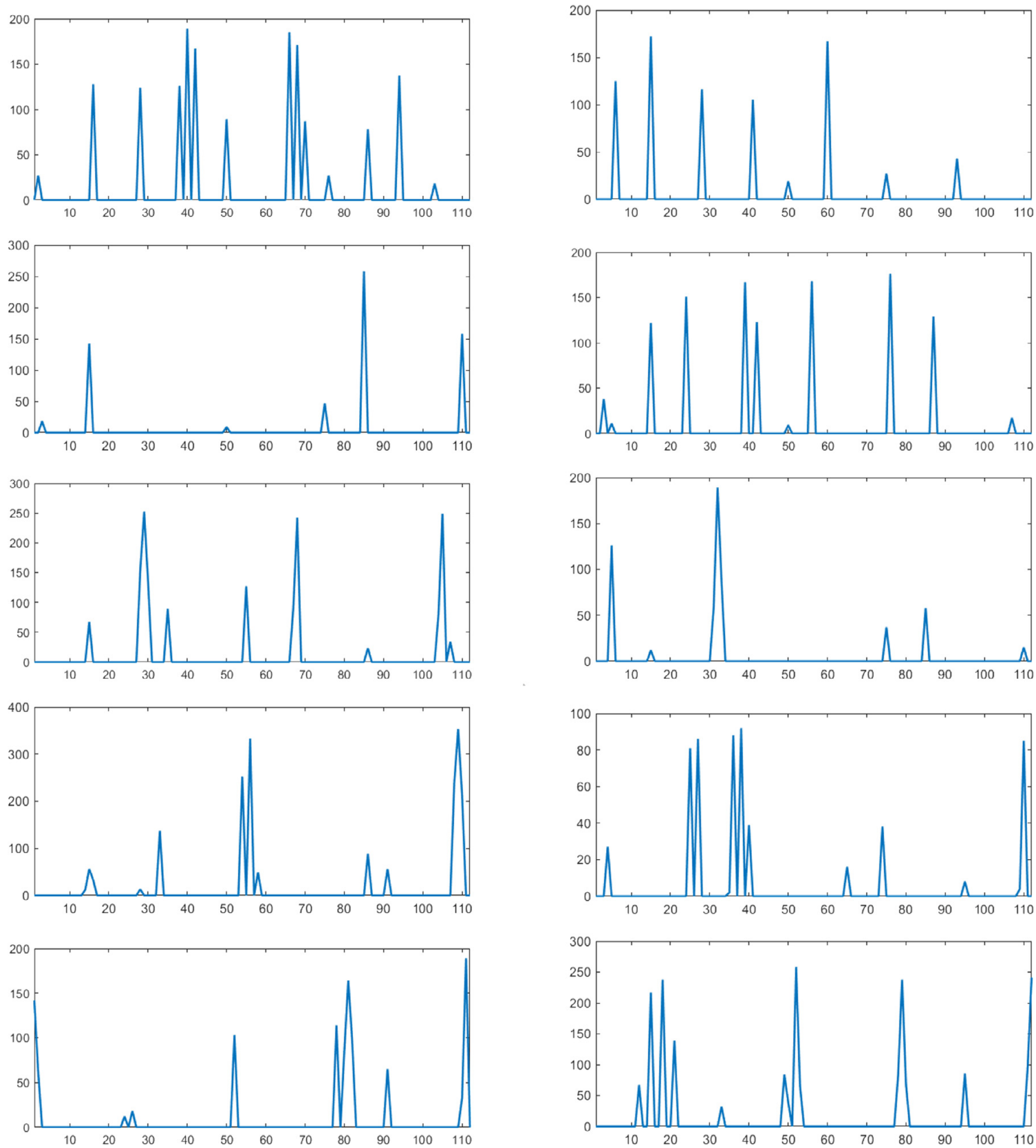

**Fig. 5.** An example of layer 7 of a sample AF ECG segment with output shape of 112 × 10.

**Table 4**
Confusion matrix for three proposed models on the MIT-BIH AF database.

| Model | Original/predicted | NSR | AF | ACC(%) | SEN(%) | SPF(%) |
|---|---|---|---|---|---|---|
| CNN-MLP | NSR | 70,486 | 3552 | 95.5 | 95.2 | 96.2 |
| | AF | 1456 | 36,856 | 95.5 | 96.2 | 95.2 |
| CNN–ENN | NSR | 71,676 | 2362 | 96.8 | 96.8 | 96.7 |
| | AF | 1264 | 37,048 | 96.8 | 96.7 | 96.8 |
| CNN–MENN | NSR | 71,909 | 2129 | 97.4 | 97.1 | 97.9 |
| | AF | 805 | 37,507 | 97.4 | 97.9 | 97.1 |

while also demonstrating the effectiveness of the MENN structure for model improvement. To illustrate the comparison among these values more intuitively, Fig. 6 was plotted based on the bar charts for three standard metrics of different models.

Fig. 7 compares the receiver operating characteristic (ROC) curves for the three models on the MIT-BIH AF database across ten-fold cross-validation. The area under ROC curve (AUC) is a standard used to measure the classification ability of model, and the greater the AUC value indicates that the model system has better classification performance. From Fig. 7, it is visible that on the one hand, the AUC value of CNN–ENN (AUC = 0.984) is slightly greater than that of CNN-MLP (AUC = 0.973), indicating that the existence of the context unit in a standard ENN do enhance the classification ability of model system compared to the traditional MLP; on the other hand, the AUC value of the CNN–MENN model (AUC = 0.991) is the largest in the three models. Therefore, it is inferred that the classification performance of model system is continuously enhanced due to the application of MENN. The main reason is that the MENN structure has more superior dynamic information memory ability such that more spatio-temporal information that exists in the network can be fully mined and utilized as much as possible.

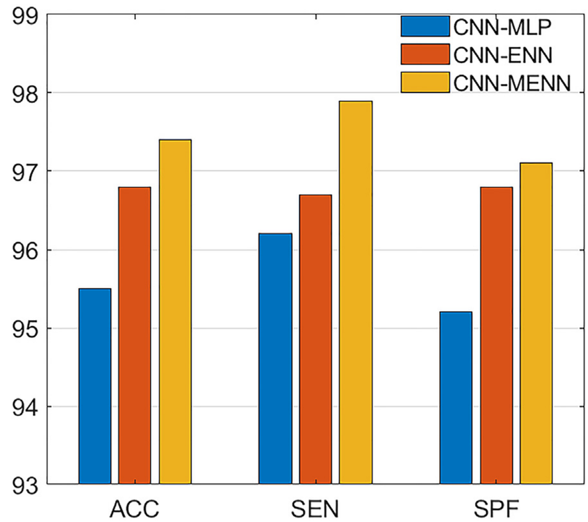


**Fig. 6.** The bar charts for three standard metrics of different models.

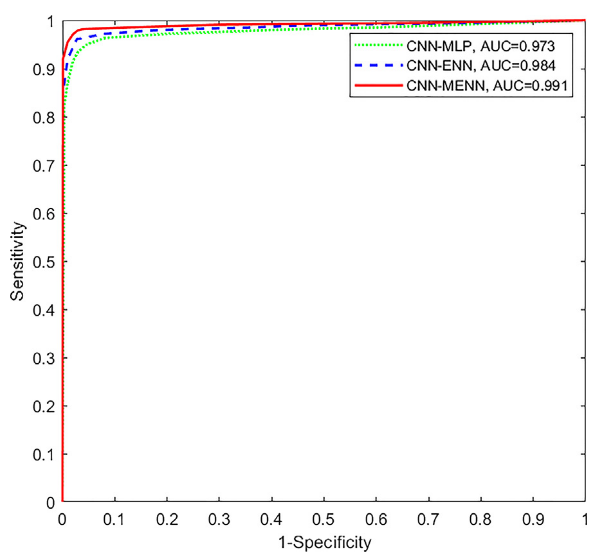


**Fig. 7.** Performance comparison of AF detection results for the ROC curves using three different models.

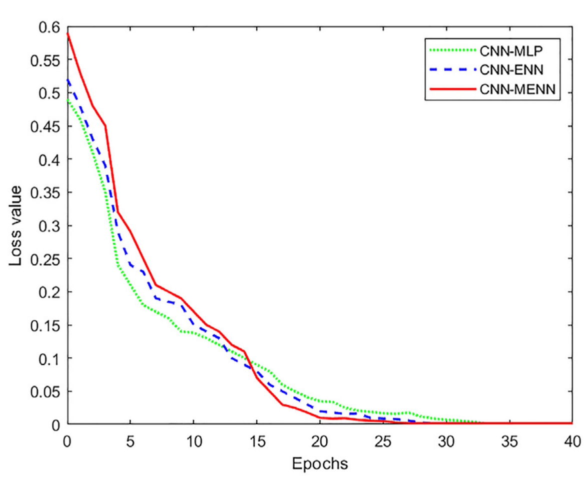


**Fig. 8.** The trend charts of the evolution of convergence process for the three models.

**Table 5**
The detailed data for convergence speed of the three models.

| Model | Stop epoch | Average time for an epoch of training (s) |
|---|---|---|
| CNN-MLP | 33 | 2313.8 |
| CNN–ENN | 29 | 2617.7 |
| CNN–MENN | 27 | 2802.1 |

**Table 6**
The detailed results of the CNN–MENN model across ten-fold cross-validation for AF detection.

| Fold | ACC (%) | SEN (%) | SPF (%) |
|---|---|---|---|
| 1 | 97.2 | 97.8 | 96.9 |
| 2 | 97.3 | 97.7 | 97.1 |
| 3 | 97.6 | 98.1 | 97.3 |
| 4 | 97.5 | 98.2 | 97.1 |
| 5 | 97.5 | 98.1 | 97.2 |
| 6 | 97.4 | 98.0 | 97.1 |
| 7 | 97.6 | 98.2 | 97.3 |
| 8 | 97.3 | 97.6 | 97.1 |
| 9 | 97.5 | 97.9 | 97.3 |
| 10 | 97.2 | 97.7 | 96.9 |

Fig. 8 illustrates the evolution of convergence process for the three models. It is clear that CNN–MENN first achieved the convergence of model system, followed by CNN–ENN, and finally CNN-MLP. In fact, CNN-MLP had converged slightly faster than CNN–ENN and CNN–MENN at the early phase of convergence process. The main reason for this phenomenon may be that both the ENN models need to handle more information feedback than CNN-MLP for convergence. In the latter phase, CNN–MENN has gradually converged by learning more comprehensive feedback information from the hidden layer, and especially from context units. Eventually, the model yielded the fastest global convergence speed after about 27 epochs, CNN–ENN and CNN-MLP took 29 and 33 epochs, respectively. Also, it can be seen from Table 5 that although CNN–MENN, CNN–ENN and CNN-MLP took an average of 2802.1 s, 2617.7 s and 2313.8 s to complete an epoch of training respectively, the proposed CNN–MENN model took the least time in the whole training process. Therefore, based on the above results, it is found that the convergence rate of the CNN–MENN model can be accelerated to some degree through adding the information feedback from the hidden layer and especially from context units, while the model often requires sufficient learning processes of data or relevant information in the early stage. However, as long as the proposed model is trained, the classification for the ECG signals is accurate and fast. It is confirmed that these findings can also be useful in other traditional static networks and deep architectures, which may contribute to the enhancement of model performance.

According to Tables 4 and 6, it can be observed that the CNN–MENN model yielded remarkable classification results across ten-fold cross-validation for AF detection. Three standard classification measures including ACC, SEN and SPF are all more than 96.9% in each fold, which indicates that almost the vast majority of the ECG segments are accurately classified. Especially to illustrate the robustness of the model, the box plots for three metrics were plotted as illustrated in Fig. 9. It is clear that these values are distributed over a tiny range of variations, which are less than 0.3%. These results explicitly demonstrate that the proposed model is robust and reliable to all the sample data. In summary, the model can automatically implement the AF detection in an accurate and stable pattern.

## 4. Discussion

In this research, an end-to-end deep network identification system based on an 11-layers CNN–MENN architecture was developed. The model not only employed CNN to extract useful

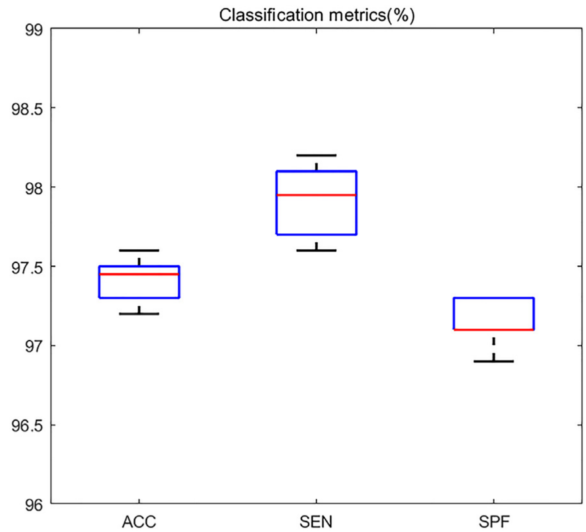


**Fig. 9.** The box plots of three classification metrics of the CNN–MENN model across ten-fold cross-validation for AF detection.

features from the raw ECG signals, but also took advantage of the dynamic information memory ability of MENN to more efficiently improve the classification performance of network. Moreover, it was evaluated across ten-fold cross-validation on the MIT-BIH AF database. To illustrate the classification performance and innovation of the proposed model, detailed comparisons with several state-of-the-art methods for AF detection are summarized in Table 7.

Firstly, the methods of Thomas et al. [9], Henzel et al. [10], Kumar et al. [12] and Boon et al. [13] are actually traditional machine learning algorithms. Feature extraction and classification are actually key factors for the implementation of these methods. However, how to construct an appropriate feature extraction mechanism for the different data source and given problem is still challenging. Furthermore, the generalization ability and robustness of the algorithms are influenced by the different types of classifiers. For instance, employing the supervised or the unsupervised classifier to evaluate the model may have different classification results. In contrast, the proposed model integrates feature extraction and classification into the model and is well suitable for automatically learning essential features against all kinds of high-dimensional data set. Thus, the above mentioned issues can be well addressed. The facts also highlight that in addition to providing the higher classification performance, the proposed model outperforms the above algorithms.

Secondly, the models of Acharya et al. [19] and Xiong et al. [20] mainly adopt CNN to implement ECG signals identification. Acharya et al. [19] performed the same 11-layers network structure as ours, but it requires the ECG heartbeat as model input. The previous study shows that it is more realistic to use the raw ECG signal for the diagnostic results than the ECG heartbeat [41]. Thus, we directly used the raw ECG segments for the classification. Xiong et al. [20] employed up to a 16-layers CNN structure with skip connection for AF signals classification. Although skip connection significantly boosts the ability to learn features and speeds up training time, this model is four layers deeper than ours and also neglects the data inconsistency due to the use of relatively small data sets.

Thirdly, the models of Faust et al. [22] and Andersen et al. [17] primarily perform LSTM to automatically detect AF segments. Faust et al. [22] obtained the high classification results with the ACC of 98.51%, SEN of 98.32% and SPF of 98.67%. However, the use of 100 RR intervals as model input does not ensure that some nonlinear properties of ECG signals are accurately captured, so the method may miss some essential information. Moreover, it is impossible to implement the AF detection in real-time by using too many RR intervals. Andersen et al. [17] combined CNN with LSTM to perform classification. Approximately the ACC of 97.80%, SEN of 98.98% and SPF of 96.95% were obtained after post-processing. In particular, the method had no denoising process and still used RR intervals instead of raw ECG signals as model input, which may affect the reliability of the model. It also involves the higher computational cost, thus there are some limitations in practical application.

In addition, the proposed model was partly motivated by the network structure of Acharya et al. [15]. The architecture of the model is also composed of 8 alternate convolutional layers and pooling layers with 3 fully connected layers, but for the classification of shockable and non-shockable ventricular arrhythmias. However, the class imbalance problems may exist due to the use of limited shockable ECG segments for training the model. In this way, the system performance is also not truly reflected by the evaluation measures obtained. So in order to overcome the above problem, in our work, first more than twice the number of ECG segments used in the original model were supplemented. In particular, AF segments account for 34.1% of the total ECG segments. Training the model on a more balanced data set not only can protect the minority class to alleviate the aforementioned issues effectively, but also is helpful to get a model with better generalization ability. Second, with the further improvement of model structure, the CNN–ENN and CNN–MENN models were constructed and verified on the MIT-BIH AF database. Especially in CNN–MENN, the model not only accounts for the full information feedback from the hidden layer, but also the information feedback from the context unit itself. Furthermore, it is demonstrated that more outstanding classification performances are obtained and the convergence speed of the model is also enhanced to some degree.

The main contribution of this study is to automatically perform the AF detection based on the three deep network models in an efficient and robust manner. Especially with the framework of the CNN–MENN model, the excellent classification results are achieved for ECG signals classification, and its advantages and innovations are as follows:

- The proposed model is the first time to diagnose AF based on CNN and MENN using raw ECG signals instead of other ECG features.
- It is an end-to-end classification mechanism. Therefore, feature extraction, feature selection and classification are not needed.
- Especially compared with several state-of-the-art methods as summarized in Table 7, the model yielded high classification performance.
- The convergence rate of the model is also accelerated to some degree due to the application of MENN instead of ENN and MLP.
- Through the analysis of over 110,000 ECG segments from all 23 different records, ten-fold cross-validation is implemented to make the model robust and reliable.

Meanwhile, some drawbacks of the proposed model are as follows:

- This model system is only focused on AF detection.
- It still requires more larger and diverse data set for learning and training.

**Table 7**
Comparison of classification performance of the proposed model with several state-of-the-art strategies on the MIT-BIH AF database.

| Authors | Year | Methodology | Classes | ACC(%) | SEN(%) | SPF(%) |
|---|---|---|---|---|---|---|
| Thomas et al. [9] | 2015 | DWT, DTCWT, morphological features + ANN | 5 | 97.86 | 88.60 | 96.18 |
| Henzel et al. [10] | 2017 | RR intervals + Generalized linear model | 2 | 93.28 | 90.34 | 95.46 |
| Acharya et al. [19] | 2017 | CNN | 4 | 92.50 | 98.09 | 93.13 |
| Xiong et al. [20] | 2017 | CNN + Skip connection | 4 | 86 | – | – |
| Kumar et al. [12] | 2018 | LEE features + RF | 2 | 96.8 | 95.8 | 97.6 |
| Boon et al. [13] | 2018 | HRV features + SVM | 5 | 87.7 | 86.8 | 88.7 |
| Faust et al. [22] | 2018 | LSTM | 2 | 98.51 | 98.32 | 98.67 |
| Andersen et al. [17] | 2019 | CNN + LSTM | 2 | 97.80 | 98.98 | 96.95 |
| **This work** | 2019 | **CNN + MLP** | **2** | **95.5** | **96.2** | **95.2** |
| | 2019 | **CNN + ENN** | **2** | **96.8** | **96.7** | **96.8** |
| | 2019 | **CNN + MENN** | **2** | **97.4** | **97.9** | **97.1** |

## 5. Conclusion

In this work, a deep learning approach for computer-aided AF detection based on an 11-layers CNN–MENN structure was presented. The structure was also evaluated on the MIT-BIH AF database across ten-fold cross-validation, while yielding high classification performance. Moreover, the results also demonstrated that the convergence rate of the model was accelerated to some degree. The significant observation of this work is that the combination of CNN and MENN for ECG signals classification not only enables the model to automatically extract useful features from raw ECG signals, but also is helpful to improve the final classification performance. Besides, the identification system is able to perform the automated AF detection and alleviate the misdiagnosis problems that confront many cardiologists without the complicated manual feature extraction. Thus, the proposed model has great potential to expand to computer software platform in the hospitals to reduce mortality and save lives. In future work, we will combine more modified ENN structures with deep learning frameworks for ECG signals analysis.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

The author would like to express his sincere gratitude to the anonymous referees, the editor, Dr. Ying Li and Prof. Zhanjie Song for many valuable suggestions and comments that helped to improve the paper. This research was supported by the National Natural Science Foundation of China under Grant No. 61379014.

## References


[1] E.J. Ciaccio, A.B. Biviano, E.Y. Wan, et al., Development of an automaton model rotational activity during atrial fibrillation, Comput. Biol. Med. 83 (2017) 161–181.

[2] G.H. Kruger, R. Latchamsetty, N.B. Langhals, et al., Bimodal classification algorithm for atrial fibrillation detection from m-health ECG recordings, Comput. Biol. Med. 104 (2019) 310–318.

[3] Y. Hagiwara, H. Fujita, O.S. Lih, et al., Computer-aided diagnosis of atrial fibrillation based on ECG signals: A review, Inform. Sci. 467 (2018) 99–114.

[4] Z. Golrizkhatami, A. Acan, ECG classification using three-level fusion of different feature descriptors, Expert Syst. Appl. 114 (2018) 54–64.

[5] G.J. Wang, C.S. Zhang, Y.P. Liu, et al., A global and updatable ECG beat classification system based on recurrent neural networks and active learning, Inform. Sci. 000 (2018) 1–20.

[6] S.K. Berkaya, A.K. Uysal, E.S. Gunal, et al., A survey on ECG analysis, Biomed. Signal Proces. 43 (2018) 216–235.

[7] M.Q. Deng, C. Wang, M. Tang, et al., Extracting cardiac dynamics within ECG signal for human identification and cardiovascular diseases classification, Neural Networks. 100 (2018) 70–83.

[8] G. Sannino, G.D. Pietro, A deep learning approach for ECG-based heartbeat classification for arrhythmia detection, Future Gener. Comput. Syst. 86 (2018) 446–455.

[9] M. Thomas, M.K. Das, S. Ari, Automatic ECG arrhythmia classification using dual tree complex wavelet based features, AEU-Int. J. Electron. C 69 (4) (2015) 715–721.

[10] N. Henzel, J. Wrbel, K. Horoba, Atrial fibrillation episodes detection based on classification of heart rate derived features, in: 2017 MIXDES-24th Inter-National Conference Mixed Design of Integrated Circuits and Systems, IEEE, 2017, pp. 571–576.

[11] U.R. Acharya, H. Fujita, M. Adam, et al., Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: A comparative study, Inform. Sci. 377 (2017) 17–29.

[12] M. Kumar, R.B. Pachori, U.R. Acharya, Automated diagnosis of atrial fibrillation ECG signals using entropy features extracted from flexible analytic wavelet transform, Biocybern. Biomed. Eng. 38 (2018) 564–573.

[13] K.H. Boon, M. Khalil-Hani, M. Malarvili, Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III, Comput. Meth. Prog. Bio. 153 (2018) 171–184.

[14] A. Hochstadt, E. Chorin, S. Viskin, et al., Continuous heart rate monitoring for automatic detection of atrial fibrillation with novel bio-sensing technology, J. Electrocardiol. 52 (2019) 23–27.

[15] U.R. Acharya, H. Fujita, S.L. Oh, et al., Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network, Future Gener. Comput. Syst. 79 (2018) 952–959.

[16] Y.D. Zhang, K. Muhammad, C. Tang, Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform, Multimed. Tools. Appl. 77 (2018) 22821–22839.

[17] R.S. Andersen, A. Peimankar, S. Puthusserypady, A deep learning approach for real-time detection of atrial fibrillation, Expert Syst. Appl. 115 (2019) 465–473.

[18] M.M.A. Rahhal, Y. Bazi, H. Alhichri, et al., Deep learning approach for active classification of electrocardiogram signals, Inform. Sci. 345 (2016) 340–354.

[19] U.R. Acharya, H. Fujita, O.S. Lih, et al., Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network, Inform. Sci. 405 (2017) 81–90.

[20] Z.H. Xiong, M.K. Stiles, J.C. Zhao, Robust ECG signal classification for detection of atrial fibrillation using a novel neural network, in: 2017 Computing in Cardiology (CinC), IEEE, 2017, pp. 1–4.

[21] S. Singh, S.K. Pandey, U. Pawar, et al., Classification of ECG arrhythmia using recurrent neural networks, Procedia Comput. Sci. 132 (2018) 1290–1297.

[22] O. Faust, A. Shenfiled, M. Kareem, et al., Automated detection of atrial fibrillation using long short-term memory network with RR interval signals, Comput. Biol. Med. 102 (2018) 327–335.

[23] J.G. Wu, H. Lundstedt, Prediction of geomagnetic storms from solar wind data using elman recurrent neural networks, Geophys. Res. Lett. 23 (4) (2013) 319–322.

[24] S. Krishnan, S. Lokesh, M.R. Devi, An efficient elman neural network classifier with cloud supported internet of things structure for health monitoring system, Comput. Netw. 151 (2019) 201–210.

[25] C.J. Yu, Y. Li, M.J. Zhang, An improved wavelet transform using singular spectrum analysis for wind speed forecasting based on elman neural network, Energy Convers. Manage. 148 (2017) 895–904.

[26] L.G.B. Ruiz, R. Rueda, M.P. Cuéllar, et al., Energy consumption forecasting based on elman neural networks with evolutive optimization, Expert Syst. Appl. 92 (2018) 380–389.

[27] H.P. Hu, H.Y. Wang, Y.P. Bai, et al., Determination of endometrial carcinoma with gene expression based on optimized elman neural network, Appl. Math. Comput. 341 (2019) 204–214.

[28] M.H.A. Shukri, M.S.A.M. Ali, M.Z.H. Noor, et al., Investigation on elman neural network for detection of cardiomyopathy, in: Control and System Graduate Research Colloquium, IEEE, 2012, pp. 328–332.
[29] A.L. Goldberger, L.A.N. Amaral, L. Glass, et al., PhysioBank, physioToolkit, and physionet components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) 215–220.
[30] L.B. Marinho, N.D.M.M. Nascimento, J.W.M. Souza, et al., A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification, Future Gener. Comput. Syst. 97 (2019) 564–577.
[31] K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybernet. 36 (4) (1980) 193–202.
[32] Y. Lecun, L. Bottou, Y. Bengio, et al., Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[33] Y.X. Hao, M. Usama, J. Yang, et al., Recurrent convolutional neural network based multimodal disease risk prediction, Future Gener. Comput. Syst. 92 (2019) 76–83.
[34] S.H. Wang, Y.D. Lv, Y.X. Sui, et al., Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling, J. Med. Syst. 42 (1) (2018) 1–11.
[35] J.L. Elman, Finding structure in time, Cogn. Sci. 14 (2) (1990) 179–211.
[36] P.H. Li, Y.G. Li, Q.Y. Xiong, et al., Application of a hybrid quantized elman neural network in short-term load forecasting, Int. J. Electr. Power 55 (2014) 749–759.
[37] K. Kolanowski, A. Świetlicka, R. Kapela, et al., Multisensor data fusion using elman neural networks, Appl. Math. Comput. 319 (2018) 236–244.
[38] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (PMLR), vol. 9, 2010, pp. 249–256.
[39] D. Kingma, J. Ba, Adam: A method for stochastic optimization, Comput. Sci. (2014).
[40] S.S. Chen, W. Hua, Z. Li, et al., Heartbeat classification using projected and dynamic features of ECG signal, Biomed. Signal Proces. 31 (2017) 165–173.
[41] C. Han, S. Li, Automated interpretable detection of myocardial infarction fusing energy entropy and morphological features, Comput. Methods Programs Biomed. 175 (2019) 9–23.

**Jibin Wang**, Ph.D. is a researcher working in School of Mathematics, Tianjin University, Tianjin 300354, China and Visual Pattern Analysis Research Lab, Tianjin University, Tianjin 300072, China. He graduated from Department of Mathematics, Northwest University of China. In the past year, he has published 3 journal papers. His main academic interests are artificial intelligence, biomedical signals, and machine learning.