Biomedical Signal Processing and Control 83 (2023) 104628

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

A novel time representation input based on deep learning for ECG classification

Youhe Huang, Hongru Li *, Xia Yu

College of Information Sciences and Engineering, Northeastern University, Shenyang 110819, PR China

ARTICLE INFO

Keywords:
Heartbeat classification
Time characteristic representation
Information Fusion
Deep learning

ABSTRACT

Electrocardiogram (ECG) is an important tool used to analyze abnormal heart activity and assess heart health, especially in remote cardiac health monitoring. Although deep learning has achieved significant results in automatic ECG classification, how to combine the characteristics of ECG physiological signals to construct inputs or features with differentiation is still a key point of classification. To this end, a novel representation input method with temporal characteristics was proposed in this paper. At first, the temporal characteristic of ECG signals was extracted and transformed into a time representation input with the original input. Subsequently, the deep learning network combining Convolutional Neural Network and Long Short-Term Memory was employed for feature extraction. Simultaneous attention mechanism was used to focus on feature differences. The proposed method was validated in the classification of five classes of heartbeats (Normal heartbeat, Left bundle branch block heartbeat, Right bundle branch block heartbeat, Atrial Premature Contraction, Premature ventricular contraction), achieving a higher average accuracy, precision, sensitivity, and specificity of 98.95%, 97.07%, 96.54%, and 99.33% respectively in the MIT-BIH arrhythmia database. The results show that our method is able to combine the periodic characteristics of ECG to construct a better temporal representation input than traditional feature fusion. This method can provide a new way to classify similar physiological signals with periodic characteristics.

1. Introduction

At present, there are many techniques available in medicine for monitoring the heartbeat, but most are used for clinical diagnosis. For example, the hospital uses 12-lead ECG equipment to check heart related diseases. And these detection methods are not suitable for remote real-time health monitoring. As a non-invasive detection method commonly used in clinical medicine, electrocardiogram (ECG) plays an important role in health diagnosis by detecting obvious or potentially abnormal cardiac activity [1,2]. With the development of technology, wearable single-lead ECG devices have been implemented for remote heart health monitoring and abnormality detection. The single-lead ECG devices have the advantage of small and lightweight, being able to wear for a long time, and monitoring arrhythmia at all times. It can provide critical information during the patient return phase or daily monitoring, facilitating daily management by physicians. Automatic arrhythmia detection can realize accurate detection of single-lead ECG devices. Compared with traditional manual detection, automatic arrhythmia detection can save a lot of time and labor cost. At present, although some products have been applied to remote arrhythmia detection, the overall effect is not satisfactory due to individual differences, data acquisition noise, algorithms, and other conditions. Hence, automatic ECG classification has obtained extensive attention from researchers.

In terms of methods, automatic ECG classification can be divided into traditional methods and deep learning methods. The traditional approaches explore a large number of manual feature representations as well as different classical classifiers. The commonly used features mainly include energy features [4,5], time domain features [6,7], frequency domain features [7,8], morphological features [7–10], etc. Classifier such as support vector machine (SVM) [11,12], random forest (RF) [11,13,14], decision tree (DT) [15], K-nearest neighbor (KNN) [11,16], artificial neural network [17] and other classifier have been utilized for classification. The pivotal task of traditional ECG classification is to construct distinguishing features for classifiers. The excellent features are clearly distinguishable. With the advent of deep learning, feature extraction and classification have been treated as a whole. As a result, the task of constructing distinctive features shifts to constructing input representations with differences. Yao et al. [18] achieved spatial

* Corresponding author.
E-mail addresses: syhuangyouhe@163.com (Y. Huang), lihongru@ise.neu.edu.cn (H. Li), yuxia@ise.neu.edu.cn (X. Yu).

https://doi.org/10.1016/j.bspc.2023.104628
Received 8 September 2022; Received in revised form 4 December 2022; Accepted 25 January 2023
Available online 14 February 2023

temporal fusion input of ECG signals using deep learning. Golrizkhatami et al. [19] extracted different levels of inputs in the deep learning model for fusion classification. However, these inputs are only the raw signal, and the differences between inputs are all extracted by deep learning model. The inputs themselves do not reflect the differences between each other. At present, the focus of ECG classification is on feature engineering. The key is to construct robust features with strong characterization of intrinsic differences in signals. The use of deep learning to extract features relies too much on algorithms, ignoring the difference and interpretation of ECG itself. Most of feature fusion methods require the construction of parallel or serial network structures, which are cumbersome and redundant with high feature dimension. Therefore, we seek to construct a representation that can distinguish or amplify the differences between different abnormal signals.

There are also many studies that construct different inputs to achieve better feature representation based on deep learning. Some studies converted 1D signals to 2D images for deep learning processing [20–22], and some scholars used a certain length of ECG signal instead of a single heart beat as input [23,24]. To better construct feature representations, Fujita et al. [25] employed differential approach to convert non-stationary signals to stationary signals for deep learning classification. Islam et al. [26] explored the difference between abnormal and normal signals in atrial fibrillation and translated this morphological irregularity into an input representation. The above studies all focus on building more distinctive features. However, there is no uniform standard on how to construct better features or inputs in the ECG classification field. As the input design becomes more and more cumbersome, the deep learning model structure becomes more and more complex, which also leads to algorithms occupying a lot of computing resources. Therefore, how to use the characteristics of ECG itself to construct a concise input that can reflect the difference between various anomalies is a direction for the development of deep learning ECG classification.

ECG is a time series signal with distinct periodicity. In conventional methods, this property is usually utilized for feature construction. Many literatures have shown that RR interval is one of the most frequently used and time dynamic ECG features. The RR interval can reflect the length of a heartbeat in the time dimension, which is the embodiment of ECG periodicity. The RR interval of abnormal heartbeat is very different from the normal RR interval, so many scholars employ the RR interval to construct features. In medicine, the heart rate variability (HRV) information based on RR interval is often used for clinical judgment. In the ECG classification, HRV information or RR interval also be used for classification [27,28]. The RR interval and other morphological features were adopted in the literature [7]. Ye et al. [29] used discrete wavelet transform (DWT) features with RR interval features to achieve four types of anomaly classification. Rahul et al. [30] performed anomaly detection using multiple classifiers based on RR interval features. In deep learning, RR interval is mainly used for fusion with deep learning features. Xiang et al. [10] fused deep learning features with morphological features for heartbeat classification. Wu et al. [11] realized the abnormal detection of atrial fibrillation (AF) by making full use of time and frequency domain features. Li et al. [31] used morphological features, RR intervals and deep learning features for abnormal ECG detection. Compared with other significant features such as QT interval, amplitude or statistical features, RR interval has the advantage of strong difference and easy measurement transformation. Consequently, a reasonable exploitation of RR interval characteristics is a key point for constructing the input or feature representation.

It is known from the existing literature that deep learning outperforms traditional machine learning methods in ECG detection of abnormal arrhythmias [19–22]. Compared with traditional feature extraction methods, deep learning can extract subtle differences in the input signal, so as to achieve better classification. Acharya et al. [32] developed a 9-layer Convolutional Neural Network (CNN) to identify five types of heartbeats. Li et al. [33] adopted deeper CNN model to achieve a better ECG classification. Moreover, the Long-Short Term Memory (LSTM) models are often employed in combination with CNN model to obtain better temporal characteristics. The traditional CNN has a good advantage in spatial feature extraction, but has general ability for temporal feature analysis. In contrast, LSTM is a recursive structured convolutional neural network with better temporal feature extraction capability. Many literatures have used LSTM structure to make up for the shortcomings of CNN. Oh et al. [34] used a combination of CNN and LSTM model to diagnose abnormal ECG signals. To detect coronary artery disease (CAD), Tan et al. [35] presented a CNN-LSTM network based on ECG signals. For the same input, various models focus on different views and extract different features. To perform better extraction and concentration of deep learning features, some studies used attention mechanism. Yao et al. [18] employed attention mechanism for deep learning feature processing to enhance the strength of temporal features. Moreover, attention mechanism was used to improve the generalization of the model in [36]. In ECG classification, designing models that match the characteristics of different inputs can be more helpful to the task. As such, the design of a model that can better extract the important features from the input information is another key point in classification.

To address these problems, this paper proposed a time representation input based on deep learning for ECG classification. The average RR interval of each heartbeat was selected as the dynamic time feature according to the ECG periodicity characteristics. The feature was then transformed into a time representation coefficient according to the temporal representation formula we designed. The time representation coefficients can amplify the differences between ECG heartbeat inputs and provide a better input representation for the classification model. Subsequently, a combined CNN and LSTM model with attention mechanism was constructed to perform spatial and temporal feature extraction on the time representation input to achieve the final classification. The attention mechanism can improve the ability of deep learning to extract features from time representation inputs. The presented time representation input can combine the time characteristics of the signal with deep learning to achieve accurate decisions.

The main contributions of this paper can be summarized as follows:

1) A novel time representation input is proposed to scale the input signal using the dynamic time signature RR interval of the ECG signal. The time representation input can measure the difference between various classes of inputs in the time dimension and will make the discrepancies between the inputs larger.
2) A deep learning network structure conforming to the temporal representation input is constructed based on CNN-LSTM with attention mechanism, which can extract spatial–temporal features and focus on critical differences. It is able to fuse the temporal characteristics of ECG signal itself with the deep learning features for better extraction of characteristics of ECG.
3) The time representation input is a new and concise fusion approach of RR time features with deep learning features, which can improve the classification effect and solve the problems such as time-consuming and redundancy of traditional deep learning feature fusion. It can provide a reference for other similar periodic physiological signals.

The rest of this paper is organized as follows. Methodology is given in Section 2. Database and experimental settings are presented in Section 3. Results and discussions are given in Section 4. At the end, conclusions are drawn in Section 5.

2. Methodology

2.1. Time representation input

The arrhythmia classification task can be described as a process of finding the corresponding output y for a given input x after a function transformation. The essential principle is the nonlinear mapping of functions. According to the representation learning theory of deep learning [37], the process can be described as the following Eq. (1) using deep learning.

$$y = fun(RPT(x_i)) \tag{1}$$

where *fun* denotes the classification function; *RPT* denotes the deep learning middle layer that transforms the input x_i into a distributed knowledge representation.

According to Eq. (1), it is clear that this process is completely data-driven. With different inputs, the intermediate feature representations and output results will change accordingly. The greater the difference between the inputs, the more the model can distinguish the difference between the inputs in the classification mission. The main purpose of the classification model is to construct better inputs or features and classifiers. The learning capability of any classifier in engineering applications is limited, including deep learning. The learning capability of any classifier in engineering applications is limited, including deep learning. Consequently, establishing more expressive inputs or features based on the attributes of the application object will greatly improve the classification effect. To obtain better input features combined with deep learning, the proposed time representation input is shown as Eq. (2). $mean(pre\_rr, pos\_rr)$ is the mean value of the RR interval before and after a heartbeat, indicating the time scale of heartbeat. $\alpha \times f$ represents the length of a normal signal heartbeat. Since the RR interval of anomalous heartbeat is usually irregular, the ratio of the RR mean to the single heartbeat length can quantify this irregularity uniformly.

$$x_t = T_i \times x_i = \frac{mean(pre\_rr, pos\_rr)}{\alpha \times f} \times x_i \tag{2}$$

where x_t is the time representation input; T_i is the time representation coefficient, x_i is the original heartbeat input; *pre_rr* is the interval between the previous heartbeat and the current heartbeat; *pos_rr* is the interval between the current heartbeat and the posterior heartbeat; *mean* is the mean calculation; α is a fixed time length of a signal heartbeat, usually taken as 0.8 s. and f is the sampling frequency of ECG signals, where it takes 360 Hz.

The time representation input can uniformly measure and compare the RR interval mean of different heartbeats. It translates the dynamic time characteristics of ECG signals (RR time interval) into measurement coefficients that can maximize the difference between various kinds of heartbeats from the input. After processed by deep learning, it is able to obtain both dynamic temporal features and deep learning features of the heartbeat, which provides a better feature representation for decision making. This approach is more concise than the traditional RR interval fusion or image input representation.

Fig. 1 shows the comparison of a normal heartbeat input before and after the time representation transformation. The time representation coefficient of this heartbeat is less than 1, so the overall decrease in the temporal representation input amplitude can be observed in the figure. It indicates a strong effect of the time representation coefficient on the input. Drawing on the idea of variable weighting and feature fusion, the value of the input data determines its impact on the decision at the end. Therefore, we designed this formula using the dynamic time characteristics of the ECG signal. Due to the variation of the input, deep learning is able to learn the dynamic properties of signals.

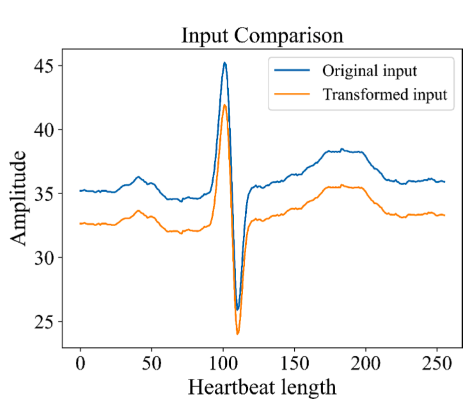


Fig. 1. Comparison of time representation input with original input.

2.2. CNN-LSTM-Attention structure

2.2.1. CNN-LSTM structure

With the development of computer technology, CNN models have been widely used in traditional classification and regression tasks. Compared with traditional classification algorithms, CNN has the ability of powerful nonlinear mapping. This is mainly because CNN has a multilayer convolutional structure. The convolution calculation can reduce the computational complexity and enable weight sharing. After convolutional operations, the convolutional kernel can abstract the low-dimensional input into a high-dimensional feature representation. In ECG classification, CNN can extract a large number of neglected detailed features, which is unmatched by manual features.

CNN models have better extraction and representation for spatial information, but they are weaker in handling time for time series. Hence, in order to cover the deficiency of CNN's insensitivity to temporal information, Recurrent neural network (RNN) model came into being. RNN is able to extract the temporal links before and after the input sample points, which is more suitable for time series analysis. One of the most representative models is the LSTM model. As an improved RNN model, the LSTM model is widely used to handle time series tasks. It solves the gradient disappearance or gradient explosion problem in RNN model. The LSTM can achieve better extraction of temporal characteristics mainly from the three gate cell control structures, as shown in Fig. 2.

The long-term memory state C_t is determined by forget gate f_t and input gatei_t, as shown in Eq. (3).

$$\begin{cases} C_t = f_t \circ C_{t-1} + i_t \circ C_t' \\ C_t' = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \end{cases} \tag{3}$$

where C_{t-1} denotes long-term memory of the previous moment, C_t' indicates the input state, W_c denotes the weight matrix, and $[h_{t-1}, X_t]$ denotes the vector consisting of the output at the previous moment and the current input.

The calculation of forget gate f_t and input gate i_t is shown in Eq. (4) and (5).

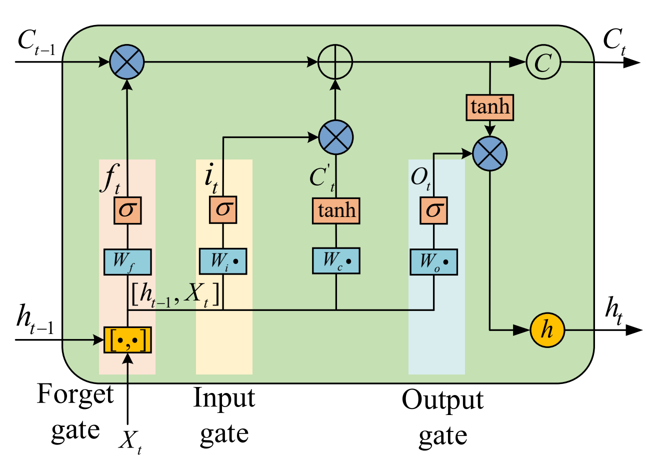


Fig. 2. Structure of LSTM memory block.

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \quad (4)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \quad (5)$$

where σ is activation function, b_f and b_i are the bias terms of forget gate and input gate, respectively.

The output value of the LSTM is determined by the output gate and the long-term memory stateC_t. The mathematical expression is given by Eq. (6):

$$h_t = O_t \circ \tanh(C_t) \quad (6)$$

where the output gate expression is shown by Eq. (7).

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (7)$$

2.2.2. Attention mechanism

Although the CNN-LSTM model can extract the spatial and temporal features of the time representation inputs, we want to spotlight the differences between the inputs of the model. Accordingly, the attention mechanism is introduced into the deep learning model. The attention mechanism is able to focus on the key information in a large number of features of deep learning and ignore the unimportant information, thus achieving better feature representation. It is due to this property that the attention mechanism is more suitable for processing the time representation input information presented in this paper. The specific calculation principle is shown in the Eq. (8).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_{keys}}}\right)V \quad (8)$$

where *Attention* denotes the attention output,Q,K, and V denotes three different input keyword matrices respectively, *softmax* denotes activation function, and d_{keys} denotes the dimension of Q and$K$.

2.3. The classification model

In this paper, a 10-layer CNN-LSTM network was designed to process the time representation input, including eight convolutional layers, one LSTM layer and one attention layer, as shown in Fig. 3. The network we adopted was improved from VGG-13 network model [14]. The existing literature and our own experiments show that ECG classification does not require a very deep network structure to process, although the data volume is large [16,22,23]. Therefore, VGG-13 model is a better choice. With a 13-layer structure, this network model can achieve good time series classification. On the basis of this network structure, we added the LSTM network and attention layer. And the network was improved according to the simulation process, such as reducing the number of layers and adjusting the parameters. Finally, the current 10-layer CNN-LSTM network model was formed. The time representation input is obtained after processing RR interval and single heartbeat, which is then fed into the network. The CNN part can better extract the spatial information in a single shot, while the LSTM part can absorb the temporal information. The addition of the attention mechanism enables focused analysis of deep learning features and improves model feature construction. Finally, the discrimination results of five categories are outputted.

The specific parameter information is shown in Table 1. For better feature extraction of 1D time series, we adopted a 1 × 32 convolution kernel instead of the traditional 1 × 3 or 1 × 7 convolution kernel in the first four layers of convolution. Compared to smaller convolutional kernels, larger width convolutional kernels are able to extract information associated in a certain range of 1D sequences. During training, the learning rate was set to 1×10^{-4} and the number of iterations was set to 100. To prevent overfitting of the model, dropout structure was adopted. Dropout is able to randomly delete the number of neurons in the network while keeping the input and output neurons unchanged.

The cross-entropy function was applied in this model to classify the five different heartbeats sorts. As a metric function, cross-entropy is widely used in deep learning classification tasks. It is able to measure the small differences between the training samples and the expectation. The calculation process can be expressed by Eq. (9).

$$H(p, q) = -\sum_{i=1}^{n} p(x_i)\log(q(x_i)) \quad (9)$$

where $H(p, q)$ stands the cross-entropy result, p stands the true distribution of the sample, q stands the predictive distribution of the model, and x_i is the input.

3. Database description and experimental settings

3.1. Database description

The MIT-BIH arrhythmia database is one of the commonest

Table 1
The specific parameter of CNN-LSTM-Attention model.

Layer types	No. of filters	Kernel size	Stride
Convolution 1	32	1 × 32	1
Convolution 2	32	1 × 32	1
Max pooling 1	1	1 × 2	2
Dropout 1	–	0.5	–
Convolution 3	64	1 × 16	1
Convolution 4	64	1 × 16	1
Max pooling 2	1	1 × 2	2
Dropout 2	–	0.5	–
Convolution 5	128	1 × 3	1
Convolution 6	128	1 × 3	1
Max pooling 3	1	1 × 2	2
Dropout 3	–	0.5	–
Convolution 7	256	1 × 3	1
Convolution 8	256	1 × 3	1
Max pooling 4	1	1 × 2	2
Dropout 4	–	0.5	–
LSTM layer	256	–	–
Attention layer	16	–	–

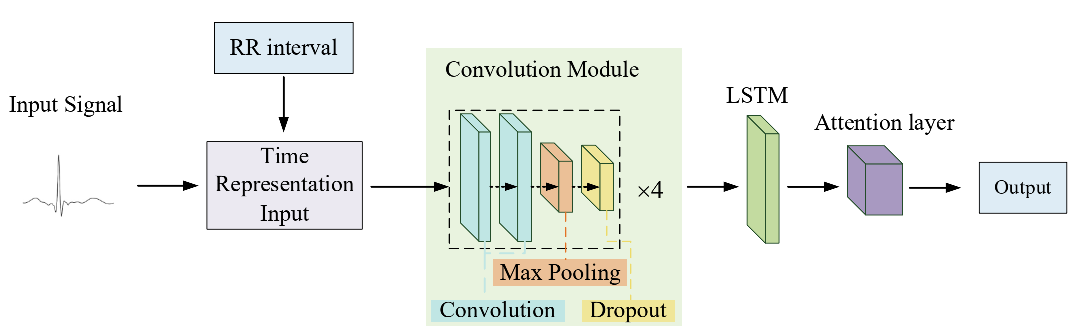


Fig. 3. The classification model.

arrhythmia databases for studying ECG signal processing. The database contains 48 records from 47 individuals, with a sampling rate of 360 Hz and a length of approximately 30 min. Each record contains two sampling channels, the first is modified II leads (MLII) and the other is usually V1 lead (or V2, V4 or V5 leads depending on the subject). Since the MLII lead sampling waveform is the most standard and the reflection characteristics are obvious, all data in this paper adopted the MLII lead signal. The entire data set was divided into DS1 and DS2 according to the inter-patient pattern, with four records with pacemakers removed [38]. For better inter-patient classification, an inter-specific strategy based on active learning was used, with the data in DS1 and the first 5 min of data from each record in DS2 as the training set and the remaining DS2 data as the test set [10,34,39]. This approach complies with the AAMI procedure specification. Meanwhile, according to the AAMI classification categories for ECG signals, we selected five types of heartbeats: Normal heartbeat (N), Left bundle branch block heartbeat (LB), Right bundle branch block heartbeat (RB), Atrial Premature Contraction (APC), Premature ventricular contraction (PVC), as samples for classification.

The delineated training and test sets are shown in Table 2. Since the ECG signals of the five categories has the highest number of class N. In order to keep the balance of the number with other arrhythmia categories, we downsampled the N class of ECG signals in the data. In the training set and test set, 220 heartbeats of N class were randomly selected from the records. If the number was less than 220, all heartbeats were extracted. Finally, 4000 *N*-type training samples and 4162 test samples were obtained. Since the number of arrhythmias in the other four categories was small, we extracted all the abnormalities in each record in the training set. Finally, 3946 LB heart beats, 3777 RB heart beats, 807 APC heart beats, and 3682 PVC heart beats were obtained. The test set was then extracted from DS2 by removing the first 5 min of each record, and a total of 3552 LB heartbeats, 2934 RB heartbeats, 1499 APC heartbeats, and 2657 PVC heartbeats were obtained.

3.2. Preprocessing

ECG signal contains baseline wander, powerline interference, electromyography (EMG) noise and other noise, which will have a certain impact on the normal signal. The frequency range of ECG signal is 0.05–100 Hz, and the key information is mainly concentrated in the range of 0.05–35 Hz. The review literature [3] shows that many researches used bandpass Butterworth filter with a frequency range of 1 $\sim$ 50 Hz for noise removal. And the main information of ECG signal was not lost. Therefore, a 3rd order bandpass Butterworth filter with a frequency range of 1 $\sim$ 50 Hz was employed to remove the noise. Afterwards, the PT algorithm was utilized to find the R-peak locations [40]. The first 100 sampling points (0.27 s) and the last 155 sampling points (0.43 s) of the R-peak were extracted, totaling 256 sampling points. In order to better classify the ECG signals, the data were processed using the Z-score normalization method.

3.3. Experimental settings

The proposed network was trained on a desktop computer with i7-6700 CPU, 16 GB RAM, and an Nvidia GTX1070Ti. Python as programming language, TensorFlow as programming tools for simulation implementation.

Table 2
The types and number of selected heartbeats.

Types	Training set	Test set
N	4000	4162
LB	3946	3552
RB	3777	2934
APC	807	1499
PVC	3682	2657
Total heartbeats	16,212	14,804

4. Results

4.1. Evaluation criterion

The evaluation indicators adopt the four metrics commonly used in the classification: Accuracy (*Acc*), Precision (*Pre*), Sensitivity (*Sen*), Specificity (*Spe*). The four variables including true positive (*TP*), true negative (*TN*), false negative (*FN*), false positive (*FP*), can be calculated for the required evaluation indicators.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{10}$$

$$Pre = \frac{TP}{TP + FP} \times 100\% \tag{11}$$

$$Sen = \frac{TP}{TP + FN} \times 100\% \tag{12}$$

$$Spe = \frac{TN}{TN + FP} \times 100\% \tag{13}$$

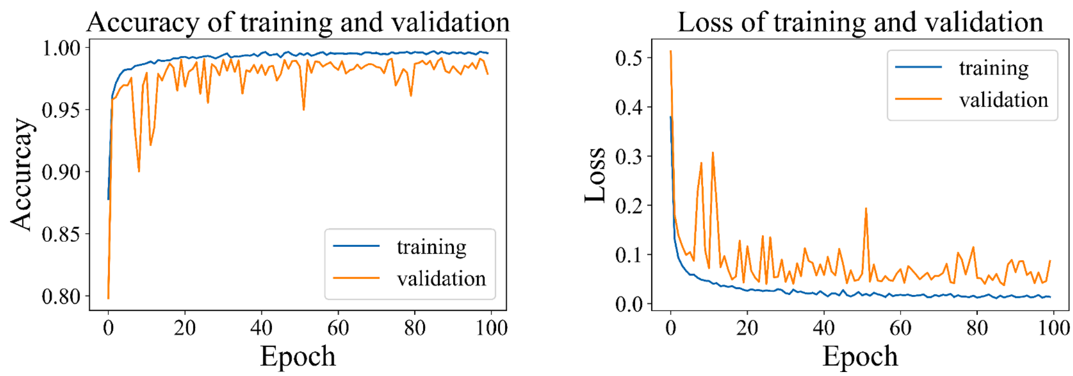
4.2. Results

Fig. 4 shows the accuracy and loss curves of the model training and validation. It can be seen from Fig. 4 that our model is a robust model with good validation accuracy and small loss after 100 iterations.

To more clearly demonstrate the effect of the time representation coefficient on the input, the time representation of the original signal is shown in Fig. 5. The five different types of raw heartbeats are graphed in Fig. 5 -a). It can be seen that the four types of heartbeats (N, LB, APC, and PVC) are close to each other near the peak of the QRS wave, and only the positions of the P-wave and *T*-wave are different. The difference between these four heartbeats and RB is significant, which means that RB and these four heartbeats can be visually distinguished from each other only by the wave. On the other hand, the N and APC heartbeats are not very distinguishable overall, making it difficult to make visual judgments. In the Fig. 5 -b), the different waveforms are distinguished to a greater extent by the transformation of the time representation coefficients. The two heartbeats N and PVC have large differences in overall amplitude from the other three heartbeats. The N and PVC beats are relatively close to each other in terms of R-peak position, but they are easily distinguishable in terms of P-wave and S-wave positions. The three categories of LB, RB, and APC heartbeats are well discriminated within a small range of amplitudes. It shows that the time representation input using RR interval has good discrimination and can reflect the different characteristics of different categories.

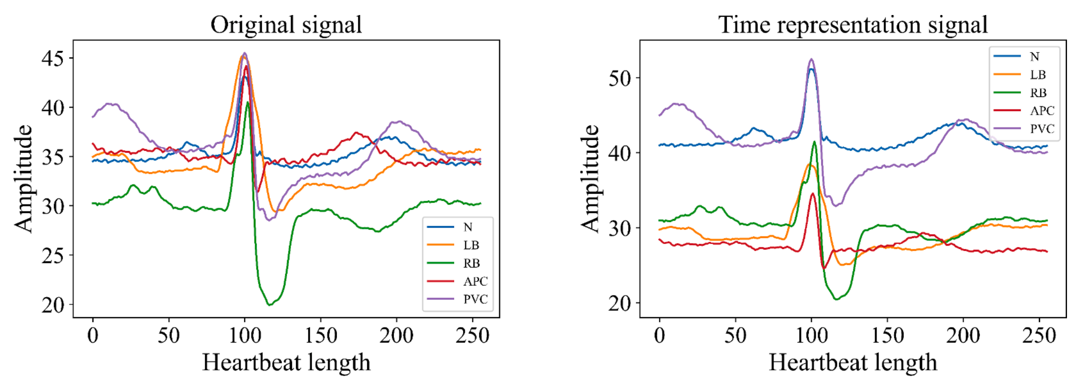
We performed a 5-fold cross validation treatment of the model, and Table 3 shows the best fold results of time representation input in 5-fold. The results obtain the highest average accuracy of 99.02 %. Moreover, the accuracy of the LB, RB, and PVC types reach more than 99.00 %, with the LB category reaching 99.78 %, while the accuracy of the N category is the worst at 98.08 %. In the precision indicator, the N class result is only 95.67 % and the APC class is 92.17 %; in the sensitivity indicator, the APC has the worst result, only 91.13 %. The main reason for this result is that some APC heartbeats are not distinguished from N heartbeats. There are 114 PVC heartbeats assigned to N class, which is the highest number of misclassified heartbeats. Moreover, 82 heartbeats in N type are classified to APC type. The interaction between the N and APC classes results in poor precision and sensitivity. This condition is most likely due to individual specificity, with different individual waveforms behave differently. Some individual PVC waveforms are closer to the N class of other individuals, which can lead to incorrect model judgments.

Table 4 shows the 5-fold average performance comparison of the



a) Accuracy of training and validation b) Loss of training and validation

Fig. 4. The accuracy and loss curves of the model.



a) Original signal b) Time representation signal

Fig. 5. Original signal time representation transformation.

Table 3
Results of time representation input model.

		Predicted label					*Acc*(%)	*Pre*(%)	*Sen*(%)	*Spe*(%)
		N	LB	RB	APC	PVC				
True label	N	4062	0	3	82	15	98.08	95.67	97.60	98.27
	LB	17	3528	0	1	6	99.78	99.77	99.32	99.93
	RB	15	0	2888	26	5	99.61	99.59	98.43	99.90
	APC	114	0	4	1366	15	98.32	92.17	91.13	99.13
	PVC	38	8	5	7	2599	99.33	98.45	97.82	99.66
Average	–	–	–	–	–	–	99.02	97.13	96.86	99.38

Table 4
Comparison of the results of different methods. The best result is in bold.

Methods	*Acc*(%)	*Pre*(%)	*Sen*(%)	*Spe*(%)	Training Time(min)
Original input	98.65	96.80	95.11	99.11	8.3
Feature Fusion-1	98.70	96.95	95.06	99.14	8.3
Feature Fusion-2	98.63	96.77	95.01	99.09	10.0
Proposed input	**98.95**	**97.07**	**96.54**	**99.38**	**8.3**

time representation input, the original input, and the two feature fusion methods in this paper. All results in the table are the averages of CNN-LSTM-Attention model with the 5-fold cross validation results. The Feature fusion-1 is a serial fusion of RR interval features with deep learning features of fully connected layer one; the Feature Fusion-2 is a parallel fusion of RR interval features with deep learning features of fully connected layer one. According to the original input performance in Table 4, both *Acc* and *Spe* results are above 98.00 %. This indicates that the current deep learning processing of the MIT-BIH database has achieved some achievements, but the improvement is more difficult. Therefore, different metrics with an improvement of more than 0.20 % are considered excellent. Compared with the previous three algorithms, the algorithm proposed in this paper achieves the best results in all metrics. It shows that the method proposed in this paper has greater effectiveness. In the *Acc* result, the proposed input is 98.95 %, which is 0.25 % higher than Feature Fusion-1. In the *Pre* and *Spe* metrics, the results of 97.07 and 99.38 are achieved, respectively. In the *Pre* metric, compared to the Feature Fusion-1 algorithm, the improvement of proposed input is 0.12 %; and compared to the original input, the improvement of proposed is 0.27 %. The *Spe* index of proposed input is improved by 0.24 % and 0.27 % compared with Feature Fusion-1 algorithm and the original input, respectively. From the perspective of improvement effect, the algorithm proposed has great advantages. In the *Sen* metric, this algorithm achieves the largest improvement, reaching 1.53 % compared to the *Sen* metric of Feature Fusion-2. Among the four metrics (*Acc*, *Pre*, *Sen*, *Spe*) of all methods, the value of *Sen* is low. The proposed input method achieves a huge improvement in this metric, which indicates the excellent performance. The training time of proposed method is 8.3 min, which is consistent with the training time of the original input. This shows that the time scaled signal has no effect on the training. So the proposed input has a high operating efficiency. Moreover, compared with the feature fusion methods (Feature Fusion-1 and Feature Fusion-2), the temporal representation input proposed in this paper is a concise feature representation by scaling the input at the very beginning stage. Like the normalization of data, it does not need to design a separate serial or parallel model. In contrast, the time representation input can not only better reflect the differences of different heart beats, realize the improvement of classification effect, and improve the algorithm running efficiency.

5. Discussions

In addition to analyzing all arrhythmias, we also focused on the results of two types of arrhythmias, APC and PVC, as shown in Table 5. These two types of arrhythmias are of greater concern to physicians and have important clinical value. In the APC anomaly classification results, the algorithm proposed in this paper achieves the best results on two metrics, 98.21 % *Acc* and 89.57 % *Sen*. The best results on the other two metrics are obtained by Feature Fusion-1, 95.86 % *Pre* and 99.60 % *Spe*. It can be seen that the proposed model in this paper does not achieve the best results in all indicators. However, compared to the original model, the proposed model improves more than 0.40 % in the *Acc* metric. In addition, it has the largest improvement among the *Sen* indicators, which is more than 6.28 % higher than original model. The results show that the proposed algorithm and Feature Fusion-1 algorithm have their own advantages in different metrics. Among the metrics for PVC anomaly classification, the proposed algorithm achieves the best results in all four indicators. Compared with original algorithms, Feature Fusion-1 and Feature Fusion-2 algorithms have decreased in all indicators, which indicates that feature fusion is not sensitive to this kind of anomaly. The algorithm proposed in this paper achieves 98.56 % and 97.63 % results in *Pre* and *Sen*. It is 0.28 % and 0.48 % higher than the original algorithm respectively. It can be concluded that this method performs well in the classification of PVC anomalies. Based on the comparison of APC and PVC classification results, the algorithm proposed in this paper has the best comprehensive effect and greater advantages.

Table 6 shows the 5-fold average impact of using different deep learning models on the four methods. As can be seen from the table, the addition of the attention mechanism improves the results for the original input, but the effect is limited. Only two of the four metrics in Feature Fusion-1 have improved. Feature fusion-2 with the addition of the attention mechanism has improved the results, with an accuracy improvement of 0.04 % and a maximum improvement of precision (0.11 %). After using the attention mechanism, the proposed time representation input in this paper improves all indicators than CNN-LSTM model. The precision is improved by 0.13 %, and sensitivity is improved the most with 0.68 %. These comparisons can show that the addition of attention mechanism has a better analysis and focus on the differences of time representation input. It also demonstrates that the CNN-LSTM-Attention mechanism is more suitable for the time representation input presented in this paper.

Table 6
Impact of different deep learning models on four methods. The best result is in bold.

Methods	Models	*Acc* (%)	*Pre* (%)	*Sen* (%)	*Spe* (%)
Original input	CNN-LSTM	98.60	96.91	94.86	99.06
	CNN-LSTM-Attention	98.65	96.80	95.11	99.11
Feature Fusion-1	CNN-LSTM	98.69	97.01	95.19	99.13
	CNN-LSTM-Attention	98.70	96.95	95.06	99.14
Feature Fusion-2	CNN-LSTM	98.59	96.66	94.80	99.07
	CNN-LSTM-Attention	98.63	96.77	95.01	99.09
Proposed input	CNN-LSTM	98.82	96.75	95.86	99.24
	CNN-LSTM-Attention	**98.95**	**97.07**	**96.54**	**99.38**

To more fully show the effectiveness of the time representation input method, a comparative analysis of existing studies is presented in Table 7. The selected literatures are classified using a patient-specific approach and MIT-BIH arrhythmia database. The results in the table

Table 7
Summary of previous studies on the arrhythmia classification.

Author, Year	Inputs or Features	Models	Classes	*Acc* (%)
Luo et al., 2017 [39]	DWT-images	CNN	4	97.50
Chen et al., 2017 [3]	RR intervals	SVM	15	98.46
Xiang et al., 2018 [10]	ECG signals + RR intervals	CNN-Attention	2	98.50
Oh et al., 2018 [34]	ECG signals	CNN-LSTM	5	98.10
Li et al., 2018 [41]	ECG + RR interval images	CNN	5	96.83
Shi et al., 2019 [42]	Region features	SVM + KNN + RF + DT	5	98.78
Oh et al., 2019 [43]	HOS + Hermite	SVM	5	98.71
Huang et al., 2019 [44]	STFT images	CNN	5	99.00
Rohmantri et al., 2020 [45]	ECG images	CNN	7	98.10
Niu et al., 2020 [46]	ECG signals + RR intervals	Multi-view CNN	2	96.4
Wang et al., 2021 [37]	ECG signals + Morphological features	DAE	5	93.87
Lu et al., 2021 [47]	ECG images	CNN-Focal loss	17	98.55
Yang et al., 2021 [48]	Local binary pattern features	RBF + RF	5	98.10
Nza et al., 2021 [49]	ECG images + entroy + CNN features	CNN + SVM	5	97.6
Feng et al., 2022 [50]	ECG signals	Unsupervised CNN	5	95.7
This study	**ECG signals + RR intervals**	**CNN-LSTM-Attention**	**5**	**98.95**

Table 5
Comparison of APC and PVC classification metrics. The best result is in bold.

Methods	APC				PVC			
	Acc(%)	*Pre*(%)	*Sen*(%)	*Spe*(%)	*Acc*(%)	*Pre*(%)	*Sen*(%)	*Spe*(%)
Original model	97.81	94.56	83.29	99.45	99.19	98.28	97.24	99.63
Feature Fusion-1	97.78	**95.86**	81.64	**99.60**	99.12	97.92	97.15	99.55
Feature Fusion-2	97.79	94.87	82.68	99.49	99.10	97.73	97.23	99.50
Proposed model	**98.21**	92.55	**89.57**	99.18	**99.32**	**98.56**	**97.63**	**99.69**

are those recorded in the literatures. As can be seen in Table 7, scholars have done various forms of transformations on the input of ECG in order to obtain a better representation of the input or features. Luo et al. [39], Li et al. [41], Huang et al. [44], Rohmantri et al. [45] and Lu et al. [47] all took ECG signals through certain transformations, and then converted them into images for classification. This approach can utilize a large number of excellent image processing models, but it will take up a lot of resources and take a long time. On the other hand, Xiang et al. [10], Wang et al. [37], Oh et al. [43] and Niu et al. [46] all used traditional manual features with ECG signal conversion features to construct a more expressive feature. Yang et al. [48] employed local binary pattern features of ECG signals for classification, and Feng et al. [50] designed an unsupervised CNN classification model. By comparison, the time representation input method proposed in this paper obtains an accuracy of 98.95 %, which is a more excellent score in the literatures. It can be shown that the time representation input we constructed better express the variability between different categories and provide a more distinct input to the classifier. This also provides a classification idea for periodic time series with similar structure, such as Phonocardiogram (PCG), Photoplethysmography (PPG), and so on.

Since the focus of this study is to establish distinctive characteristics, the type and number of arrhythmias have not been studied too much. Therefore, there are some limitations to our work. In this paper, the algorithm was currently validated using public data and the number of arrhythmia types was only five. Our approach was to construct a temporal representation input based on the RR interval. The different manifestations of arrhythmias in ECG were reflected in the RR interval. Therefore, it is possible to use this method for other arrhythmia classification. In this paper, these five categories were selected for classification because they are common and have a large number of arrhythmias. There are few other abnormalities in the MIT-BIH database. We will consider exploring more kinds of arrhythmia in our future work. And the private database collected from single-lead devices has not been tested. Therefore, our future work will focus on extending the algorithm to the analysis of more types of arrhythmias and validating it with a private database of single-lead device acquisitions. Due to the small number of APC anomaly types in the MIT-BIH database, it is easy to lead to insufficient generalization capability of the model. Therefore, subsequent work will explore ways to solve the data imbalance or consider using more datasets. In addition, the periodic physiological signals of the human body are similar, especially those related to the heart, such as phonocardiogram (PCG). Therefore, we will expand the method to these signals to see if it can help to analyze similar periodic physiological signals in future work.

6. Conclusions

In this paper, a time representation input was proposed based on the characteristics of ECG signals. The representation used the RR interval as a dynamic time scale for measuring the differences of various arrhythmias. This time representation coefficient was combined with the ECG input to obtain the time representation input. After the CNN-LSTM-Attention model processing, it was able to better extract the differences of heartbeats. The results show that the time representation input have better difference representation than the original input, which is a significant contribution to improve the classification effect. The simulation experiments revealed that the classification effect is still poor for a few categories. Since the focus of this study is to build more distinctive features, the types and number of arrhythmias have not been explored too much. Hence, future work will focus on different deep learning models, more arrhythmia types, and data imbalance problems to further improve the classification effect.

CRediT authorship contribution statement

Youhe Huang: Conceptualization, Methodology, Software, Writing – original draft. **Hongru Li:** Writing – review & editing, Supervision, Project administration, Data curation, Funding acquisition. **Xia Yu:** Writing – review & editing, Funding acquisition, Formal analysis, Validation.

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.

Data availability

The dataset is a public dataset and can be obtained free of charge

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61973067, 61903071).

References

[1] S.M.J. Dinakkarrao, A. Jantsch, M. Shafique, Computer-aided Arrhythmia Diagnosis with Bio-signal Processing: A Survey of Trends ans Techniques, ACM Comput. Surv. 52 (2) (2019) 1–37.
[2] Y. Hagiwara, H. Fujita, S.L. Oh, J.H. Tan, R.S. Tan, et al., Computer-aided diagnosis of atrial fibrillation based on ECG Signals: A review, Inf. Sci. 467 (2018) 99–114.
[3] S.K. Berkaya, A.K. Uysal, E.S. Gunal, S. Ergin, S. Gunal, et al., A survey on ECG analysis, Biomed. Signal Process. Control 43 (2018) 216–235.
[4] A. Rizwan, A. Zoha, I.B. Mabrouk, H.M. Sabbour and A.S. AI-Sumaiti, et al, A Review on the State of the Art in Atrial Fibrillation Detection Enabled by Machine Learning, IEEE Rev. Biomed. Eng., 14 (2021) 219–239.
[5] J. Mougoufan, J.S. Fouda, M. Tchuente, W. Koepf, Three-class ECG beat classification by ordinal entropies, Biomed. Signal Process. Control 67 (2021) 1–10.
[6] T. Radhakrishnan, J. Karhade, S.K. Ghosh, P.R. Muduli, R.K. Tripathy, et al., AFCNNet: Automated detection of AF using chirplet transform and deep convolutional bidirectional long short term memory network with ECG signals, Comput. Biol. Med. 137 (2021) 104783.
[7] Y. Jin, C. Qin, Y. Huang, W. zhao and C. Liu, Multi-domain modeling of atrial fibrillation detection with twin attentional convolutional long short-term memory neural networks, Knowl.-Based Syst., 193 (2020) 105460.
[8] L. Xie, Z. Li, Y. Zhou, Y. He, J. Zhou, Computational Diagnostic Techniques for Electrocardiogram Signal Analysis, Sensors 20 (21) (2020) 6318.
[9] H. Zhang, Z. Dong, M. Sun, H. Gu, Z. Wang, TP-CNN: A Detection Method for atrial fibrillation based on transposed projection signals with compressed sensed ECG, Comput. Methods Programs Biomed. 210 (2021) 106358.
[10] Y. Xiang, J. Luo, T. Zhu, S. Wang, X. Xiang, et al., ECG-Based Heartbeat Classification Using Two-Level Convolutional Neural Network and RR Interval Difference, Ieice Trans. Inform. Syst. E101D (4) (2018) 1189–1198.
[11] X. Wu, Y. Zheng, C.H. Chu, Z. He, Extracting deep features from short ECG signals for early atrial fibrillation detection, Artif. Intell. Med. 109 (2020), 101896.
[12] P. Plawiak, U.R. Acharya, Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals, Neural Comput. & Applic. 32 (15) (2019) 11137–11161.
[13] G. Chen, Z. Hong, Y. Guo, C. Pang, A cascaded classifier for multi-lead ECG based on feature fusion, Comput. Methods Programs Biomed. 178 (2019) 135–143.
[14] Y. Huang, H. Li, X. Yu, A multiview feature fusion model for heartbeat classification, Physiol. Meas. 42 (2021) 065003.
[15] M. Rizwan, B.M. Whitaker, D.V. Anderson, AF detection from ECG recordings using feature selection, sparse coding, and ensemble learning, Physiol. Meas. 39 (2018) 124007.
[16] U.R. Acharya, H. Fujita, M. Adam, S.L. Oh, V.K. Sudarshan, et al., Automated characterization and classification of coronary artery disease and myocardial infarction by decomposition of ECG signals: A comparative study, Inf. Sci. 377 (2017) 17–29.
[17] P. Plawiak, Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system, Expert Syst. Appl. 92 (2018) 334–349.
[18] Q. Yao, R. Wang, X. Fan, J. Liu, Y. Li, Multi-class Arrhythmia detection from 12-lead varied-length ECG using Attention-based Time-Incremental Convolutional Neural Network, Information Fusion. 53 (2020) 174–182.
[19] Z. Golrizkatami, S. Taheri, A. Acan, Multi-scale features for heartbeat classification using directed acyclic graph CNN, Appl. Artif. Intell. 32 (2018) 613–628.
[20] H. Zhang, C. Liu, Z. Zhang, Y. Xing, X. Liu, et al., Recurrence Plot-Based Approach for Cardiac Arrhythmia Classification Using Inception-ResNet-v2, Front. Physiol. 12 (2021) 648950.
[21] W. Lu, H. Hou, J. Chu, Feature fusion for imbalanced ECG data analysis, Biomed. Signal Process. Control 41 (2018) 152–160.

[22] Z. Dokur, T. Olmez, Heartbeat classification by using a convolutional neural network trained with Walsh functions, Neural Comput. & Applic. 32 (16) (2020) 12515–12534.
[23] U.R. Acharya, H. Fujita, S.L. Oh, Y. Hagiwara, J.H. Tan, et al., Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network, Inf. Sci. 405 (2017) 81–90.
[24] J. Zhang, J. Tian, Y. Cao, Y. Yang, X. Xu, Deep time-frequency representation and progressive decision fusion for ECG classification, Knowl.-Based Syst. 190 (2020) 105402.
[25] H. Fujita, D. Cimr, Computer Aided detection for fibrillations and flutters using deep convolutional neural network, Inf. Sci. 486 (2019) 231–239.
[26] M.S. Islam, N. Ammour, N. Alajlan, H. Aboalsamh, Rhythm-based heartbeat duration normalization for atrial fibrillation detection, Comput. Biol. Med. 72 (2016) 160–169.
[27] S. Mandal, P. Mondal, A.H. Roy, Detection of Ventricular Arrhythmia by using Heart rate variability signal and ECG beat image, Biomed. Signal Process. Control 68 (2021) 102692.
[28] S.M.P. Dinakarrao, A. Jantsch, and M. Shafique, Computer-aided Arrhythmia Diagnosis with Bio-signal Processing: A Survey of Trends and Techniques, 52 (2) 2019, 1–37.
[29] C. Ye, B.V.K.V. Kumar, M.T. Coimbra, Heartbeat Classification Using Morphological and Dynamic Features of ECG Signals, IEEE Trans. Biomed. Eng. 59 (10) (2012) 2930–2941.
[30] J. Rahul, M. Sora, L.D. Sharma, V.K. Bohat, An improved cardiac arrhythmia classification using an RR interval-based approach, Biocybernetics and Biomedical, Engineering 41 (2021) 656–666.
[31] F. Li, Y. Xu, Z. Chen, Z. Liu, Automated Heartbeat Classification Using 3-D Inputs Based on Convolutional Neural Network With Multi-Fields of View, IEEE Access 7 (2019) 76295–76304.
[32] U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, et al., A deep convolutional neural network model to classify heartbeats, Comput. Biol. Med. 89 (2017) 389–396.
[33] Y. Li, Y. Pang, J. Wang, X. Li, Patient-specific ECG classification by deeper CNN from generic to dedicated, Neurocomputing 314 (2018) 336–346.
[34] S.L. Oh, E.Y.K. Ng, R.S. Tan, U.R. Acharya, Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats, Comput. Biol. Med. 102 (2018) 278–287.
[35] J.H. Tan, Y. Hagiwara, W. Pang, I. Lim, S.L. Oh, et al., Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals, Comput. Biol. Med. 94 (2018) 19–26.
[36] A. Habib, C. Karmakar, J. Yearwood, Impact of ECG Dataset Diversity on Generalization of CNN Model for Detecting QRS Complex, IEEE Access 7 (2019) 93275–93285.
[37] J. Wang, R. Li, R. Li, B. Fu, C. Xiao, et al., Towards Interpretable Arrhythmia Classification With Human-Machine Collaborative Knowledge Representation, IEEE Trans. Biomed. Eng. 68 (7) (2021) 2098–2109.
[38] P.D. Chazal, M. O'Dwyer, R.B. Reilly, Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features, IEEE Trans. Biomed. Eng. 51 (7) (2004) 1196–1206.
[39] K. Luo, J. Li, Z. Wang, A. Cushieri, Patient-Specific Deep Architectural Model for ECG Classification, J. Healthcare Eng. 2017 (2017) 4108720.
[40] J. Pan, W.J. Tompkins, A Real-Time QRS Detection Algorithm, IEEE Trans. Biomed. Eng. 32 (3) (1997) 230–236.
[41] S. Chen, W. Hua, Z. Li, J. Li, X. Gao, Heartbeat classification using projected and dynamic features of ECG signal, Biomed. Signal Process. Control 31 (2017) 165–173.
[42] H. Shi, H. Wang, F. Zhang, Y. Huang, L. Zhao, et al., Inter-patient heartbeat classification based on region feature extraction and ensemble classifier, Biomed. Signal Process. Control 51 (2019) 97–105.
[43] S.L. Oh, E.Y. Ng, R.S. Tan, U.R. Acharya, Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types, Comput. Biol. Med. 105 (2019) 92–101.
[44] J. Huang, B. Chen, B. Yao, W. He, ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network, IEEE Access 7 (2019) 92871–92880.
[45] R. Rohmantri, N. Surantha, Arrhythmia Classification using 2D Convolutional Neural Network, Int. J. Adv. Comput. Sci. Appl. 11 (4) (2020) 201–208.
[46] J. Niu, Y. Tang, Z. Sun, W. Zhang, Inter-Patient ECG Classification with Symbolic Representations and Multi-Perspective Convolutional Neural Networks, IEEE J. Biomed. Health Inform. 24 (5) (2020) 1321–1332.
[47] Y. Lu, M. Jiang, L. Wei, J. Zhang, Z. Wang, et al., Automated arrhythmia classification using depthwise separable convolutional neural network with focal loss, Biomed. Signal Process. Control 69 (2021) 102843.
[48] P. Yang, D. Wang, W.B. Zhao, L.H. Fu, J.L. Du, et al., Ensemble of kernel extreme learning machine based random forest classifiers for automatic heartbeat classification, Biomed. Signal Process. Control 63 (2021) 102138.
[49] M. Naz, J.H. Shah, M.A. Khan, M. Sharif and M. Raza, et al., From ECG signals to images: a transformation based approach for deep learning, PeerJ computer science. (2021) e386.
[50] P. Feng, J. Fu, Z. Ge, H. Wang, Y. Zhou, et al., Unsupervised semantic-aware adaptive feature fusion network for arrhythmia detection, Inf. Sci. 582 (2022) 509–528.