SUMMARY This paper presents an automated patient-specific ECG classification algorithm, which integrates long short-term memory (LSTM) and convolutional neural networks (CNN). While LSTM extracts the temporal features, such as the heart rate variance (HRV) and beat-to-beat correlation from sequential heartbeats, CNN captures detailed morphological characteristics of the current heartbeat. To further improve the classification performance, adaptive segmentation and re-sampling are applied to align the heartbeats of different patients with various heart rates. In addition, a novel clustering method is proposed to identify the most representative patterns from the common training data. Evaluated on the MIT-BIH arrhythmia database, our algorithm shows the superior accuracy for both ventricular ectopic beats (VEB) and supraventricular ectopic beats (SVEB) recognition. In particular, the sensitivity and positive predictive rate for SVEB increase by more than 8.2% and 8.8%, respectively, compared with the prior works. Since our patient-specific classification does not require manual feature extraction, it is potentially applicable to embedded devices for automatic and accurate arrhythmia monitoring.

key words: convolutional neural network (CNN), long short-term memory (LSTM), patient-specific ECG classification

1. Introduction

Cardiovascular arrhythmia refers to any abnormal changes from electrical impulses in heart, which is a common disease that may occur suddenly and become life-threatening [1]. Electrocardiogram (ECG) signal that represents the electrical activities of heart depolarization and repolarization has been widely used in clinic to distinguish arrhythmia, owing to its low cost and non-invasiveness. However, manual arrhythmia monitoring is nearly impossible for ordinary people with little professional medical knowledge in daily life. An automated arrhythmia monitoring method with high immediacy and accuracy is therefore preferred.

Although the automated arrhythmia monitoring has a broad application prospect in the clinic, it is particularly difficult to be achieved for multiple reasons: i) morphological characteristics of ECG signals produced by the same symptoms exhibit significant dissimilarities among different patients under various circumstances [2]; ii) the imbalanced heartbeat classes of ECG signals further exacerbate the problem; iii) external noises, such as baseline wander and muscle artifact, also affect the performance of automatic classification systems [3].

Traditional arrhythmia recognition methods usually overcome these difficulties with two independent steps: feature extraction and classification. Time-domain ECG signals are converted into fixed-length distinctive feature vectors to better reveal the inherent property of morphology during the extraction step. Hermite coefficients [4], heart-beat interval [5]–[9], morphological characteristic [7], [10], [11], and features extracted by principal component analysis (PCA) [8], [12], independent component analysis (ICA) [6], [8], [9], discrete Fourier transform (DFT) [13], higher order statistics (HOS) [4], wavelet transform (WT) [6], [12], [14], [15] are independent or combined to represent ECG signals more effectively. To accomplish the final recognition, the combinations of feature vectors are processed by classifiers, such as linear discriminants (LD) [10], extreme learning machine (ELM) [5], random forests (RF) [14], type-2 fuzzy c-means (FCM) [15], optimum-path forest (OPF) [8], k-nearest neighbor (KNN) [13], [16], support vector machine (SVM) [4], [6], [13], [16] and artificial neural network (ANN) [11], [12], [15]. The extraction step refines more effective features from ECG signals through different transformations, leading to a better performance of arrhythmia monitoring.

However, the connotative characteristics underlying original ECG signals can be easily ignored during the manual feature extraction process if there are no appropriate human interference and sufficient priori knowledge [17]. Besides, since the traditional handcrafted features are merely extracted by simple transformation, more abundant inner information is needed to further improve the arrhythmia recognition performance. To solve these problems of traditional methods, the end-to-end ECG classification algorithms using convolutional neural network (CNN) or recurrent neural network (RNN) are proposed [18]–[20]. These algorithms fuse feature extraction and classification into an efficient single learning model with hierarchical network architectures. The automatic feature extraction reduces the demand of priori knowledge and covers more potential characteristics from input ECG signals, thus leading to better performance and generalization properties of classification.
algorithms.

Based on these neural network related methods, an end-to-end patient-specific ECG classification algorithm is proposed in this study by combining the bi-directional long short-term memory (BLSTM)[21] and CNN. The advantages of both network topologies are leveraged to enhance the accuracy and stability of automated arrhythmia detection, which is crucial in clinical practice. The main contributions of this paper are as follows:

1) LSTM and CNN are used in conjunction to extract both the sequence correlation and detailed morphological features of heartbeats. The state-of-the-art network architecture achieves excellent performance on patient-specific ECG classification.

2) The adaptive segmentation and re-sampling are adopted to align the heartbeats of patients with various heart rates. Heartbeats are accurately extracted from the ECG signals by these methods, and the normalization of network input format is vital to improve classification accuracy.

3) A novel clustering strategy is proposed to identify the most representative patterns from the common data. A more balanced distribution of training set is obtained through this technique.

When evaluated on the MIT-BIH arrhythmia benchmark, the proposed ECG classification algorithm shows the superior performance of recognizing both the supraventricular ectopic beats (SVEB or type-S) and ventricular ectopic beats (VEB or type-V). In particular, the sensitivity and positive predictive rate for the SVEB are improved by more than 8.2% and 8.8% over the prior algorithms, respectively. Meanwhile, the results of the VEB detection are also of the top ranks.

The rest of this paper is organized as follows. Section 2 reviews the related works mainly about the neural network based ECG classification. Section 3 presents the background knowledge about the ECG database and signal morphology. The ECG signal processing and classification strategies are described in Sect. 4. The evaluation indicators of ECG classification are introduced in Sect. 5. The performance of the proposed algorithm is tested in Sect. 6. Section 7 discusses the experimental results. Section 8 draws the conclusion of this paper.

2. Related Works

Recent ECG classification mainly focuses on neural network based approaches. In this section, some related prior arts are reviewed. According to the innovation points of improving classification performance, these methods are divided into two groups: i) methods with optimized feature extraction and ii) methods with novel network topologies. The proposed ECG classification method in this paper falls into the two categories at the same time.

2.1 Methods with Optimized Feature Extraction

Although the neural network methods can extract inner features from original ECG signals automatically, researchers believe that appropriate intervention of this process is beneficial to the performance improvement. Xu et al. [22] proposed a novel strategy for uniform heartbeat segmentation and alignment. The aligned neural network input heartbeats have the same phase shift in ECG signals, which helps to reduce the unnecessary variability in the feature vectors and is vital to the high classification performance. Zhai et al. [19] extend the ECG signal input to a 2D coupling matrix that captures both the beat morphology and beat-to-beat correlation. The features of adjacent beats are intensified to characterize the crucial anomaly beats in their works. Time-frequency analysis is also adopted to process ECG signals for improving performance [23], [24], as it can provide joint distribution features of both time domain and frequency domain. The wavelet transform is introduced in [23] to decompose ECG signals into features at different frequency sub-bands. Furthermore, the research works in [24] verify the effectiveness of ECG feature extraction by different time-frequency analysis methods, including short-time Fourier transform (STFT), continuous wavelet transform (CWT), and pseudo Wigner-Ville distribution (PWVD). Unsupervised feature extraction technique is another effective method to obtain features of high quality [25], [26]. The input ECG signals are reconstructed by auto-encoders via unsupervised learning, and the encoder output are directly used as features of classifiers.

Since the optimized feature extraction plays an important role in improving ECG classification performance, the adaptive segmentation and novel clustering strategy are proposed in our design to obtain representative features. In contrast to [22], our adaptive segmentation and re-sampling methods need not to drop sampling points or pad zeros, thus maintaining better integrity of single heartbeats. Besides, our 1D network inputs are more computation-friendly than the 2D ones in [19]. When compared with [23]–[26], no time-frequency analysis and auto-encoders are introduced in our work, which dramatically decreases the complexity of whole system.

2.2 Methods with Novel Network Topologies

Except for the feature extraction, neural network topology is also an important factor that affects the classification performance. CNN is a common means to achieve accurate ECG classification, due to its excellent ability of dealing morphological features. Kiranyaz et al. [18] adopts a 1D CNN algorithm, which provides a reliable method for real-time ECG monitoring on embedded devices. Subsequently, the carefully designed CNN architectures, such as atrous spatial pyramid pooling (ASPP) [27], are applied on the field of ECG classification for better performance. In addition, much deeper convolutional neural network models [28], [29] are designed to classify heartbeats. These deeper models achieve much higher classification accuracy by increasing neural network resources, which have the huge potential to be introduced into clinical settings as an
adjunct tool for cardiologists. Meanwhile, since LSTMs are superior in handling sequential tasks, they are proposed to deal with the ECG classification problems as well. Typically, Saadatnejad et al. [20] implements a lightweight ECG classification model by combining the wavelet transform and multiple LSTMs. The temporal dependencies existing in the heartbeat waveforms are effectively extracted by the LSTMs. Besides, the spiking neural network (SNN) is applied to recognize arrhythmia as an attempting exploration [30]. However, the superiority of SNN mainly reflects in lower resource consumption, and the classification accuracy is not quite satisfactory.

Based on these prior works, our ECG classification method achieves a nice exploration in network topology. Owing to the effectiveness of integrated LSTM and CNN, the proposed model acquires better classification performance than [18], [20], and [30] under the same evaluation indicators. When compare with [28] and [29], our model is quite lightweight and more suitable for the usage of portable arrhythmia monitoring devices.

The most related works to our method are [18], [19], and [20] in which the patient-specific classification strategy is adopted. The patient-specific strategy aims to train a personal neural network model by the relatively small common training data from all patients and the specific data collected from the corresponding patient. The patient-specific means are gaining a lot of attention for they can extract the best possible features of individuals by exploiting relatively simple network architectures. In the following discussion, we will concentrate more on those methods using patient-specific strategy.

3. Background Knowledge

3.1 Database

MIT-BIH arrhythmia database [31], which is the most recognized open source ECG-based library, is selected to be the benchmark for evaluating the heartbeat classification performance in our study. The MIT-BIH arrhythmia database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects (record 201 and 202 are from the same male subject). Continuous ECG signals are band-pass filtered at 0.1-100 Hz and digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. The R peak locations and beat class information of all recordings are independently annotated by two or more cardiologists for the computer-readable reference. Each heartbeat is identified into one of the five beat classes (normal beat (N or type-N), SVEB, VEB, fusion of a ventricular and a normal beat (F or type-F), or unknown beat type (Q or type-Q)) recommended by the AAMI standard [32].

In our study, the 44 records from the MIT-BIH arrhythmia database are used, excluding 4 records (102, 104, 107, and 217) which contain paced heartbeats. Based on the characteristics and symptoms of the subjects, the 44 records can be divided into two groups. The dataset one (DS1) contains 20 records (100 to 124) that cover representative samples of routine clinical recordings and serves as the common training dataset. The dataset two (DS2) consists of the remaining 24 records (200 to 234), which includes uncommon but clinically significant arrhythmias and works as the patient-specific dataset [33]. Following the AAMI standard [32], no more than 5 minutes of each recording in DS2 can be used for classifier training purpose. Thus in our design, the common training data from DS1 and the first 5 minutes of recordings in the patient-specific dataset DS2 are used to train the network for each patient independently. In compliance with common practice [12], [18], [19], [33], the remaining data of each recording in DS2 is used to test the performance of our classification algorithm.

3.2 ECG Signal

One typical heartbeat in ECG wave is described in Fig. 1. The most important clinical details are the P, Q, R, S, and T waves that represent the electrical activities of cardiac muscle [34]. The P wave represents the Atria depolarization (Atrial Contraction), the T wave shows the depolarization of ventricles (Ventricular relaxation), and the QRS complex represents ventricles depolarization (Ventricular contraction) [35].

The arrhythmia usually shows various disease symptoms on ECG signals. As is shown in Fig. 2, the QRS wave of S beats is almost identical to that of N beats in pattern. The difference between N beats and S Beats in morphology is that the P wave is missing in S beats, which leads to a shorter RR interval between the S beat and its previous beat. Both the detailed morphological and sequential RR interval characteristics are quite valuable to recognize the S beats accurately. The signal distortion of V beats mainly appears on the QRS wave, which is the most significant feature to detect V beats.

4. Methodologies

In this section, the details of the proposed patient-specific ECG classification algorithm are presented. First, the processing of ECG signals for removing external noises and reducing classification complexity is described. Then, the novel clustering strategy for selecting the most representative patterns from common training data is introduced. Subsequently, the neural network architecture with integrated
BLSTM and CNN is presented to explain the effectiveness of improving ECG recognition performance.

4.1 ECG Signal Processing

External noises usually cause the distortion of ECG signals and hamper the correct detection of clinical details. To solve this issue, baseline wander that severely affects the utility of recorded ECGs is first removed by subtracting the average value of the original signals from each point in both channels [36]. This approach guarantees that the processed ECG signals are averaged to zero and distributed on similar locations of all recordings. The spatially invariant inputs of ECG signals lead to a better classification performance [37].

The ECG signals after baseline wander removal need to be segmented into separated heartbeats for classification, according to the average interval between contiguous R peaks of each recording. However, the average intervals between adjacent R peaks can be quite different for individual patients. For example, there are around 90 heartbeats per minute in recording 212, but only about 50 heartbeats per minute exist in recording 231. Thus, the adaptive segmentation and re-sampling methods are proposed to reduce the negative effect from the varying heart rates among individuals.

The process of segmentation is described in Fig. 3. The R peak locations annotated in the MIT-BIH arrhythmia database are adopted, which can also be accurately determined using multiple algorithms, such as Pan-Tompkins in [38]. For the BLSTM topology in our network, one input heartbeat comprises the sample points of 1/3 of the average RR interval before the R peak and 2/3 of the average RR interval after the R peak. The imbalanced periods before and after the R peak are because the process of ventricular repolarization (causing T waves) is longer than that of sinoatrial node depolarization (causing P waves) [34]. For the CNN topology in our network, one input heartbeat comprises the sample points of 4/5 of the average RR interval both before and after the R peak. The extra sample points are used to enrich the details of P waves and T waves, which contributes to distinguishing N and SVEB. According to the background in [34], the durations of P, QRS, and T waves are usually 0.08s, 0.04s to 0.12s, and 0.16s in heartbeats. Thus, the 1/3 and 2/3 are selected in proportion to the durations of P and T waves roughly to keep the completion of one heartbeat. Considering the heartbeat duration usually lasts no shorter than 0.6s, the 4/5 keeps the QRS waves of adjacent heartbeats not appear in the sample points of current heartbeat.

In Fig. 3, $i$ indexes the sample points, $\nu(i)$ is the voltage of the ECG signal at time $i$, $t$ indexes the R peaks, $R_t$ represents the locations of R peaks, $\lfloor x \rfloor$ means the integer (floor) of $x$, and $D$ is the average RR interval of each recording (for DS2, $D$ is the average RR interval in the first five minutes).

For the BLSTM, the $m$-th sample point of the $t$-th heartbeat $H^B_t(m)$ is represented by Eq. (1),

$$H^B_t(m) = v \left( R_t + m - \left\lfloor \frac{1}{3} D \right\rfloor \right), m \in \left[ 0, \left\lfloor \frac{1}{3} D \right\rfloor + \left\lfloor \frac{2}{3} D \right\rfloor \right]$$

(1)

For the CNN, the $n$-th sample point of the $t$-th heartbeat $H^C_t(n)$ is represented by Eq. (2),

$$H^C_t(n) = v \left( R_t + n - \left\lfloor \frac{4}{5} D \right\rfloor \right), n \in \left[ 0, \left\lfloor \frac{4}{5} D \right\rfloor + \left\lfloor \frac{4}{5} D \right\rfloor \right]$$

(2)

The input lengths of our network are fixed for all patients. Taking into account both performance and complexity, the segmented ECG fragments are resampled to 300 points for the BLSTM and 100 points for the CNN by the linear interpolation. The $k$-th point $V^A(k)$ after re-sampling can be calculated by Eqs. (3) and (4),

$$V^A(k) = (V^B(j + 1) - V^B(j)) \ast \left( k \frac{J - 1}{K - 1} - j \right) + V^B(j)$$

(3)
\[ j \leq k \frac{J - 1}{K - 1} \leq j + 1 \tag{4} \]

Where the \( V^j_B(j) \) is the \( j \)-th point before re-sampling, and the \( J \) and \( K \) are the numbers of sample points before and after re-sampling. The indexes \( j \) and \( k \) satisfy

\[ 0 \leq j \leq J - 1 \tag{5} \]
\[ 0 \leq k \leq K - 1 \tag{6} \]

The adaptive segmentation and re-sampling normalize the network inputs of patients with varied heart rates, which contributes to the improvement of ECG classification.

### 4.2 Clustering Strategy

In compliance with common practice\cite{18}, \cite{19}, \cite{33}, a small part of heartbeats (75 type-N, type-S, type-V beats, and all (13) type-F, (7) type-Q) are selected from the DS1 as representatives to ensure that the features obtained from the common and patient-specific training data are comparable.

The representative type-S and type-V are hard to select, as they usually exhibit changeable characteristics and distribute unevenly among the recordings in DS1. As Table 1 shows, record 106 has 518 type-V, while record 100 has only 1. However, most of the type-V in record 106 are similar in morphology, and only the unique ones in different recordings are more informative in the training process.

Therefore, we propose a novel clustering approach to select the most representative type-S and type-V beats, in order to cover as many patterns as possible and balance their contributions. The average number of type-S or type-V to be selected in each recording (only for those containing type-S or type-V) is first calculated. For the recordings with a below-average number of type-S or type-V, all the heartbeats are added into the representative training dataset. The recordings having an above-average number of type-S or type-V are divided into clusters according to the morphology by the K-Means\cite{39}, and one heartbeat is randomly selected from each cluster for training. The number of clusters is roughly determined by the number of type-S or type-V in each recording, as Table 1 describes. In particular, the clusters of type-V are distributed more evenly to balance the contribution of recordings that have a large difference in the quantity of VEB. The proposed clustering method overcomes the occasionality of random selection, as described in \cite{18} and provides a more smooth approach to choose representative heartbeats than the independent CNN classifier in \cite{19}.

### 4.3 Integrated BLSTM and CNN

Since the BLSTM and CNN topologies are superior in handling sequential tasks and extracting detailed morphological features, respectively, they are integrated to classify the segmented ECG signals. The mini-batch gradient descent (MBGD) is adopted for the network training with a batch size of 16. The macro-architecture of the proposed classifier is described in Fig. 4.

The main model of the proposed network is divided into two separate parts named model \( \alpha \) and model \( \beta \). The model \( \alpha \) is composed of two sub-models (\( \alpha_1 \) and \( \alpha_2 \)), and each of them deals with the single channel ECG inputs by using two cascaded BLSTM layers. The current heartbeat, previous 4 heartbeats, and following 4 heartbeats make up the input steps in a time frame to extract the beat-to-beat correlation from context by utilizing the superiority of BLSTM in sequential tasks. The model \( \beta \) consists of three cascaded convolutional (CONV) layers that are followed by the rectified linear unit (ReLU) and max-pooling. The segmented heartbeats of both channels are concatenated as the CNN inputs. The otherwise ignored morphological features are extracted more effectively with CNN than BLSTM. The en...

![Fig. 4](image)

**Fig. 4** The macro-architecture of the proposed classifier. The BLSTM and CNN are integrated in parallel with different input formats.
The performance of ECG classification can be assessed using several measures. Since our network divides heartbeats into five classes following the AAMI standard [32], the confusion matrix of results is shown in Table 3 according to [10].

Four typical statistical indicators are applied, including the classification accuracy (Acc), sensitivity (Sen), specificity (Spe), and positive predictive rate (Ppr). The respective definitions of these four indicators use the true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which can be calculated as follows.

\[ TN_v = Nn + Ns + Nf + Nq + Sn + Ss + Sf + Sq + Fn + Fs + Ff + Fq + Qn + Qs + Qf + Qq \]  

(7)

\[ TN_s = Nn + Nu + Nf + Nq + Vn + Vv + Vf + Vq + Fn + Fv + Ff + Fq + Qn + Qv + Qf + Qq \]  

(8)

\[ FN_v = Vn + Vs + Vf + Vq \]  

(9)

\[ FN_s = Sn + Sv + Sf + Sq \]  

(10)

\[ TP_v = Vv \]  

(11)

\[ TP_s = Ss \]  

(12)

\[ FP_v = Nv + Sv + Fv + Qv \]  

(13)

\[ FP_s = Ns + Vs + Fs + Qs \]  

(14)

The Acc is the ratio of correctly classified patterns to the total patterns classified, Sen is the rate of correctly classified events among all events, Spe is the rate of correctly classified nonevents among all nonevents, and Ppr is the rate of correctly classified events in all detected events. They are defined as follows.

\[ Acc = \frac{TP + TN}{TP + TN + FP + FN} \]  

(15)

\[ Sen = \frac{TP}{TP + FN} \]  

(16)

\[ Spe = \frac{TN}{TN + FP} \]  

(17)

\[ Ppr = \frac{TP}{TP + FP} \]  

(18)

As N beats occupy the majority of the testing dataset, the Acc and Spe are strongly influenced by the accuracy of the type-N classification. Therefore, the Sen and Ppr are more convincing indicators in evaluating the classification performance of SVEB and VEB. Two widely used statistical measures of F1 score and G score are adopted to synthetically estimate the Sen and Ppr. The F1 score and G score represent the harmonic mean and geometric mean of the Sen and Ppr, respectively, which are defined by the following equations.

\[ F1 = \frac{2}{\frac{1}{Sen} + \frac{1}{Ppr}} \]  

(19)

\[ G = \sqrt{Sen \times Ppr} \]  

(20)

6. Results

The performance of the proposed patient-specific ECG classification algorithm is evaluated on the MIT-BIH arrhythmia database, and the results are compared with prior works.

6.1 Classification Results

Table 4 presents the confusion matrix of the patient-specific

| Layer Name | Output Size | Dropout Keep Prob |
|------------|-------------|------------------|
| CNN Input  | [200x1]     | -                |
| CONV L0    | [200x128]   | 16 SAME 1        |
| POOL L0    | [50x128]    | 4 SAME 4         |
| CONV L1    | [50x128]    | 16 SAME 1        |
| POOL L1    | [13x128]    | 4 SAME 4         |
| CONV L2    | [13x128]    | 16 SAME 1        |
| POOL L2    | [4x128]     | 4 SAME 4         |

| Layer Name | Output Size | Dropout Keep Prob |
|------------|-------------|------------------|
| BLSTM Input | [9x300] | -                |
| LSTM FW L0 | [9x600] | 0.1              |
| LSTM BW L0 | [9x600] | 0.1              |
| LSTM FW L1 | [9x600] | 0.1              |
| LSTM BW L1 | [9x600] | 0.1              |

| Layer Name | Input Size | Output Size |
|------------|------------|-------------|
| MLP        | [9*600+512] | [5]         |

Table 3 Confusion matrix definition.

| Algorithm Label | n | s | v | f | q |
|-----------------|---|---|---|---|---|
| N               | Nn|Ns|Nv|Nf|Nq|
| S               | Sn|Sv|Sf|Sq|
| V               | Vn|Vs|Vf|Vq|
| F               | Fn|Fs|Ff|Fq|
| Q               | Qn|Qs|Qv|Qf|Qq|
ECG classification results for DS2 in our study. In accordance with the AAMI standard [32], the performance on the SVEB and VEB is rather important for they can have dramatic consequences in the clinic. Approximately 85% of SVEB and 95% of VEB are correctly recognized, which confirms the effectiveness of the proposed algorithm. The Sen of SVEB is lower than that of VEB due to the lack of training patterns both in the DS1 and the first five minutes of DS2 recordings. Besides, the SVEB are prone to be misclassified as N because of their similarity on the QRS complex. In spite of the deficiencies, the performance on the SVEB is still satisfactory owing to the good feature extraction ability of our proposed networks. In addition, the novel clustering approach takes advantage of the existing patterns in the common training dataset as well.

The detailed classification results of each recording are listed in Table 5. For the VEB classification, the proposed method achieves a Sen of more than 90% on approximately 70% of the recordings and a Ppr of more than 90% on approximately 60% of the recordings (except for the recordings with no VEB). For the SVEB classification, this approach acquires a Sen of more than 80% on approximately 50% of the recordings and a Ppr of more than 80% on approximately 40% of the recordings (except for the recordings with no SVEB).

While the accuracy is satisfactory for most of the recordings, the remaining ones (such as 200, 202, 228 for SVEB and 201, 202, 234 for VEB) are not classified so well for several reasons. Firstly, the poor patient-specific patterns of SVEB and VEB largely affect the performance of networks. For example, the recording 234 has no VEB and the recording 202 has no SVEB in the first five minutes, which means that the patient-specific characteristics of these heartbeat classes cannot be accessed by networks. Using the common training dataset alone is not effective enough to detect the VEB and SVEB. Secondly, the similarity between SVEB and type-N in some recordings causes misclassification. Figure 2 shows the heartbeat morphology of both type-N and SVEB. The main difference between them is that the SVEB miss P wave before R peak. Since the P wave is not conspicuous in the waveform, the recognition of SVEB becomes extremely challenging.

### 6.2 Comparison with Others

A performance comparison with prior arts [12], [18], [19] is listed in Table 6. The results of the SVEB and VEB classification are presented in detail as they are more attainable in the clinic according to the AAMI guidance [32]. The DS2 is partitioned into two evaluation datasets. The dataset (a) contains 11 test recordings (200, 202, 210, 213, 214, 219, 221, 228, 231, 233, and 234) for VEB detection and 14 test recordings (with the addition of 212, 222, and 232) for SVEB detection. Dataset (b) is made up of all 24 recordings in DS2 (200 and onwards).
For the dataset (a), high Sen and Ppr of SVEB classification are obtained at the same time with the proposed algorithm, which dramatically increases F1 and G by 6.8 and 6.5, respectively. The performance of VEB classification is almost equal to the best one from Zhai [19]. For the dataset (b), both the SVEB and VEB classification performance of our method is the best among listed works. In particular, the Sen and Ppr for SVEB increase by more than 8.2% and 8.8%, respectively. The high recognition accuracy of SVEB dramatically improves the reliability of end-to-end patient-specific ECG classification.

7. Discussion

7.1 Common Training Data Selection

To evaluate the efficiencies of different common training data selection methodologies, i) randomly selecting method in [18], ii) dataset clustering method, and iii) the proposed recording clustering method, are applied to our networks. The difference between ii) and iii) is that ii) applies clustering on all SVEB or VEB in the common training dataset instead of on the separated recordings. The classification results based on DS2 are shown in Fig. 5. With our recording clustering method iii), the SVEB and VEB elevates the F1 by more than 3.1 and 1.4, respectively. The proposed recording clustering strategy proves to be a better choice according to the experimental results, since it not only selects the representative heartbeats but also balances the contribution of recordings.

7.2 Combining BLSTM and CNN

The BLSTM and CNN are separately employed to validate the proposed network architecture. As Fig. 6 describes, the performance deteriorates significantly when adopting either the independent BLSTM or CNN. The BLSTM is not capable enough of extracting detailed morphological features from the P and T waves, (where the main difference between N and SVEB lies), introducing huge accuracy loss in the SVEB classification. The VEB detection is also not well supported by the independent BLSTM. For the CNN only scenario, the recognition of VEB and SVEB is affected...
Fig. 7 The performance with one entire BLSTM model and two parallel BLSTM sub-models.

Table 7 Classification performance with different BLSTM sequence lengths.

| Sequence Length | VEB | SVEB |
|-----------------|-----|------|
|                 | Acc | Sen  | Spe | Ppr | F1  | G   | Acc | Sen  | Spe | Ppr | F1  | G   |
| 3               | 98.5 | 92.7 | 99.1 | 92.4 | 92.4 | 97.9 | 83.4 | 98.6 | 74.5 | 78.7 | 78.8 |
| 5               | 97.8 | 91.5 | 98.5 | 87.4 | 89.4 | 98.0 | 77.0 | 99.1 | 81.1 | 79.0 | 79.0 |
| 7               | 98.3 | 91.5 | 99.0 | 91.0 | 91.3 | 98.1 | 83.8 | 98.8 | 76.8 | 80.2 | 80.5 |
| 9               | 98.9 | 94.9 | 99.3 | 94.0 | 94.4 | 98.5 | 85.0 | 99.1 | 82.8 | 83.9 | 83.9 |
| 11              | 98.4 | 91.8 | 99.1 | 91.6 | 91.7 | 98.0 | 83.4 | 98.7 | 76.7 | 79.9 | 80.0 |

Table 8 Classification performance with different network scales.

| Re-sampling Size | VEB | SVEB |
|------------------|-----|------|
|                  | Acc | Sen  | Spe | Ppr | F1  | G   | Acc | Sen  | Spe | Ppr | F1  | G   |
| 300 100          | 98.9 | 94.9 | 99.3 | 94.0 | 94.4 | 98.5 | 85.0 | 99.1 | 82.8 | 83.9 | 83.9 |
| 200 100          | 98.1 | 92.8 | 98.7 | 88.2 | 90.4 | 90.4 | 97.9 | 81.2 | 98.7 | 75.5 | 73.2 |
| 400 100          | 98.3 | 91.9 | 99.0 | 90.6 | 91.2 | 91.2 | 97.8 | 83.9 | 98.5 | 73.6 | 78.4 |
| 300 50           | 98.4 | 95.8 | 98.7 | 89.0 | 92.2 | 92.3 | 97.4 | 79.3 | 98.3 | 70.1 | 74.4 |
| 300 150          | 98.1 | 92.8 | 98.6 | 87.9 | 90.3 | 90.3 | 97.1 | 83.1 | 97.8 | 64.6 | 72.7 |

significantly due to the absence of sufficient sequential features.

7.3 BLSTM Topology of Parallel Sub-Models

The proposed BLSTM network topology with two parallel sub-models is compared with the architecture with one unified model. The heartbeat sequences from both ECG channels are concatenated to form the inputs of the unified model, which doubles the network scale than a single sub-model. The experimental results of all recordings in DS2 are described in Fig. 7. Without the presence of parallel BLSTM sub-models, both the SVEB and VEB classification performance decreases sharply. This proves the effectiveness of the proposed parallel sub-model architecture that better utilizes the ECG signals of both channels.

7.4 Hyper-Parameters

Different hyper-parameters are experimented to tune the micro-architecture of the proposed network. Various input sequence lengths of the BLSTM are evaluated on DS2, and the performance is listed in Table 7. In general, the F1 and G of both the VEB and SVEB recognition roughly decrease with reduction of input sequence length (when it is under 9), while the performance deteriorates when the input sequence length increases from 9 to 11. Since a longer sequence is not helpful to improve the classification performance, the length of 9 is the optimal choice. Table 8 shows the performance of the network topologies with different input sizes of the BLSTM and CNN after re-sampling on DS2. Since the larger input sizes introduce a high redundancy of networks and the smaller input sizes are not capable of extracting sufficient features, the selected parameters (300 for BLSTM and 100 for CNN) are more efficient compared with other configurations.

8. Conclusions

In this paper, we propose an end-to-end patient-specific ECG classification algorithm by integrating LSTM and CNN. In contrast with the conventional machine learning methods, the feature extraction and classification are fused into a single model. The heartbeats are sent to the networks without pre-processing, which largely reduces the classification complexity. Such automated feature extraction replaces hand-craft means, hence eliminating the prerequisites of too much priori knowledge and preserving the underlying characteristics of original ECG signals.

Compared with the existing neural network methods, our approach makes use of both the BLSTM and CNN creatively. For ECG classification, CNN alone extracts poor features from the long sequence due to the limitation of filter window scales, while LSTM is not effective enough to take care of detailed morphological characteristics. Besides, the cascaded network architecture using both topologies (e.g.,
Several other techniques are also proposed to improve the recognition accuracy in our study. The adaptive segmentation and re-sampling are applied to normalize network inputs of patients, while the clustering selection technique is adopted to choose representative heartbeats in the common training dataset and reduce the effect from imbalanced distribution of patterns.

When evaluated on the MIT-BIH arrhythmia database, the performance of the proposed classification algorithm is superior over the prior works. The timeliness and high accuracy of our algorithm make it promising for long-term ECG monitoring.

Acknowledgments

The authors acknowledge the support of the National Natural Science Foundation of China under Grant 61801425.

References

[1]  A.B. de Luna, P. Coumel, and J.F. Leclercq, “Ambulatory sudden cardiac death: Mechanisms of production of fatal arrhythmia on the basis of data from 157 cases,” American Heart Journal, vol.117, no.1, pp.151–159, 1989.
[2]  R. Hoekema, G.J.H. Uijen, and A. van Oosterom, “Geometrical aspects of the interindividual variability of multilead ECG recordings,” IEEE Trans. Biomed. Eng., vol.48, no.5, pp.551–559, 2001.
[3]  V. de Pinto, “Filters for the reduction of baseline wander and muscle artifact in the ECG,” Journal of Electrocardiology, vol.25, pp.40–48, 1992.
[4]  S. Oowski, L.T. Hoai, and T. Markiewicz, “Support vector machine-based expert system for reliable heartbeat recognition,” IEEE Trans. Biomed. Eng., vol.51, no.4, pp.582–589, 2004.
[5]  J. Kim, H. Shin, Y. Lee, and M. Lee, “Algorithm for classifying arrhythmia using extreme learning machine and principal component analysis,” 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp.3257–3260, 2007.
[6]  C. Ye, B.V.K.V. Kumar, and M.T. Coimbra, “Heartbeat classification using morphological and dynamic features of ECG signals,” IEEE Trans. Biomed. Eng., vol.59, no.10, pp.2930–2941, 2012.
[7]  T.J. Jun, H.J. Park, N.H. Minh, D. Kim, and Y.-H. Kim, “Premature ventricular contraction beat detection with deep neural networks,” 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp.859–864, 2016.
[8]  V.H.C. de Albuquerque, T.M. Nunes, D.R. Pereira, E.J.d.S. Luz, D. Menotti, J.P. Papa, and I.M.R.S. Tavares, “Robust automated cardiac arrhythmia detection in ECG beat signals,” Neural Computing and Applications, vol.29, no.3, pp.679–693, 2018.
[9]  H.F. Huang, G.S. Hu, and L. Zhu, “Sparse representation-based heartbeat classification using independent component analysis,” Journal of Medical Systems, vol.36, no.3, pp.1235–1247, 2012.
[10]  P. de Chazal, M. O’Dwyer, and R.B. Reilly, “Automatic classification of heartbeats using ECG morphology and heartbeat interval features,” IEEE Trans. Biomed. Eng., vol.51, no.7, pp.1196–1206, 2004.
[11]  P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T.-S. Zhou, T. Li, and J.-S. Li, “High-performance personalized heartbeat classification model for long-term ECG signal,” IEEE Trans. Biomed. Eng., vol.64, no.1, pp.78–86, 2016.
[12]  T. Ince, S. Kiranyaz, and M. Gabouj, “A generic and robust system for automated patient-specific classification of ECG signals,” IEEE Trans. Biomed. Eng., vol.56, no.5, pp.1415–1426, 2009.
[13]  P. Plawiak, “Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system,” Expert Systems with Applications, vol.92, pp.334–349, 2018.
[14]  T. Li and M. Zhou, “ECG classification using wavelet packet entropy and random forests,” Entropy, vol.18, no.8, 285, 2016.
[15]  Y. Ozbay, R. Ceylan, and B. Karlik, “Integration of type-2 fuzzy clustering and wavelet transform in a neural network based ECG classifier,” Expert Systems with Applications, vol.38, no.1, pp.1004–1010, 2011.
[16]  S. Raj and K.C. Ray, “Sparse representation of ECG signals for automated recognition of cardiac arrhythmias,” Expert Systems with Applications, vol.105, pp.49–64, 2018.
[17]  P. Xie, G. Wang, C. Zhang, M. Chen, H. Yang, T. Lv, Z. Sang, and P. Zhang, “Bidirectional recurrent neural network and convolutional neural network (BiRCNN) for ECG beat classification,” 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp.2555–2558, 2018.
[18]  S. Kiranyaz, T. Ince, and M. Gabouj, “Real-time patient-specific ECG classification by 1-D convolutional neural networks,” IEEE Trans. Biomed. Eng., vol.63, no.3, pp.664–675, 2015.
[19]  X. Zhai and C. Lin, “Automated ECG classification using dual heartbeat coupling based on convolutional neural network,” IEEE Access, vol.6, pp.27465–27472, 2018.
[20]  S. Saadatnejad, M. Oveisi, and M. Hashemi, “LSTM-based ECG classification for continuous monitoring on personal wearable devices,” IEEE J. Biomed. Health Inform., vol.24, no.2, pp.515–523, 2020.
[21]  M. Schuster and K.K. Paliwal, “Bidirectional recurrent neural networks,” IEEE Trans. Signal Process., vol.45, no.11, pp.2673–2681, 1997.
[22]  S.S. Xu, M.-W. Mak, and C.-C. Cheung, “Towards end-to-end ECG classification with raw signal extraction and deep neural networks,” IEEE J. Biomed. Health Inform., vol.23, no.4, pp.1574–1584, 2019.
[23]  O. Yildirim, “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” Computers in Biology and Medicine, vol.96, pp.189–202, 2019.
[24]  Z. Wu, T. Lan, C. Yang, and Z. Nie, “A novel method to detect multiple arrhythmias based on time-frequency analysis and convolutional neural networks,” IEEE Access, vol.7, pp.170820–170830, 2019.
[25]  S. Nurmaini, R.U. Partan, W. Kaesarendra, T. Dewi, M.N. Rahmatullah, A. Darmawahyuni, V. Bhayyu, and F. Firdaus, “An automated ECG beat classification system using deep neural networks with an unsupervised feature extraction technique,” Applied Sciences, vol.9, no.14, 2921, 2019.
[26]  E.K. Wang, X. Zhang, and L. Pan, “Automatic classification of CAD ECG signals with SDAE and bidirectional long short-term network,” IEEE Access, vol.7, pp.182873–182880, 2019.
[27]  F. Li, Y. Xu, Z. Chen, and Z. Liu, “Automated heartbeat classification using 3-D inputs based on convolutional neural network with multi-fields of view,” IEEE Access, vol.7, pp.76295–76304, 2019.
[28]  A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, and A.Y. Ng, “Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network,” Nature Medicine, vol.25, no.1, pp.65–69, 2019.
[29]  U.R. Acharya, S.L. Oh, Y. Hagiwara, J.H. Tan, M. Adam, A. Gertych, and R.S. Tan, “A deep convolutional neural network model to classify heartbeats,” Computers in Biology and Medicine, vol.89, pp.389–396, 2017.
[30]  A. Amirshahi and M. Hashemi, “ECG classification algorithm based
on STDP and R-STDP neural networks for real-time monitoring on ultra low-power personal wearable devices,” IEEE Trans. Biomed. Circuits Syst., vol.13, no.6, pp.1483–1493, 2019.

[31] R. Mark and G. Moody, “MIT-BIH arrhythmia database directory,” Massachusetts Institute of Technology, Cambridge, 1988.

[32] Association for the Advancement of Medical Instrumentation, “Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms,” ANSI/AAMI EC38, vol.1998, 1998.

[33] W. Jiang and S.G. Kong, “Block-based neural networks for personalized ECG signal classification,” IEEE Trans. Neural Netw., vol.18, no.6, pp.1750–1761, 2007.

[34] A.A.R. Bsoul, S.-Y. Ji, K. Ward, and K. Najarian, “Detection of P, QRS, and T components of ECG using wavelet transformation,” 2009 ICME International Conference on Complex Medical Engineering, pp.1–6, 2009.

[35] P. Warrick and M.N. Homsi, “Cardiac arrhythmia detection from ECG combining convolutional and long short-term memory networks,” 2017 Computing in Cardiology (CinC), vol.44, pp.1–4, 2017.

[36] M. Blanco-Velasco, B. Weng, and K.E. Barner, “ECG signal denoising and baseline wander correction based on the empirical mode decomposition,” Computers in Biology and Medicine, vol.38, no.1, pp.1–13, 2008.

[37] M. Jaderberg, K. Simonyan, A. Zisserman, et al., “Spatial transformer networks,” Advances in Neural Information Processing Systems, pp.2017–2025, 2015.

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

[39] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding,” Proc. Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp.1027–1035, 2007.

[40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” The Journal of Machine Learning Research, vol.15, no.1, pp.1929–1958, 2014.

Jiaquan Wu received the Bachelor of Engineering degree in electronic and information engineering from Zhejiang University, in 2015. He is currently a Ph.D. student at the Institute of VLSI Design, Zhejiang University. His research interests include biomedical signal processing and neural network accelerating.

Feiteng Li was born in Shijiazhuang, Hebei, China in 1992. He received the B.S. degree in internet of things engineering from the School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China, in 2014. He is currently working towards the Ph.D. degree at the College of Electrical Engineering, Zhejiang University, Hangzhou, China. His research interests include physiological signal processing with machine learning and ultra-low-power neural network accelerator.

Zhijian Chen received the B.S. degree from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2006, and the Ph.D. degree from the College of Electrical Engineering, Zhejiang University, Hangzhou, China, in 2011. From 2011 to 2013, he was a post-doctor in the College of Electrical Engineering, Zhejiang University. Since 2013, he has been a lecturer in the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. His research interest is ultra-low-power physiological signal processor design.

Xiaoyan Xiang received the B.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2008 and 2013, respectively, all in the Institute of VLSI Design. She has been a lecturer in the State Key Lab of ASIC and System, Fudan University from 2013. Her research interests include low-voltage low-power circuits for resilient processor, and the CPU architecture optimization in IoT field.

Yu Pu received the B.S. degree from Zhejiang University, Hangzhou, China, in 2004, and the Ph.D. degree in electrical engineering from the Eindhoven University of Technology, The Netherlands, in association with the NXP Research in 2009. From 2009 to 2011, he was a research assistant professor with Sakurai Lab, the University of Tokyo, Japan. From 2011 to 2012, he was a Research Scientist with the Accelerator Team, IBM Research Zurich, Switzerland. From 2012 to 2013, he was a Principal Scientist with NXP Research, where he led R&D in ultra-low-power MCUs. From 2014 to 2019, he was with Qualcomm Research, San Diego, CA, USA. Since 2019, he has been with Alibaba DAMO computing research lab, CA, USA.