Supervised ECG Interval Segmentation Using LSTM Neural Network

Hedayat Abrishami  
Chia Han  
Xuefu Zhou  
Department of EECS  
University of Cincinnati  
Cincinnati, OH, USA  
Email: abrishht@mail.uc.edu  
Email: han@ucmail.uc.edu  
Email: zhoxu@ucmail.uc.edu

Matthew Campbell  
Children’s Hospital of Philadelphia  
Philadelphia, PA, USA  
Email: campbellm5@email.chop.edu

Richard Czosek  
Heart Institute  
Children’s Hospital Medical Center  
Cincinnati, OH, USA  
Email: Richard.czosek@cchmc.org

Abstract—Segmenting electrocardiogram (ECG) into its important components is crucial to the field of cardiology and pharmaceutical studies, because analyses of ECG segments can be used to predict heart symptoms and the effects of cardiac medications. For each study, thousands of ECG signal points need to be analyzed and segmented. Despite of the success of using deep learning (DL) methods in multiple studies on classifying the heart condition, there are still lacking DL-based methods to characterize ECG temporal features. This paper describes a novel ECG segmentation method based on the recurrent neural network (RNN) with long short-term memory (LSTM) layers. In this model, each ECG sample is classified into one of the four categories: P-wave, QRS-wave, T-wave, and neutral (others). Our work shows that DL sequence learning methods outperform a traditional Markov model in terms of accuracy and using simple local features instead of complicated features, such as wavelet encoding. Particularly on T-wave segmentation, our approach can achieve an accuracy of 90%, compared to that of 74.2% using Markov models.

I. INTRODUCTION

Analyzing critical segments of the ECG waveform in ECG signals is crucial in determining the conditions of heart for diagnosing diseases or analyzing the effects of cardiac medications. All these analytical tasks require finding a large quantity of waveform patterns that can be identified as abnormalities in ECGs. Many of the pattern features are subtle and would require an expert clinician to recognize and spot them in ECG wave intervals. Utilizing clinicians is not economically feasible on a large scale. Thus, an automated approach to segment ECG signals is very promising. In addition, an accurate automated ECG segmentation method would provide more cost-effective and reliable results for patient diagnostics, especially for mass screening applications and in drug development.

A cardiac complex is composed of several wave components, and three of them are of high significance. These three waves are the P-wave, the QRS-wave, and the T-wave, as shown in Figure 1. The main goal of ECG segmentation is to detect and localize the QRS-wave, meanwhile, other segments such as T-wave and P-wave are also of high importance. For instance, one of the symptoms of sudden cardiac arrest and eventual death relies on finding the T-wave inversion, where the T-wave is inverted rather than in its normal rising form [2]. In another case, ST depression, a cause of sudden cardiac arrest, relies on finding the end of the QRS-wave and the beginning of the T-wave [2]. Thus, a reliable automated ECG signal processing approach should be capable of segmenting ECG cardiac complexes accurately.

ECG segmentation has two main challenges. The first challenge is the variety of wave formations and various ECG abnormality patterns [3]. The second challenge is the noise generated by ECG monitoring devices that sample the electrical activities of the heart muscle [4]. A robust and reliable automated approach should overcome these two barriers. Over the years, the research community has developed two general approaches for ECG segmentation. The first type relies on identification of important peaks in the ECG waveforms and determination of other points relative to those peaks. The other type is to classify every ECG data point into one of the cardiac waves [5]. Our work focuses on segmenting ECG signals by labeling each data point into one of the four ECG cardiac waves, P-wave, QRS-wave, T-wave, and neutral.
ECG segmentation can be performed in two steps. The first step is to obtain relevant features that discriminate the wave locations from the rest of the signal. The second step is to classify these points with the information revealed from the features. These features can be obtained by various filters or transformation methods [6], [7], [8]. Usually filters are smoothing filters [9], first-order derivative and second-order derivative [10], and etc. Transformations include Fourier transform [7], wavelet transform [11], and etc. The second step in ECG segmentation has used rule-based methods [12], Hidden Markov Model (HMM) [5], neural networks [13], evolutionary algorithms [14], and Bayesian techniques [15].

This paper is organized as follows: Section II briefly reviews the related works. Section III introduces ECG-SegNet, our novel long short-term memory neural network model for ECG segmentation. Section IV presents experimental results and discussion, followed by conclusion in Section V.

II. RELATED WORKS

Scientists have developed multiple automated approaches to detect various waveforms in ECG signals. An extensive literature review on ECG segmentation can be found in [16]. Generally, there are three main approaches, namely, the derivative-based [4], [17], wavelet-filters [18], and amplitude-based methods [3]. Martinez et al. [18] used a discrete wavelet-based method for extracting temporal features. Pan-Tompkins [10], developed one of the most famous derivative-based methods to find QRS-waves. Furthermore, for the purpose of classification, machine learning methods have been used in this field of study, which include Neural Network (NN) [13], Random Forest [19], Support Vector Machine (SVM) [19], Naive Bayes [15], HMM [5], rule-based methods [12], linear discriminants [20], and logistic regression [21].

With recent emergence of DL methods in various aspects of signal processing and data analysis, DL methods have been applied to ECG pattern recognition. In [22], DL methods were used to classify ECG signals into normal and abnormal ECG. In these works, the focus was to extract hierarchical features of ECG signals in order to classify heart disease symptoms. However, there is no specific DL study for finding and locating the major wave components or segmenting the cardiac complexes. This task is of high importance in cardiology communities since every cardiologist refers to them in their diagnoses. Furthermore, the rule-based classification methods have to come with different criteria with the changes in the formation of cardiac waves. HMM is a capable time series learner and ECG signal is a time series, thus, this method can be utilized to learn cardiac wave formations. However, HMM has its limitation of learning a combination of various patterns as typically shown in ECG signals. On the other hand, LSTM RNN is a time series learner capable of learning time series patterns and their combinations. This paper introduces a newly-proposed LSTM RNN architecture to segment ECG intervals. Our findings indicated that this new architecture outperforms the traditional ECG segmentation analysis using HMMs.

III. METHODOLOGY

In this work, a LSTM RNN architecture is proposed to segment ECG intervals. In Subsection III-A, the order and attributes of ECG intervals are discussed. Subsection III-B introduces the data sets including training, validation and test data used in this work [23]. Subsection III-C describes the features extracted from ECG signal to feed into our neural network model. Section III-D reviews the Bidirectional Long Short-Term Memory Neural Network (BLSTM-RNN). Section E introduces the novel architecture of ECG-SegNet and its post-processing step. Unlike the traditional Recurrent Neural Networks (RNN), this type of network is capable of learning long temporal dependencies, which makes it suitable for ECG segmentation [24]. Finally, Subsection III-E describes experiments and the convergence of ECG-SegNet. The result demonstrates the strength of ECG-SegNet compared to the other sequence learners such as HMM for the same task of ECG interval segmentation.

A. ECG Intervals

The wave complexes contain essential information, such as shape formation, interval duration, and amplitudes. Any abnormality in the waves, such as ST depression, T-inversion, long QT-interval, and so on, can provide meaningful information for cardiac diseases. With characterization of key values of wave parameters and the information that they carry, it is possible for physicians to make decisions that lead to differential diagnoses. There are many different shapes that can be found in ECG data. Normal QRS-waves can have nine different shapes [3]. P-waves and T-waves can also appear in different forms and amplitudes. In addition, all these waves can have abnormal shapes and formations. Detecting wave complexes that can relate to serious illnesses, such as sudden cardiac death (SCD), is critical.

A cardiologist identifies cardiac waves in relation to each other. For example, following a QRS-wave location is probably an S-wave. Thus, knowledge of the waves prior location is essential to predicting the consequent wave. Likewise, each individual waves formation also affects the formation of other waves. Therefore, a method that is capable of keeping persistent memory, i.e., a recurrent neural network, can be a viable solution to this type of time-dependent problem. The recurrent neural network creates a loop to pass the information from one timestamp to another, which allows it to learn a time series [25].

B. Data Sets

The data used for this study is the QT database (QTDB). It is designed to detect and segment ECG waveforms. The QTDB was produced by PhysioNet [23] and has a large collection of recorded physiological signals sampled at 250Hz. This database includes over 105 two-channeled ECG recordings, each 15-minutes in duration and it is chosen to include a broad variety of P, QRS, ST, and T morphologies [23]. This dataset thus allows researchers to perform research on ECG signal delineation.
For our current study, every recording is divided into 500 data points sampled at 250 Hz. Within every segment, one or more cardiac complexes can be found, which makes the segmenting task more challenging. The 500 ECG data points and their extracted features are the inputs to ECG-SegNet model, and correspondingly, the related annotations are the output targets for the ECG-SegNet model. In total, there are 64,040 sets of 500 data points of ECG that have been extracted from the QTDB to be used as inputs.

In our experiments, three different sets - training, validation, and test sets - have been created using all the extracted ECG parts. These sets are mutually exclusive, indicating there are no identical segments from one recording to another. Table I illustrates the training, validation and test data we used in this paper.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of samples</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>51,419</td>
<td>55%</td>
</tr>
<tr>
<td>Validation set</td>
<td>9,350</td>
<td>10%</td>
</tr>
<tr>
<td>Testing set</td>
<td>32,721</td>
<td>35%</td>
</tr>
</tbody>
</table>

### C. Extracted Features

In this work, in addition to raw ECG data points, three other features are extracted using different filtering kernels, such as the local average of a data point and the first and second derivatives of a data point. Hence, for every data point, a feature vector of size four is created to feed to the network, as explained in Subsection III-E. As a result of applying feature extraction to 500 ECG data points, the complete input to the ECG-SegNet becomes a matrix of 500 x 4 dimension. Table II illustrates the kernels that are used to convolve with the raw ECG signal.

### TABLE II KERNELES TO EXTRACT FEATURES

<table>
<thead>
<tr>
<th>No.</th>
<th>Type</th>
<th>Kernel</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Raw ECG</td>
<td>None</td>
<td>Not applicable</td>
</tr>
<tr>
<td>2</td>
<td>Smoothing</td>
<td>[1,1,1,1,1,1,1,1]</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>First Derivative</td>
<td>[1, 1, 1, 0, -1, -1, -1]</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>Second Derivative</td>
<td>[1, 1, 1, -6, 1, 1]</td>
<td>7</td>
</tr>
</tbody>
</table>

### D. Bidirectional Long Short-Term Memory Recurrent Neural Network Review

As mentioned earlier, a shortcoming of traditional neural networks is that it fails to classify an event based on prior observations [25]; however, recurrent neural networks are intrinsically fit for segmenting ECG signals. In a conventional recurrent neural network, the input \( x = (x_1, x_2, \ldots, x_T) \) feeds to the network, and RNN computes the hidden vector sequence, \( h = (h_1, h_2, \ldots, h_T) \), and the output vector sequence, \( y = (y_1, y_2, \ldots, y_T) \), from \( t = 1, \ldots, T \) while \( T \) is the number of timestamps as,

\[
h_t = H(W_{xh}x_t + W_{hh}h_{t-1} + b_h)
\]

where, \( W \) denotes weight matrices, \( b \) denotes bias vectors, and \( H \) denotes hidden layer function.

The novelty of this work is to use recurrent neural networks to classify each data point of an ECG signal into one of the four categories, namely, the P-wave, the QRS-wave, the T-wave, and the neutral. Thus, as a result of ECG segmentation, classification of every data point can be achieved. One of the main shortcomings of the conventional recurrent neural networks in dealing with long term dependencies is that these networks encounter a problem called vanishing-exploding gradients [26], i.e., the derivative of error with respect to weights gets close to zero or infinity after a short period of the time. This problem makes the networks hard to train for long term dependencies. Hochreiter and Schmidhuber [27] were able to overcome this problem in their well-known work on Long Short-Term Memory recurrent neural networks (LSTM RNN). LSTM RNN uses trainable memory cells called LSTM cells instead of simple neurons. These memory cells have three trainable gates including input, output, and forget gates. These gates have the ability to add or remove information, thus, avoid long term dependencies. A large number of applications have performed better than their competitors using such LSTM networks [28].

Figure 2 shows a LSTM cell, the gates and their output computations are given in the following equations:

\[
i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i)
\]

\[
f_t = \sigma(W_{xf}x_t + W_{hr}h_t - 1 + W_{cf}c_{t-1} + b_f)
\]

\[
c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)
\]

\[
o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o)
\]

\[
h_t = o_t \tanh(c_t)
\]

where, \( \sigma \) is the logistic sigmoid function, and \( i, f, o, a \) and \( c \) are the input gate, forget gate, output gate, cell input activation, and cell state vectors, respectively, and all of them are the same size as the hidden vector \( h \). \( W_{xi}, W_{xf}, W_{xo} \) are weight matrices for peephole connections [25].

Another aspect of RNN is that only the prior data is used. However, in many cases, the future data is available and can be used as an informational source. Schuster et al. introduced a Bidirectional RNN (BRNN) [29], which uses both directions of the data, prior and future data points, in two separate hidden layers. Graves et al. use LSTM in a BRNN and introduce Bidirectional LSTM (BLSTM) [24]. Therefore, BLSTM is a recurrent neural network that uses LSTM cells and computes both forward and backward hidden sequences. By stacking up these types of layers, a new deep network, called the Deep Bidirectional LSTM (DBLSTM), is obtained.

Utilizing DBLSTM in ECG segmentation helps to classify a signal sample based on prior and future sample points, i.e., finding QRS-wave using prior sample points such as P-wave samples and future data points such as S-wave samples. Thus, DBLSTM becomes a very viable approach to be explored for ECG segmentation task. We will use this to define new network architecture in the next section.
Based on the aforementioned rationale for a suitable network using DBLSTM to classify ECG waveform data points, a new architecture, called ECG-SegNet, is proposed, as shown in Figure 3.

This architecture contains the following layers. From bottom up, the first layer is the input layer, \( x = (x_1, ..., x_t, x_{t+1}, ..., x_{500}) \), which takes the raw ECG signal of size 500 and three additional extracted features per data point, explained in Section III.C. Therefore, the input is \( 500 \times 4 \) time series. The hidden layer is a BLSTM layer. On every timestamp, there are two different hidden LSTM layers including forward hidden layer and one backward hidden layer. On every set, this layer has 250 LSTM cells, which suggests that it has a total of 500 hidden LSTM cells. This is followed by another BLSTM of size 250, and each hidden LSTM layer has 125 LSTM cells. The next layer is the output layer that classifies every data point in time series into four categories. This layer is called time distributed output layer and is applied on every timestamp. Each ECG signal has 500 samples, so the output dimension is \( 500 \times 4 \). For ECG signal, \( x \) of length \( T \), the network produces a length of \( T \) output sequence \( y \), where each \( y_t \) defines a probability distribution over the \( |K| \) possible states where \( K = \{1, 2, 3, 4\} \) and \( k \epsilon K \): that is, \( y_t^k \) (the \( k^{th} \) element of \( y_t \)) is the network’s estimate for the probability of observing state \( k \) at time \( t \), given \( x \). The network is trained to minimize the negative log-probability of the target sequence using a softmax output layer, Eq. 8 and Eq. 9. With length \( T \) and target sequence \( z \), the network is trained to minimize the error function shown in Eq. 9.

\[
-log Pr(z|x) = - \sum_{t=1}^{T} log y_t^z
\]  

(8)

The error derivatives at the output can be obtained as

\[
-\frac{\partial \log Pr(z|x)}{\partial y_t^k} = y_t^k - \delta_{k,x_t}
\]  

(9)

where \( \delta^k \) is the vector of output activations before they have been normalized with the softmax function. These derivatives are then fed back through the network using backpropagation through time to determine the weight gradient [25], [28].

As explained earlier, every data point in the ECG signal is classified into one of the four different categories. Segmentation implies that if the input has the temporal dimension of size \( T \), the output has the same dimension of size \( T \). In this work, the ECG input has 500 data points, which means output dimension is equal to \( 500 \times |K| \), where \( |K| \) is the number of categories. Eq. 9 computes the probability of input \( x \) at time \( t \), belonging to class \( k \), with \( k \) being one of the four categories. The final output for each data point is the class with the highest probability. In summary, Table III illustrates the architecture of the ECG segmentation network, the ECG-SegNet. It is three layers deep, consisting of the input layer, two bidirectional LSTM layers, and an output layer. More specifically, the ECG-SegNet input is a \( 500 \times 4 \) dimension, which includes 500 data points each of raw ECG signals, smoothed ECG signals, the first derivative of the ECG signal, and the second derivative of the ECG signal. The output of the signal comprises of a
500 × 4 neuron-producing probability over a 500-unit timeline and is classified into four categories: neutral, P-wave, QRS-wave, and T-wave.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Description of the layer</th>
<th>LSTM Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>ECG raw signal and its features</td>
<td>500 × 4</td>
</tr>
<tr>
<td>Layer 1</td>
<td>Bidirectional LSTM with 250 cells each</td>
<td>2 × 250</td>
</tr>
<tr>
<td>Layer 2</td>
<td>Bidirectional LSTM with 125 cells each</td>
<td>2 × 125</td>
</tr>
<tr>
<td>Output</td>
<td>Time distributed dense layer</td>
<td>500 × 4</td>
</tr>
</tbody>
</table>

After generating the output from the network, a post-processing step is applied. A filter of size 17 is applied to it. If the beginning and the end of the output under this filter belong to the same cardiac wave class, any output under this window is then assigned to the class of the start and end of this filter.

F. Training Experiment

The training set includes 51,419 ECG segments of size 500 × 4. The target data is a 500 × 4 matrix, which is the annotated class obtained from QTDB. If a data point belongs to the first class, neutral, the output data at that timestamp is [1, 0, 0, 0] vector. It gave the probability of 1 to the first class and the rest were 0. The ECG-SegNet is trained with Adam Optimizer [30] through 68 epochs using mini-batch procedure of batch size 250. The training stopped after 68 epochs because the gap between training set error was getting smaller and validation set error was getting larger and this divergence is a sign of overfitting. After training, the results showed 94.6% accuracy for training set, 93.8% accuracy for validation set, and 93.7% accuracy for test set. Figure 4 shows the accuracy rates and Figure 5 shows the error rates through 68 epochs for validation and training sets.

<table>
<thead>
<tr>
<th>Method</th>
<th>P (%)</th>
<th>QRS (%)</th>
<th>T (%)</th>
<th>Overall (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECG-SegNet</td>
<td>92.0</td>
<td>94.0</td>
<td>90.0</td>
<td>92.0</td>
</tr>
<tr>
<td>HMM on raw ECG [5]</td>
<td>5.5</td>
<td>79.0</td>
<td>83.6</td>
<td>56.03</td>
</tr>
<tr>
<td>HMM on wavelet encoded ECG [5]</td>
<td>74.2</td>
<td>94.4</td>
<td>96.1</td>
<td>88.23</td>
</tr>
</tbody>
</table>

The majority of the other researches focus on finding the cardiac complex fiducial points and not segmenting every
single data point of ECG independently. Even though the ECG-SegNet task is different than finding ECG cardiac waves location, it provides competitive accuracy in finding cardiac wave locations. Table VI shows the accuracy of finding waves regardless of segmentation using ECG-SegNet.

<table>
<thead>
<tr>
<th>Wave Identification Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>P-wave</td>
</tr>
<tr>
<td>QRS-wave</td>
</tr>
<tr>
<td>T-wave</td>
</tr>
</tbody>
</table>

Fig. 6. Two sample results

V. CONCLUSION

To the best of our knowledge, there is not a DL-based method for ECG signal segmentation yet. Our work demonstrated that ECG-SegNet is a powerful network capable of understanding the temporal ECG using only a few local features to yield very competitive results.

The ability to delineate ECG cardiac waves augments the possibility of contributing future research in cardiology. By combining the vital information of waveforms with other methods in recognizing symptoms, more accurate heart related diseases can be diagnosed, and high-throughput, automated ECG diagnostic systems can be developed to serve the need of large population screening for disease prevention [2].

ACKNOWLEDGMENT

We thank Ohio Supercomputer Center [31] for providing high performance computation power for this research.

REFERENCES


