Multi-level Information for Non-invasive Identification of Exit Site of Ventricular Tachycardia

Qiupeng Feng[1], Hongjie Hu[2] , Huafeng Liu[1]

[1]State Key Laboratory of Modern Optical Instrumentation, College of optical science and engineering , Zhejiang University, Hangzhou, China
[2]Department of Radiology, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China

Abstract

The heart problem is increasingly becoming the killer of human health. In many heart conditions, ventricular tachy-cardias are the most common cause of death. At present, we have been able to use catheter ablation to treat ventricular tachycardias clinically, but it is important to achieve the non-invasive identification of the abnormal potential point in advance. To address this issue, we proposed a novel neural network architecture, namely LGNnet, to identify the exit of ventricular tachycardia by using only 12-lead electro-cardiograms. Our LGNnet respectively introduce different modules to acquire local-level, global-level and non-local feature information. By utilizing these feature information, we finally establish a model with the smaller positioning error.

1. Introduction

Sustained ventricular tachycardias (VTs), are the most common causes of sudden cardiac death. It can be treated by catheter ablation, where invasive three-dimensional electroanatomical mapping systems such as CARTO and ENSITE are used through applying tiny probes to make point-by-point measurements on the surface of the heart to find the exit site of VT[1] . However, this procedure is still very challenging in clinical treatment, even for experienced cardiologists. Therefore, figuring out a fast and non-invasive localization of the origin of ventricular activation has attracted more and more researchers' attention and a variety of potential non-invasive solutions are emerging.

Previous studies[2] have shown that ECG which is generated by endocardial pacing can provide information about pacing areas in patients with organic heart disease. After that, various non-invasive methods have been proposed to locate the sites of abnormal activation using electrophysiological information provided by ECG, based on mathematical statistics and machine learning methods.[3] [4]

Recently, deep learning based methods have been an area of major concentration. Convolutional Neural Networks (CNN) [5] [7] have been proved as an effective structure to extract local feature information and researchers have done a lot of work on CNN's application in cardiac localization. Different from CNN, Recurrent Neural Networks (RNN) [8] [10] are the dominant solution to extract global features for ECG classification and anomaly detection. Nevertheless, both CNN and RNN deal with local neighbourhood information in time domain or space domain, which will lead to the loss of non-local information of ECG signal. Inspired by [11] , we utilize non-local information extracted by deep network for ECG anomaly location.

In this paper, we propose a novel neural network architecture for non-invasive localization of the origin of ventricular activation by using the body surface 12-lead ECG data. It is a truly end-to-end data-driven architecture based on temporal feature encoding, requiring no feature engineering or magnetic resonance imaging (MRI) /computed tomography (CT) scans. Therefore, our model can be easily applied to assist cardiologists in clinic. We introduce Conv1D layers to encode the local-level features of ECG data in the time domain from the QRS complex and further decode these features into the global-level temporal dimension with one layer of Bi-directional LSTM(BLSTM). Then we combine local and global-level features and feed into the self-attention module to acquire the non-local information. Finally, we introduce the multi-layer perceptron (MLP) to map the decoded ECG feature information to cardiac three-dimensional coordinate system to predict the coordinates of VT exits. The contributions of our work are: (i) proposing a novel neural network architecture for localization of abnormal cardiac excitation points. (ii) training our model on the benchmark dataset and achieving state-of-the-art performance without any additional information such as MRI or CT.

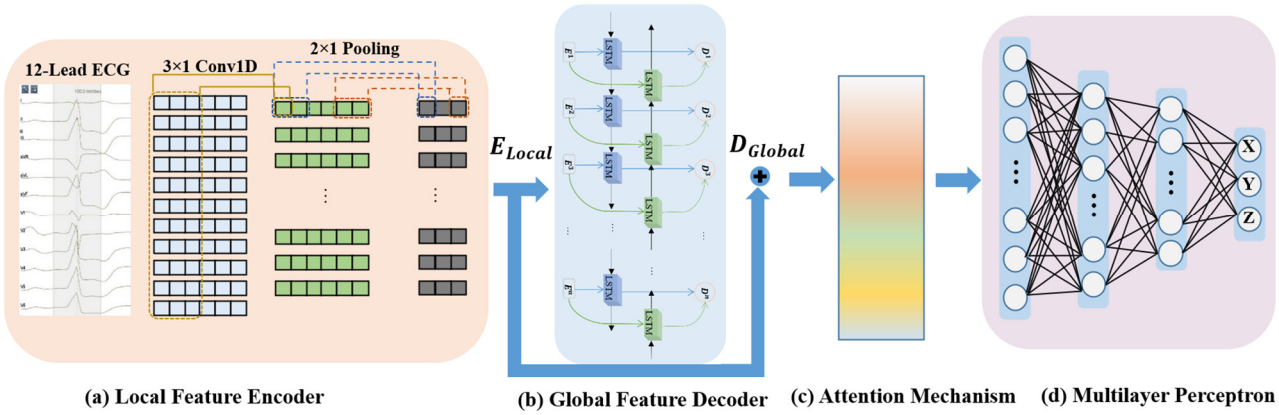

Fig.1. Diagram of Local Global Non-Local Network (LGN net).

2. Methodology

In this section, we will describe every module of our neural network architecture one-by-one. The Local Global Non-Local Network can be divided into four main modules: Local Feature Encoder module, Global Feature Decoder module, Attention Mechanism module and Multilayer Perceptron, as shown in Fig 1.

2.1. Problem Formulation

We denote the body surface electric potentials recorded by the 12-lead ECG sensor $\boldsymbol{i}$ at time $\boldsymbol{t}$ by $\boldsymbol{v_t^i}$ for all $\boldsymbol{i}$ =1, 2, …, 12 and $\boldsymbol{t}$ = 1, 2, …, 100. Let $\boldsymbol{v_t} = [\boldsymbol{v_t^1}, \boldsymbol{v_t^2}, \dots, \boldsymbol{v_t^{12}}]^{\boldsymbol{T}}$ be the 12*1 vector of measurements at time $\boldsymbol{t}$, where $\boldsymbol{S}$ represents the point of VT exits and $\boldsymbol{S}$: $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ represent the coordinate values of X, Y and Z axes. The purpose here is to estimate the exit of ventricular tachycardia $\tilde{\boldsymbol{S}}$: $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ based on measurement $\boldsymbol{v_t}$, which can be expressed as the following formula:

$$\tilde{\boldsymbol{S}}: (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \boldsymbol{LGN}(\boldsymbol{v_t}) \quad (1)$$

In other words, the objective is to make the predicted value $\tilde{\boldsymbol{S}}$ closer to the ground truth value $\boldsymbol{S}$ which is called as Euclidean distance in physical space:

$$\boldsymbol{min}\left\|\tilde{\boldsymbol{S}}: (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) - \boldsymbol{S}: (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\right\|^2 \quad (2)$$

2.2. Local Feature Encoder

During the clinical diagnosis of cardiac problems, the interpretations of 12-lead ECG signals by medical workers mostly depends on its local characteristics, such as PR Interval or QRS Complex and so on. Previous studies [6] have shown that Conv1D can be an effective structure to extract morphological information, such as Q, R, S waves. Inspired by these, we develop a Conv1D based module to learn and generalize the local-level representation of the body surface 12-lead ECG, which can be defined as:

$$\boldsymbol{E_{Local}} = \boldsymbol{Conv1D}(\boldsymbol{v_{input}}; \boldsymbol{W_{Local}}) \quad (3)$$

Here, $\boldsymbol{Conv1D}(\cdot)$ is the Conv1D network and $\boldsymbol{v_{input}}$ is the 12-lead ECG, the input of our LGN net, with corresponding weight and biases $\boldsymbol{W_{Local}}$. $\boldsymbol{E_{Local}}$ is the local-level feature encoded by the local-level feature encoder module. Fig.1(a) shows the neural layers of Conv1D module we use to extract and encode local-level feature of a given 12-lead ECG record. This module consists of three Conv1D blocks which consist of two layers of Conv1D and a layer of 1D max pooling with size 2 except for the last one. Every Conv1D layer contains one layer of 3×1 Conv1D, one batch normalization layer and one activation layer using the ELU activation function.

2.3. Global Feature Decoder

For the local-level features encoded by previous encoder module, we need a decoder to interpret their representations. A perfect choice is Bi-directional LSTM. Different from encoders, it is crucial for decoders to consider both past and future information to get a good interpretation of encoded features. BLSTMs have two independent LSTM sequence forwards and backwards to calculate two separate cell states $\boldsymbol{C_{BLSTM}^t}$, capturing the past and future information respectively. The BLSTM decoder is defined as:

$$\boldsymbol{D_{BLSTM}^t} = \boldsymbol{BLSTM}\begin{pmatrix}\boldsymbol{D_{BLSTM}^{t-1}}, \boldsymbol{D_{BLSTM}^{t+1}}, \boldsymbol{E_{Local}^t};\\ \boldsymbol{W_{BLSTM}^{t-1}}, \boldsymbol{W_{BLSTM}^{t+1}}\end{pmatrix} \quad (4)$$

Here, $\boldsymbol{BLSTM}(\cdot)$ is the BLSTM network and $\boldsymbol{D_{BLSTM}}$ is the hidden state of the BLSTM decoder. $\boldsymbol{E_{Local}^t}$ indicates that the local-level features encoded by modules before as the BLSTM's input. $\boldsymbol{W_{BLSTM}}$ represents corresponding weight and biases.

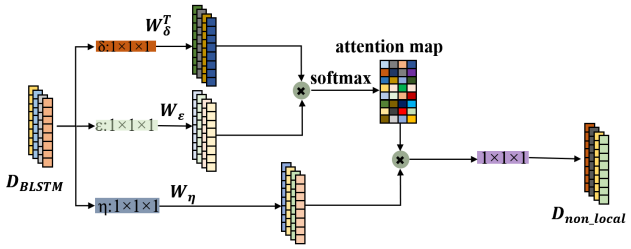

Fig.2. Detailed Illustration of Attention Mechanism Module.

2.4. Attention Mechanism

For our attention mechanism, we introduce the self-attention module to capture the relationship between widely separated temporal sequence. As in the [11] , we define the operation to extract non-local information in our model as:

$$\boldsymbol{D_{non\_local} = softmax\left(D_{BLSTM}^T W_\delta^T W_\varepsilon D_{BLSTM}\right) W_\eta D_{BLSTM}} \quad (5)$$

Here, $\boldsymbol{softmax}(\cdot)$ plays the role of the normalization operation. $\boldsymbol{W_*}$ is the weight matrix to be learned. $\boldsymbol{D_{BLSTM}}$ is the input feature and $\boldsymbol{D_{non\_local}}$ is the output non-local feature of the same size as $\boldsymbol{D_{BLSTM}}$. The whole calculation process is shown in the Fig.2

2.3. Local Global Non-Local Network

In this part, we finally can combine these modules we mentioned above into our Local Global Non-Local Network (LGNnet). The input of LGNnet is a 12-lead ECG record which is a 12-dimension vector has a length of 100 timestamps. For each ECG record, Conv1Ds calculate the local-level feature vectors based on the input of 12-lead ECG. After is, we feed the previous local-level feature vectors into the BLSTM to jointly decode the feature vectors. Then we utilize the attention mechanism to further extract the non-local information. Finally, we feed the output to one global average pooling layer (GAP) which has no parameters and much less calculation work. At last, we apply multilayer perceptron (MLP) to transform feature vectors into 3-dimensional coordinates of the predicted position of VT exit $\tilde{\boldsymbol{S}}$: $\boldsymbol{(x, y, z)}$.

$$\tilde{\boldsymbol{S}}: \boldsymbol{(x, y, z)} = \boldsymbol{MLP}\left(\boldsymbol{GAP}(\boldsymbol{D_{non\_local}}); \boldsymbol{W_{MLP}}\right) \quad (6)$$

$\boldsymbol{MLP}(\cdot)$means the operation of global average pooling. The multilayer perceptron $\boldsymbol{MLP}(\cdot)$ including the conforming weights and biases $\boldsymbol{W_{MLP}}$ maps the non-local pooled features vector $\boldsymbol{GAP}(\boldsymbol{D_{non\_local}})$ to the coordinate space to localize the coordinates of VT exits $\tilde{\boldsymbol{S}}$: $\boldsymbol{(x, y, z)}$.

3. Experiments and Results

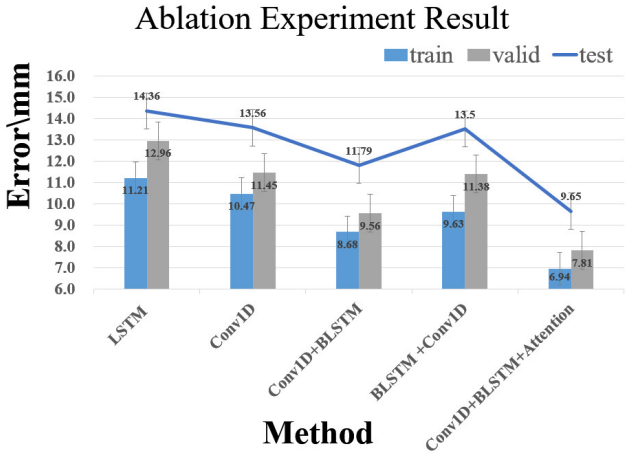

Fig. 3. Results of Ablation Study. The longitudinal axis represents the Euclidean distance error of positioning, while the horizontal axis represents different methods.

We test the performance of our method in a dataset of records of 12-lead ECGs, acquired during the pace-mapping procedure on the left ventricular (LV) endocardium with analog output of the multichannel recording system. The whole dataset was split into training and test sets. The training set consists of about 13000 data, while the test set has 4000 data.

3.1. Ablation Study

In this part, we will design the ablation experiments to verify the effectiveness of each module of our neural network architecture. We compare the performance with five baseline architecture: 1) LSTM; 2) Conv1D; 3) Conv1D+BLSTM; 4) BLSTM+Conv1D; 5) Conv1D+B-LSTM+Attention.

As the result shown in **Fig.3**, Conv1D+BLSTM models outperform the single network model LSTM and Covn1D, showing that utilizing Conv1D to extract features as encoder and BLSTM as decoder can achieve better results. As a counterexample, the results of BLSTM+Conv1D show that the more complex the network structure, maybe the worse the result. Here, we use LSTM as the encoder and Conv1D as the decoder and the worse result are obtained than a single network model. We believe that in this case the encoded feature information extracted by LSTM are not deep enough and Conv1D cannot decode features in a global time dimension well. Among all the other methods, Conv1D-BLSTM-Attention achieve the best performance. This demonstrates that jointly encoding local-level, global-level and non-local features can extract features better and significantly improve the positioning accuracy of neural network models. By comparing the results of Conv1D, Conv1D+BLSTM and Conv1D+BLSTM+Atttention, we can see the effectiveness of our different modules.

Table.1. Illustration of Non-Local Feature Decoder Module

Method	Training	Validation	Test
LGN net	6.94	7.81	**9.65**
Random Forest	10.93	11.17	13.26
LightGBM	7.55	8.66	12.722
SVM	7.16	8.83	13.46
Res-Conv1D	13.32	15.08	16.44
Seq2Seq	10.41	12.45	14.18

3.2. Comparison Experiment

In this section, we evaluated our method by comparing with various methods of machine learning and show the effect of the combination of local-level, global-level and non-local features in ECG Signal analyzing. The comparison methods are: 1) LGNnet; 2) Random forest (RF); 3) Light Gradient Boosting Machine (LightGBM) [14] ; 4) Support Vector Machine (SVM); 5) 34-layer Conv1D with residual Block [6] ; 6) Seq2Seq [15] ;
For each method, the mean positional error and its 95% confidence interval on training, validation and test set was reported in **Table.1** to highlight the advantages of our method more intuitively. Among them, LightGBM is the latest tree model structure proposed by Microsoft which also obtains second-best results after our method. However, compared with the traditional machine learning method, deep neural network has not achieved a good result in our ECG positioning problem, especially for Res-Conv1D, which was successfully used to classify 12 rhythm classes. In view of this phenomenon, we consider that the deep neural network may be more sensitive to the setting of hyper-parameters and a lower generalization characteristic for dealing with different types of tasks. For all the deep neural network models in the tables, we only change the dimension of the input and output.

4. Conclusion

In this paper, we have developed a deep learning based framework for the analysis of exit of VT from 12-leads ECG data. It is a truly end-to-end model only requiring the patient's 12-lead ECG. This approach is based upon both CNN and RNN to extract the local-level, global-level and non-local features. Experiments have been performed on and results have been compared to previously state-of-the-art methods in the literature.

Acknowledgments

This work is supported in part by the National Key Technology Research and Development Program of China (No: 2017YFE0104000, 2016YFC1300302), and by the National Natural Science Foundation of China (No: U1809204, 61525106, 81873908, 61701436).

References

[1] P. M. Van Dam et al., "Non-invasive imaging of cardiac activation and recovery," Annals of biomedical engineering, vol. 37, no. 9, pp. 1739–1756, 2009.
[2] Waxman H. L., and M. E. Josephson, "Ventricular activation during ventricular endocardial pacing: I. Electrocardiographic patterns related to the site of pacing.," Am. J. Cardiol. 50(1), 1–10, 1982.
[3] C. Liu and B. He, "Noninvasive estimation of global activation sequence using the extended kalman filter," IEEE Transactions on Biomedical Engineering, vol. 58, 2011.
[4] Y. Rudy, "Noninvasive electrocardiographic imaging of arrhyth-mogenic substrates in humans.," Circulation research, vol. 112, no. 5, pp. 863– 874, 2013.
[5] S. Ghosh and Y. Rudy. "Application of L1-norm regularization to epicardial potential solutions of the inverse electrocardiography problem.," Ann. Biomed. Eng.
[6] Y. Hannun and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network" *Nat. Med.*
[7] T. Yang, L. Yu, Q. Jin, L. Wu, and B. He, "Localization of Origins of Premature Ventricular Contraction by Means of Convolutional Neural Network from 12-lead ECG," *IEEE Trans. Biomed. Eng.*, vol. 9294, no. c, 2017.
[8] R. Silipo and C. Marchesi, "Artificial neural networks for automatic ECG analysis," in IEEE Transactions on Signal Processing, vol. 46, no. 5, pp. 1417-1425, May 1998.
[9] Chauhan, Sucheta, and Lovekesh Vig. "Anomaly detection in ECG time signals via deep long short-term memory networks." 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2015.
[10] Malhotra, Pankaj, et al. "Long short term memory networks for anomaly detection in time series." Proceedings. Presses universitaires de Louvain, 2015.
[11] Wang, Xiaolong, et al. "Non-local neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
[12] Ashish Vaswani et al. "Attention is all you need.," In Advances in Neural Information Processing Systems.
[13] Chris Dyer et al. 2015. "Transition-based dependency parsing with stack long short-term memory.," In Proceedings of ACL-2015, pages 334–343, Beijing, China, July 2015.
[14] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM : A Highly Efficient Gradient Boosting Decision Tree," no. Nips, pp. 1–9, 2017.
[15] Sutskever, I., et al, "Sequence to sequence learning with neural networks.," In Advances in neural information processing systems, pp. 3104-3112, 2014.

Address for correspondence:

Huafeng Liu
Professor
Zhejiang University, YuGu Road 38, Yuquan Campus, Hang-zhou, China
E-mail address: liuhf@zju.edu.cn