

Automatic Coordinate Prediction of the Exit of Ventricular Tachycardia From 12-Lead Electrocardiogram

Prashna K Gyawali¹, Shuhang Chen², Huafeng Liu², B. Milan Horacek³,
John L. Sapp³, Linwei Wang¹

¹ Rochester Institute of Technology, Rochester, USA

² Zhejiang University, Zhejiang, China

³ Dalhousie University, Halifax, NS, Canada

Abstract

Automatic localization of the exit site of ventricular tachycardia (VT) can improve the efficiency and efficacy of catheter ablation. Because the exit site of the VT gives rise to its QRS complex on electrocardiogram (ECG), it is possible to build a predictive model to directly localize the exit of a VT from its 12-lead ECG. In previous works, prescribed features such as QRS integrals have been used to build such models. In this paper, we propose a deep network to automatically extract more discriminative features from QRS complex to localize the origin of ventricular activation. To improve the resolution of localization compared to previous works based on a small number of pre-defined segments, we localize the origin of ventricular activation as 3D coordinates. Model training and testing were performed on 12-lead ECG data of 1012 distinct pacing sites, collected from patients during routine pace-mapping procedures. Compared with the use of prescribed QRS-integral as an input feature, the presented deep model achieved an improvement of localization accuracy by approximately 4 millimeters (~26%) on average.

1. Introduction

Sustained ventricular tachycardia (VT) often involves a life-threatening electrical “short circuit” that travels through and exits from a myocardial scar to depolarize the rest of the ventricles [1]. Radiofrequency catheter ablation is considered an effective approach to cut off the short circuit by destroying its exit [1]. The QRS complex of 12-lead electrocardiograms (ECG) can be used to infer the location of a VT exit site. In current clinical practice, this is done as a part of the routine pace-mapping procedures, where pacing is applied to different sites of the myocardium, and the VT exit is considered to be near the site at which pacing reproduces the QRS morphology of the VT on all 12 ECG leads.

Instead of this “trial-and-error” practice, a real-time au-

tomatic localization of the exit of an VT directly from its ECG data may improve the efficiency and efficacy of ablation procedures. This idea is visible in existing research works [2] [3] [4] [5] where, from a large amount of ECG data obtained during routine pace mapping procedures, a predictive model is built to infer the localization of the origin of ventricular activation. However, limited localization accuracy has been obtained due to the following factors: limited localization resolution, hand-engineered features and inter-subject variability in ECG data. First, the localization is often done on a small number of pre-defined segments that impose a low resolution in the localization results. Second, a limited number of prescribed features - typically the time-integral of the QRS complex - are used to summarize its time varying morphology. Finally, to build an accurate prediction model from a large population is challenged by the fact that the ECG signals have significant physiological and pathological variations across individuals. The recent work in [6] demonstrated the use of patient-specific models to predict the 3D coordinates of the VT exit. It addresses the issue of a low-resolution localization by departing from segment-based models, and it avoids inter-subject variations in ECG data by building a model for each subject. However, such patient-specific models are difficult to implement in clinical practice because it necessitates the collection of a sufficient amount of pace-mapping data on each patient before the prediction can be made.

In this paper, we aim to overcome the above challenges by exploiting the power of deep learning models in localizing VT exits from 12-lead ECG data. First, limited localization resolution problem is addressed by predicting the x - y - z coordinate of VT exits instead of pre-defined segments. Second, a deep framework is presented to extract the features automatically from the QRS complex of ECG signals rather than using hand-engineered features. Finally, inter-subject variations are disentangled within the deep network using our previously developed approach

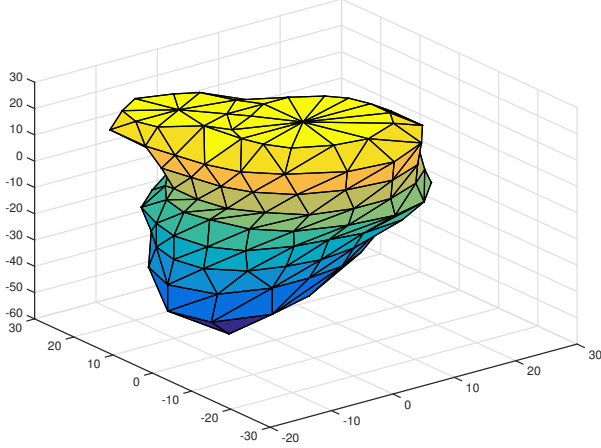


Figure 1. The standard prototype of LV endocardial surface

[7], where a supervised contrastive regularization is added to a denoising autoencoder (DAE) to separate factors of variations during the learning of hidden features. The proposed model is trained with 12-lead QRS complex along with the corresponding pacing locations. Its localization accuracy was evaluated on a separately held out dataset. The improvement of localization accuracy obtained by the presented model is demonstrated against prediction using linear regression with pre-scribed QRS integral features. Moreover, the increment in localization resolution by the presented model is presented against localizing VT exit into pre-defined segments.

2. Methods

A machine learning model is built to learn a relationship between the 12-lead QRS complexes (model input) and the origin of the corresponding ventricular activation (model output). The ECG data and their corresponding origins of ventricular activation can be obtained during routine pace mapping procedures on patients undergoing catheter ablation of VT. All these pacing sites are registered to a common LV endocardial model as shown in Fig 1.

2.1. Feature Based Approach

In existing works [5] [6], QRS integrals are common features used to localize the VT exit. As a baseline, we build the linear regression model to predict the x - y - z coordinates of the pacing sites from 120-ms QRS integrals extracted from all 12 ECG leads. QRS integrals are extracted by trapezoidal approximation. All QRS integrals from 12 ECG leads are used as input \mathbf{x} is mapped to the output coordinates \mathbf{y} via a linear model $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}$, where parameters $\{\mathbf{W}, \mathbf{b}\}$ are fitted using a least squares approach [6].

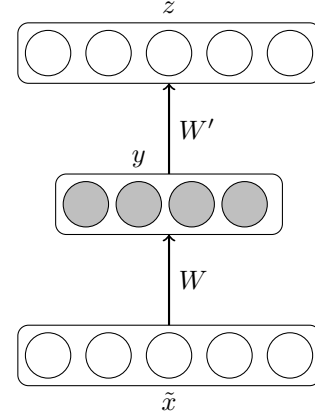


Figure 2. Denoising Autoencoder. $\tilde{\mathbf{x}}$ stands for corrupted input, \mathbf{y} is latent representation, \mathbf{z} is reconstructed input while \mathbf{W}, \mathbf{W}' are weight matrices.

2.2. Denoising Autoencoder (DAE)

Denoising autoencoder [8] learns to extract a robust low-dimensional representations of the raw input data by learning to reconstruct the clean input from a noise-corrupted version. As shown in Fig. 2, for an input vector $\mathbf{x} \in \mathbb{R}$, we first corrupt the input data into $\tilde{\mathbf{x}} \sim q_{\mathcal{D}}(\tilde{\mathbf{x}}|\mathbf{x})$, with binary masking or Gaussian additive noises. With the training data, the DAE then learns to map the corrupted $\tilde{\mathbf{x}}$ to a low-dimensional hidden representation $\mathbf{y} \in [0, 1]^{d'}$ through a deterministic transformation $\mathbf{y} = \mathbf{f}_{\theta}(\mathbf{W}^T \tilde{\mathbf{x}} + \mathbf{b})$, parameterized by $\theta = \{\mathbf{W}, \mathbf{b}\}$ where \mathbf{W} is a $d' \times d$ weight matrix, \mathbf{b} is the bias vector and \mathbf{f}_{θ} is a sigmoid function. The resulting latent representation \mathbf{y} is then “reconstructed” back to $\mathbf{z} \in \mathbb{R}$ through $\mathbf{z} = \mathbf{g}_{\theta'}(\mathbf{W}'^T \mathbf{y} + \mathbf{b}')$ where $\theta' = \{\mathbf{W}', \mathbf{b}'\}$, with an objective to minimize the average reconstruction error over n training examples:

$$\theta^*, \theta'^* = \arg \min_{\theta, \theta'} \frac{1}{n} \sum_{i=1}^n L_r(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) \quad (1)$$

The loss function L_r could be least squares or cross entropy functions. It is optimized using stochastic gradient descent [8]. In this way, a meaningful low-dimensional feature of the input data can be learned.

2.3. Two-Way DAE

We previously developed a two-way factored DAE [7] to disentangle the factor of variations in ECG signals. Specifically, the encoding process maps the input data to two different latent representations, \mathbf{h} and \mathbf{m} , representing respectively the patient-specific factor and the factor that represents the common relationship between origins of ventricular activation and QRS data. For QRS data originating from nearby locations, their hidden representations \mathbf{m} are assumed to be similar regardless if the data are collected

from the same patient; otherwise, \mathbf{m} is assumed to be different. On the other hand, for QRS data from the same patient, their hidden representation \mathbf{h} is assumed to be similar regardless of the origin of activation. To facilitate such pair-wise comparison, an additional contrastive loss [9] is used as a semi-supervised training criterion for the DAE. We randomly generate training pairs $X^p = (\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ from the training data. Each pair has labels $e^p = (e_m^p, e_h^p)$, where e_*^p is 1 if the pair shares the same label on factor $*$, and 0 otherwise. For e_h^p (patient related label), it can be easily determined from the patient label for the pair of data. Given the coordinates of an QRS data pair, e_m^p is 1 if the distance between the activation origins of the pair of data is less than 18mm, and $e_m^p = 0$ otherwise; the threshold of 18mm is determined from the average segment radius of the 16-segment model [5].

The overall objective function for training the proposed DAE consists of the standard reconstruction loss (1) and the contrastive loss (L_f):

$$L(X^p) = L_r(X^p) + \alpha L_f(X^p) \quad (2)$$

where α is tuned during optimization by stochastic gradient descent (SGD) with back propagation.

2.4. Supervised fine-tuning

Similar to [8], the learned encoding portion as described in section 2.1 or 2.2 is stacked into a three-layer deep network as shown in Fig 3. The highest-level feature learnt is used as the input to a linear regression model. The parameters of this network are fine-tuned together using stochastic gradient descent. In this way, we are yielding a deep neural network subject to supervised learning.

3. Experiments

To evaluate if a deep learning model will improve the accuracy in localizing the VT exit from 12-lead ECG, we compare the three models described in section 2 on a large set of pace-mapping data obtained from 39 patients. Specifically, the localization accuracy is measured in terms of the Euclidean distance between the predicted coordinates and those recorded during the pace-mapping procedure. Additionally, we compare the localization accuracy between 1) a deep network with standard DAE without separating factors of variations and 2) a deep network with the two-way factored DAE in first layer followed by standard DAEs and 3) separating factors of variation throughout all the layers.

3.1. Data processing

All ECG data are pre-processed for noise removal and baseline correction. Besides, the selection and extraction

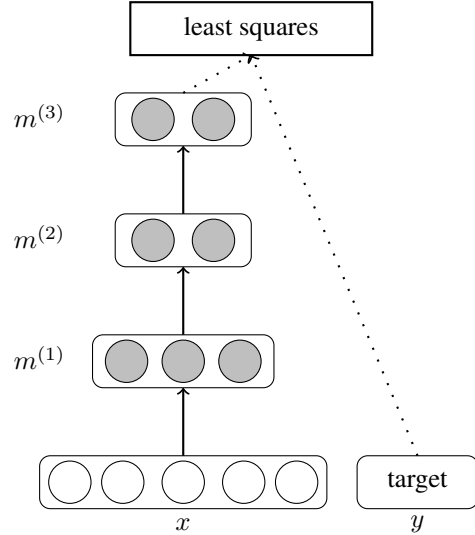


Figure 3. Fine-tuning of deep network for predicting the coordinate of VT exits. The parameters of the whole system are fine-tuned to minimize the mean square error between input and target using stochastic gradient descent. $m^{(k)}$ with $k = \{1,2,3\}$ represents the learned representation at k^{th} layer.

of QRS complexes are manually carried out to avoid motion artifacts, ectopic beats and non-capture beats. The final input signal is in the form of one QRS beat from 12 leads, each down-sampled in time to 100. Moreover, as multiple quality beats can be extracted from each ECG recording, we obtain in-total 16848 sets of ECG data with x - y - z coordinates labeled for each pacing site.

3.2. Model training

The dataset is divided into training, validation and test set: 10292, 3017 and 3539, respectively containing data from 22, 5 and 12 patients. Data from same patient is ensured to belong to only one of the three sets. The fine-tuned network consists of three hidden layers and each layer is pre-trained with both standard DAE and factor-disentangling DAE. Hyperparameters including learning rate for SGD and noise level for input corruption are selected based on their performance on the validation set. After pre-training, the whole network is fine-tuned on the clean training set as described in Section 2.4.

3.3. Results and discussion

The accuracy of the coordinate prediction by separating factors in the deep hierarchy (Deep-fDAEs) along with the other three models are presented in Table 1. Compared with the use of pre-scribed QRS-integrals as input features for linear regression, high-level features learned from the deep network of standard DAEs (sDAE) improved the prediction accuracy by approximately 3 millimeters ($\sim 23\%$)

Table 1. Prediction accuracy of the proposed approach versus three comparison methods. We report mean error (in millimeters) and its 95% confidence interval on the training, validation and test set.

Method	QRS integral	sDAE	fDAEs	Deep-fDAEs
training	15.06±0.13	11.27±0.11	11.75±0.11	10.89±0.11
validation	15.16 ±0.27	13.07±0.22	12.88±0.26	12.84±0.25
test	15.09±0.23	12.13 ±0.21	11.95±0.21	11.83±0.21

on the test set. By adding the factor-disentangling DAE in the first layer (fDAE), the prediction accuracy is further improved by $\sim 1\%$. When the factor-disentangling DAEs are stacked into a deep network, the prediction accuracy in overall is improved by 4 millimeters ($\sim 26\%$) at an average of 11.83 mm. In comparison, when localizing the exit of VT into one of the pre-defined ten [7] or sixteen segments [5], the resolution of localization cannot exceed $\sim 25\text{mm}$ or $\sim 18\text{mm}$ respectively. Therefore, an improvement in the resolution of localization was achieved compared to segment-based classification.

The large improvement of accuracy from a linear regression model to the deep DAE model demonstrates the potential of a deep learning model in automatically mining more discriminative features from ECG data compared to pre-scribed features. In comparison, the addition of factor-disentanglement achieved a smaller improvement of accuracy. This limited improvement can be associated with the fact that the model only saw 22 patients during training, a relatively small sample size for the purpose of learning inter-subject variations. Moreover, as the data are collected from actual pacing procedures during ablation, the number of pacing sites collected from each patient was limited and covered only a small portion of the myocardium in and around the region of scar. That means the training data for each patient is also limited not only in quantity but also in coverage of the prediction space. This further increases the difficulty of learning individual-level adjustments to the common relationship between QRS and the origin of ventricular activation. Future work will continue our effort in data collection on more patients in order to truly realize the potential of deep learning in this application.

4. Conclusion

In this paper, we demonstrate the use of a deep model to predict the x - y - z coordinate of the origin of ventricular activation in the left ventricular endocardium surface from 12-lead ECG data. We show that a coordinate-based prediction can improve the resolution of the localization. We also show that the use of a deep model can mine more discriminative features for better localization accuracy compared to pre-scribed features such as QRS-integrals. This improvement in resolution and accuracy may enhance the efficiency and efficacy of catheter ablation therapies.

Acknowledgements

This work is supported in part by the National Institutes of Health [No: R21HL125998], the National Science Foundation [No: ACI-1350374], the National Natural Science Foundation of China (No:61525106, 61427807) and the National Key Technology Research and Development Program of China (No:2016YFC1300302).

References

- [1] Stevenson WG. Ventricular scars and ventricular tachycardia. *Transactions of the American Clinical and Climatological Association* 2009;120:403.
- [2] Miller JM, Marchlinski FE, Buxton AE, Josephson ME. Relationship between the 12-lead electrocardiogram during ventricular tachycardia and endocardial site of origin in patients with coronary artery disease. *Circulation* 1988; 77(4):759–766.
- [3] Segal OR, Chow AWC, Wong T, Trevisi N, Lowe MD, Davies DW, Bella PD, Packer DL, Peters NS. A novel algorithm for determining endocardial vt exit site from 12lead surface eeg characteristics in human, infarctrelated ventricular tachycardia. *Journal of Cardiovascular Electrophysiology* 2007;18(2):161–168.
- [4] Yokokawa M, Liu TY, Yoshida K, Scott C, Hero A, Good E, Morady F, Bogun F. Automated analysis of the 12-lead electrocardiogram to identify the exit site of postinfarction ventricular tachycardia. *Heart Rhythm* 2012;9(3):330–334.
- [5] Sapp JL, El-Damaty A, MacInnis PJ, Warren JW, Horáček BM. Automated localization of pacing sites in postinfarction patients from the 12-lead electrocardiogram and body-surface potential maps. *Computing in Cardiology* 2012;.
- [6] Sapp JL, Bar-Tal M, Howes AJ, Toma JE, El-Damaty A, Warren JW, MaeInnis PJ, Zhou S, Horáček BM. Real-time localization of ventricular tachycardia origin from the 12-lead electrocardiogram. *JACC Clinical Electrophysiology* 2017;Accepted. In Press.
- [7] Chen S, Gyawali PK, Liu H, Horacek BM, Sapp JL, Wang L. Disentangling inter-subject variations: Automatic localization of ventricular tachycardia origin from 12-lead electrocardiograms. In *Biomedical Imaging (ISBI 2017)*, 2017 IEEE 14th International Symposium on. IEEE, 2017; 616–619.
- [8] Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*. ACM, 2008; 1096–1103.
- [9] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. In *Computer vision and pattern recognition, 2006 IEEE computer society conference on*, volume 2. IEEE, 2006; 1735–1742.

Address for correspondence:

Name: Prashna Kumar Gyawali

Full postal address: 1665 Crittenden Road, Rochester, NY, 14623

E-mail address: pkg2182@rit.edu