








ORIGINAL ARTICLE

Explainable localization of premature ventricular contraction using deep learning-based semantic segmentation of 12-lead electrocardiogram

Kota Kujime MS¹ | Hiroshi Seno PhD¹  | Kenzaburo Nakajima MD, PhD²  |
Masatoshi Yamazaki MD, PhD^{1,3}  | Ichiro Sakuma PhD¹  | Kenichiro Yamagata MD, PhD⁴  |
Kengo Kusano MD, PhD²  | Naoki Tomii PhD¹ 

¹Department of Precision Engineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

²Department of Cardiovascular Medicine, National Cerebral and Cardiovascular Center, Osaka, Japan

³Department of Cardiology, Nagano Hospital, Okayama, Japan

⁴Department of Cardiovascular Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Correspondence

Naoki Tomii, Department of Precision Engineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan.

Email: naoki_tomii@bmpe.t.u-tokyo.ac.jp

Funding information

Japan Society for the Promotion of Science, Grant/Award Number: 21H04953 and 21K18036

Abstract

Background: Predicting the origin of premature ventricular contraction (PVC) from the preoperative electrocardiogram (ECG) is important for catheter ablation therapies. We propose an explainable method that localizes PVC origin based on the semantic segmentation result of a 12-lead ECG using a deep neural network, considering suitable diagnosis support for clinical application.

Methods: The deep learning-based semantic segmentation model was trained using 265 12-lead ECG recordings from 84 patients with frequent PVCs. The model classified each ECG sampling time into four categories: background (BG), sinus rhythm (SR), PVC originating from the left ventricular outflow tract (PVC-L), and PVC originating from the right ventricular outflow tract (PVC-R). Based on the ECG segmentation results, a rule-based algorithm classified ECG recordings into three categories: PVC-L, PVC-R, as well as Neutral, which is a group for the recordings requiring the physician's careful assessment before separating them into PVC-L and PVC-R. The proposed method was evaluated with a public dataset which was used in previous research.

Results: The evaluation of the proposed method achieved neutral rate, accuracy, sensitivity, specificity, F1-score, and area under the curve of 0.098, 0.932, 0.963, 0.882, 0.945, and 0.852 on a private dataset, and 0.284, 0.916, 0.912, 0.930, 0.943, and 0.848 on a public dataset, respectively. These quantitative results indicated that the proposed method outperformed almost all previous studies, although a significant number of recordings resulted in requiring the physician's assessment.

Conclusions: The feasibility of explainable localization of premature ventricular contraction was demonstrated using deep learning-based semantic segmentation of 12-lead ECG.

Clinical trial registration: M26-148-8.

Kota Kujime and Hiroshi Seno contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Journal of Arrhythmia* published by John Wiley & Sons Australia, Ltd on behalf of Japanese Heart Rhythm Society.

KEYWORDS

automatic ECG diagnosis, deep neural network, premature ventricular contraction

1 | INTRODUCTION

Premature ventricular contraction (PVC) is known as a common arrhythmia. Repeated PVCs increase the risk of developing arrhythmias or cardiomyopathy that causes heart failure.^{1,2} PVCs are caused by abnormal excitations, which mainly originate from the right and left ventricular outflow tracts (RVOT and LVOT).^{3,4} Catheter ablation (CA) is a treatment for patients with PVC whose drug therapy is ineffective or discontinued due to unacceptable side effects.⁵ Localizing the PVC origin from the preoperative electrocardiogram (ECG) before the CA is important because it will help in choosing the appropriate approach for catheter preparation leading to avoid complications, saving procedural time, improving the success rate, and suppressing recurrent risk.^{6–10}

Various approaches have been proposed so far to assist physicians in localizing PVC origins from preoperative ECG. Rule-based algorithms using indices extracted manually from ECG waveforms have been proposed in many studies.^{6,9,11–13} For example, Yoshida et al. proposed an index calculated from the S-wave of V2 lead and the R-wave of V3 lead.¹² Some validities of those methods have been demonstrated, but the clinical application remains difficult because those require manual measurement of ECG waveform features by physicians. Moreover, these algorithms are limited to PVCs with typical waveforms because manually extracted indices and simple dichotomization are inadequate for accurate estimation.⁴

Machine learning (ML) approaches have been proposed in localizing PVC origins to overcome those limitations of rule-based algorithms, thereby expecting automatic extraction of effective ECG features.^{4,8} For example, Zheng et al. proposed the extreme gradient boosting tree model which identifies PVC origin as LVOT or RVOT, and the model achieved an accuracy of 0.976 which significantly exceeds that of manual algorithms.⁸

ML-based methods outperform rule-based algorithms, but accountability is a major and general problem for clinical application of many ML-based diagnosis methods because physicians are ultimately responsible for the diagnosis.¹⁴ Explainable models that show which features of the image were focused on using gradients of the neural network have been proposed in the field of medical image analysis.¹⁵ Conversely, ML models for localizing PVC origin output only the PVC origin outcome, and these models are supposed to be less credible for surgeons and not suitable for supporting diagnosis because of the inadequate basis of outputs.

This study aimed to realize an ML-based explainable model for localizing PVC origins from preoperative 12-lead ECG. To realize explainable localization of PVC origins, we propose a deep learning-based semantic segmentation model which classifies every sampling time of ECG signal into four categories: background (BG), sinus

rhythm (SR), PVC originating from LVOT (PVC-L), and PVC originating from RVOT (PVC-R) and a rule-based algorithm that localizes PVC origin as LVOT, RVOT, or Neutral, which is a group for the recordings requiring the physician's careful assessment before separating them into PVC-L and PVC-R (Figure 1). In this study, the accuracy of the proposed method was evaluated based on the clinical preoperative 12-lead ECG signals of patients with PVC.

2 | METHODS

2.1 | Study design

The deep learning-based semantic segmentation model was trained and evaluated on our private dataset, which includes 265 ECG recordings from 84 patients with frequent PVCs. The dataset was collected at the National Cerebral and Cardiovascular Center (NCVC dataset). Furthermore, inference using the trained model with the NCVC dataset was conducted on the public dataset used in the prior study, and the accuracy was compared with that of previous studies. This retrospective analysis of patients' clinically acquired data was approved by the ethics committee of the NCVC (M26-148-8), and all patients provided written informed consent before undergoing ablation.

2.2 | Patient selection

This study included 84 patients (37 males and 47 females; mean age 50.1 ± 17.7 years) with frequent PVCs who underwent successful CA at the NCVC from November 2010 to December 2019. A total of 14 patients had dilated cardiomyopathy, 1 had ischemic cardiomyopathy and others had no apparent heart disease. The frequency of PVC ranged from 12% to 32%, with a median of 19%. A 3D mapping system was utilized in the CA surgeries, and the origin of the PVC was judged where suppression of the PVC was finally achieved by applying radiofrequency energy. As a result of CA, 37 patients had PVC-L and 47 patients had PVC-R.

2.3 | ECG recording and measurement

Preoperative ECGs of patients with PVC were recorded in the supine position and acquired at a sampling rate of 500 Hz using LS9900 OS win 7 (Ver: 2.8a) and a CLEARSIGN amplifier (Boston Scientific, MA, USA). A total of 265 ECG recordings from 84 patients were collected, and each recording was 10 s long with a sampling rate of 500 Hz and consisted of 12 leads. Among them, 101 recordings were

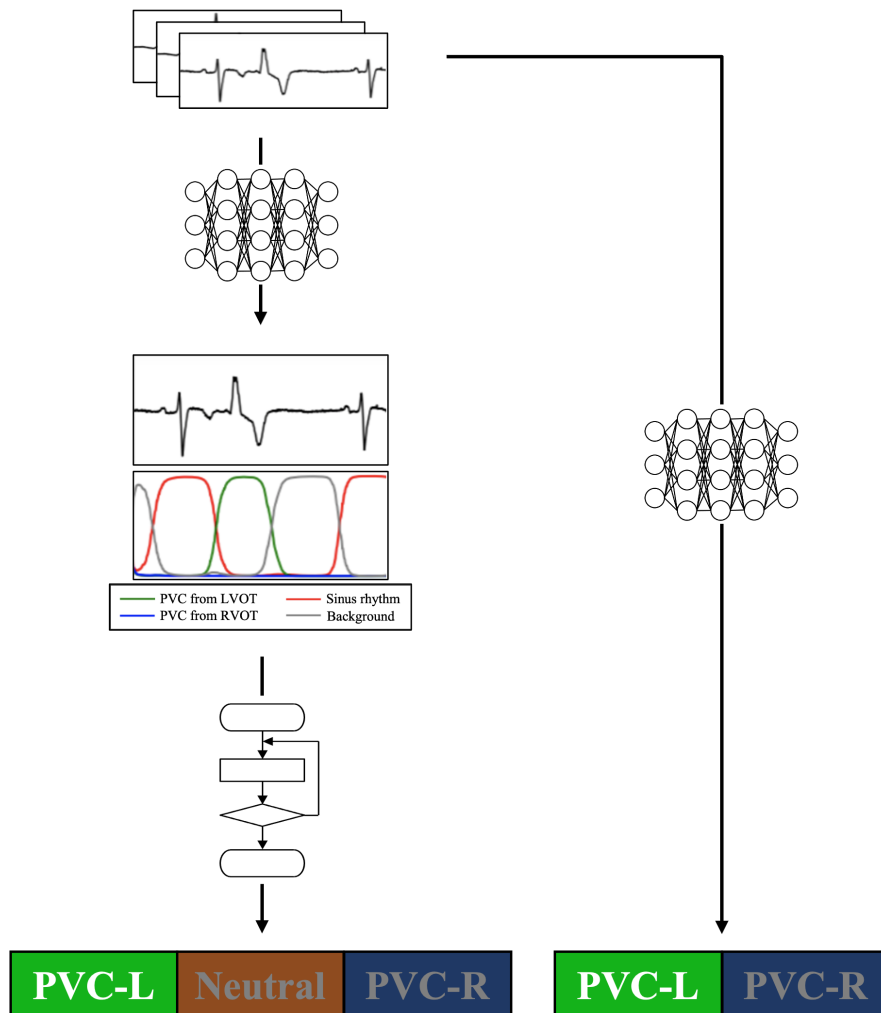


FIGURE 1 Comparison of the localization process between the proposed method and ML-based methods of previous studies. ML-based methods of previous studies determine PVC originating from LVOT (PVC-L) or PVC originating from RVOT (PVC-R) from the input ECG. Conversely, the proposed method performs deep learning-based semantic segmentation on the input ECG which classifies every sampling time into four categories (background, SR, PVC-L, and PVC-R) and a rule-based algorithm which classifies the input ECG into three categories (PVC-L, PVC-R, and Neutral) based on the preceding segmentation results. The semantic segmentation results can be used as the basis of the diagnosis.

from 37 patients with PVC-L and 164 recordings were from 47 patients with PVC-R.

2.4 | Dataset preparation

The ground truth label (BG, SR, PVC-L, and PVC-R) of each sampling time for semantic segmentation was annotated manually by a clinical expert and the ground truth label (PVC-L and PVC-R) of each recording was annotated based on the CA surgery outcome.

This study performed fivefold cross-validation to prevent variations in learning accuracy depending on the choice of data. The dataset was divided into five groups, ensuring that recordings from the same patient belonged to the same group and the number of recordings and proportion of PVC origin were equal among the groups. [Table 1](#) shows the number of recordings and patients in each group. One group was randomly selected as validation data from four groups excluding the test data in each dataset division pattern of fivefold validation, and the remaining three groups were used as training data.

In our preliminary investigation, it was observed that the accuracy of ECG semantic segmentation did not significantly differ

TABLE 1 Number of patients and recordings in fivefold validation.

	LVOT		RVOT	
	Patients	Recordings	Patients	Recordings
Group 1	7	20	9	33
Group 2	8	21	9	32
Group 3	7	20	10	33
Group 4	7	20	10	33
Group 5	8	20	9	33

between the use of chest leads alone and the inclusion of limb leads. Consequently, this study exclusively utilized chest leads and examined the optimization of their combination. [Table 2](#) shows the 14 different ECG combinations considered as inputs for a deep neural network (DNN) model, the combination of chest leads was selected to include leads focusing on the right ventricle as well as leads focusing on the left ventricle. A total of 70 patterns of the validation dataset and input ECG-lead combinations were examined. In each case, 3600 samples in the center were chosen out of 5000 samples by removing the 700 sampling at both ends, because annotated labels could be incorrect at both ends due to inadequate information.

Additionally, 3072 consecutive samples in each epoch were randomly selected out of 3600 samples.

2.5 | Development of DNN model for semantic segmentation

The developed DNN model was based on the ECG-SegNet proposed for semantic segmentation that classifies each ECG sampling time into several categories using bidirectional long short-term memory (BLSTM).¹⁶ The DNN model consisted of a 1-dimensional-convolutional neural network (1DCNN) and BLSTM which are effective for feature extraction of time series data. Figure 2 shows a detailed of the developed DNN architecture. In each layer of 1DCNN, the kernel size was set to seven, and an exponential linear unit was used as activation

function. Furthermore, a BLSTM layer with a dropout rate of 20% was implemented. Categorical cross-entropy error was used as the loss function, and an Adam optimizer with a learning rate of 0.001 was used as the optimization function. The batch size was set to 10, and the number of epochs was set to 300. The DNN model was applied to the validation data for each epoch, and the error between output and labels was calculated. The final weight of the DNN model adopted the DNN model weight of the epoch with the smallest validation error among the 300 epochs. All experiments were performed in Python using TensorFlow¹⁷ on a Linux Server (Ubuntu 18.04.3 LTS) with NVIDIA GeForce RTX 2080 GPU.

TABLE 2 Combinations of ECG-leads for estimation of PVC origin.

The number of leads	Leads pattern
1	V1
2	V1, V4
	V1, V5
	V1, V6
	V2, V4
	V2, V5
	V2, V6
3	V1, V3, V4
	V1, V3, V5
	V1, V3, V6
	V2, V3, V4
	V2, V3, V5
	V2, V3, V6
6	V1, V2, V3, V4, V5, V6

2.6 | Localization algorithm of PVC origin based on ECG segmentation

Recordings with difficult-to-determine PVC-L or PVC-R should not be forcibly classified as either one or the other, considering the reliable support of physicians in diagnosis, and such recordings should be assessed by physicians carefully. Therefore, a new class Neutral was added for these recordings, and the following rule-based algorithm classified each recording into three categories (PVC-L, PVC-R, and Neutral) based on ECG segmentation by a trained DNN model. The existence of a timing when the respective probabilities of PVC-L and PVC-R exceeded the threshold value in the analysis interval was determined for a given threshold value. The corresponding category was determined as the result of the recording classification if one of the results of either PVC-L or PVC-R was positive. Conversely, the recording was classified as Neutral if the result of both was positive or negative.

2.7 | Evaluation metrics

At first, the same algorithm, as mentioned above, was applied to the ground truth label of the segmentation to determine the ground

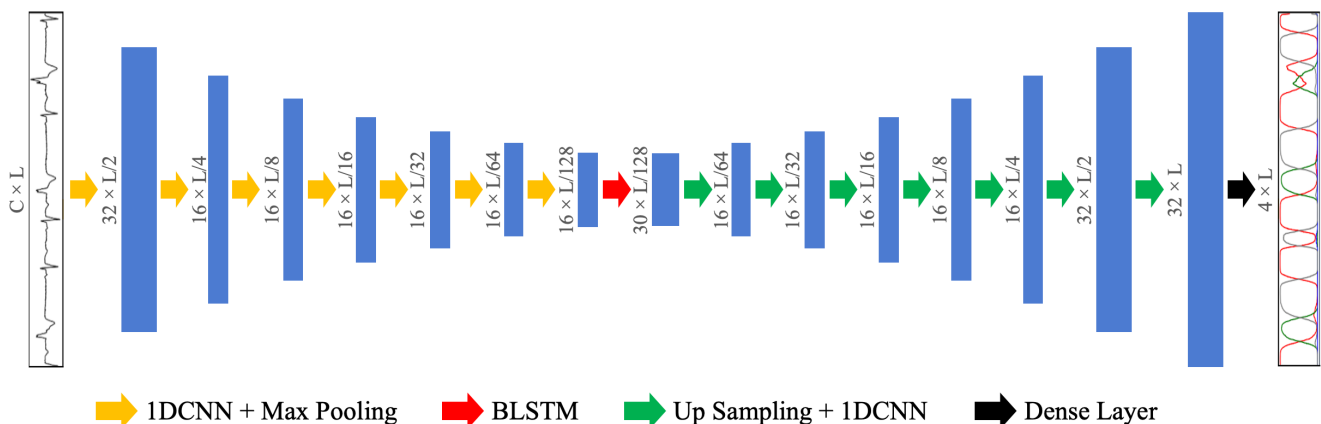


FIGURE 2 Architecture of the deep neural network (DNN) model. The DNN model classifies each ECG sampling time into four categories (background, SR, PVC originating from LVOT, and PVC originating from RVOT) via a 1-dimensional-convolutional neural network (1DCNN), bidirectional long short-term memory (BLSTM), and dense layer from the input ECGs of size $C \times L$, where C is the number of ECG leads and L is the number of ECG samplings.

truth of the recordings. The labels of the recordings were determined uniquely regardless of the threshold because the probability of the label is always 0 or 1. Then, the following metrics were used to evaluate the trained model and the localization algorithm.

Neutral rate (NR), which is the ratio of recordings classified as Neutral out of the recordings with the ground truth label of determinable ("PVC-L" or "PVC-R"), was evaluated for each trained DNN model. Toward the recordings classified as determinable, accuracy (ACC), sensitivity to "PVC-R" (SE), specificity (SP), F1-score, and the area under the curve (AUC) of the receiver operating characteristic (ROC) curve were calculated as the metrics of the classification among PVC-L and PCV-R.

2.8 | Evaluation using a public dataset

We evaluated the proposed method with a public dataset used in previous studies to compare their performance. The dataset consisted of 334 12-lead ECGs recorded at Ningbo First Hospital of Zhejiang University from 334 patients and ground truth labels (Ningbo dataset).¹⁸ The ground truth labels for the Ningbo dataset were determined based on the outcomes of CA procedures, with 77 recordings classified as PVC-L and 257 recordings classified as PVC-R.

The best combination of the lead patterns of the DNN model and the threshold of the localization algorithm for the NCVC dataset was selected to compare with previous studies. A fivefold average of NR and ACC for NCVC dataset was calculated for 14 different models with different lead patterns by varying the threshold of the localization algorithm from 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 0.95. Then, we selected the best combination that minimizes the loss in Equation (1) out of them with NR of <0.3 and ACC of >0.9.

$$\text{Loss} = (1 - \text{NR})(1 - \text{ACC}) + \lambda \text{NR} \quad (1)$$

In this study, λ was set to 0.1 because it is preferable to classify as Neutral than to make a wrong decision for recordings with difficult-to-determine PVC-L or PVC-R, considering the reliable diagnosis support.

The DNN model with the lead patterns of the best combination was trained again with whole NCVC dataset. The NCVC dataset was split into training data and validation data. The training data consisted of 88 recordings from 31 patients with PVC-L and 151 recordings from 40 patients with PVC-R. The validation data consisted of 13 recordings from six patients with PVC-L and 13 recordings from seven patients with PVC-R. The detailed learning curve is presented in Figure S1.

Then, the trained DNN model with the private dataset was evaluated using the public Ningbo dataset. The above-mentioned evaluation metrics (NR, ACC, SE, SP, F1-score, and AUC) were calculated. The recordings with a sampling rate of 2000 Hz were downsampled uniformly to 1/4 so that the sampling rate was the same as 500 Hz of training data when inputting ECG recordings of the Ningbo dataset into the trained DNN model.

3 | RESULTS

3.1 | DNN model performance when varying ECG-lead combinations and thresholds

Figure 3A–C shows the change in NR, ACC, and loss defined in Equation (1). The results of the lead combination, which has the smallest loss for each number of input leads, were shown in these figures. Both NR and ACC tended to decrease as the threshold was increased, as shown in Figure 3A,B. Moreover, NR tended to decrease and ACC tended to increase as the number of input leads increased, but using all six ECG-leads as input was not best for the DNN model. Figure 3C shows the change of loss under conditions of NR of <0.3 and ACC of >0.9. The best combination was the lead patterns of V2 and V6 lead and threshold of 0.4, as shown in Figure 3C. The proposed method with the best condition (leads: V2 and V6, threshold: 0.4) achieved NR of 0.098; ACC of 0.932; SE to PVC-R of 0.963; SP of 0.882; F1-score of 0.945 on the NCVC dataset. Figure 3D shows the ROC curve of the best performance proposed method for one of the fivefold validation, and the average AUC was 0.852.

3.2 | Inference examples of the proposed method

Figure 4 shows segmentation examples performed on the test data of the NCVC dataset using the trained DNN model with V2 and V6 leads. The top and middle images represent the ECG waveforms of V2 and V6 in each panel, respectively, which are inputs to the DNN model, and the bottom image represents the predicted probability of the presence of PVC-L and PVC-R by the trained DNN model. Figure 4A,B shows the trained DNN model predictions for typical ECG waveforms of PVC-L and PVC-R, respectively. These results revealed that the trained DNN model appropriately increased the probability of the correct category at the time of PVC occurrence while maintaining the probability of the incorrect category at almost zero. Figure 4C shows the inference results of the trained DNN model for a recording with both PVC-L and PVC-R typical waveforms. Hence, the probabilities of both PVC-L and PVC-R increased slightly at 2–3 s, but the probabilities each increased appropriately at the onset of the typical waveform. In this example, the highest probabilities of both PVC-L and PVC-R in the recording were nearly 1, and this recording was classified as Neutral for almost all threshold settings. Figure 4D shows a recording with PVC-L atypical waveforms and the prediction results of the trained DNN model. In this example, not only the PVC-L probability but also the PVC-R probability increased at the time of PVC occurrence, and the classification result for the recording was Neutral when the threshold was <0.636 and PVC-L when the threshold was >0.636.

3.3 | Performance comparison with previous studies

Table 3 shows the quantitative comparison for NR, ACC, SE, SP, F1-score, and AUC of the proposed method in this study and the

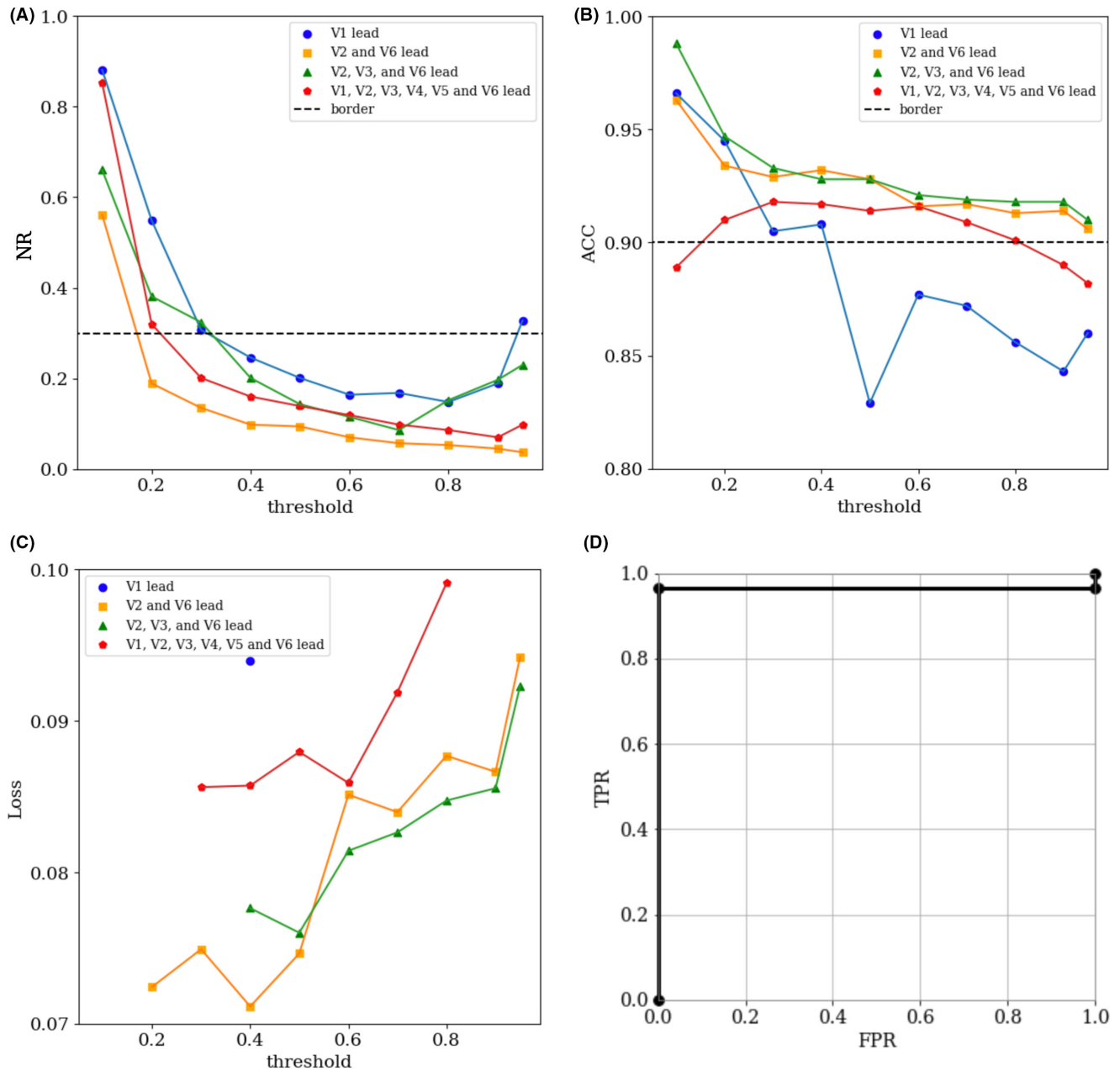


FIGURE 3 Performance of the proposed method. (A–C) Changes in Neutral rate (NR), Accuracy (ACC), and Loss defined in Equation (1) relative to threshold. From (C), the best combination was the leads pattern of V2 and V6 leads and threshold of 0.4, and the best performance DNN model achieved NR of 0.098, ACC of 0.932, SE to PVC-R of 0.963, SP of 0.882, and F1-score of 0.945. (D) ROC curve of the best performance DNN model (leads: V2 and V6, threshold: 0.4). The average AUC of the best performance DNN model was 0.852.

conventional methods of previous studies. The proposed method with V2 and V6 leads and a threshold of 0.4 achieved a NR of 0.284, ACC of 0.916, SE of 0.912, SP of 0.930, F1-score of 0.943, and AUC of 0.848 on the public Ningbo dataset. The proposed method outperformed most previous studies in all ACC, SE, SP, F1-score, and AUC although 28.4% of the recordings were classified as Neutral.

4 | DISCUSSION

4.1 | Explainable localization of PVC

This study proposed a deep learning-based segmentation model and a rule-based algorithm for explainable PVC origin localization. The proposed method provides the physician with the prediction

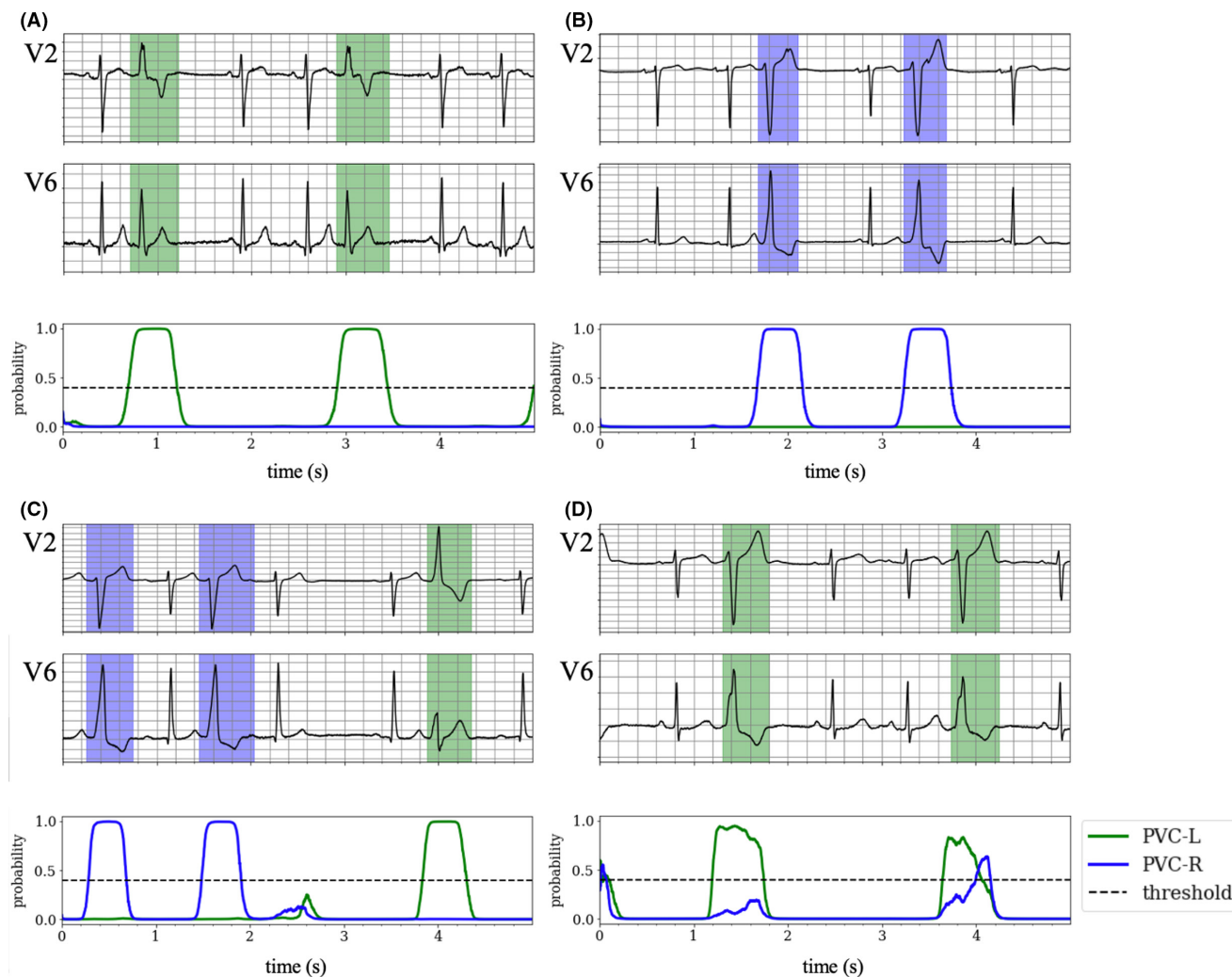


FIGURE 4 The segmentation example performed on the test data of the private dataset using the trained DNN model with V2 and V6 leads. (A, B) Typical ECG wave forms of PVC-L and PVC-R and the segmentation output of the trained DNN model. These results revealed that the trained DNN model only changed the segmentation probability of the correct category drastically while the segmentation probability of the other category was kept at almost zero. (C) A prediction example by the trained DNN model for a recording which has both PVC-L and PVC-R typical wave forms, and the probabilities of PVC-L and PVC-R increased appropriately at the onset of the typical waveform. The highest probability values for PVC-L and PVC-R in the recording were nearly 1, so this recording was classified as Neutral for almost all threshold values. (D) Atypical ECG wave forms of PVC-L and a prediction by the trained DNN model. The probabilities of both PVC-L and PVC-R increased with the timing of PVC occurrence in this case. The recording was classified as Neutral and PVC-L when thresholds were smaller and larger than 0.636, respectively.

result of PVC origin and the segmentation results, while conventional DNN models only output the PVC localization results. The segmentation results can be the basis of the diagnosis, and the proposed method can significantly increase explainability of automatic PVC origin localization. Moreover, the proposed method achieved ACC of 0.916, SE of 0.912, SP of 0.930, F1-score of 0.943, and AUC of 0.848 on the public dataset, which indicated that it outperformed almost all previous studies, although it excluded 28.4% of recordings as Neutral. Therefore, the feasibility of explainable PVC localization using deep learning-based semantic segmentation of 12-lead ECG was shown.

4.2 | Classification into the Neutral category

The three-class classification of PVC-L, PVC-R, and Neutral is one of the features of this study. For example, conventional ML-based methods had to determine PVC-L or PVC-R even for the recording containing both PVC-L and PVC-R seen in Figure 4C and the recording with atypical PVC waveforms seen in Figure 4D, while the proposed method can classify the controversial recordings as Neutral. More reliable diagnostic assistance can be achieved by classifying such recordings as Neutral. A threshold set for the predicted probability of PVC-L and PVC-R can classify the controversial recordings

TABLE 3 Comparison of performance with previous studies.

Proposed method	NR	ACC	SE	SP	F1-score	AUC
This study	0.284	0.916	0.912	0.930	0.943	0.848
Kamakura et al. ¹⁹	0		0.301	0.633		
Zhang et al. ²⁰	0		0.80			
Betensky et al. ¹³	0	0.87	0.78	0.90		
Yoshida et al. ²¹	0		0.70	0.86		
Cheng et al. ²²	0	0.88	0.76	0.91		
Yoshida et al. ¹²	0		0.76	0.53		
Nakano et al. ²³	0		0.53	0.46		
Cheng et al. ²⁴	0	0.58	0.59	0.58		
He et al. ⁶	0		0.78	0.67		0.01
Xie et al. ²⁵	0		0.68	0.58		
Di et al. ⁹	0	0.70	0.70	0.68		
Zheng et al. ⁸	0	0.98	0.97	1.00	0.98	0.99

as Neutral even in previous studies. However, this approach lacks explainability because it does not indicate how it came to the Neutral classification. Conversely, the classification of PVC-L, PVC-R, and Neutral in the proposed method is made by a rule-based algorithm based on the segmentation result, so the reason for classification can be traced. Therefore, one of the advantages of the proposed method is that it can make an explainable classification as Neutral even for controversial recordings.

4.3 | Performance comparison with previous studies

Table 3 shows the performance comparison (NR, ACC, SE, SP, F1-score, and AUC) of the proposed method and conventional methods for the public Ningbo dataset. The results revealed that the proposed method outperformed most conventional methods, while the previous study by Zheng et al. reported almost 100% for all evaluation metrics, which seems to indicate that the classification performance is higher than the proposed method. However, such near 100% performance is rather unusual to be achieved in the presence of controversial data, such as those shown in Figure 4C,D. Additionally, the proposed method of classifying such data as Neutral is considered proper in terms of diagnostic support.

4.4 | Balance between classification accuracy and NR

We used Loss (Equation 1) to select the optimal ECG-leads combination and threshold, as shown in Figure 3C. The Loss is the sum of a term related to classification accuracy and a term related to NR, and the balance between these two terms is determined by λ . A small value of λ allows the selection of a DNN model that prioritizes classification accuracy, while a large value of λ allows

the selection of a DNN model that prioritizes NR reduction. The models selected with different λ could be utilized for stepwise screening. Controversial recordings that should be confirmed by a physician may be extracted by gradually applying from a model with high accuracy to a model with low NR if the number of recordings that can be diagnosed by a physician is determined in advance. This combination of various models can achieve a highly reliable and efficient diagnosis.

4.5 | Limitations

This study proposed a deep learning-based semantic segmentation for every time sampling of ECG to establish an explainable automatic diagnosis of PVC origin localization. However, no public dataset with segmentation labels existed, so we assigned segmentation labels to the NCVC dataset and used them for training the DNN model. Therefore, the data used for training is limited to 265 ECG recordings of 84 patients, but the proposed method is expected to gain further accuracy and generalization performance by increasing the number of training data in the future. Additionally, this study conducted the comparison with previous studies using the public Ningbo dataset to evaluate the generalization performance of the proposed method. The Ningbo dataset was used for comparison with the latest previous studies, but this dataset also does not have a huge amount of data, and the generalization performance of the proposed method has not been fully examined.

Furthermore, this study investigated the optimal combinations of ECG leads for identifying PVC origins, and it was demonstrated that when using V2 and V6 leads, estimation of PVC origins with an accuracy, sensitivity, and specificity of approximately 0.9 was achieved. Since our preliminary experiments did not show a significant improvement in accuracy with the addition of limb leads, it is considered reasonable to use only chest leads to reduce the total number of lead combinations for optimization. However, since limb leads can capture excitation conduction from different

directions than those of chest leads, effectively utilizing information from limb leads may further enhance the accuracy of PVC origin estimation.

Moreover, this study proposed the Neutral classification for a controversial recording but did not consider the appropriate treatment for the cases classified as Neutral. We may consider the explainable treatment strategy by analyzing features, such as maximum and/or mean for the segmentation probabilities of PVC-L and PVC-R, if we have to determine whether PVC originating from LVOT or RVOT, as in the conventional method. Conversely, recordings classified as Neutral are likely to be atypical recordings as seen in Figure 4C,D. Additionally, such atypical recordings may differ significantly from typical PVC-L and PVC-R in the PVC location occurrence from an electrophysiological point of view. Therefore, other appropriate ablation strategies for cases classified as Neutral may be necessary to achieve a more sophisticated PVC ablation treatment.

5 | CONCLUSIONS

This study proposed a deep learning-based semantic segmentation on ECG signal and a rule-based localization algorithm based on the preceding segmentation results to realize an ML-based explainable model for localizing PVC origins from preoperative 12-lead ECG. The proposed method achieved NR of 0.284, ACC of 0.916, SE to PVC-R of 0.912, SP of 0.930, F1-score of 0.943, and AUC of 0.848 on the public dataset and outperform almost all the previous studies while it required physician's assessment for some of the ambiguous recordings. The feasibility of explainable PVC localization using deep learning-based semantic segmentation of 12-lead ECG was indicated.

ACKNOWLEDGMENTS

K. Kujime and H. Seno contributed equally to this work. This study was supported by JSPS KAKENHI 21K18036 to Dr. Tomii, and 21H04953 to Dr. Sakuma.

FUNDING INFORMATION

This study was supported by a Grant-in Aid for Scientific Research 21K18036 to Dr. Tomii, and 21H04953 to Dr. Sakuma, from the Japan Society for Promotion of Science.

CONFLICT OF INTEREST STATEMENT

None.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author, N.T. The data are not publicly available due to restrictions.








ETHICS STATEMENT

The study was approved by the ethics committee of the National Cerebral and Cardiovascular Center.

PATIENT CONSENT STATEMENT

Written informed consent was obtained from all participants prior to study inclusion.

ORCID

Hiroshi Seno  <https://orcid.org/0000-0002-7631-0940>
 Kenzaburo Nakajima  <https://orcid.org/0000-0002-1279-6558>
 Masatoshi Yamazaki  <https://orcid.org/0000-0003-3553-7193>
 Ichiro Sakuma  <https://orcid.org/0000-0002-4919-9466>
 Kenichiro Yamagata  <https://orcid.org/0000-0002-9571-7017>
 Kengo Kusano  <https://orcid.org/0000-0002-5760-9285>
 Naoki Tomii  <https://orcid.org/0000-0002-5485-4883>

REFERENCES

- de Marco F, Finlay D, Bond RR. Classification of premature ventricular contraction using deep learning. *Comput Cardiol*. 2010;2020:1-4. <https://doi.org/10.22489/CinC.2020.311>
- Talbi ML, Charef A. PVC discrimination using the QRS power spectrum and self-organizing maps. *Comput Methods Prog Biomed*. 2009;94(3):223-31. <https://doi.org/10.1016/j.cmpb.2008.12.009>
- John RM, Stevenson WG. Outflow tract premature ventricular contractions and ventricular tachycardia: the typical and the challenging. *Card Electrophysiol Clin*. 2016;8(3):545-54.
- Zhao W, Zhu R, Zhang J, Mao Y, Chen H, Ju W, et al. Machine learning for distinguishing right from left premature ventricular contraction origin using surface electrocardiogram features. *Heart Rhythm*. 2022;19(11):1781-9. <https://doi.org/10.1016/j.hrthm.2022.07.010>
- Cronin EM, Bogun FM, Maury P, Peichl P, Chen M, Namboodiri N, et al. 2019 HRS/EHRA/APHRS/LAHR expert consensus statement on catheter ablation of ventricular arrhythmias. *Europace*. 2019;21(8):1143-4. <https://doi.org/10.1093/europace/euz132>
- He Z, Liu M, Yu M, Lu N, Li J, Xu T, et al. An electrocardiographic diagnostic model for differentiating left from right ventricular outflow tract tachycardia origin. *J Cardiovasc Electrophysiol*. 2018;29(6):908-15. <https://doi.org/10.1111/jce.13493>
- Efimova E, Dinov B, Acou WJ, Schirripa V, Kornej J, Kosiuk J, et al. Differentiating the origin of outflow tract ventricular arrhythmia using a simple, novel approach. *Heart Rhythm*. 2015;12(7):1534-40. <https://doi.org/10.1016/j.hrthm.2015.04.004>
- Zheng J, Fu G, Abudayyeh I, Yacoub M, Chang A, Feaster WW, et al. A high-precision machine learning algorithm to classify left and right outflow tract ventricular tachycardia. *Front Physiol*. 2021;12:641066. <https://doi.org/10.3389/fphys.2021.641066>
- di C, Wan Z, Tse G, Letsas KP, Liu T, Efremidis M, et al. The V1-V3 transition index as a novel electrocardiographic criterion for differentiating left from right ventricular outflow tract ventricular arrhythmias. *J Interv Card Electrophysiol*. 2019;56(1):37-43. <https://doi.org/10.1007/s10840-019-00612-0>
- Kim RJ, Iwai S, Markowitz SM, Shah BK, Stein KM, Lerman BB. Clinical and electrophysiological spectrum of idiopathic ventricular outflow tract arrhythmias. *J Am Coll Cardiol*. 2007;49(20):2035-43. <https://doi.org/10.1016/j.jacc.2007.01.085>
- Ito S, Tada H, Naito S, Kurosaki K, Ueda M, Hoshizaki H, et al. Development and validation of an ECG algorithm for identifying the optimal ablation site for idiopathic ventricular outflow tract tachycardia. *J Cardiovasc Electrophysiol*. 2003;14(12):1280-6. <https://doi.org/10.1046/j.1540-8167.2003.03211.x>
- Yoshida N, Yamada T, McElderry HT, Inden Y, Shimano M, Murohara T, et al. A novel electrocardiographic criterion for differentiating a left from right ventricular outflow tract tachycardia origin: the

- V2s/V3R index. *J Cardiovasc Electrophysiol*. 2014;25(7):747–53. <https://doi.org/10.1111/jce.12392>
13. Betensky BP, Park RE, Marchlinski FE, Hutchinson MD, Garcia FC, Dixit S, et al. The V(2) transition ratio: a new electrocardiographic criterion for distinguishing left from right ventricular outflow tract tachycardia origin. *J Am Coll Cardiol*. 2011;57(22):2255–62. <https://doi.org/10.1016/j.jacc.2011.01.035>
 14. Singh A, SenGupta S, Lakshminarayanan V. Explainable deep learning models in medical image analysis. *J Imaging*. 2020;6(6):1–19. <https://doi.org/10.3390/JIMAGING6060052>
 15. Couteaux V, Nempont O, Pizaine G, Bloch I. Towards interpretability of segmentation networks by analyzing deepDreams. *Lecture Notes in Computer Science*. Vol 11797. New York: Springer International Publishing; 2019. p. 56–63. https://doi.org/10.1007/978-3-030-33850-3_7
 16. Abrishami H, Campbell M, Czosek R. Supervised ECG interval segmentation using LSTM neural network. *International Conference on Bioinformatics and Computational Biology BIOCMPUT_18*, August, pp. 71–77, 2018 [Online]. <https://csce.ucmss.com/cr/books/2018/LFS/CSREA2018/BIC3264.pdf>
 17. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. 2016 [Online]. <http://arxiv.org/abs/1603.04467>
 18. Zheng J, Fu G, Anderson K, Chu H, Rakovski C. A 12-Lead ECG database to identify outflow tract origins of idiopathic ventricular arrhythmia containing more than 300 patients. *Figshare* 2019.
 19. Kamakura S, Shimizu W, Matsuo K, Taguchi A, Suyama K, Kurita T, et al. Localization of optimal ablation site of idiopathic ventricular tachycardia from right and left ventricular outflow tract by body surface ECG. *Circulation*. 1998;98(15):1525–33. <https://doi.org/10.1161/01.CIR.98.15.1525>
 20. Zhang F, Chen M, Yang B, Ju W, Chen H, Yu J, et al. Electrocardiographic algorithm to identify the optimal target ablation site for idiopathic right ventricular outflow tract ventricular premature contraction. *Europace*. 2009;11(9):1214–20. <https://doi.org/10.1093/europace/eup231>
 21. Yoshida N, Inden Y, Uchikawa T, Kamiya H, Kitamura K, Shimano M, et al. Novel transitional zone index allows more accurate differentiation between idiopathic right ventricular outflow tract and aortic sinus cusp ventricular arrhythmias. *Heart Rhythm*. 2011;8(3):349–56. <https://doi.org/10.1016/j.hrthm.2010.11.023>
 22. Cheng Z, Cheng K, Deng H, Chen T, Gao P, Zhu K, et al. The R-wave deflection interval in lead V3 combining with R-wave amplitude index in lead V1: a new surface ECG algorithm for distinguishing left from right ventricular outflow tract tachycardia origin in patients with transitional lead at V3. *Int J Cardiol*. 2013;168(2):1342–8. <https://doi.org/10.1016/j.ijcard.2012.12.013>
 23. Nakano M, Ueda M, Ishimura M, Kajiyama T, Hashiguchi N, Kanaeda T, et al. Estimation of the origin of ventricular outflow tract arrhythmia using synthesized right-sided chest leads. *Europace*. 2014;16(9):1373–8. <https://doi.org/10.1093/europace/eut355>
 24. Cheng D, Ju W, Zhu L, Chen K, Zhang F, Chen H, et al. V3R/V7 index: a novel electrocardiographic criterion for differentiating left from right ventricular outflow tract arrhythmias origins. *Circ Arrhythm Electrophysiol*. 2018;11(11):e006243. <https://doi.org/10.1161/CIRCEP.118.006243>
 25. Xie S, Kubala M, Liang JJ, Hayashi T, Park J, Padros IL, et al. Lead I R-wave amplitude to differentiate idiopathic ventricular arrhythmias with left bundle branch block right inferior axis originating from the left versus right ventricular outflow tract. *J Cardiovasc Electrophysiol*. 2018;29(11):1515–22. <https://doi.org/10.1111/jce.13747>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kujime K, Seno H, Nakajima K, Yamazaki M, Sakuma I, Yamagata K, et al. Explainable localization of premature ventricular contraction using deep learning-based semantic segmentation of 12-lead electrocardiogram. *J Arrhythmia*. 2024;40:948–957. <https://doi.org/10.1002/joa3.13096>