




Classification of underlying paroxysmal supraventricular tachycardia types using deep learning of sinus rhythm electrocardiograms

DIGITAL HEALTH
Volume 10: 1–13
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241281200
journals.sagepub.com/home/dhj



Soonil Kwon^{1,*†}, Jangwon Suh^{2,*} , Eue-Keun Choi^{3,4} , Jimyeong Kim²,
Hojin Ju³, Hyo-Jeong Ahn³, Sunhwa Kim⁵, So-Ryoung Lee^{3,4} , Seil Oh^{3,4}
and Wonjong Rhee^{2,6}

Abstract

Background: Obtaining tachycardia electrocardiograms (ECGs) in patients with paroxysmal supraventricular tachycardia (PSVT) is often challenging. Sinus rhythm ECGs are of limited predictive value for PSVT types in patients without preexcitation. This study aimed to explore the classification of atrioventricular nodal reentry tachycardia (AVNRT) and concealed atrioventricular reentry tachycardia (AVRT) using sinus rhythm ECGs through deep learning.

Methods: This retrospective study included patients diagnosed with either AVNRT or concealed AVRT, validated through electrophysiological studies. A modified ResNet-34 deep learning model, pre-trained on a public ECG database, was employed to classify sinus rhythm ECGs with underlying AVNRT or concealed AVRT. Various configurations were compared using ten-fold cross-validation on the training set, and the best-performing configuration was tested on the hold-out test set.

Results: The study analyzed 833 patients with AVNRT and 346 with concealed AVRT. Among ECG features, the corrected QT intervals exhibited the highest area under the receiver operating characteristic curve (AUROC) of 0.602. The performance of the deep learning model significantly improved after pre-training, showing an AUROC of 0.726 compared to 0.668 without pre-training ($p < 0.001$). No significant difference was found in AUROC between 12-lead and precordial 6-lead ECGs ($p = 0.265$). On the test set, deep learning achieved modest performance in differentiating the two types of arrhythmias, with an AUROC of 0.708, an AUPRC of 0.875, an F1-score of 0.750, a sensitivity of 0.670, and a specificity of 0.649.

Conclusion: The deep-learning classification of AVNRT and concealed AVRT using sinus rhythm ECGs is feasible, indicating potential for aiding in the non-invasive diagnosis of these arrhythmias.

Keywords

Atrioventricular nodal reentry tachycardia, atrioventricular reentry tachycardia, deep learning, paroxysmal supraventricular tachycardia

Submission date: 30 March 2024; Acceptance date: 19 August 2024

¹Division of Cardiology, Department of Internal Medicine, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea

²Department of Intelligence and Information, Seoul National University, Seoul, Republic of Korea

³Division of Cardiology, Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea

⁴Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

⁵Division of Cardiology, Department of Internal Medicine, Presbyterian Medical Center, Jeonju, Republic of Korea

⁶Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Republic of Korea

* The two authors contributed equally.

† This work was conducted while the author was affiliated with Seoul National University Hospital.

Corresponding authors:

Eue-Keun Choi, Department of Internal Medicine, Seoul National University College of Medicine and Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea.
E-mail: choiek17@snu.ac.kr;

Wonjong Rhee, Department of Intelligence and Information, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea.
E-mail: wrhee@snu.ac.kr



Introduction

Tachyarrhythmia is a common type of cardiac arrhythmia and can be categorized as either supraventricular or ventricular tachyarrhythmia according to its origins. The electrocardiogram (ECG) is the most essential tool for diagnosing specific types of tachyarrhythmia. Except for unusual cases, most typical tachyarrhythmias can be confirmed by ECG if it is available during a tachycardia episode. Paroxysmal supraventricular tachycardia (PSVT) is a frequently observed cardiac arrhythmia in patients with structurally normal hearts. Since most patients with PSVT have intermittent tachycardia episodes, it is often difficult to diagnose the disease using a standard 12-lead ECG. While an ECG can confirm the presence of tachycardia, a detailed diagnosis of a specific PSVT type remains difficult due to the similar ECG characteristics shared among most PSVT types (i.e. regular, narrow, complex tachycardia). This nature of PSVT often necessitates multiple clinic visits and repetitive ECGs for the diagnosis. Therefore, an unmet need exists to detect underlying PSVT through sinus rhythm ECGs, which would significantly improve patient management.

Patients with manifest atrioventricular reentry tachycardia (AVRT) (i.e. those with a manifest accessory pathway) exhibit preexcitation during sinus rhythm, which is a diagnostic feature for predicting AVRT. However, not all PSVT types exhibit preexcitation. Among the various types of PSVT, atrioventricular nodal reentry tachycardia (AVNRT) and concealed AVRT are the most common.¹ They have a reentrant mechanism involving different circuits. AVNRT utilizes two atrioventricular nodal pathways, whereas AVRT employs both the atrioventricular nodal pathway and accessory pathway(s). If an accessory pathway only permits retrograde conduction, or if its antegrade conduction is significantly delayed compared to atrioventricular nodal conduction, no preexcitation is observed during sinus rhythm (i.e. “concealed”). Consequently, the sinus rhythm ECGs of both AVNRT and concealed AVRT typically present normal findings. Despite the differences in their mechanisms and therapeutic strategies, it remains challenging to distinguish between these two types of cardiac arrhythmias using sinus rhythm ECGs. As a result, patients with PSVT often require an electrophysiological study, which is an invasive test, for the confirmatory diagnosis.

Recent advancements in deep learning applied to 12-lead ECGs have demonstrated success across various domains of cardiovascular medicine, showing potential for detecting a range of arrhythmias.^{2–4} For instance, it has been shown that deep learning algorithms can classify ECGs with sinus rhythm, atrial fibrillation, atrial flutter, atrial tachycardia, pacing rhythm, and other conditions.^{5–7} In particular, a recent study highlighted the ability of deep-learning analysis of sinus rhythm ECGs to differentiate patients with PSVT from those without.⁸ Since deep learning algorithms automatically learn features from the provided dataset, we

hypothesize that it could differentiate between the ECGs of sinus rhythms in AVNRT and concealed AVRT by capturing yet unknown subtle ECG features. This study aims to explore the feasibility of a deep learning model to distinguish between patients with AVNRT and concealed AVRT using sinus rhythm ECGs alone.

Methods

Study design and population

Patients diagnosed with either AVNRT or concealed AVRT from Seoul National University Hospital between 2001 and 2023 were retrospectively investigated. A standard 12-lead ECG (10-second-long with a sampling rate of 500 Hz) of sinus rhythm from each patient was obtained prior to the electrophysiological study within 1 month. The diagnosis of AVNRT or concealed AVRT was confirmed through electrophysiological studies. If multiple ECGs were available for a single patient, the one closest to the date of the electrophysiological study was selected. Patients diagnosed with both AVNRT and concealed AVRT, or those with other types of cardiac arrhythmias were excluded from the analysis. This is because the deep learning model was trained to classify these two specific categories of ECGs and the rare cases of ECG with both underlying PSVT types could confuse the model and negatively impact its diagnostic performance. The dataset compiled for our study is referred to as the SNUH-PSVT dataset in the following sections.

In our study, a total of 1179 patients (833 in the AVNRT group and 346 in the concealed AVRT group) were analyzed. The SNUH-PSVT dataset was divided into two sets based on patients’ examination dates: samples collected from September 2001 to June 2021 served as the training set, while those from July 2021 to August 2023 comprised the hold-out test set. Each patient was assigned to either the training set or the test set, ensuring that no patient was included in both. The data from each patient only includes information available during the ECG measurement process, so our study design guarantees that test set performance reflects the model’s accuracy on unseen patients after the training phase. As a result, the training set contains 1001 samples (696 in AVNRT; 305 in concealed AVRT), and the test set comprises 178 samples (137 in AVNRT; 41 in concealed AVRT).

Methodology overview

In this study, we employed the transfer learning technique to enhance model performance given limited samples. Transfer learning is a technique that improves a model’s performance by learning knowledge from one task and

then applying it to a similar task in the same domain.⁹ This approach is particularly beneficial for ECG tasks when it is challenging to acquire a sufficiently large dataset with specific labels or when it is highly expensive to construct labels.¹⁰ Given that the training set comprised only 1001 samples of AVNRT and concealed AVRT, the dataset's size could potentially hinder the model's ability to learn distinguishing features between AVNRT effectively and concealed AVRT. In this context, transfer learning enables the model to learn more effectively and achieve better performance with fewer labeled samples.

Our research methodology consists of two phases: pre-training on large public datasets followed by fine-tuning on our SNUH-PSVT training dataset. A model was first pre-trained with large-scale public ECG datasets to learn the general structure of the ECG. Then, the model was fine-tuned to distinguish between sinus rhythm ECGs from patients with AVNRT and those with concealed AVRT. We employed 10-fold cross-validation to evaluate various configurations within the training set and assessed the performance of the final chosen configuration in the test set. The overall research design is depicted in Figure 1.

Pre-training and transfer learning

In this study, we used the PhysioNet/Computing in Cardiology Challenge 2021 datasets^{11,12} as a large-scale dataset for pre-training. The goal of the challenge is to solve a multilabel classification problem of cardiac abnormalities using 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead ECGs. The challenge made 88,253 annotated ECG recordings public, each identified with more than one diagnosis. The challenge selected 30 out of 133 diagnoses as target labels to assess the performance of each participating algorithm and considered four pairs of diagnoses identical, thereby reducing the total number of distinct labels to 26. Following the challenge's objective, we utilized these 26 diagnoses for pre-training. The full list of 26 diagnoses is presented in Table S1.

From the eight datasets provided by the challenge, we excluded the INCART¹³ and PTB¹⁴ datasets due to their low label variability and the longer duration of their recordings, which distinguished them from the others in terms of characteristics. Consequently, we utilized six subsets of the challenge datasets: Chapman-Shaoxing,¹⁵ CPSC,¹⁶ CPSC-Extra, G12EC, Ningbo,¹⁷ and PTB-XL.¹⁸ Samples from these datasets with target labels were randomly assigned to either a training or test set using an 80:20 split. This resulted in a total of 65,656 samples for training and 16,360 samples for testing. The occurrence of each label within these sets is detailed in Table S1.

For pre-training, we utilized complete 12-lead ECG recordings in the training set. Each recording was resampled at 500 Hz and randomly cropped to a 10-second window to align with the characteristics of our SNUH-PSVT dataset. For additional details, we primarily followed the experimental settings

of the previous research¹⁹; we applied a Butterworth bandpass filter with a frequency range of 1–45 Hz and standardized each lead of the ECG recordings. The performance of the pre-trained model in the test set and the settings for hyperparameters used in pre-training are detailed in Tables S1 and S2, respectively.

Data preprocessing, training, and evaluation process

The SNUH-PSVT dataset comprises a standard 12-lead ECGs of 10 seconds, with a sampling rate of 500 Hz. Each ECG lead was preprocessed to have a mean of 0 and a standard deviation of 1. We carefully selected high-quality 12-lead ECGs free of noise, such as motion artifacts, baseline wandering, and issues caused by poor contact between electrodes and skin. Given the absence of noise in the data and the possibility that subtle signals could be crucial for distinguishing between AVNRT and concealed AVRT, we did not apply filters during the pre-processing step. Most of the ECGs were obtained using the equipment from GE Healthcare (MAC VU360, MAC 2000, MAC 3500, and MAC 5500). All experiments were performed using Python 3.8 and PyTorch 1.8.1.²⁰ A deep learning model was trained to solve a binary classification between sinus rhythm ECGs of AVNRT and concealed AVRT. The architecture used in our experiments is a modified version of ResNet-34,²¹ which comprises 34 layers and incorporates residual connections (Figure S1). The hyperparameter settings used for training the model are listed in Table S2.

The model's performance was evaluated using ten-fold cross-validation with the training set. Each fold was split by patients. Stratified random sampling was employed to preserve each fold's class ratio between AVNRT and concealed AVRT. For each iteration of cross-validation, nine folds were designated for training, while the remaining fold was used for validation.

To assess the impact of transfer learning, we compared the model without pre-training (i.e. one trained from scratch with random initialization), and the pre-trained model. Additionally, we investigated which set of leads has distinctive features for PSVT classification by comparing models trained on various ECG lead sets: the 12-lead ECG, and an 8-lead configuration (leads I, II, and the six precordial leads), the six precordial leads (leads V1–V6), the six limb leads (leads I, II, III, aVR, aVL, aVF), and a single-lead (lead II). After identifying the optimal configuration through experiments on the training set, we assessed its performance on the test set.

Visualizing feature attributions of ECG segments

Feature attribution is a method that evaluates the importance of each input feature in contributing to a model's specific prediction. By examining feature attributions, we can identify which regions of ECGs are crucial for a model's decision-making process.²² Among various feature attribution methods, we

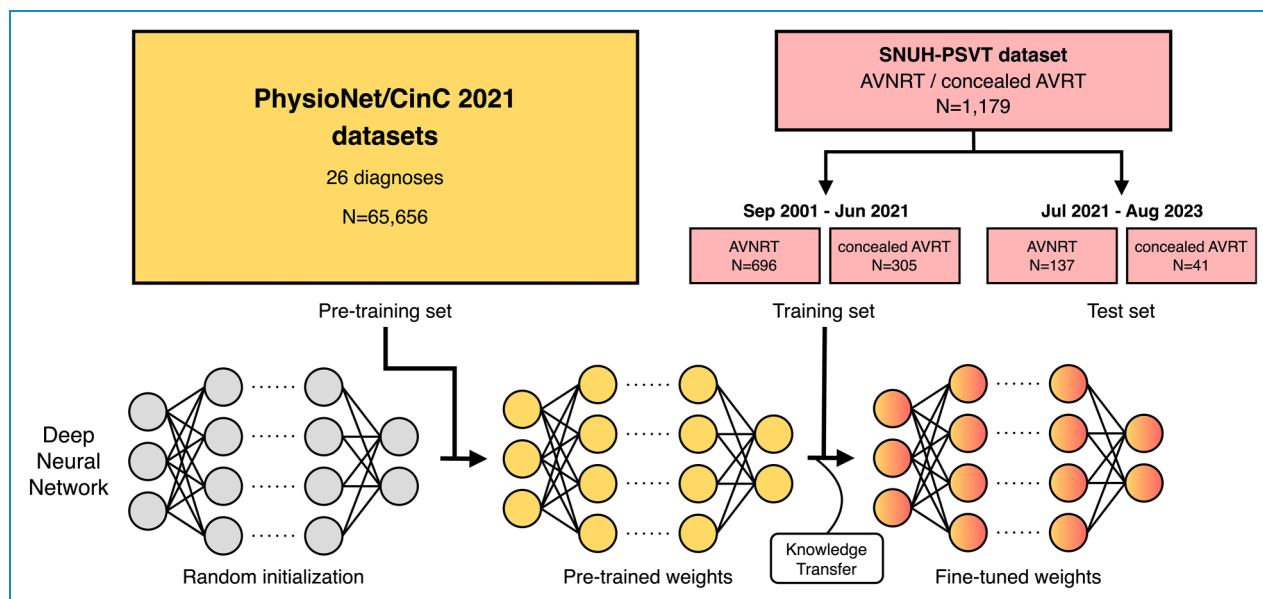


Figure 1. Overview of the research design for classifying AVNRT and concealed AVRT using sinus rhythm ECGs. The deep neural network is pre-trained on the PhysioNet/CinC 2021 datasets and further trained to differentiate between sinus rhythm ECGs with underlying AVNRT and concealed AVRT. After the model selection process using ten-fold cross-validation on the training set, the performance of the final model on the test set is reported.

chose the Integrated Gradients method, which calculates the path integral of gradients along a straight-line path from a baseline input to a particular input.²³ Previous studies have highlighted the effectiveness of this method, as evidenced by its strong performance across various evaluation metrics and its ability to produce a high-resolution attribution map.²⁴ This capability allows for a thorough exploration of how each input segment impacts the model. We used a heatmap to illustrate which segments of an ECG beat are important for a deep learning model in addressing our PSVT classification problem. The detailed experimental settings for visualizing feature attributions are described in Supplementary Methods.

Exploratory analysis

To investigate the impact of additional feature learning, the model's performance when trained with raw ECG data and additional features was compared to its performance when trained with raw ECG data alone. Also, the performances of various other deep learning architectures were compared with ours (ResNet-34) to investigate whether there are model-specific results. The details of exploratory analyses are presented in Supplementary Methods.

Model calibration analysis

The calibration of the final model (12-lead ResNet-34 with pre-training) was assessed to determine how accurately an estimated class probability by deep learning reflects the actual class proportion. The methodologic details are presented in Supplementary Methods.

Statistical analysis

The following seven metrics were calculated to evaluate the performance of deep-learning models: area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), F1-score, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The metrics were reported as the average values from the ten trials with 95% confidence intervals (CIs). The differences in AUROCs were compared using the DeLong test,^{25,26} while other metrics were compared through a paired *t*-test. The F1-score, sensitivity, specificity, PPV, and NPV were calculated based on binary predictions, using thresholds that maximize Youden's J statistic in the validation sets. Sensitivity was defined as the proportion of correctly classified AVNRT samples among all the AVNRT samples. Specificity was defined as the proportion of correctly classified concealed AVRT samples among all concealed AVRT samples. PPV and NPV were defined as the proportions of true AVNRT (or concealed AVRT) samples among samples that the model predicted as AVNRT (or concealed AVRT). Two-sided *p*-values less than 0.05 rejected the null hypothesis and were considered statistically significant.

Results

A total of 1179 patients with either AVNRT or concealed AVRT were analyzed (833 and 346 patients, respectively). The baseline characteristics of the two groups are presented

Table 1. Baseline characteristics.

	AVNRT (N= 833)	Concealed AVRT (N= 346)	P-Value
Age (years)	48.7 ± 16.2	44.2 ± 16.2	<0.001
Men	291 (35.3)	212 (62.7)	<0.001
Heart rate (per minute)	71.0 ± 9.6	69.9 ± 11.2	0.077
PR interval (ms)	154.3 ± 24.5	153.5 ± 22.7	0.625
QRS duration (ms)	91.6 ± 12.1	93.5 ± 10.8	0.010
Corrected QT interval (ms)	427.1 ± 23.3	420.2 ± 25.4	<0.001
P axis (degree)	50.0 ± 21.3	49.8 ± 21.4	0.894
R axis (degree)	49.3 ± 34.3	53.7 ± 33.5	0.043
T axis (degree)	42.7 ± 22.7	43.3 ± 26.0	0.695
Premature atrial complex	15 (1.8)	5 (1.4)	0.667
Premature ventricular complex	8 (1.0)	8 (2.3)	0.068
Left ventricular hypertrophy	13 (1.6)	8 (2.3)	0.374
Low voltage	7 (0.8)	2 (0.6)	0.637
Early repolarization	16 (1.9)	13 (3.8)	0.064
Right atrial enlargement	6 (0.7)	8 (2.3)	0.022
Left atrial enlargement	12 (1.4)	8 (2.3)	0.291
First-degree atrioventricular block	23 (2.8)	5 (1.4)	0.177
Right bundle branch block	43 (5.2)	6 (1.7)	0.007
Left bundle branch block	0 (0.0)	0 (0.0)	>0.999
Left anterior fascicular block	7 (0.8)	1 (0.3)	0.294
Left posterior fascicular block	2 (0.2)	0 (0.0)	0.362
Bifascicular block	3 (0.4)	0 (0.0)	0.264
Any infarct	32 (3.8)	13 (3.8)	0.945
Any ischemia	20 (2.4)	12 (3.5)	0.304

Data are presented as mean ± standard deviation or N (%), with percentages calculated after excluding missing values from the denominator. Corrected QT intervals were calculated using Bazett's formula.

AVNRT: atrioventricular nodal reentry tachycardia; AVRT: atrioventricular reentry tachycardia.

in Table 1. The AVNRT group was significantly older (48.7 vs. 44.2 years, $p < 0.001$) and had a lower proportion of men (35.3% vs. 62.7%, $p < 0.001$). For ECG features, the AVNRT group had a narrower QRS duration (91.6 vs. 93.5 ms, $p = 0.010$), a longer corrected QT

interval (427.1 vs. 420.2 ms, $p < 0.001$), a lower value of R axis (49.3 vs 53.7 degree, $p = 0.043$), a higher proportion of right atrial enlargement (0.7% vs 2.3%, $p = 0.022$), and a higher proportion of right bundle branch block (5.2% vs. 1.7%, $p = 0.007$).

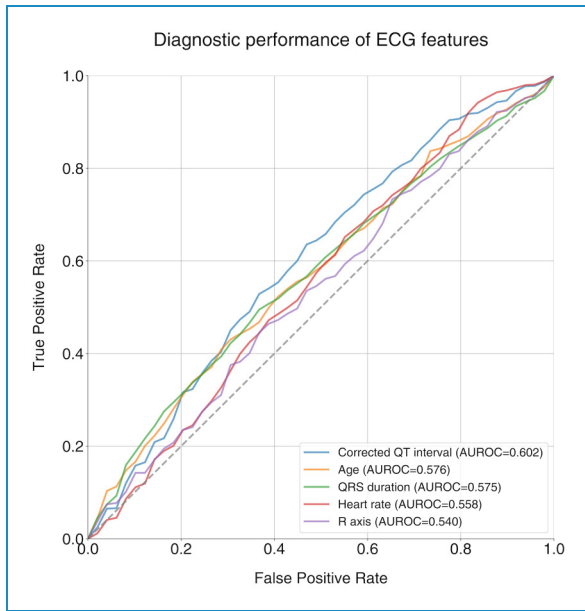


Figure 2. ROC curves and C-statistics (AUROCs) for selected ECG features. AUROC: area under the receiver operating characteristic curve; ECG: electrocardiogram.

C-statistic of selected features

The C statistics were performed using selected features that were significantly different between the two groups (age, QRS duration, corrected QT interval, R axis) plus heart rate in the training set. Among the selected features, the corrected QT interval exhibited the highest AUROC of 0.602 (95% CI 0.574–0.630) (Figure 2). The AUROCs for the other features were 0.576 (95% CI 0.537–0.615) for age, 0.575 (95% CI 0.534–0.616) for QRS duration, 0.558 (95% CI 0.526–0.590) for heart rate, and 0.540 (95% CI 0.525–0.555) for R axis, respectively.

Effect of pre-training on the classification performance

The model without pre-training achieved an AUROC of 0.668 (95% CI 0.622–0.715) and an AUPRC of 0.802 (95% CI 0.763–0.840), whereas the pre-trained model significantly improved to an AUROC of 0.726 (95% CI 0.692–0.760) and an AUPRC of 0.841 (95% CI 0.810–0.872), with p -values of 0.024 and 0.035, respectively (Table 2, Figure 3(a) and (b)). The pre-trained model demonstrated an F1-score of 0.749 (95% CI 0.702–0.796), a sensitivity of 0.681 (95% CI 0.609–0.754), a specificity of 0.711 (95% CI 0.651–0.771), a PPV of 0.845 (95% CI 0.823–0.866), and an NPV of 0.507 (95% CI 0.465–0.549).

Comparison of deep-learning performance across different ECG leads

The highest AUROC and AUPRC values were observed with the 12-lead set, achieving 0.726 (95% CI 0.692–0.760) and 0.841

Table 2. Effect of pre-training on the performance of a deep learning model.

	AUROC (95% CI)	AUPRC (95% CI)	F1-score (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Without pre-training	0.668 (0.622–0.715)	0.802 (0.763–0.840)	0.730 (0.670–0.789)	0.683 (0.579–0.787)	0.626 (0.518–0.733)	0.814 (0.784–0.845)	0.494 (0.431–0.557)
With pre-training	0.726* (0.692–0.760)	0.841* (0.810–0.872)	0.749 (0.702–0.796)	0.681 (0.609–0.754)	0.711 (0.651–0.771)	0.845 (0.823–0.866)	0.507 (0.465–0.549)

* Significance testing compared to values of “Without pre-training”: $p < 0.05$. AUPRC: area under the precision–recall curve; AUROC: area under the receiver operating characteristic curve; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

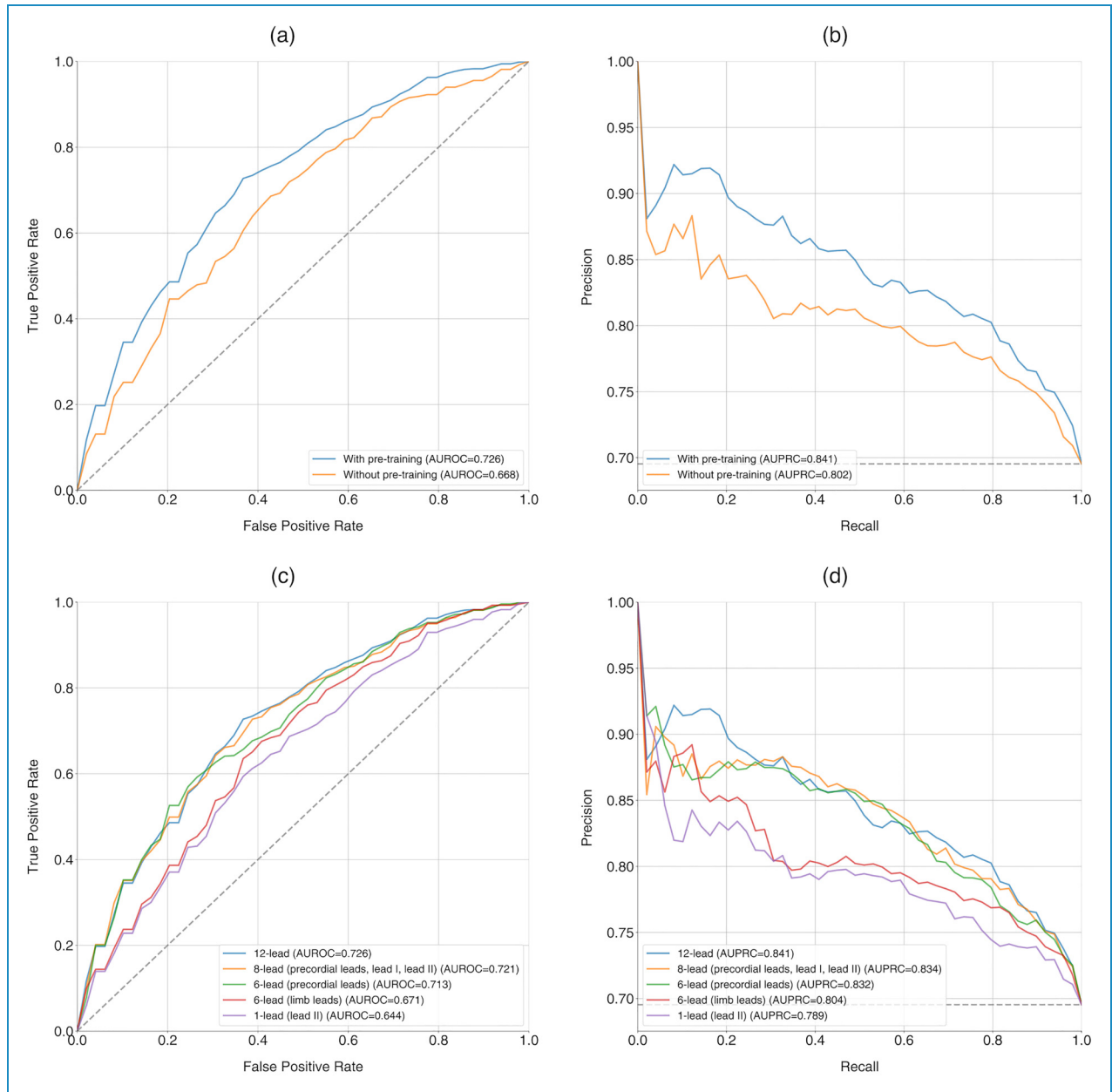


Figure 3. Performance of deep learning in ten-fold cross-validation for PSVT classification. (Upper row) Comparison of the model with and without pre-training, including (a) receiver operating curves (AUROCs) and (b) precision-recall curves (AUPRCs). (Lower row) Comparison of the model using different sets of leads, including (c) receiver operating curves (AUROCs) and (d) precision-recall curves (AUPRCs). AUPRC: area under the precision-recall curve; AUROC: area under the receiver operating characteristic curve; ECG: electrocardiogram.

(95% CI 0.810–0.872), respectively (Figure 3(c) and (d)). The eight-lead set and the six precordial lead set showed AUROCs similar to that of the 12-lead set, with values of 0.721 (95% CI 0.679–0.762) and 0.713 (95% CI 0.676–0.749), and p -values of 0.676 and 0.265. However, the six limb lead set and lead II alone demonstrated significantly lower AUROCs compared to the 12-lead set, with p -values of 0.002 and <0.001 . A summary of the diagnostic performances based on the trained lead sets is presented in Table 3, indicating a trend of decreasing diagnostic performance with fewer ECG leads.

Evaluation of the test performance

A summary of the diagnostic performance of the selected configuration on the test set is presented in Table 4. Our modified ResNet-34 with pre-training, which processes 12-lead ECG recordings, is evaluated. The model exhibited an AUROC of 0.708 (95% CI 0.696–0.721), a slight decrease from the ten-fold cross-validation AUROC of 0.726, while its' AUPRC of 0.875 (95% CI 0.864–0.885) showed a slight increase compared to the validation

Table 3. Performance of the deep learning model based on the trained ECG leads.

	AUROC (95% CI)	AUPRC (95% CI)	F1-score (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
12-Lead	0.726 (0.692-0.760)	0.841 (0.810-0.872)	0.749 (0.702-0.796)	0.681 (0.609-0.754)	0.711 (0.651-0.771)	0.845 (0.823-0.866)	0.507 (0.465-0.549)
8-Lead (precordial leads, lead I, lead II)	0.721 (0.679-0.762)	0.834 (0.798-0.871)	0.740 (0.693-0.787)	0.674 (0.593-0.755)	0.699 (0.613-0.785)	0.844 (0.807-0.882)	0.499 (0.448-0.550)
6-Lead (precordial leads)	0.713 (0.676-0.749)	0.832 (0.802-0.863)	0.745 (0.709-0.781)	0.685 (0.604-0.767)	0.680 (0.554-0.806)	0.842 (0.805-0.880)	0.508 (0.449-0.567)
6-Lead (limb leads)	0.671** (0.638-0.703)	0.804 (0.774-0.834)	0.752 (0.703-0.801)	0.723 (0.644-0.802)	0.577 (0.502-0.652)	0.798 (0.780-0.817)	0.492 (0.452-0.532)
1-Lead (lead II)	0.644*** (0.611-0.678)	0.790 (0.764-0.815)	0.706 (0.660-0.752)	0.645 (0.557-0.733)	0.628 (0.510-0.745)	0.807 (0.779-0.836)	0.446 (0.418-0.475)

** Significance testing compared to values of "12-lead"; $p < 0.01$.

*** Significance testing compared to values of "12-lead"; $p < 0.001$.

AUPRC: area under the precision-recall curve; AUROC: area under the receiver operating characteristic curve; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

Table 4. Performance of the deep learning model on test set.

Metric	Value (95% CI)
AUROC	0.708 (0.696–0.721)
AUPRC	0.875 (0.864–0.885)
F1-score	0.750 (0.704–0.797)
Sensitivity	0.670 (0.603–0.738)
Specificity	0.649 (0.610–0.688)
PPV	0.864 (0.855–0.873)
NPV	0.382 (0.346–0.418)

AUPRC: area under the precision-recall curve; AUROC: area under the receiver operating characteristic curve; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

AUPRC of 0.841. This variation in AUPRC is likely due to the higher baseline present in the test set, which is attributable to differences in the distribution of class samples. The model's test performance indicated an F1 score of 0.750 (95% CI 0.704–0.797), a sensitivity of 0.670 (95% CI 0.603–0.738), a specificity of 0.649 (95% CI 0.610–0.688), a PPV of 0.864 (95% CI 0.855–0.873), and an NPV of 0.382 (95% CI 0.346–0.418).

Analyzing ECG segment influence on PSVT classification

Utilizing the Integrated Gradients method, the visualization of ECG feature attributions is presented in Figure 4. The average beat of high-probability samples is displayed as a cyan line for each lead, with the standard deviation represented as a shaded area. The bright regions in the background indicate high attribution values. For a given beat, the deep-learning model concentrated on the *T*-waves and PR segments of sinus rhythm ECGs with an underlying AVNRT. In contrast, it focused on the terminal portions of the *P*-waves in ECGs with underlying concealed AVRT.

Discussion

This study explored the capability of a deep-learning model to differentiate patients with underlying AVNRT from those with concealed AVRT using sinus rhythm ECGs. The AVNRT group and the concealed AVRT group exhibited different characteristics in terms of several features: age, sex, QRS duration, corrected QT interval, *R* axis, the presence of right atrial enlargement, and the presence of right bundle branch block (Table 1). The observed differences in age and sex align with previous studies, which suggest that AVNRT is more prevalent in women than

AVRT and patients with AVNRT tend to be older (47 vs. 37 years).^{27,28} Although the C statistics of the selected features were better than random chance, their performances did not reach a level of clinical significance (Figure 2). Consequently, we explored the feasibility of using a deep learning model with raw ECG signals to address the PSVT classification problem.

To enhance learning efficiency, we utilized pre-training with large-scale datasets of diverse ECG diagnoses. This approach significantly improved the performance of our deep-learning model in this classification task, where the available sample size is limited (Table 2, Figure 3(a) and (b)). Through comparisons of training results using different sets of leads, we infer that the precordial leads likely contain the necessary information to distinguish between sinus rhythm ECGs with underlying AVNRT and concealed AVRT. Additionally, using a greater number of leads might increase the ability to detect subtle signals essential for classification in ECG recordings (Table 3, Figure 3(c) and (d)). By using 12-lead ECGs, our model achieved an AUROC of 0.708 and an AUPRC of 0.875 in the test dataset. These results indicate that, while not perfect, the deep-learning model can identify some distinctive features between AVNRT and concealed AVRT from sinus rhythm ECGs (Table 4). The subtle differences between these two arrhythmias, which are indistinguishable to human cardiologists, suggest that the model offers new possibilities for addressing classification challenges previously considered intractable by human experts.

We also investigated whether the performance of the deep learning model was solely attributable to the differences in traditional ECG features. The sex difference between the two groups (AVNRT and concealed AVNRT) may introduce differences in ECG features based on sex. However, the model's performance did not differ between men and women in terms of AUROCs (0.695 vs 0.680, $p=0.745$) (Figure S2), and adding additional feature learning to the deep learning model did not improve diagnostic performance (Table S3). Therefore, we hypothesize that deep learning models can identify subtle differences in an ECG signal beyond traditional ECG features. Figure 4 indicates that the PR interval, *P* wave, and *T* wave may be "important" to the deep-learning model to differentiate between the two arrhythmias. Based on these findings, we hypothesize that there could be differences in atrioventricular nodal conduction physiology, atrial depolarization, and ventricular repolarization between the two arrhythmias.

Table S4 presents the performance of other models. The deep neural networks, HeartNet and HeartNetIEEE, exhibited lower performance compared to our ResNet34 without pre-training. The machine learning models achieved better scores than our ResNet34, but the differences were insignificant. Additionally, due to the smaller capacity of machine learning models compared to deep neural

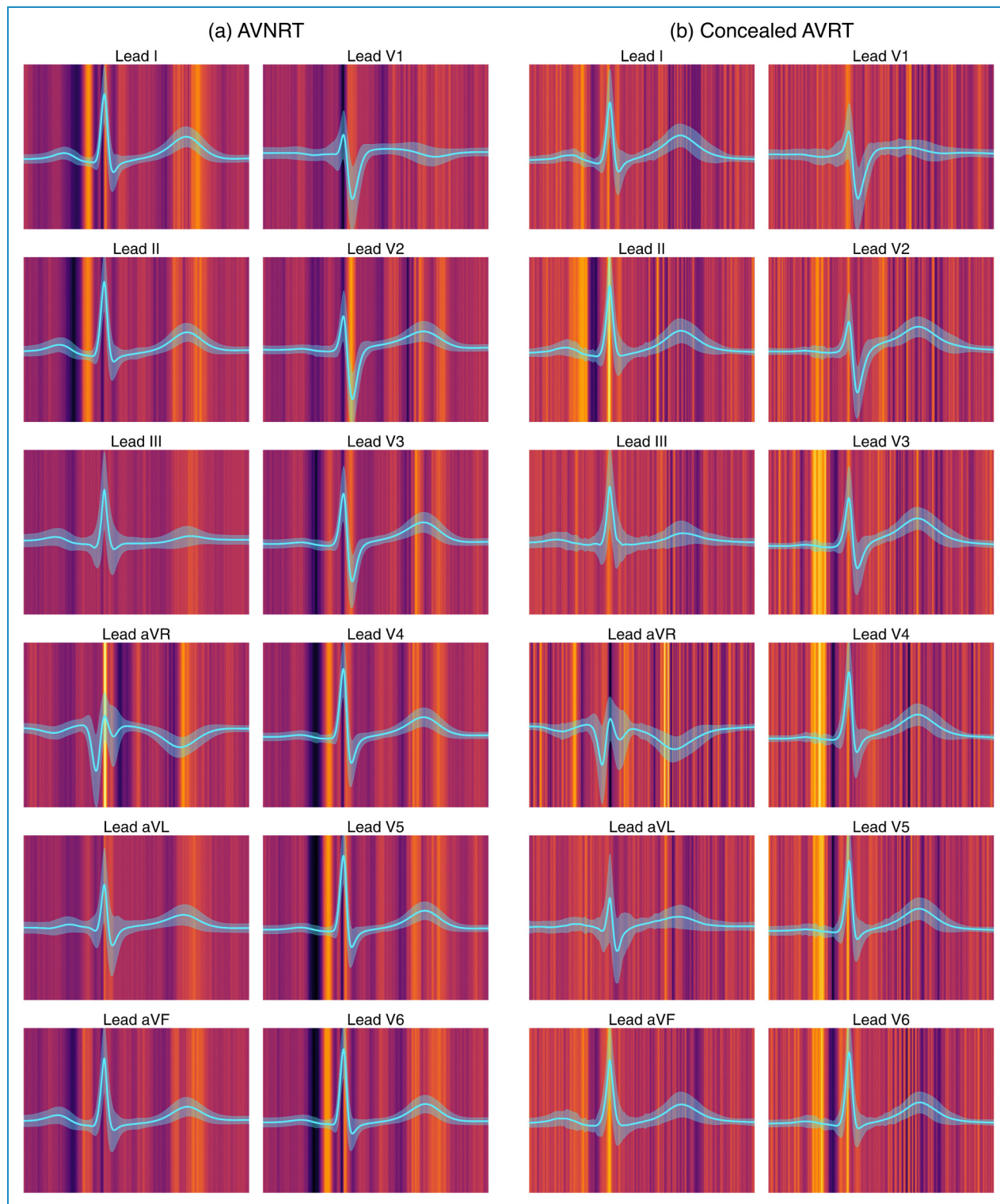


Figure 4. Visualization of feature attribution on ECG segments for PSVT classification. AVNRT: atrioventricular nodal reentry tachycardia; AVRT: atrioventricular reentry tachycardia; ECG: electrocardiogram; PSVT: paroxysmal supraventricular tachycardia.

networks, the transfer learning technique cannot be easily applied to them. Therefore, we concluded that the ResNet34 model with pre-training is appropriate for classifying the two PSVT types with raw ECG data.

The calibration of the model's performance is presented in Figure S3. Our results showed that the deep learning model was accurately calibrated for AVNRT-likely samples (i.e. a predicted probability of AVNRT ≥ 0.5).

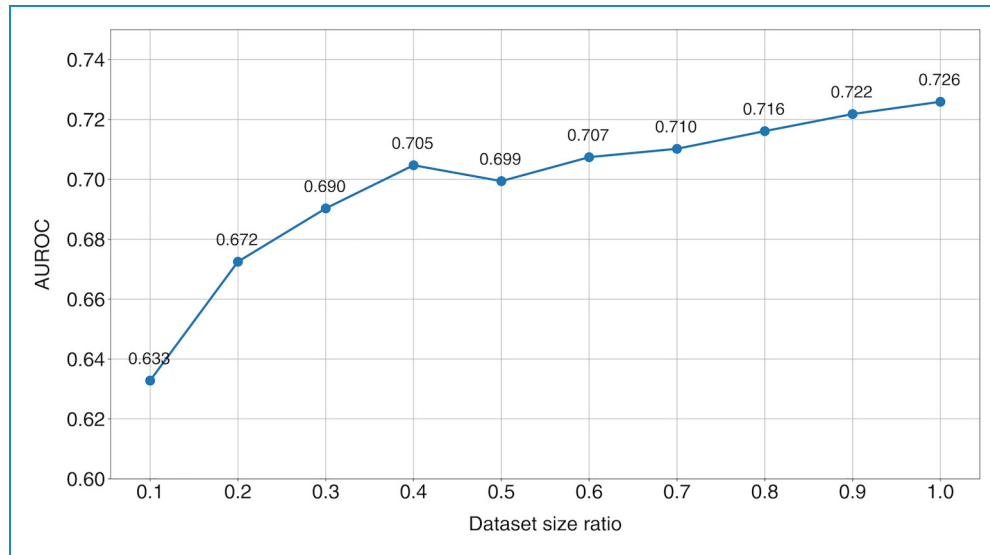


Figure 5. Performance of the deep learning model with varying training dataset sizes. AUROC: area under the receiver operating characteristic curve.

However, it underestimated the probability of AVNRT for the AVNRT-unlikely samples (i.e. probability <0.5). This underestimation might be partly due to the relatively small size of the training dataset, or it could be that sinus rhythm ECGs with underlying concealed AVRT have fewer learnable patterns.

Despite the successful application of transfer learning, the size of the dataset remains a limiting factor for our model's performance. Figure 5 examines whether increasing the dataset size could enhance our model's performance by varying the size of the training set. A gradual improvement in the AUROC was observed with an increase in training set size, suggesting that our model could improve predictive performance with access to a larger dataset. Future research should explore whether the findings of this study can be replicated using a more extensive dataset of PSVTs.

Physicians often find diagnosing and confirming the type of PSVT challenging. Due to PSVT's paroxysmal nature, intermittent ECG tests are less effective, leading many patients with PSVT to repeatedly visit clinics for testing. Therefore, technology that can predict PSVT types using sinus rhythm ECGs can improve diagnostic efficiency and ultimately enhance patient management. Recently, it has been reported that PSVT can be detected using deep-learning analysis of sinus rhythm ECGs.⁸ However, patients with manifest AVRT exhibit preexcitation during sinus rhythm, a strong ECG feature for predicting AVRT. Consequently, sinus rhythm ECGs of manifest AVRT might be more easily distinguished from those of AVNRT by deep learning. In contrast, our study demonstrated that deep-learning classification of sinus rhythm ECGs with underlying AVNRT and concealed AVRT is feasible. Predicting PSVT types can guide medical treatment by enabling the selection of appropriate drugs

based on the specific type of PSVT and assisting interventional electrophysiologists in developing effective ablation strategies. Additionally, our deep learning model operates using only a 12-lead ECG of sinus rhythm routinely obtained in PSVT clinics. This makes our deep learning model highly practical for daily clinical use. However, at the current stage, improvement in diagnostic performance is necessary to enhance the clinical utility of our model further. Therefore, a future study with a more extensive dataset and efficient algorithms is desired.

Limitation

The limitations of our study are as follows. First, the total number of samples available for this study was 1179, which may not suffice to efficiently train a high-capacity deep-learning model. Moreover, the number of AVNRT samples exceeded the number of concealed AVRT samples by more than double, potentially leading to a class imbalance issue. To address these challenges, we utilized a transfer learning technique to enhance learning efficiency. Second, the study only included patients diagnosed with either AVNRT or concealed AVRT. Other PSVT types, such as paroxysmal atrial tachycardia and junctional tachycardia, were not included in this study's scope. Third, there were differences in some baseline characteristics between the AVNRT and concealed AVRT groups. Subtle ECG changes potentially related to age or sex might introduce bias into our results. Nonetheless, we have provided evidence suggesting that our model's performance does not solely depend on these baseline characteristics. Fourth, our study did not include external validation with datasets collected from other institutions.

To our knowledge, no publicly available datasets include sinus rhythm ECGs from patients with underlying AVNRT and concealed AVRT. Introducing an external validation process would enhance the validity of our research findings. To minimize the correlation between samples, we constructed the hold-out test set based on the examination dates. Also, ECG features were obtained from the automated readings provided by the equipment's algorithm in our study. This approach may introduce potential inaccuracies in the measurement of some features. Finally, the deep-learning algorithm demonstrated modest performance, which might limit its clinical utility. However, our study has confirmed the feasibility of classifying underlying AVNRT and concealed AVRT using sinus rhythm ECGs and indicated that collecting more data could improve the model's performance.

Conclusion

This study investigated the feasibility of a deep learning algorithm to differentiate between sinus rhythm ECGs with underlying AVNRT and concealed AVRT. To address the challenge of the small dataset size, a transfer learning approach was used, which significantly enhanced the model's performance. The deep learning algorithm proved capable of classifying underlying PSVT types using sinus rhythm ECGs. It has the potential to assist physicians in suspecting specific types of PSVT when tachycardia ECGs are not observed. Further research is needed to refine and validate the algorithm, allowing for the diagnosis of other types of cardiac arrhythmias using sinus rhythm ECGs.

AUPRC:	area under the precision-recall curve
AUROC:	area under the receiver operating characteristic curve
AVNRT:	atrioventricular nodal reentry tachycardia
AVRT:	atrioventricular reentry tachycardia
CI:	confidence intervals
ECG:	electrocardiogram
NPV:	negative predictive value
PPV:	positive predictive value
PSVT:	paroxysmal supraventricular tachycardia

Acknowledgements: The authors are grateful to all the participants for their involvement in this study.

Consent to participate: The IRB exempted the study from requiring informed consent because the study only utilized deidentified data retrospectively and did not interact with study participants.

Contributorship: S. Kwon and J. Suh contributed equally and are co-first authors. E.-K. Choi and W. Rhee are the co-corresponding authors. Concept and design were done by S. Kwon, J. Suh, E.-K. Choi, J. Kim, and W. Rhee. Acquisition, analysis, and interpretation of data were done by S. Kwon, J. Suh, E.-K. Choi, J. Kim, H. Ju, H.-J. Ahn, S. Kim, and W. Rhee. Drafting of the manuscript was done by S. Kwon and J. Suh. Critical revision of

the manuscript for important intellectual content was done by E.-K. Choi, S.-R. Lee, S. Oh, and W. Rhee. Statistical analysis was done by S. Kwon, J. Suh, and J. Kim. Funding acquisition was done by E.-K. Choi and W. Rhee. Administrative, technical, or material support were done by E.-K. Choi and W. Rhee. Supervision was done by E.-K. Choi and W. Rhee.

Data and code availability: The datasets analyzed in this study are not publicly available due to patient privacy and security concerns. The code for this study is publicly accessible at <https://github.com/SNU-DRL/classification-psvt>.


Declaration of conflicting interests: The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: S. Kwon, J. Suh, J. Kim, H. Ju, H.-J. Ahn, S. Kim, S.-R. Lee, and S. Oh: None to disclose. W. Rhee: received research grants from Samsung Electronics, NAVER, SK Hynix, and SK Telecom outside the submitted work. E.-K. Choi: received grants from Bayer, BMS/Pfizer, Biosense Webster, Chong Kun Dang, Daewoong Pharmaceutical Co, Daiichi-Sankyo, DeepQure, Dreamtech Co., EIL Pharmaceutical Co., Medtronic, Sanofi-Aventis, Samjinpharm, Seers Technology, and Skylabs outside the submitted work.


Ethical approval: The study protocol was approved by the Seoul National University Hospital Institutional Review Board and adhered to the Declaration of Helsinki revised in 2013 (IRB No: H-2108-141-1246).

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the Korea Medical Device Development Fund grant funded by the Korean government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health and Welfare, the Ministry of Food and Drug Safety) (Project Nos.: HI20C1662, 1711194311, RS-2020-KD000173), by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1A2C2007139), and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [No. RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)].

Guarantor: S. Kwon.

ORCID iDs: Jangwon Suh  <https://orcid.org/0000-0002-6293-0444>

Eue-Keun Choi  <https://orcid.org/0000-0002-0411-6372>

So-Ryoung Lee  <https://orcid.org/0000-0002-6351-5015>

Supplemental material: Supplemental material for this article is available online.

References

1. Ganz LI and Friedman PL. Supraventricular tachycardia. *N Engl J Med* 1995; 332: 162–173.
2. Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J Am Coll Cardiol* 2019; 73: 1317–1335.

3. Siontis KC, Noseworthy PA, Attia ZI, et al. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol* 2021; 18: 465–478.
 4. Somani S, Russak AJ, Richter F, et al. Deep learning and the electrocardiogram: review of the current state-of-the-art. *EP Europace* 2021; 23: 1179–1191.
 5. Dhananjay B, Kumar RP, Neelapu BC, et al. A Q-transform-based deep learning model for the classification of atrial fibrillation types. *Phys Eng Sci Med* 2024; 47: 621–631.
 6. Budaraju D, Neelapu BC, et al. Stacked machine learning models to classify atrial disorders based on clinical ECG features: a method to predict early atrial fibrillation. *Biomed Tech (Berl)* 2023; 68: 393–409.
 7. Chen C, Hua Z, Zhang R, et al. Automated arrhythmia classification based on a combination network of CNN and LSTM. *Biomed Signal Process Control* 2020; 57: 101819.
 8. Wang L, Dang S, Chen S, et al. Deep-learning-based detection of paroxysmal supraventricular tachycardia using sinus-rhythm electrocardiograms. *J Clin Med* 2022; 11: 4578.
 9. Weiss K, Khoshgoftaar TM and Wang D. A survey of transfer learning. *J Big Data* 2016; 3: 1–40.
 10. Weimann K and Conrad TO. Transfer learning for ECG classification. *Sci Rep* 2021; 11: 5251.
 11. Reyna MA, Sadr N, Alday EAP, et al. Will two do? Varying dimensions in electrocardiography: the PhysioNet/computing in cardiology challenge 2021. In: *2021 computing in cardiology (CinC)*, pp.1–4. IEEE.
 12. Reyna MA, Sadr N, Alday EAP, et al. Issues in the automated classification of multilead ECGs using heterogeneous labels and populations. *Physiol Meas* 2022; 43: 084001.
 13. Tihonenko V, Khaustov A, Ivanov S, et al. St Petersburg INCART 12-lead arrhythmia database. PhysioBank PhysioToolkit and PhysioNet 2008.
 14. Bousseljot R, Kreiseler D and Schnabel A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. 1995.
 15. Zheng J, Zhang J, Danioko S, et al. A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. *Sci Data* 2020; 7: 48.
 16. Liu F, Liu C, Zhao L, et al. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *J Med Imaging Health Inform* 2018; 8: 1368–1373.
 17. Zheng J, Chu H, Struppa D, et al. Optimal multi-stage arrhythmia classification approach. *Sci Rep* 2020; 10: 2898.
 18. Wagner P, Strodthoff N, Bousseljot R-D, et al. PTB-XL, a large publicly available electrocardiography dataset. *Sci Data* 2020; 7: 1–15.
 19. Suh J, Kim J, Lee E, et al. Learning ECG representations for multi-label classification of cardiac abnormalities. In: *2021 computing in cardiology (CinC)*, pp.1–4. IEEE.
 20. Paszke A, Gross S, Massa F, et al. Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 2019; 32: 8024–8035.
 21. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition 2016*, pp.770–778.
 22. Zhou Y, Booth S, Ribeiro MT, et al. Do feature attribution methods correctly attribute features? In: *Proceedings of the AAAI conference on artificial intelligence 2022*, pp.9623–9633.
 23. Sundararajan M, Taly A and Yan Q. Axiomatic attribution for deep networks. In: *International conference on machine learning 2017*, pp.3319–3328. PMLR.
 24. Suh J, Kim J, Jung E, et al. Evaluating feature attribution methods for electrocardiogram. arXiv preprint arXiv:2211.12702 2022.
 25. Sun X and Xu W. Fast implementation of DeLong’s algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process Lett* 2014; 21: 1389–1393.
 26. DeLong ER, DeLong DM and Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44(3): 837–845.
 27. Goyal R, Zivin A, Souza J, et al. Comparison of the ages of tachycardia onset in patients with atrioventricular nodal reentrant tachycardia and accessory pathway—mediated tachycardia. *Am Heart J* 1996; 132: 765–767.
 28. Liuba I, Jönsson A, Säfström K, et al. Gender-related differences in patients with atrioventricular nodal reentry tachycardia. *Am J Cardiol* 2006; 97: 384–388.
-