Machine Learning Framework for Early Prediction of Ventricular Tachycardia Using Single-Lead Electrocardiogram Signals

Shu-Yi Yeh

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Bioengineering

University of Washington

2025

Committee:

Patrick M. Boyle

Austin Baird

Brody H. Foy

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Machine Learning Framework for Early Prediction of Ventricular Tachycardia Using Single-Lead Electrocardiograms Signals

Shu-Yi Yeh

Chair of the Supervisory Committee:
Patrick M. Boyle
Department of Bioengineering

This study presents a machine learning framework designed for early prediction of ventricular tachycardia (VT) using single-lead electrocardiogram (ECG) signals collected from portable monitoring devices. Unlike prior studies that relied heavily on hospital-based data or incorporated additional demographic information, this work enhances applicability for out-of-hospital patient care. Three modeling approaches were explored: a Long Short-Term Memory (LSTM) model, 2D Convolutional Neural Networks (CNN) trained on spectrograms, and a Support Vector Machine (SVM) using features extracted by a Variational Auto-Encoder (VAE). Among these, the VAE-SVM model demonstrated superior performance, achieving an F1 score of 0.66 and a recall of 0.77. Explainable AI techniques, latent space traversal, and correlation analysis were applied to interpret model behavior and identify physiologically meaningful features associated with VT onset. These findings highlight a valuable opportunity for developing wearable-based VT detection tools that can be integrated into daily health monitoring systems.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Dr. Patrick Boyle, for his invaluable guidance and support throughout my thesis work. His constant encouragement inspired me to explore new ideas and challenge myself. I would also like to thank my committee members, Dr. Austin Baird and Dr. Brody Foy, for their insightful feedback and thoughtful suggestions, which helped me stay on track and significantly improved the quality of this thesis.

I would like to thank MESA for providing such an outstanding dataset. I'm thankful to Dr. Arun Sridhar for his assistance in conception of project and providing clinical perspectives, and to Prof. Susan Heckbert for navigating receipt of data from MESA project. I also thank Dr. Jake Mayfield for the manual review of device-annotated ECG data. Special thanks to Amber Chen for the development of preliminary version of AI/ML models. Also, I truly appreciate Lahari Gorantla, Farzana Mohamedali, Stephanie Osorio-Tristan, and Sanika Joshi for their preliminary work on data processing.

I would like to thank Dr. Surbhi Sharma, who was always there to help me brainstorm ideas and patiently answered all my questions about machine learning. I am also grateful to Matthew Magoon and Neha Arunkumar for their help with data processing. Many thanks to Dr. Chelsea Gibbs for sharing helpful tips on presentation skills and job searching. Finally, I would like to thank everyone in the CardSS Lab for their support, encouragement, and for creating such a collaborative research environment.

Lastly, I would like to thank my family. I wouldn't have had the opportunity to study at UW without their support and love. I am also deeply grateful to my boyfriend, Hao-Yang. Even though

he is in Taiwan, he has always stood by me, offering constant support and encouragement for every decision I made.

I will always cherish the valuable experiences and personal growth I have gained over the past two years. Although studying abroad could sometimes feel lonely, I was fortunate to receive tremendous support and kindness from the people around me. I am truly grateful to everyone who has been part of this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1. INTRODUCTION

1.1 Background and Motivation

Ventricular Tachycardia (VT) is a type of arrhythmia characterized by an abnormally high heart rate originating from the ventricles, defined as three or more consecutive beats at a rate of more than 100 beats per minute. VT is primarily caused by ischemic heart disease or other types of structural heart disease that are associated with a risk of sudden death [1]. While brief episodes of VT may not pose significant risks, prolonged occurrences can escalate into ventricular fibrillation, potentially leading to cardiac arrest [1] [2]. The most critical challenge with VT is its unexpected nature. Patients are often unable to foresee its onset, leaving them vulnerable to sudden and life-threatening events. Early prediction could enable patients to modify their behavior to avoid dangerous situations. For instance, if a patient is driving, they could stop the vehicle before the onset of VT, preventing potential car accidents. Therefore, developing methods for the early detection of VT is essential, particularly for individuals using portable electrocardiogram (ECG) devices outside hospital settings.

Since 2010, advancements in machine learning have increasingly been applied to medical data [3]. In 2013, Jabbar et al., used the K-Nearest Neighbor algorithm for binary heart disease detection, including conditions such as coronary heart disease, cardiomyopathy, cardiovascular disease, and heart failure, among others [4]. Similarly, in 2019, Alarsan et al. employed machine learning models such as Random Forest and Gradient-Boosted Trees on ECG features to detect abnormal heartbeats, using data that included both inpatients and outpatients from the MIT-BIH Arrhythmia Database [5]. Techniques like Convolutional Neural Networks (CNN) and Variational Autoencoders (VAE) also started being adopted in this field. For instance, Leur et al. applied these methods to enhance the explainability of ECG analysis [6]. Building on the advancements in

machine learning frameworks and their success in analyzing medical signals, this study explores the feasibility of applying these techniques to predict VT onset. Specifically, single-lead ECG data collected from the Multi-Ethnic Study of Atherosclerosis [7][8] using the portable ECG device Zio Patch will be analyzed. Unlike prior studies that primarily used hospital-based ECG data, our study addresses a fundamentally different challenge by leveraging data that reflect real-world monitoring in outpatient settings. As an in-the-wild data source, it better reflects natural physiological variability and supports the development of early warning systems that can proactively alert patients in everyday life.

1.2 Review of Relevant Literature

Many researchers have employed machine learning frameworks to ECG data to classify and/or predict various heart diseases. Numerous studies have demonstrated strong performance in classifying or predicting conditions such as ventricular fibrillation (VF) [9] and atrial fibrillation (AF) [10]. Recently, there has been a growing interest in predicting the onset of VT [11][12]. However, most existing studies perform prediction or classification based on heart rate variability (HRV) features extracted from 12-lead ECGs using conventional methods, rather than applying deep learning models directly to the raw ECG signals. Moreover, most of these studies typically rely on hospital-acquired data rather than data from portable devices, as the former is generally more accurate and suitable for predictive modeling. While their approaches are well-suited to controlled clinical environments, our study focuses on a different context that involves real-world, ambulatory data collected through wearable devices. These differences in data characteristics lead to distinct modeling challenges and objectives, but the methodologies and architectures developed in prior work still offer valuable insights for model design in our setting.

Lee et al. proposed an artificial neural network architecture to predict VT using data collected one hour before occurrence [11]. They used 14 parameters including heart rate variability derived from single lead ECG and respiratory rate variability from respiratory signals. The model was trained using 52 recordings from 41 patients in cardiovascular intensive care unit (CCU), comprising both control and event data (1 hour before VT), each lasting 5 minutes. The model achieved a good performance, with an AUC of 0.93, accuracy of 0.85, sensitivity of 0.88, and specificity of 0.82. However, the study has notable limitations. The dataset is relatively small and includes repeated recordings from the same patients, which may introduce bias and limit the generalizability of the results. Also, the inclusion of respiratory data, a key feature in the model, significantly contributes to its performance. Consequently, this research does not sufficiently demonstrate that high predictive accuracy can be achieved solely using single-lead ECG data. Moreover, patients in the CCU commonly present with multiple or more severe comorbidities, which could artificially enhance the model's apparent predictive performance. In such cases, the model's prediction of VT may actually be driven by indicators of overall clinical severity rather than VT-specific signals. This limits the model's generalizability and its utility in out-of-hospital settings.

In the past few years, the development of machine learning models has accelerated rapidly. Many groups have started using CNNs as the model architecture of choice due to their exceptional ability to handle image and spatial data. Taye et al. proposed a model with a one-dimensional convolutional neural network (1-D CNN) architecture to predict the occurrence of ventricular tachyarrhythmia (VTA) [13]. The dataset, sourced from the PhysioNet database known as the Spontaneous Ventricular Tachyarrhythmia Database [14], comprised 135 pairs of RR interval time series obtained from 78 patients. These were recorded by implantable cardioverter defibrillators

(ICDs) and included data from 360 to 60 seconds before VTA onset for training, and data from 60 seconds to the onset of VTA for testing. The CNN model used RR intervals as input and automatically extracted features to do prediction. In contrast, classical machine learning models such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Artificial Neural Networks (ANN) relied on 11 heart rate variability (HRV) features extracted using traditional methods, since these models do not possess the ability to learn features automatically from raw input data. The CNN model demonstrated superior performance, achieving a more reliable AUC of 0.78 and an accuracy of 0.85. This study introduced an innovative approach to predicting heart disease by using RR intervals as input for the CNN, allowing the model to autonomously identify and select predictive features. However, the model aimed to predict VTA, a broader category that includes both VT and VF. This difference in signal characteristics highlights a potential limitation of the model, as its high performance might be partially attributed to the easier detection of VF events rather than the more challenging task of distinguishing VT. Additionally, the superior quality of data obtained from ICDs, which are directly implanted, likely played a role in achieving such promising results. These limitations highlight the need of developing VT-specific prediction models based on out-of-hospital datasets to improve generalizability and clinical applicability.

In recent years, telemetry-based systems have emerged as a promising solution for real-time monitoring, especially for patients who neither require hospitalization nor have an ICD. Portable ECG devices enable continuous health tracking in everyday settings, offering significant convenience. Economou Lundeberg et al. presented a study that used data from mobile cardiac telemetry to predict the risk of VT within 30 days based on features extracted from 24 hours of ECG recordings and demographic data [12]. The model was Elastic Net, a regularized regression method that linearly combines Lasso and Ridge regression. The dataset was derived from a cohort

of 19,781 patients wearing portable ECG devices. The results demonstrated that the model achieved an AUC of 0.76, even without including age and sex, and a 98.2% negative predictive value for low-risk patients. Although the signal quality from portable ECG devices is inferior to hospital-grade data, which may affect the reliability of these results, this study remains valuable in the medical field. However, the imbalanced dataset and extended prediction window (30 days) limit the clinical value of this study, emphasizing the need for short-term VT prediction using balanced data distributions.

Addressing the limitations of these studies, a machine learning framework was developed to improve the prediction performance of VT. A key aspect of this approach is the use of data from portable ECG devices, which enhances its potential for broader applicability. Furthermore, this study explores whether cutting-edge technologies, such as machine learning, can further enhance prediction accuracy, explainability and overall effectiveness.

1.3 Specific Goals

The objective of this study is to investigate the potential of VT onset prediction by using several machine learning frameworks. Through transformation and feature extraction on ECG data, three types of inputs were used. Machine learning models such as Random Forest, VAE, and CNN were applied to assess performance. The outcomes were evaluated as binary prediction, distinguishing between normal and pre-VT data.

Chapter 2. METHODOLOGY

2.1 DATASET

The dataset was derived from the Multi-Ethnic Study of Atherosclerosis (MESA), a study focused on the characteristics of subclinical cardiovascular disease and the risk factors that predicted progression to clinically overt cardiovascular disease [7][8].

In this research, 488 participants who wore the portable continuous ECG device, Zio Patch (iRhythm Technologies, Inc, San Francisco, CA), for at least 7 days were initially included. However, 66 participants were excluded due to low-quality signals caused by artifacts, signal loss, or noise. Therefore, a total of 422 participants were included in the study, each with a pair of control and event data. **Figure 2.1** presents the workflow diagram for this study. For all participants, an episode of VT was recorded. VT was defined by iRhythm as a run of at least four ventricular ectopic beats based on the company's interpretation of the Zio Patch recordings. Data collected immediately before the onset of VT was labeled as the event group. In contrast, data collected either 24 hours before or 24 hours after the onset of VT was labeled as the control group. For model training and evaluation, multiple time intervals were extracted from both control and event groups to perform predictive analysis.

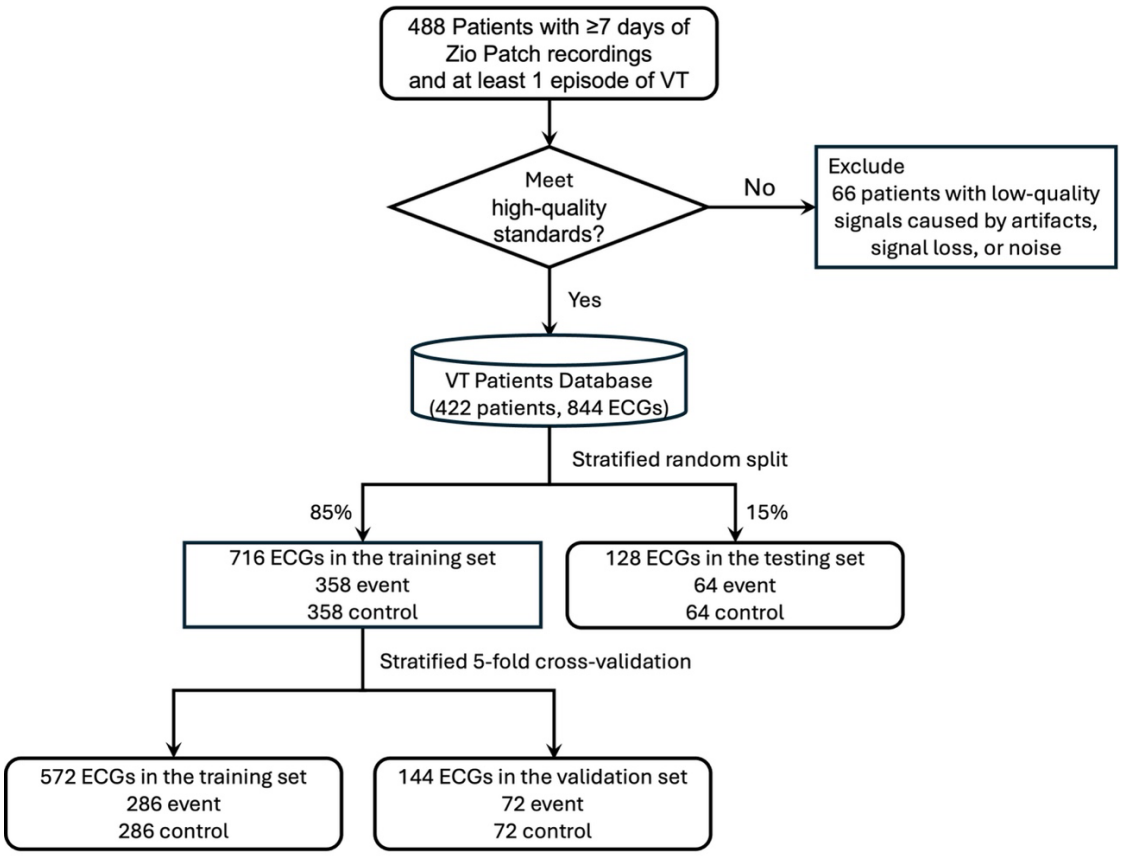

Figure 2.1 Workflow Diagram. A total of 422 patients with 844 ECG segments are included in this study. The database of VT patients was used to train the machine learning model for VT prediction task.

2.2 Data Extraction and Preprocessing

The ECG segments of 844 patients were extracted from Zio Patch devices. The ECG information was obtained from Zio Patch reports, which recorded the timing of VT episodes. Based on this information, multiple pre-VT time intervals were selected, including 0 to 10 seconds, 0 to 30 seconds, 0 to 1 minute, 0 to 2 minutes, 0 to 3 minutes, and 0 to 5 minutes before the onset of VT. For the control group, ECG segments were extracted starting exactly 24 hours before or after the VT episode. Additionally, VT signals from each patient were collected for the classification task.

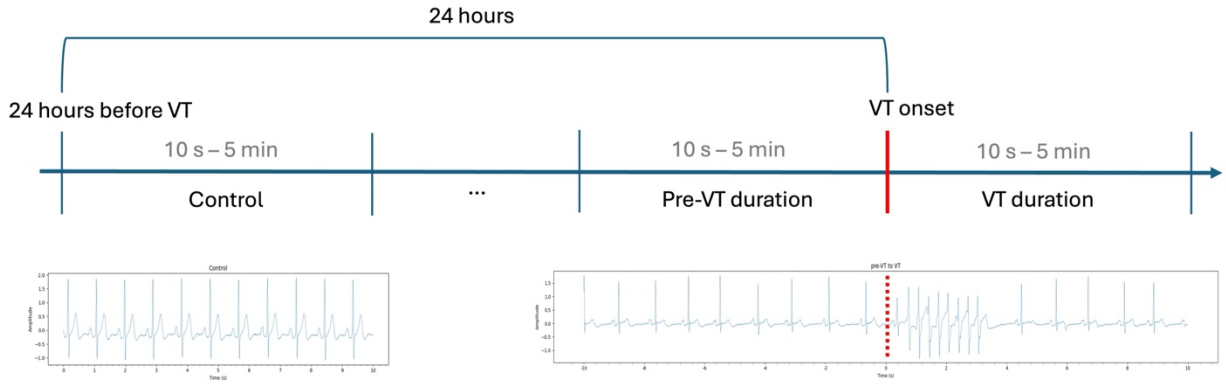

Figure 2.2 Data Segmentation and Examples. This figure presents the segmentation of ECG data into three categories: control, pre-VT, and VT, based on the timing relative to VT onset (red line).

The ECG signals were sampled at a rate of 199.805 Hz, as determined by the acquisition device. **Figure 2.2** illustrates the ECG data segmentation process and provides examples of control, pre-VT, and VT signals. In this figure, the control segments are extracted starting from 24 hours prior to VT onset, and each segment shown has a duration of 10 seconds.

2.3 Machine Learning Models

2.3.1 *Benchmark Model*

The benchmark model represented a simple model, where heart rate variability (HRV) metrics were used as input features, and a random forest classifier was employed for binary prediction [15]. The model used a total of 12 HRV features, including Mean Normal-to-Normal Interval (Mean NN), Standard Deviation of Normal-to-Normal Intervals (SDNN), Root Mean Square of Successive Differences (RMSSD), Percentage of Successive Normal-to-Normal Intervals Greater than 50 ms (pNN50), Mean Heart Rate, Standard Deviation of Heart Rate, Minimum Heart Rate, Maximum Heart Rate, Maximum – Minimum Heart Rate Difference, Poincaré plot standard deviation perpendicular the line of identity (SD1), Poincaré plot standard deviation along the line

of identity (SD2), and the Ratio of SD1 to SD2 (SD1/SD2 Ratio), all derived from the ECG signals. These HRV features were selected based on the methodology proposed in [11]. The benchmark model was intended to establish a baseline performance for comparison with more advanced models. In this context, random forest was selected for the benchmark due to its strong performance with small feature sets and low sensitivity to hyperparameter tuning, making it a practical and reliable baseline model.

2.3.2 *Model for 1D ECG Signals*

Two models were employed for the processing of 1D signal data and VT prediction.

(a) LSTM Model for VT Prediction using ECG Signals

The first approach employed a Long Short-Term Memory (LSTM) network [16], which was specifically designed to handle time-series data, such as ECG signals. LSTM was well-suited for capturing temporal patterns and dynamics within raw 1D signal data, making it well-suited for tasks that involved time-dependent features.

(b) VAE-Based Feature Learning Followed by SVM for VT Prediction

The second approach involved using a VAE to extract latent features from the input signal. This strategy was motivated by the limited discriminative power observed in traditional HRV features and raw ECG signals, highlighting the need for a more expressive, data-driven representation of cardiac dynamics. The VAE model leverages unsupervised learning to uncover latent representations that may be more relevant to pre-VT detection. This model was first introduced by Diederik et al. [17]. The input data consisted of a single median beat computed from a specific time interval after noise removal. These latent features were then given as input to a SVM classifier for binary prediction [18].

The VAE model used in this study was based on the TimeVAE model, originally proposed by Desai et al. [19], which was specifically designed to handle time series data. TimeVAE employs a combination of convolutional layers and residual connection to model the components of time-series data. The architecture consisted of two main parts: the encoder and the decoder. The encoder compressed time-series data into independent latent features, while the decoder reconstructed the original data using these features. The encoder architecture consisted of three 1D convolutional layers with 50, 100 and 200 filters, respectively. These convolutional layers were followed by a flattening layer and dense layers that outputted the latent features. The decoder consisted of deconvolution layer with residual connections for improved training while addressing the vanishing gradient problem. The model was optimized using the Evidence Lower Bound (ELBO) loss function, which improved the model by minimizing both the reconstruction loss and KL divergence. Reconstruction loss measured the difference between the original and reconstructed data, while KL divergence quantified how much the model's probability distribution differed from the true probability distribution, which was set to a standard normal distribution. **Figure 2.3** shows the visualization of the VAE architecture, along with a real ECG median beat example from both the control and event groups, as well as their corresponding reconstructions after VAE modeling.

After extracting the latent features, SVM was applied to do the prediction of VT. SVM is a conventional supervised machine learning algorithm used for classification and prediction. SVM was used with an objective was to construct an optimal decision boundary, known as a hyperplane, that maximized the margin between the cases and controls. When the data is not linearly separable, SVM employs a kernel function to map it into a higher-dimensional space where a suitable hyperplane could be identified. SVM is particularly well-suited for binary classification tasks.

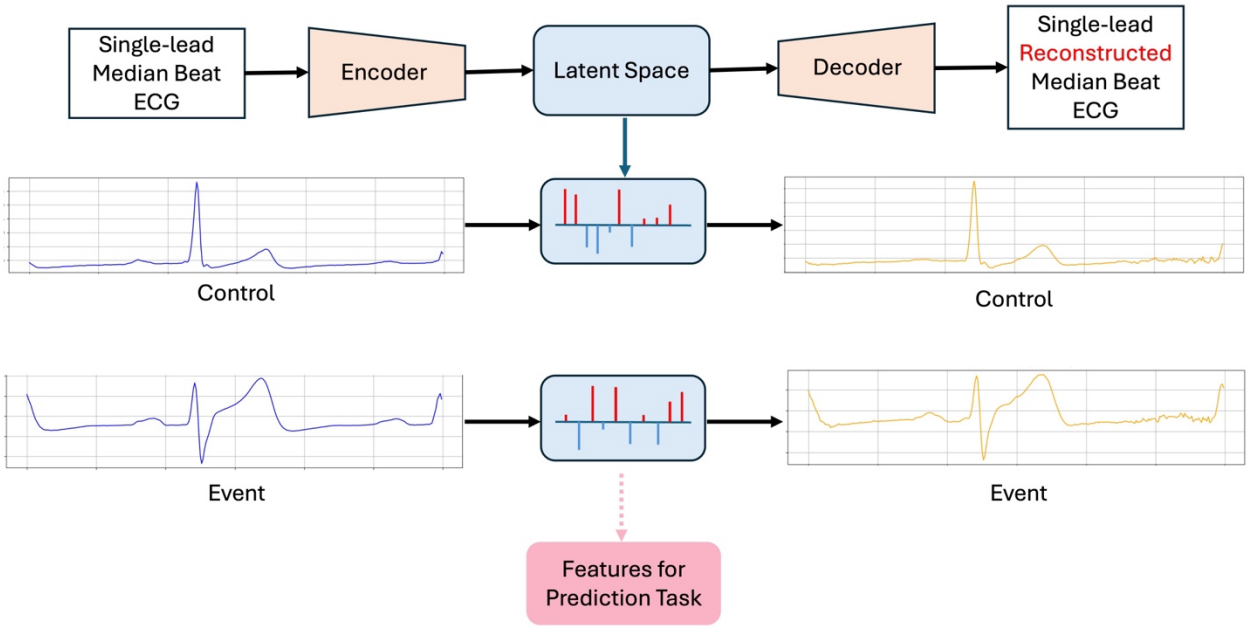

Figure 2.3 Visualization of VAE. This image illustrates the structure of the VAE model implemented for feature extraction. It depicts the key components of the network, including the encoder, which compresses ECG median beat into latent features, and the decoder, which reconstructs the original data from these features.

While other models, such as Random Forest and XGBoost [20], were also explored for prediction using VAE features, SVM was ultimately selected due to its superior performance in this task.

2.3.3 *Model for 2D Spectrogram*

Since CNNs are well-suited for capturing local spatial features, we hypothesized that applying a CNN to our data might help extract informative patterns and enhance predictive accuracy. To enable this, the one-dimensional ECG signals were converted into two-dimensional spectrograms using the Short-Time Fourier Transform (STFT) [21], which represents the variation of the signal's frequency content over time, shown in **Figure 2.4**. Subsequently, various 2D CNN models were employed for VT prediction.

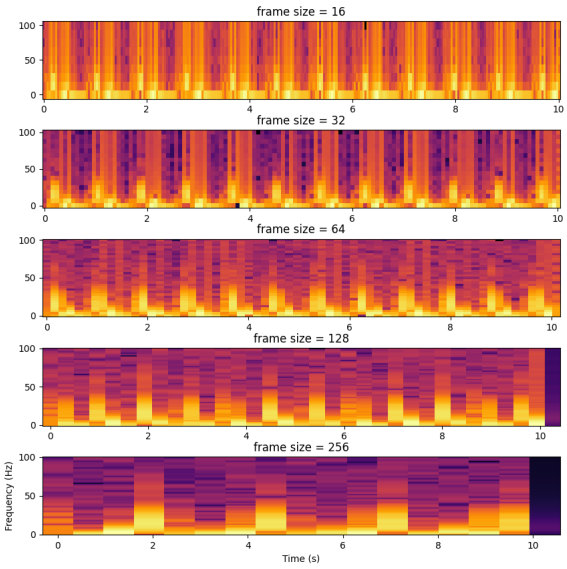

Figure 2.4 Spectrogram of ECG signal. This figure shows spectrograms obtained after performing STFT on a 10-second ECG signal. Five spectrograms were presented using different frame sizes, which represented varying segment lengths.

(a) 2D CNN Model

CNN model is one of the most common models used for image classification and pattern recognition tasks recently [22]. A CNN model was applied on the spectrogram to analyze whether it showed better prediction capability than a 1D model. The 2D CNN model consisted of two convolutional layers with 16 and 32 filters, respectively, both using a 3×3 kernel. Each convolutional layer was followed by batch normalization and ReLU activation [23]. The extracted features were then flattened and passed through two fully connected layers with 256 and 32 neurons, respectively. A dropout layer with a rate of 0.2 was added to prevent overfitting. Finally, a sigmoid activation was applied to the output layer for binary prediction. The model was developed using Keras [24], a deep learning API, with TensorFlow as the backend [25] and Scikit-learn for machine learning tasks [26][27]. **Figure 2.5** illustrates the architecture of the model, generated using a Python package called VisualKeras [28].

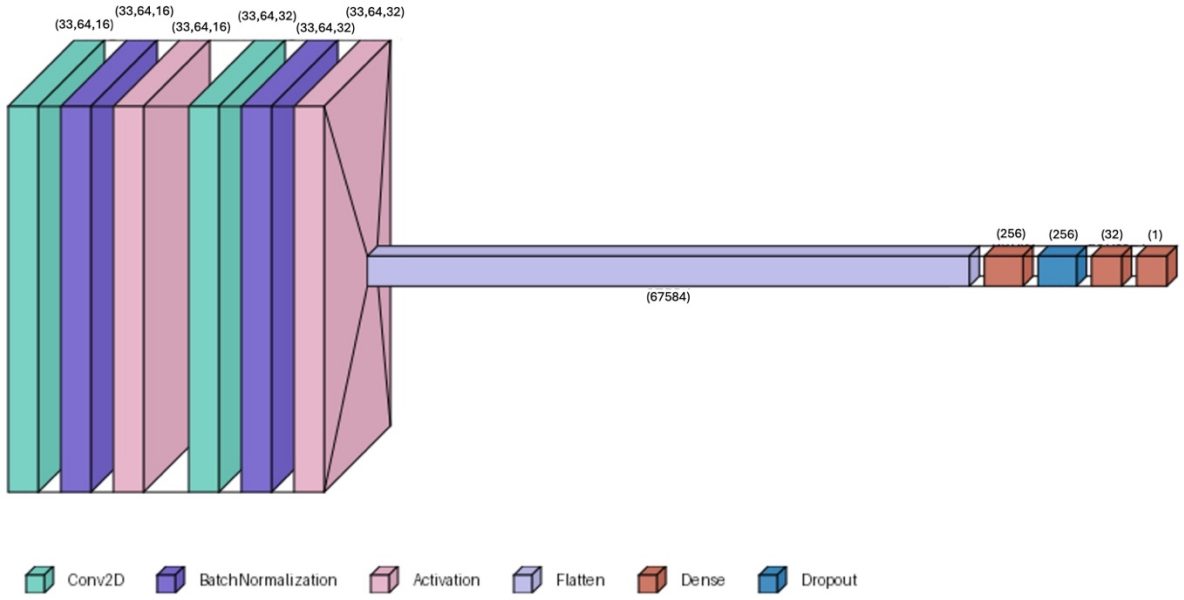

Figure 2.5 Visualization of 2D CNN architecture. This image illustrates the structure of the 2D CNN model that was implemented for the task. It shows the layers of the network, including convolutional layers, batch normalization, activation functions, flatten, dropout, and dense layers.

(b) 2D CNN Pre-trained Model

In addition to the custom CNN model that was trained entirely from scratch using our dataset, transfer learning model was also applied to test whether pre-trained parameters could improve prediction performance. Transfer learning is a machine learning technique which applies the parameters learned from a task to improve the performance on a related task. For the pre-trained model, ResNet101V2 [29], InceptionV3 [30], and DenseNet121 [31] were used as the backbone architectures. These models were initialized with ImageNet pre-trained weights. To preserve the general features learned from large-scale data, all layers were frozen except for the last five layers, which remained trainable and were fine-tuned on our dataset. The extracted features from the backbone were then flattened and passed through fully connected layers with 256 and 32 neurons,

followed by a dropout layer (rate = 0.2) to avoid overfitting. A sigmoid activation was used in the final layer for binary prediction.

2.4 Model Training

For deep learning models including a 1D LSTM and 2D CNN, the Adam optimizer was used for training, and binary cross-entropy was adopted as the loss function. Training was conducted for up to 100 epochs with a batch size of 32. To improve convergence, the ReduceLROnPlateau method was applied to monitor validation loss and dynamically reduce the learning rate by a factor of 0.2 when no improvement was observed for 5 consecutive epochs, with a minimum learning rate of 0.0001. The dataset was split into 85% training/validation and 15% testing. Stratified 5-fold cross-validation was used on the training/validation set to ensure balanced class distribution across folds.

For the VAE model, a reconstruction weight of 3.0 was applied to balance the reconstruction loss and KL divergence. The model was trained using the Adam optimizer. To dynamically adjust the learning rate, the ReduceLROnPlateau callback was employed, which reduced the learning rate by a factor of 0.5 if the reconstruction loss did not improve for 30 consecutive epochs (factor=0.5, patience=30, mode='min'). These settings were chosen empirically to strike a balance between stable convergence and training efficiency. Training was performed with a batch size of 8 for up to 100 epochs. To prevent overfitting, early stopping was applied by monitoring the reconstruction loss on the training data. If the loss failed to improve by at least 0.01 for 50 consecutive epochs (min_delta=0.01, patience=50, mode='min'), training was automatically halted.

For traditional machine learning models, including SVM, Random Forest, and XGBoost, the same 85:15 split was applied for training/validation and testing, with stratified 5-fold cross-validation used within the training set. For Random Forest, hyperparameter tuning was conducted

using RandomizedSearchCV with 5-fold cross-validation. The search space included the number of trees (n_estimators, range from 50 to 500) and tree depth (max_depth, range from 1 to 20). oob_score=True was enabled to obtain an internal estimate of generalization accuracy using out-of-bag samples. For XGBoost, in addition to the number of trees and tree depth, the learning rate was also tuned to control the contribution of each boosting round. For the SVM model, a radial basis function (RBF) kernel was used to capture non-linear relationships. GridSearchCV was employed to optimize hyperparameters C and gamma, with search ranges based on prior studies and preliminary testing.

2.5 Assessment of Model Performance

To assess the performance of all kinds of the model, two types of metrics were used, traditional evaluation metrics and explainable AI.

2.5.1 *Evaluation Metrics*

The performance of most machine learning model was evaluated by several common metrics, including Precision, Recall, F1 score, Accuracy, Receiver Operating Characteristic (ROC) curve, Area Under the Curve (AUC), and Confusion Matrix. All of these metrics were calculated by averaging across 5-fold cross-validation. The optimal threshold of the model is selected by maximizing Youden's J statistic, which ensures a balance between sensitivity and specificity. True Positive Rate (TPR) and False Positive Rate (FPR) are calculated from the formula below:

$$TPR = \frac{TP}{TP + FN} \tag{2.1}$$

$$FPR = \frac{FP}{FP + TN} \tag{2.2}$$

where, TP means true positive, FN means false negative, FP means false positive, and TN means true negative.

The formula of Precision, Recall, F1 Score, and Accuracy are listed below:

$$Precision = \frac{TP}{TP + FP} \tag{2.3}$$

$$Recall\ (TPR) = \frac{TP}{TP + FN} \tag{2.4}$$

$$F1\ Score = 2 \times \frac{Recall\ \times\ Precision}{Recall + Precision} \tag{2.5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.6}$$

The ROC curve is a commonly used metrics for evaluating the performance of a classifier model. It visualizes the trade-off between the TPR and the FPR across different threshold settings. To quantify the performance represented by the ROC curve, AUC measures the overall area under the ROC curve. The Confusion Matrix is a table that displays the counts of TP, TN, FP, and FN. This table helps in analyzing the performance of a prediction model by showing the distribution of correctly and incorrectly classified cases. It provides a clearer view of the proportion of Type I error (False Positives) and Type II error (False Negatives) among all cases.

2.5.2 *Explainable AI*

For the VAE model, to better assess the importance of the 10 latent spaces extracted by the model, an explainable AI method, SHAP (SHapley Additive exPlanations) was applied. SHAP [32] is a game theoretic approach which can help explain the output of machine learning model. The SHAP summary plot was generated to visualize how each latent space contributes to the model's output. Additionally, the distribution of SHAP values provides insights into the stability of these latent spaces, helping to evaluate their consistency and impact on predictions.

Chapter 3. RESULTS

3.1 HRV Statistical Analysis and Performance of Benchmark Model

Prior to building benchmark models, statistical analysis was conducted to evaluate whether any of the 12 HRV features showed significant differences between the control and event (pre-VT) groups. A paired t-test was performed for each feature, comparing values between control and event phases within the same subjects, as shown in **Table 3.1**. None of the features showed statistically significant differences, with all *p*-values exceeding 0.05. This result suggests that conventional HRV metrics may lack sensitivity in capturing physiological changes preceding VT events. Building on this observation, a baseline machine learning model was constructed to further explore whether non-linear relationships or feature interactions could enhance predictive performance.

The benchmark model employed a Random Forest classifier using all 12 HRV features as input. Hyperparameters were optimized (max_depth = 5, n_estimators = 486), and model evaluation was conducted using 5-fold cross-validation. As shown in **Table 3.2**, the model yielded nearly identical results across all time intervals. All performance metrics remained close to 0.5, indicating that the model lacked predictive power and performed comparably to random guessing.

These findings suggest that HRV features, whether analyzed through linear statistical analysis or modeled with non-linear classifiers like Random Forest, lack the discriminative power to distinguish pre-VT from control segments.

Table 3.1 Descriptive statistics and p-values of HRV between control and pre-VT groups.

Feature	Control (Mean± SD)	Event (Mean ± SD)	p-value
Mean NN	872.09 ± 178.44	870.77 ± 190.23	0.8722
SDNN	79.87 ± 80.86	82.72 ± 79.31	0.439
RMSSD	99.17 ± 118.29	101.59 ± 112.7	0.6212
pNN50	22.27 ± 27.42	22.94 ± 26.97	0.52
Mean HR	73.22 ± 16.15	73.81 ± 18.11	0.5038
SD HR	9.06 ± 11.59	9.23 ± 10.25	0.8027
Min HR	59.62 ± 15.26	59.16 ± 15.26	0.5043
Max HR	112.46 ± 70.66	113.34 ± 64.22	0.8345
HR Range	52.84 ± 70.03	54.18 ± 62.2	0.7477
SD1	70.1 ± 83.63	71.81 ± 79.67	0.6231
SD2	83.35 ± 80.92	87.51 ± 81.48	0.3066
SD1/SD2 Ratio	0.76 ± 0.46	0.75 ± 0.41	0.5285

Table 3.2 Evaluation metrics of the benchmark model with different time interval before VT.

Evaluation Metrics	Time before VT onset					
	0-10 s	0-30 s	0-1 min	0-2 min	0-3 min	0-5 min
F1 Score	0.49	0.49	0.50	0.45	0.50	0.48
Accuracy	0.50	0.48	0.50	0.49	0.49	0.49
Precision	0.51	0.47	0.51	0.49	0.50	0.48
Recall	0.51	0.49	0.49	0.42	0.50	0.48
AUC	0.51	0.50	0.48	0.49	0.48	0.48

3.2 Performance of LSTM Model

The LSTM model performed slightly better than the benchmark, as most evaluation metrics were above 0.5, as shown in **Table 3.3**. However, even its best performance, achieved from the 0 to 1 minute window before VT, only yielded an average F1-score of 0.56 and an accuracy of 0.55. These values indicate that the model was not sufficiently accurate for clinical applications. The LSTM model directly processed raw ECG signals as input and relied on its architecture to capture temporal dependencies in the data. However, the results suggest that raw ECG signals alone may lack sufficient discriminative power for VT prediction, possibly due to high variability and noise in real-world data. Further investigation into feature selection, model architecture, and advanced data preprocessing techniques could potentially enhance predictive accuracy.

Table 3.3 Evaluation metrics of the LSTM model with different time interval before VT.

Evaluation Metrics	Time before VT onset					
	0-10 s	0-30 s	0-1 min	0-2 min	0-3 min	0-5 min
F1 Score	0.53	0.48	0.56	0.52	0.54	0.51
Accuracy	0.56	0.54	0.56	0.55	0.55	0.55
Precision	0.57	0.61	0.57	0.56	0.56	0.58
Recall	0.56	0.54	0.56	0.55	0.56	0.55
AUC	0.53	0.50	0.54	0.52	0.53	0.52

3.3 Performance of SVM Model with VAE Feature Extraction

3.3.1 *Performance of Prediction*

Using VAE for feature extraction, four different feature dimensions 5, 10, 20, and 30 were selected for comparison. These extracted latent features were then fed into an SVM model for prediction. **Figure 3.1** shows the median ECG beat for both classes, along with their reconstructed signals using different numbers of features extracted by the VAE. It could be observed that when the number of features was too small, such as 5, the P wave couldn't be effectively captured and reconstructed. Also, in the control signal, the T wave failed to be reconstructed with the same amplitude as the original signal. However, as the feature dimension increased, the reconstructed signals became more consistent across different feature settings. Notably, the reconstruction introduced some noise in the last 50 data points. However, since this occurred after the T wave, which is less critical for classification, it is unlikely to have significant impact on model's performance. Furthermore, since the noise appeared in both the control and event classes, its effect on the prediction results was expected to be negligible.

To support these visual insights with quantitative evidence, **Table 3.4** presents the test-set reconstruction mean squared errors (MSE) across latent feature dimensions. The reconstruction MSE for 5 features was substantially higher (0.1010) compared to that of 10 features (0.0629), indicating that a very low-dimensional latent space may not adequately capture key morphological details. In contrast, the difference in reconstruction error between 10, 20, and 30 features was minimal, suggesting that when the feature dimension is sufficiently large, such as 10 or more, the reconstructed signals achieve comparable quality across different settings.

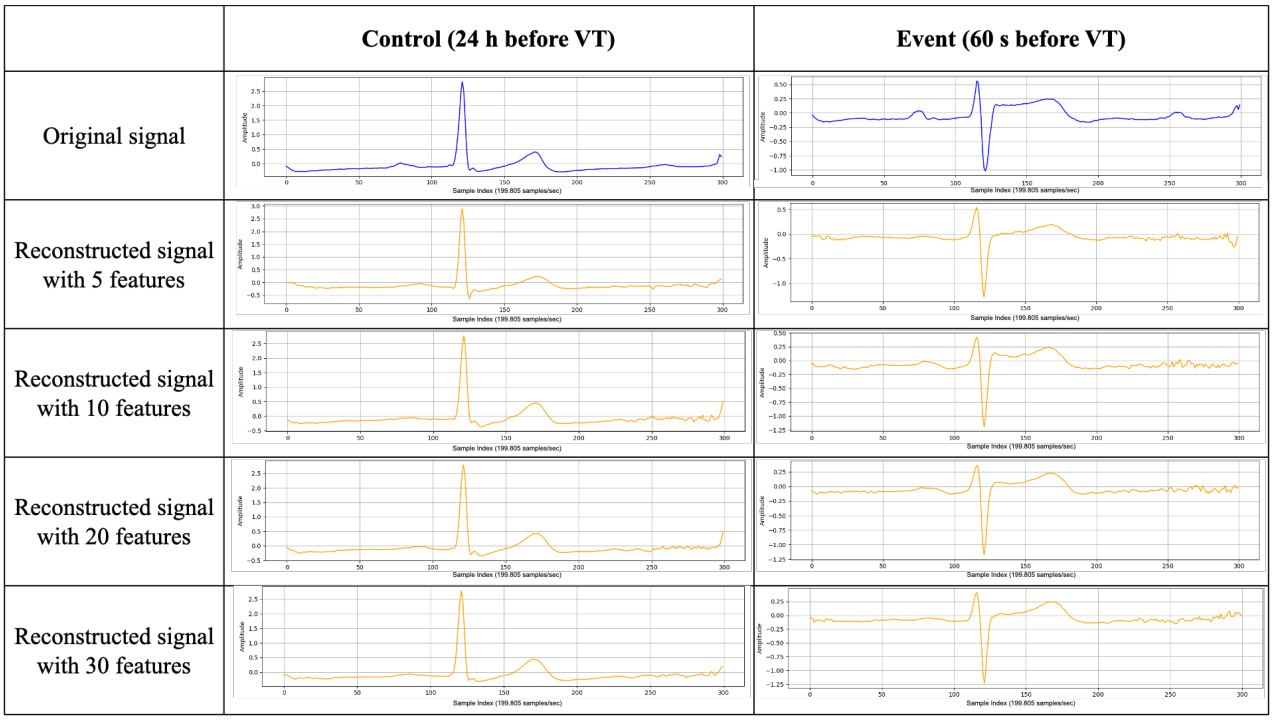

Figure 3.1 Comparison of reconstruction with different feature dimensions of VAE. The figure illustrates the ECG median beats for both classes and their corresponding reconstructions using different numbers of features extracted by VAE.

Table 3.4 Test-Set Reconstruction Mean Squared Error (MSE) Across Different Latent Feature Dimensions.

Latent Feature Dimension	Test Set Reconstruction MSE
5	0.1010
10	0.0629
20	0.0529
30	0.0524

Table 3.5 summarizes the evaluation metrics of SVM prediction across different feature dimensions and time intervals. The results indicate that when using 10 features extracted by VAE and a time interval of 0 to 1 minute before VT, the model achieved the best performance, with an F1 score of 0.66 and a recall of 0.77. The best SVM parameters ($C = 1.1$, $\gamma = 0.09$) were selected via GridSearchCV. Interestingly, these results suggest that increasing the latent dimensionality does not necessarily lead to better performance. This may be due to the introduction of redundant or noisy features in higher-dimensional latent spaces. Furthermore, with limited training data, higher-dimensional representations may increase the risk of overfitting, leading the model to capture noise rather than meaningful patterns.

The **Figure 3.2** and **Figure 3.3** illustrates the corresponding ROC curve and confusion matrix, providing further insights into the model's prediction performance. **Figure 3.2** shows the ROC curves from the 5-fold cross-validation, with individual AUC values ranging from 0.56 to 0.62. The solid blue line represents the mean ROC curve across the 5 folds, with an AUC of 0.59. **Figure 3.3** shows the confusion matrix, derived from the best-performing fold evaluated on the test set. The model correctly identified 52 out of 64 positive samples and 28 out of 64 negative samples, resulting in a true positive rate of 81.25% and an accuracy of 0.63. These two figures reflect the model's predictive performance on previously unseen data.

Table 3.5 Evaluation metrics of the SVM model with VAE feature extraction of different time interval before VT.

Feature of VAE	Evaluation metrics	Time before VT onset					
		0-10 s	0-30 s	0-1 min	0-2 min	0-3 min	0-5 min
5	F1 Score	0.59	0.49	0.60	0.58	0.53	0.45
	Accuracy	0.57	0.56	0.58	0.59	0.57	0.56
	Precision	0.57	0.62	0.57	0.59	0.57	0.62
	Recall	0.65	0.54	0.69	0.60	0.58	0.42
	AUC	0.54	0.53	0.55	0.58	0.54	0.52
10	F1 Score	0.48	0.59	0.66	0.57	0.59	0.48
	Accuracy	0.58	0.59	0.61	0.60	0.59	0.56
	Precision	0.64	0.59	0.59	0.62	0.59	0.60
	Recall	0.43	0.62	0.77	0.57	0.64	0.48
	AUC	0.56	0.57	0.59	0.60	0.59	0.54
20	F1 Score	0.49	0.53	0.59	0.57	0.59	0.58
	Accuracy	0.56	0.57	0.57	0.61	0.59	0.58
	Precision	0.60	0.59	0.56	0.63	0.59	0.60
	Recall	0.47	0.56	0.66	0.54	0.61	0.62
	AUC	0.53	0.55	0.54	0.60	0.59	0.55
30	F1 Score	0.42	0.46	0.62	0.58	0.42	0.58
	Accuracy	0.55	0.57	0.58	0.60	0.56	0.59
	Precision	0.63	0.67	0.57	0.61	0.63	0.61
	Recall	0.39	0.40	0.74	0.58	0.36	0.58
	AUC	0.52	0.54	0.55	0.60	0.52	0.55

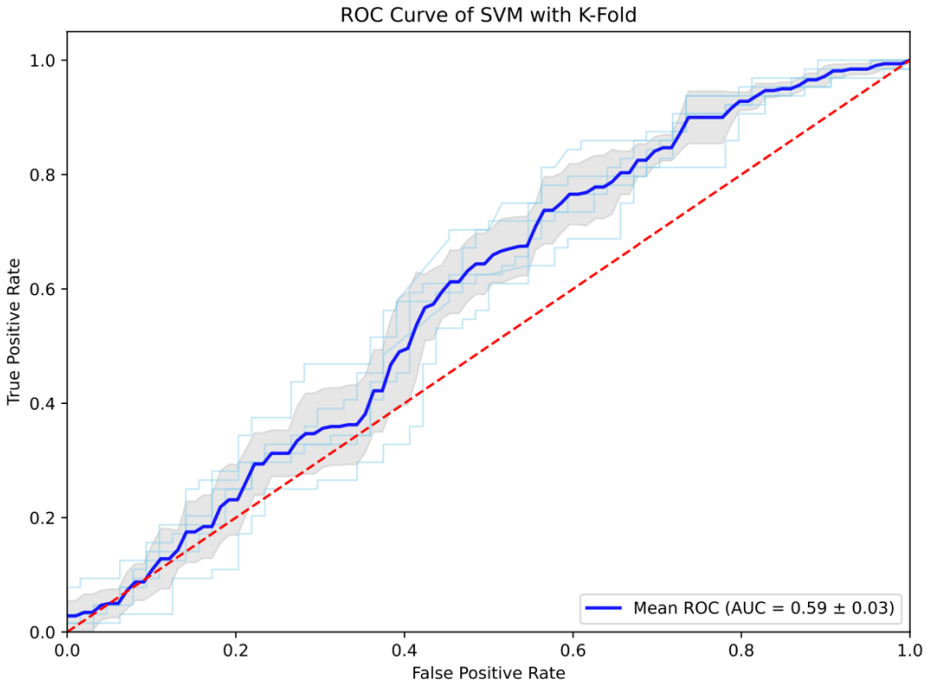

Figure 3.2 ROC curves from 5-fold cross-validation. This figure illustrates the ROC curves obtained from the five folds in the cross-validation process. The solid blue line represents the average ROC curve with an AUC of 0.59, accompanied by a standard deviation of 0.03.

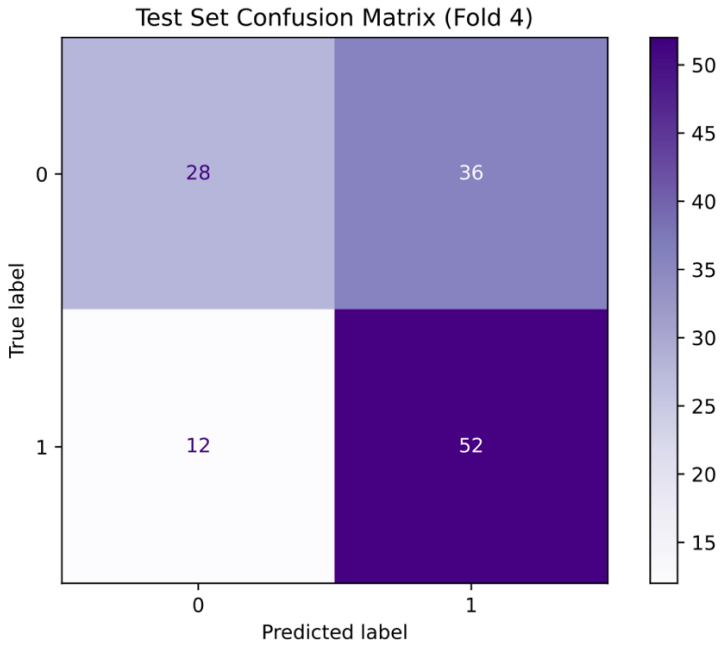

Figure 3.3 Confusion matrix from test set. This figure displays the confusion matrix obtained from test set.

3.3.2 *Correlation Between Features and ECG Measurements*

For the optimal parameter selection, where 10 features were identified as the best choice, their correlation with various ECG measurements were analyzed, including the PR interval, QRS complex, and QT interval. These measurements were obtained from median beats using the Python package NeuroKit2 [33]. NeuroKit2 is an advanced biosignal processing package that extracts the onset points of P, Q, R, S, and T waves and calculates the three intervals accordingly. The correlation plot was presented as a heatmap in **Figure 3.4**, where postitive value indicates a positive Spearman correlation coefficient and negative value represents a negative correlation. For each ECG measurement, the VAE feature with the highest absolute correlation was identified. Specifically, PR interval exhibited the strongest correlation with feature 6, QRS complex was most correlated with feature 7, and QT interval showed the highest correlation with both feature 1 and feature 7.

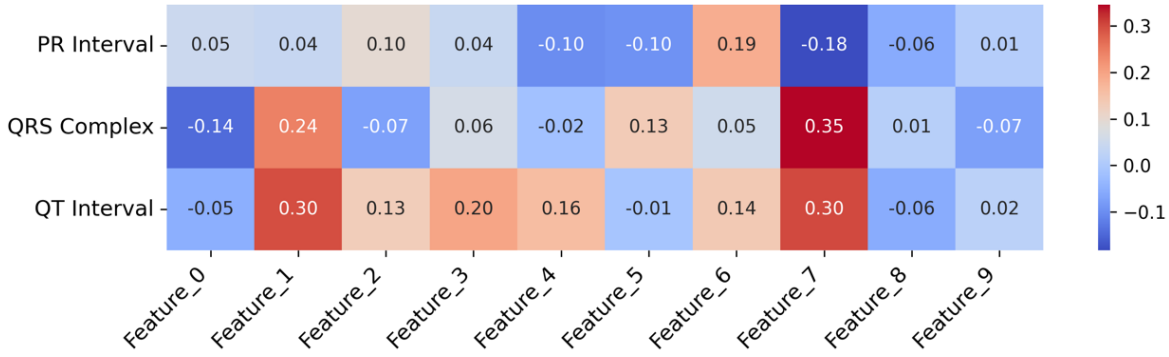

Figure 3.4 Heatmap of VAE Feature-ECG Measurement Correlations. This figure presents the Spearman correlation between individual VAE-derived features and three ECG measurements. The color represents the Spearman correlation coefficient, with values>0 indicating positive correlations and values<0 indicating negative correlations.

For the feature that were most correlated with PR interval, QRS complex, and QT interval, further latent traversal analysis was conducted by reconstructing the ECG signals using VAE decoder with different offsets added on those features, ranging from -5 to 5. This method was used to evaluate how variations in individual latent features affected ECG measurements. The correlation was evaluated by calculating the mean ECG measurements of the median beat from the reconstructed signals across all patients for each offset. Subsequently, the values across different offsets were integrated, and a Pearson correlation test was conducted to assess the relationship between the offsets and the ECG measurements, as shown in the left panel of **Figure 3.5**. In addition, another heatmap was generated without averaging the data. This plot directly visualized all individual data points without computing the mean ECG measurements. By presenting the complete dataset, it offers a detailed view of variations across different offsets and patients, as shown in the right panel of **Figure 3.5**.

Figure 3.5(a) represents the correlation between the PR interval and VAE feature 6. Although the heatmap in Figure 9 indicated a positive correlation coefficient, the correlation observed with offsets was not apparent, and the heatmap appeared disorganized, failing to provide clear and conclusive information about the relationship.

Figure 3.5(b) represents the correlation between the QRS complex interval and VAE feature 7. The Pearson correlation coefficient ($r = 0.95$) in the left figure indicated a very strong positive linear relationship between the feature 7 and QRS complex. Additionally, the p-value was smaller than 0.05, which confirmed that this correlation was statistically significant. The heatmap also demonstrated that the QRS complex value generally increased as the value of feature 7 increased.

Figure 3.5(c) shows the correlation between the QT interval and both VAE feature 1 and feature 7, as they both have the highest correlation coefficient values. While the average QT

interval showed a strong correlation with feature 7 (second row of (c), r = 0.95), the heatmap for feature 1 appeared more concentrated (first row of (c)), indicating a better distribution of all data. Specifically, in the negative offset of feature 7 in the heatmap, there was a clear separation into two distinct values, whereas for feature 1, individual variations are smaller. This suggested that while feature 7 showed a strong correlation with the QT interval, its distribution might introduce instability in the model. In contrast, feature 1 provided a more uniform distribution, which could serve as a more stable indicator for further analysis.

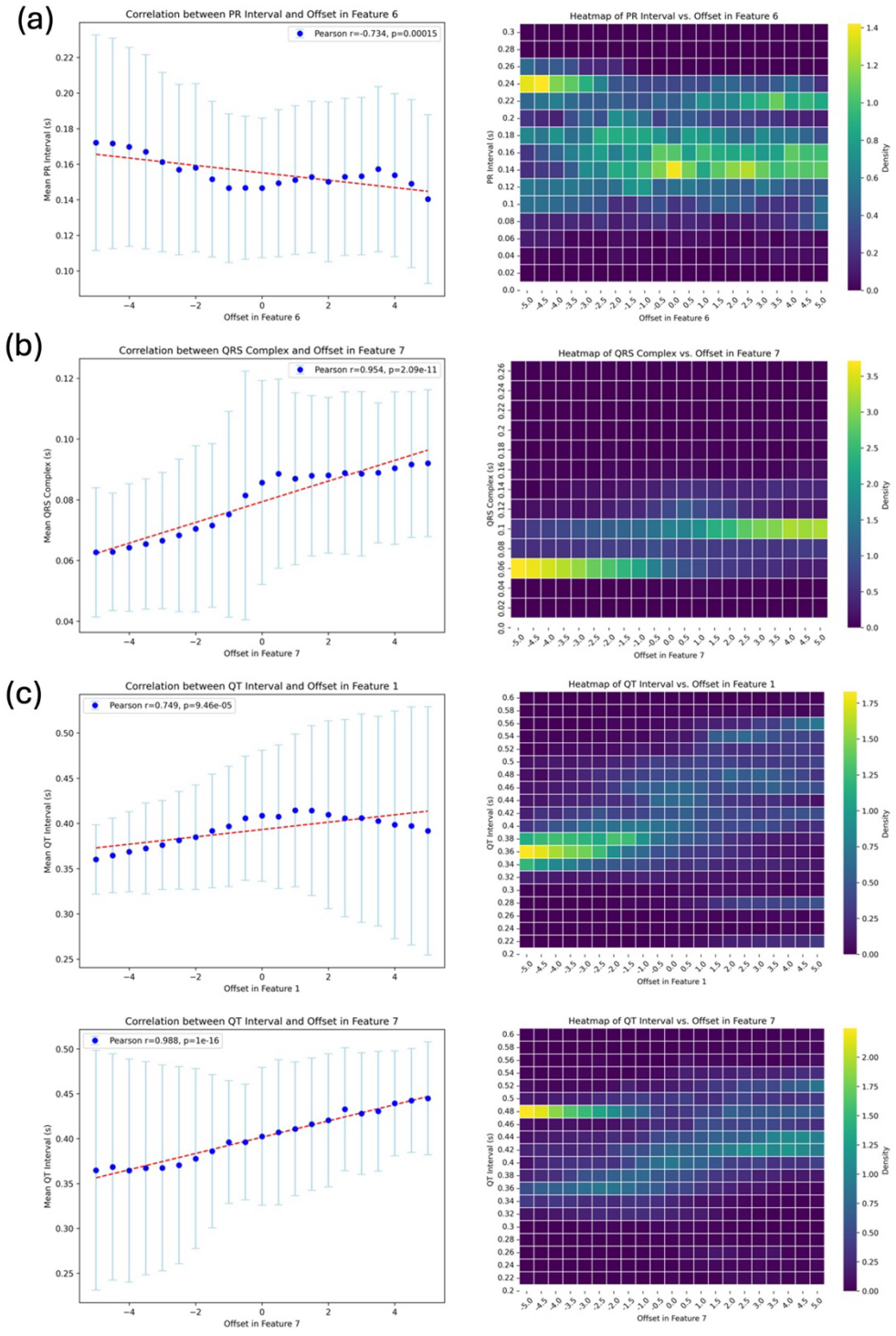

Figure 3.5 ECG Measurement Variations with Offsets on the Most Correlated VAE Features. (a) PR interval with feature 6, (b) QRS complex with feature 7, (c) QT interval with features 1 and 7. (Left: Pearson correlation of mean ECG measurements across offsets. Right: Heatmap showing ECG measurement variations across patients and offsets.)

For the group with the highest correlation in **Figure 3.4**, which is between Feature 7 and the QRS complex, ECG measurements were quantified, and the actual reconstruction results with varying offsets are plotted in **Figure 3.6**. This visualization allowed for the observation of subtle changes in the median beat during latent traversal. The black line represented the reconstructed signal without offsets, the red line corresponded to reconstruction with a positive offset applied to the feature, and the blue line showed reconstruction with a negative offset. These results clearly demonstrated that the duration of QRS complex was influenced by the value of Feature 7. Specifically, as the feature's value increased, distinct shifts in the Q wave onset and R peak position were observed. For instance, the positive offset caused the Q wave to begin much earlier (around point 90) compared to the negative offset (around point 110). It indicated that Feature 7 played a key role in modulating the timing and shape of the QRS complex, potentially encoding important structural characteristics. This analysis also suggests that the latent features have captured meaningful physiological patterns, with some correlating to known physiological metrics, and are therefore not merely random noise.

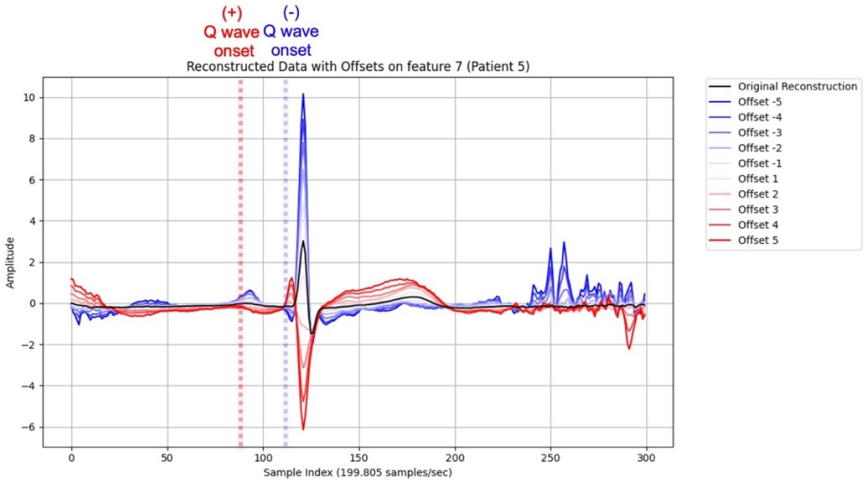

Figure 3.6 ECG Median Beat Reconstruction by VAE Decoder with Offsets on Feature 7. This figure illustrates that applying different offsets to Feature 7 results in changes to the reconstruction of the ECG, primarily affecting the duration and onset position of the QRS complex.

All features were also compared on the same patient, as shown in **Figure 3.7**. From the reconstructed signals with offsets applied to different features, it was observed that each feature influenced a specific aspect of the median beat. For example, feature 6 and feature 7 primarily affected the magnitude of the R peak, while feature 5 mainly affected the magnitude of the Q wave. However, although the SVM prediction experiment suggested that using 10 features yielded the best results, these features were not entirely independent. If they had been truly independent, each feature would have corresponded to a distinct region of the median beat. Instead, multiple features were observed to influence the same region of the ECG. For instance, features 1, 2, 5, and 6 exhibit strong changes in both the magnitude and position of the T wave, indicating that these features collectively influence the T wave. There might be two explanations for this observation. First, it could indicate redundancy in the latent space, where several features capture overlapping aspects of the signal. But alternatively, it may reflect that the model has uncovered subtle, physiologically meaningful patterns in the waveform that are not easily visible to the human eye or identifiable through conventional ECG metrics.

Additionally, it was notable that applying the offset amplified the noise outside the PQRST complex of the median beat. Although this did not affect the primary waveform, it may have had potential influence on the prediction results.

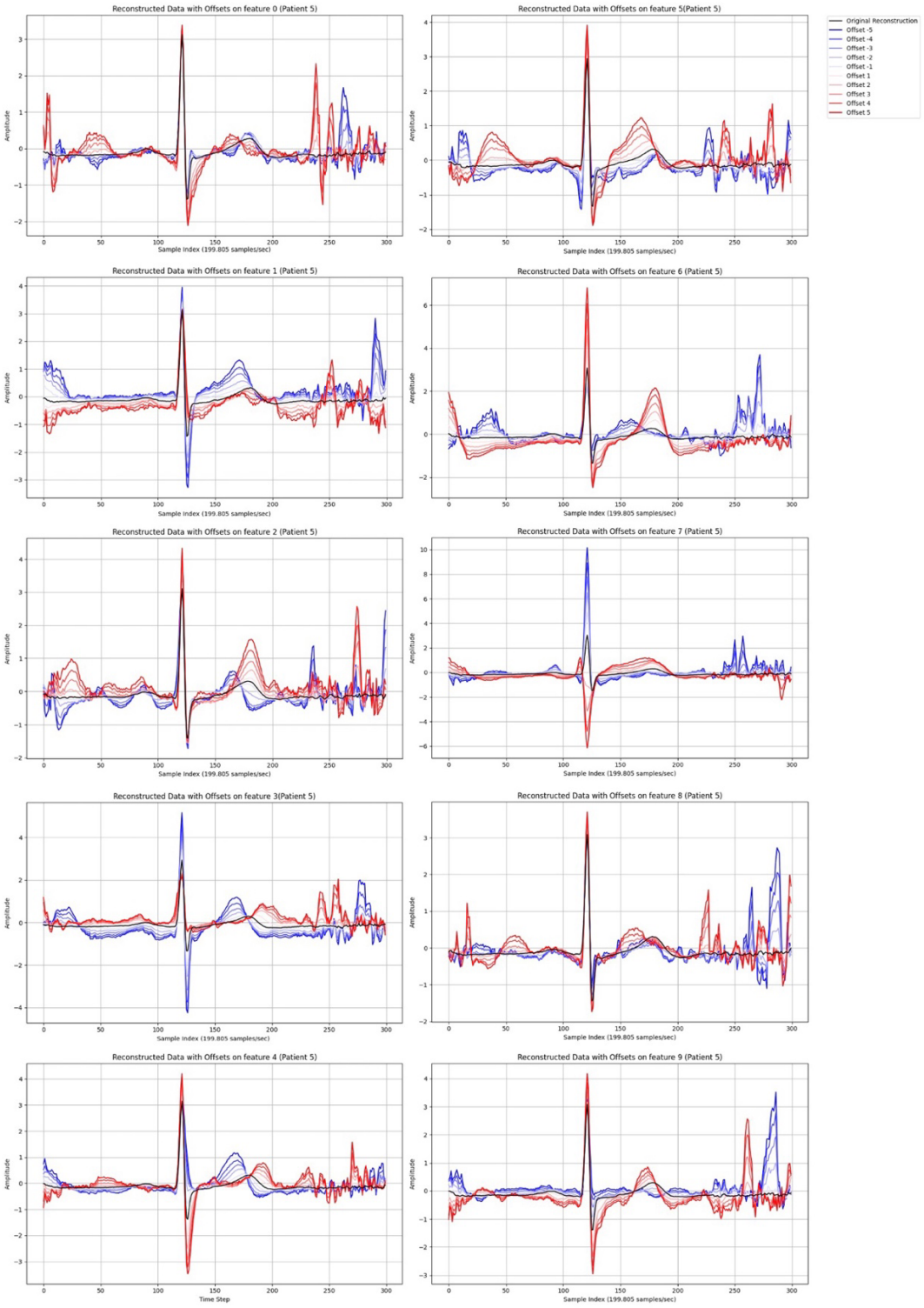

Figure 3.7 Visualization of ECG Median Beat Reconstruction by VAE decoder with Latent Traversal. This figure shows the reconstruction of all VAE features with latent traversal. It indicates that applying different offsets to each feature affects different parts of the median beat respectively.

3.3.3 *SHAP Summary for Features*

The explainable AI technique, SHAP, was used to evaluate the contribution of each VAE feature to the prediction. The summary plot ranked feature by importance and visualized the distribution of SHAP values, as shown in **Figure 3.8**. Each point represented a SHAP value for a specific instance, with red indicating high feature values and blue indicating low ones, helping to reveal their influence on the model's decisions. It was observed that some features, such as feature 7, exhibited a clear pattern in which higher feature values negatively impacted the model output. In contrast, some features, like feature 9, didn't show a distinct color distribution, suggesting minimal or no influence on the model's predictions.

This figure could be compared to **Figure 3.7** to identify patterns in the most important features determined by SHAP, providing deeper insights into how their values influenced the model's predictions. For example, SHAP identified feature 7 as the most influential, and latent traversal analysis revealed that it is related to the QRS complex. This suggests that variations in QRS duration may be important for identifying pre-VT cases.

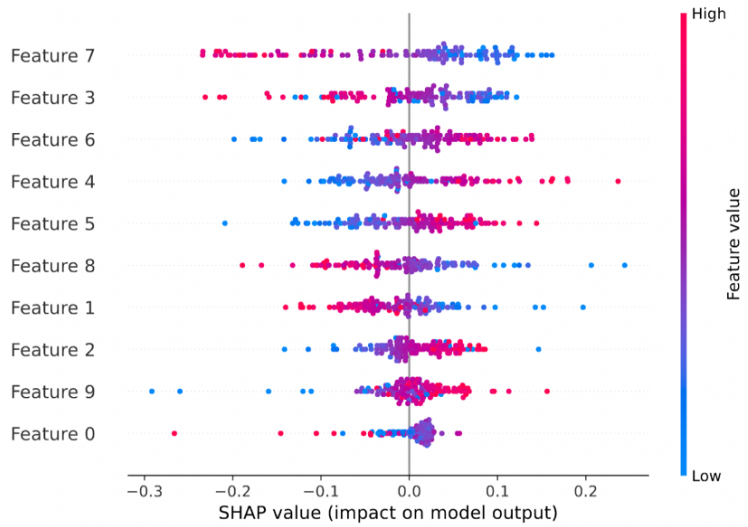

Figure 3.8 SHAP Summary Plot of All Features. This figure illustrates the impact of each feature on the model's predictions, with features ranked by importance from top to bottom.

3.4 Performance of 2D CNN Model

In addition to the model that uses 1D signals or features as input, the data was also transformed into a spectrogram, which was then used as input for a 2D CNN model. The best results from this 2D model are presented in **Table 3.6**. As shown in the table, even the highest F1 score achieved was only 0.5, with a corresponding AUC of 0.5, indicating limited performance.

To further evaluate whether models with pre-trained parameters could yield better results, three pre-trained architectures, DenseNet121, InceptionV3, and ResNet121V2, were applied, as summarized in **Table 3.7**. Among all the results, the DenseNet121 model DenseNet121 demonstrated the best performance when applied to data from 0 to 10 seconds prior to VT onset. It's worth noting that the CNN model, which was trained specifically on our dataset, had slightly lower performance compared to DenseNet121. However, despite some improvement with the pre-trained models, the results are still insufficient for real-world clinical applications, where higher accuracy and reliability are necessary for patient diagnosis and decision-making.

Table 3.6 Evaluation metrics of the 2D CNN model with different time interval before VT.

Evaluation metrics	Time before VT onset					
	0-10 s	0-30 s	0-1 min	0-2 min	0-3 min	0-5 min
F1 Score	0.45	0.46	0.50	0.44	0.41	0.41
Accuracy	0.53	0.53	0.54	0.52	0.53	0.52
Precision	0.60	0.49	0.57	0.51	0.39	0.38
Recall	0.53	0.53	0.54	0.53	0.53	0.52
AUC	0.48	0.50	0.50	0.49	0.52	0.51

Table 3.7 Evaluation metrics of the three pre-trained 2D CNN model, which are DenseNet121, InceptionV3, and resnet121v2.

Evaluation metrics	Time before VT onset					
	0-10 s	0-30 s	0-1 min	0-2 min	0-3 min	0-5 min
DenseNet121						
F1 Score	0.52	0.41	0.42	0.48	0.38	0.40
Accuracy	0.55	0.52	0.52	0.52	0.51	0.51
Precision	0.58	0.61	0.50	0.47	0.37	0.57
Recall	0.55	0.52	0.52	0.52	0.51	0.51
AUC	0.54	0.46	0.47	0.49	0.49	0.45
InceptionV3						
F1 Score	0.46	0.47	0.41	0.46	0.49	0.50
Accuracy	0.52	0.52	0.52	0.52	0.53	0.54
Precision	0.50	0.53	0.64	0.58	0.55	0.55
Recall	0.52	0.52	0.52	0.52	0.53	0.54
AUC	0.49	0.47	0.46	0.49	0.49	0.51
Resnet121v2						
F1 Score	0.50	0.48	0.46	0.47	0.45	0.49
Accuracy	0.55	0.54	0.54	0.53	0.52	0.55
Precision	0.62	0.58	0.45	0.55	0.55	0.51
Recall	0.55	0.54	0.54	0.53	0.52	0.54
AUC	0.52	0.48	0.48	0.48	0.48	0.52

3.5 Comprehensive Comparison of All Methods

Overall, the best result from each method is presented in **Figure 3.9**. The top panel shows the results across multiple evaluation metrics (F1 score, accuracy, precision, and recall) for all models. Notably, the VAE+SVM model achieved the best overall performance with an F1 score of 0.66 and a recall of 0.77, which are the highest among all models. The bottom panel focuses on mean AUC scores obtained via 5-fold cross-validation, with 95% confidence intervals shown as error bars. Again, the VAE+SVM model achieved the highest AUC (0.59 ± 0.03), demonstrating the robustness of its performance.

These findings suggest that the latent representations learned through the VAE framework effectively capture clinically relevant features for VT prediction.

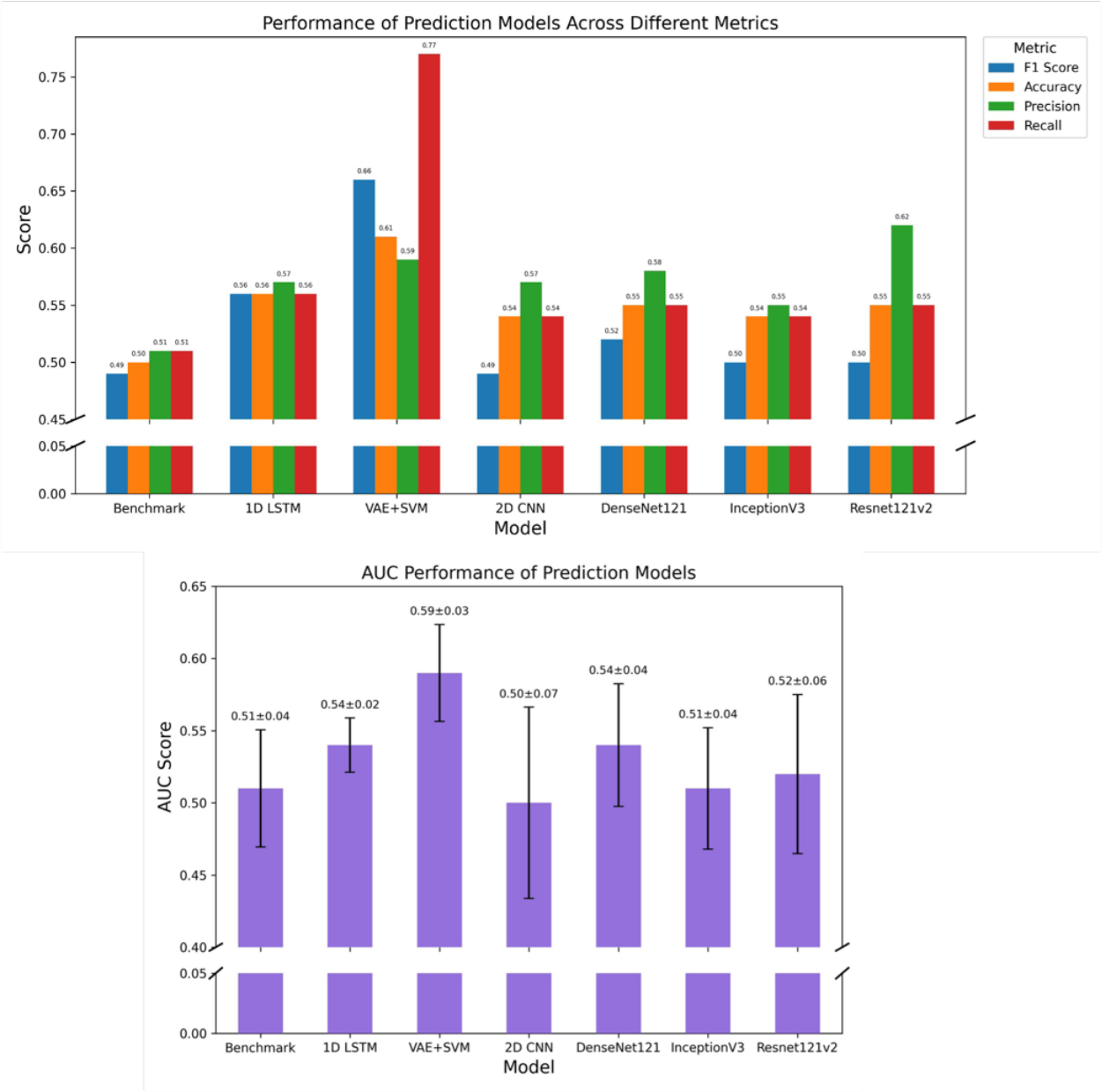

Figure 3.9 Summary of VT Prediction Performance. This figure summarizes the best results from each model for VT prediction task.

To highlight the distinction between our prediction task and the conventional classification of VT and non-VT, an additional comparative analysis was conducted between VT and control groups using the same models. This comparison enabled an evaluation of model performance in a more traditional and relatively simpler classification setting, thereby providing a contrast to the more challenging predictive task. The results are shown in **Figure 3.10**. Notably, the VAE model was not applicable to the classification task, as VT episodes were very brief, and the resulting median beats tended to be dominated by normal rhythms, leading to a loss of critical VT-related information.

Interestingly, the results of classification differed significantly from those of the prediction task. The most basic model, the benchmark model, achieved a remarkably high F1 score of 0.91 and an AUC of 0.96. The results indicate that due to the significant difference between normal and VT beats, even the simplest model can effectively classify the cases, sometimes outperforming more complex models. However, for tasks involving more subtle signal variations, more advanced techniques such as data preprocessing, transformation, and feature extraction are required to capture underlying patterns. In such cases, more sophisticated models, such as VAE and CNN are likely to be more effective.

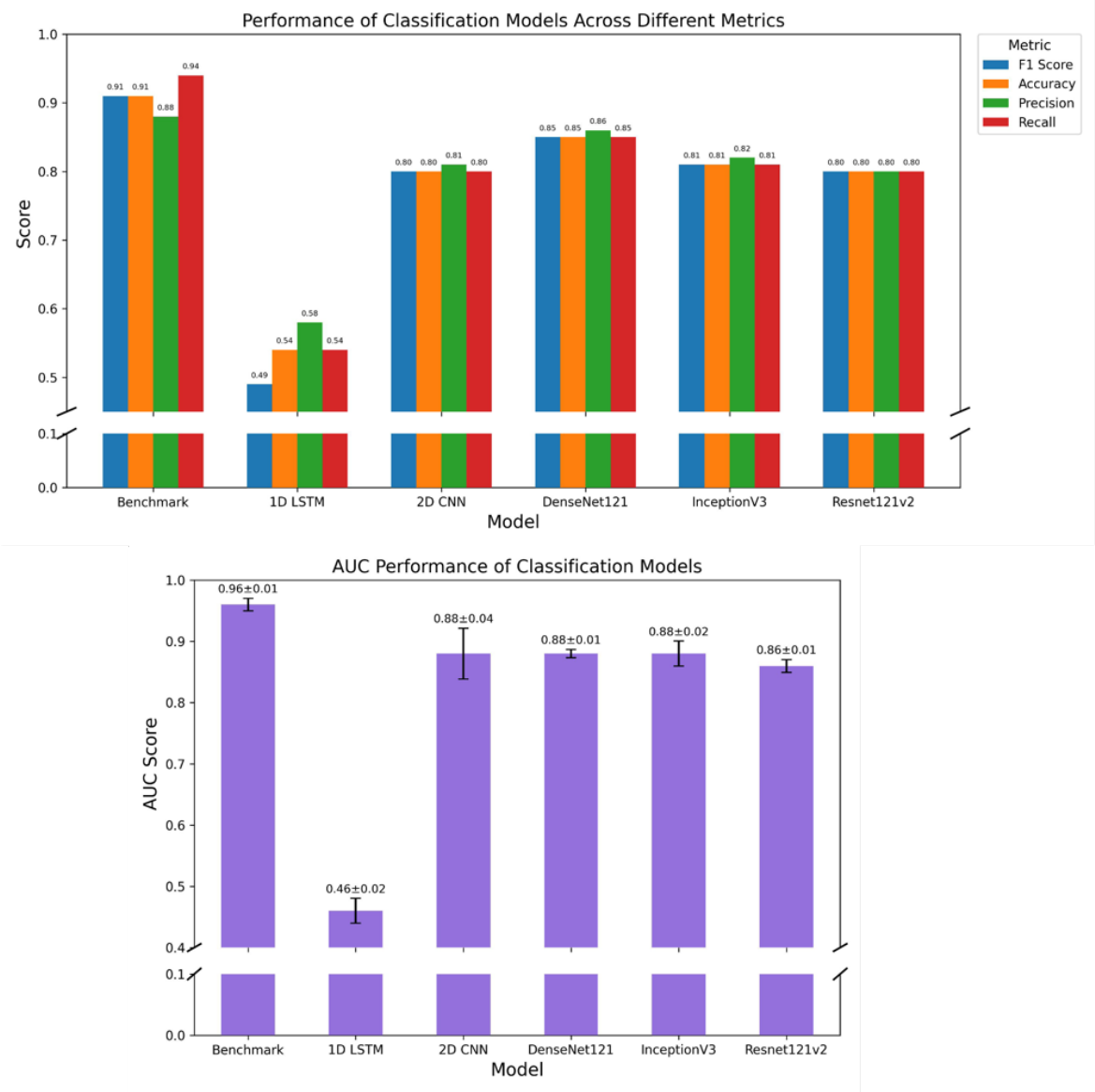

Figure 3.10 Summary of VT Classification Performance. This figure summarizes the best results from each model for VT classification task.

Chapter 4. DISCUSSION AND CONCLUSION

4.1 DISCUSSION

This study demonstrates the potential of VAE-based machine learning framework in predicting the onset of VT. By analyzing the segments of ECG data collected 1 minute before VT from a portable monitoring device in out-of-hospital setting, the model is able to estimate the possibility of an imminent VT episode.

We would like to discuss a potential explanation for why using data from 1 minute prior to the onset of VT leads to higher predictive accuracy compared to using data from closer segments, such as 10 seconds before the event. In our dataset, VT events primarily occurred in patients who were relatively stable and outside the hospital setting. This suggests that many of the observed episodes were likely focal VT, which is more common in such patients. Focal VT is usually driven by triggered activity, particularly delayed afterdepolarizations (DADs), which are closely associated with intracellular calcium (Ca^{2+}) overload [34]. Although the onset of focal VT may appear sudden, the underlying Ca^{2+} imbalance may develop progressively, potentially causing subtle electrophysiological changes prior to the event. These gradual changes may begin earlier than the final seconds before VT onset, making the 0-1 minute window more informative and predictive than the 0-10 second window.

The VAE model, through its latent space representation, appears well-suited to capture these early, nuanced temporal dynamics. Compared to other supervised models, the VAE's unsupervised encoding allows it to extract more informative patterns from the signal. Given that SVM performs well in separating non-linear patterns, the combination of VAE and SVM likely enables more effective discrimination between pre-VT and normal segments. This synergy between feature

extraction and robust classification may explain the superior performance observed in our experiments.

These results are particularly meaningful in the context of real-world applications, where access to high-quality clinical data is often limited. Existing research has primarily relied on hospital-acquired data, which tend to be more precise, or incorporated additional demographic and respiratory information to enhance model performance. In contrast, this study focuses solely on single-lead ECG data obtained from a portable device, making it more practical for patients outside of clinical environments. Although these individuals may not face immediate danger, a sudden VT episode remains a serious and potentially life-threatening event. Therefore, this study provides a valuable contribution toward developing early warning systems for remote monitoring.

In addition to improving predictive performance, the VAE model also improves explainability through latent space visualization. By observing how changes in individual latent dimensions affect the morphology of the ECG signal, clinicians can gain insight into what the machine learning model has learned. This interpretability is vital for clinical acceptance and real-world applicability.

Although the overall AUC and accuracy were modest, the results demonstrated that using VAE-based feature extraction significantly improved model performance compared to the benchmark model. An important consideration, however, is the trade-off between sensitivity and specificity. Our VAE model was designed to prioritize sensitivity with a high recall of 0.77 as it is crucial in medical risk prediction, particularly for life-threatening conditions like VT, to avoid missing potential events. However, this prioritization comes at the cost of increased false positives. In continuous monitoring settings, this may lead to alarm fatigue and reduced user trust. Since patients are in non-VT states for most of the time, even a modest false positive rate can result in a

large number of false alarms. Consequently, while the model shows strong potential for early warning, additional refinement or post-processing strategies may be necessary to reduce false positives and ensure practical utility in real-world deployment.

Overall, while this approach is not yet ready for clinical implementation, it offers a promising direction for future research into practical early warning systems for out-of-hospital patient populations. Further investigation with larger datasets and clinical validation is needed to confirm these preliminary results and assess their applicability in real-world healthcare settings.

4.2 LIMITATIONS

There are some limitations to this research.

First, the dataset is relatively small. After filtering out noisy recordings, only 422 patients remained, each with one VT episode validated by physicians. Although deep learning models can still operate with limited samples, their performance is inevitably constrained compared to models training on larger datasets.

Second, the available data consists solely of ECG signals. While this represents a key strength and novelty of the study, it also increases the challenge of accurately predicting VT in the absence of additional clinical or demographic information. For instance, if demographic information were available, further analyses, such as examining correlations between age and outcomes, could be conducted to potentially enhance the model's predictive power.

Finally, the quality of the data is limited, as it is collected from a portable ECG device. Unlike hospital-monitored ECGs, portable devices are more susceptible to long-term noise, often caused by improper placement or poor contact. This issue increases the difficulty of accurate prediction, given that manual inspection of every recording is impractical. Although data preprocessing

techniques were applied to mitigate most of the noise, some residual artifacts may remain, potentially impacting model performance.

4.3 Future Goals

To further improve model performance, the data preprocessing pipeline should be optimized. Given that feature extraction has significantly enhanced predictive accuracy, various techniques, including Transformer-based models, Recurrent Neural Networks (RNNs), and other advanced architectures, will be explored. By evaluating their effectiveness on ECG signals, the most suitable method for enhancing VT prediction can be identified.

To address the limitation of insufficient data, future work may focus on expanding the dataset or using meta-learning approaches. Meta learning [35], often referred to as "learning to learn", enables models to generalize more efficiently by leveraging prior knowledge from related tasks, rather than training from scratch for each new task. In addition, incorporating demographic variables such as age and sex, along with other clinical data, may further enhance model performance, as these factors are known to influence cardiac physiology and arrhythmia risk. This would also support subgroup analysis to explore risk variations across different populations.

Furthermore, additional explainable AI techniques will be applied to the ECG features extracted by the VAE. For example, Goettling et al. employed Deep Taylor Decomposition, an explainable AI method, to visualize the key regions of ECG beats that contributed to model prediction [36]. Applying similar visualization techniques to the current model may enhance interpretability, as it would allow physicians to understand the reasoning behind the model's decisions. This would not only increase physician trust in the system but also transition the model from a black-box approach to a more transparent and clinically acceptable tool.

4.4 CONCLUSION

This study presents a novel and practical approach for the early prediction of VT using single-lead ECG data collected from portable monitoring devices. Although the current method may not yet meet the standards required for clinical deployment, it demonstrates strong potential as an early warning system for patients in out-of-hospital settings. By identifying high-risk periods in advance, such a system could provide individuals with valuable time to prepare or seek medical assistance.

Despite current limitations in model performance and data quality, the proposed framework offers a promising foundation for future development. With continued refinement, validation, and the integration of explainable AI techniques, this approach holds significant potential to improve risk management in large-scale populations at high risk of VT.

REFERENCE

[1] B. A. Koplan and W. G. Stevenson, “Ventricular Tachycardia and Sudden Cardiac Death,” *Mayo Clin. Proc.*, vol. 84, no. 3, pp. 289–297, Mar. 2009, doi: 10.4065/84.3.289.

[2] W. G. Stevenson, “VENTRICULAR SCARS AND VENTRICULAR TACHYCARDIA”.

[3] A. Mincholé, J. Camps, A. Lyon, and B. Rodríguez, “Machine learning in the electrocardiogram,” *J. Electrocardiol.*, vol. 57, pp. S61–S64, Nov. 2019, doi: 10.1016/j.jelectrocard.2019.08.008.

[4] M. A. jabbar, B. L. Deekshatulu, and P. Chandra, “Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm,” *Procedia Technol.*, vol. 10, pp. 85–94, 2013, doi: https://doi.org/10.1016/j.protcy.2013.12.340.

[5] F. I. Alarsan and M. Younes, “Analysis and classification of heart diseases using heartbeat features and machine learning algorithms,” *J. Big Data*, vol. 6, no. 1, p. 81, Aug. 2019, doi: 10.1186/s40537-019-0244-x.

[6] R. R. van de Leur *et al.*, “Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders,” *Eur. Heart J. - Digit. Health*, vol. 3, no. 3, pp. 390–404, Jul. 2022, doi: 10.1093/ehjdh/ztac038.

[7] M. J. Blaha and A. P. DeFilippis, “Multi-Ethnic Study of Atherosclerosis (MESA),” *J. Am. Coll. Cardiol.*, vol. 77, no. 25, pp. 3195–3216, Jun. 2021, doi: 10.1016/j.jacc.2021.05.006.

[8] “The Multi-Ethnic Study of Atherosclerosis.” [Online]. Available: https://internal.mesa-nhlbi.org/about

[9] L.-M. Tseng and V. S. Tseng, “Predicting Ventricular Fibrillation Through Deep Learning,” *IEEE Access*, vol. 8, pp. 221886–221896, 2020, doi: 10.1109/ACCESS.2020.3042782.

[10] S. Khurshid *et al.*, “ECG-Based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation,” *Circulation*, vol. 145, no. 2, pp. 122–133, Jan. 2022, doi: 10.1161/CIRCULATIONAHA.121.057480.

[11] H. Lee, S.-Y. Shin, M. Seo, G.-B. Nam, and S. Joo, “Prediction of Ventricular Tachycardia One Hour before Occurrence Using Artificial Neural Networks,” *Sci. Rep.*, vol. 6, no. 1, p. 32390, Aug. 2016, doi: 10.1038/srep32390.

[12] J. Economou Lundeberg *et al.*, “Ventricular tachycardia risk prediction with an abbreviated duration mobile cardiac telemetry,” *Heart Rhythm O2*, vol. 4, no. 8, pp. 500–505, Aug. 2023, doi: 10.1016/j.hroo.2023.06.009.

[13] G. T. Taye, H.-J. Hwang, and K. M. Lim, “Application of a convolutional neural network for predicting the occurrence of ventricular tachyarrhythmia using heart rate variability features,” *Sci. Rep.*, vol. 10, no. 1, p. 6769, Apr. 2020, doi: 10.1038/s41598-020-63566-8.

[14] “PhysioNet. Spontaneous Ventricular Tachyarrythmia Database.” [Online]. Available: https://physionet.org/physiobank/database/mvtdb/

[15] A. Jovic and N. Bogunovic, “Random Forest-Based Classification of Heart Rate Variability Signals by Using Combinations of Linear and Nonlinear Features,” in *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010*, vol. 29, P. D. Bamidis and N. Pallikarakis, Eds., in IFMBE Proceedings, vol. 29. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 29–32. doi: 10.1007/978-3-642-13039-7_8.

[16] J. Schmidhuber and F. Cummins, “Learning to Forget: Continual Prediction with LSTM”.

[17] D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” Dec. 10, 2022, *arXiv*: arXiv:1312.6114. doi: 10.48550/arXiv.1312.6114.

[18] M. Awad and R. Khanna, “Support Vector Machines for Classification,” in *Efficient Learning Machines*, Berkeley, CA: Apress, 2015, pp. 39–66. doi: 10.1007/978-1-4302-5990-9_3.

[19] A. Desai, C. Freeman, Z. Wang, and I. Beaver, “TimeVAE: A Variational Auto-Encoder for Multivariate Time Series Generation,” Dec. 07, 2021, *arXiv*: arXiv:2111.08095. doi: 10.48550/arXiv.2111.08095.

[20] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[21] M. Munteanu, C. Rusu, L. Vladareanu, D. Petreus, V. Rusu, and M. Dobra, “EKG Analysis Using STFT Phase,” in *International Conference on Advancements of Medicine and Health Care through Technology*, vol. 26, S. Vlad, R. V. Ciupa, and A. I. Nicu, Eds., in IFMBE Proceedings, vol. 26. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 231–234. doi: 10.1007/978-3-642-04292-8_51.

[22] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” Dec. 02, 2015, *arXiv*: arXiv:1511.08458. doi: 10.48550/arXiv.1511.08458.

[23] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines”.

[24] “Keras.” [Online]. Available: https://github.com/keras-team/keras

[25] M. Abadi *et al.*, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” Mar. 16, 2016, *arXiv*: arXiv:1603.04467. doi: 10.48550/arXiv.1603.04467.

[26] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *Mach. Learn. PYTHON*.

[27] L. Buitinck *et al.*, “API design for machine learning software: experiences from the scikit-learn project,” Sep. 01, 2013, *arXiv*: arXiv:1309.0238. doi: 10.48550/arXiv.1309.0238.

[28] “VisualKeras.” [Online]. Available: https://github.com/paulgavrikov/visualkeras

[29] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 10, 2015, *arXiv*: arXiv:1512.03385. doi: 10.48550/arXiv.1512.03385.

[30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[31] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” Jan. 28, 2018, *arXiv*: arXiv:1608.06993. doi: 10.48550/arXiv.1608.06993.

[32] S. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” Nov. 25, 2017, *arXiv*: arXiv:1705.07874. doi: 10.48550/arXiv.1705.07874.

[33] D. Makowski *et al.*, “NeuroKit2: A Python toolbox for neurophysiological signal processing,” *Behav. Res. Methods*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021, doi: 10.3758/s13428-020-01516-y.

[34] S. M. Markowitz and B. B. Lerman, “Mechanisms of focal ventricular tachycardia in humans,” *Heart Rhythm*, vol. 6, no. 8, pp. S81–S85, Aug. 2009, doi: 10.1016/j.hrthm.2009.02.034.

[35] J. Vanschoren, “Meta-Learning,” in *Automated Machine Learning*, F. Hutter, L. Kotthoff, and J. Vanschoren, Eds., in The Springer Series on Challenges in Machine Learning. , Cham: Springer International Publishing, 2019, pp. 35–61. doi: 10.1007/978-3-030-05318-5_2.

[36] M. Goettling, A. Hammer, H. Malberg, and M. Schmidt, “xECGArch: a trustworthy deep learning architecture for interpretable ECG analysis considering short-term and long-term features,” *Sci. Rep.*, vol. 14, no. 1, p. 13122, Jun. 2024, doi: 10.1038/s41598-024-63656-x.