

Predicting Ventricular Arrhythmia in Myocardial Ischemia Using Machine Learning

Anna Busatto^{1,2,3}, Jake A Bergquist^{1,2,3}, Tolga Tasdizen^{1,4}, Benjamin A Steinberg^{5,6,7}, Ravi Ranjan^{2,3,8}, Rob S MacLeod^{1,2,3}

¹ Scientific Computing and Imaging Institute, University of Utah, SLC, UT, USA

² Nora Eccles Cardiovascular Research and Training Institute, University of Utah, SLC, UT, USA

³ Department of Biomedical Engineering, University of Utah, SLC, UT, USA

⁴ Department of Electrical and Computer Engineering, University of Utah, SLC, UT, USA

⁵ Department of Medicine, University of Utah, SLC, UT, USA

⁶ Department of Medicine, Denver Health Medical Center, Denver, CO, USA

⁷ Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

⁸ Division of Cardiology, University of Utah, SLC, UT, USA

Abstract

Ventricular arrhythmia frequently complicates myocardial ischemic events, sometimes to devastating ends. Accurate arrhythmia prediction in this setting could improve outcomes, yet traditional models struggle with the temporal complexity of the data. This study employs a Long Short-Term Memory (LSTM) network to predict the time to the next premature ventricular contraction (PVC) using high-resolution experimental data. We analyzed electrograms from 11 large animal experiments, identifying 1832 PVCs, and computed time-to-PVC. An LSTM model (247 inputs, 1024 hidden units) was trained on 10 experiments, with one held out for testing, achieving a validation MAE of 8.6 seconds and a test MAE of 135 seconds (loss 68.5). Scatter plots showed strong validation correlation and a positive test trend, suggesting the potential of this approach.

1. Introduction

Myocardial ischemia, a condition characterized by reduced blood flow to the heart muscle, often leads to severe and dangerous electrophysiological abnormalities. Premature ventricular contractions (PVCs) are ectopic heartbeats originating from the ventricles that can serve as precursors to more severe arrhythmias, including ventricular tachycardia or fibrillation [1, 2]. The unpredictable nature of ischemia-induced ventricular arrhythmias poses a significant challenge for timely clinical intervention, yet accurate prediction of their occurrence could enable proactive management strategies, possibly reducing the risk of adverse

cardiac events.

Traditional statistical approaches, such as linear regression or time-series analysis, often fail to capture the complex temporal dynamics inherent in electrophysiological data, particularly when dealing with high-dimensional signals like those recorded from multiple channels during experimental studies. These limitations have driven the exploration of advanced machine learning techniques, particularly deep learning models, which are better equipped to handle sequential and non-linear patterns. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, are particularly well-suited for this task due to their ability to model long-term dependencies in time-series data [3].

Prior research in arrhythmia prediction has often focused on classification tasks, such as detecting the presence of arrhythmia risk in ECG signals, using methods such as convolutional neural networks (CNNs) [4]. However, predicting the precise timing of the next arrhythmic event remains underexplored, despite the inherent benefit to inform clinical decision-making. This study addresses the gap in arrhythmia prediction by developing a machine learning pipeline to predict the time to the next PVC using LSTM-based time series analysis. We leverage high-resolution electrograms from *in situ* large animal experiments, specifically targeting ischemia-induced PVCs. Our dataset spans 11 experiments, from which we identified 1832 PVCs. Our objective is to create a robust, real-time predictive tool that can enhance clinical prediction capabilities, ultimately improving patient outcomes in the context of myocardial ischemia.

2. Methods

Data Acquisition and Preprocessing

The dataset comprises electrophysiological recordings from 11 large animal experiments, designed to study ischemia-induced arrhythmias, described previously [5]. Briefly, in each experiment, recordings were made simultaneously at 1 kHz from the epicardial surface, within the myocardium via plunge needle arrays, and the torso surface. For the purposes of this study, the epicardial recordings were used. Recordings were segmented into 15-second runs, capturing the electrophysiological changes during interventions of induced myocardial ischemia. Intervention periods—characterized by coronary artery occlusion and either electrical or pharmacological stimulation to replicate ischemic supply-demand mismatches—were identified, along with episodes containing PVCs. Across all experiments, a total of 1832 PVCs were detected, reflecting a diverse set of arrhythmic events.

For each heartbeat in a non-PVC run, we computed the time to the next PVC within the same intervention period, providing a continuous target variable for prediction. This time-to-PVC metric was calculated by identifying the temporal difference between the start of a given beat and the next run containing a PVC, ensuring alignment with the experimental protocol.

Model Training

An LSTM model was developed to predict the time to the next PVC, leveraging the sequential nature of the electrophysiological data. The model architecture included 247 input channels, corresponding to the number of recordings from the epicardial surface, and 1024 hidden units in a single layer. A softplus activation function was used at the output to ensure non-negative predictions, as time-to-PVC values are inherently positive. The model was trained on data from 10 experiments, with one experiment held out for testing to evaluate generalizability.

The training spanned 75 epochs, using an AdamW optimizer with a learning rate of 0.001 and a weight decay of 1e-5 to regularize the model [6]. A hybrid loss function was employed, combining mean squared logarithmic error (MSLE) and mean absolute error (MAE) with a weighting factor of 0.5, to balance sensitivity to both small and large prediction errors. This hybrid approach was chosen to address the wide range of time-to-PVC values observed in the dataset, which spanned from a few seconds to over 1000 seconds. A learning rate scheduler adjusted the learning rate dynamically based on validation performance to ensure convergence. Hyperparameters, including the hidden unit size, number of hidden layers, learning rate, loss function alpha, and batch size, were optimized through a systematic sweep to maximize validation performance.

Model Testing

The trained model was evaluated on the held-out experiment to assess its performance on unseen data, simulating a real-world scenario where the model must generalize to new experimental conditions. The best model checkpoint, selected based on the lowest validation loss achieved during training, was used for testing. Predictions were generated for the time-to-PVC values in the test set, and performance was measured using both MAE and the hybrid loss function used during training.

3. Results

The full dataset included 32,009 samples, representing non-PVC runs across the 11 experiments. The 10 experiments used in training were split into a training dataset of 24,773 samples and a validation dataset of 4,372 samples, reflecting an 85:15 ratio. Training achieved a validation mean absolute error (MAE) of 8.6 seconds at epoch 75, demonstrating strong predictive capability on the validation set. The validation scatter plot (Fig 2, top) shows a tight correlation between predicted and true time-to-PVC values, with most predictions aligning closely with the identity line ($R^2=0.9759$), indicating robust learning across the experiments. The distribution of errors reveals that 80.10% of validation predictions had an absolute error of less than 10 seconds, and 90.85% were within 20 seconds, highlighting the model's precision for shorter prediction horizons. However, for time-to-PVC values above 800 seconds, the model showed slightly increased dispersion, with an MAE of 49.5 seconds and a median absolute error of 22.2 seconds in this range.

Testing was performed on the held-out experiment, which was not included in the training or validation sets. The test set included 2,864 samples, representing approximately 9% of the full dataset. The test results yielded an MAE of 135 seconds and a loss of 68.5. The test scatter plot (Fig 2, bottom) exhibits a positive trend, with many points following the expected diagonal ($R^2=0.4880$), particularly for time-to-PVC values below 200 seconds. The MAE was 61.5 seconds, and the median absolute error was 53.8 seconds. However, greater dispersion is observed for larger time-to-PVC values, with errors increasing for values above 400 seconds, where the MAE rose to 306.8 and the median absolute error to 308.1 seconds. Overall, the model correctly predicted the time-to-PVC within a 10-second window for 5.52% of the test samples, within a 50-second window for 25.52% of the samples, and within a 100-second window for 45.39% of the samples. These metrics highlight the model's performance on unseen data and its ability to generalize, albeit with increased error compared to validation.

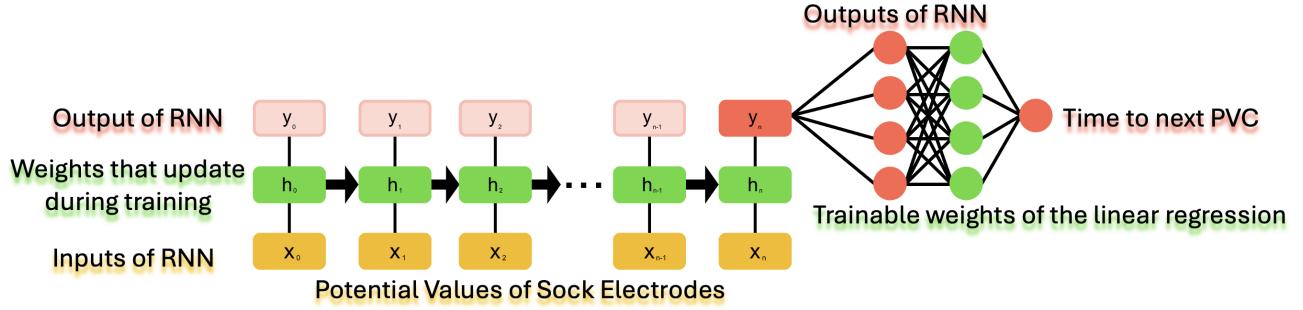


Figure 1: Network diagram.

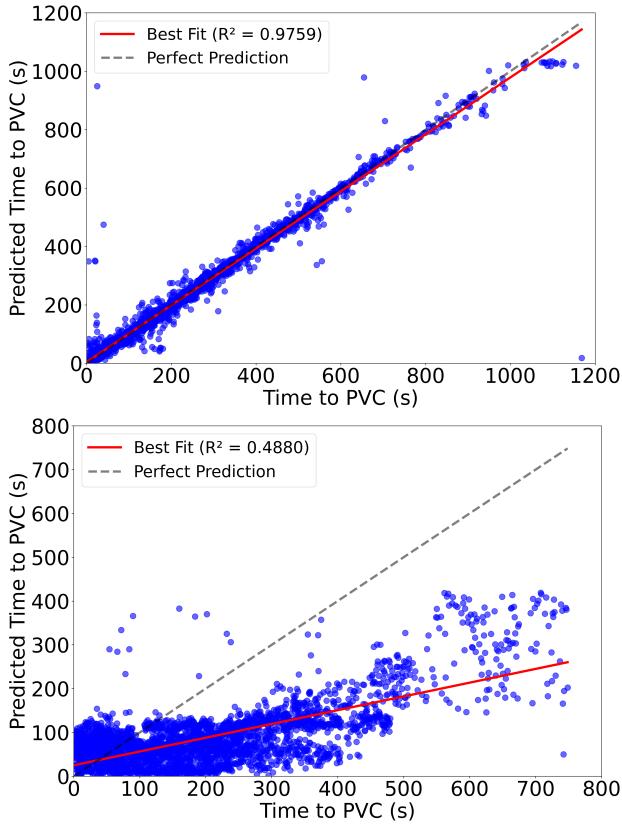


Figure 2: Scatter plots of predicted versus true time-to-PVC for validation (top) and for testing (bottom). Validation MAE: 8.6 s; Test MAE: 135 s, Loss: 68.5.

4. Discussion and Conclusions

The LSTM model's validation MAE of 8.6 seconds at epoch 75, based on a validation set of 4372 samples, underscores its ability to accurately predict time-to-PVC across experiments (Fig 2, top). The tight clustering along the identity line suggests the model effectively captures temporal dependencies in electrophysiological data. This performance is particularly notable given the complexity of the task, as predicting the precise timing of PVCs requires

modeling both short-term and long-term patterns in high-dimensional electrophysiological signals [7, 8]. Compared to prior work that achieved high accuracy in real-time arrhythmia detection using hybrid convolutional neural networks [9], our study advances the field by focusing on the timing of PVCs, a challenging task that provides actionable insights for clinical intervention. The validation performance, with an MAE of 8.6 seconds, is promising, as a 10-second prediction window could guide real-time interventions in myocardial ischemia, such as adjusting pacing strategies, administering anti-arrhythmic drugs, or preparing for potential escalation to more severe arrhythmias like ventricular tachycardia.

The test results, with an MAE of 135 seconds and a loss of 68.5 on the held-out experiment, indicate a positive trend (Fig 2, bottom), as many predictions align with true values, particularly for shorter intervals. The model's ability to predict within a 100-second window for 45.39% of the test samples is encouraging, showing that it retains predictive capability on unseen data. This trend is particularly evident for time-to-PVC values below 200 seconds, where the model achieves an MAE of 61.5 seconds and a median absolute error of 53.8 seconds, suggesting that it can provide reliable predictions for near-term events. However, the elevated test MAE highlights limitations, particularly for longer time-to-PVC values, where errors are higher, with a MAE of 306.8 seconds and a median absolute error of 308.1 seconds for values above 400 seconds. This discrepancy may stem from inter-experiment variability, such as differences in ischemic severity, electrode placement, or animal-specific physiological responses, which were not fully captured in the training data. Additionally, the under-representation of longer time-to-PVC events in the training data, as the majority of samples had time-to-PVC values below 400 seconds, likely contributes to the increased error for these cases. The model may also be overfitting to patterns specific to the training experiments, despite the use of weight decay for regularization.

The hybrid loss function, combining MSLE and MAE with a weighting factor of 0.5, balanced sensitivity to small and large errors, supporting the model's ability to learn

across a wide range of time intervals. The MSLE component ensured that the model prioritized relative accuracy for smaller time-to-PVC values, which are more clinically actionable, while the MAE component provided robustness for larger values, ensuring that the model did not overly penalize small relative errors in larger predictions. However, the high test loss suggests that further refinement is needed to improve generalizability. The network diagram (Fig 1) illustrates the model's architecture, highlighting its ability to process sequential inputs from the electrode array and produce continuous predictions through a linear regression layer. This design choice, while effective for capturing temporal dependencies, may benefit from additional mechanisms, such as attention layers, to better focus on critical time steps in the sequence, particularly for longer prediction horizons where the model currently struggles.

The clinical implications of this work are relevant, particularly in the context of myocardial ischemia, where timely intervention can prevent progression to life-threatening arrhythmias. PVCs can act as a harbinger of more severe arrhythmic events in ischemic conditions, emphasizing the need for predictive tools to enable proactive management [2]. These may include more aggressive interventions for ischemia or more aggressive anti-arrhythmic medical therapy. Even the test performance could provide a useful early warning system, allowing clinicians to monitor high-risk patients more closely and intervene prior to life-threatening arrhythmias.

Future improvements will focus on reducing test error by expanding the dataset beyond the current 29,145 samples, potentially by including more experiments with a broader range of time-to-PVC values to better represent longer prediction horizons. Incorporating additional features, such as beat morphology, could enhance the model's ability to capture complex patterns and improve prediction accuracy for longer intervals. For instance, beat morphology could provide insights into the electrophysiological changes preceding a PVC, and subsequently more significant downstream arrhythmia. Fine-tuning the model with more diverse test data could further enhance its generalizability across different experimental and clinical settings.

This pipeline demonstrates the potential of LSTM-based modeling for arrhythmia prediction in the setting of myocardial ischemia, offering a foundation for advanced arrhythmia management. By providing a predictive window, this approach could enable clinicians to intervene proactively, potentially reducing the risk of sustained, life-threatening arrhythmias. Continued refinement of the model and dataset will be critical to translating this technology into clinical practice.

Acknowledgments

Support for this research came from the Center for Integrative Biomedical Computing (www.sci.utah.edu/cibc), NIH/NIGMS grants P41 GM103545 and R24 GM136986, NIH/NIBIB grant U24EB029012, NIH/NHLBI R21HL172288, and the Nora Eccles Harrison Foundation for Cardiovascular Research.

References

- [1] Chan AK, Dohrmann ML. Management of Premature Ventricular Complexes. *Missouri Medicine* 2010;107(1):39–43. ISSN 0026-6620.
- [2] Sánchez J, Llorente-Lipe I, Espinosa CB, Loewe A, Hernández-Romero I, Vicente-Puig J, Ros S, Atienza F, Carta-Bergaz A, Climent AM, Guillen MS. Enhancing premature ventricular contraction localization through electrocardiographic imaging and cardiac digital twins. *Computers in Biology and Medicine* May 2025;190:109994. ISSN 00104825.
- [3] Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Comput* November 1997;9(8):1735–1780. ISSN 0899-7667.
- [4] Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Information Sciences* September 2017;405:81–90. ISSN 0020-0255.
- [5] Zenger B, Good W, Bergquist J, Burton B, Tate J, Berkenbile L, Sharma V, MacLeod R. Novel experimental model for studying the spatiotemporal electrical signature of acute myocardial ischemia: a translational platform. *J Physiol Meas* Feb 2020;41(1):015002.
- [6] Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. September 2018; URL <https://openreview.net/forum?id=Bkg6RiCqY7>.
- [7] Sun A, Hong W, Li J, Mao J. An Arrhythmia Classification Model Based on a CNN-LSTM-SE Algorithm. *Sensors* January 2024;24(19):6306. ISSN 1424-8220.
- [8] Mohebbanaaz, Sai YP, Kumari LVR. A novel inference system for detecting cardiac arrhythmia using deep learning framework. *Neural Computing and Applications* March 2025;ISSN 1433-3058.
- [9] Bollepalli SC, Sevakula RK, Au-Yeung WM, Kassab MB, Merchant FM, Bazoukis G, Boyer R, Isselbacher EM, Armandas AA. Real-Time Arrhythmia Detection Using Hybrid Convolutional Neural Networks. *Journal of the American Heart Association* December 2021;10(23):e023222. ISSN 2047-9980.

Address for correspondence:

Anna Busatto
University of Utah
72 Central Campus Dr, Salt Lake City, UT 84112
anna.busatto@sci.utah.edu