

Journal Pre-proof

Predicting Ventricular Arrhythmia in Myocardial Ischemia Using Deep Learning

Anna Busatto, PhD, Jake A. Bergquist, PhD, Tolga Tasdizen, PhD, Benjamin A. Steinberg, MD, PhD, FHRS, Ravi Ranjan, MD, PhD, FHRS, Rob S. MacLeod, PhD



PII: S2666-5018(26)00064-4

DOI: <https://doi.org/10.1016/j.hroo.2026.02.009>

Reference: HROO 979

To appear in: *Heart Rhythm O2*

Received Date: 14 November 2025

Revised Date: 3 February 2026

Accepted Date: 4 February 2026

Please cite this article as: Busatto A, Bergquist JA, Tasdizen T, Steinberg BA, Ranjan R, MacLeod RS, Predicting Ventricular Arrhythmia in Myocardial Ischemia Using Deep Learning, *Heart Rhythm O2* (2026), doi: <https://doi.org/10.1016/j.hroo.2026.02.009>.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Heart Rhythm Society. Published by Elsevier Inc.

Predicting Ventricular Arrhythmia in Myocardial Ischemia Using Deep Learning

Anna Busatto^{1,2,3,*} PhD^{0000-0002-3577-460X}, Jake A. Bergquist^{1,2,3} PhD⁰⁰⁰⁰⁻⁰⁰⁰²⁻⁴⁵⁸⁶⁻⁶⁹¹¹, Tolga Tasdizen^{1,5} PhD⁰⁰⁰⁰⁻⁰⁰⁰¹⁻⁶⁵⁷⁴⁻⁰³⁶⁶,

Benjamin A. Steinberg^{6,7,8} MD, PhD, FHRS⁰⁰⁰⁰⁻⁰⁰⁰²⁻⁴⁷²⁹⁻⁷⁸²⁰, Ravi Ranjan^{2,3,6} MD, PhD, FHRS⁰⁰⁰⁰⁻⁰⁰⁰²⁻³³²¹⁻²⁴³⁵,

Rob S. MacLeod^{1,2,3} PhD⁰⁰⁰⁰⁻⁰⁰⁰²⁻⁰⁰⁰⁰⁻⁰³⁵⁶

¹ Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, Utah, United States of America

² Nora Eccles Cardiovascular Research and Training Institute, University of Utah, Salt Lake City, Utah, United States of America

³ Department of Biomedical Engineering, University of Utah, Salt Lake City, Utah, United States of America

⁴ Department of Mathematics, University of Utah, Salt Lake City, Utah, United States of America

⁵ Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, Utah, United States of America

⁶ Division of Cardiology, University of Utah, Salt Lake City, Utah, United States of America

⁷ Department of Medicine, Denver Health Medical Center, Denver, Colorado, United States of America

⁸ Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, United States of America

* Corresponding Author: Anna Busatto
Scientific Computing and Imaging (SCI) Institute, University of Utah
72 S Central Campus Dr, Salt Lake City, UT 84112, USA
Email: anna.busatto@utah.edu

Short Title: Predicting PVC Timing in Myocardial Ischemia

Disclosures Benjamin A. Steinberg reports salary support from the NIH/NHLBI (K23HL143156, R56HL168264, R21HL172288) and AHA/PCORI (18SFRN34110489), and other research support from Abbott, Cardiva, Sanofi, and AltaThera; and consulting to Sanofi, InCarda, Milestone, Pfizer, and AltaThera. Tolga Tasdizen reports salary support from NIH/NHLBI (R21HL172288). Ravi Ranjan is supported by R01HL174056 and R01HL162353, and is a consultant to Abbott and Biosense Webster. The remaining co-authors did not report any relevant disclosures.

1 Abstract

Background: Myocardial ischemia can trigger ventricular arrhythmias with life-threatening consequences. Current monitoring is largely reactive, limiting opportunities for preventive intervention.

Objective: To determine whether high-resolution epicardial electrograms contain predictive signatures that enable forecasting the timing of premature ventricular contractions (PVCs) during acute ischemia, and to quantify subject-specific data requirements for effective personalization.

Methods: We analyzed epicardial sock electrograms (247 electrodes, 1 kHz) from 21 porcine acute ischemia experiments comprising 2,252 spontaneous PVCs. Signals were segmented into overlapping sequences of 3, 5, or 7 consecutive non-PVC beats with a continuous target of time-to-next PVC. A 6-layer Long Short-Term Memory (LSTM) network (hidden size 128) with temporal attention was

trained using mean absolute error (MAE). Performance was evaluated in (A) pooled 80/10/10 cross-validation and (B) leave-one-experiment-out testing with subject-specific fine-tuning using 10% or 15% of held-out data.

Results: In Paradigm A, MAE decreased with longer context (6.50 s for 3 beats, 5.97 s for 5 beats, 4.73 s for 7 beats) with excellent calibration ($R^2 > 0.996$). In Paradigm B, increasing fine-tuning from 10% to 15% reduced mean MAE by 9.6–14.6 s and flattened error growth with prediction horizon, improving the fraction of predictions within 30–60 s windows.

Conclusion: Epicardial electrograms support accurate PVC time-to-event forecasting during acute ischemia, and modest subject-specific adaptation substantially improves generalization, motivating development of real-time predictive monitoring tools.

2 Keywords

Arrhythmia prediction; myocardial ischemia; epicardial electrograms; LSTM; time-to-event modeling

3 Introduction

Myocardial ischemia arises from an imbalance between coronary blood flow, cardiac metabolic demand, and metabolic waste removal and is a major trigger of electrical instability in the heart. This instability can manifest as premature ventricular contractions (PVCs), ectopic beats that sometimes precede far more dangerous arrhythmias such as ventricular tachycardia or fibrillation [1–5]. Since ventricular arrhythmias contribute substantially to sudden cardiac death worldwide, early identification of their precursors has been a long-standing clinical goal [6–8]. Existing monitoring strategies, from surface electrocardiograms to implantable defibrillators, tend to operate *reactively*, delivering alarms or therapy only after the arrhythmia has started. A *predictive* framework capable of anticipating ventricular arrhythmias before they occur would offer a valuable opportunity to intervene proactively.

Creating models of PVC occurrence based on available electrocardiographic signals is a natural, preliminary approach to provide such predictions. Conventional tools such as linear models or spectral analysis capture only limited aspects of the dynamics underlying arrhythmogenesis [9]. This limitation reflects the high-dimensional, non-linear nature of ischemic electrophysiological signals and their dependencies across multiple time scales. This complexity has motivated the growing use of machine learning approaches [10], and in particular deep learning [11–14], which are well suited to extracting patterns from sequential data. Long Short-Term Memory (LSTM) networks [15–18], a type of recurrent neural network designed to capture long-range temporal dependencies, are especially promising for electrophysiological prediction tasks [15–18].

Most prior work has focused on classification problems, such as arrhythmia detection or PVC identification using deep learning models. However, predicting the timing of an arrhythmic event, rather than simply its presence [20–22], has received little attention despite its promise for providing actionable clinical insights. The ability to estimate the timing of an event would provide a window for deploying preventive measures, bridging the gap between passive monitoring and active risk management.

In this study, we focused on two central questions. First, do high-resolution electrograms contain signatures that can be used to forecast PVCs in the setting of acute ischemia? Second, how much subject-specific information is required for a model to achieve useful performance? The answers have practical consequences: if predictive features are absent even in epicardial recordings, then non-invasive ECGs are unlikely to succeed. Conversely, if such features are present and require only modest subject-specific adaptation, then the approach may be feasible in monitored clinical settings where ischemic stress is evolving.

Although the recordings analyzed here are epicardial and not directly obtainable in routine clinical care, they provide a high-fidelity test bed for assessing feasibility. Our clinical motivation is acute ischemic settings in which ventricular ectopy emerges before definitive coronary evaluation or

intervention, where estimating time-to-next PVC could support time-sensitive monitoring without replacing guideline-directed care.

The ultimate goal of this study was to determine whether electrophysiological signals contain sufficient predictive information to forecast PVC timing during acute myocardial ischemia, and to assess the degree of subject-specific data required for reliable personalization.

4 Methods

4.1 Signal Acquisition and Preprocessing

The dataset for this study contained electrophysiological recordings from 21 large-animal experiments, specifically designed to investigate arrhythmias caused by myocardial ischemia, as previously described by Zenger *et al.* [23]. These experiments employed a well-established porcine model of acute myocardial ischemia, in which ischemia was induced by reducing perfusion through the left anterior descending coronary artery (LAD) and then increasing the myocardial demand by means of controlled atrial pacing. All experiments were performed in healthy adult swine without pre-existing myocardial infarction or scar, with the goal of modeling acute ischemia rather than chronic ischemic cardiomyopathy. Baseline ventricular function was physiologically normal prior to ischemia induction; formal stratification by left ventricular ejection fraction was not performed, as we assumed from blood pressure monitoring that animals had normal systolic function. The extent of ischemia was controlled experimentally via graded LAD flow reduction and demand elevation, rather than adjusted for analytically post hoc. Each experiment consisted of between three and five controlled episodes of acute myocardial ischemia, referred to as interventions, each lasting 15 minutes, with a fixed level of occlusion determined specifically for each animal to create predictable and reversible ischemia. Each 15-minute ischemic episode was divided into five 3-minute stages with incremental increases in either pacing rate or the infusion of a pharmacological agent known to increase cardiac stress (dobutamine). A 30-minute rest period followed each ischemic episode to allow the heart to return to baseline. During these experiments, high-fidelity electrocardiographic signals were captured simultaneously from multiple sites, including the epicardial surface, the myocardium via plunge needle arrays, and the body surface. Specifically, electrograms were recorded from a 247-electrode epicardial sock, sampled at 1 kHz [24], and subsequently filtered, baseline-corrected, and fiducialized using the open-source software PFEIFER [25]. For this analysis, the recordings from the epicardial surface were utilized, consistent with the experimental focus on epicardial surface data. All experiments were approved by the Institutional Animal Care and Use Committee of the University of Utah, protocol number 20-11001.

From the resulting comprehensive set of experimental signals, a total of 2,252 premature ventricular contractions (PVCs) were identified across all 21 animals. To account for possible temporal dependencies, the recordings were segmented into contiguous 15-second runs within each intervention, and then further into overlapping sequences of 3, 5, or 7 consecutive non-PVC beats. Rather than treating each heartbeat as an isolated event, this multi-beat design provided the model with a more complete view of the evolving electrophysiological state leading up to a PVC. For each sequence, the continuous target variable, “time-to-PVC”, was defined as the elapsed time from the start of the middle beat of the sequence to the start of the first subsequent 15-second run that contained a PVC occurring within the same coronary artery occlusion or stimulation episode, as shown in Figure 1. This sliding window strategy (*e.g.*, beats 1-2-3, beats 2-3-4, beats 3-4-5) produced the following numbers of samples per configuration: 3 beats = 86 870, 5 beats = 86 012, and 7 beats = 85 191. These samples were used for model training and evaluation. Each input sample was represented as a matrix $\mathbf{X} \in \mathbb{R}^{C \times T}$, where C is the number of epicardial electrodes (channels) and $T = B \times L$ is the total number of time points obtained by concatenating $B \in \{3, 5, 7\}$ consecutive beats along the time axis, with L samples per beat (1 kHz sampling).

4.2 PVC Identification

Premature ventricular contractions (PVCs) occurred spontaneously during the ischemia interventions (LAD perfusion reduction with demand elevation via controlled pacing and/or dobutamine); there was

no need to elicit them using programmed ventricular stimulation. PVC events were identified from the recorded electrocardiographic signals using annotation procedures established within the lab [26]. This study did not aim to localize the spatial site-of-origin of PVCs. While the epicardial sock provides high spatial sampling that could support localization via earliest-activation mapping, our objective here was forecasting the *time-to-next PVC* during acute ischemia. Accordingly, we treated PVCs as time-stamped ventricular events within each intervention and focused our analyses on their temporal occurrence rather than their spatial origin. The spatial characterization of PVC origin is an important direction for future work that we highlight below.

4.3 Model Architecture and Training

We modeled time-to-event prediction as a single-output regression from short sequences of epicardial electrograms. For each configuration (*i.e.*, 3, 5, or 7 consecutive non-PVC beats, as described in the previous subsection), the input sequence was passed to a stacked Long Short-Term Memory (LSTM) network with six recurrent layers (hidden size = 128)[15]. The final hidden representation was mapped to a single scalar via a fully connected layer, yielding the predicted time-to-PVC for the center beat of the input sequence. We used dropout with $p = 0.3$ between the six recurrent backbone layers to improve generalization.

To emphasize the most informative beats within each input sequence, we appended a single-head additive attention layer on top of the LSTM outputs. With $\{\mathbf{h}_t\}_{t=1}^T$ as the hidden states from the last LSTM layer over a T -beat sequence ($T \in \{3, 5, 7\}$), with $\mathbf{h}_t \in \mathbb{R}^D$, the layer had the following linear scoring function:

$$e_t = \mathbf{w}^\top \mathbf{h}_t + b, \mathbf{w} \in \mathbb{R}^D, b \in \mathbb{R}.$$

If a binary validity mask $\{m_t\}_{t=1}^T$ is provided (where $m_t = 1$ indicates a valid time step), we suppressed invalid positions by

$$\tilde{e}_t = \begin{cases} e_t, & m_t = 1, \\ -\infty, & m_t = 0. \end{cases}$$

Attention weights were then computed with a softmax over time,

$$\alpha_t = \frac{\exp(\tilde{e}_t)}{\sum_{k=1}^T \exp(\tilde{e}_k)},$$

so that $\alpha_t = 0$ for all t with $m_t = 0$ and $\sum_{t=1}^T \alpha_t = 1$. We formed a context vector by a weighted sum of the hidden states,

$$\mathbf{c} = \sum_{t=1}^T \alpha_t \mathbf{h}_t \in \mathbb{R}^D.$$

Finally, the context \mathbf{c} was passed through dropout and a linear prediction head to produce the time-to-PVC estimate,

$$\hat{y} = \mathbf{W}_o \text{Dropout}(\mathbf{c}) + \mathbf{b}_o,$$

where $\mathbf{W}_o \in \mathbb{R}^{1 \times D}$ and $\mathbf{b}_o \in \mathbb{R}$ (implemented outside the attention module). An overview of the end-to-end architecture and training protocols for both evaluation paradigms is shown in Figure 2.

We trained all models to minimize mean absolute error (MAE). We selected MAE because it is robust to the heavy-tailed error distribution observed in myocardial ischemia (occasional long inter-event intervals), and because absolute deviations align directly with the intended clinical utility (how many

seconds early or late a prediction is). Models were optimized with AdamW [27] (learning rate 3×10^{-4}) with weight decay for regularization. Unless otherwise specified, we trained for 75 epochs with a batch size of 256. Pilot learning-curve sweeps (50–100 epochs) showed validation of MAE plateauing between ~ 60 and 80 epochs; fixing 75 standardized trainings across conditions while remaining in the stable regime. To limit overfitting, we selected the checkpoint with the lowest validation MAE (minimum mode; top-3 retained) and applied dropout between the recurrent backbone and the output layer ($p = 0.2$). In Paradigm B (described below), fine-tuning used 20 additional epochs at a $10\times$ lower learning rate, further constraining adaptation on the smaller subject-specific set.

Two practical choices improved stability and reproducibility: (i) we normalized inputs per recording to reduce inter-experiment scale differences, and (ii) we used consistent sequence packing across experiments (beats concatenated along time within a sample) so that the backbone learned temporal structure rather than channel ordering artifacts.

For (i), let $x_r(t)$ denote the raw epicardial potential at time t from channel r within a given intervention. For each recording and channel r , we computed the mean μ_r and standard deviation σ_r , using only the training split for that condition (no test/validation leakage), aggregating all time points that contribute to the training windows. We then standardized all splits with

$$x'_r(t) = \frac{x_r(t) - \mu_r}{\max(\sigma_r, 10^{-6})}.$$

In Paradigm B, μ_r and σ_r for the held-out subject were estimated from its fine-tuning subset and applied unchanged to that subject's validation/test windows. Fine-tuning experiments (Paradigm B below) started from the pre-trained weights and continued supervised optimization on the held-out subject for 20 additional epochs; all parameters were unfrozen during adaptation because freezing the backbone degraded performance in our setting.

4.4 Evaluation Metrics and Calibration

We evaluated regression performance using mean absolute error (MAE),

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|,$$

where y_i is the true time to PVC and \hat{y}_i is the predicted time to PVC. We chose this metric because it is robust to the heavy-tailed error distribution (occasional long inter-event intervals) and maps directly to clinical interpretability (seconds early/late).

As a scale-normalized complement to MAE, we report mean relative error (MRE),

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \times 100\%,$$

computed on samples with $y_i > 0$ (cases with $y_i = 0$ were excluded). MRE highlights proportional error relative to the true horizon and is therefore larger when many test samples have short time-to-PVC values.

To quantify the scale and offset bias in predictions, we regress predicted on true time-to-PVC in pooled cross-validation (Paradigm A) separately for each beats setting:

$$\hat{y} = \beta_0 + \beta_1 y.$$

Perfect calibration corresponds to $\beta_1 = 1$ and $\beta_0 = 0$; $\beta_1 < 1$ indicates range compression (long horizons underestimated, short overestimated), while $\beta_1 > 1$ indicates range expansion. We also report R^2 as a goodness-of-fit summary. Figures show the identity line ($\hat{y} = y$); Table 1 reports (β_0, β_1, R^2) .

For clinical interpretability, we report the proportion of test samples with absolute error within $\tau \in \{15, 30, 60, 120\}$ s for each condition:

$$\text{Acc}_{\leq \tau} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(|\hat{y}_i - y_i| \leq \tau),$$

where $\mathbb{1}(\cdot)$ is the indicator function that equals 1 if the condition is true and 0 otherwise, and n is the total number of test samples.

4.5 Error-Horizon Analysis

We analyzed how the error changed with prediction horizon h (the true time-to-PVC). For each condition, we binned the test set by h using $q = 10$ quantile bins (fallback to equal-width bins if quantiles collapse due to ties), computed mean absolute error per bin, and plotted MAE versus bin midpoints. We then fit a simple linear model

$$\text{MAE}(h) = \alpha + \gamma h,$$

and reported the slope γ in units of seconds of error per 100 s of horizon, together with its standard error from the ordinary least-squares fit (Table 2).

4.6 Experimental Paradigms

We evaluated the approach in two complementary settings designed to probe population-level generalization and subject-specific adaptation.

Paradigm A: Intra-Experiment Cross-Validation: For each sequence length (3, 5, or 7 beats), we pooled all samples from the 21 experiments and drew a stratified random split of 80%/10%/10% for training/validation/testing, respectively. This ensured that each experiment contributed to all three partitions while keeping individual samples disjoint across splits. The model was trained on the training set, with its hyperparameters tuned to optimize performance on the validation set. The final test set, which contained samples from all experiments that were unseen by the model during training and validation, was then used to provide an unbiased measure of performance on the overall dataset. This approach simulates a scenario in which a model is trained on a broad range of data and is evaluated on a random, held-out subset of that same data.

Paradigm B: Subject-Specific Fine-Tuning: To simulate adapting a pre-trained model to a new subject, we used a leave-one-experiment-out protocol across all 21 experiments. In each fold, the model was pre-trained on data from 20 experiments (90% training, 10% validation) for 75 epochs. We then fine-tuned on a small, random baseline subset (either 10% or 15%) of the held-out experiment for 20 additional epochs, updating all parameters. The remaining samples from the held-out experiment were reserved strictly for testing. This design quantifies the benefit of limited subject-specific data and matches the intended clinical workflow in which a model is adapted upon initial monitoring rather than trained from scratch.

4.7 Implementation Details

The model was built using PyTorch Lightning, with Hydra configuration managing hyperparameters and training workflows for reproducibility. Training ran on SLURM-managed multi-GPU nodes. Typical allocations per job included 16 CPU cores and 250 GB host memory, with one of the following GPU topologies: $4 \times$ NVIDIA H200 (NVL), $4 \times$ NVIDIA A800 40GB (active), or $8 \times$

NVIDIA L40S. Checkpoint resumption was enabled to tolerate preemption/interruption across nodes. For Paradigm A, the pooled samples from all 21 experiments were processed in one run, with the best model selected based on the lowest validation MAE and evaluated on the test set. For Paradigm B, the leave-one-experiment-out protocol across 21 folds, all layers were unfrozen, the fine-tuning learning rate was reduced by a factor of 0.1, and the folds were parallelized using a SLURM job array (indices 0–20). Performance metrics, including MAE and mean relative error, were logged via Weights & Biases, with checkpoints saved based on validation MAE (minimum mode, top-3 retained). Test predictions and errors were exported to CSV and JSON files in a structured results directory for further analysis.

4.8 Statistical Analysis

We analyzed fold-aggregated Mean Absolute Error (MAE) from Paradigm B (fine-tuning) to compare (i) fine-tuning percentages (10% vs. 15%) within each beat setting (3, 5, or 7 beats) and (ii) beat counts within each fine-tuning percentage. Non-parametric, paired statistical tests were used due to the potential non-Gaussian and heavy-tailed nature of fold-wise MAE distributions and to control for fold-specific variability by treating each fold as its own control. Parametric alternatives, such as paired t -tests or repeated-measures ANOVA, were avoided due to their reliance on normality and variance homogeneity assumptions, which are often violated in MAE data.

Paired fine-tuning comparison (10% vs. 15%).

For each beat setting (3, 5, or 7), we formed paired MAE samples across cross-validation folds to compare models fine-tuned with 10% versus 15% of the training data. The Wilcoxon matched-pairs signed-rank test, a non-parametric method robust to non-normality and outliers, was used to test the null hypothesis of equal median MAE. A negative median difference, $\tilde{\Delta}_{15-10}$ (MAE of 15% minus 10%), indicates superior performance for 15% fine-tuning.

To control the family-wise error rate across the three beat-specific comparisons, we applied the Holm–Bonferroni correction, which scales p -values by the number of remaining tests in ascending order of significance, ensuring monotonicity. We report the adjusted p -values, $\tilde{\Delta}_{15-10}$, and the rank-biserial correlation coefficient:

$$r_{\text{rb}} = \frac{W_+ - W_-}{n(n+1)/2},$$

where W_+ and W_- are the sums of ranks for positive and negative paired differences, respectively, and n is the number of non-zero differences. The coefficient r_{rb} , ranging from -1 to 1 , quantifies the effect size, with $|r_{\text{rb}}| \approx 0.1$, 0.3 , and 0.5 indicating small, medium, and large effects, respectively. A positive r_{rb} suggests that 15% fine-tuning yields lower MAE, complementing statistical significance with a measure of practical importance.

Beats comparison within fine-tuning percentage.

For each fine-tuning percentage (10% or 15%), we compared MAE across beat settings (3, 5, and 7 beats) using the Friedman test, a non-parametric alternative to repeated-measures ANOVA, suitable for matched folds and robust to non-normal distributions. This test assesses whether median MAE differs across beat counts. If significant ($p < 0.05$), we conducted post-hoc Wilcoxon matched-pairs signed-rank tests for pairwise comparisons (3 vs. 5, 3 vs. 7, 5 vs. 7 beats) to pinpoint differences. Each pairwise test was adjusted using the Holm–Bonferroni correction to control the family-wise error rate within this family. We report the adjusted p -values, the median paired difference $\tilde{\Delta}_{b-a}$ (MAE of the higher beat count minus the lower, with negative values indicating better performance for the higher beat count), and the rank-biserial correlation coefficient (r_{rb}), as defined above. A positive r_{rb}

indicates that the higher beat count yields lower MAE, providing a standardized measure of effect magnitude.

Multiple comparisons.

The Holm-Bonferroni correction was applied to control the family-wise error rate within two test families: (i) beats-fixed comparisons (10% vs. 15% fine-tuning within each of 3, 5, and 7 beats) and (ii) percentage-fixed comparisons (3, 5, and 7 beats within each of 10% and 15% fine-tuning). For each family, p-values are ranked from smallest to largest, scaled by the number of remaining tests, and adjusted to ensure monotonicity (non-decreasing adjusted p-values). This approach balances the need to minimize false positives with the preservation of statistical power, ensuring robust inference across multiple hypotheses.

5 Results

5.1 Intra-Experiment Cross-Validation (Paradigm A)

In the pooled cross-validation setting of Paradigm A, in which electrograms from all 21 experiments were combined and randomly split into training, validation, and test sets, the model demonstrated strong overall performance across different input sequence lengths. Table 3 presents the testing MAE and MRE for 3, 5, and 7 beats. A key finding is the progressive reduction in MAE as the number of input beats increased: from 6.50 seconds for 3 beats, to 5.97 seconds for 5 beats, and further to 4.73 seconds for 7 beats. Similarly, the MRE decreased from 15.1% to 7.9%, indicating not only lower absolute errors but also proportionally more accurate predictions relative to the true time-to-PVC values.

Figure 3 illustrates the calibration of predictions through scatter plots of predicted versus true time-to-PVC for each beat configuration test set. The points cluster tightly around the identity line (dashed), with minimal deviation across the full range of horizons (up to approximately 1250 seconds). This visual alignment suggests that the model avoids systematic over- or underestimation, producing predictions that are well-calibrated regardless of the input length. Quantitative confirmation comes from Table 1, which reports the linear regression parameters for each case. Slopes were consistently close to 1 (ranging from 0.992 to 0.997), intercepts were near zero (from -0.50 to 0.88 seconds), and R^2 values exceeded 0.996, underscoring excellent goodness-of-fit and minimal bias. For instance, the 7-beat model achieved the best calibration with a slope of 0.997 and R^2 of 0.9985, implying that nearly all variance in true values was explained by the predictions.

Table 4 shows the proportion of test samples where absolute errors fell within predefined thresholds (15, 30, 60, and 120 seconds). Performance was notably high, with over 91.9% of predictions within 15 seconds for all configurations, rising to 96.1% for 7 beats. At 30 seconds, accuracies ranged from 97.8% to 99.4%, and at 60 seconds, they exceeded 99.5% across the board. These thresholds demonstrate the model's reliability for short-term forecasting. The trend of improvement with more beats was evident here as well, with the 7-beat model achieving near-perfect accuracy (99.9%) within 120 seconds.

Finally, the dependency of error on prediction horizon was examined in Figure 4 (top-left panel) and Table 2. MAE is binned by true time-to-PVC deciles and plotted against bin midpoints on a logarithmic scale. For all beat lengths, MAE remained relatively flat across short to medium horizons (10^1 to 10^2 seconds), with an increase only at the longest bins (around 10^3 seconds). The 7-beat curve lay uniformly below the others, confirming its superior performance. Linear fits yielded shallow slopes of 0.36–0.46 seconds of additional error per 100 seconds of horizon, with small standard errors (0.06–0.09), indicating stable predictive accuracy even for extended intervals. This flat profile suggests that the pooled model handles varying PVC sparsity effectively, with errors dominated by irreducible variability rather than horizon-dependent bias.

5.2 Subject-Specific Fine-Tuning (Paradigm B)

Paradigm B evaluated the model’s ability to adapt to novel subjects via fine-tuning, simulating clinical personalization. Table 5 summarizes fold-level MAE and MRE across the 21 leave-one-out folds, for both 10% and 15% fine-tuning percentages. It also shows the results for 0% fine-tuning as a comparison. Mean MAE values were higher than in Paradigm A (50.42–65.54 s), reflecting the harder setting in which each held-out experiment was only partially seen during fine-tuning (10–15% of its data), and the model had to generalize to the remaining 85–90%. However, increasing fine-tuning data from 10% to 15% consistently lowered MAE: by approximately 9.58 seconds for 3 beats (60.00 to 50.42 s), 9.86 seconds for 5 beats (62.70 to 52.84 s), and 14.61 seconds for 7 beats (65.54 to 50.93 s). Standard deviations (28.06–55.43 s) and ranges indicated variability across folds, likely due to differences in ischemia severity or PVC frequency per experiment. MRE followed a similar pattern, dropping from 140.24% to 130.29% for 3 beats, and more substantially to 106.18% for 7 beats at 15%, emphasizing the benefit of additional subject data for proportional accuracy.

As a qualitative calibration check on a single held-out experiment, Figure 5 shows predicted versus true time-to-PVC for fold 3 of 21 at 15% fine-tuning for 3, 5, and 7 beats. Points concentrate around the identity line, and ordinary least-squares fits (solid lines) are close to the dashed identity, illustrating typical behavior on one fold.

Figure 6 visualizes these distributions as notched boxplots, with medians (black lines), means (orange diamonds), and outliers (dark gray points). The notches (95% CI of medians) show clear separation between 10% and 15% for 5 and 7 beats, while whiskers ($1.5 \times \text{IQR}$) highlight the spread. For example, the 7-beat 15% condition had the lowest median, underscoring robust gains from fine-tuning.

Statistical validation in Table 6 confirmed these differences. Within-beat comparisons (10% vs. 15%) yielded median reductions of -1.23 s (3 beats), -6.72 s (5 beats), and -7.00 s (7 beats). Holm-adjusted Wilcoxon p-values were non-significant for 3 beats (0.2029) but highly significant for 5 and 7 beats (0.0002 and 0.0012), with large negative rank-biserial effects (-0.896 and -0.801), indicating strong practical improvements. Comparisons across beats within percentages (via Friedman and post-hoc Wilcoxon) showed no significant differences, suggesting that fine-tuning benefits are more pronounced than beat-length effects in this setting.

Threshold accuracies in Table 4 provided clinical context for Paradigm B. At 10% fine-tuning, 72.6–74.6% of predictions fell within 60 seconds, improving to 76.7–79.7% at 15%. For 120 seconds, accuracies reached 87.7–93.0%, with the highest (93.0%) for 3 beats at 15%. These gains (*e.g.*, +7.3 percentage points within 60 s for 5 beats) translated to more samples in actionable windows, particularly for longer contexts.

Horizon dependency is detailed in Figure 4 (right and bottom panels) and Table 2. MAE rose more steeply than in Paradigm A (slopes 2.54–8.67 s per 100 s), with wider shaded errors at long horizons due to fewer samples. However, 15% fine-tuning flattened these curves across all beats (*e.g.*, 8.67 to 6.02 s/100 s for 7 beats), reducing standard errors and attenuating the increase at extended bins. This pattern reveals that additional subject data primarily enhanced long-horizon predictions, where baseline variability was highest.

Table 1: Calibration for pooled 80–10–10 models (Paradigm A). Linear regression of predicted on true time-to-PVC; ideal calibration is Slope = 1, Intercept = 0.

Beats	Slope	Intercept (s)	R^2
3	0.992	0.88	0.9963
5	0.995	-0.50	0.9981
7	0.997	-0.27	0.9985

Table 2: Slope of MAE vs. prediction horizon (bin midpoints). Slopes are reported as seconds of error per 100 s of horizon; SE from OLS fit.

Paradigm	Beats	FT (%)	Slope (s/100 s)	SE	Bins
A	3	—	0.44	0.07	10
A	5	—	0.46	0.09	10
A	7	—	0.36	0.06	10
B	3	10	5.48	1.30	10
B	3	15	2.94	1.10	10
B	5	10	3.43	0.92	10
B	5	15	2.54	0.87	10
B	7	10	8.67	2.11	10
B	7	15	6.02	1.35	10

Table 3: Paradigm A: Mean absolute error (MAE) and mean relative error (MRE) for testing set.

Beats	MAE (s)	MRE (%)
3	6.50	15.11
5	5.97	11.01
7	4.73	7.93

Table 4: Proportion of test samples within absolute error thresholds by condition. Paradigm A used pooled 80–10–10 CV; Paradigm B used leave-one-experiment-out with subject-specific fine-tuning (FT).

Paradigm	Beats	FT (%)	≤ 15 s (%)	≤ 30 s (%)	≤ 60 s (%)	≤ 120 s (%)
A	3	—	91.9	97.8	99.5	99.9
A	5	—	93.3	98.6	99.7	99.9
A	7	—	96.1	99.4	99.8	99.9
B	3	10	26.9	49.7	73.9	89.8
B	3	15	28.5	51.3	76.7	93.0
B	5	10	30.8	52.4	74.6	89.4
B	5	15	36.4	59.7	79.7	92.5
B	7	10	28.5	50.1	72.6	87.7
B	7	15	35.1	57.4	79.2	91.7

Table 5: Paradigm B: fold-level MAE by beats and fine-tuning percentage. Values are mean ± SD across folds; range shown in brackets. **Abbreviation:** FT = fine-tuning (percentage of subject data used for adaptation).

Beats	FT (%)	MAE (s)	Range (s)	MRE (%)	Range (%)
3	0	204.96 ± 67.89	[71.94, 317.38]	366.00 ± 213.44	[91.54, 875.88]
3	10	60.00 ± 45.27	[39.39, 80.61]	140.24 ± 82.46	[30.17, 363.44]
3	15	50.42 ± 28.06	[37.64, 63.19]	130.29 ± 56.60	[55.00, 245.28]
5	0	211.48 ± 85.25	[58.49, 400.16]	423.44 ± 297.45	[877.87, 1250.89]

5	10	62.70 ± 37.58	[45.59, 79.81]	144.67 ± 75.03	[44.66, 314.31]
5	15	52.84 ± 33.56	[37.56, 68.12]	117.62 ± 64.82	[35.40, 293.53]
7	0	203.37 ± 82.04	[84.07, 335.51]	342.54 ± 208.95	[102.78, 952.60]
7	10	65.54 ± 55.43	[40.31, 90.78]	138.39 ± 81.40	[39.06, 347.86]
7	15	50.93 ± 39.36	[33.01, 68.84]	106.18 ± 56.80	[26.55, 226.17]

Table 6: Statistical comparisons with Holm-adjusted p -values. Multigroup tests use Friedman; pairwise tests use Wilcoxon matched-pairs signed-rank. Median Δ in seconds (15% – 10%). r_{rb} is rank-biserial effect size. **Abbreviations:** mg = multigroup, ph = post hoc.

Factor	Contrast	Method	Pairs	p_{Holm}	Med. Δ	r_{rb}
Beats within % (10%)	3, 5, 7 beats (mg)	Friedman	–	–	–	–
Beats within % (15%)	3, 5, 7 beats (mg)	Friedman	–	–	–	–
Beats within % (10%)	3 vs 5 beats (ph)	Wilcoxon	21	1.0000	-2.98	–
Beats within % (10%)	3 vs 7 beats (ph)	Wilcoxon	21	1.0000	3.96	–
Beats within % (10%)	5 vs 7 beats (ph)	Wilcoxon	21	1.0000	1.79	–
Beats within % (15%)	3 vs 5 beats (ph)	Wilcoxon	21	1.0000	-2.43	–
Beats within % (15%)	3 vs 7 beats (ph)	Wilcoxon	21	0.8166	-5.05	–
Beats within % (15%)	5 vs 7 beats (ph)	Wilcoxon	21	1.0000	-0.70	–
% within Beats (3 beats)	10% vs 15%	Wilcoxon	21	0.2029	-1.23	-0.325
% within Beats (5 beats)	10% vs 15%	Wilcoxon	21	0.0002*	-6.72	-0.896
% within Beats (7 beats)	10% vs 15%	Wilcoxon	21	0.0012*	-7.00	-0.801

* $p < 0.05$ indicates significance; p -values are Holm-adjusted.

6 Discussion

6.1 The Predictive Value of Electrophysiological Data

Our results affirmatively address the first central question posed in the introduction: do high-resolution epicardial electrograms contain predictive signatures that can be leveraged to forecast the timing of PVCs in the setting of myocardial ischemia? This claim is supported by the strong performance observed in both experimental paradigms. In Paradigm A, the LSTM model achieved low MAE ranging from 4.73 to 6.50 seconds, with near-perfect calibration ($R^2 > 0.996$) and high accuracies within clinically relevant thresholds (*e.g.*, over 99.5% within 60 seconds). These metrics indicate that the model successfully extracted time-dependent patterns from the electrograms, patterns that traditional linear or spectral methods often fail to capture [11, 12]. The progressive improvement with longer input sequences, a 27.3% improvement from 3 to 7 beats, highlights that incorporating more temporal context allowed the model to better capture the electrophysiological dynamics leading to PVCs, aligning with the complex dynamics of ischemic arrhythmogenesis described earlier [1, 2].

In Paradigm B, the model’s adaptability to novel subjects reinforced its value, with 15% fine-tuning reducing MAE by up to 14.61 seconds and flattening the error-horizon slopes. This result demonstrated that even limited subject-specific data can refine predictions, addressing inter-subject variability in ischemic responses. Collectively, these findings validate epicardial electrograms as a rich substrate for arrhythmia forecasting, echoing prior studies on the electrophysiological insights provided by such signals [5, 8]. By establishing feasibility in this high-fidelity setting, our work lays

the groundwork for translating these capabilities to non-invasive modalities, directly responding to the clinical need for proactive rather than reactive monitoring.

6.2 Interpretation of Model Performance and Personalization

The results generally underscore the efficacy of LSTM networks in handling the sequential nature of electrophysiological signals, a core challenge in arrhythmia prediction. The attention mechanism's role in weighting informative beats likely contributes to the observed gains with longer sequences, allowing the model to focus on subtle precursors to PVCs amid ischemic instability [15, 16]. In Paradigm A, the flat error-horizon profiles (slopes of 0.36–0.46 s per 100 s) indicate that the model maintains consistent accuracy across varying prediction intervals, a critical attribute for real-world applications where PVC sparsity can fluctuate.

Paradigm B's fine-tuning results make a strong claim for personalization: minimal adaptation (10–15% of data) yielded significant improvements (*e.g.*, large effect sizes of -0.801 to -0.896), particularly for longer horizons where inter-subject differences were amplified. This finding ties back to our second question, *how much subject-specific information is required for a model to achieve useful performance?*, quantifying that, in our case, as little as 15% subject-specific data sufficed for useful performance, simulating clinical scenarios like initial monitoring periods. Such performance positions our approach as a bridge between population-level models and individualized care, consistent with emerging trends in personalized cardiac AI [13, 14].

To complement aggregate metrics, we examined test segments with similar true horizons but divergent errors (Figure 7). Visually, these examples were not meaningfully distinct, suggesting that the model relies on subtle, distributed cues that are difficult for human experts to discern by eye. This observation supports the use of sequence models and motivates targeted interpretability analyses (*e.g.*, attention weights or saliency over beats) to clarify which signal characteristics drive accurate versus inaccurate predictions.

6.3 Challenges and Limitations

Despite these advances, several challenges underscore our focus on fundamental feasibility and highlight areas for refinement. We focused on forecasting the temporal occurrence of PVCs rather than attempting to localize their spatial site of origin, a straightforward step given the resolution of our epicardial mapping measurements. Instead, our goal was to predict the future occurrence of PVCs based on recordings of normal epicardial potentials. Future studies will evaluate whether annotating the spatial origin of PVCs improves forecasting performance or enables mechanistic stratification. By the very nature of the preparation (a healthy heart in a swine), ventricular tachycardia was observed only infrequently following PVCs in these experiments. Such sparse data precluded a systematic analysis of PVC–VT transitions, which leads naturally to additional future studies.

The sparsity of training data is a recurring theme in the medical applications of machine learning, and our results showed related limitations. The steeper error-horizon slopes in Paradigm B (2.54–8.67 s per 100 s) reflect the inherent stochasticity of PVCs in ischemia, where long intervals introduce greater uncertainty, a limitation inherent to time-to-event modeling in sparse arrhythmic data [6]. Inter-fold variability (*e.g.*, MAE SD up to 55.43 s) underscores subject heterogeneity, potentially arising from experimental variations in ischemia protocols or animal physiology, which may not fully mirror human pathophysiology.

While epicardial electrograms provide a high signal-to-noise benchmark, their invasive nature restricts direct clinical use; translation to surface ECGs might degrade performance due to signal attenuation and noise, as noted in related electrophysiological studies [7]. The reliance on MAE, while robust to heavy-tailed errors, does not quantify prediction uncertainty, potentially leading to overconfidence in high-risk scenarios. The dataset, though unique (21 experiments, 2,252 PVCs), is modest in scale and derived from animal models, limiting generalizability to diverse human populations or varying ischemic etiologies. Retrospective evaluation assumes static conditions, but real-time drifts (*e.g.*, evolving reperfusion) could exacerbate errors.

There are some additional challenges that often arise in the setting of approaches requiring training based on limited data. Since samples are constructed as overlapping beat windows, distinct train/validation/test partitions can still share context even when the center beat used for supervision is unique to a split. In Paradigm A, windows near split boundaries may reuse neighboring beats across partitions; in Paradigm B, fine-tune and test windows from the same experiment can likewise share non-center beats. We explicitly excluded duplicate center beats across splits, but residual context overlap might have introduced a slight optimistic bias. Two practical mitigations would be “exclusion margins” that also hold out $\pm \lfloor (T-1)/2 \rfloor$ beats around each labeled center (for $T \in \{3, 5, 7\}$), and block-wise splitting by contiguous time (*e.g.*, intervention segments) to ensure non-overlapping context between splits. In our study, the limited fine-tune fraction in Paradigm B and the center-beat uniqueness constraint reduced, though did not eliminate, this effect. Finally, the qualitative examples underscore that the forecasting cues are often not obvious to human inspection, reinforcing the need for interpretable analyses to expose which features the model exploits and to guide future preprocessing or architectural choices.

6.4 Future Directions

From a translational perspective, although this study was conducted in a large-animal model, the experimental conditions were designed to replicate key physiological features of human acute myocardial ischemia, including controlled coronary flow reduction, increased myocardial demand, and the emergence of spontaneous ventricular ectopy. Porcine cardiac anatomy, electrophysiology, and ischemic responses share key similarities with those of humans, making this a well-established translational model for studies of ischemia and ventricular arrhythmias. Importantly, the goal of this work was not to produce a clinically deployable predictor or to alter current standards of care, but to establish whether electrophysiological signals contain sufficient information to forecast PVC timing under ischemic stress. Demonstrating feasibility in high-fidelity epicardial recordings provides a necessary foundation for future translational studies.

From a clinical perspective, the findings of this study are most relevant to settings involving acute ischemic stress, rather than chronic ischemic cardiomyopathy with established scar. This framework is not intended to replace standard coronary evaluation, revascularization, or guideline-directed therapy, but to motivate future investigation into whether arrhythmic vulnerability during evolving ischemia can be identified earlier through electrophysiological monitoring. Potential downstream applications, pending substantial validation, may include acute coronary syndromes prior to intervention, peri-procedural ischemia, or monitored settings such as stress testing, the emergency department, or intensive care unit.

A plausible future use case is a patient undergoing continuous ECG monitoring during suspected or evolving ischemia (*e.g.*, chest pain in the emergency department while awaiting coronary evaluation, high-risk stress testing, or peri-procedural ischemia). In this scenario, an ECG-based version of the present approach could run continuously and output a rolling estimate of *time-to-next PVC* (and ideally an uncertainty measure) that updates as new beats arrive. The intended value would be lead time: a sustained shortening of predicted horizons, or repeated predictions within a short-window threshold, could serve as a decision-support signal of heightened electrical instability and prompt clinicians to increase vigilance or readiness to escalate care (*e.g.*, intensified rhythm surveillance, preparation for defibrillation/pacing, reassessment of reversible triggers such as electrolytes and ischemic burden, and prioritization of urgent evaluation/intervention when clinically indicated). Conversely, persistently long and stable predicted horizons could support continued observation without additional alarms. Crucially, any alerting logic would require prospective validation to define safe operating points (thresholds, false-alarm burden, and patient selection), and would be integrated with, not substituted for, standard clinical assessment and guideline-directed management. Application to chronic ischemia or scar-mediated ventricular tachycardia was outside the scope of the present study and would require dedicated investigation.

7 Conclusions

This study shows that deep learning, specifically LSTM models, can forecast the timing of PVCs during myocardial ischemia from short sequences of epicardial electrograms. The model's accuracy increased as the input sequence length grew from 3 to 7 beats, while its calibration remained highly accurate, suggesting that the remaining prediction errors are primarily due to natural variations in physiological responses rather than flaws in the model that could be corrected. We further demonstrated that a pre-trained model can be efficiently adapted to a new subject with a small amount of subject-specific data (15%), yielding substantial, statistically reliable error reductions, especially with 5–7 beats, and increasing the fraction of predictions within clinically meaningful 30–60 s windows. These findings establish the feasibility of predictive modeling from high-resolution electrophysiological signals, paving the way for real-time, personalized tools that shift arrhythmia management toward proactive prevention and improved patient outcomes in ischemic heart disease [?].

8 Acknowledgments

Computational support for this research came from the Center for Integrative Biomedical Computing (www.sci.utah.edu/cibc). We thank the Scientific Computing and Imaging Institute for providing computational resources for this project. NIH grants R01HL174056-01 and 1R03NS145583, VA grant 5I01CX002758-02, NIH/NHLBI R21HL172288, and the Nora Eccles Harrison Foundation for Cardiovascular Research.

9 Declarations

9.1 Data Availability Statement

The data that support the findings of this study are not publicly available, but are available from the corresponding author upon reasonable request.

9.2 Author Contributions

A.B., J.A.B., T.T., B.A.S., R.R., and R.S.M. contributed to the conception and design of the study. A.B., J.A.B., T.T., B.A.S., and R.S.M. were involved in the acquisition, analysis, or interpretation of data. All authors participated in drafting the manuscript or revising it critically for important intellectual content, approved the final version, and agree to be accountable for all aspects of the work to ensure the accuracy and integrity of any part of the research is appropriately investigated and resolved. All individuals who meet authorship criteria are listed.

9.3 Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author(s) utilized ChatGPT-5 to enhance readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

References

- [1] A. K. Chan and M. L. Dohrmann, "Management of Premature Ventricular Complexes," *Missouri Medicine*, vol. 107, no. 1, pp. 39–43, 2010.
- [2] E. P. Gerstenfeld and T. De Marco, "Premature Ventricular Contractions," *Circulation*, vol. 140, pp. 624–626, Aug. 2019.
- [3] R. Scorza, M. Jonsson, L. Friberg, M. Rosenqvist, and V. Frykman, "Prognostic implication of premature ventricular contractions in patients without structural heart disease," *EP Europace*, vol. 25, pp. 517–525, Feb. 2023.

- [4] L. Calo', M. Tatangelo, G. Panattoni, C. Crescenzi, M. Squeglia, F. Fanisio, F. Romeo, F. Toto, E. de Ruvo, and M. Rebecchi, "Unlocking the enigma: decoding premature ventricular complexes for effective clinical assessment and risk management," *European Heart Journal Supplements*, vol. 26, pp. i23–i28, Apr. 2024.
- [5] J. Sánchez, I. Llorente-Lipe, C. B. Espinosa, A. Loewe, I. Hernández-Romero, J. Vicente-Puig, S. Ros, F. Atienza, A. Carta-Bergaz, A. M. Climent, and M. S. Guillem, "Enhancing premature ventricular contraction localization through electrocardiographic imaging and cardiac digital twins," *Comput. Biol. Med.*, vol. 190, p. 109994, May 2025.
- [6] S. S. Chugh, "Early identification of risk factors for sudden cardiac death," *Nature Reviews. Cardiology*, vol. 7, pp. 318–326, June 2010.
- [7] R. M. John, U. B. Tedrow, B. A. Koplán, C. M. Albert, L. M. Epstein, M. O. Sweeney, A. L. Miller, G. F. Michaud, and W. G. Stevenson, "Ventricular arrhythmias and sudden cardiac death," *The Lancet*, vol. 380, pp. 1520–1529, Oct. 2012.
- [8] P. T. Tang, M. Shenasa, and N. G. Boyle, "Ventricular Arrhythmias and Sudden Cardiac Death," *Cardiac Electrophysiology Clinics*, vol. 9, pp. 693–708, Dec. 2017.
- [9] N. A. Trayanova, D. M. Popescu, and J. K. Shade, "Machine learning in arrhythmia and electrophysiology," *Circulation Research*, vol. 128, no. 4, pp. 544–566, 2021.
- [10] J. A. Bergquist, B. Zenger, J. Brundage, R. S. MacLeod, T. J. Bunch, R. Shah, X. Ye, A. Lyons, M. Torre, R. Ranjan, T. Tasdizen, and B. A. Steinberg, "Performance of off-the-shelf machine learning architectures and biases in low left ventricular ejection fraction detection," *Heart Rhythm O2*, vol. 5, no. 9, pp. 644–654, 2024.
- [11] Z. I. Attia, P. A. Noseworthy, F. Lopez-Jimenez, S. J. Asirvatham, A. J. Deshmukh, B. J. Gersh, R. E. Carter, X. Yao, A. A. Rabinstein, B. J. Erickson, S. Kapa, and P. A. Friedman, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *The Lancet*, vol. 394, pp. 861–867, Sept. 2019.
- [12] P. A. Noseworthy, Z. I. Attia, L. C. Brewer, S. N. Hayes, X. Yao, S. Kapa, P. A. Friedman, and F. Lopez-Jimenez, "Assessing and Mitigating Bias in Medical Artificial Intelligence," *Circulation: Arrhythmia and Electrophysiology*, vol. 13, p. e007988, Mar. 2020. Publisher: American Heart Association.
- [13] R. Kabra, S. Israni, B. Vijay, C. Baru, R. Mendu, M. Fellman, A. Sridhar, P. Mason, J. W. Cheung, L. DiBiase, S. Mahapatra, J. Kalifa, S. A. Lubitz, P. A. Noseworthy, R. Navara, D. D. McManus, M. Cohen, M. K. Chung, N. Trayanova, R. Gopinathannair, and D. Lakkireddy,

“Emerging role of artificial intelligence in cardiac electrophysiology,” *Cardiovascular Digital Health Journal*, vol. 3, pp. 263–275, Dec. 2022.

[14] S. Khurshid, S. Friedman, C. Reeder, P. Di Achille, N. Diamant, P. Singh, L. X. Harrington, X. Wang, M. A. Al-Alusi, G. Sarma, A. S. Foulkes, P. T. Ellinor, C. D. Anderson, J. E. Ho, A. A. Philippakis, P. Batra, and S. A. Lubitz, “ECG-Based Deep Learning and Clinical Risk Factors to Predict Atrial Fibrillation,” *Circulation*, vol. 145, pp. 122–133, Jan. 2022. Publisher: American Heart Association.

[15] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, pp. 1735–1780, Nov. 1997.

[16] Ö. Yildirim, “A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification,” *Computers in Biology and Medicine*, vol. 96, pp. 189–202, May 2018.

[17] L. Wang and X. Zhou, “Detection of Congestive Heart Failure Based on LSTM-Based Deep Network via Short-Term RR Intervals,” *Sensors (Basel, Switzerland)*, vol. 19, p. 1502, Mar. 2019.

[18] H. Bilal, Y. Tian, A. Ali, Y. Muhammad, A. Yahya, B. A. Izneid, and I. Ullah, “An Intelligent Approach for Early and Accurate Predication of Cardiac Disease Using Hybrid Artificial Intelligence Techniques,” *Bioengineering*, vol. 11, p. 1290, Dec. 2024. Publisher: Multidisciplinary Digital Publishing Institute.

[19] T. K. Revathi, S. Balasubramaniam, V. Sureshkumar, and S. Dhanasekaran, “An Improved Long Short-Term Memory Algorithm for Cardiovascular Disease Prediction,” *Diagnostics*, vol. 14, p. 239, Jan. 2024.

[20] 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,” *Scientific Reports*, vol. 6, p. 32390, Aug. 2016.

[21] M. Z. H. Kolk, S. Ruipérez-Campillo, C. P. Allaart, A. A. M. Wilde, R. E. Knops, S. M. Narayan, F. V. Y. Tjong, and DEEP RISK investigators, “Multimodal explainable artificial intelligence identifies patients with non-ischaeic cardiomyopathy at risk of lethal ventricular arrhythmias,” *Scientific Reports*, vol. 14, p. 14889, June 2024.

[22] L. Fiorina, T. Carbonati, K. Narayanan, J. Li, C. Henry, J. P. Singh, and E. Marijon, “Near-term prediction of sustained ventricular arrhythmias applying artificial intelligence to single-lead ambulatory electrocardiogram,” *European Heart Journal*, vol. 46, pp. 1998–2008, June 2025.

[23] B. Zenger, W. Good, J. Bergquist, B. Burton, J. Tate, L. Berkenbile, V. Sharma, and R. MacLeod, “Novel experimental model for studying the spatiotemporal electrical signature of acute myocardial ischemia: a translational platform,” *J. Physiol. Meas.*, vol. 41, p. 015002, Feb 2020.

- [24] B. Zenger, J. A. Bergquist, W. W. Good, B. Steadman, and R. S. MacLeod, “High-capacity cardiac signal acquisition system for flexible, simultaneous, multidomain acquisition,” in *2020 Computing in Cardiology*, pp. 1–4, 2020.
- [25] A. Rodenhauser, W. W. Good, B. Zenger, J. Tate, K. Aras, B. Burton, and R. S. MacLeod, “PFEIFER: Preprocessing framework for electrograms intermittently fiducialized from experimental recordings,” *The Journal of Open Source Software*, vol. 3, p. 472, 2018.
- [26] A. Rodenhauser, W. Good, B. Zenger, J. Tate, K. Aras, B. Burton, and R. MacLeod, “PFEIFER: Preprocessing framework for electrograms intermittently fiducialized from experimental recordings,” *J. Open Source Software*, vol. 3, p. 472, Sept. 2018.
- [27] I. Loshchilov and F. Hutter, “Decoupled Weight Decay Regularization,” *CoRR*, vol. abs/1711.05101, Jan. 2019.

Figure 1: Schematic representation of the temporal relationship between premature ventricular contractions (PVCs) and intervention periods with induced myocardial ischemia. The figure illustrates two intervention episodes (Intervention 1 and Intervention 2), each consisting of sequences of beats leading up to PVC events. PVCs are marked by vertical red lines (PVC 1, PVC 2, PVC 3 in Intervention 1; PVC 1, PVC 2 in Intervention 2). Colored segments represent beats preceding PVCs within each intervention period, with arrows indicating progression toward the next PVC. A recovery period separates the two interventions. Crossed-out segments denote periods excluded from analysis.

Figure 2: LSTM-based framework for time-to-PVC prediction from epicardial electrograms. (A) Model architecture: sequences of 3, 5, or 7 consecutive non-PVC epicardial beats are processed through a 6-layer stacked LSTM network with an attention mechanism for temporal weighting. The final representation passed through dropout regularization and a fully connected layer to produce a continuous time-to-PVC prediction in seconds. (B) Paradigm A (Intra-Experiment Cross-Validation): All samples from 21 experiments ($n=2,252$ PVCs; 85,000–88,000 total samples) were pooled and stratified randomly into training (80%), validation (10%), and test (10%) sets, ensuring all experiments contribute to each partition. (C) Paradigm B (Subject-Specific Fine-Tuning): Leave-one-experiment-out validation protocol, where the model was pre-trained on 20 experiments for 75 epochs, then fine-tuned on 10–15% of the held-out experiment for 20 additional epochs, with the remaining 85–90% reserved for testing. This approach simulates clinical adaptation to individual patients with limited subject-specific data.

Figure 3: Pooled 80–10–10 predicted vs. true time-to-PVC for 3, 5, and 7 beats. The dashed line is the identity ($\hat{y} = y$); proximity indicates calibration.

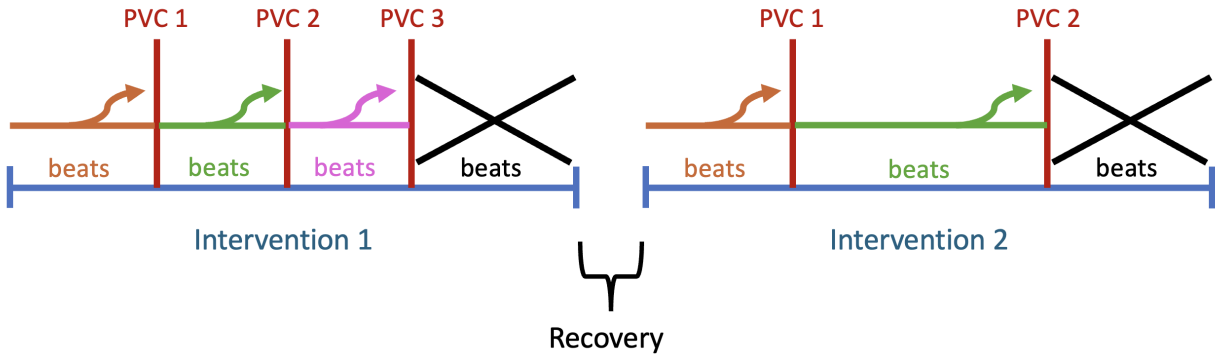
Figure 4: Prediction error as a function of horizon. Curves show mean absolute error (MAE, s) computed in decile bins of the true time-to-PVC; the x -axis is logarithmic and points are plotted at bin midpoints. **Top-left:** Paradigm A for input contexts of 3, 5, and 7 beats. **Top-right/bottom panels:** Paradigm B for 3- (top-right), 5- (bottom-left), and 7-beat (bottom-right) models, comparing fine-tuning with 10% vs. 15% of subject data. Shaded ribbons denote ± 1 standard error within each bin. Fewer samples at the longest horizons yield wider bands.

Figure 5: Paradigm B, fine-tuned with 15% of the held-out experiment. Predicted vs. true time-to-PVC for 3-, 5-, and 7-beat inputs shown for a single illustrative fold (fold 3 of 21). The dashed line marks the identity ($\hat{y} = y$); the solid line is an ordinary least squares fit with the equation inset. These panels are examples; variability across folds is observed.

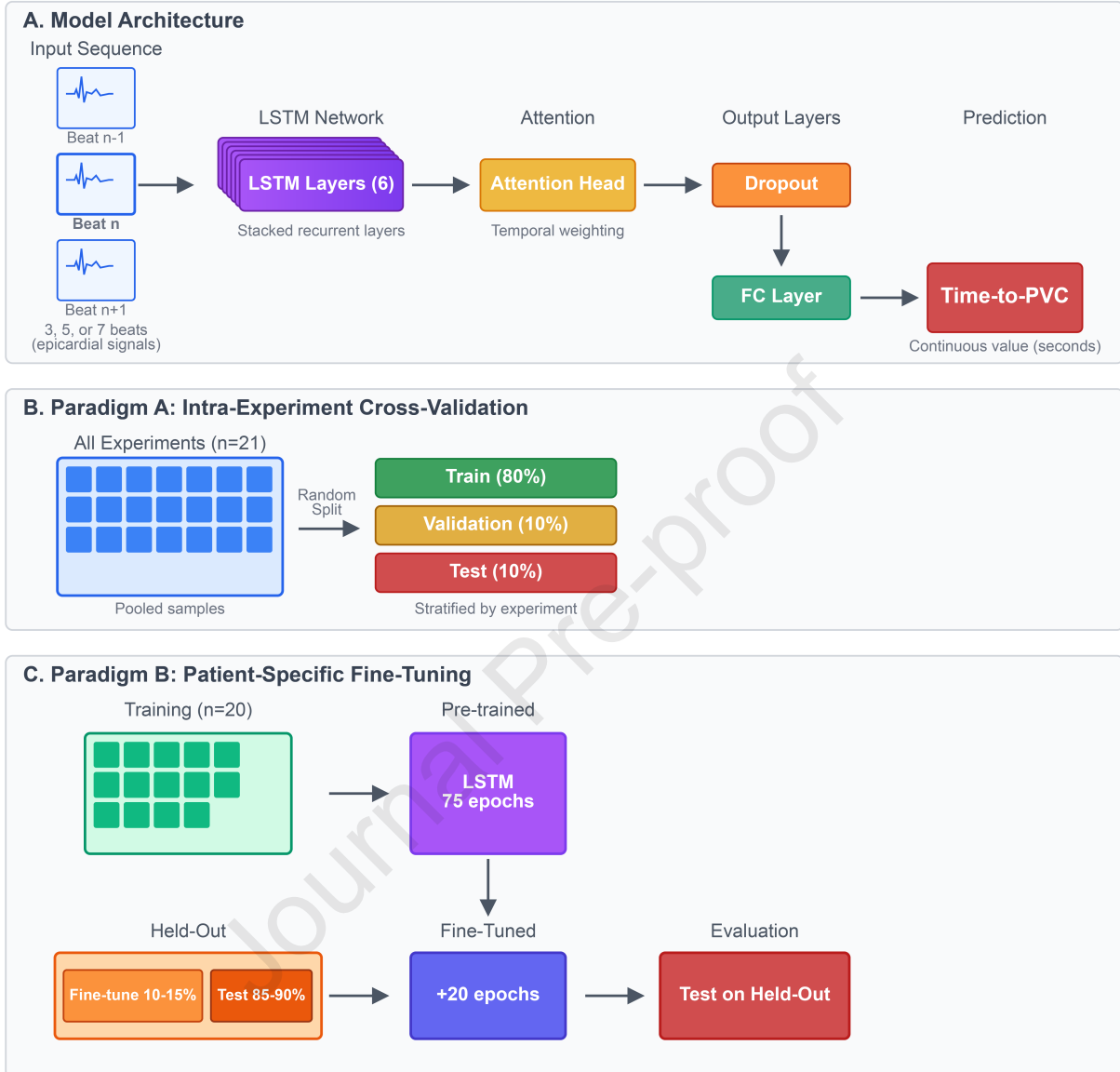
Figure 6: Distribution of test MAE across the 21 leave-one-experiment-out folds in Paradigm B, stratified by input context (3/5/7 beats) and fine-tuning percentage (10% or 15%). For each condition,

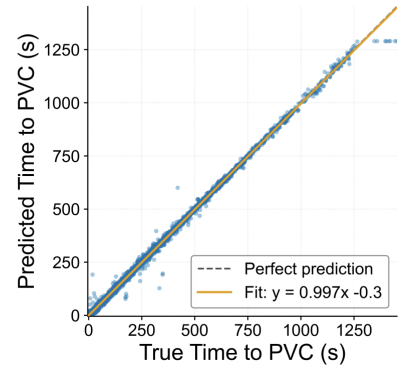
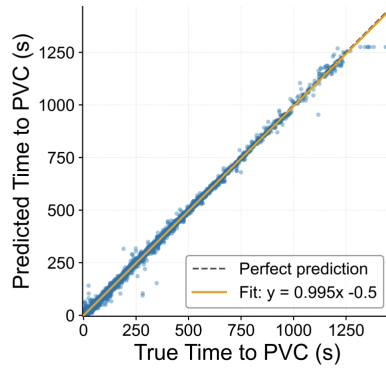
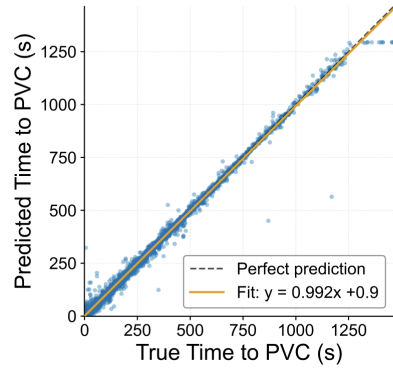
a single box summarizes the 21-fold-level MAEs, where each fold's MAE was computed on that fold's test set after fine-tuning on the specified fraction of the held-out experiment. Notched boxes indicate an approximate 95% confidence interval for the median (black line); the box spans the interquartile range (IQR); whiskers follow the Tukey rule ($1.5 \times \text{IQR}$); orange diamonds denote the mean; gray points are outliers. Fine-tuning percentages refer to the proportion of data from the held-out experiment used for adaptation; the remainder was used exclusively for testing.

Figure 7: Qualitative examples from the test set (Paradigm B, 15% fine-tuning, fold 4). Each panel shows a short epicardial electrogram segment (Potential, mV, versus Time, ms). **Top**: large-error case with *true* time-to-PVC = 57.541 s, *predicted* = 151.262 s (absolute error = 93.721 s; relative error = 162.9%). **Bottom**: small-error case with *true* = 56.291 s, *predicted* = 73.163 s (absolute error = 16.872 s; relative error = 30.0%). Despite nearly identical true horizons (≈ 56 –58 s), the traces do not exhibit obvious visual differences that would explain the disparity in error, underscoring the subtlety of the predictive cues.

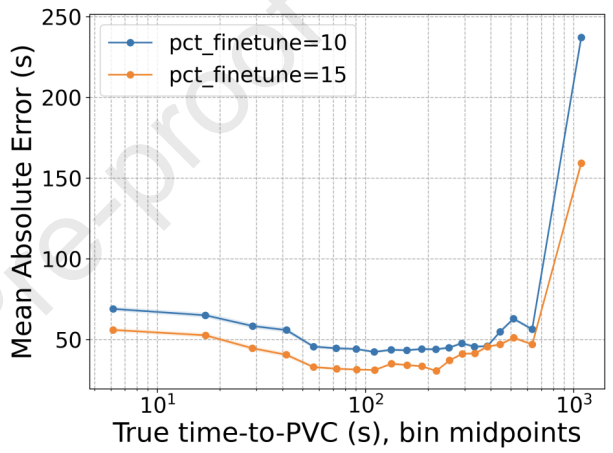
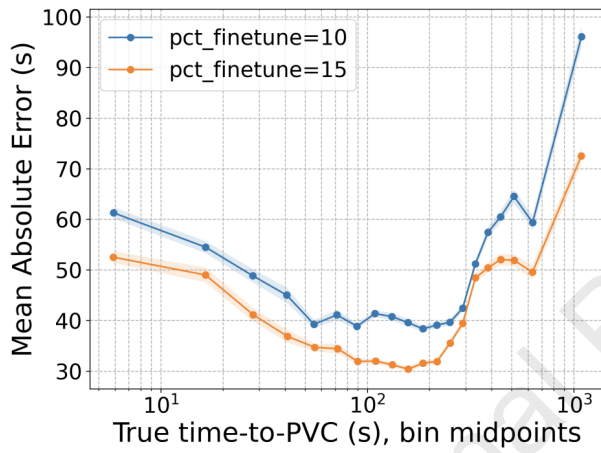
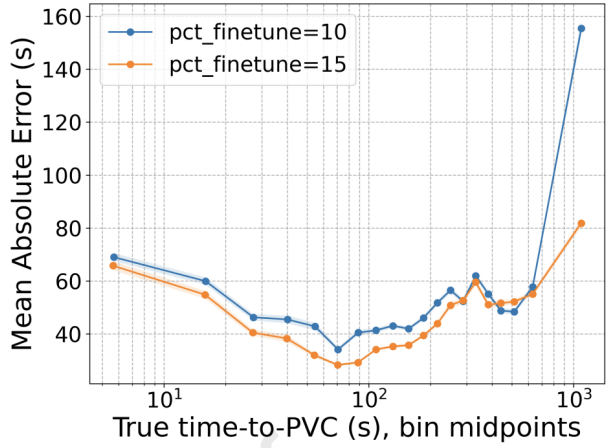
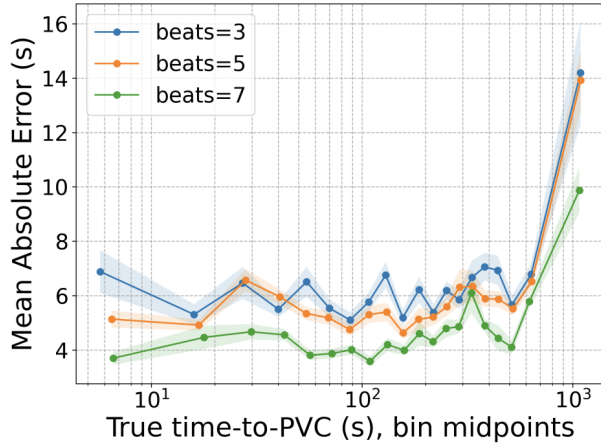


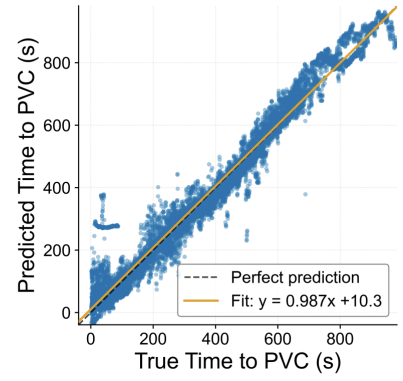
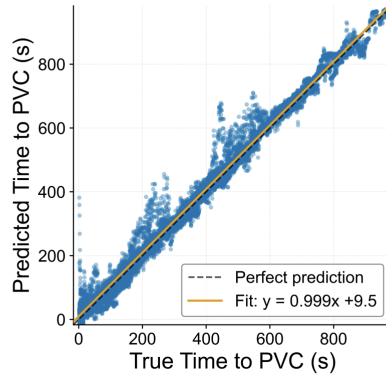
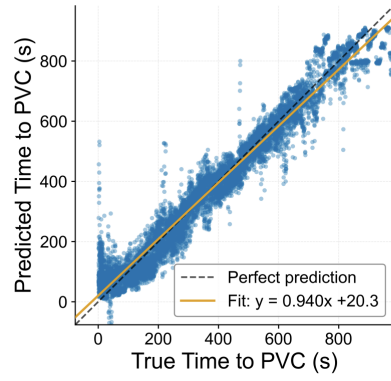
Journal Pre-proof



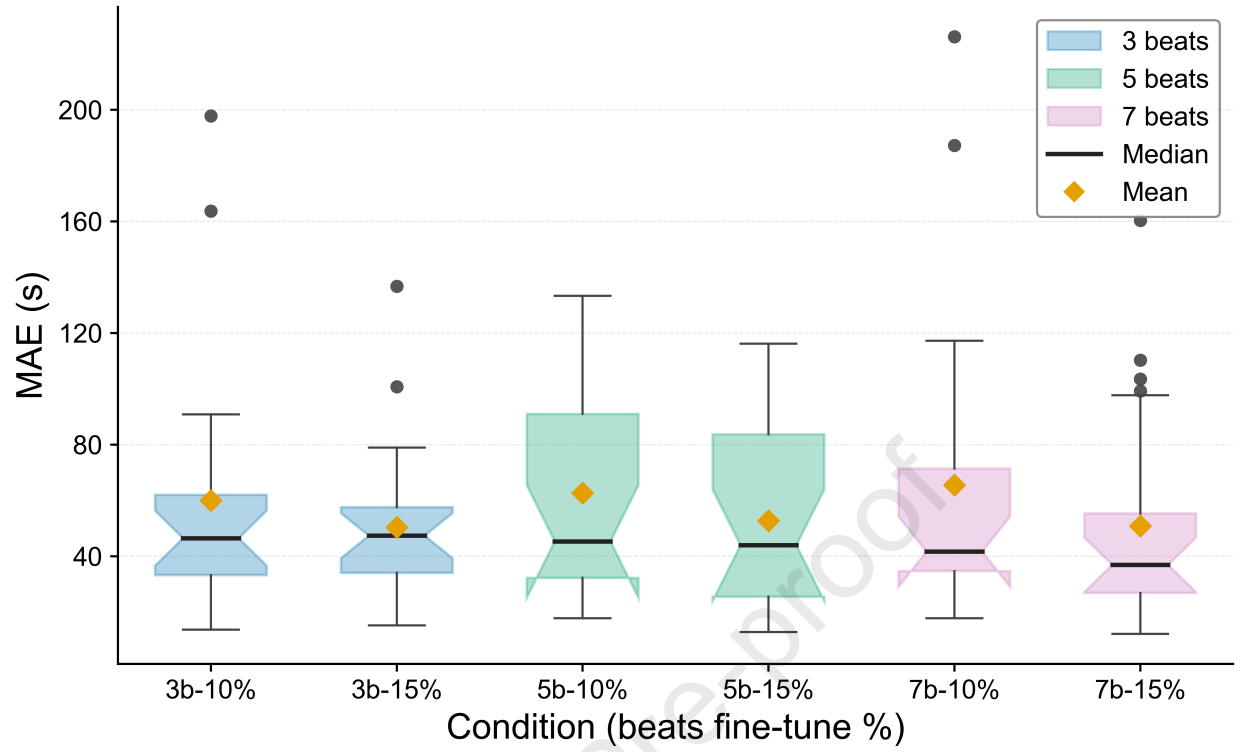


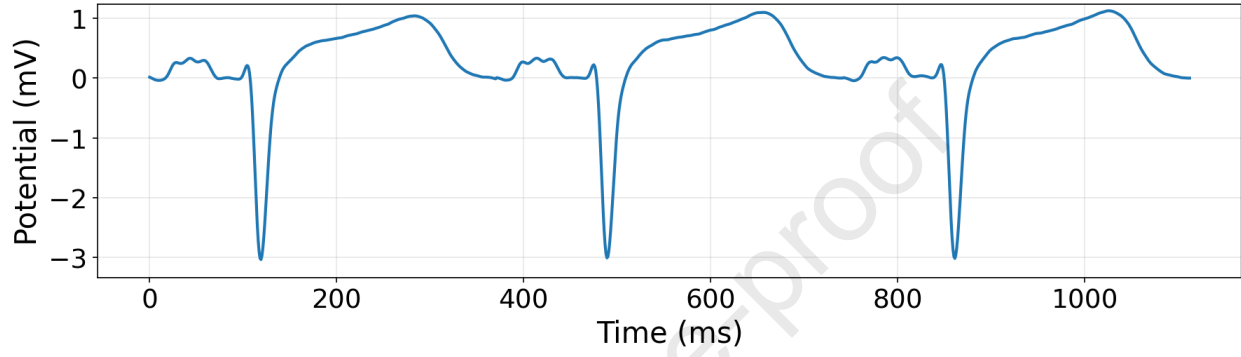
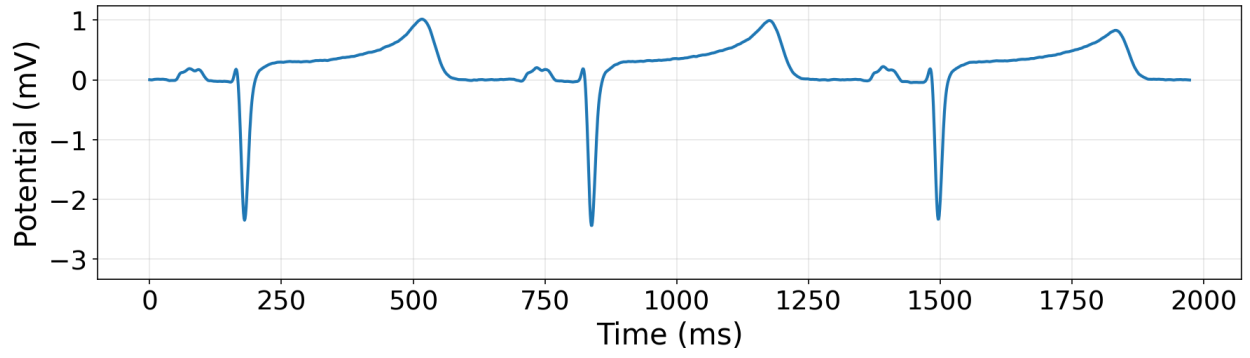
Journal Pre-proof





Journal Pre-proof





Key Findings

- Deep-learning models (6-layer LSTM with attention) accurately forecast the timing of premature ventricular contractions (PVCs) during acute myocardial ischemia from short epicardial electrogram sequences; pooled cross-validation achieved low error and excellent calibration (e.g., MAE 6.50→4.73 s as input increased from 3→7 beats; $R^2 \geq 0.996$).
- Providing more temporal context improves accuracy: using 7 beats lowered MAE by ~27% vs 3 beats and yielded the best calibration and threshold performance.
- In a subject-specific setting (leave-one-experiment-out), limited fine-tuning substantially improves performance, adapting with 15% of subject data reduced MAE by ~10-15 seconds versus 10%, and flattens error-vs-horizon slopes, with the strongest gains observed for 5-7 beats.
- Prediction accuracy is high within clinically actionable windows: in pooled testing, 96-99% of estimates were within 30 seconds and $\geq 99.5\%$ within 60 seconds; with personalization, 60-second accuracy improved to ~77-80% at 15% fine-tuning.
- Findings establish the feasibility of proactive, real-time arrhythmia risk management: epicardial signals contain predictive signatures, and only modest subject-specific data are needed to personalize forecasting models.