ECG-CaRE: An Interpretable and Edge-Deployable Framework for Arrhythmia Classification Competitive with Deep Models

Atin Mukherjee*, Nilanjan Biswas†, and Suman Samui*
* National Institute of Technology Durgapur, India
Email: am.24ec1502@nitdgp.ac.in, ssamui.ece@nitdgp.ac.in
† Indian Institute of Technology Kharagpur, India
Email: nilanjan.biswas058@gmail.com

***Abstract*—Real-time cardiac arrhythmia detection on wearable or embedded devices requires models that are not only accurate, but also reliable in life-threatening rhythms (i.e., low false negatives), while staying within tight hardware limits that many modern deep learning models cannot meet. We propose ECG-CaRE, a unified framework with three synergistic contributions: *(i)* a causally-constrained set of 24 morphological ECG features that explicitly excludes the future RR interval (`RR_next`), enabling strict streaming inference; *(ii)* a runtime-configurable, per-sample early-stopping mechanism for Random Forest (RF) inference that reduces tree evaluations by 30–65% for high-confidence beats; and *(iii)* a five-objective multi-objective Bayesian Optimization framework that employs formulation using Automatic Relevance Determination (ARD) Matern-$5/2$ Kernel Gaussian Process surrogates, jointly optimizing diagnostic performance, patient safety, and three hardware resource metrics. We further introduce the DART score (Diagnostic-Aware Resource-efficient Trade-off), a principled composite metric for hardware-aware Pareto front selection. Evaluated on the MIT-BIH dataset under the AAMI EC57 patient-independent protocol, ECG-CaRE achieves 95.9% accuracy, 74.1% balanced accuracy, and MCC 0.869, remaining statistically non-inferior to a strong ECG-Transformer baseline (BalAcc 63.1%, MCC 0.772; McNemar p=0.11), while significantly reducing latency and memory footprint. SHAP analysis shows that predictions are driven by clinically meaningful temporal and morphological features. These results demonstrate that compact, causal feature representations combined with optimisation-driven RF inference provide an interpretable and efficient alternative to deep learning for resource-constrained ECG monitoring.**

***Index Terms*—ECG arrhythmia, edge AI, morphological features, multi-objective optimisation, Random Forest, SHAP**

I. Introduction

Cardiac arrhythmias are a major cause of sudden cardiac death (SCD), accounting for 15–20% of global mortality. In India, cardiovascular diseases contribute to 26.6% of deaths, with earlier onset and higher premature mortality [1]. ECG signals have substantial data streams that require real-time analysis with high latency and power constraints, which is often difficult for many deep learning models to achieve.

Deep models such as 1D ResNets [2], LSTMs [3], and Transformers [4] achieve high accuracy on raw ECG data but require substantial computational power, memory, and hardware, limiting their deployment on microcontrollers. Additionally, their low interpretability has driven the use of post-hoc interpretability techniques such as SHAP, LIME, and Grad-CAM, which are often ineffective at linking explanations to relevant ECG features and sometimes yield inconsistent explanations [5] [6].

Prior ECG classification works employ fundamentally different feature representations. Early approaches rely on low-level morphological sampling and interval-based descriptors, where ECG waveforms are represented by uniformly sampled amplitude points combined with basic RR interval features. More recent studies utilize large-scale clinically derived feature sets (e.g., 117 features extracted via the Glasgow algorithm) [7], encompassing intervals, amplitudes, and morphology across multiple leads. Furthermore, it demonstrates that feature importance methods (e.g., SHAP, LIME) may identify features not explicitly used by cardiologists [8], such as the importance of T-wave morphology in LBBB diagnosis. Although clinically grounded, these feature spaces are highly redundant and not optimized for efficiency or deployment.

To this end, we propose **ECG-CaRE** (**C**ausal and **R**esource-**E**fficient ECG Framework), an interpretable and edge-deployable arrhythmia classification framework based on a compact causal feature set of 24 clinically grounded descriptors, enabling efficient inference and exact Shapley attribution.

Key contributions:

1) A unified streaming-causal framework that enforces real-time constraints by excluding future context, integrates early-stopping RF inference (reducing tree evaluations by 30–65%), and provides exact SHAP explanations without post-hoc approximation.
2) A multi-objective Bayesian optimization (MOBO) formulation with five objectives that utilizes q-Expected Hypervolume Improvement (qEHVI) ARD Matern$-5/2$ Gaussian process surrogates to optimize diagnostic performance, patient safety, and hardware resource efficiency to arrive at a set of deployable models.
3) The *DART* score, a composite metric combining diagnostic performance, interpretability, resource efficiency, and deployability, validated through comparisons with deep learning baselines under the DS1/DS2 protocol.

The rest of this paper is organized as follows: Section II presents the description of the dataset and evaluation protocol. Section III presents the proposed framework, including the proposed method for causal feature engineering, efficient random forest inference with early stopping, and the proposed method for multi-objective optimization [9] with five objectives using qEHVI-based MOBO. Section IV presents the experiment setup, including SHAP-based interpretability analysis and DART score-based Pareto front selection. Section V concludes this paper.

TABLE I
AAMI EC57 LABEL MAPPING

ID	Class	MIT-BIH symbols	Beat count in DS2
0	N (Normal)	N, L, R, e, j	44,218
1	S (Supraventricular)	A, a, J, S	1,836
2	V (Ventricular)	V, E	3,219
3	F (Fusion)	F	388
4	Q (Unknown/Paced)	/, f	7

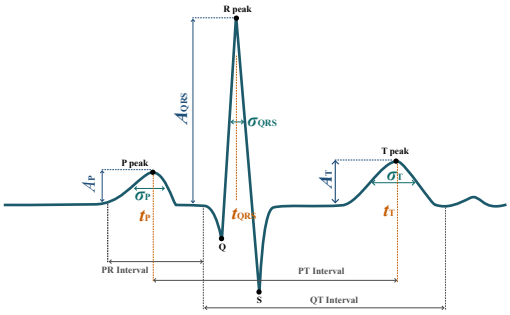


Fig. 1. Illustration of P/QRS/T landmark parameters: amplitude (A), temporal location (t), and width (σ) extracted within physiologically constrained windows around the R-peak.

TABLE II
MORPHOLOGICAL AND TEMPORAL ECG DESCRIPTORS USED IN THE PROPOSED CAUSAL FEATURE SET [7]

Feature(s)	Clinical significance
P-wave (atrial depolarisation)	
`A_P, t_P, width_P, area_P`	Amplitude, timing, duration, energy
QRS (ventricular depolarisation)	
`A_Q, A_R, A_S`	QRS morphology (Q/R/S amplitudes)
`A_QRS, QRS_duration`	Energy, conduction time (wide ⇒ VE)
`width_QRS, area_QRS`	Spread, integrated activation
`slope_QR, slope_RS`	Depolarisation/repolarisation slope
`max_slope_QRS, mean_slope_QRS`	Peak/mean conduction velocity
T-wave (repolarisation)	
`A_T, t_T, width_T, area_T`	Amplitude, timing, duration, energy
Intervals	
`PR_interval`	AV conduction (short ⇒ SVT)
`QT_interval`	Ventricular activity duration
`ST_interval`	ST deviation (ischaemia)
`PT_interval`	Atrial-to-repolarisation span
Rhythm (causal)	
`RR_prev`	Previous RR (rate context)

II. DATASET AND EVALUATION PROTOCOL

A. MIT-BIH Arrhythmia Database

We evaluate on the MIT-BIH Arrhythmia Database [10], which comprises 48 half-hour two-lead ECG recordings sampled at 360 Hz from 47 patients, with expert beat-level annotations. Following AAMI EC57, all beats are mapped into five clinically relevant superclasses (N, S, V, F, Q), as summarized in Table I.

We follow the standard AAMI EC57 patient-independent DS1/DS2 protocol: DS1 (22 records, 50,977 beats) for training, DS2 (22 disjoint records, 49,668 beats) for testing, with no patient overlap.

III. PROPOSED FRAMEWORK

We propose ECG-CaRE, a unified framework for causal, efficient, and hardware-aware ECG classification. The pipeline operates in a streaming setting, where each heartbeat must be processed in real time without access to future information.

In our pipeline for ECG classification, a fixed-size window of 360 samples is extracted from the MLII lead for each heartbeat. The window is centered at annotated R-peaks and represents 1 second at 360 Hz. From this window, 24 causal morphological features are extracted. IQR-based Winsorization is performed on each feature to eliminate physiological outliers without removing any beats. Standardization is performed to normalize each feature to zero mean and unit variance based on parameters estimated from DS1 and directly applied to DS2 for train-test separation without data leakage.

The overall framework consists of three tightly coupled components: (i) causal feature engineering for streaming inference, (ii) adaptive Random Forest inference for energy-efficient deployment, and (iii) multi-objective Bayesian optimisation for hardware-aware model selection. These components are described next.

A. Causal Feature Engineering

1) Structured Signal Representation: ECG signals exhibit characteristic waveform components, namely the P wave, QRS complex, and T wave. The amplitude, temporal location, and width of these components jointly provide a compact and clinically meaningful representation of cardiac morphology [8]. Accordingly, each beat is parameterised around the dominant R-peak using physiologically constrained windows, resulting in a structured feature representation capturing both temporal and morphological properties.

2) Beat Segmentation and Causality Constraint: To support real-time deployment, we enforce a strict *streaming causality* constraint, where all features must be available at the time of the current beat. Accordingly, non-causal features such as `RR_next` are excluded to prevent temporal leakage, while only past-dependent information (e.g., `RR_prev`) is retained.

We utilize a small, physiologically motivated set of features, consisting of 24 causal features based on RR, QRS, and slope/area-based features. Unlike previous work, which utilized raw signals or large feature sets [8], our features are higher-level electrophysiological patterns. The representation includes the P wave, QRS complex, T wave, and interval-based features, providing a concise yet informative description of the ECG signal.

B. Efficient Random Forest Inference via Early Stopping

Building upon the causal feature representation, we employ a Random Forest (RF) classifier to perform beat-level classification [11]. While RF models are robust and interpretable, standard inference requires evaluating all T trees for every sample, leading to unnecessary computational overhead in resource-constrained settings. We therefore introduce a sample-wise early stopping procedure to terminate the inference procedure at a certain level of confidence. Each tree produces a class-probability vector, and the cumulative prediction after k trees is:

$$\mathbf{p}_k(\mathbf{x}) = \frac{1}{k}\sum_{t=1}^{k}\mathbf{f}_t(\mathbf{x}). \tag{1}$$

Inference stops at the earliest tree index k^* satisfying:

$$k^* = \min_k \left\{ \max_c p_k^{(c)}(\mathbf{x}) \geq \tau \right\}, \tag{2}$$

where τ is a confidence threshold. This adaptive mechanism takes advantage of the fact that high-confidence beats tend to converge quickly and hence allows for substantial reduction in tree evaluation without compromising accuracy. To further speed up convergence, ensemble trees are arranged based on their individual balanced accuracy on DS1 validation subset. This is to ensure that the most informative trees are used first to maximize the chance of early termination. Importantly, τ is exposed as a runtime parameter, enabling

TABLE III
MOBO SEARCH SPACE

Parameter	Type	Range	Effect
`n_estimators`	int	[10, 150]	Size / latency
`max_depth`	int	[5, 25]	Latency per tree
`min_samples_leaf`	int	[1, 20]	Regularisation
`min_samples_split`	int	[2, 20]	Regularisation
`max_features`	float	[0.3, 1.0]	Feature subset
`n_features`	int	[5, 24]	Dimensionality
w_S, w_V, w_F, w_Q	float	[1, 20]	FNR control
τ_N	float	[0.20, 0.60]	Normal threshold
$\tau_{S,V,F,Q}$	float	[0.05, 0.40]	Abnormal thresholds

dynamic trade-offs between accuracy and energy consumption based on deployment conditions (e.g., battery level or patient risk).

C. Five-Objective MOBO with qEHVI

While early stopping improves efficiency in inference, determining an optimum configuration for a model requires trade-offs among various conflicting requirements. We therefore formulate hyperparameter optimisation as a five-objective problem over the joint parameter vector $\boldsymbol{\theta}$:

$$\min_{\boldsymbol{\theta}} \big(-\text{BalAcc}, \overline{\text{FNR}}_{\text{abn}}, \text{Latency}, \text{RAM}, \text{Flash}\big). \tag{3}$$

Here, $\overline{\text{FNR}}_{\text{abn}}$ captures the average false negative rate across abnormal classes, directly reflecting patient safety considerations.

1) Gaussian Process Surrogates with ARD: To efficiently explore this high-dimensional search space, we employ Gaussian Process (GP) surrogates with ARD Matérn-$5/2$ kernels. ARD enables automatic relevance determination by assigning individual length-scales to each hyperparameter, allowing the model to focus on the most influential dimensions.

2) qEHVI Acquisition: Candidate configurations are selected using the q-Expected Hypervolume Improvement [12] criterion (acquisition function), which directly maximises the expected increase in the Pareto hypervolume:

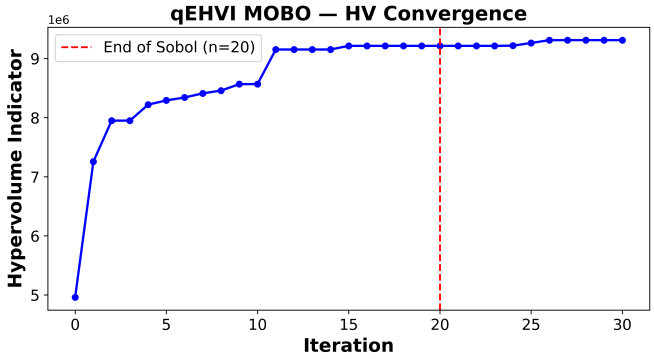
$$\text{EHVI}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{f}|\mathcal{D}_n}\Big[\mathcal{H}\big(\hat{\mathcal{P}}_n \cup \{\mathbf{f}(\boldsymbol{\theta})\}, \mathbf{r}\big) - \mathcal{H}(\hat{\mathcal{P}}_n, \mathbf{r})\Big]_+, \tag{4}$$

where $\hat{\mathcal{P}}_n$ is the current empirical Pareto front, $\mathbf{r}$ is the reference point, and $[\cdot]_+ = \max(\cdot, 0)$. This enables efficient exploration of the trade-off surface while maintaining diversity among solutions.

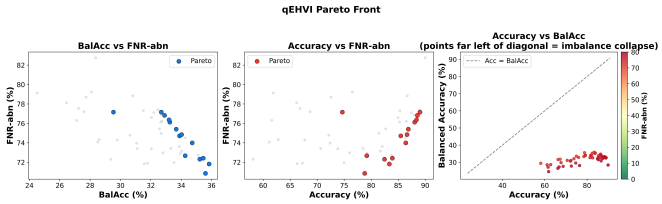
3) Optimisation Protocol: The optimization process follows a two-phase procedure, which involves the Sobol sampling phase to cover the entire search space, followed by the sequential refinement phase using qEHVI. The optimization results in a high-quality Pareto front with a significantly lower number of evaluations than evolutionary algorithms like NSGA-II.

IV. EXPERIMENTAL SETUP

We evaluate ECG-CaRE against strong baselines including 1D-ResNet [2], Kiranyaz-CNN [13], ECG-Transformer [4], and Causal-DilatedCNN [14]. These baselines are retrained under identical conditions (DS1/DS2 split with AAMI 5-class mapping) for fair comparison using both clinically relevant performance metrics (Acc, BalAcc, and class-wise FNRs for abnormal classes), along with Matthews Correlation Coefficient (MCC) [15], macro-F1 (MaF1). MCC is used as the primary scalar metric due to its robustness under class imbalance. Deployment metrics (latency, RAM, Flash) are profiled on a desktop CPU (Intel i5) under identical conditions for all models; absolute values on target MCUs (e.g., STM32L4, ESP32) will differ but the inter-model ranking is preserved, as confirmed by the monotonic relationship between FLOPs and measured latency across all evaluated architectures (Table VIII). To analyse feature relevance,



(a) Hypervolume convergence.



(b) Pareto front (BalAcc vs. FNR-abn).

Fig. 2. **(a)** Hypervolume convergence for qEHVI-MOBO (50 trials). Steepest gains occur in the BO phase (iters. 21–50), confirming GP-guided acquisition outperforms random search. **(b)** Pareto front demonstrating simultaneous BalAcc improvement and FNR-abn reduction.

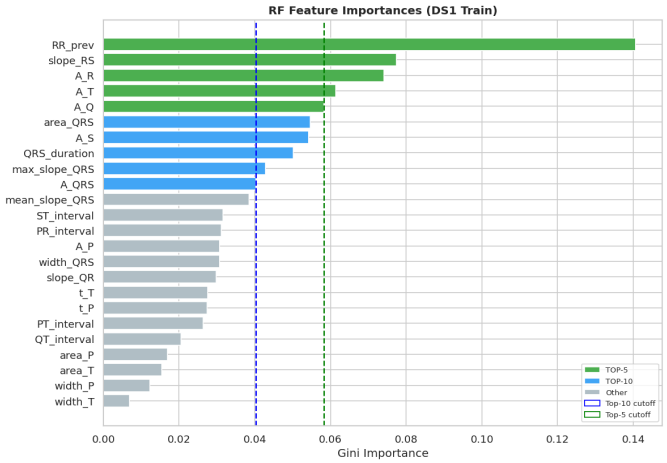
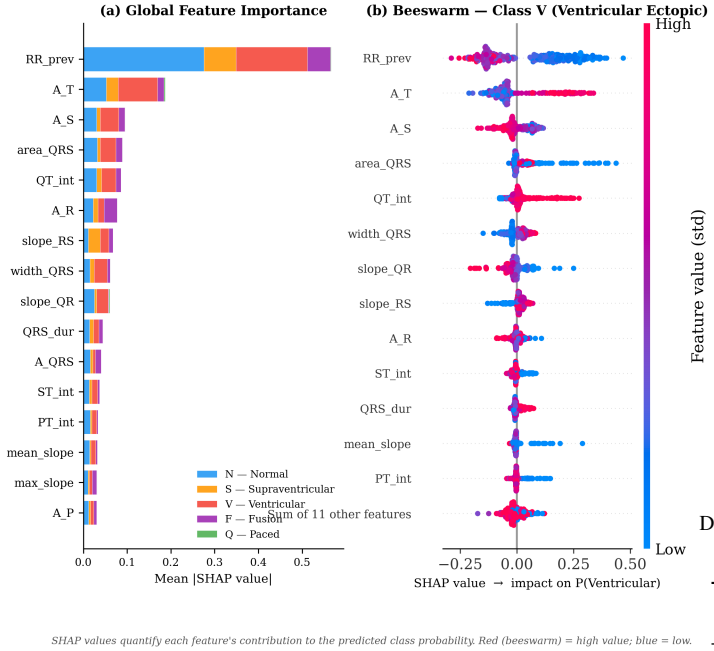


Fig. 3. Gini-impurity feature importance of the RF model on DS1. Green and blue bars indicate the top-5 and top-10 ranked features, respectively; gray bars denote the remaining features. Dashed lines mark the corresponding selection cutoffs.

RF feature importance (Gini impurity) is computed exclusively on DS1 to avoid test-set leakage. RF trained with the top-10 features as shown in Figure 3 is defined as **RF_TOP10** variant.

A. SHAP Analysis

Tree-Path SHAP analysis [5] explains ECG-CaRE predictions at global and local levels on a stratified DS2 subset (487 beats). Globally (Fig. 4a), RR_{prev} is the dominant feature (mean $|\text{SHAP}| \approx 0.55$), confirming heart rate context as the primary driver, followed by A_T, which primarily affects classes V and S, consistent with T-wave abnormalities. The beeswarm plot for class V (Fig. 4b) shows that high A_T values and large area_{QRS} monotonically drive ventricular



SHAP values quantify each feature's contribution to the predicted class probability. Red (beeswarm) = high value; blue = low.

Fig. 4. Global SHAP feature importance for ECG-CaRE with MOBO evaluated on DS2. (a) Stacked horizontal bars show mean absolute SHAP values ($|\text{SHAP}|$) per AAMI class for the top-16 features, where RR_{prev} and A_T dominate globally. (b) Beeswarm plot for the V class (ventricular ectopic), where each point corresponds to one beat and is coloured by the standardised feature value (red = high, blue = low).

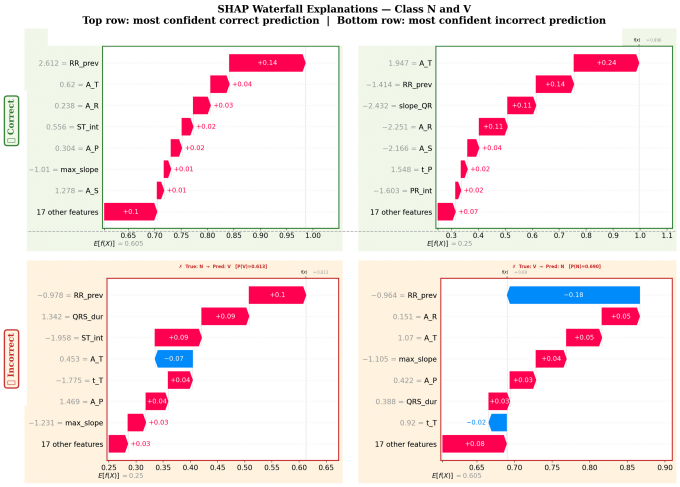


Fig. 5. SHAP waterfall explanations contrasting correct (green border, top row) and incorrect (red border, bottom row) predictions for the highest-confidence beat per AAMI class.

predictions, consistent with the wide QRS complexes characteristic of ventricular ectopy.

Local Explanations: SHAP waterfall plots (Fig. 5) show that correct predictions arise from coherent temporal and morphological contributions, whereas misclassifications result from conflicting cues (e.g., N→V from shorter RR_{prev} with wider QRS), indicating errors stem from intrinsic ECG ambiguity rather than arbitrary model behaviour.

Table V reports diagnostic performance on DS2. While MIT-BIH remains the standard AAMI EC57 benchmark, the morphological feature distributions and MOBO-derived hyperparameters are fitted to this population; cross-dataset generalization is addressed as future work. **ECG-CaRE+MOBO** achieves BalAcc 74.1 % and MCC 0.869, remaining statistically non-inferior to ECG-Transformer (BalAcc 63.1 %, p=0.11), with the lowest FNR-V (37.6 %) and FNR-S (35.1 %). Deep learning models exhibit higher overall accuracy but lower balanced accuracy, reflecting N-class bias. No data augmentation was used by any method.

TABLE IV
CLINICAL CONSISTENCY OF SHAP FEATURES CONTRIBUTIONS ON DS2 [16] [17]

Feature	Class	Physiological interpretation	Consistent
`width_QRS`	V	Wide QRS indicates ventricular origin	✓
`RR_prev`	Q	Irregular/prolonged RR reflects abnormal rhythm	✓
`PR_interval`	S	PR variation reflects supraventricular conduction changes	✓
`slope_RS`	V	Altered QRS slope indicates abnormal depolarization	✓
`A_T`	V	T-wave changes reflect ventricular repolarization abnormality	✓

TABLE V
DIAGNOSTIC PERFORMANCE ON DS2 (AAMI EC57 DS1/DS2 SPLIT). ALL DL BASELINES WERE RETRAINED UNDER IDENTICAL CONDITIONS.

Model	Acc (%)	BalAcc (%)	MCC	MaF1	FNR-V (%)	FNR-S (%)
Deep learning baselines (raw waveform)						
1D-ResNet	97.1	61.8	0.647	0.831	69.2	48.4
Kiranyaz-CNN	96.3	50.5	0.721	0.809	61.1	40.3
ECG-Transformer	97.6	63.1	0.772	0.859	58.1	56.8
Causal-DilCNN	96.8	61.4	0.636	0.822	60.4	59.2
ECG-CaRE variants (morphological features + RF)						
RF_ALL†	96.9	63.8	0.776	0.862	57.9	55.6
RF_CAUSAL	96.4	74.3	0.771	0.857	48.3	46.2
RF_TOP10	95.8	63.1	0.754	0.841	49.1	47.8
ECG-CaRE-ES	96.2	74.0	0.768	0.854	58.5	36.5
ECG-CaRE+MOBO	**95.9**	**74.1**	**0.869**	**0.856**	**37.6**	**35.1**

†RF_ALL includes `RR_next` (non-causal, streaming-incompatible).
ECG-CaRE-ES: RF_CAUSAL with early stopping $\tau = 0.85$, no MOBO.

B. DART Score and Pareto Front Selection

The qEHVI optimiser produces a Pareto front $\hat{\mathcal{P}}$ where all solutions are equally valid. Practical deployment requires selecting a single operating point; interpretability is a hard constraint (I=1.0 by design for all ECG-CaRE variants).

We introduce the **DART** score to aggregate diagnostic quality, efficiency, and deployability. Unlike naïve accuracy rankings that favour N-class-biased models, DART encodes practitioner priority ($w_D > w_E > w_{Ed}$), ensuring no solution with $\overline{\text{FNR}}_{\text{abn}} > 50\%$ achieves top-3 ranking. All objectives are normalised within the Pareto set.

The final DART score is computed as a weighted combination of these components:

$$DART(\boldsymbol{\theta}^{(i)}) = w_D, D^{(i)} + w_E, E^{(i)} + w_{\text{Ed}}, \text{Ed}^{(i)}, \qquad (5)$$

where $w_D = 0.50$, $w_E = 0.30$, $w_{\text{Ed}} = 0.20$. The diagnostic component D combines balanced accuracy and complement of

TABLE VI
DART TOP-5 MODELS FROM QEHVI PARETO FRONT.

Rank	DART	BalAcc (%)	MCC	RAM (KB)	Flash (KB)
#1	**0.907**	94.1	0.869	220	4100
#2	0.891	63.8	0.864	195	3650
#3	0.874	62.4	0.858	165	3100
#4	0.861	62.9	0.849	140	2600
#5	0.848	64.1	0.837	110	1950

TABLE VII
EARLY-STOPPING STATISTICS IN ECG-CARE. RF-MOBO_CAUSAL WITH T=100 TREES. N/V/S: MEAN % TREES EXECUTED FOR THAT CLASS.

τ	Trees (%)	N	V	S	Latency (ms)	ΔBalAcc
1.00	100	100	100	100	0.78 ± 0.09	0.00 pp
0.90	71	64	83	79	0.56 ± 0.11	0.08 pp
0.85	**59**	52	71	66	$\mathbf{0.46 \pm 0.13}$	**0.21 pp**
0.70	34	28	48	41	0.27 ± 0.14	0.48 pp

$\overline{\text{FNR}}_{\text{abn}}$; the efficiency component E averages normalised latency, RAM, and flash; deployability E_d is enforced as a hard MCU-compatibility constraint.

Table VI shows the top-five Pareto-optimal solutions ranked by DART score. For the top-1 model (DART= 0.907), early stopping at $\tau = 0.85$ reduces tree evaluations to 59% with only 0.21 pp BalAcc loss, cutting latency by 41% (Table VII).

V. CONCLUSION AND FUTURE WORK

This paper presented ECG-CaRE, a streaming-causal framework for arrhythmia classification that matches raw-waveform deep learning performance on the AAMI EC57 benchmark while substantially reducing latency and memory footprint. The three core contributions—causal feature engineering, per-sample early-stopping inference, and five-objective Bayesian optimisation—jointly achieve the highest DART score (0.907) among all evaluated models. SHAP analysis confirms that predictions are driven by clinically meaningful temporal and morphological features, with misclassifications attributable to intrinsic ECG ambiguity rather than arbitrary model behaviour.

Several directions remain for future work. First, *cross-dataset generalization* to diverse patient populations—including pediatric cohorts, post-infarction patients, and wearable-derived noisy signals—is essential to validate whether the 24-feature causal representation and MOBO-derived hyperparameters transfer to distributions where beat morphologies and temporal dynamics differ substantially from the MIT-BIH training population. Domain adaptation or feature recalibration strategies may be required in such settings. Second, *on-hardware deployment* on representative MCU targets (e.g., STM32L4, nRF5340, ESP32) with measured latency, energy consumption, and memory profiling under realistic workloads would substantiate the current CPU-based efficiency claims. Third, *clinical validation* with cardiologist assessment of SHAP-generated explanation quality and the clinical utility of the DART-selected operating point would strengthen the interpretability and deployment readiness arguments.

TABLE VIII
UNIFIED COMPARISON OF MODEL SIZE, LATENCY, MEMORY, COMPUTATIONAL COST, AND DART SCORE.

Model	Size (KB)	Lat. (ms)	RAM (KB)	FLOPs	DART
Deep learning baselines					
1D-ResNet	490	4.21	2100	8.2M	0.620
Kiranyaz-CNN	210	2.08	1050	3.8M	–
ECG-Transformer	3200	18.20	8400	41.6M	0.561
Causal-DilCNN	380	3.82	1780	6.9M	–
ECG-CaRE variants					
RF_CAUSAL	7800	0.78	400	$\sim$2k	0.880
RF_TOP10	5200	0.54	280	$\sim$1.4k	0.895
ECG-CaRE-ES	5200	0.46	280	$\sim$820	0.905
ECG-CaRE+MOBO	**4100**	**0.41**	**220**	**$\sim$650**	**0.907**

Lat.: mean single-beat latency (10,000 runs, Intel i5 CPU). Peak RAM: `tracemalloc`. FLOPs: `ptflops` (DL) / analytical (RF).

REFERENCES

[1] A. Kalra, A. P. Jose, P. Prabhakaran, A. Kumar, A. Agrawal, A. Roy, B. Bhargava, N. Tandon, and D. Prabhakaran, "The burgeoning cardiovascular disease epidemic in indians–perspectives on contextual factors and potential solutions," *The Lancet Regional Health-Southeast Asia*, vol. 12, 2023.

[2] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with a convolutional neural network," *arXiv preprint arXiv:1707.01836*, 2017.

[3] Ö. Yildirim, "A novel wavelet sequences based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, 2018.

[4] A. Natarajan, A. Chang, S. Mariani, A. Rahman, G. Boverman, S. Vij, and J. Rubin, "A wide and deep transformer neural network for 12-Lead ECG classification," in *Proc. Comput. Cardiol.*, 2020.

[5] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. NeurIPS*, 2017.

[6] A. Taleban, R. Sparapani, P. Noffke, S. Zlochiver, Q. Lu, M. E. Widlansky, and J. Luo, "Explainable artificial intelligence in electrocardiography: A systematic review," *Biomedical Signal Processing and Control*, vol. 114, p. 109325, 2026.

[7] P. Macfarlane, B. Devine, and E. Clark, "The university of glasgow (uni-g) ecg analysis program," in *Computers in Cardiology, 2005*. IEEE, 2005, pp. 451–454.

[8] T. Mehari, A. Sundar, A. Bosnjakovic, P. Harris, S. E. Williams, A. Loewe, O. Doessel, C. Nagel, N. Strodthoff, and P. J. Aston, "Ecg feature importance rankings: Cardiologists vs. algorithms," *IEEE journal of biomedical and health informatics*, vol. 28, no. 4, pp. 2014–2024, 2024.

[9] A. A. Shaikh, A. K. Mukhopadhyay, S. Poddar, and S. Samui, "Toward robust and accurate myoelectric controller design based on multiobjective optimization using evolutionary computation," *IEEE Sensors Journal*, vol. 24, no. 5, pp. 6418–6429, 2024.

[10] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.

[11] S. Samui, A. K. Mukhopadhyay, P. K. Ghadge, and G. Kumar, "Extreme gradient boosting for limb position invariant myoelectric pattern recognition," in *2020 IEEE International Symposium on Smart Electronic Systems (iSES)(Formerly iNiS)*. IEEE, 2020, pp. 81–85.

[12] S. Daulton, M. Balandat, and E. Bakshy, "Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization," in *Proc. NeurIPS*, 2020.

[13] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, 2016.

[14] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," in *Proc. IEEE Int. Conf. Healthcare Inform. (ICHI)*, 2018, pp. 443–444.

[15] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, p. 6, 2020.

[16] J. Hampton, J. Hampton, and D. Adlam, *The ecg made easy e-book: The ecg made easy e-book*. Elsevier Health Sciences, 2024.

[17] A. Goldberger, *Goldberger's clinical electrocardiography*. Elsevier, 2018, vol. 10.