PINNAC: Theory-Trained Neural Networks for Robust Arrhythmia Classification

Dhruvika Solanki[1], Al Zadid Sultan Bin Habib[1,2,+], Md Younus Ahamed[1,2,+], Brijesh Patel[3], and Donald A. Adjeroh[1,2,*]

[1]AI+Health Engineering Center, West Virginia University, Morgantown, WV 26506, USA
[2]Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506, USA
[3]Indiana University School of Medicine, Indianapolis, IN 46074, USA
[*]donald.adjeroh@mail.wvu.edu
[+]these authors contributed equally to this work

ABSTRACT

Electrocardiogram (ECG) based arrhythmia screening must remain reliable under physiologic artifacts, data imbalance, and distribution shift. We introduce PINNAC, a theory-trained neural network for arrhythmia classification that couples a standard neural backbone (CNN or MLP) with physics-informed compact excitable-cell surrogates governed by the Aliev Panfilov and FitzHugh Nagumo biophysical models. These physics modules enforce electrophysiological consistency through differentiable constraints or distillation losses, yielding latent states that are concatenated with learned features for classification. A confidence-weighted, warm-up schedule balances the physics and data terms during training, and test-time aggregation reduces prediction variance. We evaluate PINNAC on the MIT–BIH Arrhythmia Database and Noise Stress Test Database across clean, noisy, and reduced-training scenarios, as well as on a feature-engineered tabular variant with Gaussian perturbations. The physics-informed models consistently outperform strong baselines, showing higher accuracy and stronger class-balanced performance under noise and data scarcity. Ablation analyses show that using both physics priors jointly improves generalization and robustness compared to single-model constraints. These results demonstrate that embedding compact electrophysiological priors into deep neural network classifiers improves resilience to signal corruption, supports interpretable latent dynamics, and provides a practical path toward robust, physiology-aware ECG diagnostics.

Introduction

Cardiac arrhythmias remain a major contributor to morbidity and mortality worldwide[1,2]. Atrial fibrillation alone affected more than 59 million people globally in 2019 and continues to rise with aging populations. Cardiac-related sudden death remains a leading killer, accounting for an estimated 10–20% of all deaths[3,4]. Recent inhalation studies show that the physicochemical properties of particulate-ozone mixtures drive oxidative burden and airway dysfunction, a pathway with clear implications for cardiovascular health via cardiopulmonary interactions[5,6]. Early identification of cardiac arrhythmia from ECGs can reduce adverse events, yet deployment at scale is hindered by signal variability, patient heterogeneity, class imbalance, and physiologic artifacts such as electrode motion (EM), muscle activity (MA), and baseline wander (BW)[1]. These factors induce distribution shifts across devices and cohorts, limiting the reliability of learned models outside their training conditions. Traditional diagnostic methods[7–13] are based on medical theory and remain interpretable, but they lack robustness and are difficult to apply reliably in clinical settings. Their performance depends on accurate feature extraction, since ECG analysis requires precise detection of key points in the signal. Small errors in locating these points can grow and affect the final diagnosis. In addition, the shape of ECG waveforms varies across patients and arrhythmia types, making it hard for these models to capture small but important differences. Together, these issues limit diagnostic accuracy and often lead to higher misclassification rates in real-world deployment. Recent advancements in deep learning have exhibited remarkable performance and driven significant advances in automatic feature learning for arrhythmia classification, especially with architectures such as Deep Neural Networks (DNNs)[14], Convolutional Neural Networks (CNNs)[15], and Recurrent Neural Networks (RNNs)[16].

Nonetheless, purely data-driven models can overfit to the training distribution, struggle under physiologic noise and limited data, and offer limited insight into the mechanisms behind their decisions[1]. Conversely, mechanistic electrophysiology provides structure and interpretability from biophysical tissue models to compact excitable-cell surrogates such as the FitzHugh Nagumo (FHN)[17,18] and the Aliev Panfilov (AP)[19,20] models. However, these physics-based theoretical models are computationally intensive in their full form and, when used alone, can be too coarse to reproduce sharp QRS morphology or to adapt to sensor

artifacts and patient variability. We pursue a domain-knowledge–aware alternative that fuses data-driven feature learning with physiology. A compact theory-trained Physics-Informed Neural Network (PINN) is coupled with a 1D convolutional classifier and with a multilayer perceptron, yielding end-to-end models (1D CNN+PINN and MLP+PINN). The PINN is a differentiable surrogate that estimates latent excitable-cell states $(\hat{V}, \hat{W})$ and is trained to satisfy compact electrophysiological relations; these states are concatenated with CNN/MLP features for classification. A hybrid objective enforces label fit together with a physics-consistency loss on the PINN residuals, while a lightweight reconstruction auxiliary stabilizes the feature space. Training stability is improved by a simple warmup schedule on the physics weight; optional knowledge distillation transfers calibrated decision margins and test-time aggregation (TTA) reduces variance by querying the PINN across a small grid of time and an applied-current parameter I_{app}. This design yields three practical advantages. First, physics-consistent latent states act as an inductive bias that regularizes representations, improving generalization under physiologic perturbations or corruption, as well as when labeled data are limited. Second, the PINN exposes interpretable trajectories in an excitable-cell phase plane that can be inspected alongside ECG morphology, aiding failure analysis and clinical sense-making. Third, the framework remains lightweight and end-to-end: all components are trained jointly without bespoke solvers at inference, enabling reproducible evaluation across different protocols.

Key Contributions: The main contributions of this work are summarized as follows:

- **Dual-Physiology Framework:** We introduce PINNAC, a hybrid physics-informed deep learning framework that integrates two biophysical cardiac models, namely, the Aliev Panfilov (AP) and FitzHugh Nagumo (FHN) systems into standard neural network models for robust classification of arrhythmia. Each physics module constrains latent state dynamics during training, enforcing electrophysiological consistency while retaining the discriminative power of deep models.

- **Model-Agnostic Integration:** The proposed physics heads are incorporated into both multilayer perceptron (MLP) and one-dimensional convolutional neural network (1D-CNN) architectures, demonstrating that the physics-informed regularization generalizes across backbone types.

- **Robustness Evaluation:** Using the raw, feautre extracted MIT-BIH[21] arrhythmia dataset and two diagnostic subsets, we conduct five-fold cross-validation and evaluate model performance under realistic electromyographic noise levels from the Noise Stress Test Database (NSTDB)[22]. The results show consistent gains in accuracy and macro-F1 compared with strong baselines (XGBoost, CatBoost, LGBM, Random Forest, TabNet, standard CNN/MLP, etc.).

Related Work

Machine Learning for Analyzing Cardiac Signals. Early machine learning methods for in-clinic cardiovascular prediction relied on tabular risk factors and classical ensembles, including boosting and hard-voting pipelines, which demonstrated respectable accuracy on coronary and cardiovascular endpoints but required careful feature curation and balancing[23,24]. More recently, broad surveys have charted the field's shift toward deep architectures that learn directly from ECG, outlining trends in convolutional, recurrent, and transformer models as well as open challenges around noise, class imbalance, and deployment[1,25]. Within supervised ECG modeling, strong baselines span improved residual CNNs for inter-patient arrhythmia classification, hybrid CNN-LSTM pipelines with channel attention, transformer-based frameworks, and CNN-Transformer hybrids operating on time-frequency representations without R-peak detection[26–29]. To mitigate label scarcity and enhance transfer, recent work has explored self-supervised pretraining tailored to ECG's temporal-spatial structure, narrowing the gap to fully supervised performance with limited labels[30–33]. Beyond rhythm classification, large-scale studies now use 12-lead ECG to estimate ejection fraction, highlighting both the promise of deep models and the importance of fairness and generalization across diverse clinical populations[34].

Generalized PINN Models. General-purpose physics-informed learning has advanced along three complementary axes. Foundational analyses identified why standard PINNs can underfit stiff or multiscale regimes despite sufficient capacity, pointing to loss-term imbalance, spectral bias, and ill-conditioned landscapes, and motivating curriculum or seq2seq-style training[35]. Subsequent work proposed principled fixes at the objective level, including self-adaptive loss balancing to tune residual, boundary, and data terms during training[36] and replacing the ubiquitous L_2 physics loss when stability or approximation guarantees favor alternative norms for nonlinear control partial differential equations (PDEs)[37]. On the optimization side, multi-objective gradient methods align per-term updates to avoid adverse interference, e.g., dual-cone descent that enforces nonnegative inner products with both PDE and boundary gradients, improving stability on failure cases[38]. Beyond pointwise residuals, region-optimized PINNs extend supervision from collocation points to their neighborhoods to tighten generalization and satisfy hidden higher-order constraints without extra differentiations[39]. Architecturally, sequence models inject long-range dependencies: a transformer formulation recasts PINN supervision over pseudo-sequences with attention and tailored activations,

yielding better propagation of initial conditions and accuracy on high-dimensional PDEs[40]. In addition to these instance-wise advances, operator-learning hybrids combine neural operators with physics guidance to learn families of parametric PDEs efficiently and then refine instance-wise with residual constraints, reducing data needs and improving stability[41]. Parameterized PINNs further amortize solutions across PDE parameters via learned embeddings, accelerating multi-query inference and mitigating known training failures on convection diffusion reaction families[42]. Finally, studies on loss-landscapes formalize the conditioning challenges and motivate mixed first/second-order schedules that attain lower residuals and better solutions, underscoring the role of optimization in reported gains[43].

PINNs for Arrythmia Analysis. Most physics-informed approaches to cardiac electrophysiology target inverse or surrogate modeling rather than supervised rhythm recognition. PINNs have been used to reconstruct activation times from sparse intracardiac measurements by enforcing eikonal-wave physics, improving map accuracy and enabling active measurement selection[44]. Electrophysiological (EP)-PINNs then incorporated monodomain dynamics with AP surrogates to recover action-potential fields and estimate tissue parameters such as diffusion and excitability from limited observations, including optical mapping data[45]. Subsequent studies pushed to more realistic settings, demonstrating PINN reconstructions in complex 3D geometries and fibrillatory regimes and coupling to more detailed ionic models for parameter identification[46,47]. Other efforts embed physics constraints into deep active learning pipelines for efficient sensor placement and robust spatiotemporal prediction[48], and develop physics-informed residual learning with local spatiotemporal support to stabilize training for inverse ECG reconstruction under noise and geometric complexity[49]. A recent narrative review synthesizes these trends across physiological signals, highlighting opportunities and open challenges in loss design, optimization, and data quality[50].

Positioning. The related work motivates a supervised classifier that is explicitly regularized by cardiac electrophysiology while retaining the practical strengths of modern deep models. PINNAC contributes such a formulation by combining two complementary priors: FHN and AP within a single training objective that augments cross entropy with dual physics terms and learns end to end on labeled ECG. Concretely, the loss in (10) couples the data term with both FHN and AP penalties, uses confidence-aware weighting to avoid over-constraining uncertain batches, and applies a warm-up schedule to stabilize optimization. We instantiate this in two backbones: a Dense (MLP (multilayer perceptron)) variant that enforces physics via residual penalties computed by automatic differentiation, and a CNN (convolutional neural network) variant whose physics heads are trained by distillation from pretrained PINNs and fused with learned features for prediction. This design addresses gaps in machine learning in the clinic (robustness under noise and distribution shift), and leverages insights from generalized PINN training (balancing data and physics, improved optimization). It also extends cardiac PINNs beyond inverse modeling to direct arrhythmia recognition, yielding competitive accuracy on clean-signals while delivering improved noise robustness in cross-dataset evaluation.

Methodology

Overview. PINNAC is a supervised ECG classifier regularized by cardiac electrophysiology. The core idea is to couple a data term (cross entropy) with two complementary physics priors: the FHN and AP dynamics, inside one training objective. This yields a unified formulation in which the classifier learns from labels while being softly constrained to produce representations consistent with membrane excitation and recovery. We instantiate PINNAC with two backbones: an MLP and a CNN. Both minimize the same combined loss; they differ only in how the physics terms are realized, with the MLP computing residual penalties by automatic differentiation and the CNN distilling from pretrained physics solvers via regression heads.

Problem Setup. Let $\mathscr{D} = \{(x_i, y_i)\}_{i=1}^{N}$ denote ECG segments with class labels $y_i \in \{1, \ldots, C\}$. A classifier f_θ produces logits $z = f_\theta(x) \in \mathbb{R}^C$ and probabilities $p = \mathrm{softmax}(z)$. The supervised term is the average cross entropy as shown in Eq. 1.

$$\mathscr{L}_{\mathrm{CE}}(\theta) = -\frac{1}{N}\sum_{i=1}^{N} \log p_{y_i}(x_i;\theta) \tag{1}$$

Where N is the number of samples; C the number of classes; $\mathscr{D} = \{(x_i, y_i)\}_{i=1}^{N}$ the labeled ECG dataset with segment x_i and label $y_i \in \{1, \ldots, C\}$; θ are model parameters; f_θ is the classifier mapping x to logits $z = f_\theta(x) \in \mathbb{R}^C$; $p = \mathrm{softmax}(z)$ are class probabilities with component $p_{y_i}(x_i;\theta)$; and $\mathscr{L}_{\mathrm{CE}}(\theta)$ is the average cross-entropy loss in Eq. (1). PINNAC is instantiated with two backbones. PINNAC-MLP computes physics residuals by automatic differentiation inside the network. PINNAC-CNN uses a convolutional backbone with two physics heads trained by distillation from separately trained physics solvers and then fused with the backbone features for classification.

Physics Priors - FitzHugh-Nagumo (FHN). The FitzHugh–Nagumo model is a two-state caricature of excitable membrane dynamics with a fast activator (membrane potential) $V(t)$ and a slow inhibitor (recovery) $W(t)$[17]. In nondimensional form, it is written as in Eq. (2) with parameters $k, a, b, \varepsilon > 0$. The cubic nonlinearity in V creates an N/S shaped V-nullcline that yields

a threshold: small perturbations relax to a quiescent fixed point, whereas supra-threshold inputs produce a large excursion (the "spike") before slow return along the recovery branch. The small parameter ε enforces a fast-slow time-scale separation, capturing refractoriness and the plateau of the action potential surrogate, while k controls spike gain, a is a shift parameter, and b biases the recovery current. In the phase plane, trajectories follow canard-like segments consistent with excitation-recovery timing; violations of physiologically plausible upstroke/repolarization (e.g., overly rapid recovery or inverted curvature) produce large residuals in our penalty $\mathscr{R}_{\text{FHN}}$. As a prior, FHN therefore imposes a stringent constraint on excitable timing and refractoriness, utilizing only two states, which maintains autograd-based residual evaluation stability and affordability during training[18]. Here, t is time; $V(t)$ is the fast membrane potential (activator); $W(t)$ is the slow recovery variable (inhibitor); $\dot{V} = \mathrm{d}V/\mathrm{d}t$ and $\dot{W} = \mathrm{d}W/\mathrm{d}t$ are time derivatives; $k > 0$ sets activation gain/scaling; $a > 0$ shapes/shifts the cubic nonlinearity; $b > 0$ biases the recovery nullcline; and $\varepsilon > 0$ is the fast-slow time-scale ratio (small).

Physics Priors - Aliev Panfilov (AP). The Aliev Panfilov model enhances the two-variable excitable system by incorporating a reaction term that sharpens activation and an explicit voltage recovery coupling that governs restitution[19]. In our surrogate form (Eq. (3)), the voltage obeys $\dot{V} = kV(V-a)(1-V) - VW + I_{\text{app}}$, while the recovery variable follows $\dot{W} = \varepsilon(bV - W)$. The cubic reaction $kV(V-a)(1-V)$ steepens the upstroke and enforces saturation, the multiplicative coupling $-VW$ tightens the loop between activation and recovery, and the drive I_{app} enables stimulus-locked responses with tunable strength and latency. The recovery equation ties W to V, producing action potential duration adaptation and physiologically realistic V–W phase relationships. Although the full AP model supports wave propagation and spiral dynamics in tissues, in our ODE surrogate, it serves as a compact prior on single-site morphology: it penalizes implausible shapes (overly broad or undershot pulses) and misaligned recovery kinetics. Combined with FHN, AP complements the fast-slow timing constraint by anchoring stimulus-dependent morphology and restitution[20]. In PINNAC we exploit both priors by sampling times t, forming predictions $Y_\theta(x,t) = [\widehat{V}_\theta, \widehat{W}_\theta]$, and applying residual/distillation losses; this yields complementary regularization while preserving training stability and low computational cost. Here, t is time; $V(t)$ the membrane potential (activator); $W(t)$ the recovery variable (inhibitor); $\dot{V} = \mathrm{d}V/\mathrm{d}t$ and $\dot{W} = \mathrm{d}W/\mathrm{d}t$ are time derivatives; $k > 0$ is the reaction/activation gain; a ($0 < a < 1$ typically) shapes the activation cubic (threshold shift); $b > 0$ couples recovery to voltage; $\varepsilon > 0$ sets the fast–slow time-scale ratio; and I_{app} is the applied stimulus current (external drive).

Electrophysiology Priors. Two compact models provide temporal constraints on membrane voltage $V(t)$ and a recovery variable $W(t)$ under an applied current I_{app}. The FHN surrogate used in the implementation as Eq. 2 with parameters $k, a, b, \varepsilon > 0$. The AP surrogate used for training is in Eq. 3. These ODE forms serve as the physics reference in both variants.

$$\dot{V} = kV(1-V)\left(V - \frac{W+b}{a}\right), \quad \dot{W} \quad = \varepsilon\left(-W - kV(V-1)\right) \tag{2}$$

$$\dot{V} = kV(V-a)(1-V) - VW + I_{\text{app}}, \quad \dot{W} \quad = \varepsilon\left(bV - W\right) \tag{3}$$

Where t is time; $V(t)$ is membrane voltage (activator); $W(t)$ is recovery variable (inhibitor); $\dot{V} = \mathrm{d}V/\mathrm{d}t$, $\dot{W} = \mathrm{d}W/\mathrm{d}t$ time derivatives; $k > 0$ activation/reaction gain; a cubic shift/threshold parameter; $b > 0$ recovery bias/coupling; $\varepsilon > 0$ fast-slow time-scale ratio; I_{app} applied stimulus current.

Physics Terms. Let $\widehat{Y}_\theta(x,t) = [\widehat{V}_\theta(x,t), \widehat{W}_\theta(x,t)]$ denote the network's physics output conditioned on a time input t drawn uniformly from $[t_{\min}, t_{\max}]$. In PINNAC-MLP, physics is enforced by residual penalties obtained by substituting $\widehat{Y}_\theta$ into the right-hand sides of (2)-(3) and differentiating $\widehat{Y}_\theta$ with respect to t via autograd in Eq. 4[17,18] and 5[19,20]. Averaging over mini-batch samples and random times yields Eq. 6. In PINNAC-CNN, physics heads are trained by regression to trajectories produced by separately trained PINNs that satisfy (2)–(3). Denoting the teacher trajectories by $\Phi_{\text{FHN}}(t)$ and $\Phi_{\text{AP}}(t, I_{\text{app}})$, the distillation losses are described in Eq. 7.

$$\mathscr{R}_{\text{FHN}}(x,t;\theta) = \left\| \frac{\partial \widehat{V}_\theta}{\partial t} - k\widehat{V}_\theta(1-\widehat{V}_\theta)\left(\widehat{V}_\theta - \frac{\widehat{W}_\theta + b}{a}\right) \right\|_2^2 + \left\| \frac{\partial \widehat{W}_\theta}{\partial t} - \varepsilon\left(-\widehat{W}_\theta - k\widehat{V}_\theta(\widehat{V}_\theta - 1)\right) \right\|_2^2 \tag{4}$$

$$\mathscr{R}_{\text{AP}}(x,t;\theta) = \left\| \frac{\partial \widehat{V}_\theta}{\partial t} - k\widehat{V}_\theta(\widehat{V}_\theta - a)(1-\widehat{V}_\theta) + \widehat{V}_\theta\widehat{W}_\theta - I_{\text{app}} \right\|_2^2 + \left\| \frac{\partial \widehat{W}_\theta}{\partial t} - \varepsilon(b\widehat{V}_\theta - \widehat{W}_\theta) \right\|_2^2 \tag{5}$$

$$\mathscr{L}_{\text{FHN}}^{\text{res}}(\theta) = \mathbb{E}_{x,t}\left[\mathscr{R}_{\text{FHN}}(x,t;\theta)\right], \qquad \mathscr{L}_{\text{AP}}^{\text{res}}(\theta) = \mathbb{E}_{x,t}\left[\mathscr{R}_{\text{AP}}(x,t;\theta)\right] \tag{6}$$

$$\mathscr{L}^{\text{dist}}_{\text{FHN}}(\theta) = \mathbb{E}_{x,t}\|\widehat{Y}_\theta(x,t) - \Phi_{\text{FHN}}(t)\|_2^2, \qquad \mathscr{L}^{\text{dist}}_{\text{AP}}(\theta) = \mathbb{E}_{x,t,I_{\text{app}}}\|\widehat{Y}_\theta(x,t,I_{\text{app}}) - \Phi_{\text{AP}}(t,I_{\text{app}})\|_2^2 \tag{7}$$

Where x ECG input; θ model parameters; $t \sim \text{Unif}[t_{\min}, t_{\max}]$ time; $\widehat{Y}_\theta(x,t) = [\widehat{V}_\theta(x,t), \widehat{W}_\theta(x,t)]$ network physics output; $\partial\widehat{V}_\theta/\partial t$, $\partial\widehat{W}_\theta/\partial t$ time derivatives; $k, a, b, \varepsilon > 0$ ODE parameters; I_{app} applied current; $\|\cdot\|_2^2$ squared ℓ_2 norm; $\mathscr{R}_{\text{FHN}}(x,t;\theta)$, $\mathscr{R}_{\text{AP}}(x,t;\theta)$ residual penalties (Eqs. (4)–(5)); $\mathbb{E}_{x,t}[\cdot]$ expectation over minibatch samples and sampled times; $\mathscr{L}^{\text{res}}_{\text{FHN}}$, $\mathscr{L}^{\text{res}}_{\text{AP}}$ residual losses (Eq. (6)); $\Phi_{\text{FHN}}(t)$, $\Phi_{\text{AP}}(t, I_{\text{app}})$ teacher trajectories; and $\mathscr{L}^{\text{dist}}_{\text{FHN}}$, $\mathscr{L}^{\text{dist}}_{\text{AP}}$ distillation losses (Eq. (7)).

Algorithm 1. Training PINNAC-MLP

initialize parameters θ
for $e = 1$ to E_{pre} **do**
  update θ by minimizing $\mathscr{L}_{\text{CE}}(\theta)$ in (1)
end for
for $e = 1$ to E_{ft} **do**
  set $\lambda^{(e)}_{\text{FHN}}, \lambda^{(e)}_{\text{AP}}$ using (9)
  for each minibatch (x, y) **do**
    compute logits and probabilities; obtain $s(x;\theta)$ via (8)
    draw times t and currents when used
    evaluate $\mathscr{L}^{\text{res}}_{\text{FHN}}$ and $\mathscr{L}^{\text{res}}_{\text{AP}}$ using (6)
    minimize $\mathscr{L}^{(e)}(\theta)$ in (10)
  end for
end for

Algorithm 2. Training PINNAC-CNN

pretrain physics solvers to obtain Φ_{FHN} and Φ_{AP}
initialize CNN backbone, physics heads, and fusion classifier
for $e = 1$ to E **do**
  set $\lambda^{(e)}_{\text{FHN}}, \lambda^{(e)}_{\text{AP}}$ using (9)
  for each minibatch (x, y) **do**
    forward pass to obtain logits and physics predictions
    compute $s(x;\theta)$ using (8) and draw (t, I_{app})
    compute $\mathscr{L}^{\text{dist}}_{\text{FHN}}$ and $\mathscr{L}^{\text{dist}}_{\text{AP}}$ using (7)
    minimize $\mathscr{L}^{(e)}(\theta)$ in (10)
  end for
end for

Confidence Weighting and Warm-up. The physics strength is modulated by the classifier's confidence and by a linear warm-up across fine-tuning epochs. For a sample x, we define Eq. 8 with a floor $s_{\min} \in (0, 1]$. For epoch index e within a fine-tuning phase of length E, the physics weights are given in Eq. 9.

$$s(x;\theta) = \text{clip}\Big(\max_c p_c(x;\theta),\, s_{\min},\, 1\Big) \tag{8}$$

$$\lambda^{(e)}_{\text{FHN}} = \frac{e}{E}\lambda^\star_{\text{FHN}}, \qquad \lambda^{(e)}_{\text{AP}} = \frac{e}{E}\lambda^\star_{\text{AP}} \tag{9}$$

where θ denotes model parameters; $p_c(x;\theta)$ predicted probability for class c; $\max_c p_c(x;\theta)$ classifier confidence; $\text{clip}(u, s_{\min}, 1)$ clamps u to $[s_{\min}, 1]$ with floor $s_{\min} \in (0, 1]$; e epoch index within fine-tuning of length E; $\lambda^{(e)}_{\text{FHN}}$, $\lambda^{(e)}_{\text{AP}}$ epoch-specific physics weights; and $\lambda^\star_{\text{FHN}}$, $\lambda^\star_{\text{AP}}$ target (end-of-warm-up) weights used to scale the linear schedule e/E.

Combined Objective. The final objective joins cross entropy with both physics terms using confidence weighting and warm-up. Let $(\mathscr{L}_{\text{FHN}}, \mathscr{L}_{\text{AP}})$ denote either the residual losses in (6) for PINNAC-MLP or the distillation losses in (7) for PINNAC-CNN. The epoch-specific loss is in Eq. 10. Using both FHN and AP simultaneously in a single objective is the central design choice in PINNAC, with the MLP variant applying residual penalties and the CNN variant applying supervised distillation from pretrained PINNs.

$$\mathscr{L}^{(e)}(\theta) = \mathscr{L}_{\text{CE}}(\theta) + \mathbb{E}_x\Big[s(x;\theta)\Big(\lambda^{(e)}_{\text{FHN}}\mathscr{L}_{\text{FHN}}(\theta) + \lambda^{(e)}_{\text{AP}}\mathscr{L}_{\text{AP}}(\theta)\Big)\Big] \tag{10}$$

Where $\mathscr{L}^{(e)}(\theta)$ is the epoch-e total loss; θ model parameters; $\mathscr{L}_{\text{CE}}$ cross-entropy term; $\mathbb{E}_x[\cdot]$ expectation over training samples (mini-batch); $s(x;\theta)$ confidence weight (Eq. 8); $\lambda^{(e)}_{\text{FHN}}$, $\lambda^{(e)}_{\text{AP}}$ epoch-wise warm-up weights (Eq. 9); and $\mathscr{L}_{\text{FHN}}, \mathscr{L}_{\text{AP}}$ the physics losses (residual for PINNAC-MLP or distillation for PINNAC-CNN).

Unified Formulation. Equations (1)-(10) define one training objective for PINNAC. The only implementation difference between the two variants is the realization of the physics terms $(\mathscr{L}_{\text{FHN}}, \mathscr{L}_{\text{AP}})$: PINNAC-MLP uses residual penalties from Eqs. (4)-(6), while PINNAC-CNN uses supervised distillation from teacher trajectories in Eq. (7). All other components namely, the confidence weighting, warm up, and final fusion into the classifier are identical. Figure 1 presents our end-to-end architecture for PINNAC, our proposed framework for robust analysis of cardiac arrhythmia.

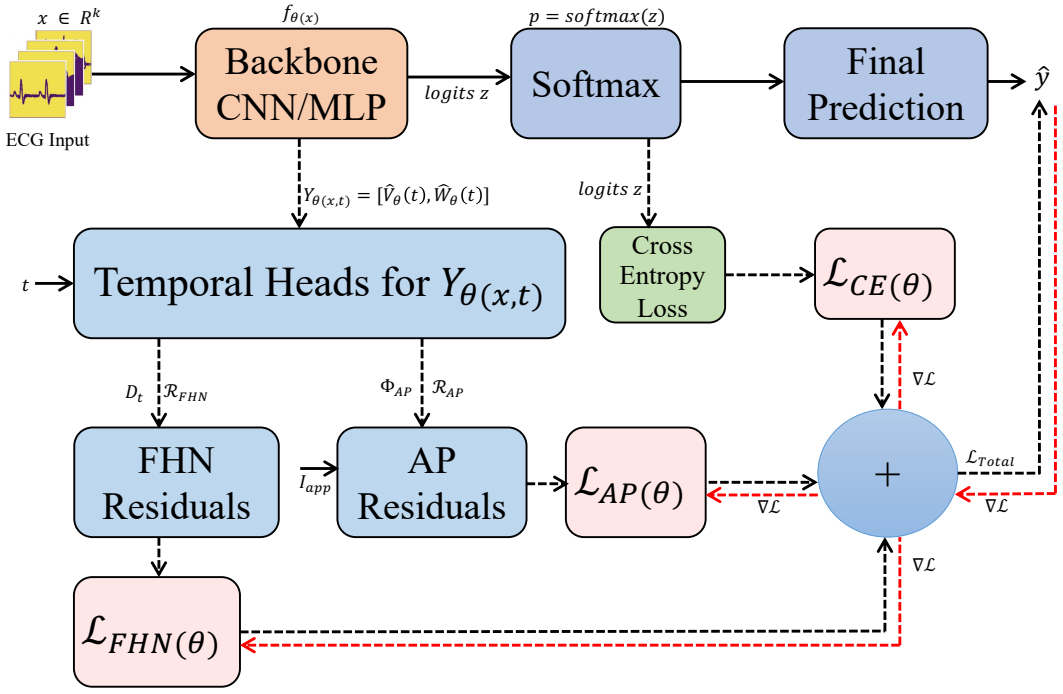


Figure 1. End-to-end architecture for PINNAC. A backbone f_θ encodes ECG x to logits z and probabilities p for the cross-entropy term $\mathscr{L}_{\text{CE}}$. In parallel, temporal heads produce $Y_\theta(x,t) = [\widehat{V}, \widehat{W}]$ to impose electrophysiology via FHN residuals (autograd D_t) and AP residuals/distillation with input current I_{app}, yielding $\mathscr{L}_{\text{FHN}}$ and $\mathscr{L}_{\text{AP}}$, respectively. The total loss combines $\mathscr{L}_{\text{CE}}$ with confidence-weighted, warm-up physics losses; red dashed gradients backpropagate through all branches to the backbone.

Backbones and Training Schedules. PINNAC-MLP follows the same routing as the CNN variant but implements all blocks with fully connected layers. An encoder maps a fixed-length ECG feature vector to a shared representation $h(x)$ through stacked Dense layers with nonlinearities and dropout. Two physics heads take $(h(x), t)$ and $(h(x), t, I_{\text{app}})$, respectively, each built from Dense layers with tanh activations followed by a linear layer that outputs $(\widehat{V}, \widehat{W})$. A fusion layer concatenates $[h(x), \widehat{Y}_{\text{FHN}}, \widehat{Y}_{\text{AP}}]$ and feeds a final Dense classifier producing logits z. Training proceeds in two phases: classification pretraining that minimizes (1), followed by fine-tuning on (10) with confidence weighting (8) and warm-up (9). In this MLP variant the physics terms in (10) are instantiated as the residual penalties in (6) computed by automatic differentiation, while the architecture remains purely Dense throughout. PINNAC-CNN begins by training separate PINNs that solve (2) and (3). The CNN backbone produces features $h(x)$, two physics heads regress to the teacher trajectories using (7), and a fusion layer concatenates backbone features with predicted physics variables before the final classifier. The entire network is optimized with (10) where the physics terms are the distillation losses. Algorithm 1 and Algorithm 2 shows the respective pseudocodes for the two variants of PINNAC.

Experiments and Results

Dataset. Our experiments use both raw ECG signals and a feature-extracted tabular variant. For raw signals, we draw from the MIT-BIH Arrhythmia Database[21,51,52] and employ the MIT-BIH Noise Stress Test Database (NSTD) to introduce controlled noise conditions for robustness evaluation[22,53]. In addition, we use a feature-extracted version of MIT-BIH that provides 32 engineered features per beat and 100,689 samples[54,55]. Table 1 summarizes the sample counts for all datasets and subsets used in our experiments. To study noise robustness, electrode motion (EM) artifacts were added to the raw dataset at signal-to-noise ratios (SNR) of 20 dB, 10 dB, and 5 dB, followed by the addition of zero-mean Gaussian noise at multiple levels (0.1–0.8) to the feature-extracted dataset. To address the pronounced class imbalance in MIT-BIH, we further split the task into five balanced binary cases, ensuring equal class counts whenever possible, which enables fairer training and clearer model comparisons.

Data Scarcity. To analyze the effect of limited data availability, we derive two smaller subsets from the raw MIT-BIH signals. The first subset (Subset1) includes 15,426 beats representing a minimal sample scenario while the second subset (Subset 2) contains 19,676 beats. Both subsets preserve class balance and sampling diversity. This setup allows us to study how performance scales with data quantity and how models behave when only a fraction of the full 90,829 samples is accessible.

Table 1. Dataset and subset sample counts used in the experiments.

Dataset	Total Samples
Feature Extracted Dataset	100,689
Raw Dataset	90,829
Subset 1	15,426
Subset 2	19,676

Baselines. We benchmarked a broad suite of baselines spanning classical machine learning, gradient boosting, and deep tabular models. The classical set includes Naive Bayes, KNN, SVM, Decision Tree, Lasso, Random Forest (RF), and AdaBoost[56]. For boosting-based learners, we evaluated GBM[57], LGBM[58], XGBoost[59], and CatBoost[60]. Deep tabular baselines comprise a 1D CNN, TabNet[61], TabSeq[62], and a standard multilayer perceptron (MLP). To study the effect of physics-informed objectives, we considered two PINNAC variants with CNN and MLP, each using modified loss terms. We also evaluated a variation that integrated both AP and FHN-based physics constraints.

Computational Complexity. Let n be the number of samples, d the feature count, B the batch size, p the number of PINNAC-MLP parameters, L pretrain epochs (CE only), F finetune epochs, and τ the interval between physics steps. A CE step costs $O(Bp)$, so pretraining and the CE part of finetuning together cost $O\big((L+F)\,n\,p\big)$. Each physics step (AP and FHN losses with autograd) costs $O(Bp_{\text{phys}})$, performed about $(n/B)/\tau$ times per epoch, adding $O\big(F\,n\,p_{\text{phys}}/\tau\big)$ cost. Thus total training time is $O\big((L+F)\,n\,p+F\,n\,p_{\text{phys}}/\tau\big)$. Inference time is in $O(p)$ per sample (or $O(np)$ for a dataset), and peak memory is $O\big(p+B(d+s)\big)$ where s is the stored activation size (with a small extra constant factor for the physics head). For PINNAC-CNN, let q be the total parameters of the CNN backbone plus physics heads, T the segment length, and $\{k_\ell, C_{\ell-1}, C_\ell\}_{\ell=1}^{L}$ the 1D-convolution kernel widths and channels. Per epoch, training time scales as $\Theta\big(nT\sum_{\ell=1}^{L} k_\ell C_{\ell-1} C_\ell\big) + O(n\,d_f)$ for the fused dense work d_f, with a one-time amortized overhead from teacher PINN pretraining. Inference time is in $\Theta\big(T\sum_{\ell=1}^{L} k_\ell C_{\ell-1} C_\ell + d_f\big)$ per sample, and peak memory is $\Theta(BT\,C_{\max}) + O(q)$ (plus optional cached teacher trajectories), where $C_{\max} = maxC_l, l = 0,1,...,L$. In this case, n samples; d feature count; B batch size; p trainable parameters in PINNAC-MLP; L pretrain epochs (CE only); F finetune epochs; τ interval between physics steps; CE = cross-entropy step (supervised loss); p_{phys} parameters/ops attributable to physics heads and autograd; s stored activation size; for PINNAC-CNN, q total parameters (CNN backbone+physics heads); T segment length; k_ℓ kernel width at layer ℓ; $C_{\ell-1}, C_\ell$ in/out channels at layer ℓ; d_f fused dense compute; $C_{\max} = \max_\ell C_\ell$; $O(\cdot)$ and $\Theta(\cdot)$ denote asymptotic cost bounds.

Implementation Details. We tuned all models with Optuna[63] for 150 trials per dataset following standard tabular deep learning practice of Gorishniy et al.[64–67]. Experiments ran on PyTorch 2.4.1 + cu121 with AMP across two systems: a lab workstation (Windows, Intel CPU, 16 GB NVIDIA GPUs, 64 GB RAM, RTX 2000 Ada 16 GB) and the TITAN facility (8 clusters, 192 GB NVIDIA GPUs; nodes with x86_64 Linux, 188 GB RAM, TITAN RTX 24 GB). We used Distributed Data Parallel (DDP), which was beneficial for memory/time, and Data Parallel (DP) on the workstation. Full configs and baseline code references and PINNAC hyperparameters are in Supplementary Material.

Evaluation Metrics. We use stratified 5 fold cross validation and report mean $\pm$ std dev. for every dataset and method. To compare methods across datasets, we compute average ranks[68] over all benchmarks in each suite and highlight the top methods for further analysis. We visualize Critical Difference (CD) diagrams[69] and assess significance using the Friedman test with Nemenyi post hoc[70] and Wilcoxon signed rank tests with Holm correction[71] at $\alpha = 0.05$.

Experiment w/ Raw Dataset. We first evaluate model performance on the raw MIT–BIH dataset, without feature engineering. This setting directly tests the ability of different architectures to extract informative representations from high-dimensional signals, and performance under low data regimes (data scarcity problem).

Clean Scenario. On the raw dataset, PINNAC-MLP achieves the best overall performance, securing the top rank (1.0 ± 0.0). Accuracy remains above 99% across all splits (99.29% on the full set, 99.69% on Subset 1, 99.76% on Subset 2), with the highest F1 scores as well (97.4%, 93.7%, 83.1%). In contrast, strong boosting baselines such as XGBoost and LGBM achieve similar accuracies (99.2–99.5%) but much lower F1 on the hardest split (65.2% and 47.4%, respectively). PINNAC-CNN, however, shows the opposite profile: it ranks last in accuracy (98–99%, avg. rank 7.0 ± 1.0), yet achieves mid-tier F1 performance (95.4% on the full set, 76.5% on Subset 1, 71.6% on Subset 2), outperforming several baselines in class-balanced terms. Thus, while PINNAC-MLP dominates across both accuracy and F1, the CNN variant highlights a trade-off, underperforming in raw accuracy but offering comparatively stronger balance of precision and recall than some tree-based models.

EM 20 dB Noise. Introducing 20 dB noise does not substantially alter the hierarchy observed in the clean setting. PINNAC-MLP again secures the top rank (1.0 ± 0.0), with accuracy holding above 99% (99.24 full, 99.64 Subset 1, 99.75 Subset 2) and F1 scores remaining strongest (97.21, 90.40, 81.55). The margins relative to the leading tree–boosting baselines widen under noise: while XGBoost and LGBM sustain accuracies near 99%, their F1 scores drop more sharply. This suggests that the physics-informed inductive bias not only preserves high accuracy but also mitigates imbalance in class-wise performance. The behavior of PINNAC-CNN contrasts with the MLP but remains consistent with its performance under clean-data scenario. Accuracy places it in the lower tier (6.33 ± 0.58 average rank), yet F1 is comparatively robust, ranking second overall (2.33 ± 0.58).

EM 10 dB Noise. At higher noise, the advantage of physics-informed models becomes clearer. PINNAC-MLP remains first (1.0 ± 0.0), with accuracy near 99% and the top F1 (94.6 full, 79.5 Subset 1, 79.8 Subset 2). Unlike in the clean setting, PINNAC-CNN rises to second overall, with F1 (91.5, 70.0, 67.1) well above all boosting baselines, whose scores fall into the 50–70 range. Thus, while PINNAC-MLP continues to dominate, PINNAC-CNN shows marked robustness under stronger noise, establishing itself as the next-best alternative.

EM 5 dB Noise. At the highest noise level, the separation between physics-informed models and baselines is most evident. PINNAC-MLP sustains the top rank (1.0 ± 0.0), though both accuracy (95.1 full, 96.5 Subset 1, 97.3 Subset 2) and F1 (85.4, 70.0, 70.5) decline under heavy distortion. Importantly, PINNAC-CNN consolidates its role as the secondary model: while accuracy is lower (90.8 full, 91.4 Subset 1, 85.9 Subset 2), its F1 (74.4, 55.6, 55.4) remains well above all boosting baselines, whose scores collapse into the 25–50 range. Thus, under extreme noise, the CNN variant previously mid-tier in the clean scenario emerges as the only consistent alternative to the MLP, outperforming every baseline on class-balanced metrics. Together, the two PINNAC models demonstrate resilience, with MLP preserving accuracy and CNN providing robustness in recall-oriented evaluation where tree-based methods fail.

Table 2. Mean accuracy ($\mu \pm \sigma$) for top baselines and PINNAC variants on the raw MIT-BIH dataset under Clean and Electromagnetic (EM) noise scenarios. Each block shows results for Full, Subset 1, and Subset 2, with rankings based on cross-split generalization (Here, n denotes the number of samples).

Clear Scenario									
Dataset	n	1. PINNAC-MLP	2. XGBoost	3. LGBM	4. TabNet	5. RF	6. CatBoost	7. GBM	8. PINNAC-CNN
Full	90829	99.29 ± 0.08	99.16 ± 0.04	99.22 ± 0.03	98.98 ± 0.12	98.93 ± 0.04	98.68 ± 0.09	98.37 ± 0.05	98.72 ± 0.05
Subset 1	15426	99.69 ± 0.06	99.51 ± 0.05	99.53 ± 0.08	99.17 ± 0.24	99.33 ± 0.04	99.43 ± 0.05	99.18 ± 0.11	99.10 ± 0.19
Subset 2	19676	99.76 ± 0.03	99.50 ± 0.10	90.56 ± 11.28	99.47 ± 0.07	99.43 ± 0.12	99.44 ± 0.11	99.24 ± 0.08	97.98 ± 2.49
Avg. Rank	–	1.0 ± 0.0	2.67 ± 0.58	4.0 ± 3.46	4.67 ± 2.08	5.0 ± 0.0	5.0 ± 1.73	6.67 ± 1.15	7.0 ± 1.0
EM 20 dB Scenario									
Dataset	n	1. PINNAC-MLP	2. XGBoost	3. LGBM	4. RF	5. CatBoost	6. TabNet	7. PINNAC-CNN	8. GBM
Full	90829	99.24 ± 0.08	98.89 ± 0.06	99.02 ± 0.04	98.69 ± 0.04	98.52 ± 0.06	98.69 ± 0.12	98.64 ± 0.05	98.02 ± 0.07
Subset 1	15426	99.64 ± 0.07	99.31 ± 0.04	99.35 ± 0.10	99.05 ± 0.11	99.27 ± 0.03	98.63 ± 0.38	99.02 ± 0.23	98.82 ± 0.23
Subset 2	19676	99.75 ± 0.03	99.39 ± 0.12	89.12 ± 12.73	99.29 ± 0.13	99.26 ± 0.09	99.21 ± 0.20	97.75 ± 2.85	98.99 ± 0.21
Avg. Rank	–	1.00 ± 0.00	2.67 ± 0.58	4.00 ± 3.46	4.17 ± 1.04	5.00 ± 1.73	5.83 ± 1.89	6.33 ± 0.58	7.00 ± 1.00
EM 10 dB Scenario									
Dataset	n	1. PINNAC-MLP	2. PINNAC-CNN	3. XGBoost	4. RF	5. CatBoost	6. LGBM	7. TabNet	8. GBM
Full	90829	98.53 ± 0.13	97.43 ± 0.29	92.62 ± 0.20	92.83 ± 0.19	92.57 ± 0.37	92.62 ± 0.17	92.04 ± 0.41	87.85 ± 0.83
Subset 1	15426	99.22 ± 0.17	97.91 ± 0.54	94.96 ± 0.42	94.84 ± 0.49	94.55 ± 0.45	95.00 ± 0.55	89.01 ± 3.28	90.33 ± 1.27
Subset 2	19676	99.49 ± 0.10	94.98 ± 5.37	93.23 ± 0.38	92.61 ± 0.43	94.78 ± 0.29	82.62 ± 12.83	92.05 ± 2.12	90.13 ± 1.30
Avg. Rank	–	1.00 ± 0.00	2.00 ± 0.00	4.17 ± 0.29	4.33 ± 1.15	5.00 ± 1.73	5.17 ± 2.57	7.00 ± 1.00	7.33 ± 0.58
EM 5 dB Scenario									
Dataset	n	1. PINNAC-MLP	2. PINNAC-CNN	3. XGBoost	4. RF	5. LGBM	6. CatBoost	7. TabNet	8. GBM
Full	90829	95.13 ± 0.12	90.76 ± 0.83	85.28 ± 0.10	86.38 ± 0.16	86.82 ± 0.17	84.14 ± 0.38	81.27 ± 0.40	67.43 ± 1.57
Subset 1	15426	96.50 ± 0.26	91.41 ± 1.89	86.64 ± 1.39	84.92 ± 1.11	86.57 ± 1.40	83.87 ± 0.76	76.47 ± 3.08	72.35 ± 0.97
Subset 2	19676	97.30 ± 0.55	85.86 ± 9.08	84.46 ± 0.85	82.90 ± 0.39	75.50 ± 11.75	84.20 ± 0.64	78.88 ± 2.01	73.60 ± 1.98
Avg. Rank	–	1.00 ± 0.00	2.00 ± 0.00	3.67 ± 1.15	4.67 ± 0.58	4.67 ± 2.08	5.33 ± 1.15	6.67 ± 0.58	8.00 ± 0.00

Class Imbalance and Binary Case Construction. The original dataset was highly imbalanced: class N contained 90083 samples, while VEB, SVEB, F, and Q had 7009, 2779, 803, and 15 samples, respectively. To mitigate bias and stabilize training, we reframed the task into five balanced binary classification problems, each with equal class counts whenever possible: Case 1 (N vs VEB+SVEB+F+Q; $n = 21212$), Case 2 (VEB vs SVEB+F+Q+N; $n = 14018$), Case 3 (SVEB vs F+Q+N+VEB; $n = 5558$), Case 4 (F vs Q+VEB+SVEB+N; $n = 1606$), and Case 5 (Q vs N+VEB+SVEB+F; $n = 30$). This design reduces dominance of the majority class, simplifies decision boundaries for each subproblem, and provides a clearer view of model behavior across clinically relevant contrasts.

Experiment w/ Feature Extracted Dataset. This section reports results on the Feature Extracted Dataset of MIT-BIH. In the regular (clean) setting, XGBoost attains the best cross-dataset average rank (1.83 ± 0.75), narrowly ahead of CatBoost

Table 3. Mean F1 ($\mu \pm \sigma$) for top baselines and PINNAC variants on the raw MIT-BIH dataset under Clean and Electromagnetic (EM) noise scenarios. Each block shows results for Full, Subset 1, and Subset 2, with rankings based on cross-split generalization (Here, n denotes feature dimensionality).

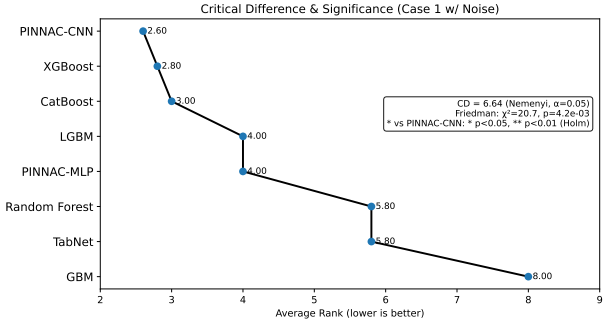
Clear Scenario									
Dataset	n	1. PINNAC-MLP	2. XGBoost	3. PINNAC-CNN	4. LGBM	5. TabNet	6. CatBoost	7. RF	8. GBM
Full	90829	97.37 ± 0.26	96.12 ± 0.31	95.38 ± 0.07	96.31 ± 0.21	95.47 ± 0.56	93.90 ± 0.53	94.94 ± 0.30	92.59 ± 0.29
Subset 1	15426	93.68 ± 3.17	81.14 ± 3.81	76.48 ± 7.09	81.11 ± 7.97	73.51 ± 11.92	73.47 ± 4.49	70.83 ± 2.86	72.29 ± 2.66
Subset 2	19676	83.05 ± 7.62	65.22 ± 0.48	71.55 ± 4.15	47.36 ± 23.29	65.14 ± 0.31	65.10 ± 0.44	64.93 ± 0.52	64.46 ± 0.29
Avg. Rank	–	1.00 ± 0.00	2.67 ± 0.58	3.67 ± 1.53	4.33 ± 3.21	4.33 ± 0.58	6.0 ± 1.00	6.67 ± 1.15	7.33 ± 0.58
EM 20 dB Scenario									
Dataset	n	1. PINNAC-MLP	2. PINNAC-CNN	3. XGBoost	4. LGBM	5. TabNet	6. CatBoost	7. RF	8. GBM
Full	90829	97.21 ± 0.29	95.06 ± 0.11	94.92 ± 0.34	95.39 ± 0.25	94.16 ± 0.54	93.29 ± 0.46	93.94 ± 0.31	91.29 ± 0.42
Subset 1	15426	90.40 ± 4.50	74.94 ± 4.59	72.49 ± 6.09	72.64 ± 1.91	71.97 ± 8.79	70.31 ± 4.83	64.26 ± 0.30	69.85 ± 2.98
Subset 2	19676	81.55 ± 5.74	71.17 ± 4.45	65.00 ± 0.53	46.44 ± 23.79	64.15 ± 0.86	64.43 ± 0.34	64.50 ± 0.65	63.82 ± 0.77
Avg. Rank	–	1.00 ± 0.00	2.33 ± 0.58	3.67 ± 0.58	4.33 ± 0.21	5.33 ± 0.58	6.00 ± 1.00	6.00 ± 2.00	7.33 ± 0.58
EM 10 dB Scenario									
Dataset	n	1. PINNAC-MLP	2. PINNAC-CNN	3. RF	4. CatBoost	5. XGBoost	6. LGBM	7. TabNet	8. GBM
Full	90829	94.59 ± 0.37	91.52 ± 0.88	74.57 ± 0.72	72.79 ± 1.29	72.56 ± 0.59	72.46 ± 0.78	72.38 ± 2.39	63.22 ± 1.54
Subset 1	15426	79.45 ± 4.53	70.00 ± 2.05	54.75 ± 2.83	52.48 ± 1.66	54.46 ± 1.24	55.74 ± 2.86	44.05 ± 10.08	50.50 ± 1.27
Subset 2	19676	79.82 ± 7.99	67.11 ± 6.08	50.07 ± 1.24	52.12 ± 0.76	50.98 ± 0.94	37.02 ± 17.96	46.40 ± 3.83	45.43 ± 2.06
Avg. Rank	–	1.00 ± 0.00	2.00 ± 0.00	4.00 ± 1.00	4.33 ± 1.53	4.67 ± 0.58	5.67 ± 2.52	7.00 ± 1.00	7.33 ± 0.58
EM 5 dB Scenario									
Dataset	n	1. PINNAC-MLP	2. PINNAC-CNN	3. XGBoost	4. RF	5. LGBM	6. CatBoost	7. GBM	8. TabNet
Full	90829	85.41 ± 0.33	74.41 ± 2.27	48.30 ± 0.44	49.07 ± 1.11	49.67 ± 0.75	46.99 ± 0.87	36.56 ± 0.77	45.10 ± 3.46
Subset 1	15426	70.00 ± 2.51	55.63 ± 2.96	36.42 ± 2.62	31.38 ± 2.79	34.86 ± 3.30	29.60 ± 0.74	30.83 ± 1.04	24.91 ± 3.97
Subset 2	19676	70.54 ± 2.23	55.37 ± 5.91	34.71 ± 1.72	31.65 ± 0.72	26.36 ± 11.02	34.28 ± 0.60	28.36 ± 1.45	31.09 ± 2.99
Avg. Rank	–	1.00 ± 0.00	2.00 ± 0.00	3.67 ± 1.15	4.67 ± 0.58	5.00 ± 2.65	5.67 ± 1.53	7.00 ± 1.00	7.00 ± 1.00

(2.00 ± 1.26). RF (3.67 ± 1.21) and LGBM (3.83 ± 2.93) form the next tier, while PINNAC-MLP is competitive but trails the top tree-boosting methods (4.50 ± 1.38). Consistent with this ranking, accuracies on MIT-BIH proper are ≈99% for the strongest models, and the case splits remain in the high-90s except for the tiny Case 5. Under additive Gaussian noise (0.1-0.8), the ordering shifts in favor of the physics-informed model: PINNAC-MLP achieves the lowest average rank (1.60 ± 0.89), followed by CatBoost (2.00 ± 0.71) and XGBoost (2.20 ± 0.84). Despite increasing noise, top models degrade gracefully; even at 0.8 noise, PINNAC-MLP remains at 96.02% while CatBoost and XGBoost stay mid-95%, indicating strong robustness for the PINN-guided approach and overall resilience of tree-boosting baselines. In Fig. 2f, critical-difference (Nemenyi) rank plot on the feature-extracted MIT–BIH under additive noise (0.1-0.8) illustrates that average ranks (lower is better) place PINNAC-MLP first (1.60), then CatBoost (2.00) and XGBoost (2.20); the Friedman test is significant ($\chi^2 = 32.7$, $p = 3.0 \times 10^{-5}$) with Nemenyi CD = 6.64 pairwise gaps beyond CD are significant.

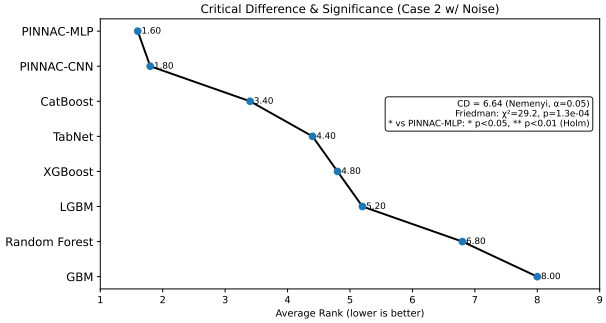
Table 4. Mean accuracy ($\mu \pm \sigma$) for top PINNAC baselines under Regular and Noisy scenarios for feature-extracted MIT-BIH. Gaussian noise levels of 0.1–0.8 are applied in the noisy case. Rankings are based on cross-dataset performance within each scenario (Here, n = dimensionality). See Table S3 and S4 in the supplementary material for full results.

Regular Scenario									
Dataset	n	1. XGBoost	2. CatBoost	3. RF	4. LGBM	5. PINNAC-MLP	6. TabNet	7. GBM	8. PINNAC-CNN
MIT-BIH	100689	99.29 ± 0.03	99.22 ± 0.02	99.03 ± 0.04	98.05 ± 0.20	98.84 ± 0.06	98.96 ± 0.24	98.46 ± 0.07	96.32 ± 0.24
Case 1	21212	98.51 ± 0.24	98.29 ± 0.21	98.18 ± 0.24	98.42 ± 0.21	97.80 ± 0.37	97.28 ± 0.20	96.58 ± 0.30	96.86 ± 0.24
Case 2	14018	98.45 ± 0.13	98.50 ± 0.10	98.02 ± 0.17	98.41 ± 0.18	98.08 ± 0.27	97.32 ± 0.26	97.28 ± 0.41	97.56 ± 0.14
Case 3	5558	97.57 ± 0.50	97.19 ± 0.54	97.30 ± 0.25	97.63 ± 0.61	96.64 ± 0.45	95.66 ± 0.34	96.29 ± 0.40	94.48 ± 0.68
Case 4	1606	96.14 ± 0.70	96.58 ± 0.56	95.14 ± 0.90	96.51 ± 0.64	94.46 ± 1.07	94.42 ± 1.28	95.33 ± 0.56	92.68 ± 0.76
Case 5	30	73.33 ± 13.33	76.67 ± 8.16	73.33 ± 24.94	50.00 ± 0.00	73.33 ± 8.16	66.67 ± 13.33	66.67 ± 10.54	73.33 ± 16.33
Avg. Rank	–	1.83 ± 0.75	2.00 ± 1.26	3.67 ± 1.21	3.83 ± 2.93	4.50 ± 1.38	6.17 ± 1.17	6.33 ± 1.51	6.50 ± 2.35
Noisy Scenario									
Dataset	Noise	1. PINNAC-MLP	2. CatBoost	3. XGBoost	4. RF	5. GBM	6. LGBM	7. TabNet	8. PINNAC-CNN
MIT-BIH	0.1	99.02 ± 0.02	99.02 ± 0.01	99.06 ± 0.01	98.70 ± 0.02	97.86 ± 0.04	97.35 ± 0.28	97.24 ± 0.24	96.22 ± 0.24
MIT-BIH	0.2	98.34 ± 0.04	98.76 ± 0.05	98.73 ± 0.01	98.21 ± 0.04	97.43 ± 0.15	97.36 ± 0.50	96.68 ± 0.14	95.78 ± 0.22
MIT-BIH	0.4	97.28 ± 0.10	97.20 ± 0.05	97.19 ± 0.07	96.77 ± 0.05	96.23 ± 0.06	96.66 ± 0.43	95.68 ± 0.68	94.48 ± 0.25
MIT-BIH	0.6	96.18 ± 0.12	96.10 ± 0.09	96.11 ± 0.12	95.37 ± 0.06	94.99 ± 0.08	94.63 ± 1.00	94.68 ± 0.10	93.76 ± 0.28
MIT-BIH	0.8	96.02 ± 0.06	95.44 ± 0.05	95.25 ± 0.06	94.49 ± 0.04	94.27 ± 0.07	94.38 ± 0.17	94.12 ± 0.14	92.94 ± 0.38
Avg. Rank	–	1.60 ± 0.89	2.00 ± 0.71	2.20 ± 0.84	4.00 ± 0.00	5.40 ± 0.55	5.80 ± 0.84	6.80 ± 0.45	8.00 ± 0.00

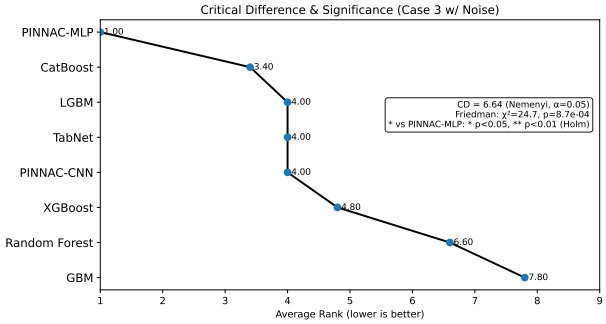
Cases w/ Noises and Statistical Significance. Across the five noisy cases, PINNAC variants dominate average ranks. Case 1 (Fig. 2a) is led by PINNAC-CNN (avg. rank ≈2.6), with XGBoost and CatBoost close behind; the Friedman test is significant ($\chi^2 = 20.7$, $p = 4.2 \times 10^{-3}$) but no Nemenyi pairwise difference exceeds the CD (CD = 6.64). Case 2 (Fig. 2b) is



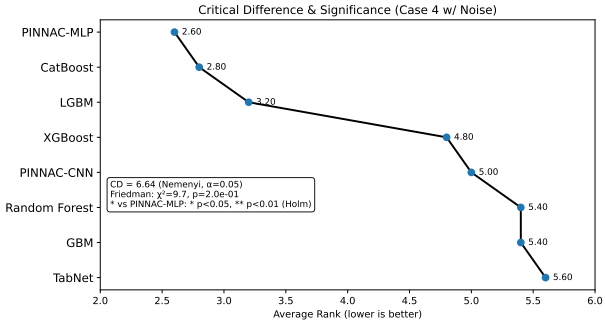
(a) Case 1 w/ Noise



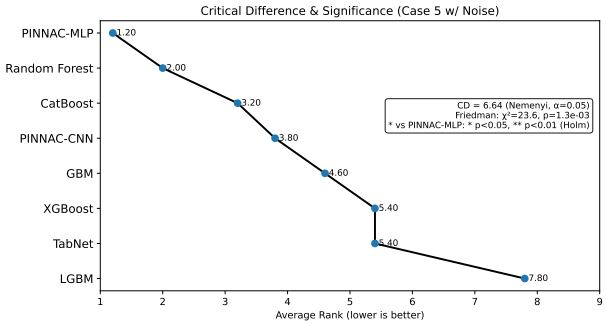
(b) Case 2 w/ Noise



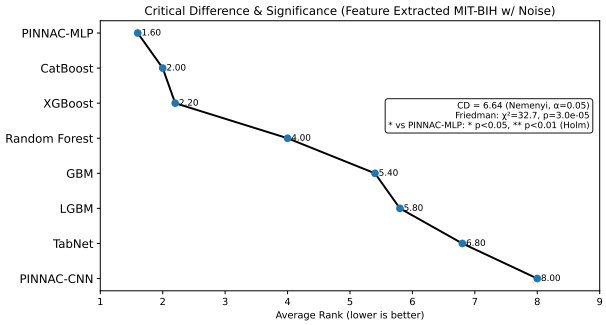
(c) Case 3 w/ Noise



(d) Case 4 w/ Noise



(e) Case 5 w/ Noise



(f) Feature Extracted MIT-BIH w/ Noise

Figure 2. Critical-difference (CD) rank plots (Nemenyi) for the five cases and feature extracted MIT-BIH with noises. Points show average ranks (lower is better). Each panel reports the Nemenyi critical difference and the Friedman test; pairwise gaps exceeding the CD are significant at α=0.05 (Holm-adjusted notes in panels). See Table S4-S9 in the supplementary material for full results.

topped by PINNAC-MLP (≈ 1.6) followed by PINNAC-CNN (≈ 1.8) and CatBoost; the omnibus test is significant ($\chi^2 = 29.2$, $p = 1.3\times10^{-4}$), yet post-hoc gaps remain below the CD. In Case 3 (Fig. 2c), PINNAC-MLP leads clearly (≈ 1.0), ahead of CatBoost and the mid-tier trio (LGBM/TabNet/PINNAC-CNN at ≈ 4.0); the Friedman test is significant ($\chi^2 = 24.7$, $p = 8.7\times10^{-4}$), and only the contrast to GBM marginally exceeds the Nemenyi CD, while others do not. Case 4 (Fig. 2d) again favors PINNAC-MLP (≈ 2.6), then CatBoost and LGBM, but the omnibus test is not significant at α=0.05 ($\chi^2 = 9.7$, $p = 0.20$). Case 5 (Fig. 2e) shows PINNAC-MLP first (≈ 1.2), with RF and CatBoost next; the Friedman test is significant ($\chi^2 = 23.6$, $p = 1.3\times10^{-3}$), yet pairwise gaps are below the CD. Overall, PINNAC-MLP is best in four of five noisy cases (2-5) and PINNAC-CNN wins Case 1, indicating that physics-informed training yields robust rankings as noise grows, while tree-boosted baselines remain competitive but rarely surpass PINNAC.

Ablation Studies. To assess the contribution of individual physics-informed components, we perform an ablation study by comparing models trained with single PINN constraints (FHN or AP) against the combined configuration used in PINNAC. The goal is to quantify how each physical prior influences model performance across signal-to-noise conditions. Table 5 and Table 6 show the ablation results. Across most scenarios (Clean, EM 20 dB, EM 10 dB, and EM 5 dB), both FHN and AP-constrained

models consistently improve over their non-physics baselines (MLP and CNN). This demonstrates that even a single governing equation can guide the model toward smoother, more physiologically consistent representations. For instance, under the Clean scenario on the Subset 1 dataset, FHN-MLP and AP-MLP improve accuracy from 98.38% to 99.33% and 99.24%, respectively, with similar trends observed for CNN-based counterparts. The combined PINNAC variant, which integrates both FHN and AP priors within a unified constraint layer, yields the highest overall performance across all noise levels. Under EM 10 dB and EM 5 dB conditions, where signal degradation is most severe, the joint model preserves notable accuracy improvements and F1 gains relative to single PINN versions. This highlights that the two physiological formulations capture complementary aspects of signal dynamics, with FHN emphasizing excitable membrane behavior and AP modeling restitution and recovery which together enhance both predictive stability and robustness to noise. In summary, single PINN models provide meaningful physical regularization, but the combined configuration achieves superior generalization under both clean and noisy regimes. This validates the design choice of fusing multiple physics-based constraints within the PINNAC framework.

Table 5. Mean accuracy ($\mu \pm \sigma$) for ablation variants on the raw MIT-BIH dataset under Clean and EM noise scenarios. Results compare non-physics baselines (MLP, CNN), single-constraint models (FHN, AP), and the combined PINNAC configuration across Full, Subset 1, and Subset 2 splits.

Dataset	MLP	FHN-MLP	AP-MLP	PINNAC-MLP	CNN	FHN-CNN	AP-CNN	PINNAC-CNN
Clean								
Full	98.33 ± 0.19	98.38 ± 0.33	98.25 ± 0.17	99.29 ± 0.08	96.40 ± 0.72	96.52 ± 2.19	96.91 ± 0.61	98.72 ± 0.05
Subset 1	98.38 ± 0.30	99.33 ± 0.07	99.24 ± 0.35	99.69 ± 0.06	93.82 ± 7.26	98.17 ± 1.30	97.71 ± 1.27	99.10 ± 0.19
Subset 2	98.75 ± 0.33	99.49 ± 0.21	99.52 ± 0.10	99.76 ± 0.03	94.08 ± 4.66	98.36 ± 0.69	98.11 ± 1.09	97.98 ± 2.49
EM 20 dB								
Full	98.18 ± 0.19	98.25 ± 0.36	98.10 ± 0.23	99.24 ± 0.08	96.05 ± 0.76	96.05 ± 2.54	96.50 ± 0.72	98.64 ± 0.05
Subset 1	97.77 ± 0.77	99.20 ± 0.10	99.16 ± 0.43	99.64 ± 0.07	90.14 ± 8.85	98.01 ± 1.47	97.52 ± 1.43	99.02 ± 0.23
Subset 2	98.61 ± 0.38	99.42 ± 0.28	99.49 ± 0.11	99.75 ± 0.03	93.22 ± 5.32	98.16 ± 0.90	97.86 ± 1.26	97.75 ± 2.85
EM 10 dB								
Full	94.91 ± 0.62	95.74 ± 0.68	94.72 ± 0.62	98.53 ± 0.13	91.57 ± 1.32	90.26 ± 6.27	91.09 ± 2.56	97.43 ± 0.29
Subset 1	94.46 ± 1.70	97.70 ± 0.39	97.88 ± 0.80	99.22 ± 0.17	87.75 ± 9.16	95.83 ± 3.42	94.08 ± 3.27	97.91 ± 0.54
Subset 2	97.23 ± 1.04	98.88 ± 0.41	98.92 ± 0.21	99.49 ± 0.10	87.18 ± 7.62	95.12 ± 3.39	94.66 ± 2.62	94.98 ± 5.37
EM 5 dB								
Full	87.48 ± 1.45	88.95 ± 0.56	86.65 ± 1.01	95.13 ± 0.12	77.75 ± 2.50	75.84 ± 11.12	76.87 ± 4.59	90.76 ± 0.83
Subset 1	90.52 ± 2.13	94.00 ± 1.36	94.55 ± 0.76	96.50 ± 0.26	80.98 ± 6.49	85.32 ± 10.71	82.98 ± 7.26	91.41 ± 1.89
Subset 2	91.70 ± 1.92	95.34 ± 1.41	96.05 ± 0.63	97.30 ± 0.55	74.81 ± 8.94	84.22 ± 5.91	83.42 ± 3.37	85.86 ± 9.08

Table 6. Mean F1 ($\mu \pm \sigma$) for ablation variants on the raw MIT-BIH dataset under Clean and EM noise scenarios. Results compare non-physics baselines (MLP, CNN), single-constraint models (FHN, AP), and the combined PINNAC configuration across Full, Subset 1, and Subset 2 splits.

Dataset	MLP	FHN-MLP	AP-MLP	PINNAC-MLP	CNN	FHN-CNN	AP-CNN	PINNAC-CNN
Clean								
Full	92.83 ± 0.85	93.09 ± 1.45	92.65 ± 0.77	97.37 ± 0.26	86.40 ± 2.00	87.84 ± 5.28	88.41 ± 1.58	95.38 ± 0.07
Subset 1	75.34 ± 2.77	86.31 ± 1.89	88.25 ± 4.76	93.68 ± 3.17	65.47 ± 4.29	72.64 ± 6.90	66.39 ± 5.31	76.48 ± 7.09
Subset 2	67.24 ± 1.06	76.19 ± 8.46	71.69 ± 7.87	83.05 ± 7.62	56.54 ± 10.44	63.66 ± 1.01	61.96 ± 2.82	71.55 ± 4.15
EM 20 dB								
Full	92.25 ± 0.88	92.49 ± 1.52	92.03 ± 0.92	97.21 ± 0.29	85.47 ± 1.88	86.52 ± 5.82	87.20 ± 1.81	95.06 ± 0.11
Subset 1	72.21 ± 2.41	82.40 ± 1.19	86.54 ± 6.01	90.40 ± 4.50	63.18 ± 2.45	70.98 ± 7.65	64.85 ± 4.39	74.94 ± 4.59
Subset 2	66.86 ± 0.93	74.23 ± 6.85	68.33 ± 6.22	81.55 ± 5.74	55.20 ± 9.19	63.46 ± 1.10	61.84 ± 3.26	71.17 ± 4.45
EM 10 dB								
Full	82.98 ± 1.26	84.83 ± 1.57	82.48 ± 1.25	94.95 ± 0.37	76.41 ± 1.75	75.31 ± 9.61	75.78 ± 4.23	91.52 ± 0.88
Subset 1	64.38 ± 1.79	70.28 ± 1.81	71.50 ± 3.48	79.45 ± 4.53	57.68 ± 3.25	60.15 ± 6.94	55.90 ± 7.07	70.00 ± 2.05
Subset 2	63.42 ± 1.38	68.01 ± 3.49	65.75 ± 2.70	79.82 ± 7.99	48.13 ± 6.19	57.58 ± 3.99	55.91 ± 4.21	67.11 ± 6.08
EM 5 dB								
Full	69.34 ± 1.57	70.69 ± 0.28	68.14 ± 0.83	85.41 ± 0.33	55.59 ± 3.27	52.34 ± 10.04	54.12 ± 3.63	74.41 ± 2.27
Subset 1	54.56 ± 1.45	58.20 ± 3.41	58.87 ± 1.52	70.00 ± 2.51	41.29 ± 5.59	37.83 ± 12.96	35.51 ± 12.29	55.63 ± 2.96
Subset 2	53.25 ± 2.31	58.63 ± 3.24	57.65 ± 2.32	70.54 ± 2.23	34.18 ± 3.64	42.33 ± 7.38	39.37 ± 2.72	55.37 ± 5.91

Weight Sensitivity Analysis. To obtain the final PINNAC model, we experimented with different weighting configurations $(\lambda_{\text{FHN}}, \lambda_{\text{AP}})$ for combining the FHN and AP components within the physics-informed network. These weights control the relative contribution of each physiological constraint to the joint optimization objective. Table 7 summarizes the resulting Macro-F1 scores under varying EM noise conditions, showing that moderate weight ratios yield consistently strong generalization

and robustness across all scenarios. Among all configurations, the pair $(\lambda_{\text{FHN}}, \lambda_{\text{AP}}) = (0.03, 0.03)$ yielded the best overall performance across clean and noisy conditions.

Table 7. Macro-F1 (%) across EM noise levels for different $(\lambda_{\text{FHN}}, \lambda_{\text{AP}})$ weight pairs.

Weight Pair $(\lambda_{\text{FHN}}, \lambda_{\text{AP}})$	**Clean**	**EM 20 dB**	**EM 10 dB**	**EM 5 dB**
(0.03, 0.03)	97.37	97.21	94.95	85.41
(0.05, 0.05)	97.30	97.19	94.81	85.11
(0.03, 0.02)	97.29	97.12	94.88	84.98
(0.30, 0.10)	97.27	97.15	94.60	84.59
(0.08, 0.01)	97.25	97.09	94.66	84.94

Discussion

This study introduces PINNAC, a supervised classifier that couples standard neural backbones with compact physics-informed excitable cell surrogates for robust arrhythmia classification. The central finding is that explicitly encoding electrophysiology in the training objective confers three practical benefits for arrhythmia recognition from ECG signals. First, physics consistency acts as an inductive bias that improves generalization under distribution shift, particularly when signals are corrupted or labels are scarce. Second, the model exposes interpretable latent state trajectories in a phase space that align with membrane excitation and recovery. Third, the approach remains lightweight and end to end, which facilitates reproducibility and integration with existing clinical machine learning toolchains. Across both raw waveform and feature extracted settings, PINNAC matches or exceeds the accuracy of strong baselines on clean data, while delivering larger gains in macro F1 for minority arrhythmia classes under noise. These gains are most pronounced when electromyographic and related artifacts degrade morphology, where purely data driven learners tend to preserve accuracy through dominant class predictions at the expense of recall on ectopic beats. By contrast, enforcing FHN and AP residuals or distilling from physics trajectories regularizes the temporal embedding toward plausible excitation and recovery, which stabilizes decision boundaries and preserves class balance. The ablation study supports this interpretation. Each surrogate on its own improves robustness relative to plain MLP and CNN backbones, and their joint use yields the strongest average ranks across noise conditions. This suggests that the two surrogates capture complementary aspects of the ECG signal under arrhythmia, with FHN emphasizing fast-slow timing and refractoriness, and AP emphasizing stimulus linked morphology and recovery coupling.

The comparison between PINNAC-MLP and PINNAC-CNN highlights an additional practical point. On clean waveforms, the MLP variant tends to dominate in both accuracy and macro F1, consistent with the relative ease of fitting well segmented beats in a compact feature space. Under moderate to severe corruption, however, the CNN variant becomes comparatively more competitive on class balanced metrics. This pattern reflects different error modes. Convolutional encoders retain more local morphological information that can aid recall when the feature space is distorted by noise, while the MLP variant, which relies on a tabular or compressed embedding, benefits more from physics penalties that prevent collapse toward majority classes. In practice, these profiles are complementary. An ensemble or a backend selection policy can trade a small loss of accuracy on clean datasets for improved robustness in deployment settings where artifacts are common. A key advantage of the proposed design is the availability of physics driven latent summaries that support error analysis. The physics heads output latent variables that can be inspected alongside ECG to probe failure cases and to relate predictions to mechanistic hypotheses. For instance, when the model misclassifies supraventricular ectopy, the latent trajectory often reveals an exaggerated recovery loop or a blunted upstroke that mirrors attenuation in the input. Such visual diagnostics are not a substitute for model explanation in the strict sense, but they provide an intuitive handle for clinicians and engineers to reason about model behavior and to compare predictions across devices, preprocessing pipelines, and cohorts. They also enable simple sanity checks. When trajectories depart strongly from the feasible manifold learned by the surrogates, performance typically degrades, which signals that upstream segmentation or denoising may need attention.

Although PINNAC shows improved performance in robust classification of arrhythmia under scenarios with noise, or with limited data, there are still several limitations, and hence room for improvement. First, the surrogates are ordinary differential equations that approximate single cell dynamics. They do not model spatial propagation, restitution heterogeneity, or lead specific projections. This choice is deliberate, as it keeps training stable and fast, but it also constrains the scope of the physics prior. Extending the framework with low rank operator constraints or simplified reaction diffusion terms may capture propagation effects without incurring the full cost of partial differential equation solvers. Second, the evaluation relies on widely used public datasets, which are known to exhibit cohort specific biases and protocol artifacts. Although we test with controlled noise and class balanced splits, broader external validation with data from an independent health system and multi lead recordings is required before clinical deployment. Third, while the confidence weighting and warm up schedule

mitigate loss term imbalance, PINN objectives can still be sensitive to hyperparameters and to the sampling of collocation times. More principled multi-objective optimization or self-adaptive balancing strategies could reduce tuning effort and improve convergence on difficult regimes. The design also raises questions about calibration and uncertainty. Physics consistency and distillation from surrogate trajectories should, in principle, produce better calibrated probabilities under shift. Our preliminary analyses suggest improved calibration relative to plain backbones in noisy settings, or under data scarcity, but a comprehensive study with temperature scaling, conformal risk control, and shift aware reliability diagrams is warranted. Relatedly, integrating uncertainty estimates from the physics residuals into the decision policy could provide a natural abstention mechanism for out of distribution beats or segments with severe artifact burden.

Future work will explore three directions. First, multi-lead and longer context modeling to capture spatial and temporal dependencies that influence arrhythmia morphology. Second, personalization through hierarchical priors or meta learning, allowing subject specific adjustments to excitability and recovery parameters without full retraining. Third, unified pretraining that blends self-supervised objectives on large unlabeled ECG datasets with physics-guided consistency, followed by task-specific finetuning. We anticipate that these extensions will further improve robustness, reduce data requirements, and enhance diagnostic utility. In summary, PINNAC demonstrates that compact electrophysiology priors can be combined with modern neural encoders to produce accurate, noise robust, and diagnostically useful arrhythmia classifiers. The approach aligns with a broader trend toward hybrid modeling in deploying machine learning in the clinic, where domain knowledge is used not to replace learning but to shape it. With rigorous external validation, careful calibration, and prospective testing in workflow, physics-informed ECG analysis has the potential to improve medical decision-making, screening reliability, and clinician trust in real world settings.

References

1. Ansari, Y. *et al.* Deep Learning for ECG Arrhythmia Detection and Classification: An Overview of Progress for Period 2017–2023. *Frontiers in Physiology* **14**, 1246746 (2023).
2. Garg, K. o. Global Health Inequities in Electrophysiology Care: A State-of-the-Art Review. *JACC: Advances* **3**, 101387 (2024).
3. Linz, D. *et al.* Atrial Fibrillation: Epidemiology, Screening and Digital Health. *The Lancet Regional Health–Europe* **37** (2024).
4. Hayashi, M. *et al.* The Spectrum of Epidemiology Underlying Sudden Cardiac Death. *Circulation Research* **116**, 1887–1906 (2015).
5. Majumder, N. *et al.* Aerosol Physicochemical Determinants of Carbon Black and Ozone Inhalation Co-Exposure Induced Pulmonary Toxicity. *Toxicological Sciences* **191**, 61–78 (2023).
6. Majumder, N. *et al.* Oxidant-Induced Epithelial Alarmin Pathway Mediates Lung Inflammation and Functional Decline Following Ultrafine Carbon and Ozone Inhalation Co-Exposure. *Redox Biology* **46**, 102092 (2021).
7. Mohonta, S. C. *et al.* Electrocardiogram Based Arrhythmia Classification Using Wavelet Transform with Deep Learning Model. *Sensing and Bio-Sensing Research* **37**, 100502 (2022).
8. Martis, R. J. *et al.* ECG Beat Classification Using PCA, LDA, ICA and Discrete Wavelet Transform. *Biomedical Signal Processing and Control* **8**, 437–448 (2013).
9. Liu, N. *et al.* A Support Vector Machine Approach for AF Classification from A Short Single-Lead ECG Recording. *Physiological Measurement* **39**, 064004 (2018).
10. Dai, H. *et al.* QRS Residual Removal in Atrial Activity Signals Extracted from Single Lead: A New Perspective Based on Signal Extrapolation. *IET Signal Processing* **10**, 1169–1175 (2016).
11. Asl, B. M. *et al.* Support Vector Machine-Based Arrhythmia Classification Using Reduced Features of Heart Rate Variability Signal. *Artificial Intelligence in Medicine* **44**, 51–64 (2008).
12. Li, Q. *et al.* Ventricular Fibrillation and Tachycardia Classification Using A Machine Learning Approach. *IEEE Transactions on Biomedical Engineering* **61**, 1607–1613 (2013).
13. Tsipouras, M. G. *et al.* An Arrhythmia Classification System Based on the RR-Interval Signal. *Artificial Intelligence in Medicine* **33**, 237–250 (2005).
14. Kim, J. W. *et al.* Automatic Parameter Acquisition of 12 Leads ECG Using Continuous Data Processing Deep Neural Network. *Journal of Biomedical Engineering Research* **41**, 107–119 (2020).
15. Feyisa, D. W. *et al.* Lightweight Multireceptive Field CNN for 12-Lead ECG Signal Classification. *Computational Intelligence and Neuroscience* **2022**, 8413294 (2022).

16. Sampath, A. & Sumithira, T. Sparse Based Recurrent Neural Network Long Short Term Memory (RNN-LSTM) Model for the Classification of ECG Signals. *Applied Artificial Intelligence* **36** (2022).

17. FitzHugh, R. Impulses and Physiological States in Theoretical Models of Nerve Membrane. *Biophysical Journal* **1**, 445–466 (1961).

18. Nagumo, J. *et al.* An Active Pulse Transmission Line Simulating Nerve Axon. *Proceedings of the IRE* **50**, 2061–2070 (2007).

19. Aliev, R. R. & Panfilov, A. V. A Simple Two-Variable Model of Cardiac Excitation. *Chaos, Solitons @AND@ Fractals* **7**, 293–301 (1996).

20. Gani, M. O. & Ogawa, T. Stability of Periodic Traveling Waves in the Aliev-Panfilov Reaction-Diffusion System. *Communications in Nonlinear Science and Numerical Simulation* **33**, 30–42 (2016).

21. Moody, G. MIT-BIH Arrhythmia Database v1.0.0 (2005). [Online; accessed 2025-10-07].

22. Moody, G. MIT-BIH Noise Stress Test Database v1.0.0 (1999). [Online; accessed 2025-10-07].

23. Habib, A.-Z. S. B. *et al.* A Study on Coronary Disease Prediction Using Boosting-Based Ensemble Machine Learning Approaches. In *2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*, 1–6 (IEEE, 2019).

24. Habib, A.-Z. S. B. & Tasnim, T. An Ensemble Hard Voting Model for Cardiovascular Disease Prediction. In *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, 1–6 (IEEE, 2020).

25. Reshad, A. I. *et al.* Deep Learning-Based Detection of Arrhythmia Using ECG Signals–A Comprehensive Review. *Vascular Health and Risk Management* 685–703 (2025).

26. Li, Y. *et al.* Inter-Patient Arrhythmia Classification with Improved Deep Residual Convolutional Neural Network. *Computer Methods and Programs in Biomedicine* **214**, 106582 (2022).

27. Sun, A. *et al.* An Arrhythmia Classification Model Based on a CNN-LSTM-SE Algorithm. *Sensors* **24**, 6306 (2024).

28. El-Ghaish, H. & Eldele, E. ECGTransForm: Empowering Adaptive ECG Arrhythmia Classification Framework with Bidirectional Transformer. *Biomedical Signal Processing and Control* **89**, 105714 (2024).

29. Kim, D. *et al.* A Novel Hybrid CNN-Transformer Model for Arrhythmia Detection without R-peak Identification using Stockwell Transform. *Scientific Reports* **15**, 7817 (2025).

30. Chen, W. *et al.* Temporal and Spatial Self Supervised Learning Methods for Electrocardiograms. *Scientific Reports* **15**, 6029 (2025).

31. Le, D., Truong, S., Brijesh, P., Adjeroh, D. A. & Le, N. scl-st: Supervised contrastive learning with semantic transformations for multiple lead ecg arrhythmia classification. *IEEE J. Biomed. Heal. Informatics* **27**, 2818–2828 (2023).

32. Bui, N.-T. *et al.* Tsrnet: Simple framework for real-time ecg anomaly detection with multimodal time and spectrogram restoration network. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 1–4 (IEEE, 2024).

33. Phan, T. *et al.* Multimodality multi-lead ecg arrhythmia classification using self-supervised learning. In *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 01–04 (IEEE, 2022).

34. Devkota, A. *et al.* AI Analysis for Ejection Fraction Estimation from 12-Lead ECG. *Scientific Reports* **15**, 13502 (2025).

35. Krishnapriyan, A. *et al.* Characterizing Possible Failure Modes in Physics-Informed Neural Networks. *Advances in Neural Information Processing Systems* **34**, 26548–26560 (2021).

36. Xiang, Z. *et al.* Self-Adaptive Loss Balanced Physics-Informed Neural Networks. *Neurocomputing* **496**, 11–34 (2022).

37. Wang, C. *et al.* Is L^2 Physics Informed Loss Always Suitable for Training Physics Informed Neural Network? *Advances in Neural Information Processing Systems* **35**, 8278–8290 (2022).

38. Hwang, Y. & Lim, D. Dual Cone Gradient Descent for Training Physics-Informed Neural Networks. *Advances in Neural Information Processing Systems* **37**, 98563–98595 (2024).

39. Wu, H. *et al.* RoPINN: Region Optimized Physics-Informed Neural Networks. *Advances in Neural Information Processing Systems* **37**, 110494–110532 (2024).

40. Zhao, Z., Ding, X. & Prakash, B. A. PINNsFormer: A Transformer-Based Framework For Physics-Informed Neural Networks. In *The Twelfth International Conference on Learning Representations* (2024).

41. Li, Z. *et al.* Physics-Informed Neural Operator for Learning Partial Differential Equations. *ACM/IMS Journal of Data Science* **1**, 1–27 (2024).

42. Cho, W. *et al.* Parameterized Physics-informed Neural Networks for Parameterized PDEs. In *International Conference on Machine Learning*, 8510–8533 (PMLR, 2024).

43. Rathore, P. *et al.* Challenges in Training PINNs: A Loss Landscape Perspective. In *International Conference on Machine Learning*, 42159–42191 (PMLR, 2024).

44. Sahli Costabal, F. *et al.* Physics-Informed Neural Networks for Cardiac Activation Mapping. *Frontiers in Physics* **8**, 42 (2020).

45. Herrero Martin, C. *et al.* EP-PINNs: Cardiac Electrophysiology Characterisation Using Physics-Informed Neural Networks. *Frontiers in Cardiovascular Medicine* **8**, 768419 (2022).

46. Chiu, C.-E. *et al.* Physics-Informed Neural Networks Can Accurately Model Cardiac Electrophysiology in 3D Geometries and Fibrillatory Conditions. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, 98–109 (Springer, 2024).

47. Chiu, C.-E. *et al.* Characterisation of Anti-Arrhythmic Drug Effects on Cardiac Electrophysiology using Physics-Informed Neural Networks. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 1–5 (IEEE, 2024).

48. Xie, J. & Yao, B. Physics-Constrained Deep Active Learning for Spatiotemporal Modeling of Cardiac Electrodynamics. *Computers in Biology and Medicine* **146**, 105586 (2022).

49. Zhu, L. *et al.* Physics-Informed Residual Learning with Spatiotemporal Local Support for Inverse ECG Reconstruction. *Scientific Reports* **15**, 31747 (2025).

50. Zhao, A. *et al.* Physics-Informed Neural Networks for Physiological Signals Processing and Modeling: A Narrative Review. *Physiological Measurement* (2025).

51. Moody, G. B. & Mark, R. G. The Impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine* **20**, 45–50 (2001).

52. Goldberger, A. L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **101**, e215–e220 (2000).

53. Moody, G. B. *et al.* A Noise Stress Test for Arrhythmia Detectors. *Computers in Cardiology* **11**, 381–384 (1984).

54. Sakib, S. *et al.* Harnessing Artificial Intelligence for Secure ECG Analytics at the Edge for Cardiac Arrhythmia Classification. In *Secure Edge Computing*, 137–153 (CRC Press, 2021).

55. Sakib, S. ECG Arrhythmia Classification Dataset. [Online; accessed 2025-10-07].

56. Freund, Y. & Schapire, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences* **55**, 119–139 (1997).

57. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* **29**, 1189–1232 (2001).

58. Ke, G. *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems* **30** (2017).

59. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (2016).

60. Prokhorenkova, L. *et al.* CatBoost: Unbiased Boosting with Categorical Features. *Advances in Neural Information Processing Systems* **31** (2018).

61. Arik, S. Ö. & Pfister, T. TabNet: Attentive Interpretable Tabular Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, 6679–6687 (2021).

62. Habib, A. Z. S. B. *et al.* TabSeq: A Framework for Deep Learning on Tabular Data via Sequential Ordering. In *International Conference on Pattern Recognition*, 418–434 (Springer, 2024).

63. Akiba, T. *et al.* Optuna: A Next-Generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631 (2019).

64. Gorishniy, Y. *et al.* TabM: Advancing Tabular Deep Learning with Parameter-Efficient Ensembling. In *The Thirteenth International Conference on Learning Representations* (2025).

65. Gorishniy, Y. *et al.* TabR: Tabular Deep Learning Meets Nearest Neighbors. In *The Twelfth International Conference on Learning Representations* (2024).

66. Gorishniy, Y. *et al.* On Embeddings for Numerical Features in Tabular Deep Learning. *Advances in Neural Information Processing Systems* **35**, 24991–25004 (2022).

67. Gorishniy, Y. *et al.* Revisiting Deep Learning Models for Tabular Data. *Advances in Neural Information Processing Systems* **34**, 18932–18943 (2021).

68. Friedman, M. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association* **32**, 675–701 (1937).

69. Demšar, J. Statistical Comparisons of Classifiers Over Multiple Data Sets. *Journal of Machine Learning Research* **7**, 1–30 (2006).

70. Nemenyi, P. B. *Distribution-Free Multiple Comparisons* (Princeton University, 1963).

71. Demšar, J. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research* **7**, 1–30 (2006).