ECGWavePuzzle as Morphology-Aware Auxiliary Supervision for Multitask Arrhythmia Classification

Guilherme Silva[1], Arthur Negrão[1], Pedro Silva[2], Eduardo Luz[2]

[1] Postgraduate Program in Computer Science – Federal University of Ouro Preto – Brazil
guilherme.lopes@aluno.ufop.edu.br

[2]Computing Department – Federal University of Ouro Preto – Brazil
eduluz@ufop.edu.br

Abstract. *Arrhythmia detection with deep learning often relies on heavy pre-processing and synthetic oversampling, obscuring whether models truly learn cardiac patterns. We propose a multitask self-supervised framework that jointly optimizes arrhythmia classification, RR-interval regression, and the ECG-WavePuzzle pretext task. This design encourages the model to capture both temporal rhythm and intra-beat morphology directly from raw ECG signals. Evaluated on MIT-BIH, PTB-XL, and IEGM, the approach improves stability and predictive performance over single-task baselines without synthetic data generation. Our results suggest that physiologically grounded auxiliary objectives provide a more principled path to robust ECG representation learning.*

1. Introduction

Cardiovascular diseases represent the leading cause of death globally, establishing the electrocardiogram (ECG) signal as an indispensable tool for their diagnosis [WHO 2021, Cohen 2019]. In this context, automating the detection of arrhythmias, which are abnormalities in heart rhythm, is critical for improving patient outcomes [Luz et al. 2016]. While deep learning models, particularly Convolutional Neural Networks (CNNs), have shown promising results in this domain [Hannun et al. 2019, Chen et al. 2024b], their true performance is often obscured by aggressive signal pre-processing and simplistic data balancing techniques.

The underlying challenge can be deconstructed into two distinct yet interrelated problems: modeling the intricate intra-beat morphology (the shape of a single heartbeat) and understanding the broader inter-beat temporal context (the rhythm and sequence of beats). The reliance of many state-of-the-art models on techniques like the Synthetic Minority Over-sampling Technique (SMOTE) makes it difficult to ascertain whether performance gains stem from a genuine understanding of cardiac patterns or are mere artifacts of the pre-processing pipeline itself [Luz et al. 2016, Fernández et al. 2018]. This reliance is particularly problematic, as SMOTE can introduce noisy, non-physiological samples, potentially corrupting the very morphological features it aims to augment [Fernández et al. 2018].

To advance in this field, it is important to build robust representations of single-beat characteristics without introducing confounding artifacts. Promising alternatives emerge from enriching the learning process with auxiliary tasks. For instance, estimating RR intervals provides vital temporal information, while self-supervised methods like the ECGWavePuzzle, where a model learns to reassemble a correctly ordered heartbeat from its shuffled segments, can promote a deep understanding of its morphology [Silva et al. 2024b]. Our central hypothesis is that forcing a model to learn these temporal and morphological features concurrently with the primary classification task can develop richer and more generalizable representations from the original data alone, thereby exposing the need for artificial sample generation. Multitask learning (MTL) provides the ideal framework for this paradigm, enabling a single model to optimize multiple correlated objectives simultaneously [Caruana 1997]. Preliminary studies have already confirmed the viability of this approach, such as the HydraNet proposed by [Silva et al. 2024a] for arrhythmia detection in embedded systems and the ResNet-18-based ECG-MACE model by [Lin et al. 2025] for predicting long-term cardiovascular events.

Building on these premises, this work proposes a multitask model that integrates arrhythmia classification with RR interval regression and the ECGWavePuzzle self-supervised task, extending the work of [Silva et al. 2024a]. We adopt an experimental setup that avoids complex pre-processing and over-sampling methods like SMOTE, allowing us to isolate and evaluate the true contribution of the learning architecture. Our investigation is structured to answer a central research question: To what extent can a multitask learning framework, which integrates temporal and self-supervised morphological objectives, enhance arrhythmia classification performance on imbalanced datasets without resorting to artificial data augmentation?

Our main contributions are summarized as follows: **(1)** We propose a multitask architecture that jointly learns arrhythmia classification, RR-interval regression, and the ECGWavePuzzle self-supervised task, capturing both temporal and morphological ECG features. **(2)** We evaluate the method on the MIT-BIH Arrhythmia Dataset [Moody and Mark 2001] under the AAMI standard and the inter-patient protocol of [De Chazal et al. 2004], without synthetic oversampling. **(3)** Experiments show that multitask learning consistently outperforms single-task baselines, indicating that physiologically grounded auxiliary tasks provide a promising alternative to synthetic data augmentation in the evaluated setup.

2. Related Works

The application of Multitask learning to the analysis of ECG signals is a growing research area, with strategies aimed at improving the efficiency and generalization of deep learning models. Recent works can be categorized by the clinical goal of the multitask application, spanning from signal preprocessing and arrhythmia diagnosis to long-term risk prediction.

Initial approaches have focused on optimizing ECG preprocessing by combining correlated tasks. [Chen et al. 2024a] proposed the MTL-NET based on Bidirectional LSTMs and an attention module to simultaneously perform Signal Quality Assessment (SQA) and denoising. By training both tasks in parallel, the model demonstrated performance superior to single-task methods, achieving a PRD of 20.46% in the denoising task and high precision (98.81%) and recall (98.84%) in identifying unacceptable quality signals. This approach shows that intrinsically linked preprocessing tasks mutually benefit within an MTL framework.

In the field of arrhythmia diagnosis, MTL has been applied both to increase efficiency in embedded systems and to refine classification accuracy. Focus-ing on resource-constrained devices like Implantable Cardioverter Defibrillators (ICDs), [Silva et al. 2024a] developed a lightweight model named HydraNet that concurrently performs ventricular arrhythmia detection (classification) and RR-interval estimation (regression). The inclusion of the auxiliary regression task helped the model achieve an F_β score of 0.88 and an overall accuracy of 93.55% with a low inference latency of 59.96 ms on a microcontroller. Complementarily, rather than combining tasks of different natures, [Geng et al. 2023] explored MTL to enhance a single primary diagnostic task. They created a synthetic auxiliary task from the arrhythmia class hierarchy itself and utilized an architecture with a Contextual Transformer (CoT) block. The method achieved an average F1-score of 0.833 on the PTB-XL dataset and 0.827 on the CPSC2018 dataset, outperforming single-task strategies.

Expanding the scope from immediate diagnosis to long-term risk prediction, [Lin et al. 2025] applied MTL to predict multiple Major Adverse Cardiovascular Events (MACE) within one year from a single 12-lead ECG. Their proposed ECG-MACE model was trained to simultaneously predict four distinct outcomes: (i) myocardial infarction, (ii) heart failure, (iii) ischemic stroke, and (iv) all-cause mortality. The multitask approach proved superior to single-task models across all categories, achieving high AUROC values of 0.90 for heart failure, 0.85 for myocardial infarction, and 0.89 for mortality. The benefit of MTL was particularly evident in the challenging task of ischemic stroke prediction, where the AUROC increased from 0.66 in the single-task model to 0.76 in the MTL version.

However, to the best of our knowledge, the application of a self-supervised technique like the ECGWavePuzzle, proposed by [Silva et al. 2024b], within a multitask framework that simultaneously addresses three distinct ECG-related tasks remains an underexplored area.

3. Methodology

We propose a multitask self-supervised learning framework for arrhythmia detection from ECG signals. The pipeline consists of three components: (i) ECG segmentation using annotated R-peaks, (ii) a sequence-level self-supervised pretext task (*ECGWavePuzzle*), and (iii) a multitask architecture (`HydraNet`) that jointly learns arrhythmia classification, RR-interval regression, and the puzzle task.

3.1. ECG Signal Segmentation

Continuous ECG recordings from the three dataset used in the experiments (MIT-BIH, PTB-XL and IEGM) are segmented into fixed-length sequences using the annotated R-peaks as fiducial points. Following [Shen et al. 2024], each input sample consists of a window of 1,250 points.

For each annotated R-peak, a segment is extracted starting 150 samples before the peak and extending for 1,099 additional samples, producing a 1,250-sample window (Figure 1). Because every R-peak is used as a reference, consecutive segments overlap, effectively increasing the number of training samples.

A key requirement of our approach is that all beats within a segment must be complete. Following Silva *et al.* [Silva et al. 2024a], the heartbeat is defined as a 300-sample window centered on its R-peak (150 samples before and 149 after). To guarantee

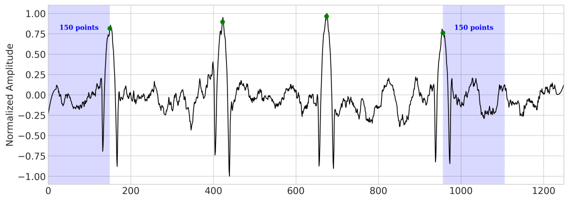
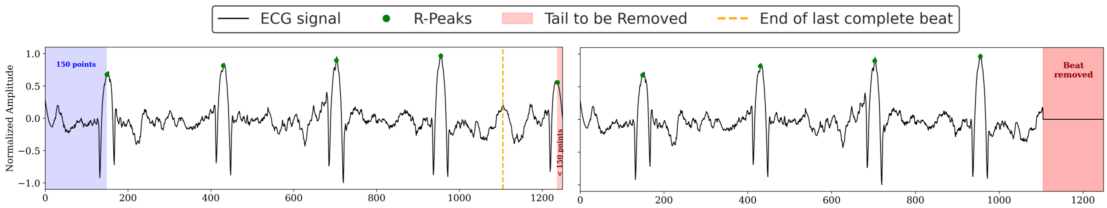


Figure 1. Example of a valid 1,250-sample ECG segment containing only complete beats.

this condition, two validation checks are applied: **(1) First Beat Validation:** the first R-peak must have at least 150 preceding samples; and **(2) Last Beat Validation:** the final R-peak must have at least 150 subsequent samples.

Segments failing the second criterion contain an incomplete final beat. In this case, the signal portion after the last complete beat is masked by setting its amplitude to zero (Figure 2). This prevents the model from learning artifacts associated with truncated beats while ensuring that the incomplete beat appears as a valid beat in the next overlapping segment.



(a) The last R-peak has fewer than 150 subsequent points.

(b) Zeroed out the incomplete portion of the signal.

Figure 2. Processing of an invalid ECG segment.

Each segment receives a binary label based on the annotations of the beats it contains: a segment is labeled *normal* if all beats are normal and *arrhythmic* otherwise.

3.2. Sequence-Level *ECGWavePuzzle*

To exploit structural information in ECG signals, we extend the *ECGWavePuzzle* pretext task proposed by [Silva et al. 2024b]. While the original formulation operates at the single-beat level, we adapt it to operate on entire ECG segments. Given a segmented 1,250-sample window, individual beats are extracted using the annotated R-peaks. Each beat corresponds to a 300-sample window centered on its R-peak. A predefined permutation rule is then applied uniformly to all beats in the sequence, as shown in Figure 3.

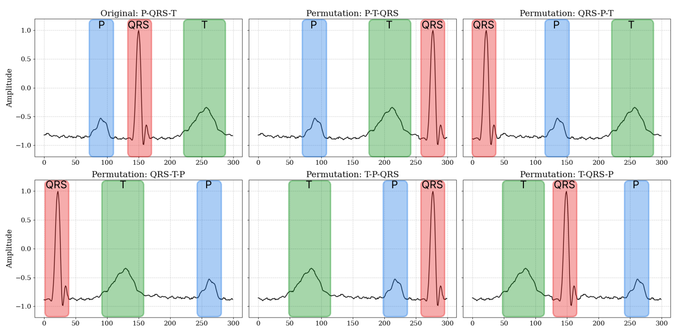


Figure 3. Sequence-level *ECGWavePuzzle*. The same permutation is applied to all beats in the segment.

Because the same transformation is applied to the entire segment, a single permutation label can be assigned to the sequence. The model is therefore trained to identify the correct ordering of ECG waveform components. This task encourages the network to learn morphological and temporal dependencies within ECG signals.

3.3. Baseline Architectures

We evaluate our approach using two CNN-based architectures that serve both as standalone classifiers and as feature extraction backbones for the multitask model: `VANet`, `LightNet`

VANet: The `VANet` architecture [Shen et al. 2024] consists of a downsampling block followed by five convolutional blocks and an output layer. The downsampling stage applies `MaxPool1d` with stride 10 to reduce temporal resolution. Each convolutional block contains a `Conv1d` layer followed by batch normalization and ReLU activation. The numbers of filters are $\{3, 2, 2, 8, 1\}$ with kernel sizes $\{3, 3, 3, 3, 6\}$. In our multitask framework, the downsampling layer and the first four convolutional blocks are used as a shared feature extractor.

LightNet: A compact and lightweight CNN architecture that first downsamples the input signal using `MaxPool1d` with a stride of 10, followed by two convolutional blocks composed of `Conv1d`, batch normalization, and ReLU activation. Channel dimensions increase from 1 to 2 and from 2 to 4. The extracted features are flattened and processed by a fully connected classifier with a 64-unit hidden layer, ReLU activation, and dropout ($p = 0.5$), followed by the final classification layer. In the multitask setting, the convolutional blocks form the shared encoder.

3.4. HydraNet: multitask Learning

The HydraNet is a multitask architecture designed to jointly learn complementary ECG representations. The model contains a shared feature extractor followed by multiple task-specific heads (Figure 4). The backbone can be instantiated using any baseline architecture described above. The shared backbone processes the 1,250-sample ECG segment and produces a feature map that is flattened into a vector of dimension N. This representation is fed to task-specific branches.

Two configurations are evaluated. In the two-task configuration, the network jointly performs arrhythmia classification (AC) and RR-interval regression. Each branch contains two fully connected layers (`Linear(N,64)` and `Linear(64,32)`) with ReLU activation and dropout. The AC head outputs two logits (`Linear(32,2)`), while the RR head predicts a continuous value (`Linear(32,1)`). The three-task configuration extends the model by adding the *ECGWavePuzzle* branch, which predicts one of six permutation classes (`Linear(32,6)`).

To balance the tasks during training, we adopt uncertainty-based loss weighting. Each task is associated with a learnable variance parameter σ^2, allowing the model to adaptively adjust task importance. The total loss is defined as:

$$L_{total} = \frac{L_{AC}}{2\sigma^2_{AC}} + \frac{L_{RR}}{2\sigma^2_{RR}} + \frac{L_{Puz}}{2\sigma^2_{Puz}} + \log(\sigma_{AC}) + \log(\sigma_{RR}) + \log(\sigma_{Puz}), \tag{1}$$

where L_{AC} is the cross-entropy loss for arrhythmia classification, L_{RR} is the mean absolute error for RR regression, and L_{Puz} is the cross-entropy loss for the puzzle task.

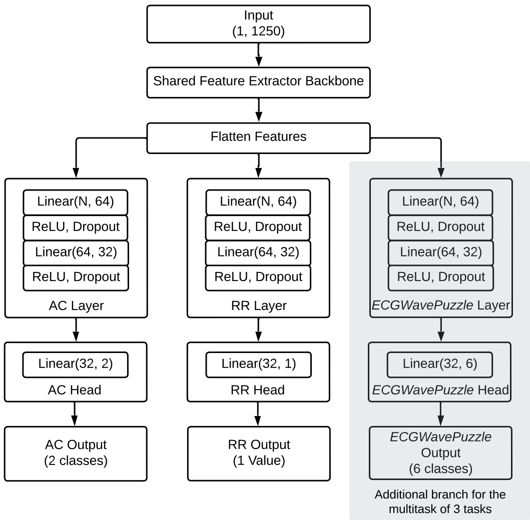


Figure 4. Proposed multitask HydraNet architecture with the shared feature extractor backbone and task-specific layers and heads.

This formulation enables stable joint optimization and allows the network to dynamically balance task contributions during training.

4. Experiment and Results

This section presents the datasets, evaluation protocols, and experimental design used to assess the proposed multitask self-supervised framework for arrhythmia classification. Source code to fully reproduce the work is available in https://github.com/<omittedduetopeerreview>.

4.1. Datasets and Evaluation Protocols

We evaluate the method on three cardiac signal benchmarks with complementary properties: MIT-BIH for beat-level arrhythmia classification, PTB-XL for multi-label diagnostic ECG classification, and IEGM for ventricular arrhythmia detection from intracardiac recordings.

PTB-XL Dataset. PTB-XL is a large-scale public 12-lead ECG benchmark with nearly 21,000 recordings and detailed diagnostic annotations [Wagner et al. 2020]. We use the standard 12-lead setting sampled at 500 Hz and follow the official split: 17,439 recordings for training, 2,180 for validation, and 2,180 for testing [Strodthoff et al. 2020]. Table 1 presents the number of records of each class.

Table 1. Class distribution in PTB-XL.

Number of Records	Class	Description
7185	NORM	Normal ECG
3232	CD	Conduction Disturbance
3064	STTC	ST/T Change
2936	MI	Myocardial Infarction
815	HYP	Hypertrophy

The downstream task is multi-label classification over the five diagnostic superclasses: NORM, MI, STTC, CD, and HYP [Wagner et al. 2020]. This setting is clinically realistic because multiple abnormalities may co-occur. It is also substantially imbalanced, especially for HYP, making PTB-XL a useful benchmark for evaluating the robustness of auxiliary supervision.

The PTB-XL recordings, acquired at 500 Hz, were converted to 360 Hz to ensure uniform temporal resolution. Next, the signal was partitioned into individual heartbeats using R-peaks as temporal references. Additionally, each beat was normalized to a fixed length of 1250 samples. Finally, we cast the task as binary classification by merging CD, STTC, MI, and HYP into a single arrhythmic class, as they represent types of arrhythmia.

MIT-BIH Arrhythmia Database. The MIT-BIH Arrhythmia Database is a standard benchmark for arrhythmia classification [Luz et al. 2016, Moody and Mark 2001]. It contains 48 two-lead ECG recordings of 30 minutes each, sampled at 360 Hz, from 47 subjects. Following the ANSI/AAMI EC57 standard [ANSI/AAMI 2008], beats are grouped into five superclasses: Normal (N), Supraventricular Ectopic Beat (S), Ventricular Ectopic Beat (V), Fusion Beat (F), and Unknown Beat (Q). In this work, we focus on three primary AAMI classes (N, S, and V) and cast the task as binary classification by merging S and V into a single arrhythmic class, as both represent types of arrhythmia.

We adopt the inter-patient protocol of [De Chazal et al. 2004] and exclude the four pacemaker recordings according to ANSI/AAMI recommendations [ANSI/AAMI 2008, Silva et al. 2025]. The training set (DS1) contains records 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, and 230, while the test set (DS2) contains records 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, and 234. Under this binary formulation, MIT-BIH is relatively balanced compared with many arrhythmia benchmarks. Therefore, the main question is whether auxiliary objectives improve representation quality and generalization, rather than whether they compensate for severe class skew. All signals are used in their raw form, without handcrafted preprocessing.

IEGM Database. The IEGM dataset is derived from intracardiac electrograms provided by SingularMedical [Singular Medical Technologies Inc. 2023] and used in the 2022 ACM/IEEE TinyML Design Contest [Jia et al. 2023]. It includes recordings from 90 patients with single-chamber implantable cardioverter-defibrillators, sampled at 250 Hz. Signals are segmented with a 5-second sliding window and a 150-sample overlap, yielding windows of 1250 samples.

The IEGM dataset comprised 30,190 samples in total, with 24,568 samples allocated to training (DS1), including 12,741 from class 0 and 11,827 from class 1, and 5,622 samples allocated to testing (DS2), including 3,235 from class 0 and 2,387 from class 1.

Expert annotations define two high-level classes: ventricular arrhythmias (VAs) and non-ventricular arrhythmias (non-VAs), which are further divided into eight rhythm types: AFb, AFt, SR, SVT, VFb, VFt, VPD, and VT. We follow the original patient-independent split [Jia et al. 2023], with approximately 85% of the data for training and 15% for testing.

4.2. Results

We compare four training regimes for each backbone (VANet and LightNet) on PTB-XL, MIT-BIH, and IEGM: (i) single-task classification, (ii) classification with RR-interval regression, (iii) classification with ECGWavePuzzle, and (iv) a 3-task setting combining classification, RR regression, and ECGWavePuzzle. Results are reported as mean ± standard deviation over two seeds.

4.2.1. PTB-XL

Table 2 shows that auxiliary supervision substantially improves VANet, whose baseline is unstable. The best VANet results are obtained with the 3-task setting, with accuracy $\mathbf{0.7605 \pm 0.0080}$ and F1-score $\mathbf{0.7611 \pm 0.0080}$, while the 2-task model yields the highest precision and specificity. ECGWavePuzzle alone recovers most of the gain, indicating that morphology-aware self-supervision is highly effective even without RR regression.

Table 2. PTB-XL results (mean ± std). Both the 2-task and 3-task multitask setups use HydraNet. Best values per backbone are in bold (higher is better).

Backbone	Regime	Acc.	Prec.	F1	Spec.	Time (s)
VANet	Baseline	0.6500 ± 0.0858	0.5352 ± 0.2170	0.5717 ± 0.1649	0.6222 ± 0.1223	38.395 ± 4.085
VANet	2-task multitask	0.7591 ± 0.0023	$\mathbf{0.7719 \pm 0.0034}$	0.7599 ± 0.0023	0.7662 ± 0.0029	125.020 ± 20.760
VANet	ECGWavePuzzle-class.	0.7598 ± 0.0091	0.7617 ± 0.0098	0.7604 ± 0.0092	0.7587 ± 0.0101	108.270 ± 20.390
VANet	3-task multitask	$\mathbf{0.7605 \pm 0.0080}$	0.7625 ± 0.0083	$\mathbf{0.7611 \pm 0.0080}$	0.7594 ± 0.0085	165.410 ± 26.700
LightNet	Baseline	0.7692 ± 0.0053	0.7821 ± 0.0045	0.7700 ± 0.0054	0.7765 ± 0.0049	29.535 ± 3.545
LightNet	2-task multitask	0.7724 ± 0.0012	$\mathbf{0.7894 \pm 0.0000}$	0.7730 ± 0.0013	0.7817 ± 0.0007	113.735 ± 0.295
LightNet	ECGWavePuzzle-class.	0.7734 ± 0.0038	0.7781 ± 0.0036	$\mathbf{0.7742 \pm 0.0038}$	0.7753 ± 0.0036	120.815 ± 2.625
LightNet	3-task multitask	0.7714 ± 0.0031	0.7741 ± 0.0035	0.7721 ± 0.0031	0.7714 ± 0.0036	183.625 ± 10.335

For LightNet, which is already strong as a baseline, improvements are smaller but consistent. ECGWavePuzzle-classification achieves the best accuracy and F1-score, whereas RR regression gives the highest Precision and Specificity. Table 3 shows that ECGWavePuzzle also improves class-wise balance. For VANet, the baseline is heavily biased, whereas ECGWavePuzzle produces a more balanced sensitivity profile. For LightNet, the main effect is improved class 1 sensitivity.

Table 3. PTB-XL per-class sensitivity (recall), mean ± std. Both the 2-task and 3-task multitask setups use HydraNet. Best values per backbone and class are in bold.

Regime	VANet		LightNet	
	Sens. (Class 0)	Sens. (Class 1)	Sens. (Class 0)	Sens. (Class 1)
Baseline	0.4059 ± 0.4059	$\mathbf{0.8385 \pm 0.1615}$	0.8336 ± 0.0018	0.7194 ± 0.0081
2-task multitask	$\mathbf{0.8219 \pm 0.0074}$	0.7105 ± 0.0016	$\mathbf{0.8545 \pm 0.0033}$	0.7089 ± 0.0047
ECGWavePuzzle-class.	0.7495 ± 0.0181	0.7679 ± 0.0021	0.7908 ± 0.0028	0.7599 ± 0.0045
3-task multitask	0.7515 ± 0.0119	0.7674 ± 0.0050	0.7719 ± 0.0074	$\mathbf{0.7710 \pm 0.0002}$

4.2.2. MIT-BIH

Table 4 summarizes the MIT-BIH results. For VANet, the 2-task setting yields the best Accuracy and F1-score, showing that RR regression is highly informative for beat-level discrimination. ECGWavePuzzle improves Precision and Specificity, while the 3-task setting provides the most conservative operating point. For LightNet, the 3-task model achieves the best Accuracy, F1-score, and Specificity, and ECGWavePuzzle-classification gives the highest Precision.

Because MIT-BIH is relatively balanced in our binary formulation, these gains cannot be attributed mainly to imbalance mitigation. Instead, they suggest that auxiliary

Table 4. MIT-BIH results (mean ± std). Both the 2-task and 3-task multitask setups use HydraNet. Best values per backbone are in bold (higher is better).

Backbone	Regime	Acc.	Prec.	F1	Spec.	Time (s)
VANet	Baseline	0.4358 ± 0.1852	0.3963 ± 0.3334	0.3734 ± 0.2728	0.5719 ± 0.0719	38.385 ± 25.905
VANet	2-task multitask	0.6323 ± 0.0035	0.7444 ± 0.0066	$\mathbf{0.6567 \pm 0.0033}$	0.6627 ± 0.0082	24.095 ± 6.975
VANet	ECGWavePuzzle-class.	0.6195 ± 0.0055	0.7656 ± 0.0026	0.6439 ± 0.0053	0.6812 ± 0.0044	39.770 ± 1.420
VANet	3-task multitask	0.6157 ± 0.0031	$\mathbf{0.7715 \pm 0.0009}$	0.6397 ± 0.0030	$\mathbf{0.6854 \pm 0.0020}$	45.215 ± 1.685
LightNet	Baseline	0.6887 ± 0.0232	0.7880 ± 0.0118	0.7092 ± 0.0215	0.7252 ± 0.0192	12.185 ± 1.305
LightNet	2-task multitask	0.6927 ± 0.0254	0.7812 ± 0.0124	0.7126 ± 0.0232	0.7185 ± 0.0201	21.540 ± 5.960
LightNet	ECGWavePuzzle-class.	0.6922 ± 0.0161	$\mathbf{0.8044 \pm 0.0020}$	0.7124 ± 0.0150	0.7433 ± 0.0032	39.000 ± 0.690
LightNet	3-task multitask	0.7107 ± 0.0063	0.8030 ± 0.0007	$\mathbf{0.7296 \pm 0.0057}$	$\mathbf{0.7475 \pm 0.0008}$	44.600 ± 2.890

tasks improve representation learning even when class skew is limited. Table 5 confirms that auxiliary tasks improve class-wise balance, especially for VANet.

Table 5. MIT-BIH per-class sensitivity (recall), mean ± std. Both the 2-task and 3-task multitask setups use HydraNet. Best values per backbone and class are in bold.

Regime	VANet		LightNet	
	Sens. (Class 0)	Sens. (Class 1)	Sens. (Class 0)	Sens. (Class 1)
Baseline	0.2990 ± 0.2990	$\mathbf{0.8448 \pm 0.1552}$	0.6520 ± 0.0273	0.7986 ± 0.0110
2-task multitask	$\mathbf{0.6017 \pm 0.0012}$	0.7238 ± 0.0176	0.6669 ± 0.0308	0.7700 ± 0.0092
ECGWavePuzzle-class.	0.5573 ± 0.0065	0.8051 ± 0.0022	0.6406 ± 0.0290	$\mathbf{0.8460 \pm 0.0227}$
3-task multitask	0.5456 ± 0.0042	0.8251 ± 0.0004	$\mathbf{0.6736 \pm 0.0120}$	0.8214 ± 0.0104

4.2.3. IEGM

Table 6 reports the IEGM results. For VANet, all auxiliary variants dramatically outperform the unstable baseline, and the 3-task model achieves the best values across all metrics. This indicates strong complementarity between RR regression and ECGWavePuzzle in ventricular arrhythmia detection from intracardiac signals.

Table 6. IEGM results (mean ± std). Both the 2-task and 3-task multitask setups use HydraNet. Best values per backbone are in bold (higher is better).

Backbone	Regime	Acc.	Prec.	F1	Spec.	Time (s)
VANet	Baseline	0.6613 ± 0.2367	0.5398 ± 0.3595	0.5751 ± 0.3220	0.6947 ± 0.1947	24.970 ± 5.090
VANet	2-task multitask	0.9172 ± 0.0012	0.9173 ± 0.0012	0.9171 ± 0.0010	0.9135 ± 0.0000	31.825 ± 0.065
VANet	ECGWavePuzzle-class.	0.9153 ± 0.0043	0.9162 ± 0.0035	0.9154 ± 0.0042	0.9154 ± 0.0027	36.780 ± 0.270
VANet	3-task multitask	0.9181 ± 0.0020	0.9185 ± 0.0016	0.9182 ± 0.0020	0.9173 ± 0.0007	44.385 ± 0.005
LightNet	Baseline	0.9335 ± 0.0172	0.9336 ± 0.0171	0.9333 ± 0.0174	0.9302 ± 0.0195	19.345 ± 0.035
LightNet	2-task multitask	0.9420 ± 0.0119	0.9422 ± 0.0121	0.9421 ± 0.0119	0.9412 ± 0.0131	29.540 ± 0.070
LightNet	ECGWavePuzzle-class.	0.9438 ± 0.0110	0.9442 ± 0.0109	0.9440 ± 0.0109	0.9437 ± 0.0113	32.405 ± 0.175
LightNet	3-task multitask	0.9416 ± 0.0120	0.9417 ± 0.0122	0.9415 ± 0.0121	0.9406 ± 0.0136	40.615 ± 0.085

For LightNet, the baseline is already strong, but ECGWavePuzzle-classification still achieves the best Accuracy, Precision, F1-score, and Specificity. Thus, morphology-aware self-supervision remains useful even with a high-capacity backbone. Table 7 shows that auxiliary objectives also improve class-wise sensitivity, especially for VANet.

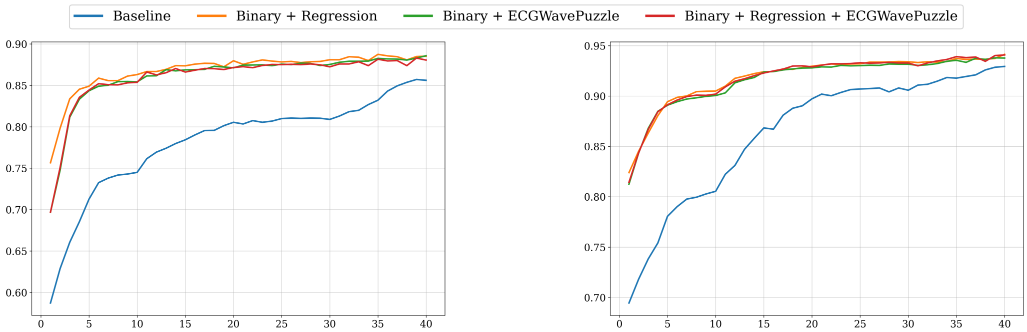
4.3. Results discussion

Figures 5 and 6 also confirm faster convergence and better separability under auxiliary supervision in IEGM dataset with VANet and LightNet. In all experiments, the multi-objective variants performed better than the baseline. This behavior is also observed in PTB-XL and MIT-BIH datasets.

Across datasets, three trends are consistent. First, auxiliary supervision is most

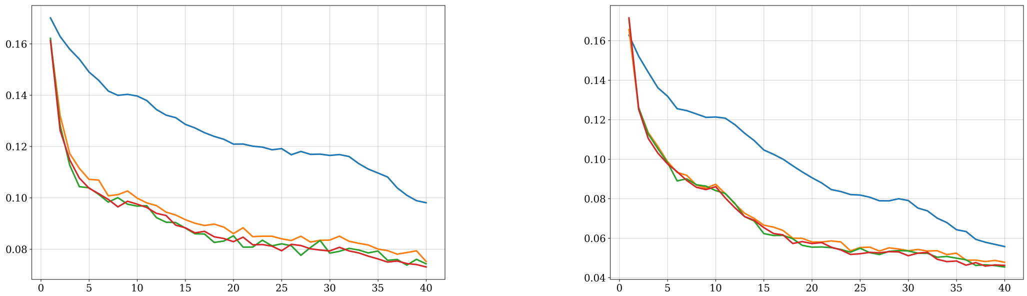
Table 7. IEGM per-class sensitivity (recall), mean ± std. Both the 2-task and 3-task multitask setups use HydraNet. Best values per backbone and class are in bold.

Backbone	Regime	Sens. (Class 0)	Sens. (Class 1)
VANet	Baseline	0.4729 ± 0.4729	0.9164 ± 0.0836
VANet	2-task multitask	0.9377 ± 0.0073	0.8894 ± 0.0071
VANet	ECGWavePuzzle-class.	0.9141 ± 0.0142	0.9168 ± 0.0090
VANet	3-task multitask	0.9224 ± 0.0096	0.9123 ± 0.0081
LightNet	Baseline	0.9522 ± 0.0051	0.9081 ± 0.0338
LightNet	2-task multitask	0.9468 ± 0.0052	0.9355 ± 0.0210
LightNet	ECGWavePuzzle-class.	0.9446 ± 0.0086	0.9428 ± 0.0140
LightNet	3-task multitask	0.9471 ± 0.0031	0.9340 ± 0.0241



(a) Accuracy with VANet.

(b) Accuracy with LightNet.



(c) Train loss with VANet.

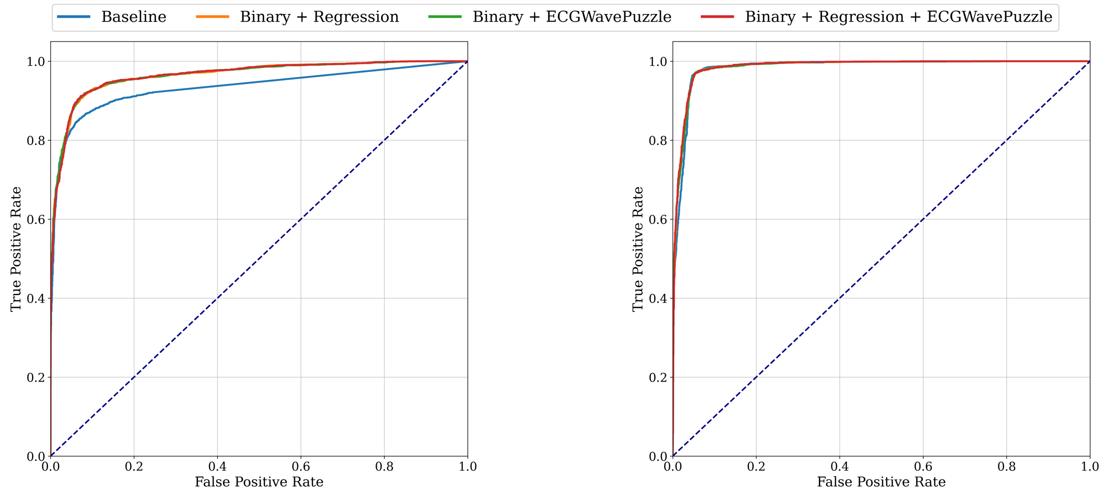
(d) Train loss with LightNet.

Figure 5. Comparison of the baseline and multi-objective variants on IEGM dataset. In (a) and (b), the validation accuracy (y-axis) vs. epochs (x-axis). In (c) and (d), the train loss (y-axis) over the epochs (x-axis).

beneficial when the baseline model is weak or unstable. Second, the results are consistent with the hypothesis that RR regression and ECGWavePuzzle provide complementary supervisory signals: the former emphasizes temporal structure, while the latter encourages morphology-aware representations. Third, the 3-task formulation is robust in some cases, although the best regime still depends on dataset characteristics and backbone capacity.

5. Conclusion

In the evaluated protocols, competitive performance can be achieved without synthetic oversampling or heavy handcrafted preprocessing. Instead, by jointly learning arrhythmia classification, RR-interval regression, and the self-supervised ECGWavePuzzle task, HydraNet extracts richer temporal and morphological representations directly from raw ECG signals. Across MIT-BIH, PTB-XL, and IEGM, auxiliary supervision improves performance in most evaluated settings and notably stabilizes weak baselines such as VANet, with RR regression contributing temporal context, ECGWavePuzzle strengthen-



(a) IEGM with VANet. **(b) IEGM with LightNet.**

Figure 6. ROC curves on IEGM for the best selected model in each regime.

ing morphology-aware features, and their combination often providing the most robust overall behavior.

More broadly, our findings suggest that physiologically motivated auxiliary supervision is a useful design axis alongside data processing choices. By aligning auxiliary tasks with cardiac physiology, multitask self-supervised learning yields models that are more stable in the evaluated public benchmarks and promising for future real-world validation. These findings position multitask self-supervision as a principled foundation for the next generation of reliable ECG intelligence.

Acknowledgments

The authors acknowledge the support of the *Fundação de Amparo à Pesquisa do Estado de Minas Gerais* (FAPEMIG, project APQ-01768-24), the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), the C*oordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), the *Universidade Federal de Ouro Preto* (PROPPI/UFOP), and its Graduate Program in Computer Science (PPGCC/UFOP).

References

ANSI/AAMI (2008). Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. American National Standards Institute, Inc. (ANSI), Association for the Advancement of Medical Instrumentation (AAMI). ANSI/AAMI/ISO EC57, 1998-(R)2008.

Caruana, R. (1997). Multitask learning. *Machine learning*, 28(1):41–75.

Chen, M., Li, Y., Zhang, L., Zhang, X., Gao, J., Sun, Y., Shi, W., and Wei, S. (2024a). Multitask learning-based quality assessment and denoising of electrocardiogram signals. *IEEE Transactions on Instrumentation and Measurement*, 73:1–13.

Chen, Z., Yang, D., Cui, T., Li, D., Liu, H., Yang, Y., Zhang, S., Yang, S., and Ren, T.-L. (2024b). A novel imbalanced dataset mitigation method and ecg classification model based on combined 1d_cbam-autoencoder and lightweight cnn model. *Biomedical Signal Processing and Control*, 87:105437.

Cohen, A. (2019). *Biomedical Signal Processing: Volume 2: Compression and Automatic Recognition*. CRC press.

De Chazal, P., O'Dwyer, M., and Reilly, R. B. (2004). Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE transactions on biomedical engineering*, 51(7):1196–1206. Publisher: IEEE.

Fernández, A., Garcia, S., Herrera, F., and Chawla, N. V. (2018). Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, 61:863–905.

Geng, Q., Liu, H., Gao, T., Liu, R., Chen, C., Zhu, Q., and Shu, M. (2023). An ecg classification method based on multi-task learning and cot attention mechanism. In *Healthcare*, volume 11, page 1000. MDPI.

Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., and Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1):65.

Jia, Z., Li, D., Liu, C., Liao, L., Xu, X., Ping, L., and Shi, Y. (2023). Tinyml design contest for life-threatening ventricular arrhythmia detection. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 43(1):127–140.

Lin, C.-H., Liu, Z.-Y., Chu, P.-H., Chen, J.-S., Wu, H.-H., Wen, M.-S., Kuo, C.-F., and Chang, T.-Y. (2025). A multitask deep learning model utilizing electrocardiograms for major cardiovascular adverse events prediction. *npj Digital Medicine*, 8(1):1.

Luz, E. J. d. S., Schwartz, W. R., Cámara-Chávez, G., and Menotti, D. (2016). Ecg-based heartbeat classification for arrhythmia detection: A survey. *Computer methods and programs in biomedicine*, 127:144–164.

Moody, G. B. and Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50. Publisher: IEEE.

Shen, C., Chen, C., and Zhang, M. (2024). Vanet: A solution for ventricular arrhythmias detection of iegm on embedded devices. *IEEE Embedded Systems Letters*, 17(3):176–179.

Silva, G., Negrão, A., Moreira, G., Luz, E., and Silva, P. (2024a). An embedding multitask neural network for efficient arrhythmia detection. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 412–423. SBC.

Silva, G., Silva, P., Moreira, G., Freitas, V., Gertrudes, J., and Luz, E. (2025). A systematic review of ecg arrhythmia classification: Adherence to standards, fair evaluation, and embedded feasibility. *arXiv preprint arXiv:2503.07276*.

Silva, G., Silva, P., Moreira, G., and Luz, E. (2024b). Bridging the gap in ecg classification: Integrating self-supervised learning with human-in-the-loop amid medical equipment hardware constraints. In *International Symposium on Applied Reconfigurable Computing*, pages 63–74. Springer.

Singular Medical Technologies Inc. (2023). Iegm dataset. `https://www.singularmedical.net`. Accessed: 2023-05-01.

Strodthoff, N., Wagner, P., Schaeffter, T., and Samek, W. (2020). Deep learning for ecg analysis: Benchmarks and insights from ptb-xl. *IEEE journal of biomedical and health informatics*, 25(5):1519–1528.

Wagner, P., Strodthoff, N., Bousseljot, R.-D., Kreiseler, D., Lunze, F. I., Samek, W., and Schaeffter, T. (2020). Ptb-xl, a large publicly available electrocardiography dataset. *Scientific data*, 7(1):154.

WHO (2021). Cardiovascular diseases (cvds). `https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)`.