XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Automatic Detection of Cardiac Pathologies from ECG Signals Using a Hybrid Transformer-Supervised Variational Autoencoder (VAE)

Elis De Mori Ravagnani, Daniel Cruz Cavalieri

Dept. of Control and Automation Engineering
Instituto Federal do Espírito Santo
Av. dos Sabiás,330, 29166-630, Serra/ES, Brazil
elisravagnani@gmail.com, daniel.cavalieri@ifes.edu.br

Abstract. Cardiovascular diseases remain a leading global health concern, with electrocardiogram (ECG) analysis serving as a critical diagnostic tool. This study presents a hybrid model: Transformer-Supervised Variational Autoencoder (VAE) that addresses challenges in automated ECG analysis: accurate arrhythmia classification, interpretable latent space representation, and synthetic ECG generation for data augmentation. Key findings emerge: The model excels (99.19% accuracy) in discriminating normal beats (N) and ventricular ectopic beats (V), but shows limitations with supraventricular ectopic beats (S), a challenge mirroring real-world ECG interpretation, as evidenced by confusion matrices and latent space visualizations. UMAP projections (supervised and unsupervised) reveal that S-beats persistently overlap with normal beats due to shared morphologies, aligning with cardiological knowledge. Generated ECG signals retain critical P-QRS-T features despite elevated reconstruction loss (68.56%), demonstrating potential for data augmentation in clinical scenarios. The work advances ECG analysis by introducing a Transformer-VAE hybrid for joint representation learning and classification and providing visual evidence of pathological clustering in latent space. Future work should address S-beat discrimination through rhythm-specific attention mechanisms and also address the improvement of reconstruction fidelity.

Keywords: Variational Autoencoder (VAE), Cardiac Anomaly Detection, Transformer, Latent Space Visualization.

1 Introduction

The electrocardiogram (ECG) is the most commonly used tool for diagnosing heart diseases [1]. This non-invasive exam evaluates the electrical activity of the heart through electrodes attached to the skin, providing information about heart rhythm and the number of beats per minute.

ECG interpretation plays a critical role in the diagnosis of several heart conditions, such as arrhythmias, myocardial infarction, and pericarditis. It is vital for the early diagnosis, prompt treatment and improved patient outcomes.

Each wave and segment of the ECG signal, such as the P wave, QRS complex, and T wave, provides fundamental insights for determining the type and severity of the diseases [2].

Due to the challenges of manual ECG interpretation, such as its time-consuming nature and variability between experts, AI-powered tools are being developed to provide efficient and accurate automated analysis.

Among the emerging technologies in this area, deep generative models such as Variational Autoencoders (VAE) have shown great potential for ECG analysis. Unlike traditional methods, VAEs are capable of reconstructing original inputs and generating synthetic data to augment datasets, improve generalization, and facilitate the discovery of patterns. Therefore, potentially enhancing the quality of automated diagnostic systems, particularly in cases with data imbalance.

In light of these developments, this paper presents an intelligent system for the automatic detection of cardiopathies from ECG signals, based on advanced deep learning techniques. Specifically, it's introduced the use of a Transformer Supervised Variational Autoencoder (VAE), a hybrid architecture that combines the representational power of Transformer-based VAEs with the discriminative capabilities of supervised classifiers. This approach aims to unite the strengths of both generative modeling and deep sequence learning, offering a robust solution for ECG interpretation.

2 Related Work

Over the past decades, a wide range of methods has been proposed for automated ECG analysis, particularly for heartbeat classification and arrhythmia detection. These approaches typically involve signal preprocessing, feature extraction, and classification using machine learning or deep learning models.

Sahoo et al.[3] presented a systematic review of state-of-the-art methods for detecting cardiac arrhythmias using ECG signals, covering signal decomposition, feature extraction, and automated detection. They highlight the strengths and limitations of existing approaches and emphasize the importance of real-time, computer-aided systems.

Kizenetsov et al.[4] proposed a variational autoencoder to generate ECG signals by encoding with a minimal number of latent variables. Their method produced realistic synthetic ECGs. This approach addresses the shortage of labeled ECG data for supervised learning.

Tang et al.[5] proposed a hierarchical transformer architecture for ECG diagnosis, addressing limitations of standard transformers in processing multi-lead signals. Their model combines a depth-wise convolutional encoder, a three-stage transformer with multi-scale feature aggregation via a CLS token, and an attention-gated module for lead dependency learning. The model outperformed CNNs and existing transformer approaches, achieving superior FBeta and challenge scores.

Van der Valk et al.[6] developed a supervised VAE model to analyze ECG signals for heart function (LVF). The model was split into two parts using adversarial learning: one focused on heart function and another for general ECG features. Their method achieved state-of-the-art performance and provided visualizable ECG feature attributions through latent-space manipulation.The model's outputs correlated strongly with survival outcomes, matching echocardiogram labels.

In contrast to prior studies, this Transformer-Supervised VAE integrates arrhythmia classification, latent space interpretability, and ECG generation in a single framework. While Tang et al. focus on transformer-based classification and Van der Valk et al. on disentangled representations, our model combines both approaches. In this work, misclassifications were analyzed through visual latent space exploration, and the potential of synthetic ECG generation for data augmentation is demonstrated.

3 Methods

In this section, we present the methodology developed for the automatic classification of ECG signals. The overall workflow of the proposed approach is outlined in Figure 1. Each component of the methodology is described in the following subsections.

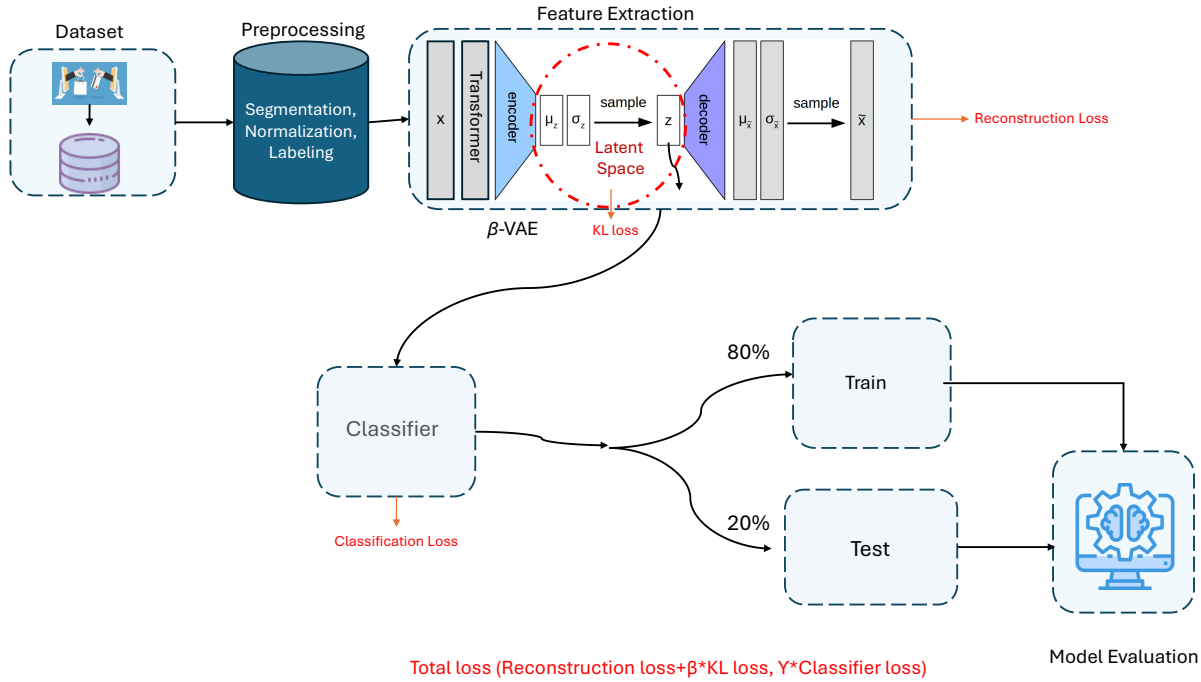


Figure 1. Proposed Methodology

3.1 Dataset

In this study, the PhysioNet MIT-BIH Arrhythmia Database was used to assess the effectiveness of the proposed methodology. This dataset comprises 48 half-hour ECG recordings, each sampled at 360 Hz and containing two channels. Expert annotators manually classified the heartbeats into 15 distinct classes based on the detected arrhythmia types. Consistent with common practice in the literature, the first channel was utilized for heartbeat detection and subsequent analysis. The richness and diversity of this dataset make it a standard benchmark for evaluating automatic arrhythmia detection systems.

3.2 Preprocessing

As described in Figure 1, the preprocessing pipeline included segmentation, normalization, and labeling of the ECG signals. Each heartbeat was extracted using annotated R-peaks, with a fixed window of 1 second centered around each R-peak. To avoid misalignment, beats with PR intervals exceeding 300 ms were excluded, and zero-padding was applied to shorter beats. All ECG signals underwent min-max normalization, thereby reducing amplitude-related bias and ensuring consistency across samples.

Heartbeats were categorized into three clinically significant groups: (1) N (Non-ectopic): normal beats, bundle branch blocks, and atrial aberrations; (2) S (supraventricular ectopic): premature atrial/junctional beats; and (3) V (ventricular ectopic): premature ventricular contractions and ventricular escape beats. Rare classes, such as fusion or paced beats, were excluded to minimize noise and improve model robustness.

3.3 Feature Extraction

The Feature Extraction stage leverages a Transformer Supervised Variational Autoencoder (β-VAE) where the segmented and normalized beats are first processed by a Transformer encoder equipped with positional encoding and multi-head self-attention mechanisms. This encoder projects each input sequence into a latent space parameterized by a mean(μ) and standard deviation (σ), as shown in the "Latent Space" component of Figure 1. The latent representations are then sampled and passed to a decoder for signal reconstruction, as well as to a classifier for beat categorization. Training the model involved jointly optimizing three loss functions:

- Reconstruction Loss: Optimizes the regeneration of the original input from the latent space;

$$\text{Reconstruction Loss} = \frac{1}{M}\sum_{i=1}^{M}\text{Binary crossentropy}(x^{(i)}\hat{x}^{(i)}) \tag{1}$$

- Kullback-Leibler (KL) Divergence Loss: Regularizes the structure of the latent space, encouraging meaningful representations;

$$\text{KL Divergence Loss} = -\frac{1}{2}\sum_{j=1}^{D}\left(1 + z_{\log_var_j} - z^2_{\text{mean}_j} - e^{z_{\log_var_j}}\right) \tag{2}$$

- Classification Loss: Guides the correct assigning of classes labels based on latent features.

$$\text{Classification loss} = \frac{1}{M}\sum_{i=1}^{M}\text{Sparse categorical crossentropy}(y_i, \hat{y}_i) \tag{3}$$

These components are combined into a total loss function given by:

$$\text{Total Loss} = \text{Reconstruction Loss} + \beta \times \text{KL Divergence loss} + \gamma \times \text{Classification Loss}, \tag{4}$$

where β and γ are weighting coefficients that balance unsupervised and supervised objectives. As shown in Figure 1, gradients are propagated jointly through the encoder, decoder, and classifier during training, optimizing all components simultaneously.

The dataset was randomly divided into training (85%) and testing (15%) subsets, as shown in Figure 1, following an intra-patient split to ensure robust evaluation. Table 1 presents a summary of the hyperparameters used.

Table 1. Model Hyperparameters

Hyperparameter	**Value**	**Description**
Number of Epochs	25	Number of training iterations
β (beta)	0.001	Weight of the KL divergence term
γ (gamma)	2.0	Weight of the classification loss term
Input Shape	140	Length of each beat sample
Latent Dimension	3	Size of the latent space
Learning Rate	0.001	Step size of the optimizer

3.4 Evaluation Metrics

Model performance is assessed using quantitative metrics and diagnostic visualizations, as outlined in the final stage of Figure 1. Predictions are generated in 10 batches, and for each batch, the model reports the loss components and overall accuracy. Global performance is measured by aggregating accuracy of all batches, while per-batch metrics provide insight into model consistency and potential overfitting.

A detailed visual analysis is conducted using confusion matrices for each class, which help quantify misclassifications and identify any class-specific biases. Additionally, the structure of the learned latent space is visualized using both unsupervised and supervised dimensionality reduction techniques, with encoder outputs projected into two or three dimensions and color-coded by class label.

4 Experimental Results

The model achieves a test accuracy of 99.19% (Table 2) on the MIT-BIH Arrhythmia Database, with a confusion matrix (Figure 2) demonstrating strong class separation for normal beats (N) and ventricular ectopic beats (V). Figure 3 presents the training loss curves across 25 epochs. The model demonstrates stable convergence, with a consistent decline in KL Divergence Loss.

Table 2. Overall Statistics

Metric	Value
Loss	0.7472
Reconstruction Loss	0.6856
KL Loss	2.5905
Classification Loss	0.0295
Accuracy	0.9919
Overall Recall	0.6425

However, the model fails to reliably classify supraventricular ectopic beats (S), showing significantly higher misclassification rates for this class. This limitation aligns with known clinical challenges in ECG interpretation of S-type beats. Batch-wise evaluation confirms the model's stability across test subsets, though the persistent S-class errors suggest a fundamental classification weakness for this arrhythmia type.

The intrinsic structure of the learned representations is shown in Figure 4, which examines the first three dimensions of the VAE's latent space. While the raw projection reveals a natural organization of cardiac beats by electrophysiological properties, the separation between classes is not entirely distinct. Ventricular ectopic beats (light blue) form partially isolated regions, suggesting the encoder captures some pathological patterns, while supraventricular beats (brown), despite their limited sample size, tend to aggregate near normal beats (blue), consistent with their shared QRS morphology. The overlapping boundaries between clusters, particularly for supraventricular beats, indicate that the unsupervised latent space encodes intermediate physiological relationships. This organization confirms the VAE's ability to learn medically meaningful features without explicit labels.

Figure 5 presents the 2D UMAP projection of the learned latent space, showing clustered but partially

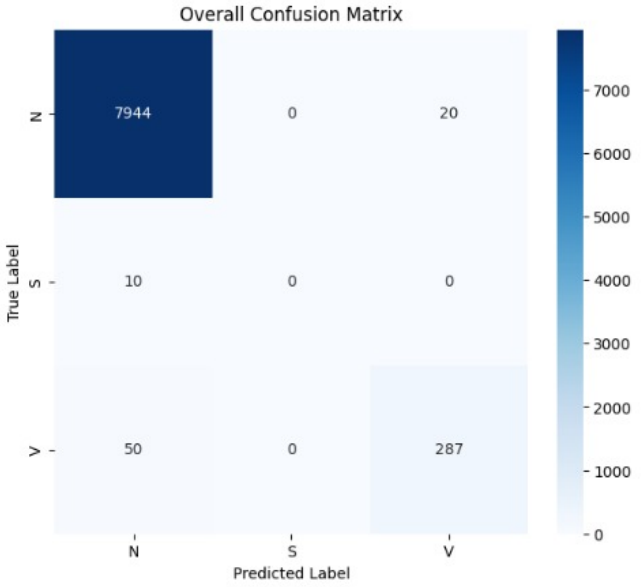


Figure 2. Confusion Matrix

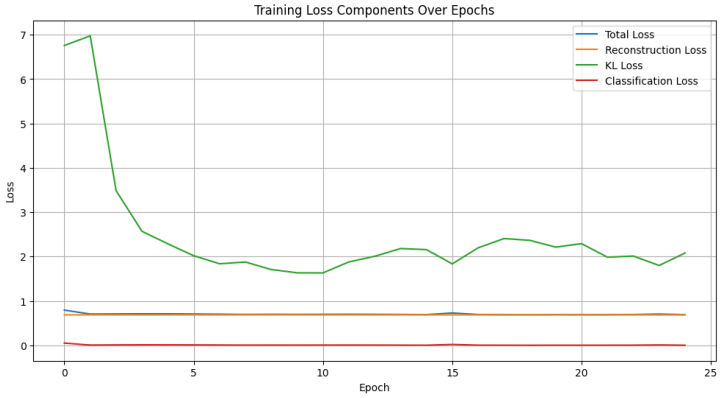


Figure 3. Training Loss Behaivor

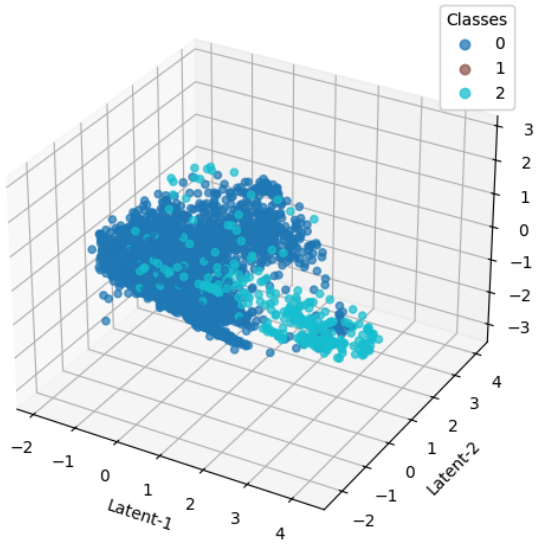


Figure 4. 3D unprocessed Latent Space Visualization

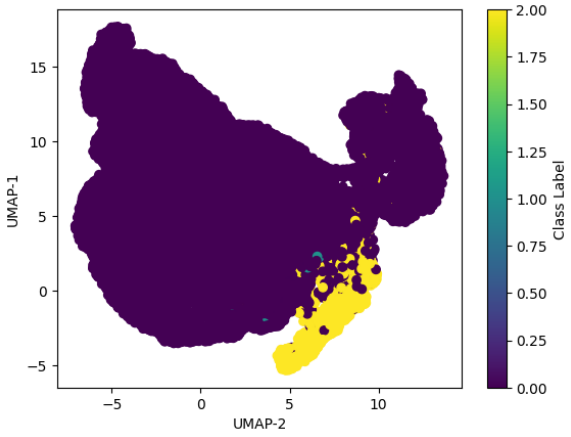


Figure 5. 2D Umap Enhanced Latent Space Visualization

overlapping arrhythmia distributions. While normal beats (purple) and ventricular ectopic beats (yellow) form discernible concentrations, their transitional zones suggesting gradual electrophysiological variations. Supraventricular beats (blue) occupy an intermediate position in latent space, forming a diffuse continuum between normal and ventricular ectopic clusters. This bridging pattern reflects their variable morphological similarity to normal sinus rhythms, and occasional pathological features that approach ventricular ectopic characteristics. The lack of clear demarcation suggests the model encodes the clinical reality where supraventricular arrhythmias often represent borderline cases in ECG diagnostics. Notably, the persistence of these overlaps reflects the biological spectrum of cardiac rhythms, where clear-cut boundaries rarely exist in clinical practice. The projection nevertheless maintains clinically meaningful organization, with ventricular ectopic beats remaining relatively distinct due to their pathological signatures. Figure 6 shows the latent space in 3D using unsupervised UMAP. This nonlinear reduction preserves geometric relationships while revealing arrhythmia subtypes; normal beats (blue) aggregate into partially overlapping regions with supraventricular beats (green), while ventricular ectopic beats (red) occupy relatively distinct, though not fully isolated—zones.This projection preserves the pathological relationships evident in the raw latent space, confirming that the VAE encodes a spectrally organized representation. Spatial distances remain clinically interpretable, with ventricular beats showing the greatest separation from other groups.

Figure 7 illustrates how supervised UMAP (target weight=0.7) reshapes the latent space while preserving physiological relationships. Ventricular ectopic beats (purple) coalesce into denser, more defined regions compared to the unsupervised projection, though still connected to other clusters through transitional zones. Supraventricular beats (yellow) maintain their bridging position between normal (red) and ventricular groups.

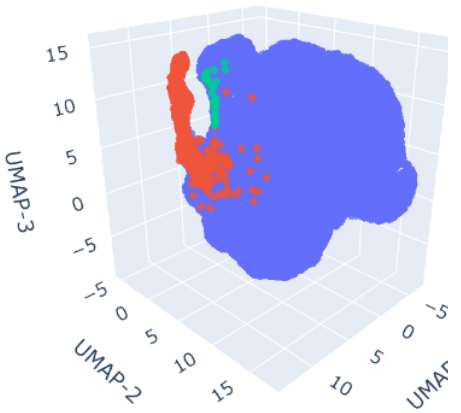


Figure 6. Unsupervised 3D UMAP Projection of VAE Latent Space

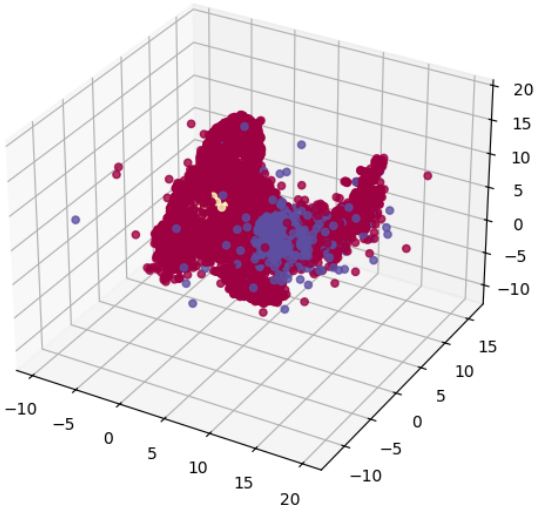


Figure 7. Supervised 3D UMAP Projection

Generated ECG samples (Figure 8) maintain realistic wave morphologies, though with smoothed waveforms compared to raw data. Reconstructed signals preserve key diagnostic features, with a mean reconstruction loss of 68.54%, suggesting the model struggles to create similar signals.

5 Conclusions and future research

The Transformer-Supervised VAE model presents a robust hybrid framework for ECG analysis, achieving 99.04% test accuracy and clinically interpretable latent representations. While the model distinguishes normal beats (N) and ventricular ectopic beats (V) with excellence, its inability to reliably classify supraventricular ectopic beats (S) mirrors real-world diagnostic challenges, as evidenced by both confusion matrices and latent space topology. The 3D latent visualization reveals pathological clustering, with some overlap between classes, with S-beats overlapping normal beats due to morphological similarities—a finding that aligns with cardiological knowledge. Reconstruction quality (mean loss: 68.54%) indicates limitations in waveform fidelity, though generated samples retain diagnostically critical P-QRS-T features.

These results suggest that while the architecture effectively captures global ECG patterns, future work should address S-beat classification and better wave reconstruction.

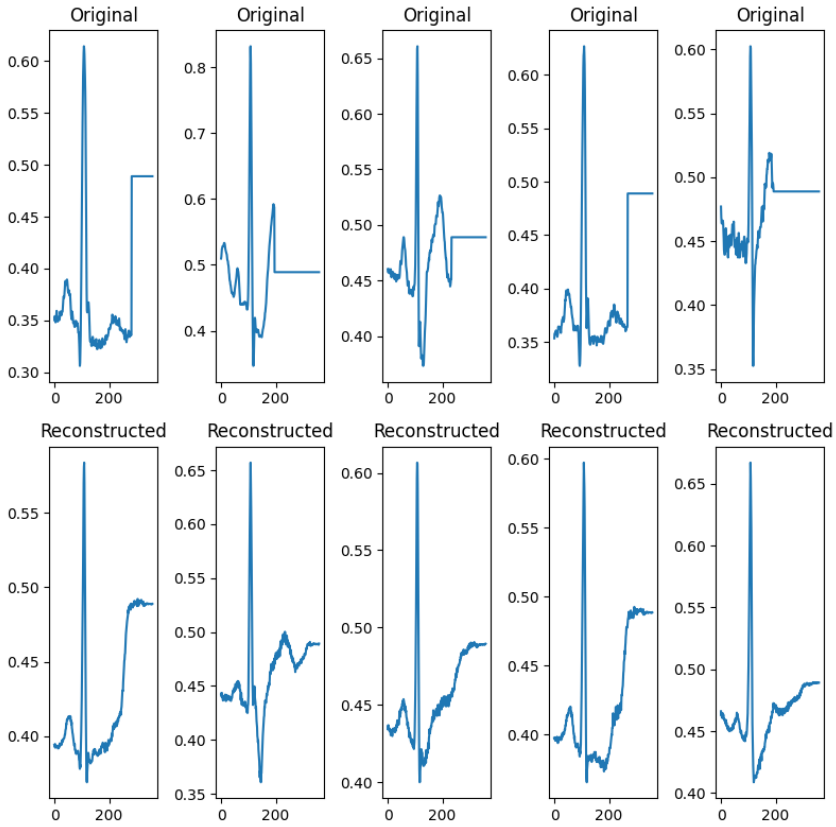


Figure 8. Original and Reconstructed ECG signals

Acknowledgements. The authors would like to express their sincere thanks to the Universidade Aberta Capixaba (UnAC), a program of the Government of the State of Espírito Santo, linked to the Secretaria de Estado da Ciência, Tecnologia, Inovação e Educação Profissional (Secti), for the institutional and financial support (T.O. 1064/2023) that enabled the realization of this study. We also thank the Federal Institute of Espírito Santo (Ifes) for the infrastructure provided.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] A. Mincholé, J. Camps, A. Lyon, and B. Rodríguez. Machine learning in the electrocardiogram. *Journal of Electrocardiology*, vol. 57, pp. S61–S64, 2019.

[2] M. Wasimuddin, K. Elleithy, A.-S. Abuzneid, M. Faezipour, and O. Abuzaghleh. Stages-based ecg signal analysis from traditional signal processing to machine learning approaches: A survey. *IEEE Access*, vol. 8, pp. 177782–177803, 2020.

[3] S. Sahoo, M. Dash, S. Behera, and S. Sabut. Machine learning approach to detect cardiac arrhythmias in ecg signals: A survey. *IRBM*, vol. 41, n. 4, pp. 185–194, 2020a.

[4] V. Kuznetsov, V. Moskalenko, D. Gribanov, and N. Zolotykh. Interpretable feature generation in ecg using a variational autoencoder. *Frontiers in Genetics*, vol. 12, pp. 638191, 2021.

[5] X. Tang, J. Berquist, B. A. Steinberg, and T. Tasdizen. Hierarchical transformer for electrocardiogram diagnosis. arXiv preprint arXiv:2411.00755v1, 2024.

[6] S. Sahoo, M. Dash, S. Behera, and S. Sabut. Interpretable ecg feature extraction by disentanglement of the vae latent space. *IRBM*, vol. 41, n. 4, pp. 185–194, 2020b.