

An Attention-Augmented VAE-BiLSTM Framework for Anomaly Detection in 12-Lead ECG Signals

Marc Garreta Basora^{1*} and Mehmet Oguz Mulayim^{2,1}

¹*Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès, Spain*

²*Artificial Intelligence Research Institute (IIIA-CSIC), Cerdanyola del Vallès, Spain*

Abstract

Anomaly detection in 12-lead electrocardiograms (ECGs) is critical for identifying deviations associated with cardiovascular disease. This work presents a comparative analysis of three autoencoder-based architectures: convolutional autoencoder (CAE), variational autoencoder with bidirectional long short-term memory (VAE-BiLSTM), and VAE-BiLSTM with multi-head attention (VAE-BiLSTM-MHA), for unsupervised anomaly detection in ECGs. To the best of our knowledge, this study reports the first application of a VAE-BiLSTM-MHA architecture to ECG anomaly detection. All models are trained on normal ECG samples to reconstruct non-anomalous cardiac morphology and detect deviations indicative of disease. Using a unified preprocessing and evaluation pipeline on the public China Physiological Signal Challenge (CPSC) dataset, the attention-augmented VAE achieves the best performance, with an AUPRC of 0.81 and a recall of 0.85 on the held-out test set, outperforming the other architectures. To support clinical triage, this model is further integrated into an interactive dashboard that visualizes anomaly localization. In addition, a performance comparison with baseline models from the literature is provided.

Keywords: Anomaly Detection, ECG, Autoencoder, Variational Autoencoder, Attention Mechanism, Unsupervised Learning, Visualization Dashboard

1 INTRODUCTION

Anomaly detection (AD) refers to the process of identifying patterns that deviate from an expected or normal behavior in the data [1]. The importance of this data identification lies in the fact that these variations, known as anomalies, may lead to critical actionable information [2]. For instance, AD plays a crucial role in the healthcare domain, as they can, when properly processed, indicate potential diseases or critical health events in patients. More specifically, cardiovascular diseases (CVDs) are the leading cause of death globally [3], accounting for over one third of all deaths every year [4]. Early detection of cardiac issues is therefore essential, as it improves patients' quality of life by providing early warning of upcoming health events, reduces the economic burden on healthcare systems, and even saves lives [5]. Among various diagnostic tools, the electrocardiogram (ECG) is one of the most common, non-invasive methods that is used as a diagnostic tool. Because it records key information such as heart rhythm, heart rate and cardiac axis information [6], the ECG is an important data for early recognition of various cardiac conditions such as coronary artery disease (CAD), heart failure (HF), arrhythmia (ARR), and other heart diseases.

This time-series signal captures the electrical activity of the heart, reflecting how electrical impulses propagate through cardiac tissues and can be detected via electrodes placed on the skin [7]. In this work, we will focus on the 12-lead ECG, which provides different information from various parts of a patient's body.

Regardless of its diagnostic importance, interpreting multi-lead ECG samples is time-consuming, making even trained physicians misclassify subtle variations that can lead to health diseases [5, 8]. This limitation motivates the development of an automated approach that can assist medical professionals by reducing the time of detection, as well as possibly enhancing the accuracy of identifying cardiac anomalous patterns from ECG data.

To understand these anomalies, it is important to understand the fundamental ECG waveform components. As shown in Figure 1, a normal ECG consists of different intervals and waves that represent the electrical activity of the heart. First, there is the P-wave, continued by the QRS complex, which reflects the contraction of the ventricles when the heart pumps. After that, the T-wave indicates the heart recovery. There are also intervals such as PR, QT or TP, which provide information about the electrical signals flow.

In recent research, machine learning (ML) has been applied to anomaly detection across various domains. In the healthcare context, ML models show strong capabili-

* Corresponding author: Marc.GarretaB@autonoma.cat

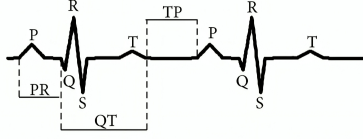


Fig. 1: ECG morphology of two normal beats. Reproduced from Zhang et al. [9] with permission.

ties in processing large volumes of physiological multivariate time-series data, enabling those architectures to learn complex patterns that may be difficult to detect and time-consuming through manual professional inspection. In particular, unsupervised and self-supervised approaches are gaining prominence due to the scarcity of labeled data in this domain.

Among the proposed ML architectures, variational autoencoder (VAE)-based architectures are one of those that have proved effective in distinguishing between normal and abnormal 12-lead ECG recordings [10, 11]. Additionally, adding attention mechanisms into these architectures has also demonstrated significant benefits in other anomaly detection application domains [12].

Particularly, this work focuses on autoencoder-based architectures to detect anomalies in multivariate ECG time-series data. Specifically, three models are implemented and compared for the task: the Convolutional Autoencoder (CAE), the Variational Autoencoder with Bidirectional Long Short-Term Memory (VAE-BiLSTM), and the VAE-BiLSTM with Multi-Head Attention (VAE-BiLSTM-MHA).

Contributions

This paper, to the best of our knowledge, proposes the first application of an architecture that integrates Multi-Head Attention into a VAE-BiLSTM framework for 12-lead ECG anomaly detection. The contributions of this work are as follows:

- We present a novel approach to 12-lead ECG anomaly detection by using a VAE-BiLSTM-MHA framework, which integrates multi-head attention into a variational autoencoder model.
- We conduct a comparative analysis with custom implementations of two other autoencoder-based architectures: CAE and VAE-BiLSTM.
- We develop an interactive dashboard that provides visually interpretable results based on model decisions.
- We provide all code publicly for further research.

Document Structure

The remainder of this paper is structured as follows: Section 2 briefly surveys classical and deep learning techniques for ECG anomaly detection. Section 3 describes the methodology followed, including the problem formulation in Section 3.1, the preprocessing pipeline in Section 3.2, the model architectures in Section 3.3 and the anomaly detection strategy in Section 3.4. Section 4 presents the experimental setup and results, covering the used datasets in

Section 4.1, the applied evaluation metrics in Section 4.2, the results in Section 4.3, and the interactive dashboard in Section 4.4. Finally, Section 5 summarizes the conclusions and outlines future work.

2 STATE OF THE ART

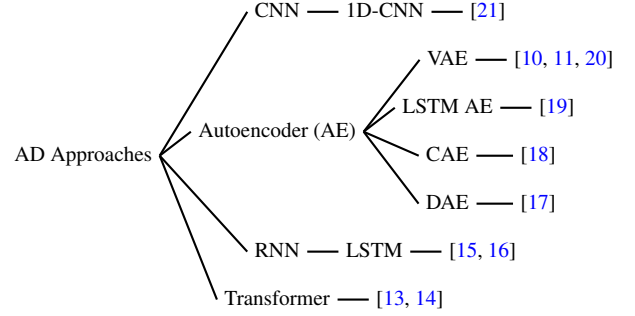


Fig. 2: Approaches for Anomaly Detection (AD) in ECGs

The field of anomaly detection in healthcare has evolved over time. Early methods relied on rule-based systems [22] and small-scale medical corpora, which were limited when handling the complexity of multivariate time-series data such as ECG or electroencephalogram (EEG) signals. With the increase in data from electronic health records (EHRs) [23] and advances in deep learning techniques, these latest approaches have overcome the previous ones.

This literature review focuses mostly on unsupervised and self-supervised learning as these methods are considered to tackle the lack of labeled anomalous data in ECG databases [24, 25].

Figure 2 illustrates the anomaly detection approaches that used ECGs as data. Gu et al. [21] design a CNN with depthwise convolutions and 8-bit quantization, specifically adapted for Field-Programmable Gate Array (FPGA) deployment in wearable devices. Their implementation processes 4-second ECG windows in real time, achieving 97.69 % accuracy on the MIT-BIH [26] arrhythmia dataset and demonstrating that high-fidelity anomaly detection pipelines can be applied under power and area constraints.

Recurrent architectures have also been widely explored for ECG anomaly detection. Chauhan and Vig [15] apply a deep Long Short-Term Memory (LSTM) to raw ECG time signals, learning temporal representations that identify irregular beats with high detection accuracy. Additionally, Thill et al. [16] propose a stacked LSTM approach that models the multivariate prediction-error distribution across the ECG leads via a Gaussian model, achieving high recall with low false-alarm rates on the MIT-BIH arrhythmia dataset.

Autoencoder (AE)-based approaches have shown promising results regarding the reconstruction of normal ECG morphology and detection of anomalies through reconstruction errors. Particularly, variational autoencoders (VAEs) have demonstrated to be effective for ECG anomaly detection [10, 11, 20]. Liu et al. [10] used a vector quantized VAE (VQ-VAE) to perform synthetic data augmentation in order to classify between several cardiac

anomalies. Jang et al. [11] applied a convolutional autoencoder trained with unlabeled data to extract ECG features. Atamny et al. [20] employed a variational autoencoder, which outperformed the other unsupervised models they evaluated. However, the integration of multi-head attention mechanisms within VAE-based models remains unexplored in the context of ECG data. This apparent gap is one of the motivations of this study to explore this autoencoder variant within other architectures, providing a comparative analysis of its performance against similar approaches.

Among other autoencoder approaches, several studies are relevant. Lomoio et al. [18] propose a 1D convolutional autoencoder trained on synthetic ECG segments to learn “normal” patterns, reporting AUROC (Area Under the ROC Curve) of 97.82 % on simulated data and a AUROC of 0.80% on the CPSC-2018 12-lead ECG test set [27]. They also provide reconstruction-error heatmaps over input data for explainability, validated against cardiologist annotations. Choi et al. [19] introduce a segment-wise LSTM autoencoder that processes PreQ, QRS, and PostS intervals separately—corresponding to atrial conduction, ventricular depolarization, and ventricular repolarization phases of the heart’s cycle, respectively—achieving AUROCs up to 0.96 per segment and an overall Atrial Fibrillation (AF) detection AUROC of 0.98 when the three anomaly scores are fused via an XGBoost classifier, and report a AUROC of 0.74% on the CPSC-2018 12-lead ECG test set. Hribar and Torkar [17] develop a denoising autoencoder for 12-lead ECG that removes the need for band-pass filters—which limit the frequency of the ECG signal—and notch filters—which remove narrow-band interference—, attaining 0.81 accuracy and 0.74 recall on the PhysioNet/CinC 2021 challenge dataset [28], with saliency overlays pinpointing the temporal origins of anomalies. In a comprehensive comparative study, Atamny et al. [20] benchmark standard AEs, VAEs, diffusion models, normalizing flows, and Gaussian mixture models on the CPSC-2018 12-lead ECG challenge, finding the VAE as the best performant with AUROC = 0.83 while even the simplest AE achieves AUROC = 0.76.

Regarding the utilization of VAE-based models in other domains, Fu et al. [29] proposed OmniAnomaly, a VAE-based architecture enhanced with adversarial training and probability reconstruction generation in industrial sensor streams. Correia et al. [12] introduced MA-VAE, which integrates multi-head attention into the VAE framework, showing promising results in complex temporal datasets in Automotive Endurance Powertrain Testing.

Transformer-based architectures are also used in the task of anomaly detection and, this variety of models excel at modeling multivariate time-series data-ECG data-by capturing both temporal and spatial dependencies between multiple leads and its duration-timesteps. For instance, Hu et al. [13] presents a hybrid CNN-Transformer network that achieves state-of-the-art arrhythmia classification on single-lead data by combining local feature extraction with global self-attention. Similarly, Kim et al. [14] introduces S-transform-augmented CNN-Transformer that preprocesses the input into a time-frequency representation before attention pooling, producing an improvement when detecting subtle waveform anomalies. In parallel, Tuli et al. [30] proposes TranAD, a Transformer-based model that captures long-term dependencies for robust anomaly

detection.

Motivated by the high performance of autoencoder-based ECG anomaly detection systems, as explained above, the absence of any study implementing multi-head attention in VAE architectures for ECG data emphasizes the novelty of this work and encourages its comparative evaluation against similar models.

3 METHODOLOGY

This section describes the main steps taken to design and implement the anomaly detection systems for 12-lead ECG signals, including the formulation of the problem, data pre-processing, model development, and the anomaly detection process.

3.1 Problem Formulation

A multivariate time series is a sequence of data points along m dimensions. In the context of ECG signals, each dimension corresponds to one of the 12 leads, resulting in $m = 12$. Therefore, a 12-lead ECG dataset can be represented as a multivariate time series:

$$\mathcal{T} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}, \quad \mathbf{x}_t \in \mathbb{R}^m \quad [30] \quad (1)$$

where each observation $\mathbf{x}_t \in \mathbb{R}^{12}$ is composed by all 12 leads at time t .

In this unsupervised learning setting, the autoencoder-based models are trained exclusively on normal samples $\mathcal{T}$ to basically learn a compact representation of the healthy signal manifold and to overcome the imbalance between normal and anomalous ECG samples in clinical corpora [24, 25]. During inference or testing, given an unseen sample $\hat{\mathbf{x}}_t$, the task is to compute how much a 12-lead ECG test sample deviates from the learned representation to decide whether that sample diverges from the normal-signal manifold or not. That is, if the unseen sample lies too far from the normal representation, it could be considered an anomaly.

To measure this difference, an anomaly score S_t is defined, which is compared against a threshold τ to assign a binary anomaly label y_t :

$$y_t = \begin{cases} 1, & \text{if } S_t > \tau, \\ 0, & \text{otherwise,} \end{cases} \quad [15] \quad (2)$$

The value of S_t is computed per each window along each of the 12th leads, so at the end, you end up with several scores per lead that are averaged to obtain a single score per sample (See 3.2.2 for more details regarding the Windowing process).

The value of τ is not fixed but estimated on the validation split that contains only normal ECG recordings. Four strategies for deciding the adequate threshold between a normal representation and an anomalous sample are explored:

1. **95th percentile.** It is an unsupervised rule that its threshold is fixed at the 95th percentile of the validation scores from normal samples.
2. **F₁-optimisation.** This approach keeps the value τ that maximises the F₁ score during validation.

3. **Youden’s J statistic** [31]. This technique optimizes the Receiver Operating Characteristic (ROC) curve during validation, selecting the threshold that maximizes $J = \text{TPR}(\text{TruePositiveRate}) - \text{FPR}(\text{FalsePositiveRate})$.
4. **Peaks-Over-Threshold (POT)** [29, 32]. It uses one of the above techniques to choose a baseline threshold u . Then, a Generalised Pareto Distribution [33] is fit to the extreme tail of the validation scores to set τ .

In this study, the above procedure is repeated for every architecture (i.e., CAE, VAE-BiLSTM, VAE-BiLSTM-MHA), obtaining a model-specific threshold that is used during testing with the goal of minimizing the “false alarm” [34] principle, a concept used in the medical domain that occurs when a system erroneously classifies a normal sample as anomalous.

3.2 Data Preprocessing

3.2.1 Input Data

Two ECG databases are used to train and validate each of the proposed models. Data curation was performed over all recordings labeled as Sinus Rhythm—normal heartbeats—from both datasets to ensure a refined and combined dataset containing only normal samples. The first dataset, PTB-XL [35], comprises twelve-lead recordings from PhysioNet [36], from which it was either cropped or padded to have an input data of ten-seconds segments to guarantee consistency within the second dataset. Each ECG sample is re-sampled at 500 Hz, providing 5,000 data points per lead, and a total of 8,900 normal samples were used with a 80/20 training and validation split (See Table 1). The second dataset, MIMIC-IV ECG [37], was introduced to perform real-world data augmentation after observing that the used architectures would perform better given a high-quality and high-quantity of ECG normal samples. After resampling all recordings to ten-second segments and filtering for valid normal rhythms, over 92,000 samples—labeled as Sinus Rhythm—were selected from more than 800,000 total samples available to include only those ECGs labeled uniquely as non-anomalous, excluding other data with additional diagnoses or ambiguous labeling systems. As shown in Figure 3, each sample consists of twelve-leads—each capturing ten seconds of the heart’s electrical activity.

Characteristic	PTB-XL	MIMIC-IV ECG
Total ECGs	21,799	> 800,000
Sinus rhythm ECGs	$\approx 8,900$	$\approx 92,000$
Number of patients	18,869	> 160,000
Access	Public	Restricted

TABLE 1: Input datasets overview

3.2.2 Window Segmentation

Window segmentation is used during the experiments proposed in Section 4. This sample partition technique serves as a way to capture localized patterns within large ECG signals [15]. The training data set is matched by overlapping

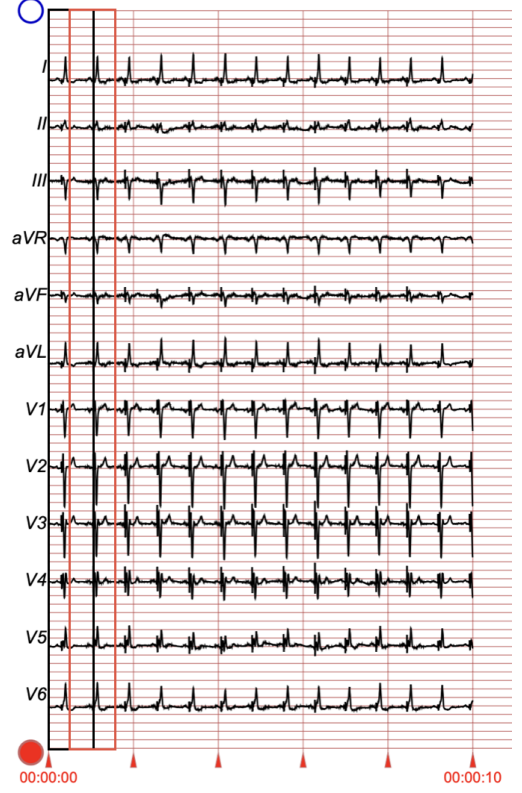


Fig. 3: Example of a raw 12-lead ECG sample from the MIMIC-IV ECG dataset. The red and black boxes show consecutive windows extracted for training.

windows W , where each window represents a fixed-length segment of the ECG signal:

$$W = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}, \quad \mathbf{w}_i \in \mathbb{R}^{L \times m} \quad (3)$$

where L is the number of ECG leads and each window $\mathbf{w}_i$ is an $L \times m$ matrix containing a fixed-length segment from all leads. The total number of windows per sample n is

$$n = \left\lfloor \frac{T - m}{s} \right\rfloor + 1, \quad s < m \quad (4)$$

In this case, T is the total length of the recording, m the window length (e.g., 500 samples), and s the hop size (e.g., 250 samples for 50 % overlap). As shown in Figure 3, the black and red-bordered rectangles over the first two seconds of an ECG recording illustrate how two windows are created and how the window segmentation is applied.

3.2.3 Filtering and Normalization

After reviewing the literature and experimenting with different filtering and normalization configurations, the following techniques were chosen as optimal:

1. ECG signals are cleaned using a combination of band-pass and notch filters [38, 39, 40]. Specifically, a 3rd-order Butterworth bandpass filter with cutoff frequencies of 0.5 Hz and 100 Hz is applied to remove baseline wander. Then, a notch filter centered at 60 Hz removes power-line interference. These filters are applied locally to each ECG lead to preserve signal quality and

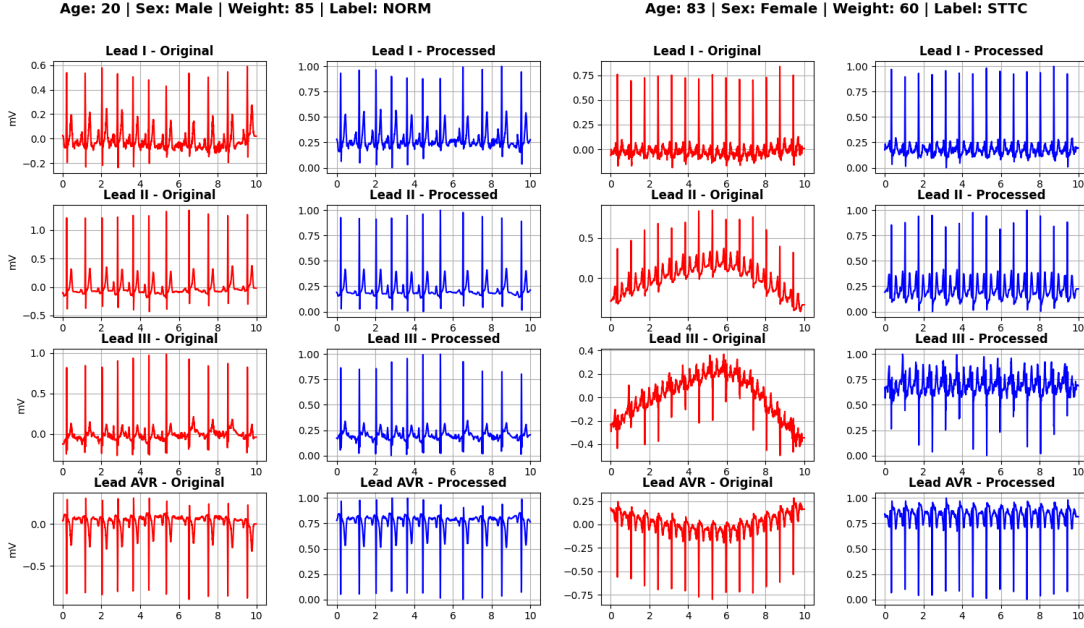


Fig. 4: Comparison of non-processed and preprocessed ECG signals

process them in a personalized way so that each lead is treated accordingly to its frequency distribution.

2. After filtering, the signals are normalized using z-score normalization [41] in order to have a mean of 0 and a standard variation of 1 on a per-lead basis to ensure consistent amplitude scaling across the dataset, which facilitates effective training of machine learning models.

$$x_{t,i}^{(z)} = \frac{x_{t,i} - \mu_i}{\sigma_i + \epsilon} \quad (5)$$

where i is the lead index, and the constant vector ϵ is introduced to the denominator to prevent division by zero when the lead has zero variance.

In Figure 4, the effects of filtering and normalization (in blue) are observable compared to the raw signals (in red). This preprocessing step reduces baseline drift and power-line interference. However, for anomalous samples, the same operations can also mask subtle deviations, consequently increasing the risk of false negatives. For example, in the raw versus preprocessed comparison of the second sample, the low-amplitude ST-segment deviations visible in the raw signals are attenuated by the previously explained processing techniques. Consequently, there is a “trade-off between noise suppression and anomaly/morphology preservation” [42], as processing is an essential step before training a model to input non-noisy data, but it can also delete subtle anomalous patterns, thereby increasing the risk of false negatives.

3.3 Model Architectures

Convolutional Autoencoder (CAE)

The proposed Convolutional Autoencoder, illustrated in Figure 5, is used as the baseline model in our comparative analysis and was adapted from open-source code [18]. It processes twelve-lead ECG windows of shape $[B, 12, T]$,

where B is the batch size, 12 is the number of ECG leads per sample, and T is the window length.

The model consists of two main blocks: an encoder that maps data x into a latent representation z , and a decoder that maps this compressed space z back into the original data representation x' to perform a reconstruction of the original signal. The encoder compresses and reduces the temporal information using convolutional layers, while increasing channel depth. Afterwards, the decoder produces the inverse of this action, reconstructing x' to match the original input shape.

During training, the model minimizes the Mean-Squared Error (MSE) loss between each input sample and its reconstructed version. This loss function penalizes high-amplitude reconstruction deviations more heavily. In a 12-lead ECG, the largest per-sample amplitudes occur within the QRS complex, particularly at the R-peaks, where the signal changes sharply. Consequently, the autoencoder is driven to reproduce these segments with high fidelity, while small deviations in flatter regions (e.g., PR or ST segments) have a minor effect on the total loss.

Variational Autoencoder Bidirectional Long Short-Term Memory (VAE-BiLSTM)

The proposed VAE-BiLSTM model, illustrated in Figure 6, encodes each twelve-lead ECG window of shape $[B, 12, T]$ (batch size B , leads 12, timesteps T)—with a permutation to convert it to the required shape $[B, T, 12]$ —with a bidirectional LSTM encoder block that generates a latent mean vector $\mu_z \in \mathbb{R}^{B \times d}$ and a log-variance vector $\log \sigma_z^2 \in \mathbb{R}^{B \times d}$, where d is the latent dimension. Afterwards, the reparameterization trick¹ is used to convert a latent sample z from a non-differentiable sampling step to a differentiable one. Then, z is then repeated T times

¹The reparameterization trick is used in VAE models to enable gradient propagation with a differentiable transformation expressing a deterministic transformation (see Eq. 6).

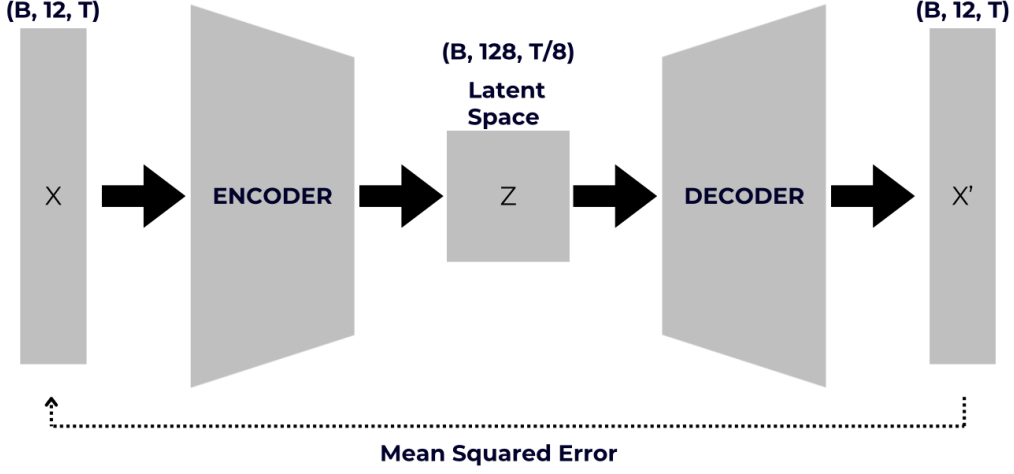


Fig. 5: CAE architecture

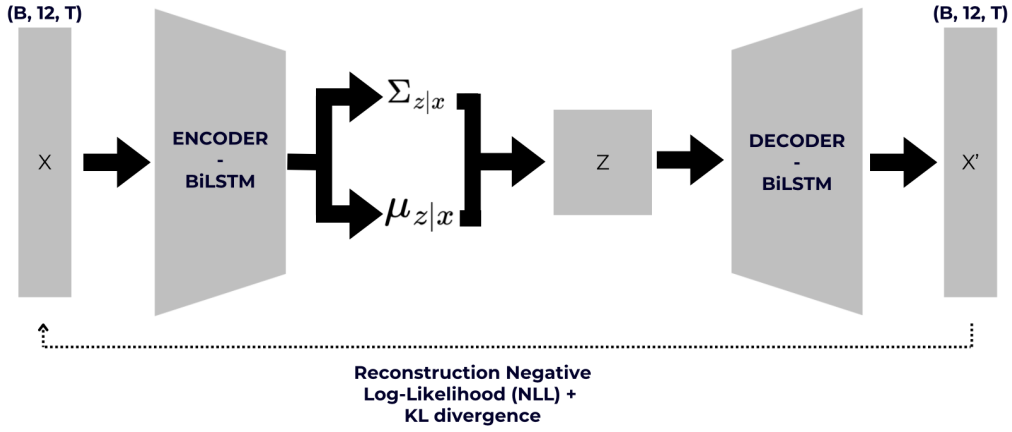


Fig. 6: VAE-BiLSTM architecture

in a loop process and it is used as input to a unidirectional LSTM decoder, which outputs per-lead reconstruction means $\hat{x}_t$ and log-variances $\log \hat{\sigma}_t^2$, with tensors of shape $[B, 12, T]$.

During training, the model minimizes the negative log-likelihood (NLL) of the output plus a Kullback-Leibler (KL) divergence multiplied by an annealed coefficient β . The architecture is inspired by the original VAE formulation of Kingma and Welling [43] and by the OmniAnomaly framework for time-series anomaly detection [29].

This stochastic approach introduces variability in the latent representation, rather than relying on a point-wise reconstruction error—MSE—used in the previous deterministic model.

VAE-BiLSTM with Multi-Head Attention (VAE-BiLSTM-MHA)

Inspired by the original VAE architecture of Kingma and Welling [43], the OmniAnomaly time-series architecture [29], and the MA-VAE design [12], an extension of the previous model, VAE-BiLSTM, is presented in order to incorporate two attention mechanisms: a lead-wise attention to capture inter-lead dependencies and a multi-head attention to enhance the latent sequence representation as illustrated in Figure 7.

Each 12-lead window $\mathbf{x} \in \mathbb{R}^{B \times T \times 12}$ is encoded by two bidirectional BiLSTM layers, producing hidden states $\mathbf{h} \in \mathbb{R}^{B \times T \times 2h}$, where h is the second LSTM's hidden size. A fully connected layer (MLP) maps $\mathbf{h}$ to per-timestep latent mean and log-variance $[\mu_z, \log \sigma_z^2] \in \mathbb{R}^{B \times T \times 2d}$, with latent dimension d . Latent samples are then generated using the reparameterization trick [43]:

$$\mathbf{z} = \mu_z + \exp(0.5 \log \sigma_z^2) \odot \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (6)$$

To add inter-lead correlations, a lead-wise attention block embeds every (B, T) slice of the input leads and applies a 4-head self-attention layer, resulting in lead context vectors $\tilde{\mathbf{h}} \in \mathbb{R}^{B \times T \times d}$. At the end, an enriched representation is obtained given the combination of the lead-wise attention with the latent space $\mathbf{z}^* = \mathbf{z} + \tilde{\mathbf{h}}$.

Then, the latent $\mathbf{z}^*$ is used as *values* in a 8-head attention layer whose *queries* and *keys* are linear projections of the raw input window. This produces a context-aware sequence representation $\mathbf{A} \in \mathbb{R}^{B \times T \times d}$ that now encodes both global latent information and lead-specific saliency.

Finally, $\mathbf{A}$ is passed through two bidirectional LSTMs and an output MLP, providing per-sample reconstruction mean and logvar statistics.

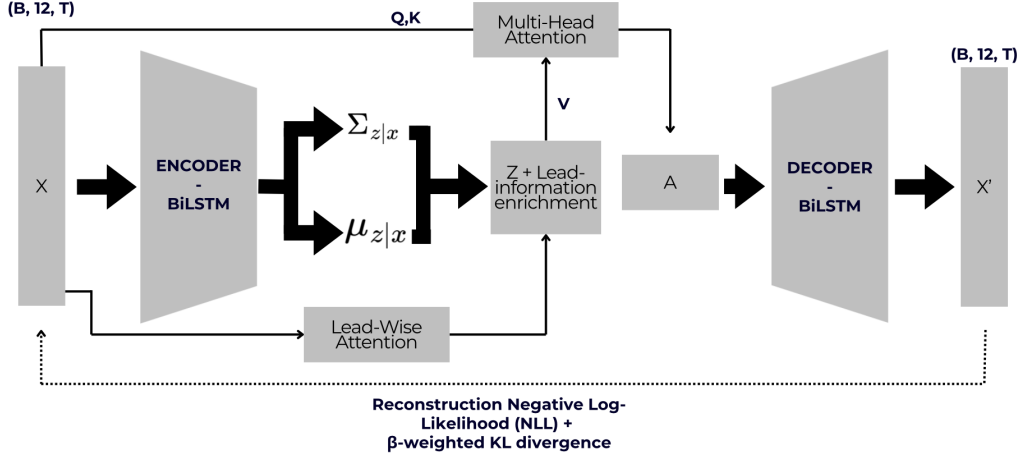


Fig. 7: VAE-BiLSTM with Multi-Head Attention architecture

3.4 Anomaly Detection Task

Given an unseen multi-lead ECG sample of length T , the overlapping window segmentation is applied to obtain n windows per sample (See Section 3.2.2 for the window process):

$$W_x = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}, \quad (7)$$

$$n = \left\lfloor \frac{T-m}{s} \right\rfloor + 1, \quad s < m. \quad (8)$$

where

- T is the total length of the ECG recording in samples,
- m is the length of each window in samples,
- s is the stride between successive windows,
- n is the total number of windows

Each window $\mathbf{w}_i$ is then passed through the trained anomaly detector model, which produces a per-window anomaly score

$$s_i = \mathcal{S}(\mathbf{w}_i) \quad (9)$$

where the scoring function $\mathcal{S}$ depends on the model architecture:

- For the CAE model, the anomaly score for each window is computed as the mean-squared reconstruction error:

$$s_i^{\text{CAE}} = \|\mathbf{w}_i - \hat{\mathbf{w}}_i\|_2^2 \quad (10)$$

where the MSE measures point-wise differences between the input and its reconstruction [44]. Consequently, windows with large reconstruction errors are considered as anomalous signals.

- For the VAE-BiLSTM model, each window's anomaly score is computed by the negative evidence lower bound (ELBO [43]), where the MSE is used as the reconstruction term, which can be derived from maximizing the Gaussian log-likelihood:

$$s_i^{\text{VAE-BiLSTM}} = \|\mathbf{w}_i - \hat{\mathbf{w}}_i\|_2^2 + \text{KL}(q_\phi(z | \mathbf{w}_i) \| p(z)) \quad (11)$$

where the first section measures the MSE reconstruction error and the second section is the KL divergence between the approximate posterior $q_\phi(z | \mathbf{w}_i)$ and the prior $p(z)$, enforcing a regularized latent space. Consequently, a window's high score can lead to an unlikely latent representation and a poor reconstructed sample.

- For the VAE-BiLSTM-MHA model, we define the anomaly score as an attention-weighted ELBO:

$$s_i^{\text{VAE-BiLSTM-MHA}} = \underbrace{\sum_{t=1}^m \alpha_{i,t} \|\mathbf{w}_{i,t} - \hat{\mathbf{w}}_{i,t}\|_2^2}_{\text{Attention-weighted reconstruction}} + \underbrace{\text{KL}(q_\phi(z | \mathbf{w}_i) \| p(z))}_{\text{Latent regularization}} \quad (12)$$

where $\alpha_{i,1}, \dots, \alpha_{i,m}$ considers normalized attention weights produced per each window, the $\|\mathbf{w}_{i,t} - \hat{\mathbf{w}}_{i,t}\|_2^2$ determines the MSE reconstruction error and the other section regularizes the latent space against the prior as in the case of VAE-BiLSTM above.

The attention-weighted reconstruction term is based on the assumption that the self-attention block inside the model “attends” to those time steps where there are relevant variations. Consequently, low-information segments contribute less than in salient zones regarding the MSE. The correlation between high-attention regions and high reconstruction error will be illustrated in the visualizations presented in Section 4.3.

Decision Rule Once each window $\mathbf{w}_i$ has been assigned a score s_i , a single score is obtained for the entire recording by taking the mean:

$$S = \frac{1}{n} \sum_{i=1}^n s_i \quad (13)$$

As explained in Section 3.1, a threshold τ is established during validation. Then, the final anomaly decision is given

by

$$\hat{y} = \begin{cases} \text{anomalous,} & S > \tau, \\ \text{normal,} & S \leq \tau. \end{cases}$$

If the average score S exceeds τ , the signal is labeled as anomalous; otherwise, it is labeled as normal. This simple rule leverages the validation-tuned threshold to balance sensitivity and specificity.

4 EXPERIMENTS

The three proposed autoencoder-based models (CAE, VAE-BiLSTM, and VAE-BiLSTM-MHA) are evaluated against four literature baselines listed below, which also report results on the same dataset. The comparison is made under the closest possible conditions:

- ConvAE [18]
- VAE-AE Hybrid [20]
- AE + Peak Detection [19]
- MSGformer [45]

Below, Tables 2–6 summarize the hyperparameter configurations for each of our proposed models. All parameters were selected manually, aided by visual analysis.

TABLE 2: CAE Hyperparameters

Parameter	Value
Window size	500
Stride	$0.5 \times \text{window size}$ (250)
Learning rate	1×10^{-3}
Number of epochs	100
Criterion	MSE loss
Optimizer	Adam

TABLE 3: VAE-BiLSTM Hyperparameters (excluding β)

Parameter	Value
Window size	500
Stride	$0.5 \times \text{window size}$ (250)
Learning rate	5×10^{-3}
Number of epochs	100
Latent dimension	64
Hidden dimension	128
Criterion	Recon. + KL
Optimizer	Adam

TABLE 4: VAE-BiLSTM-MHA Hyperparameters (excluding β)

Parameter	Value
Window size	500
Stride	$0.5 \times \text{window size}$ (250)
Learning rate	1×10^{-4}
Number of epochs	100
Latent dimension	64
Hidden dimension	128
Number of attention heads	8
Dropout (encoder)	0.1
Gaussian noise (input)	$\sigma = 0.01$
Criterion	Attn. + Recon. + KL
Optimizer	Adam

TABLE 5: β -Annealing Schedule for VAE-BiLSTM

Epoch (t)	β_t
$t \leq 10$	$\beta_t = \frac{t}{10}$
$t > 10$	$\beta_t = 1.0$

TABLE 6: β -Scheduling for VAE-BiLSTM-MHA (cyclical ramp from 10^{-8} to 10^{-2} over epochs [11–100])

Epoch (t)	β_t
$t \leq 10$	$\beta_t = 10^{-8}$
$10 < t \leq 100$	$\beta_t = 10^{-8} + \frac{t-10}{90} [10^{-8}, 10^{-2}]$

4.1 Datasets

The primary evaluation dataset is the publicly available CPSC 2018 challenge corpus [27], which comprises 12-lead ECG recordings acquired at 500 Hz for a nominal duration of 10 s per patient. The dataset is balanced in terms of gender (approximately 50 % female, 50 % male). For the binary anomaly detection task, each recording has been re-labeled as either *Sinus Rhythm* (i.e., normal) or *Anomalous*, obtaining over 2,000 ECG samples—each recording from a unique patient—resulting to over 2,000 patients [46].

The preprocessing pipeline remained identical to that used for the MIMIC-IV ECG data, except for the notch filter, whose center frequency was lowered from 60 Hz to 50 Hz to match the mains frequency of the region in which the new dataset was recorded.

To ensure consistency with the training preprocessing and windowing (Section 3.2.3), the testing set passed through identical preprocessing steps—bandpass filtering to remove baseline wander and power-line interference, followed by z-score normalization per lead—except for the notch filter, in which case its maximum frequency was lowered from 60 Hz to 50 Hz to match the main frequency of the region in which the CPSC dataset was recorded. Additionally, a selection of 10 s duration segments at 500 Hz was performed. This curation and preprocessing pipeline ensures that both training and test sets follow a standard

TABLE 7: Performance Comparison of the Proposed and (cited) Baseline Models on the CPSC Dataset

Model	Prec.	Rec.	F1	AUROC	AUPRC
CAE	0.64	0.82	0.72	0.77	0.80
VAE-BiLSTM	0.70	0.76	0.73	0.78	0.81
VAE-BiLSTM-MHA	0.75	0.85	0.80	0.80	0.81
ConvAE [18]	—	—	0.79	0.80	—
VAE-AE Hybrid [20]	—	—	—	0.65	—
AE + Peak Detection [19]	—	0.56	0.56	0.74	—
MSGformer [45]	—	—	0.847	0.88	—

preprocessing approach, as used in several research articles [38, 39].

4.2 Evaluation Metrics

To evaluate the performance of the proposed anomaly detection model, a set of standard metrics commonly used in anomaly detection literature was selected, including precision, recall, F1 score, AUPRC, and AUROC.

Precision, Recall, and F1 Score: The primary evaluation metrics are precision (P), recall (R), and F1 score (F_1):

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F_1 = 2 \frac{PR}{P + R} \quad (14)$$

where TP , FP , and FN are the number of True Positives (correctly detected anomalies), False Positives (normal samples incorrectly labeled as anomalies), and False Negatives (missed anomalies), respectively.

AUPRC: The Area Under the Precision-Recall Curve determines the model’s ability to correctly detect anomalies (TP) while avoiding false alarms (FP).

AUROC: The Area Under the Receiver Operating Characteristic Curve evaluates the trade-off between the true positive rate ($TPR = \frac{TP}{TP + FN}$) and the false positive rate ($FPR = \frac{FP}{FP + TN}$) across different thresholds. That is to say, a higher AUROC value shows a stronger discriminative ability between healthy and anomalous ECG samples.

4.3 Results

Table 7 summarizes the anomaly detection performance of all models evaluated on the CPSC 2018 test set. The results of the proposed models are obtained under a unified evaluation process, whereas the baseline scores are directly extracted from the respective original publications and reflect their reported evaluation conditions. Below we detail the findings of the experiments with each model.

Convolutional Autoencoder (CAE)

The convolutional autoencoder achieves a solid performance on the quantitative anomaly detection task. Additionally, the interpretability of this model is obtained

through the saliency map² using the model’s MSE reconstruction error.

Figure 8 illustrates an original normal ECG signal (black) and its reconstruction (orange) per each lead, with three extra information: the per-window anomaly score (“A”), the saliency heat-map, and the MSE reconstruction error values. As observed in the figure, the highest MSE values and the darkest saliency are located in every QRS. It means that the model has learned to focus on these rapid and large changes in the signals. More importantly, even though these spikes obtain the highest MSE error, the global anomaly score remains below the defined threshold, meaning that the CAE model interprets the spikes within the normal morphology.

Variational Autoencoder BiLSTM (VAE-BiLSTM)

Compared with the CAE model, the VAE-BiLSTM achieves higher F1, AUROC, and AUPRC, since its stochastic latent variables force the model to learn representations that generalize from normal-only training data. However, in early experiments the model suffered from *posterior collapse*, where the approximate posterior collapsed onto the prior as formulated in Equation 15:

$$q_\phi(\mathbf{z} | \mathbf{x}) \approx p(\mathbf{z}) \quad (15)$$

At collapse, the KL divergence term fell to nearly zero, the encoder outputs $\mu_z \approx \mathbf{0}$ and $\log \sigma_z^2 \approx \mathbf{0}$, and the decoder didn’t receive information about the latent space z , converting it into a deterministic auto-encoder (since $\beta \text{KL} \approx 0$).

To solve this problem, a technique called cyclical KL annealing [12] was applied (see Table 6), increasing β over time starting at epoch 11—letting the model also learn a well-reconstructed normal manifold—over a 10-epoch schedule and then repeating the cycle. This technique allowed the model to distribute importance between reconstruction and regularization phases—MSE and KL divergence accordingly. A dropout was also added for further regularization.

Overall, this combined strategy prevented posterior collapse as each latent dimension maintained a minimum of information flow, and the posterior parameters μ_z and $\log \sigma_z^2$ stayed separated from the prior values, thus, the encoder did not collapse to $\mu_z = 0$ and $\log \sigma_z^2 = 0$.

²A saliency map is a visualization that shows which parts of an input the model considers most important.

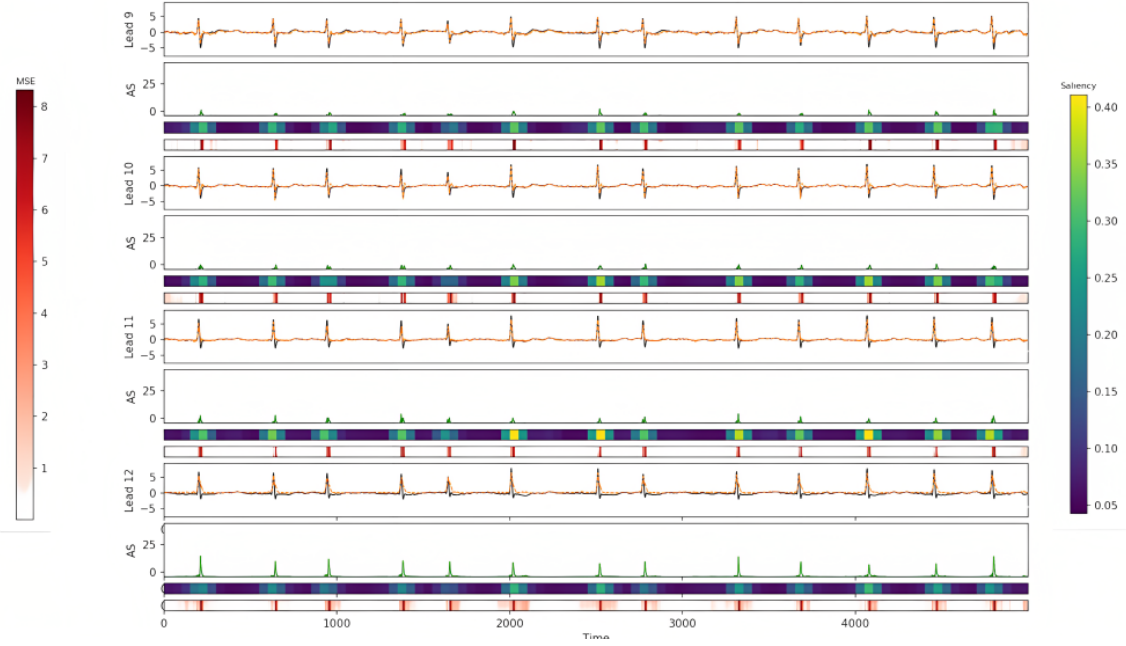


Fig. 8: CAE interpretability of a normal ECG sample

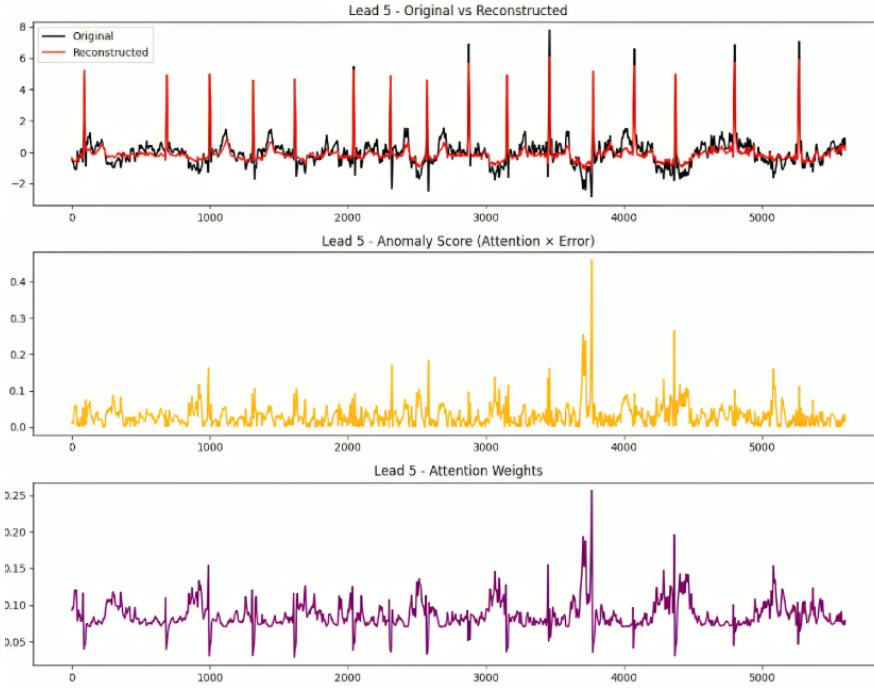


Fig. 9: VAE-BiLSTM-MHA interpretability of an anomalous ECG sample

VAE-BiLSTM with Multi-Head Attention (VAE-BiLSTM-MHA)

The last model achieves improved performance on the quantitative task, showing higher Precision, Recall, F1-score, and AUROC compared to the previous models. Additionally, its interpretability is enhanced through the attention mechanisms. Figure 9 shows a 12-lead anomalous ECG sample (black original vs. red reconstruction), together with the per-window anomaly score in yellow and the attention weights in purple. It illustrates that between time steps 3,500 and 4,000 the model produces higher reconstruction error, higher anomaly scores, and increased at-

tention weights in the same section. This correlation indicates that the model not only struggles to reconstruct the signal in that specific region, but also assigns more importance to that region through its attention mechanism. The combination of reconstruction error and attention weights probably means that the model is highlighting a segment it both finds difficult to reproduce and finds it relevant for the overall result.

Figure 10 focuses on a 500 timestep window of the previous anomalous ECG sample, showing the original vs reconstructed signal, the anomaly score and the attention weights within a heatmap. The peaks in reconstruction error and anomaly score are strongly correlated and they are located

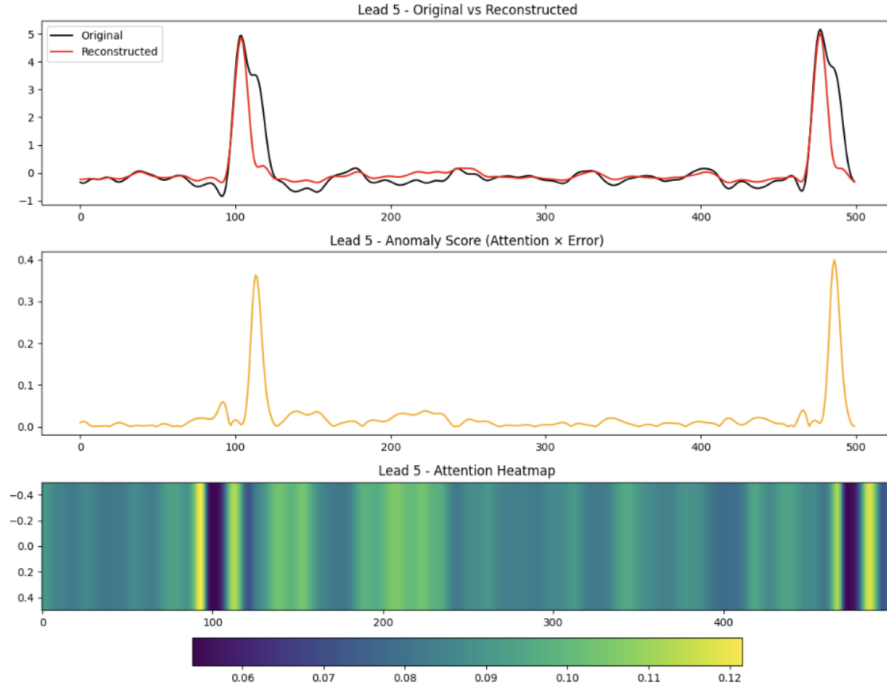


Fig. 10: VAE-BiLSTM-MHA interpretability of Figure 9, focused on a window size of 500 samples

in QRS peaks. Interestingly, the heatmap shows that the model does not focus exactly on the sharp error peaks, but rather on the segments before and after them. This indicates that the model considers the surrounding waveform morphology, rather than focusing only on the sharp peaks. Such behavior leads to a more informative explanation, as subtle variations before and after the QRS complex can indicate anomalous patterns.

4.4 Dashboard

A local web application is implemented as the last objective of this work, it uses the Dash framework within Plotly integration to create visualizations of these last interpretable models—both CAE and VAE-BiLSTM-MHA. As it is observed in Figure 11, the user interface offers a simple and personalized framework where you can upload an ECG signal in standard formats—.npz or .dat/.hea. Afterwards, clicking on “Analyze ECG Sample”, the interface will render the 12-lead ECG signal, predict an anomaly decision, and provide as well as a detailed plot to interpret model’s reasoning.

The dashboard also offers configuration possibilities such as choosing the leads that you wanna visualize, choosing between a CAE and a VAE-BiLSTM-MHA models, and deciding the anomaly threshold—though it is not recommend to alter the default threshold value as it is computed on the validation set (Section 3.1). In the main plot section, there are several displays per each lead. The original signal (black) and the reconstructed signal (red) are found in the first graphic of each lead. Then, there is another graphic that shows the anomaly score (orange) in each timestep. Moreover, there are two horizontal heat-map bars that compute the lead-wise attention weights and the point-wise MSE per leads along time steps.

5 CONCLUSIONS

In this work, a comparative analysis was conducted among three autoencoder-based architectures for anomaly detection in multivariate 12-lead electrocardiogram (ECG) time-series data. Specifically, a convolutional autoencoder (CAE), a variational autoencoder with bidirectional long short-term memory (VAE-BiLSTM), and a VAE-BiLSTM augmented with multi-head attention (VAE-BiLSTM-MHA) architectures were evaluated on the CPSC 2018 dataset. The results demonstrated that the proposed novel approach VAE-BiLSTM-MHA outperformed the other models, achieving an AUPRC of 0.81 and a recall of 0.85. Beyond quantitative performance, this model also exhibited qualitative strengths regarding the interpretability of anomaly diagnosis. An additional contribution of this work was the development of a user-friendly interactive web application that provides a simple way to observe ECG signals and indicate possible anomaly deviations with interpretable graphics, helping users identify and understand cardiac anomalies.

Despite these advances, limitations are present in this study. First, the evaluation on a single benchmark dataset with customized configurations defined within this work may limit the generalizability of the results. Second, the models were trained and evaluated in an offline setting, without facing real-time ECG monitoring challenges. Finally, the absence of medical professionals in the process of model creation and its evaluation restricted the ability to fully assess the clinical interpretability of the models’ outputs.

Future research directions include extending the current VAE-BiLSTM-MHA framework for real-time, online anomaly detection in collaboration with medical professionals, to facilitate deeper clinical insights into ECG data and its anomalies. Such partnerships would foster inno-



Fig. 11: Interactive Dashboard

vation by combining technical expertise with the domain knowledge of healthcare practitioners. Additionally, exploring advanced architectures such as Transformer-based models and integrating additional patient-related data with ECG signals may improve both the accuracy of anomaly detection and the interpretability of diagnostic outcomes.

SOFTWARE AVAILABILITY

The source code has been made publicly available at: <https://github.com/marcgarreta/12-lead-ECG-AD>

ACKNOWLEDGMENTS

The authors would like to extend their special thanks to Universitat Autònoma de Barcelona for providing the computational resources that supported this research.

REFERENCES

- [1] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, “Deep learning for anomaly detection: A review,” *ACM Comput. Surv.*, vol. 54, no. 2, Mar. 2022.
- [2] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation forest,” in *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422.
- [3] E. G. Nabel, “Cardiovascular disease,” *The New England Journal of Medicine*, vol. 349, no. 1, pp. 60–72, 2003.
- [4] W. H. Organization. (2025) Cardiovascular diseases.
- [5] S. B. Dunbar, O. A. Khavjou, T. Bakas, G. Hunt, R. A. Kirch, A. R. Leib, R. S. Morrison, D. C. Poehler, V. L. Roger, and L. P. Whitsel, “Projected costs of informal caregiving for cardiovascular disease: 2015 to 2035: A policy statement from the american heart association,” *Circulation*, vol. 137, no. 19, pp. e558–e577, 2018.
- [6] Y. Sattar and L. Chhabra, “Electrocardiogram,” in *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing, 2023.
- [7] E. A. Ashley and J. Niebauer, *Cardiology Explained*. Remedica, 2004, chapter 3, pp. 15–34.
- [8] D. A. Cook, S.-Y. Oh, and M. V. Pusic, “Accuracy of physicians’ electrocardiogram interpretations: A systematic review and meta-analysis,” *JAMA Internal Medicine*, vol. 180, no. 11, pp. 1461–1471, 2020.
- [9] Y. Zhang, I. Silva, G. B. Moody, L. A. Celi, D. J. Stone, and G. D. Clifford, “Robust and accurate anomaly detection in ecg artifacts using time series motif discovery,” *Journal of Healthcare Engineering*, vol. 6, no. 1, pp. 1–22, 2015.
- [10] F. Liu, H. Deng, Y. Wang, and Z. Wang, “Using the vq-vae to improve the recognition of abnormalities in short-duration 12-lead electrocardiogram records,” *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105906, 2021.
- [11] J.-H. Jang, T. Y. Kim, H.-S. Lim, and D. Yoon, “Unsupervised feature learning for electrocardiogram data using the convolutional variational autoencoder,” *PLoS ONE*, vol. 16, no. 12, p. e0260612, 2021.
- [12] L. Correia, J.-C. Goos, P. Klein, T. Bäck, and A. V. Kononova, “Ma-vae: Multi-head attention-based variational autoencoder for anomaly detection in multivariate time series,” *arXiv preprint arXiv:2309.02253*, 2023.

- [13] R. Hu, J. Chen, and L. Zhou, "A transformer-based deep neural network for arrhythmia detection using continuous ecg signals," *Computers in Biology and Medicine*, vol. 144, p. 105325, 2022.
- [14] D. Kim, K. R. Lee, D. S. Lim, K. H. Lee, J. S. Lee, D.-Y. Kim, and C.-B. Sohn, "A novel hybrid cnn-transformer model for arrhythmia detection without r-peak identification using stockwell transform," *Scientific Reports*, vol. 15, no. 1, p. 7817, 2025.
- [15] S. Chauhan and L. Vig, "Anomaly detection in ecg time signals via deep long short-term memory networks," in *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, Oct. 2015, pp. 1–7.
- [16] M. Thill, S. Däubener, W. Konen, and T. Bäck, "Anomaly detection in electrocardiogram readings with stacked lstm networks," in *CEUR Workshop Proceedings*, vol. 2473, 2019, pp. 1–9.
- [17] R. Hribar and D. Torkar, "Explainable anomaly detection of 12-lead ecg signals using denoising autoencoder," in *Intelligent Secure Trustable Things*, ser. Studies in Computational Intelligence, M. Karner, Ed. Springer Nature, 2024, vol. 1147, pp. 127–140.
- [18] U. Lomoio, P. Vizza, R. Giancotti, S. Petrolo, S. Flesca, F. Boccuto, P. H. Guzzi, P. Veltri, and G. Tradigo, "A convolutional autoencoder framework for ecg signal analysis," *Heliyon*, vol. 11, p. e41517, 2025.
- [19] S. Choi, K. Choi, H. K. Yun, S. H. Kim, H.-H. Choi, Y.-S. Park, and S. Joo, "Diagnosis of atrial fibrillation based on ai-detected anomalies of ecg segments," *Heliyon*, vol. 10, no. 1, p. e23597, 2023.
- [20] O. Atamny, A. Saguner, R. Abaecherli, and E. Konukoglu, "Outlier detection in ecg," in *Proceedings of Computing in Cardiology*, vol. 50, 2023, pp. 1–4.
- [21] M. Gu, Y. Zhang, Y. Wen, G. Ai, H. Zhang, P. Wang, and G. Wang, "A lightweight convolutional neural network hardware implementation for wearable heart rate anomaly detection," *Computers in Biology and Medicine*, vol. 155, p. 106623, 2023.
- [22] M. N. Imtiaz and N. Khan, "Pan-tompkins++: A robust approach to detect r-peaks in ecg signals," *arXiv preprint arXiv:2211.03171v3*, 2022.
- [23] B. Theodorou, C. Xiao, and J. Sun, "Synthesize high-dimensional longitudinal electronic health records via hierarchical autoregressive language model," *Nature Communications*, vol. 14, no. 1, p. 5305, 2023.
- [24] A. Jiang, C. Huang, Q. Cao, Y. Xu, Z. Zeng, K. Chen, Y. Zhang, and Y. Wang, "Anomaly detection in electrocardiograms: Advancing clinical diagnosis through self-supervised learning," *arXiv preprint arXiv:2404.04935v1*, Apr. 2024.
- [25] J. Qin, F. Gao, Z. Wang, D. C. Wong, Z. Zhao, S. D. Relton, and H. Fang, "A novel temporal generative adversarial network for electrocardiography anomaly detection," *Artificial Intelligence in Medicine*, vol. 136, p. 102489, 2023.
- [26] G. B. Moody and R. G. Mark, "The mit-bih arrhythmia database," 1988, physioNet Digital Library.
- [27] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, J. Li, and E. N. Y. Kwee, "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368–1373, 2018.
- [28] M. A. Reyna, N. Sadr, E. A. Perez Alday, A. Gu, A. J. Shah, C. Robichaux, A. B. Rad, A. Elola, S. Seyedi, S. Ansari, H. Ghanbari, Q. Li, A. Sharma, and G. D. Clifford, "Will two do? varying dimensions in electrocardiography: The physionet/computing in cardiology challenge 2021," in *2021 Computing in Cardiology (CinC)*. IEEE, 2021, pp. 1–4.
- [29] C. Fu, D. Zhang, P. Huang, S. Liu, Y. Qiao, and H. Xu, "Robust anomaly detection for multivariate time series via generative adversarial networks," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*, 2019, pp. 2828–2837.
- [30] S. Tuli, G. Casale, and N. R. Jennings, "Tranad: Deep transformer networks for anomaly detection in multivariate time series data," *arXiv preprint arXiv:2201.07284*, 2022.
- [31] E. M. M. López and E. L. Molina, "Inference of the Youden index and associated threshold using empirical likelihood for quantiles," *Statistics in Medicine*, vol. 30, no. 19, pp. 2467–2480, 2011.
- [32] A. Siffer, P. Fouque, A. Termier, and C. Largouët, "Anomaly detection in streams with extreme value theory," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*, 2017, pp. 1067–1075.
- [33] MathWorks. (2025) Generalized pareto distribution. MATLAB Documentation.
- [34] W. A. Gacek and W. Pedrycz, "Ecg signal processing, classification and interpretation: A comprehensive framework of computer-aided diagnosis," *International Journal of Bioelectromagnetism*, vol. 17, no. 1, pp. 1–28, 2015.
- [35] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000, rRID:SCR.007345.
- [36] PhysioNet, "Physionet: An archive for physiologic signals," <https://physionet.org/>, accessed: 2025-06-23.

- [37] PhysioNet. (2025) Mimic-iv ecg (version 1.0). PhysioNet Dataset.
- [38] L. Koch, B. Shickel, M. Harer, J. Ma, M. Ahmad, M. Aslam, J. Seeliger, A. Wood, A. Goyal, P. Rajpurkar, L. Zimmer, U. J. Muehlematter, M. Jermyn, E. Michailidis, R. Arnold, A. V. Dalca, J. M. Kwon, S. R. Steinhubl, W. T. O’Neal, Z. I. Attia, M. M. Churpek, and M. P. Lungren, “A foundational vision transformer improves diagnostic performance for electrocardiograms,” *npj Digital Medicine*, vol. 6, p. 147, 2023.
- [39] O. Linschmann, M. Rohr, K. Leonhardt, and C. Antink, “Multi-label classification of cardiac abnormalities for multi-lead ecg recordings based on auto-encoder features and a neural network classifier,” in *Proceedings of the Computing in Cardiology Conference*, vol. 48, 2021, pp. 1–4.
- [40] S. H. El-Khafif and M. A. El-Brawany, “Artificial neural network-based automated ecg signal classifier,” *International Scholarly Research Notices*, vol. 2013, pp. 1–6, 2013.
- [41] S. Zhang, Y. Fang, and Y. Ren, “Ecg autoencoder based on low-rank attention,” *Scientific Reports*, vol. 14, p. 12823, 2024.
- [42] J. Ródenas, M. García, J. J. Rieta, and R. Alcaraz, “An efficient algorithm based on wavelet transform to reduce powerline noise from electrocardiograms,” in *Computing in Cardiology (CinC)*, Maastricht, The Netherlands, 2018, p. 200.
- [43] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [44] T. Zhang, Y. Zhang, B. Zhang, H. Zhou, Q. Xu, Y. Zhang, and C. Xu, “Unsupervised deep anomaly detection for multi-sensor time-series signals,” *arXiv preprint arXiv:2107.12626*, 2021.
- [45] C. Ji, L. Wang, J. Qin, and Z. Wang, “Msgformer: A multi-scale grid transformer network for 12-lead ecg arrhythmia detection,” *Biomedical Signal Processing and Control*, vol. 87, p. 105499, 2024, online ahead of print.
- [46] Z. Li and H. Zhang, “Automatic detection for multi-labeled cardiac arrhythmia based on frame blocking preprocessing and residual networks,” *Frontiers in Cardiovascular Medicine*, vol. 8, p. 616585, 2021.