Anomaly Detection in Patient Vital Signs for Early Warning of Critical Health Events

Marc Garreta Basora

July 1, 2025

Abstract

Anomaly detection in 12-lead electrocardiograms (ECGs) identifies critical deviations indicative of cardiovascular disease. This report aims to conduct a comparative analysis of autoencoder-based architectures–convolutional autoencoder (CAE), long short-term memory 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. Each model is trained over normal samples to reconstruct non-anomalous cardiac morphology and locate deviations in possible cardiovascular disease. Using a unified preprocessing and evaluation pipeline on the public China Physiological Signal Challenge (CPSC) dataset, the attention-augmented VAE emerges as the top performer and is deployed in an interactive dashboard that visualizes anomaly localization to aid clinical triage. This model achieves an AUC-PR of 0.81 and a recall of 0.85 on the held-out test set, outperforming the remaining architectures.

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, improving patients' quality of life by providing early warning of upcoming health events, reducing the economic burden on health-care systems, and even saving 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 shows different information from differents parts of the body of a patient.

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 motivate the development of an automated approach that can assist medical professionals to reduce 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 indicate the heart recovery. There are also intervals such as PR, QT or TP, which provide information about the electrical signals flow.

The following sections detail the project objectives and the overall structure of this document.

Main objective. The main objective of this Bachelor's

• Contact E-mail: 1636444@uab.cat
• Supervised by: Mehmet Oguz Mulayim (Department of Computer Science)
• Academic Year 2024/25

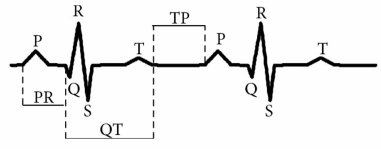

Fig. 1: ECG morphology of two normal beats. Reproduced from Zhang et al. [9] with permission

Degree Final Project is to develop an automatic system for anomaly detection in multivariate healthcare time-series data, with a focus on 12-lead ECG signals, to enhance the early identification of critical health events, specifically cardiovascular diseases.

Specific objectives.

- Understanding anomaly detection role in healthcare.
- Review and analyze the main anomaly detection techniques for time-series data, with emphasis on binary classification.
- Procurement of healthcare time-series data: in particular, collecting and preprocessing ECG signals data.
- Implement and train AI models to detect anomalies in multivariate ECG time-series data.
- Evaluate the performance of the trained models using suitable metrics for time-series data.
- Develop a visualization tool to display detected anomalies in a clear way for healthcare professionals.

Particularly, this report focuses on autoencoder-based architectures to detect anomalies in multivariate ECG time-series data. Specifically, three models are implemented and compared: 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).

Document Structure

The document is structured as follows: Section 2 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 datasets used in Section 4.1, the evaluation metrics applied in Section 4.2, the results in Section 4.3, and the interpretable dashboard in Section 4.4. Finally, Section 5 summarises the conclusions and outlines future work.

2 Related Work

The field of anomaly detection in healthcare has evolved over time. Early methods relied on rule-based systems [21] and small-scale medical corpora, which were limited when handling the complexity of multivariate time-series data

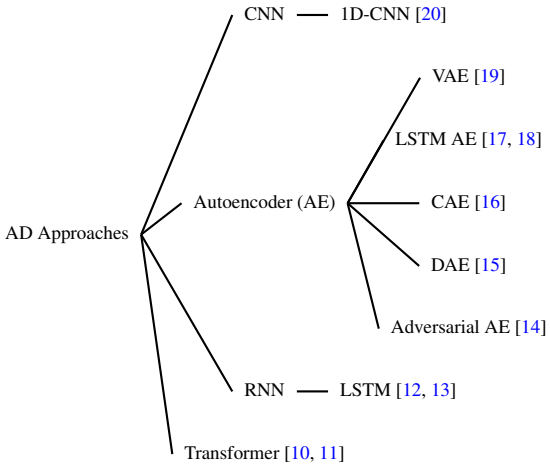

Fig. 2: Anomaly Detection (AD) Approaches in ECGs

such as ECG or electroencephalogram (EEG) signals. With the increase in data from electronic health records (EHRs) [22] 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 [23, 24].

Figure 2 illustrates the anomaly detection approaches that used ECGs as data. Gu et al. [20] 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 [25] 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 [12] apply a deep Long Short-Term Memory (LSTM) autoencoder to raw ECG time signals, learning temporal representations that identify irregular beats with high detection accuracy. Additionally, Thill et al. [13] propose a stacked LSTM approach that models the multivariate prediction-error distribution across leads via a Gaussian model, achieving high recall with low false-alarm rates on the MIT-BIH arrhythmia dataset.

Autoencoder–based approaches have shown promising results regarding the reconstruction of normal ECG morphology and detection of anomalies through reconstruction errors. More specifically, variational autoencoders (VAEs) have demonstrated to be effective for ECG anomaly detection. However, the integration of multi-head attention mechanisms within VAE-based models remains unexplored in the context of ECG data. This confirmed gap motivates the exploration of this autoencoder variant within other architectures, providing a comparative analysis of its performance against similar approaches.

Among autoencoder approaches, several studies are relevant. Lomoio et al. [16] propose a 1D convolutional autoencoder trained on synthetic ECG segments to learn “normal” patterns, reporting ROC AUCs of 97.82 % on simu-

lated data and a ROC AUC of 0.80% on the CPSC-2018 12-lead ECG test set [26]. They also provide reconstruction-error heatmaps over input data for explainability, validated against cardiologist annotations. Choi et al. [17] 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 ROC AUC of 0.74% on the CPSC-2018 12-lead ECG test set. Hribar and Torkar [15] 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 narrowband interference–, attaining 0.81 accuracy and 0.74 recall on the PhysioNet/CinC 2021 challenge dataset [27], with saliency overlays pinpointing the temporal origins of anomalies. In a comprehensive comparative study, Atamny et al. [28] benchmark standard AEs, VAEs, diffusion models, normalizing flows, and Gaussian mixture models on the CPSC-2018 12-lead ECG challenge, finding the MAE-based VAE leads with AUC = 0.83 while even the simplest AE achieves AUC = 0.76. Additionally, a vector quantized (VQ-VAE) is used to perform synthetic data augmentation in order to classify between several cardiac anomalies [19].

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, a hybrid CNN-Transformer network achieves state-of-the-art arrythmia classification on single-lead data by combining local feature extraction with global self-attention [10]. Similarly, S-transform–augmented CNN–Transformer preprocesses the input into a time–frequency representation before attention pooling, producing an improvement when detecting subtle waveform anomalies [11].

Motivated by the high performance of autoencoder-based ECG anomaly detection systems–as explained previously–the absence of any study implementing multi-head attention in VAE architectures for ECG data emphasizes the novelty of this work and motivates 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 preprocessing, 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 can be represented as a multivariate time series:

$$\mathcal{T} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T\}, \quad \mathbf{x}_t \in R^m \quad [29] \tag{1}$$

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

In this unsupervised learning setting, the autoencoder-based models are trained exclusively on normal samples 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 [23, 24]. 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 from the training set T 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_t to assign a binary anomaly label y_t:

$$y_t = \begin{cases} 1, & \text{if } S_t > \tau, \\ 0, & \text{otherwise,} \end{cases} \quad [12] \tag{2}$$

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. **F_1-optimisation.** This approach keeps the value τ that maximises the F_1 score during validation.

3. **Youden's J statistic [30].** This technique optimizes the Receiver Operating Characteristic (ROC) curve during validation, selecting the threshold that maximises $J = \text{TPR}(\text{TruePositiveRate}) - \text{FPR}(\text{FalsePositiveRate})$.

4. **Peaks-Over-Threshold (POT) [31, 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.

The procedure is repeated for every architecture (Convolutional-Autoencoder, Variational Autoencoder-BiLSTM, Variational Autoencoder with Multi-Head Attention), 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 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.

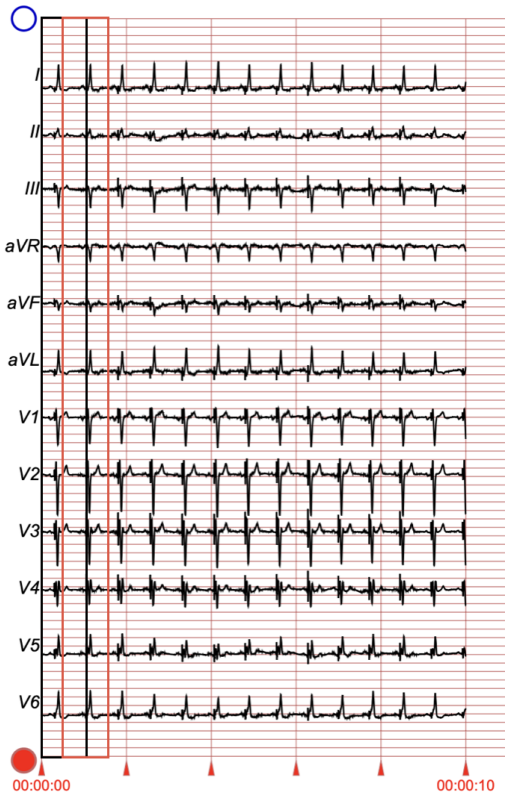

Fig. 3: Example of a raw ECG sample from the MIMIC-IV ECG dataset. The red and black boxes show consecutive windows extracted for training.

Characteristic	PTB-XL	MIMIC-IV ECG
Total ECGs	21 799	> 800000
Sinus rhythm ECGs	≈ 8900	≈ 92000
Number of patients	18 869	> 160000
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 [12]. The training data set is matched by overlapping windows W, where each window represents a fixed-length segment of the ECG signal:

$$W = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}, \quad \mathbf{w}_i \in R^{L \times m} \tag{3}$$

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

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

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 bandpass 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 powerline interference. These filters are applied locally to each ECG lead to preserve signal quality and process them in an personalized way so that each lead is treated accordingly of 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.

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

By normalizing the data using these values, each lead's signal is transformed to have an approximate mean of 0 and a standard variation of 1. The constant vector ϵ is introduced to the denominator to prevent division by zero when a feature has zero variance.

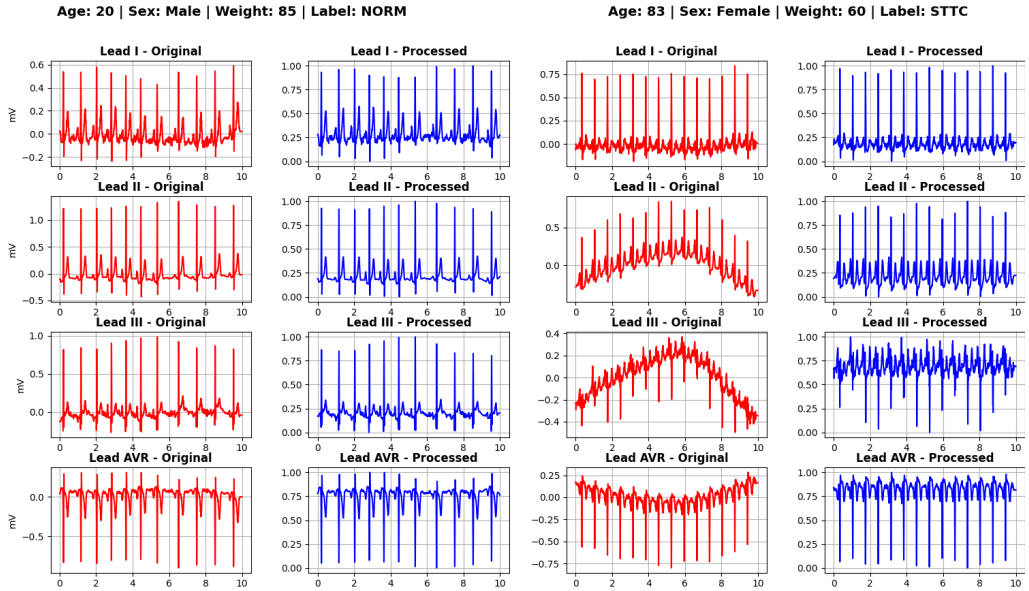

Fig. 4: Comparison of non-processed and pre-processed ECG signals

As observed in Figure 4, the effects of filtering and normalization (in blue) are observable to the raw signals (in red). This pre-processing 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 lightweight 1D convolutional autoencoder (See Figure. 5) processes twelve-lead ECG windows independently and it serves as the main baseline of the autonencoder-based models comparative analysis. It is composed by two main blocks, an encoder that maps data x into a latent representation z and a decoder that maps z back into the original data representation x' as closely as possible. More specifically, each input of shape $[B, 12, L]$, where B is the batch size, 12 is the number of leads, and L is the window size, passes through three convolutions that increase the depth of the channel while reducing at half the temporal resolution at each step ($L \rightarrow L/2 \rightarrow L/4 \rightarrow L/8$), finalizing its way in a latent tensor Z of shape $[B, 128, L/8]$. This architecture–an adaptation from Zhang et al. (2021) [43]–is the first used in this project as it introduces an unsupervised approach without a high complexity structure, providing a simple experimentation process in early stages of this project.

During training, the model minimises the point-wise Mean-Squared Error (MSE) between each input sample and its reconstruction. Because the error is squared, high-amplitude discrepancies dominate the loss: doubling the sample-wise error quadruples its contribution. 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 comparatively minor effect on the total loss. During training, the model minimises the point-wise Mean Squared Error (MSE) between each input and its reconstructed version. In ECG signals, this loss highlights regions with rapid changes–for instance, in the QRS complex–because as it is a squared error, even small reconstruction errors end up being high penalizations.

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

The proposed VAE-BiLSTM model, illustrated in Fig. 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 $\boldsymbol{\mu}_z \in R^{B\times d}$ and a log-variance vector $\log \boldsymbol{\sigma}_z^2 \in R^{B\times d}$, where d is the latent dimension. Afterwards, the reparameterization trick [1] is used to convert a latent sample z from a non-differentiable sampling step to a differentiable one. Then, $\mathbf{z}$ is then repeated T times in a loop process and it is used as input to a unidirectional LSTM decoder, which outputs per-lead reconstruction means $\widehat{\mathbf{x}}_t$ and log-variances $\log \widehat{\boldsymbol{\sigma}}_t^2$, with tensors of shape $[B, 12, T]$.

During training, the model minimises the negative log-likelihood (NLL) of the output plus a Kullback-Leibler

[1] The reparameterization trick is used in VAE models to enable gradient propagation with a differentiable transformation expressing a deterministic transformation

$$\mathbf{z} = \boldsymbol{\mu}_z + \exp\big(0.5 \log \boldsymbol{\sigma}_z^2\big) \odot \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{6}$$

[44]

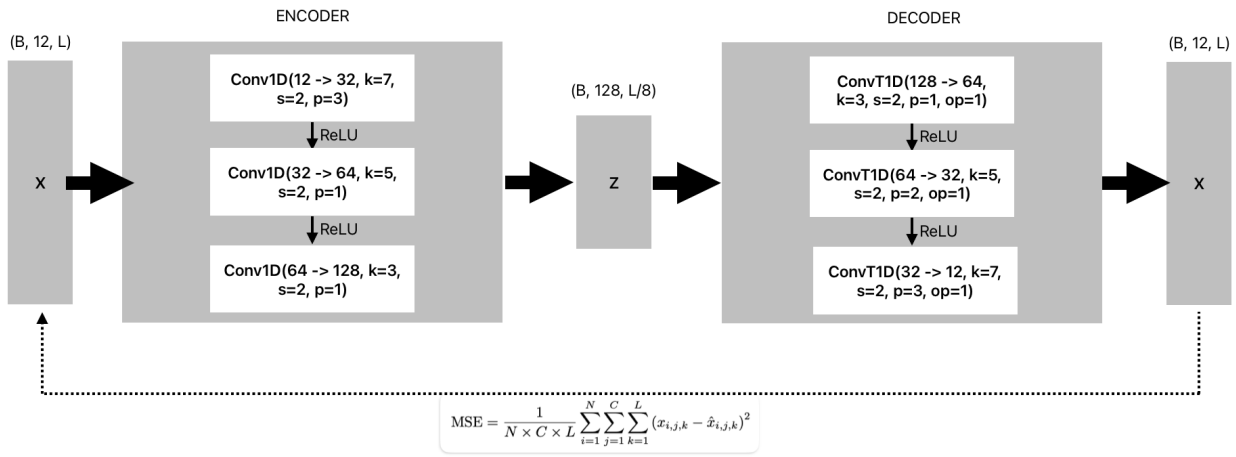

Fig. 5: CAE architecture

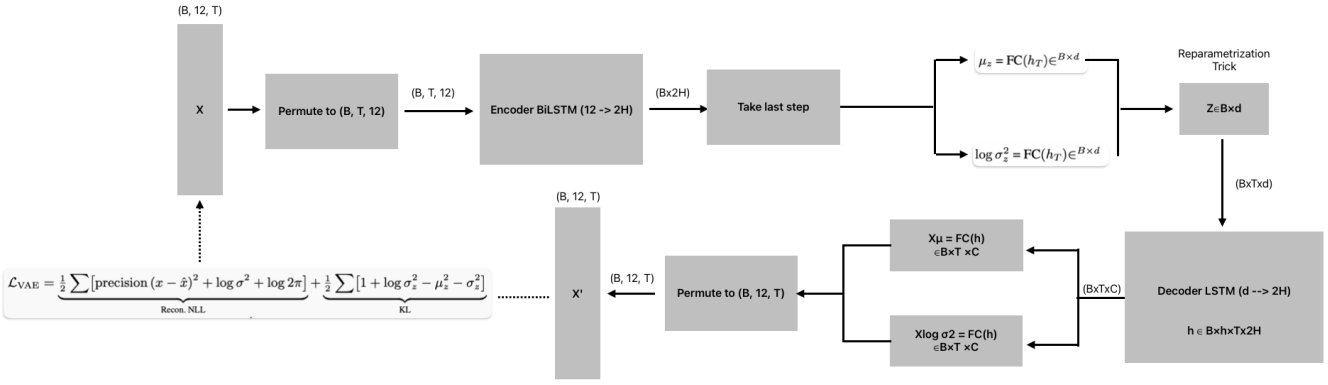

Fig. 6: VAE-BiLSTM architecture

(KL) divergence multiplied by an annealed coefficient β (see Eq. L_{VAE} in Fig. 6). The architecture is inspired by the original VAE formulation of Kingma and Welling [44] and by the OmniAnomaly framework for time-series anomaly detection [32].

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

Multi-Head-Attention VAE (MHA-VAE)

Inspired by the original VAE architecture of Kingma and Welling [44], the OmniAnomaly time–series architecture [32], and the MA-VAE design [45], an extension of the previous model–VAE-BiLSTM–is presented in order to introduce attention. This new architecture has two attention modules–a lead-wise attention and a multi-head sequence attention–between the encoder and decoder (Fig. 7).

First, each 12-lead window $\mathbf{x} \in R^{B \times T \times 12}$ passes through two bidirectional LSTMs, producing hidden states $\mathbf{h} \in R^{B \times T \times 2h}$, where h is the second LSTM's hidden size. An MLP maps $\mathbf{h}$ to per-timestep latent mean and logvar $[\boldsymbol{\mu}_z, \log \boldsymbol{\sigma}_z^2] \in R^{B \times T \times 2d}$, with latent dimension d.

Afterward, a late sample is generated via the reparameterization trick,

$$\mathbf{z} = \boldsymbol{\mu}_z + \exp\big(0.5 \, \log \boldsymbol{\sigma}_z^2\big) \odot \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{7}$$

so that gradients–differentiable gradients–can flow through the stochastic layer during back-propagation.

Then, to extract inter-lead correlations, every (B, T) slice of the input (12×1) is linearly embedded to d channels and fed to a 4-head self-attention layer. The resulting lead context vector $\tilde{\mathbf{h}} \in R^{B \times T \times d}$ is averaged across heads and added to $\mathbf{z}$, resulting in $\mathbf{z}^\star = \mathbf{z} + \tilde{\mathbf{h}}$.

The latent $\mathbf{z}^\star$ is used as *values* in an 8-head attention layer whose queries and keys are linear projections of the raw input window. This produces a context-aware sequence representation $\mathbf{A} \in R^{B \times T \times d}$ 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.

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):

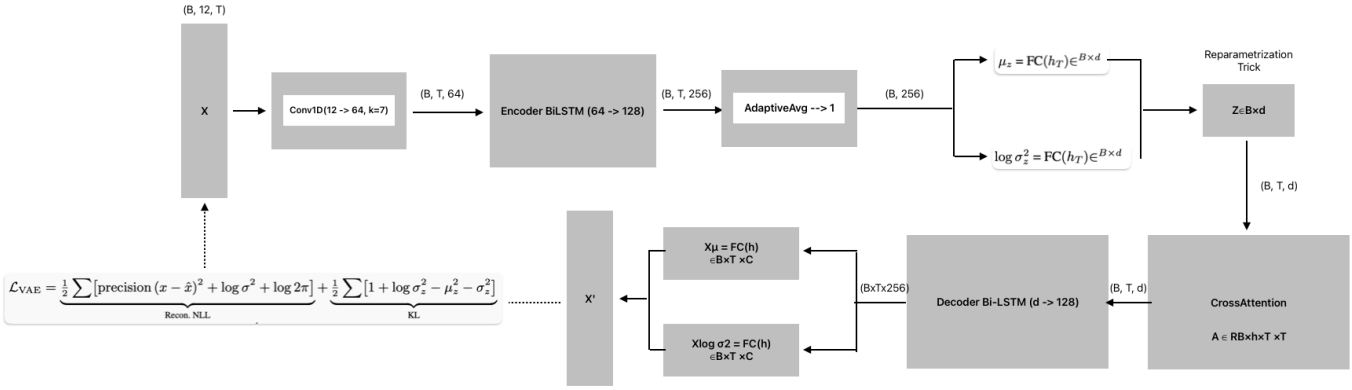

Fig. 7: VAE with Multi-Head Attention architecture

$$W_x = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}, \tag{8}$$

$$n = \left\lfloor \frac{T-m}{s} \right\rfloor + 1, \quad s < m. \tag{9}$$

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

$$s_i = \mathcal{S}(\mathbf{w}_i) \tag{10}$$

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:

$$s_i^{\mathrm{CAE}} = \|\mathbf{w}_i - \hat{\mathbf{w}}_i\|_2^2 \tag{11}$$

  where the MSE measures point-wise differences between the input and its reconstruction [46]. 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 [44]):

$$s_i^{\text{VAE-BiLSTM}} = \|\mathbf{w}_i - \hat{\mathbf{w}}_i\|_2^2 + \mathrm{KL}\big(q_\phi(z \mid \mathbf{w}_i) \,\|\, p(z)\big) \tag{12}$$

  where the first section measures the MSE reconstruction error and the second section regularizes the latent space against the prior. Consequently, a window's high score can lead to an unlikely latent representation and a poor reconstructed sample.

- For the VAE with multi-head attention, we define the anomaly score as an attention-weighted ELBO:

$$s_i^{\text{VAE-MHA}} = \underbrace{\sum_{t=1}^{m} \alpha_{i,t} \left\|\mathbf{w}_{i,t} - \hat{\mathbf{w}}_{i,t}\right\|_2^2}_{\text{Attention-weighted reconstruction}} + \underbrace{\mathrm{KL}\big(q_\phi(z \mid \mathbf{w}_i) \,\|\, p(z)\big)}_{\text{Latent regularization}} \tag{13}$$

  where $\alpha_{i,1}, \ldots, \alpha_{i,m}$ 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 denotes the KL divergence between the approximate posterior $q_\phi(z \mid \mathbf{w}_i)$ and the prior $p(z)$, enforcing a regularized latent space.

  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:

$$S = \frac{1}{n}\sum_{i=1}^{n} s_i \tag{14}$$

As explained in Section 3.1, a threshold τ is established during validation. Finally, the final decision is then given by

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

If the average score S overpasses 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 are evaluated (CAE, VAE-BiLSTM, VAE-BiLSTM with Multi-Head Attention) against four literature baselines on the same dataset and under an identical partitioning configuration:

- ConvAE [16]
- VAE–AE Hybrid [28]
- AE + Peak Detection [17]
- Transformer [47]

Below, Tables 2–6 summarize the hyperparameter configurations for each model. All parameters were selected manually aided by visualization analysis.

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

TABLE 2: CAE Hyperparameters

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

TABLE 3: VAE–BiLSTM Hyperparameters (excluding β)

Parameter	Value
Window size	500
Stride	0.5×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 4: VAE–MHA Hyperparameters (excluding β)

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

TABLE 5: β-Annealing Schedule for VAE–BiLSTM

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

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

4.1 Datasets

The primary evaluation dataset is the publicly available CPSC 2018 challenge corpus [26], 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 relabeled as either *Sinus Rhythm* or *Anomalous*, obtaining over 2000 ECG samples–each recording from a unique patient–resulting to over 2000 patients [48].

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, that its maximum frequency was lowered from 60 Hz to 50 Hz to match the main frequency of the region in which the CSPC 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 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, AUC-ROC, and AUC-PR (PR AUC).

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

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F_1 = 2\,\frac{P\,R}{P + R} \tag{15}$$

where:

- TP: True Positives (correctly detected anomalies)

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

Model	Prec.	Rec.	F1	AUC-ROC	AUC-PR
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 (Novel)	**0.75**	**0.85**	**0.80**	**0.80**	**0.81**
ConvAE [16]	–	–	0.79	0.80	–
VAE–AE Hybrid [28]	–	–	–	0.65	–
AE + Peak Detection [17]	–	0.56	0.56	0.74	–
MSGformer [47]	–	–	**0.847**	**0.88**	–

- FP: False Positives (normal samples incorrectly labeled as anomalies)
- FN: False Negatives (missed anomalies)

AUC-PR: The Area Under the Precision-Recall Curve (AUC-PR) measures the model's ability to detect anomalous signals without producing a high-quantity of false alarms over normal signals. Precision specify the proportion of samples that contain anomalies, and recall indicates the fraction of all anomalous samples the model successfully detects. As an overall, the AUC-PR reflects the balance between finding anomalies–with the assumption to be low in the dataset–against the risk of alarm fatigue. This value tells that as higher it is, the lower the false alarms are produced.

AUC-ROC: The Area Under the Receiver Operating Characteristic curve (AUC-ROC) measures the trade-off between the true positive rate and the false positive rate across various decision thresholds. AUC-ROC provides an overall assessment of the model's discriminative power in distinguishing between normal and anomalous samples.

4.3 Results

Table 7 summarizes the anomaly detection performance of all models evaluated on the CPSC 2018 test set:

Convolutional Autoencoder (CAE) – Results

The convolutional autoencoder achieves a solid performance on the quantitative anomaly detection task. Additionally, the interpretability of this model is obtained through the saliency map [2] using the model's MSE reconstruction error.

Figure 8 illustrates the original (black) and its reconstruction (orange) per each lead, with three extra information: the per-window anomaly score ("AS"), 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 interpret them as normal morphology. With a closer look, in Figure 9, the same pattern is observed. The highest MSE error goes to the QRS complex where the beat is located. As well as that, there is also a region during the first quarter where there is a subtle shape deviation.

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

Variational Autoencoder BiLSTM (VAE-BiLSTM) – Results

Compared with the CAE baseline, the VAE–BiLSTM achieves higher F1, AUC–ROC, and AUC–PR, since its stochastic latent variables force the model to learn representations that generalise from normal-only training data. However, in early experiments the model suffered from *posterior collapse*, where the approximate posterior collapsed onto the prior (See Equation 15):

$$q_\phi(\mathbf{z} \mid \mathbf{x}) \approx p(\mathbf{z}) \tag{15}$$

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

To solve this problem, a technnique called cyclical KL annealing [45] was (Table 6) was applied, 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–it did not collapse to $\mu_z = 0$ and $\log \sigma_z^2 = 0$.

Variational Autoencoder BiLSTM with Multi-Head Attention

The VAE BiLSTM with Multi-Head Attention extends the previous model–a VAE-BiLSTM–with two multi-head self-attention modules–one for the lead level and one at the sequence level–to force the model learn inter-lead correlations (Fig. 7 and to provide more interpretability when deciding whether an ECG signal is considered anomalous or not. After the generation of latent parameters μ_z and $\log \sigma_z^2$ from the encoder, and the sampling of z, a 4-head lead-wise attention is applied to produce local context vectors and average attention weights per each of the twelve leads. Additionally, the original input x is linearly projected into the latent dimensions to map both queries and keys into the latent space, while the updated latent vectors $\mathbf{z}^\star$ are used as values. These three parameters are then fed into an 8-head multi-head attention block, creating a sequence-level attention that combine the global latent representation with the input features across all leads and time-steps.

Figure 10 shows a 12-lead anomalous ECG sample (black original vs. orange reconstruction), with with the per-window anomaly score, the MSE heat-map–blue scale–, and the multi-head attention saliency–yellow hues indicate higher attention. In this particular case, attention does attend in two leads in the particular region in time where there is a possible anomalous pattern, located in the region where the anomaly score shows a high score. However,

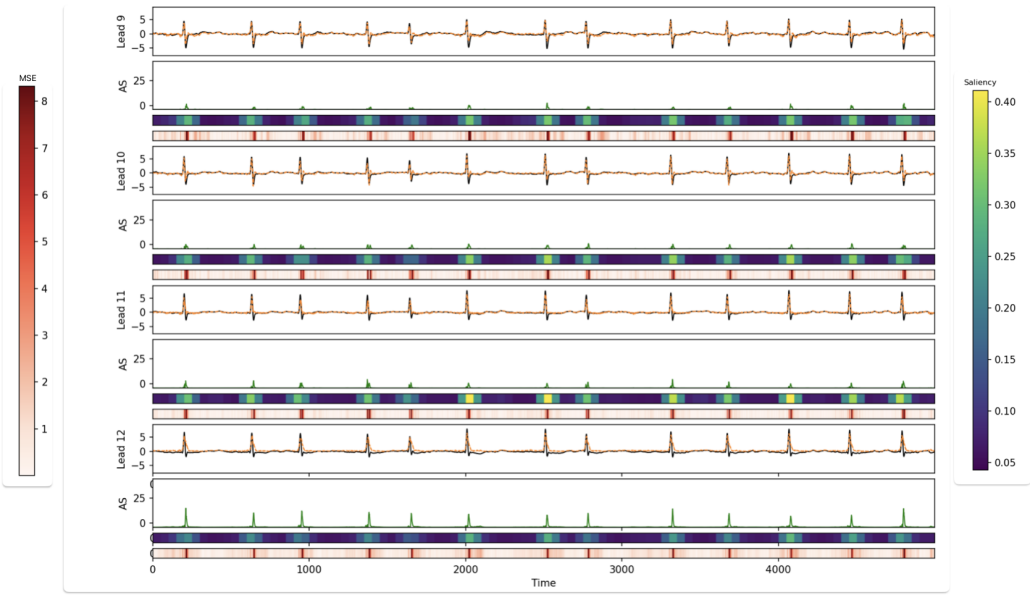

Fig. 8: CAE interpretability of a normal ECG sample

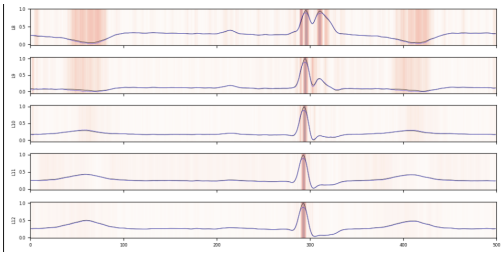

Fig. 9: CAE interpretability given a window

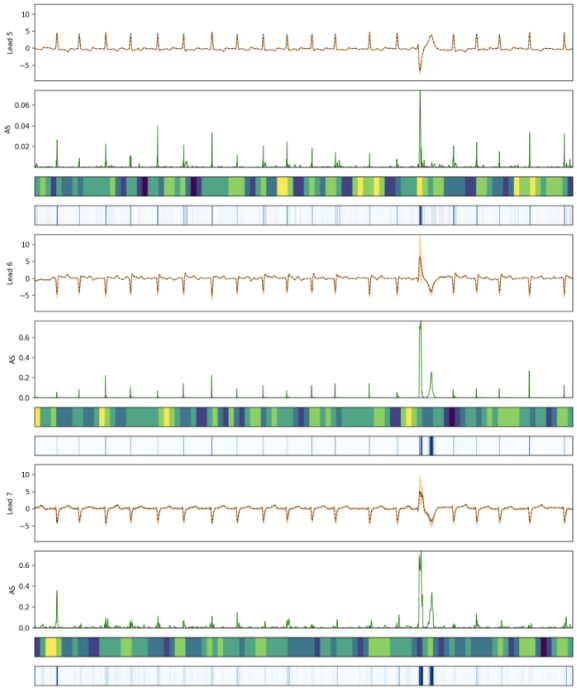

Fig. 10: Example of a possible anomalous pattern in a 12-lead ECG sample.

these attention do not also accomplish its task of focusing over those patterns that might be considered abnormal as there are yellow attentande–particularly in Lead 6–where at the beginning there are yellow hues where the ECG is showing a normal morphology. Figure 11 zooms into a single beat from another anomalous ECG sample, confirming that the highest attention weights and reconstruction errors align precisely with the subtle morphological deviations located at the last region. By integrating attention, the model not only performs better in terms of quantitative results–detecting an anomaly over a 12-lead ECG sample–but it is also able to provide interpretable colormaps that highlight the important regions for the anomaly decision.

4.4 Dashboard

A local web application is implemented as the last objective of this project, 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 12, the user interface offers a simple and personalized framework where you can upload an ECG signal in standard formats–`.npy` or `.dat/.hea`. Afterwards, when clicking on "Analyze ECG Sample", the interface will render the 12-lead ECG signal to offer an anomaly decision as well as its interpretable plot. There are also configuration possibilities such as choosing the leads that you wanna visualize, choosing between a CAE-based and a VAE-based model with Attention or deciding the anomaly threshold–though it is not recommend to alter the default 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 tells the anomaly score in each timestep. Moreover, there are two horizontal heat-map bars that compute the lead-wise attention weights and the point-wise MSE.

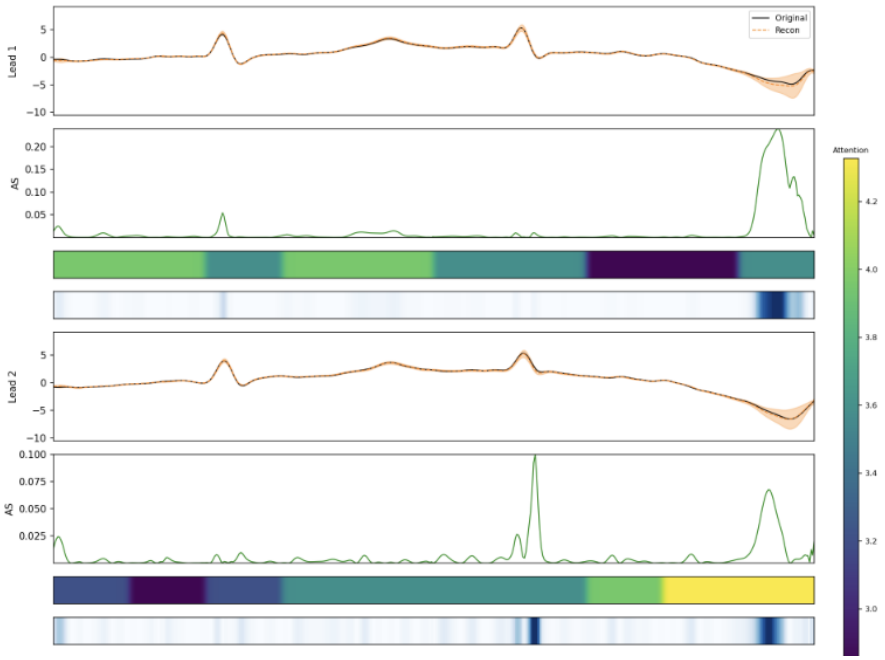

Fig. 11: Example of a possible anomalous pattern in a 12-lead ECG sample at window level.

Fig. 12: User Interface

5 Conclusions

In this project, 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 enhanced with multi-head attention were evaluated using the CPSC 2018 dataset. The iterative experimentation process–modeling and evaluating these models–addressed challenges identified during testing, resulting in a VAE with multi-head attention that significantly improved both quantitative–anomaly detection task and qualitative—anomaly diagnosis and interpretability task–results. This last model achieved an AUC-PR score of 0.81 and a recall of 0.85, outperforming the other compared models. Additionally, the creation of a user-friendly interactive web application provides an interpretable and simple way to observe signals and their possible anomaly deviations visually, helping users in identifying and understanding cardiac anomalies.

Future research directions include extending the current framework to handle with real-time online anomaly detection in collaboration with medical professionals, to enhance the understanding of ECG data and its anomalies. This association would foster innovation from technical roles and the key understanding of healthcare professionals. Furthermore, exploring advanced systems such as Transformer-based models and incorporating additional patient-related data with ECG signals could improve both the quantitative results of anomaly detection and the interpretability of anomaly diagnostic.

Software Availability

The code is made publicly available on https://github.com/marcgarreta/12-lead-ECG-AD

Acknowledgments

Special thanks to my supervisor, Mehmet Oguz Mulayim, for his supervision and constant support, and to UAB for providing computational resources.

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] 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.

[11] 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.

[12] 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.

[13] 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.

[14] S. Papini, M. Montanari, F. Rundo, A. Abate, A. Zorzi, and A. Menciassi, "Ecg anomaly detection with unsupervised deep learning: A comprehensive review," *Frontiers in Physiology*, vol. 13, p. 961724, 2022.

[15] 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.

[16] 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.

[17] 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.

[18] M. Roy, S. Majumder, A. Halder, and U. Biswas, "Ecg-net: A deep lstm autoencoder for detecting anomalous ecg," *Engineering Applications of Artificial Intelligence*, vol. 124, p. 106484, 2023.

[19] 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.

[20] 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.

[21] M. N. Imtiaz and N. Khan, "Pan-tompkins++: A robust approach to detect r-peaks in ecg signals," *arXiv preprint arXiv:2211.03171v3*, 2022.

[22] 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.

[23] 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.

[24] 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.

[25] G. B. Moody and R. G. Mark, "The mit-bih arrhythmia database," 1988, physioNet Digital Library.

[26] 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.

[27] 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.

[28] 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.

[29] Y. L. Agarwal and B. Ravindran, "Tranad: Deep transformer networks for anomaly detection in multivariate time series data," *arXiv preprint arXiv:2201.07284*, 2022.

[30] 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.

[31] 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.

[32] 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.

[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] T. Zhang, Y. Zhang, B. Zhang, H. Zhou, Q. Xu, Y. Zhang, and C. Xu, "Ecg-anomaly: Anomaly detection for electrocardiogram using self-supervised contrastive learning and adversarial training," *arXiv preprint arXiv:2107.12626*, 2021. [Online]. Available: https://arxiv.org/abs/2107.12626

[44] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013. [Online]. Available: https://arxiv.org/abs/1312.6114

[45] 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. [Online]. Available: https://arxiv.org/abs/2309.02253

[46] 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. [Online]. Available: https://arxiv.org/abs/2107.12626

[47] 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.

[48] 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.