UAB
Universitat Autònoma
de Barcelona

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

Marc Garreta Basora
1636444@uab.cat
Bachelor's Degree in Artificial Intelligence

Tutor: Oguz Mulayim
oguz@iiia.csic.es

1 Problem Definition

Anomaly detection (AD) refers to the process of identifying patterns that deviate from expected or normal behavior in the data [CBK09, PSCH21]. The importance of this data analysis lies in the fact that these deviations, known as anomalies [1], may lead to critical actionable information [LTZ08]. 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 mortality worldwide, accounting approximately one third of global deaths [LDD+22, Org21]. Apart from the impact on health, CVDs also have economic consequences as, in the United States, for example, the total cost of CVDs was estimated at 616 billion dollars in 2015 and is projected to reach 1.2 trillion dollars by 2035, factoring in caregiver costs, healthcare services, medication, and lost productivity due to mortality [DKB+18].

2 Objectives

2.1 Main objective

The main objective of the Bachelor's Degree Final Project is to develop an AI-based system for anomaly detection in healthcare time-series data, with a focus on electrocardiogram (ECG) signals, to enhance the early identification of critical health events, specifically cardiovascular diseases.

This work explores the ECG signal, as its processing is one of the most common methods for early detection in CVDs providing information about cardiac electrophysiology. This time-series signal is widely used in medical diagnosis because of its non-invasive and effective advantages in cardiologist tools [AG11].

Specifically, the ECG measures the electrical activity of the heart, following the principle that these electrical impulses propagate through the body and can be detected on the skin [AN04]. Its graphical representation, shown in Figure 4, consists of 12 leads (channels), each providing a unique electrical perspective of the heart. Anomalies in ECG signals can reveal various cardiac conditions such as coronary artery disease (CAD), heart failure (HF), arrhythmia (ARR), and other cardiac diseases. Given the important role of early warning in healthcare, anomaly detection approaches have been applied within ECG data, helping with the identification of anomalous patterns.

2.2 Specific objectives

- Understanding anomaly detection role in healthcare
- Review and analyze the main anomaly detection techniques for time-series data
- Procurement of healthcare time-series data: in particular, collecting and pre-processing ECG signals data

[1] Data patterns that have different data characteristics from normal instances

- Implement and train AI models to detect anomalies in 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

3 Methodology

Literature Review and Analysis of Anomaly Detection Approaches:

- Literature Review on Anomaly Detection Techniques: Perform a search given reliable and quality academic papers using academic search engines such as Google Scholar and Scopus to understand the task at hand and the necessity of anomaly detection in time series data. More specifically, physiological data such as ECG will be the main focus to relate it within anomaly detection state-of-the-art (SoA) approaches applied to time series data in healthcare.
- Challenges in Anomaly Detection for Healthcare Time-Series Data: Analyze the main challenges in anomaly detection when using physiological data to overcome them during the length of the work, e.g. imbalanced data, unlabeled data, noisy data, etc.
- Comparative Analysis of Anomaly Detection Techniques in the Healthcare domain: Analyze and compare different actual anomaly detection techniques given literature and open source code to determine the feasibility of the task at hand. The goal is to obtain those studies that show the most promising approaches to deal with physiological data and the challenges that it comprises.

Procurement of Healthcare Dataset

- Identification and Selection of time-series datasets: A review of existing datasets mainly used in approaches that deal with this task or similars will be performed, including both open-access and restricted databases for this study using repositories such as *Physionet*[2]. The criteria of selecting the most promising dataset will consider dataset quality, availability, realism, and ideally, the linkage within patient metadata such as demographic information or other electronic health records.
- Data Collection, Exploratory Data Analysis (EDA), and Preprocessing:
  - Collect data from various repositories to compare them regarding the above mentioned criteria
  - Use of programming languages such as *Python* and libraries such as *wfdb*[3] to load and visualize the time series data
  - Perform exploratory data analysis (EDA) using Python visualization tools such as *matplotlib* and *wfdb* to understand the complexity of this typology of data.
  - Implement ECG pre-processing techniques such as noise removal, normalization and pattern extraction to multivariate time-series.

Implementation of AI Models

- Model Selection and Training: Utilize machine learning frameworks such as *PyTorch* and *TensorFlow* to implement selected models for anomaly detection. Additionally, a monitoring process through frameworks such as *Weights and Biases*[4] will be used for real-time tracking of model performance.
- Hyperparameter Optimization: Hyperparameter tuning will be used to ideally encounter the perfect parameters such as learning rate, batch size or architecture design to be used within the task at hand.

Development of a Visualization Dashboard (Streamlit or Flask)

- Storyboard and Proof of Concept (PoC) Development: A conceptual end-user design will be provided to make an initial overview of the final application. This PoC will be used so that the user interface aligns with the project objectives.

[2] https://physionet.org
[3] https://physionet.org/content/wfdb-python/4.1.0/
[4] https://wandb.ai/site/

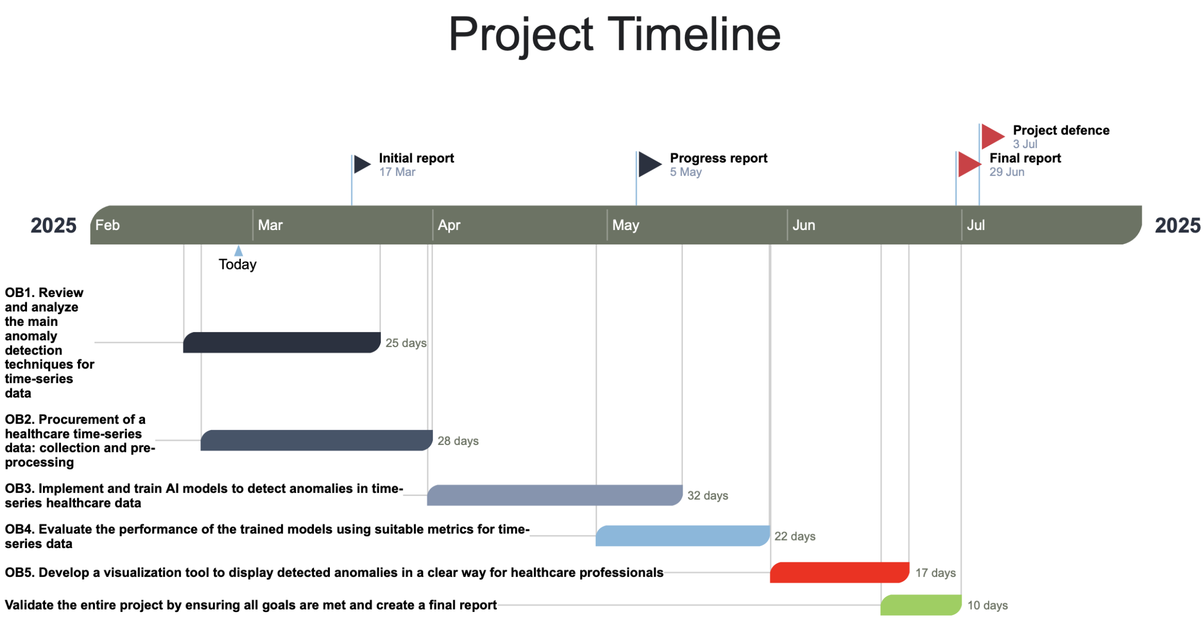

Figure 1: Provisional Timeline for the TFG

- User-Centric Interface for Medical Professionals: Develop a front-end using *Streamlit*[5] or *Flask*[6], to provide a valuable experience for clinicians to:
  - Upload data via the user dashboard.
  - Visualize detected anomalies in those files that contain an ECG waveform format for explainability reasons.
  - Provide a report of possible risks given the identified anomaly.

4 Planning

4.1 Timeline

As shown in Figure 1, the planning of this project follows a waterfall-like methodology, where each stage progresses step by step in a structured order aligned with the objectives. Therefore, it does not include rigid dependencies so that every step can be flexible throughout time. The process begins with a literature review to conduct a feasibility study of the project, once the background is established, the focus changes into a practical direction, where data exploration and model implementation is performed. Following this, after training, the models are evaluated using appropriate metrics. Finally, a user-friendly UI is developed to enable healthcare professionals to visualize the anomalies detected in the data provided by them.

5 Related Work

The field of anomaly detection in healthcare has evolved over time. First approaches relied heavily on rule-based systems and small-scale medical documentation, which were limited when handling the overall complexity of multivariate time-series data such as ECG or EEG signals. Nowadays, recent methods use machine learning techniques to deal with complex data in healthcare originating from the abundant electronic health records (EHRs). We can categorize all approaches in four major categories as below:

[5]https://streamlit.io
[6]https://flask.palletsprojects.com

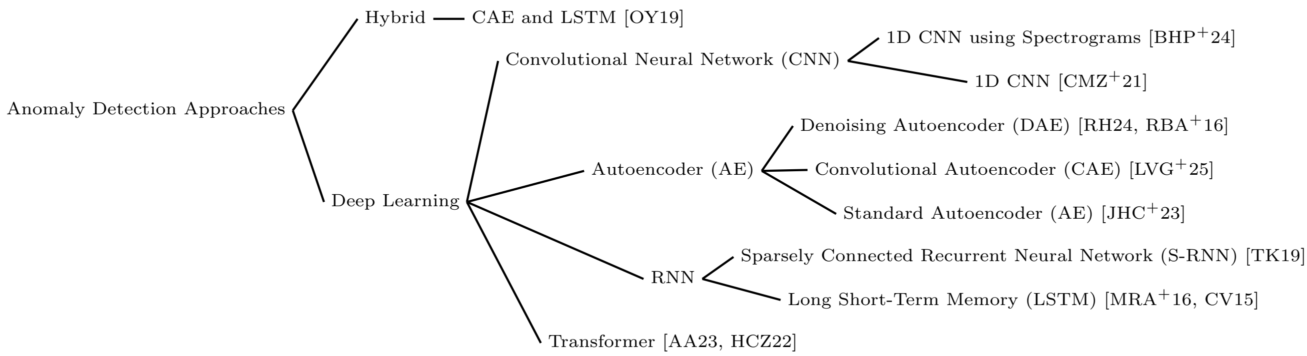

Figure 2: Anomaly Detection Approaches in ECGs

5.1 Deep Learning Approaches

Transformer Models: Transformer-based methods have been used for anomaly detection to address the lack of labeled data in the medical domain, more specifically, in the detection of anomalies in ECG. Besides, the high-capacity when dealing with long time intervals of data is also a challenge where transformers have been demonstrated to perform well [AA23, HCZ22].
Recurrent Neural Networks (RNN): Long Short-Term Memory (LSTM) networks have been used in this typology of tasks because of their long-term temporal dependencies capabilities [MRA+16, CV15]. As well as that, other RNN-based approaches such as Sparsely Connected Recurrent Neural Networks (S-RNN) have been used [TK19].

5.2 Autoencoder-based Methods

Autoencoders (AE): Autoencoders are neural networks that learn to reconstruct the input data, capturing the important features of normal patterns. In the context of ECG anomaly detection, normal patters are defined as consistent P-QRS-T[7] wave morphology, a stabilized heart rate and proper intervals of these waves [Her22]. So, autoencoders are effective when dealing with imbalanced datasets, where normal ECG recording are more common than anomalous ones. During the testing phase, AEs can identify anomalies without having to be trained with anomalous data [JHC+23].
Convolutional Autoencoders (CAE): Convolutional autoencoders apply the concept of convolutional layers to autoencoders so that it is more effective for feature extraction. This method is strong for creating spatial relationships in the signal [LVG+25].
Denoising Autoencoders (DAE): Denoising autoencoders increase the robustness of the model by adding noise to input signals, so it can be resilient to noise in cases where signals are prominently noisy. E.g. an ECG can contain noise because of the machine that is recording the signal. [RH24, RBA+16].

5.3 Convolutional Neural Networks (CNN)

1D CNN: A 1D convolutional neural network can be applied in the detection of arrythmia processing ECG waveforms as time-series data, where each signal is represented as a one-dimensional vector[CMZ+21]. In this approach, convolutional filters are applied to capture both spatial and temporal features.
1D CNN with Spectrograms: A similar version of the above mentioned method, a 1D CNN that uses frequency-based representations instead of ingesting time-series ECG signal as input. [BHP+24].

5.4 Hybrid Approaches

CAE and LSTM Combination: Hybrid models are also interesting approaches. For instance, a combination of a convolutional autoencoder with Long Short Term Memory (LSTM) can outperform other single-technique approaches [OY19].

[7]The basic pattern of electrical activity across the heart[AN04]

6 Datasets

The PTB-XL dataset is the main focus for this study because of its comprehensive data linkeage within patient data and its realism.
It includes over 21,000 clinical ECG recordings from nearly 19,000 unique patients, making it one of the largest databases available in medical repositories regarding ECG signals. Additionally, the metadata collected in each ECG linked to patient history makes it perfect to explore multimodal approaches to combine different types of health data to improve the anomaly detection task.

Dataset	Number of ECGs	Number of Patients	ECG Type	Sampling Rate
PTB-XL[1]	21,799	18,869	12-lead ECG	500 Hz / 100 Hz
MIMIC-IV ECG[2]	800,000+	160,000+	12-lead ECG	500 Hz
ECG5000[3]	10,000	N/A	2-lead ECG	N/A
MIT-BIH Arrhythmia[4]	48	47	2-lead ECG	360 Hz

Table 1: Comparison of ECG Datasets with Focus on PTB-XL

7 Storyboard

As observed in Figure 3, the storyboard provides an overview of how the user interface (UI) will function for the end-user.

- **Document Upload:** Users can upload patient-related files, including ECG recordings (e.g., .hea and .dat files) and demographic details.
- **ECG Visualization:** The interface displays two ECG images:
  - The original, unaltered ECG graph.
  - An annotated version highlighting potential anomalies.
- **Annotation and Editing:** Users can interact with the annotated ECG:
  - Amplify specific channels for better visualization.
  - Adjust AI-generated annotations and add new markers.
  - Zoom into sections and modify display settings.
- **Export and Save:** The final annotated ECG can be exported as a report or stored for future reference.

[1] `https://physionet.org/content/ptb-xl/1.0.3/`
[2] `https://physionet.org/content/mimic-iv-ecg/1.0/`
[3] `https://www.timeseriesclassification.com/description.php?Dataset=ECG5000`
[4] `https://www.physionet.org/content/mitdb/1.0.0/`

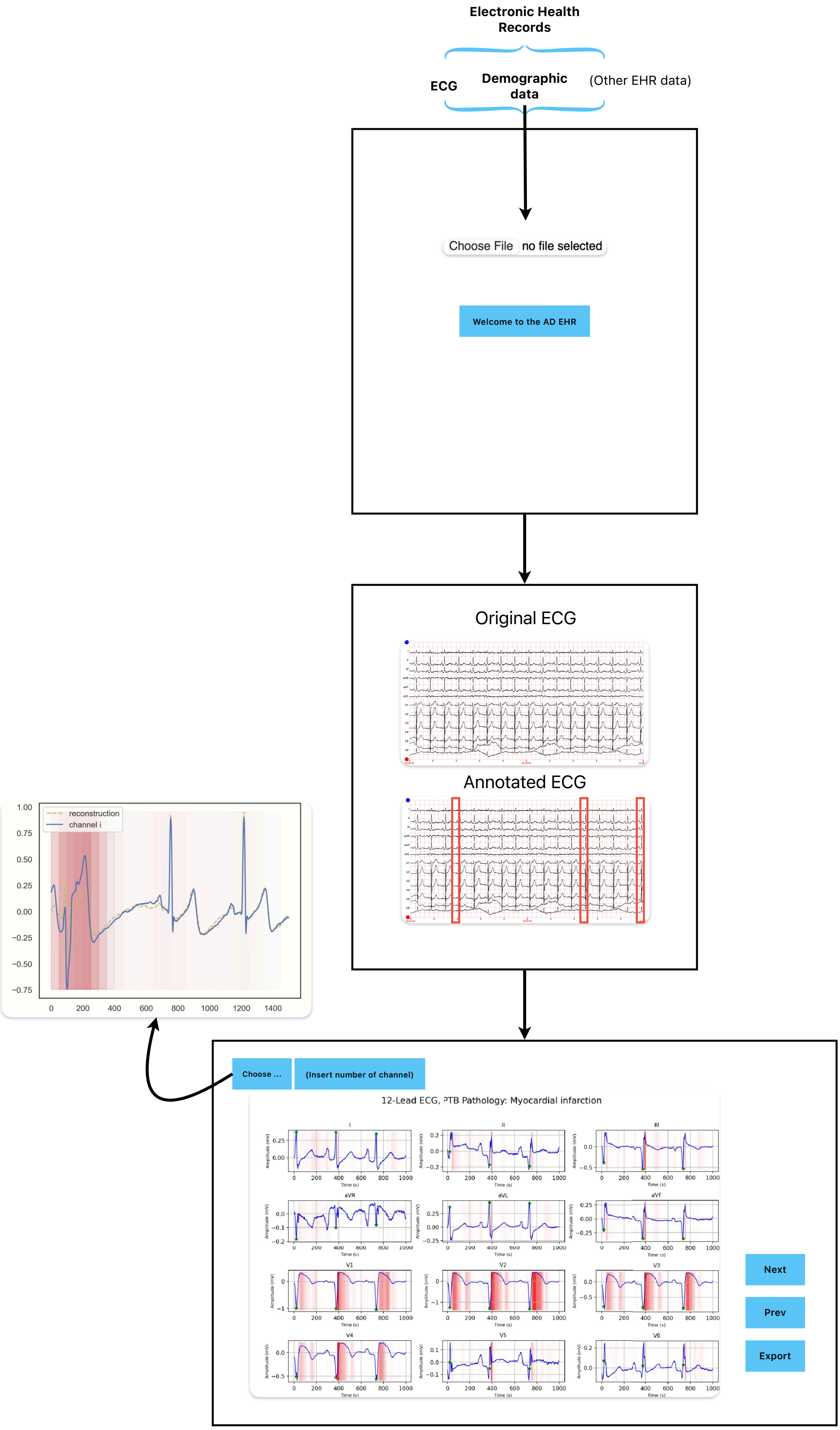

Figure 3: Provisional Storyboard for the UI

References

[AA23] Abrar Alamr and Abdelmonim Artoli. Unsupervised transformer-based anomaly detection in ecg signals. *Algorithms*, 16(3), 2023.

[AG11] Witold Pedrycz Adam Gacek. *ECG Signal Processing, Classification and Interpretation*. Springer Science Business Media, Springer Science Business Media, 09 2011.

[AN04] Euan A. Ashley and Josef Niebauer. *Cardiology Explained*. Remedica, Remedica, 2004. Chapter 3, pp. 15–34.

[BHP+24] Nhat-Tan Bui, Dinh-Hieu Hoang, Thinh Phan, Minh-Triet Tran, Brijesh Patel, Donald Adjeroh, and Ngan Le. Tsrnet: Simple framework for real-time ecg anomaly detection with multimodal time and spectrogram restoration network. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–4, 2024.

[CBK09] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3), July 2009.

[CMZ+21] Omar Cheikhrouhou, Redowan Mahmud, Ramzi Zouari, Muhammad Ibrahim, Atef Zaguia, and Tuan Nguyen Gia. One-dimensional cnn approach for ecg arrhythmia analysis in fog-cloud environments. *IEEE Access*, 9:103513–103523, 2021.

[CV15] Sucheta Chauhan and Lovekesh Vig. Anomaly detection in ecg time signals via deep long short-term memory networks. In *2015 IEEE international conference on data science and advanced analytics (DSAA)*, pages 1–7. IEEE, 2015.

[DKB+18] Sandra B. Dunbar, Olga A. Khavjou, Tamilyn Bakas, Gail Hunt, Rebecca A. Kirch, Alyssa R. Leib, R. Sean Morrison, Diana C. Poehler, Veronique L. Roger, and Laurie P. Whitsel. Projected costs of informal caregiving for cardiovascular disease: 2015 to 2035: A policy statement from the american heart association. *Circulation*, 137(19):e558–e577, 2018.

[HCZ22] Rui Hu, Jie Chen, and Li Zhou. A transformer-based deep neural network for arrhythmia detection using continuous ecg signals. *Computers in Biology and Medicine*, 144:105325, 2022.

[Her22] Rebeca Alina Watson Hernandez. Interpretación del electrocardiograma normal. *Revista Ciencia y Salud Integrando Conocimientos*, 6(5):85–91, Oct 2022.

[JHC+23] Aofan Jiang, Chaoqin Huang, Qing Cao, Shuang Wu, Zi Zeng, Kang Chen, Ya Zhang, and Yanfeng Wang. Multi-scale cross-restoration framework for electrocardiogram anomaly detection. In Hayit Greenspan, Anant Madabhushi, Parvin Mousavi, Septimiu Salcudean, James Duncan, Tanveer Syeda-Mahmood, and Russell Taylor, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 87–97, Cham, 2023. Springer Nature Switzerland.

[LDD+22] Megan Lindstrom, Nicole DeCleene, Henry Dorsey, Valentin Fuster, Catherine O. Johnson, Kate E. LeGrand, George A. Mensah, Christian Razo, Benjamin Stark, Justine Varieur Turco, and Gregory A. Roth. Global burden of cardiovascular diseases and risks collaboration, 1990-2021. *JACC*, 80(25):2372–2425, 2022.

[LTZ08] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422, 2008.

[LVG+25] Ugo Lomoio, Patrizia Vizza, Raffaele Giancotti, Salvatore Petrolo, Sergio Flesca, Fabiola Boccuto, Pietro Hiram Guzzi, Pierangelo Veltri, and Giuseppe Tradigo. A convolutional autoencoder framework for ecg signal analysis. *Heliyon*, 11(2):e41517, 2025.

[MRA+16] Pankaj Malhotra, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff. Lstm-based encoder-decoder for multi-sensor anomaly detection, 2016.

[Org21] World Health Organization. Cardiovascular diseases (cvds). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), 11 June 2021.

[OY19] Ru-San Tan Edward J. Ciaccio U. Rajendra Acharya Ozal Yildirim, Ulas Baran Baloglu. A new approach for arrhythmia classification using deep coded features and lstm networks. *Computer Methods and Programs in Biomedicine*, 176:121–133, Jul 2019.

[PSCH21] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Comput. Surv.*, 54(2), March 2021.

[RBA$^+$16] M.M. Al Rahhal, Yakoub Bazi, Haikel AlHichri, Naif Alajlan, Farid Melgani, and R.R. Yager. Deep learning approach for active classification of electrocardiogram signals. *Information Sciences*, 345:340–354, 2016.

[RH24] Drago Torkar Rok Hribar. Explainable anomaly detection of 12-lead ecg signals using denoising autoencoder. *Studies in Computational Intelligence*, pages 127–140, 2024.

[TK19] Chenjuan Guo-Christian S. Jensen Tung Kieu, Bin Yang. Outlier detection for time series with recurrent autoencoder ensembles. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pages 2725–2732, Aug 2019.

A Annex A: Additional Tables

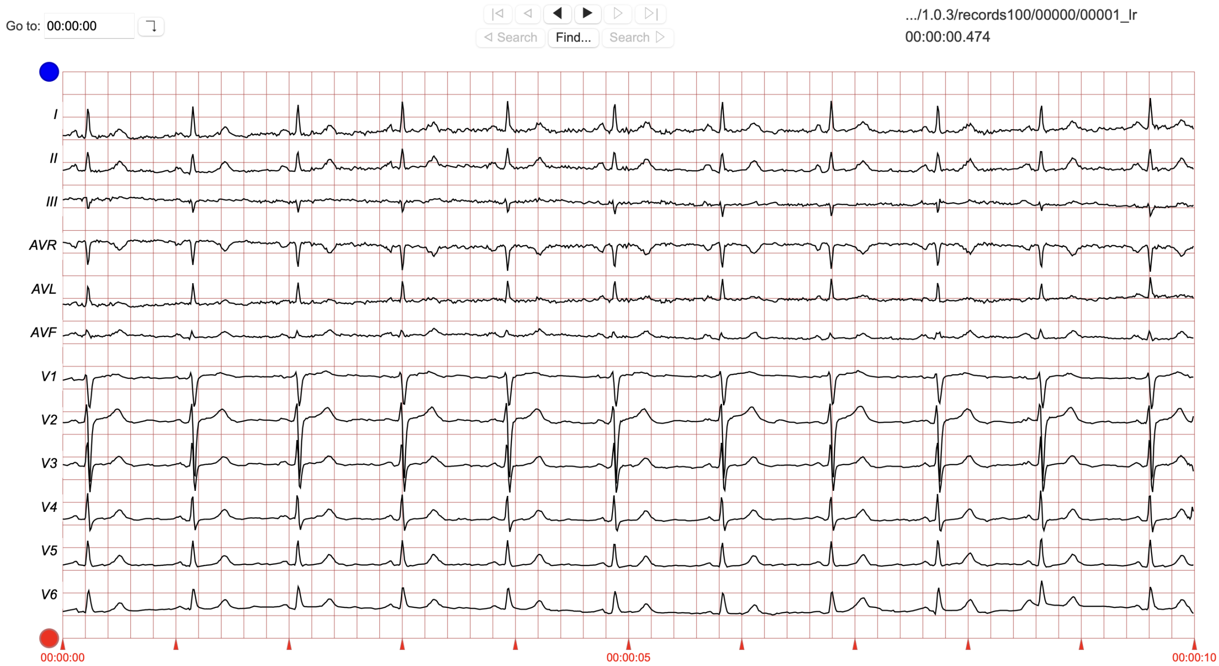

Figure 4: ECG sample