UNIVERSITÀ DEGLI STUDI DI SALERNO

UNIVERSITÀ DEGLI STUDI DI SALERNO

DIPARTIMENTO DI INFORMATICA

DIPARTIMENTO DI ECCELLENZA

University of Salerno – Department of Computer Science

Meta-Learning for ECG Anomaly Detection

Leonardo Falanga
Student ID: 0522502019
Master's Degree in Artificial Intelligence
Department of Computer Science
University of Salerno
Supervisor: Prof. Loredana Caruccio

Abstract

This project explores the use of meta-learning to improve anomaly detection in electrocardiogram (ECG) time-series data. The goal is to train a meta-model that learns to choose the most appropriate base anomaly detection model K-Nearest Neighbors (KNN) or Isolation Forest (IForest) for each ECG segment, based on statistical features extracted using the `tsfresh` library. The project results demonstrate that meta-learning can significantly enhance accuracy and adaptiveness in time-series analysis, opening possibilities for implementation on embedded platforms like AIfES.

1. Introduction

Time-series analysis plays a crucial role in many fields, especially in biomedical signal processing. ECG signals are commonly used to monitor heart activity, and detecting anomalies in such signals is vital for early diagnosis and preventive healthcare. Traditional anomaly detection methods rely on one pre-defined algorithm. However, ECG signals are highly dynamic and non-stationary, and no single algorithm performs optimally across all scenarios.

Meta-learning, often described as "learning to learn", shifts the paradigm. Instead of selecting the best model manually or using a static configuration, a meta-learner is trained to autonomously choose the best model depending on the characteristics of each data segment. In this work, we applied this principle to ECG anomaly detection.

2. Meta-Learning Explained

Meta-learning is an advanced machine learning concept where the system learns how to adapt or select other models. The key idea is that each data instance (in our case, an ECG block) may require a different strategy. The meta-model is trained on meta-data: data about the performance of other models.

In our setup:

- Each ECG block is analyzed using both KNN and Isolation Forest.

- Their performance is evaluated, and the better model is recorded.
- A meta-classifier is then trained on extracted features to predict which model is likely to perform better.

This layered architecture allows the system to generalize its decision-making process and adapt in real-time to unseen patterns in data.

3. The Notebook – Code Walkthrough

The Jupyter notebook provided with this report, `MetaLearning_ECG_clean.ipynb`, executes the full pipeline:

1. **Loading ECG Data:** The notebook begins by reading ECG files from a dataset directory. Each file contains labeled time-series data.
2. **Splitting into Blocks:** The data is divided into fixed-length segments (blocks), enabling localized analysis.
3. **Feature Extraction:** Using `tsfresh`, over 100 statistical features are extracted from each block. These features capture time-domain and frequency-domain information.
4. **Anomaly Detection Evaluation:** Each block is analyzed using KNN and IForest. Their labels are compared to ground truth to compute accuracy.
5. **Meta-Labeling:** The model with the highest accuracy for each block is assigned as the target label for the meta-model.
6. **Training the Meta-Model:** A Random Forest is trained using the features as input and the best model label as output.
7. **Testing and Evaluation:** The notebook evaluates the meta-model on a held-out test set and displays the final accuracy and confusion matrix.

4. Key Models Used

K-Nearest Neighbors (KNN): A non-parametric method that classifies a data point based on the majority label among its nearest neighbors. Effective in well-clustered datasets.

Isolation Forest (IForest): An ensemble method that isolates outliers by randomly selecting features and split values. Anomalies are more susceptible to isolation.

Random Forest (Meta-Model): A supervised ensemble learning method using decision trees. It generalizes well and handles high-dimensional feature spaces.

5. Results

- Meta-model test accuracy: **0.908**
- Confusion matrix illustrates consistent performance across model prediction classes.

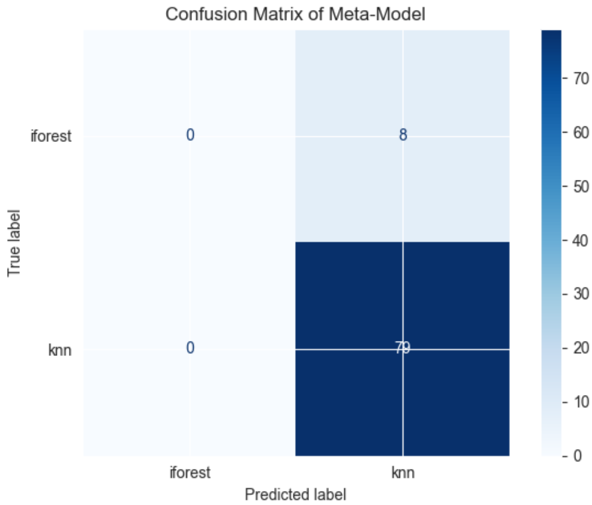

Figure 1: Confusion Matrix for Meta-Model

6. Future Applications with AIfES

AIfES (Artificial Intelligence for Embedded Systems) is a lightweight C++ machine learning framework designed to run on microcontrollers such as Arduino, STM32, and ESP32. This work considers future deployment on embedded systems using AIfES, a C-based machine learning library developed by the Fraunhofer Institute for Microelectronic Circuits and Systems (IMS). AIfES is subject to proprietary licensing.
Integrating this meta-learning logic into AIfES could allow real-time model selection directly on edge devices.

Such an approach would:

- Enable real-time ECG monitoring in wearable medical devices.
- Reduce reliance on cloud-based computation.
- Offer energy-efficient, intelligent anomaly detection in embedded IoT applications.

Conclusion

Meta-learning offers a powerful paradigm shift in anomaly detection by adding a layer of intelligence that learns how to choose. The presented approach successfully demonstrates this on real ECG data with promising accuracy. With further optimization, this system can be embedded in ultra-low power devices using AIfES, making it ideal for mobile and medical edge applications.

Supervisor: Prof. Loredana Caruccio