In21-S5-CS3501 Data Science and Engineering Project
Department of Computer Science and Engineering
University of Moratuwa

Robust Identification Model Using Single-Channel ECG Data

Team Members:
Sajeenthiran P - 210553J
Ranaweera HK - 210523T
Rathnayaka W.T - 210536K

Group ID:23
Project ID: 5
Mentor:Dr. Uthayasanker
Teaching Assistant: Mr Vithurabiman Senthuran

1. Executive Summary:

This project aims to develop a robust identification model using single-channel ECG (Electrocardiogram) data. The problem addressed is the need for a reliable biometric identification system that remains accurate despite distortions in ECG signals and should also protect the user’s privacy. We are planning to apply deep learning techniques along with representation learning. The expected outcome is a high-accuracy identification model that can be integrated into security systems, wearable devices, and other applications requiring robust biometric identification.

2. Problem Statement:

The project addresses the challenge of developing an ECG biometric identification model resilient to ECG signal distortions. Traditional ECG biometric models often suffer from reduced accuracy due to noise and other distortions in the data. This problem is significant as it affects the reliability of security systems, healthcare applications, and other areas where accurate personal identification is crucial. Mostly these distortions arise in wearable devices that measure ECG signals. The goal is to create a model that maintains high identification accuracy despite these challenges.

The project also addresses the issue of patient privacy by implementing a methodology to secure the patient's ECG signals.

3. Data Description:

The data used in this project will be single-channel ECG recordings. The source of the data will be publicly available ECG databases such as PhysioNet. This data is structured and includes key features such as heart rate variability, QRS complex, P waves, and T waves. Data collection will involve extracting relevant features from the raw ECG signals and preprocessing them to remove noise and artifacts.

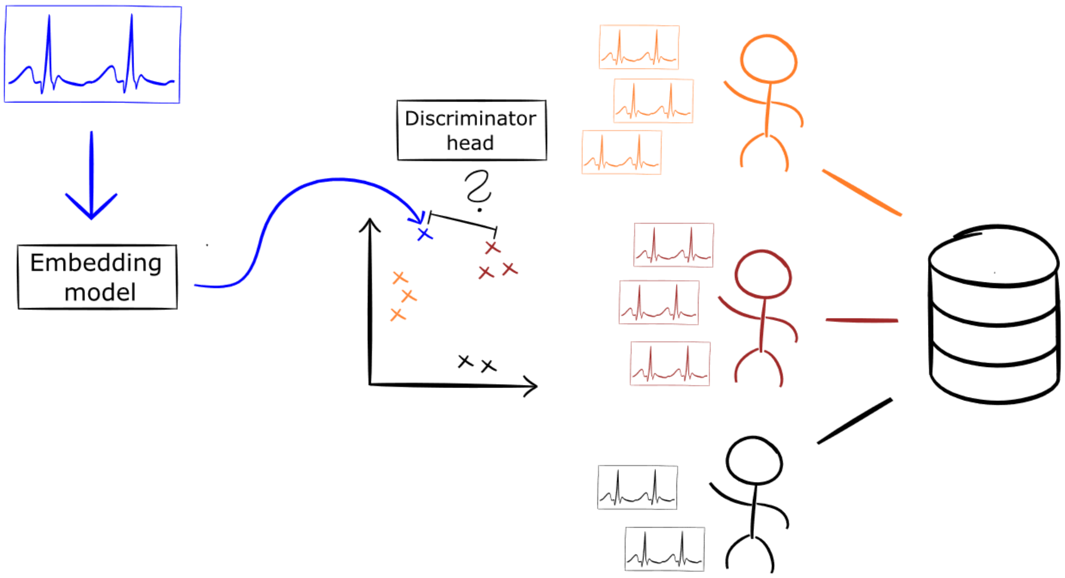
4. Methods:

We will employ various data science and machine learning techniques to develop our identification model. The approach includes:

- **Data Cleaning:** Removing noise and artifacts from the raw ECG data.
- **Feature Extraction:** Extracting relevant features such as QRS complex, P waves, T waves, and heart rate variability.

- **Data Exploration:** Analyzing the features to understand their distributions and relationships.
- **Machine Learning:** Training models using algorithms such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).
- **Model Architecture:** Consists of an Embedding model that generates embeddings for each ECG signal in a separate vector and a Discriminator head Which identifies the embeddings of each person.
- **Evaluation:** Assessing the model's performance using accuracy, precision, recall, and F1-score metrics.

Fig. 1. The overview of our ECG-based patient identification method.



5. Expected Outcomes and Success Criteria:

Expected Outcomes

1. **High-Accuracy Identification model**
   The primary outcome is to develop a robust identification model that accurately identifies individuals based on their ECG data. We aim for the model to achieve a high accuracy rate despite common ECG signal distortions.

2. **Distortion Resilience**
   The model will be evaluated for its resilience against various types of noises and distortions. This ensures our model's performance under different levels and types of noise to ensure reliability in real-world scenarios.

3. **Privacy protection**
   Implementation of privacy-preserving techniques to secure patient ECG data. This will involve using methods that anonymize the ECG data, ensuring that personal health information is protected throughout the identification process.

Success Criteria

1. **Accuracy metrics**
   The model should achieve an identification accuracy of at least 95% on the test dataset. Additionally, metrics such as precision, recall and F1-score should reflect high performance, particularly in scenarios involving noisy and distorted ECG data.

2. **Noise Robustness**
   The model's performance should not be deducted from the different kind of noise levels.This includes maintaining at least 90% of its accuracy under controlled noise conditions.

3. **Privacy Measures**
   The implementation of privacy measures will be considered successful if the model can effectively anonymize or encrypt ECG data without compromising identification accuracy.

6. Preliminary Bibliography:

1. Seják, M., Sido, J., & Žahour, D. (2023). ElectroCardioGuard: Preventing patient misidentification in electrocardiogram databases through neural networks. *Knowledge-based Systems*, *280*, 111014. https://doi.org/10.1016/j.knosys.2023.111014
2. Gupta, A., Huerta, E., Zhao, Z., & Moussa, I. (2021). Deep Learning for Cardiologist-Level Myocardial Infarction Detection in Electrocardiograms. *IFMBE Proceedings*, *80*, 341–355. https://doi.org/10.1007/978-3-030-64610-3_40
3. Mousavi, S., & Afghah, F. (2019). Inter- and Intra- Patient ECG Heartbeat Classification for Arrhythmia Detection: A Sequence to Sequence Deep Learning Approach. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1308–1312. https://doi.org/10.1109/ICASSP.2019.8683140
4. Al-Jibreen, A., Al-Ahmadi, S., Islam, S., & Artoli, A. M. (2024). Person identification with arrhythmic ECG signals using deep convolution neural network. *Scientific Reports*, *14*(1), 4431. https://doi.org/10.1038/s41598-024-55066-w
5. Butt, F. S., Wagner, M. F., Schafer, J., & Ullate, D. G. (2022). Toward Automated Feature Extraction for Deep Learning Classification of Electrocardiogram Signals. *IEEE Access*, *10*, 118601–118616. https://doi.org/10.1109/ACCESS.2022.3220670