ECG-Based Authentication Interface Project Report

Team member	ID	Department
محمد عبد الرؤوف أحمد عبد الشافي	20191700557	SC
محمد محمد خليفة حسن	20201701161	SC
كارولين عطا سامي برسوم	20191700445	SC
عبدالرحمن إمام فضل إمام	20201700432	SC

Team ID: SC_06

1. Introduction

Authentication interfaces based on electrocardiogram (ECG) signals have gained significant attention due to their potential for secure and user-friendly authentication. ECG signals, captured from the heart's electrical activity, offer unique biometric features that can be leveraged for identity verification.

In this project, we explore the development of an ECG-based authentication system. By analyzing ECG data, extracting relevant features, and employing machine learning classifiers, we aim to create a robust and efficient interface for user authentication.

Stay tuned for further details on our dataset, methodology, implementation, and results!

2. Dataset Description

- **Dataset Name**: PTB Diagnostic ECG Database v1.0.0
    - The database contains 549 records from 290 subjects (aged 17 to 87, mean 57.2; 209 men, mean age 55.5, and 81 women, mean age 61.6; ages were not recorded for 1 female and 14 male subjects). Each subject is represented by one to five records. There are no subjects numbered 124, 132, 134, or 161. Each record includes 15 simultaneously measured signals: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) together with the 3 Frank lead ECGs (vx, vy, vz). Each signal is digitized at 1000 samples per second, with 16 bit resolution over a range of ± 16.384 mV. On special request to the contributors of the database, recordings may be available at sampling rates up to 10 KHz.
    - Within the header (.hea) file of most of these ECG records is a detailed clinical summary, including age, gender, diagnosis, and where applicable, data on medical history, medication and interventions, coronary artery pathology, ventriculography, echocardiography, and hemodynamics. The clinical summary is not available for 22 subjects.

The diagnostic classes of the remaining 268 subjects are summarized below:

Diagnostic class	Number of subjects
Myocardial infarction	148
Cardiomyopathy/Heart failure	18
Bundle branch block	15
Dysrhythmia	14
Myocardial hypertrophy	7
Valvular heart disease	6
Myocarditis	4
Miscellaneous	4
Healthy controls	52

- **Recording Specifications**:
  - Non-commercial PTB prototype recorder
  - 16 input channels (14 for ECGs, 1 for respiration, 1 for line voltage)
  - Input voltage: ±16 mV
  - Resolution: 16-bit with 0.5 µV/LSB
  - Bandwidth: 0 - 1 kHz

- **Clinical Information**:
  - Detailed clinical summaries available for most records (age, gender, diagnosis, medical history, etc.)
  - Diagnostic classes include myocardial infarction, cardiomyopathy, bundle branch block, dysrhythmia, and more[1].

➢ ECG Preprocessing

To ensure the quality and reliability of our ECG data, we applied the following preprocessing techniques:

- **Butterworth Bandpass Filtering**:

  We used a Butterworth bandpass filter to remove noise and retain relevant frequency components.

  The filter was designed to allow frequencies within a specific range (e.g., 0.5 Hz to 50 Hz) while attenuating others.

  The function butter_bandpass_filter(data, low_cut, high_cut, fs_, order_) implemented this filter.

  Parameters:

  data: Raw ECG signal

  low_cut: Lower frequency cutoff (e.g., 0.5 Hz)

  high_cut: Upper frequency cutoff (e.g., 50 Hz)

  fs_: Sampling frequency

  order_: Filter order

- **Normalization**:

  We standardized the ECG data using z-score normalization.

  The function Normalize(df) performed this normalization.

  Standardization ensures that features have similar scales, which is crucial for machine learning algorithms.

3.Feature Extraction

Extracting relevant features from ECG signals is essential for subsequent classification. Here's how we extracted fiducial features:

- **Segmentation**:

  We segmented the ECG signal around R-peaks (QRS complexes) to isolate individual heartbeats.

  The function Segmentation(channel_record) identified R-peaks using the Pan-Tompkins algorithm.

  Parameters:

  channel_record: ECG signal from a specific lead (e.g., lead II)

  Segments were typically centered around R-peaks and had fixed lengths (e.g., 600 samples).

- **Fiducial Points Extraction**:
  - ✓ For each segmented heartbeat, we located key fiducial points:
  - ✓ R peak (R-wave peak)
  - ✓ Q peak (Q-wave peak)
  - ✓ S peak (S-wave peak)
  - ✓ QRS onset and offset
  - ✓ P peak (P-wave peak)
  - ✓ P onset and offset
  - ✓ T peak (T-wave peak)
  - ✓ T onset and offset

  The function findFiducialPoints(segment_) identified these points based on specific criteria.

- **Feature Calculation**:

- From the fiducial points, we computed various features:
    - Time intervals (e.g., R-R interval, QRS duration)
    - Amplitude differences (e.g., R peak to S peak)
    - Slopes (e.g., P-wave slope)
    - Morphological features (e.g., T-wave shape)
    - The function find Features (fiducialPoints, signal) calculated these features.
    - Features were used as input for subsequent classification models.

By combining robust preprocessing and feature extraction, we prepared the ECG data for accurate classification.

4.classifier used

We tried 3 classifiers on the extracted features.

- **parameters SVM classifier:**
    - Kernel: RBF
    - Regularization parameter (C): 1.0
    - Random state: 42
    - Probability estimates: Enabled
- **parameters KNN classifier:**
    - Number of neighbors (n_neighbors): 4

- **parameters Decision tree classifier:**
    - Random state: 42

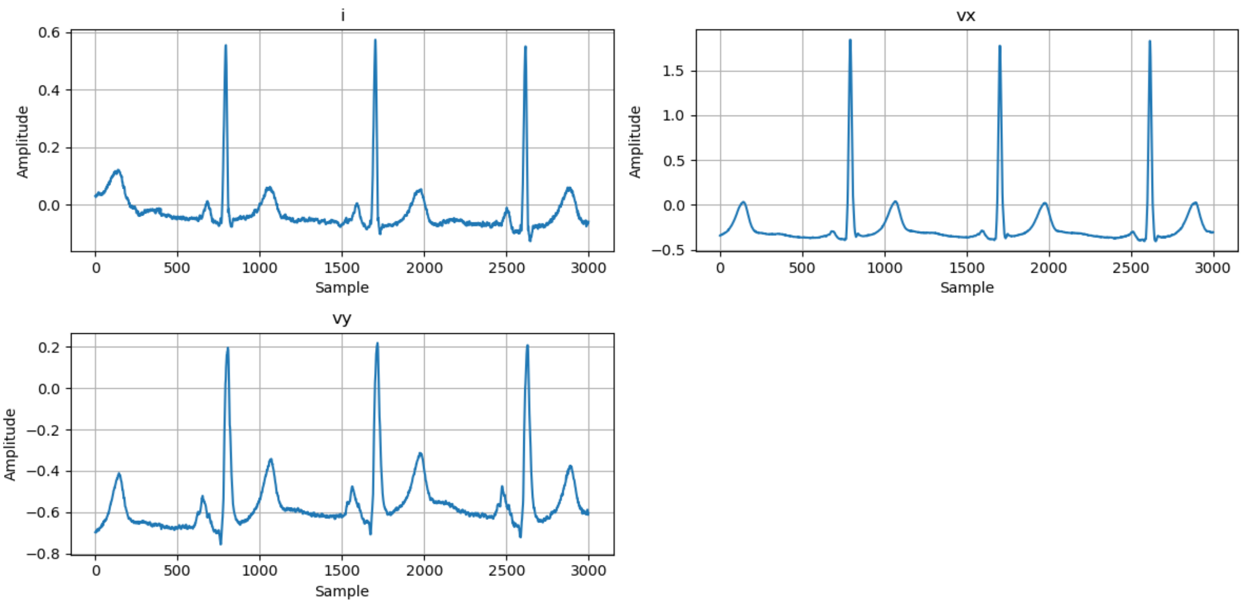
5.classifier results

Model	Accuracy
SVM	0.986
KNN	1.00
Decision tree	0.993

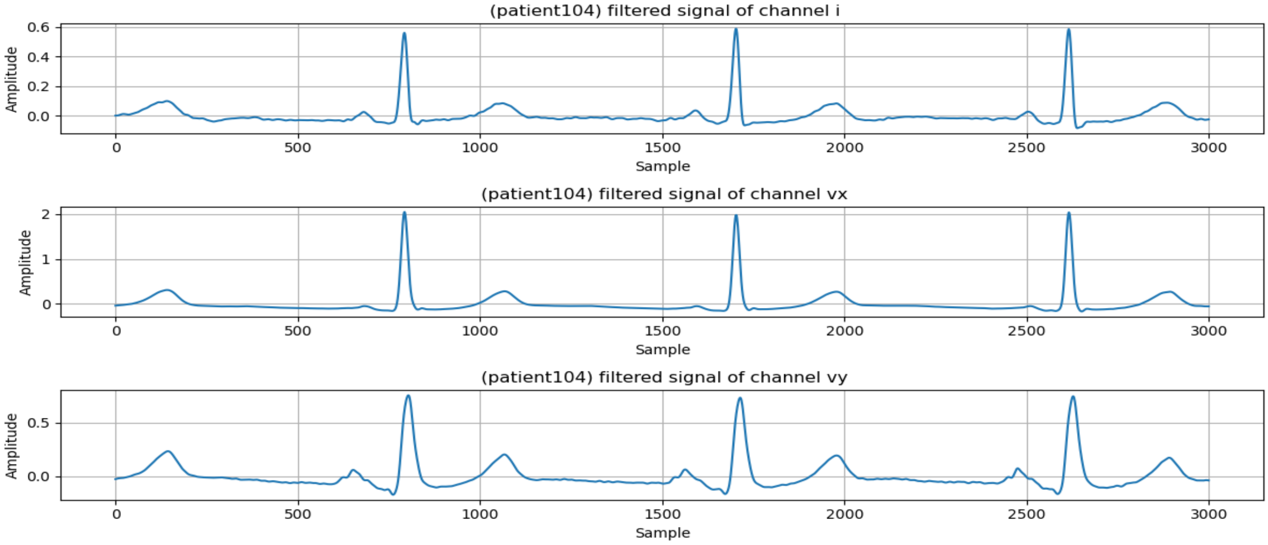
6. Screen shots for your running interface

- **Plots:**

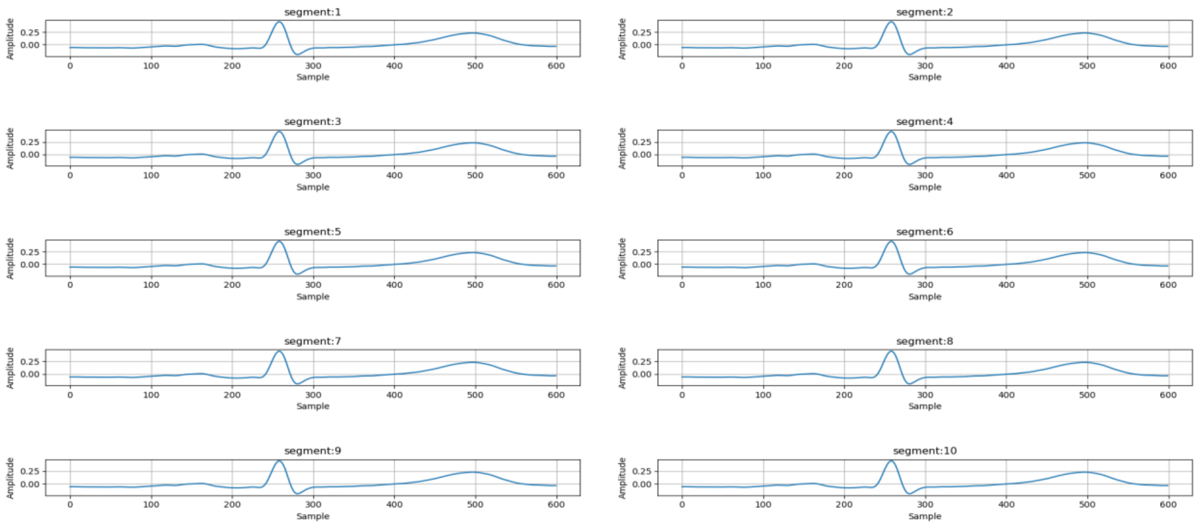
1. Sample of raw data after reading:



2. Sample of preprocessed data:



3. Sample of data after segmentation:



4. Sample of segment and fiducial points on it: