DTEK0086 Biosignal Analytics

Arrhythmia detection using electrocardiography

Background

Cardiac arrhythmias are medical conditions in which the heartbeat is irregular. They result from variations in the timing and frequency of the heart activity. Arrhythmias can cause lightheadedness, chest pain, and shortness of breath. They might be life-threatening and, in extreme cases, lead to sudden death. Most arrhythmias can be effectively treated, e.g., by medication or pacemaker, if they are diagnosed. Machine learning algorithms have been utilized for arrhythmia detection in several studies [1].

Objective

The objective of this project is to perform arrhythmia detection using features extracted from the shape of the heart cycles in ECG signals. Using machine learning, you need to differentiate the ECG signals into two classes: i.e., 1 for the signals that contain arrhythmia and 0 for the signals with normal heart cycles. The analysis should be done in Python (more details in the Instruction Section).

For this course project, you need to:

1. Submit your Python script and your report of the observations, graphs, and conclusions made upon analyzing the given signals. It is suggested to submit a Jupyter Notebook file, including your code and report.
2. Give a 20-minute presentation about your work. Your presentation should include descriptions of
    a. The problem and the biosignal
    b. The steps in your analysis: e.g., what pre-processing methods you use, which features you extract, which machine-learning algorithms you use
    c. The results that you obtain: e.g., the accuracy of two machine learning methods
    d. Your evaluation and conclusion on the findings and methods

Data collection setup

The Lead II ECG was measured from the limbs (left leg – right arm). The single-channel ECG was collected from the subjects with arrhythmias. The signal was digitized at 360 Hz with 11-bit resolution over a 10mV range. The corrupted signals due to noises or loose electrode collections were removed. The data annotation was performed by human experts. The data is extracted from the Physionets MIT-BIH arrhythmia database (mitdb).

Structure of the data

The project includes the ECG of 25 subjects. Each record (i.e., file) consists of 10 seconds of ECG signals, and it corresponds to a normal ECG or an ECG with arrhythmias. The dataset includes two folders as “normal” and “abnormal.” The “normal” folder contains 3855 records of normal ECG. The “abnormal” folder contains 4046 ECG signals with arrhythmias. Each record is saved as a CSV file, including a column representing the Lead II ECG. The filename includes the event number and the subject ID. For example, 0 is the event number, and 100 is the subject ID in “sample_ID100_0.csv.”

Instruction

For the analysis, you should:

1. Use pre-processing techniques (such as filtering) if necessary.
2. Detect R-peaks of each ECG record
3. Extract features from the shape of the heart cycles: e.g., amplitude and duration of QRS complex, duration of the ST segment, T-wave duration (see ecg features).
4. Divide your dataset into the training set and test set (70-30% split).
5. Standardize your data: i.e., use the mean and standard deviation of the training data to standardize the training data and the test data.
6. Select two supervised machine learning algorithms and train two classifiers using the train set. Each classifier should predict 1 in the presence of arrhythmia or 0 otherwise.
7. Compare the two classifiers by evaluating the results using the test set.
    a. Obtain the confusion matrix, accuracy, precision, recall, and F1-score. These can be calculated from the predicted and true values.

Hint: you can utilize packages like scipy for the pre-processing, biosppy for R-peak detection, and scikit-learn for machine learning.

[1] Soman, Thara, and Patrick O. Bobbie. "Classification of arrhythmia using machine learning techniques." WSEAS Transactions on computers 4.6 (2005): 548-552.