University of Glasgow | School of Computing Science

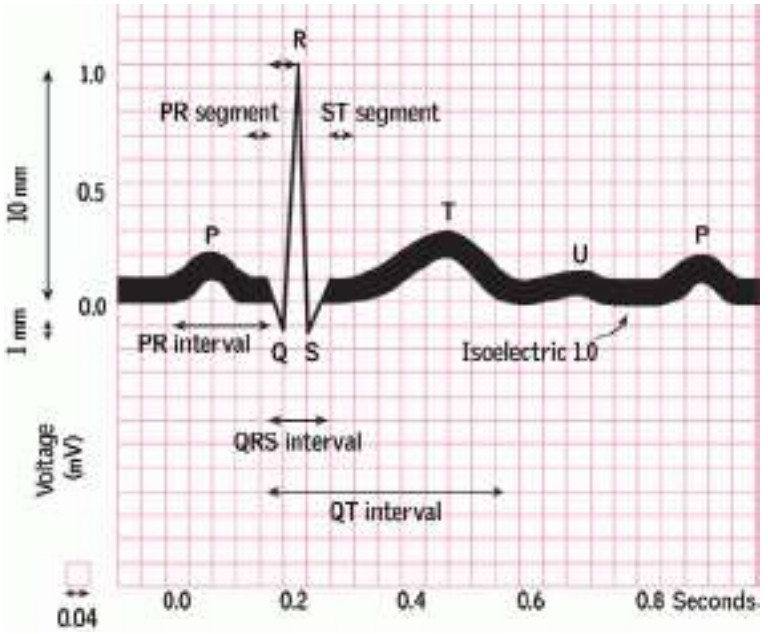
THE AWARDS 2020 | UNIVERSITY OF THE YEAR

Preprocessing of ECG Signal

Dr. Fani Deligianni,
fani.deligianni@glasgow.ac.uk
Lecturer (Assistant Professor)
Lead of the Computing Technologies for Healthcare Theme
https://www.gla.ac.uk/schools/computing/staff/fanideligianni

Electrocardiogram (ECG)

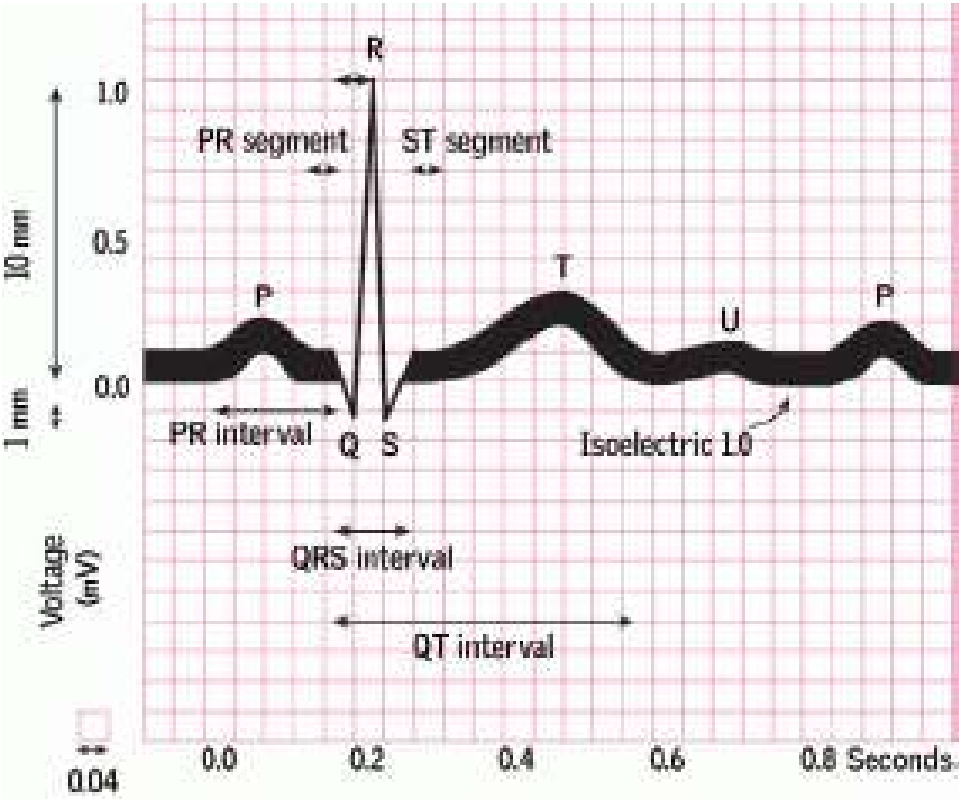
- An ECG test consist of collecting data through the electrical activity of the human cardiovascular system
- ECG consist of three key features which represent distinct stages of the heartbeat.
  - **P-wave:** Depolarization of the atria.
  - **QRS complex:** Depolarization of the ventricles.
  - **T-wave:** Re-polarization of the ventricles.



Cardiovascular Medicine. In: Oxford handbook of clinical medicine, 10th edition

ECG Classification

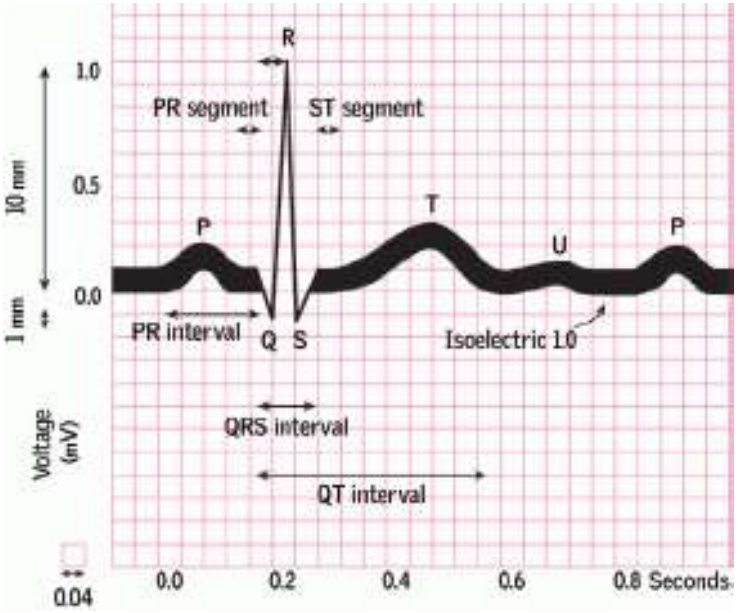
- Manual ECG analysis is time-consuming and error prone
- ECG abnormalities may require continuous monitoring
- Machine learning has been extensively applied in ECG classification



Cardiovascular Medicine. In: Oxford handbook of clinical medicine, 10th edition

Noise Interference

- The ECG signals are extremely susceptible to high and low frequency noise. These noise usually occur from:
  - Baseline wander
  - Misplaced electrode contact
  - Motion artifacts
  - Power line interference



Cardiovascular Medicine. In: Oxford handbook of clinical medicine, 10th edition

MIT-BIH ECG Dataset

- The MIT-BIH dataset used for this investigation is a public database consisting of a large number of annotated beats.
- It is frequently used for time-series classification research.

- The MIT-BIH Arrhythmia Database contains sections of ambulatory ECG recordings:
  - From 47 subjects, digitized at 360 samples per second per channel.
  - 11-bit resolution at 10-mV range on two channels.
  - Here 23 recordings were picked at random from a set of 4000 24-hour ECG recordings.
  - Collected from a population 60% of inpatients and 40% outpatients.

MIT-BIH ECG Dataset

- This data has been pre-annotated and labelled by cardiologists.
- These different annotations refer to various normal and abnormal ECG signals which represent different types of arrhythmia.
- The dataset consists of ECG signals of various classes, but the eight classes used for this investigation are ’N’, ’L’, ’R’, ’V’, ’A’, ’F’, ’f’, ’/’.
- The table shows the description and numerical identification values assigned to these classes.

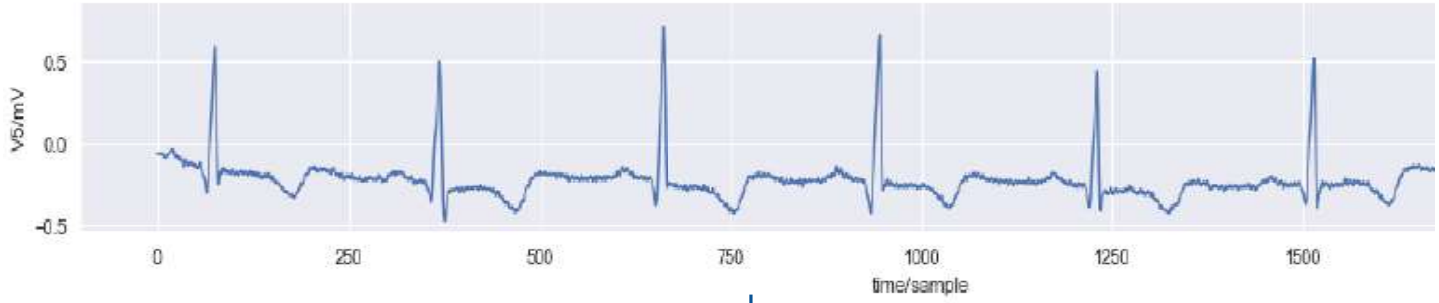
Class	ID	Beat Description
N	1	Normal
L	2	Left Bundle Branch Block
R	3	Right Bundle Branch Block
V	4	Premature Ventricular Contraction
A	5	Atrial Premature
F	6	Fusion of Ventricular and Normal
f	7	Fusion of Paced and Normal
/	8	Paced

Data Pre-processing

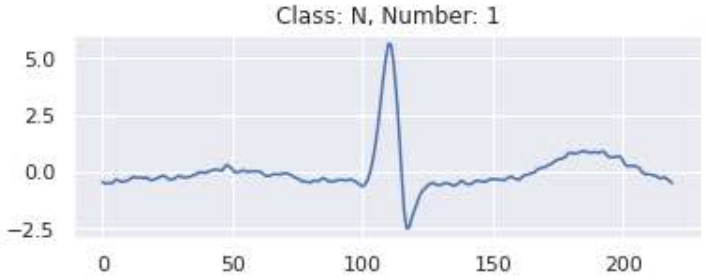
Raw Data

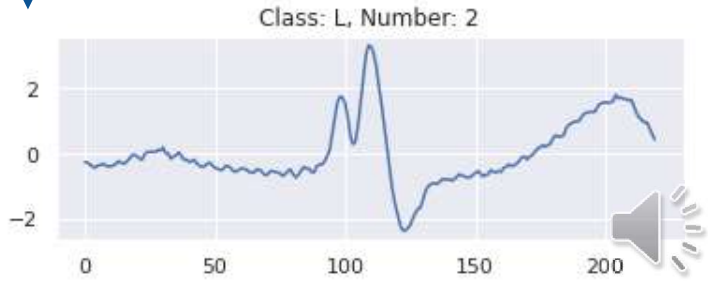
Data Pre-processing

Input Data



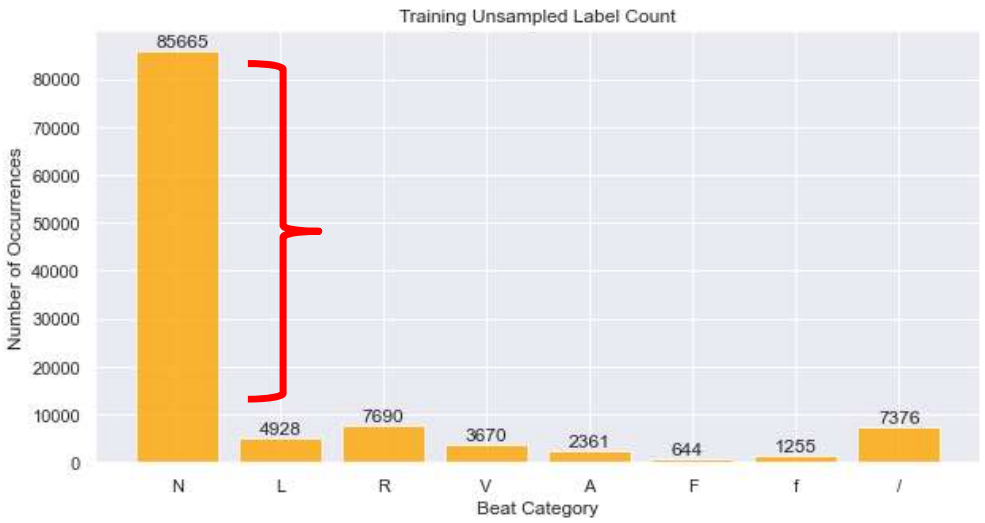
Individual Beat, Centre R-peak, Standardize, Beat Annotation





Class Imbalance

- The normal class is over-represented in the data
- Resampling is based on a **bootstrap method** which resamples a dataset with replacement, iteratively
- For **up-sampling and down-sampling**, the sample value was calculated by taking the mean values of the total number of beats of the abnormal classes.



Summary

- Preprocessing of the ECG signal include:
  - Filtering to remove noise
  - Annotation of the R-peaks
  - Segmentation of the recordings into ECG beats
  - Resampling the data to address the imbalance problem

References

- Mark RG et al. ‘An annotated ECG database for evaluating arrhythmia detectors’, IEEE Transactions on Biomedical Engineering 29(8):600, 1982
- Moody et al. ‘The impact of the MIT-BIH arrhythmia database’, IEEE Engineering in Medicine and Biology Magazine 20(3), 45-50, 2001
- Yola et al. ‘Improving ECG Classification Interpretability using Saliency Maps’, IEEE BIBE, 2020.