

## REPORT PROJECT I

### 1. PROBLEM

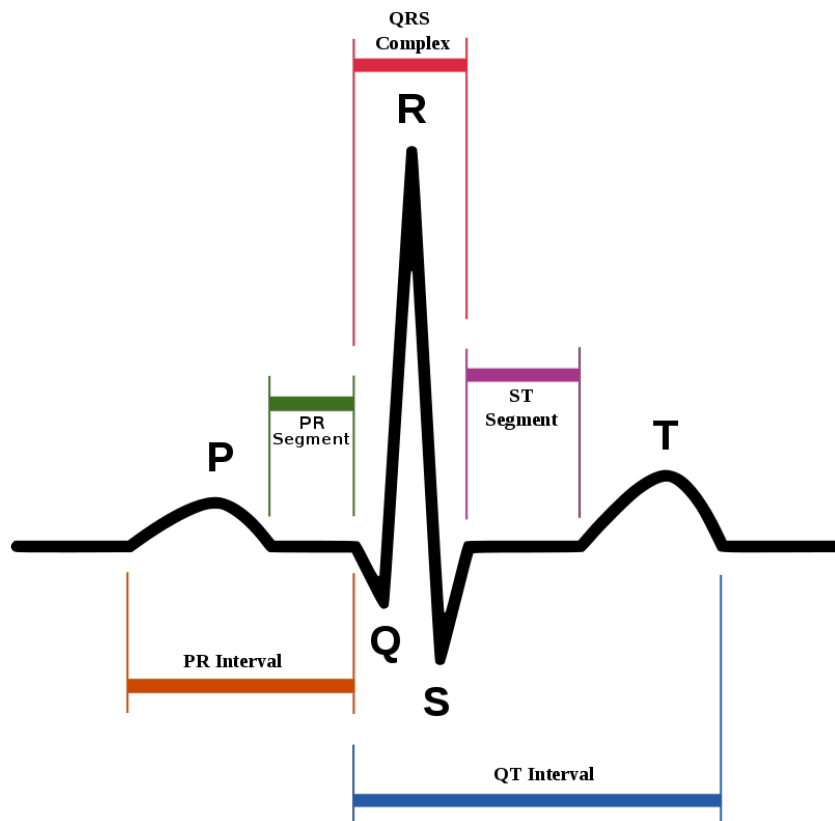
---

Prediction and Classification of Cardiac Arrhythmia.

#### a. Introduction:

Arrhythmia can be diagnosed by measuring the heart activity using an instrument called Electrocardiogram (ECG) and then analyzing the recorded data. The ECG is an established technique in cardiology for the analysis of cardiac condition of the patients. In its basic definition, ECG is the electrical representation of the contractile activity of the heart, and can be recorded fairly easily by using surface electrodes on the limbs or chest of the patient. The ECG is one of the most recognized and used biomedical signal in the field of medicine. The rhythm of the heart in terms of beats per minute (bpm) can be easily calculated by counting the R peaks of the ECG wave during one minute of recording.

Below is a schematic sample of a normal ECG.



## REPORT – CAPSTONE PROJECT I

Irregularity in heart beat may be life threatening. Hence both accurate detection of presence as well as classification of arrhythmia are important.

Different parameter values can be extracted from the ECG waveforms and can be used along with other information about the patient like age, medical history, etc to detect arrhythmia. However, sometimes it may be difficult for a doctor to look at these long duration ECG recordings and find minute irregularities. Therefore, using machine learning for automating arrhythmia diagnosis can be very helpful.

This project aims using different machine learning algorithms for predicting and classifying arrhythmia into different categories.

## 2. DATA WRANGLING

---

### a. Examining the Features and Samples

#### Features:

F.Nu.	Name	Explanation	Lin./Nom.
1	Age	Age in years	Linear
2	Sex	0 = male; 1 = female	nominal
3	Height	Height in centimeters	Linear
4	Weight	Weight in kilograms	Linear
5	QRS duration	Average of QRS duration in msec.	Linear
6	P-R interval	Average duration between onset of P and Q waves in msec.	linear
7	Q-T interval	Average duration between onset of Q and offset of T waves in msec.	Linear
8	T interval	Average duration of T wave in msec.	Linear

## REPORT – CAPSTONE PROJECT I

9	P interval	Average duration of P wave in msec.	linear
	Vector angles in degrees on front plane of:		
10	QRS		linear
11	T		linear
12	P		linear
13	QRST		linear
14	J		linear
15	Heart rate	Number of heart beats per minute	Linear
	Of channel DI: Average width, in msec.		
16	Q wave		linear
17	R wave		linear
18	S wave		linear
19	R' wave	small peak just after R	linear
20	S' wave		linear
21	Number of intrinsic deflections		linear
22	Existence of ragged R wave		nominal
23	Existence of diphasic derivation of R wave		nominal
24	Existence of ragged P wave		Nominal
25	Existence of diphasic derivation of P wave		nominal
26	Existence of ragged T wave		nominal
27	Existence of diphasic derivation of T wave		nominal
28 .. 39	Of channel DII	(similar to 16 .. 27 of channel DI)	
40 .. 51	Of channels DIII		
52 .. 63	Of channel AVR		
64 .. 75	Of channel AVL		
76 .. 87	Of channel AVF		
88 .. 99	Of channel V1		
100..111	Of channel V2		
112..123	Of channel V3		
124..135	Of channel V4		

136..147	Of channel V5		
148..159	Of channel V6		
	Of channel DI	Amplitude , * 0.1 milivolt, of	
160		JJ wave	Linear
161		Q wave	Linear
162		R wave	Linear
163		S wave	Linear
164		R' wave	Linear
165		S' wave	linear
166		P wave	linear
167		T wave	linear
168	QRSA	Sum of areas of all segments divided by 10,( Area= width * height / 2 )	linear
169	QRSTA	QRSTA = QRSA + 0.5 * width of T wave * 0.1 * height of T wave. (If T is diphasic then the bigger segment is considered)	Linear
170..179	Of channel DII		
180..189	Of channel DIII		
190..199	Of channel AVR		
200..209	Of channel AVL		
210..219	Of channel AVF		
220..229	Of channel V1		
230..239	Of channel V2		
240..249	Of channel V3		
250..259	Of channel V4		
260..269	Of channel V5		
270..279	Of channel V6		

**b. Class Distribution:**

Class code	Class	Number of instances
01	Normal	245

## REPORT – CAPSTONE PROJECT I

02	Ischemic changes (Coronary Artery Disease)	44
03	Old Anterior Myocardial Infarction	15
04	Old Inferior Myocardial Infarction	15
05	Sinus tachycardy	13
06	Sinus bradycardy	25
07	Ventricular Premature Contraction (PVC)	3
08	Supraventricular Premature Contraction	2
09	Left bundle branch block	9
10	Right bundle branch block	50
11	1. degree AtrioVentricular block	0
12	2. degree AV block	0
13	3. degree AV block	0
14	Left ventricle hypertrophy	4
15	Atrial Fibrillation or Flutter	5
16	Others	22

There are (452) rows, each representing medical record of a different patient. There are 279 attributes like age, weight and patient's ECG related data.

The data set is labeled with 16 different classes. Classes 2 to 15 correspond to different types of arrhythmia. Class 1 corresponds to normal ECG with no arrhythmia and class 16 refers to unlabeled patient.

The data set is heavily biased towards the no arrhythmia case with 245 instances belonging to class 1 and 185 instances being split among the 14 arrhythmia classes and the rest 22 are unclassified.

The main challenges in processing this data set are the limited number of training examples compared to the number of features, heavy bias towards the case of normal ECG, missing feature values and feature values belonging to both continuous and categorical types.