CAN WE TEACH A MACHINE TO BE CARDIOLOGIST?

Hien Long Nguyen
A1798520

Supervisor: Prof. Derek Abbott

Co-supervisor: Mohsen Dorraki

Submitted in partial fulfilment of the requirements for the degree of
Bachelor of Engineering (Honours)

in

Electrical and Electronic

The University of Adelaide
2021

Keywords

Electrocardiogram, Machine learning, Deep Learning, Neural Network, Convolutional Neural Network, Support Vector Machine, Long-Short Term Memory

Abstract

Electrocardiogram (ECG) is one of the most popular biomedical signals that cardiologists use to diagnose cardiovascular diseases. It can clearly reflect the cardiac system of human body thanks to sophisticated 12-lead ECG recording system. However, this technique is non-invasive method to measure electrical activity of the heart and output graphical results as image of difference in electrical potential between parts of the body, especially around the heart. By analysing feature in the ECG image, expect can predict abnormality of the heart and give early advice to patients, avoiding premature death. However, the diagnosing process can be challenging duel to noise from hospital environment. Therefore, human is deploying computers to do the job of cardiologists. Thanks to their computational power and effective machine learning algorithms, computer can deal with long/noisy recordings and predict cardiac problems with high accuracy.

Therefore, it is possible to teach a machine to be a cardiologist. To do that, the ECGs must first be examined and denoised to for analysing and feature extraction. From useful attributes, machine learning algorithms can be utilized to effectively learn hidden patterns from ECG data and classify classes based on difference patterns. Finally, a validation process is conducted to assess the performance of classifiers. Throughout these steps, a computer can imitate the way human learn and help us to analyse physiological signal. With machine computational power, it can help reduce the burden of biomedical expert as well as monitor human health, improving our quality of life.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
AF	Atrial Fibrillation
ANN	Artificial Neural Network
ARR	Arrhythmia
AUC	Area Under Region of Convergence Curve
CNN	Convolutional Neural Network
CVD	Cardiovascular Disease
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
HHT	Hilbert-Huang Transform
IF	Instantaneous Frequency
LSTM	Long-Short Term Memory
ML	Machine Learning
MoV	Moment of Velocity
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristics
SC-LNLMS	Self-Correcting Leaky Normalised Least Mean Squares
SE	Sensitivity
SVM	Support Vector Machine
WHO	World Health Organisation

Acknowledgements

I would like to express my special gratitude to my supervisors Professor Derek Abbott and Mohsen Dorraki for their precious advice and support throughout this interesting project. Without your valuable guidance we could not complete this research.

For Sonia, my partner, it is an honour to have an opportunity to study and work with a proactive and supportive teammate like you. Throughout the project we have performed a good teamwork and overcome challenge to achieve the final goal.

For my family and friends, I am grateful for your endless courage and support.

Chapter 1: Introduction

1.1 CONTEXT AND MOTIVATION

Cardiovascular diseases (CVDs) took the majority of lives in Australia, responsible for 26% of all deaths, and influenced a total of 4 million Australians since 2019 [1]. Financially, these diseases cost more than any other illnesses, 5 billion was paid annually for heart failure treatment [1] .Globally, these heart disorders contributed to more than 17.9 million deaths, taking the majority 31% of total death causes, particularly in low- and middle-income countries [2]. In order to reduce the burden of these health problems on human resources as well as the economy, early heart abnormality detection and treatment are necessary.

With that primary goal, this project will aim to develop machine-learning algorithms to reflect the cardiac health of human body. This is a promising and low-cost solution since it uses electrocardiogram (ECG) and deploys computation power of computers to monitor people cardiovascular system, and hence give early advice on their health, avoiding premature deaths.

1.2 ELECTROCARDIOGRAM

Electrocardiogram is a non-invasive and standard tool to diagnose and report electrical activity of the heart via electrode nodes connected people skin. This type of signal is recorded mainly on the surface of human body [3]. The morphology and beat intervals time series analysed through ECG waveform represent the cardiac systems. Any irregularity in morphological patterns and time series leads to indications of cardiac arrhythmia. The waveform contains valuable information about the heart activities: depolarization and repolarization of ions named Na^+ and K^- .

1.3 ECG WAVEFORM AND HEART FUNCTION

There are four champers in the heart named: left atrium, right atrium, left ventricle, right ventricle and some nodes called sinoatrial and atrioventricular as illustrated in Figure 1. The atria are electrically isolated with the two ventricles below by fibrous - a non-conductive tissue. To receive oxygen, the right atrium and

ventricle work together to form a circular blood flow from the heart to lungs. The blood containing poor oxygen is received by inferior vena cava and superior veins, it then travels into the heart through the right atrium. The blood is forced down to the right ventricles by the contraction of the atrium above and it is later bumped through the lungs when the right ventricles is fully sketched and contracts. With the same mechanism, the left atrium and left ventricles work together to receive the enrich-oxygen blood from the lungs via the pulmonary veins and force it to flow through the rest of the human body [4].

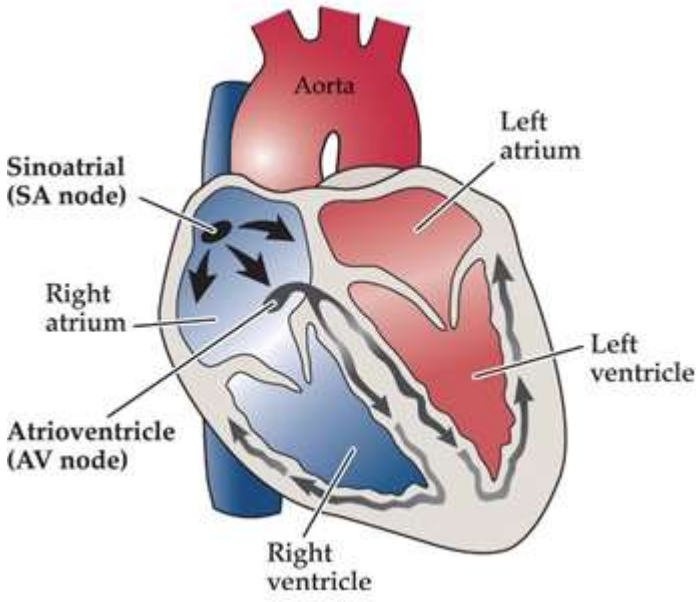

Figure 1. Heart conducting system [3]

The activities of the heart are controlled mainly by Sinoatrial (SA) and Atrioventricular (AV) node. The SA node is the natural pacemaker of heart rate since it initializes the heat beat. It generates electrical impulses used to initialise the contraction of the two atria. Similarly, the generated current propagates to the AV node and control the ventricles contraction. These control signals are generated and passed through myocardium thanks to a process called depolarization and repolarization as shown in Figure 2. In a resting stage cardiac cells are polarized, meaning that there is an electrically balanced state between potassium and sodium inside and outside the cardiac cells respectively, no current flows. In the depolarization, the cells' membrane is modified for the ion exchange, sodium is pulled inward, and potassium is pushed outward, therefore the current is generated. These impulses circulate mainly inside the heart, however there still a significant

portion flows to the body surface and can be detected by the electrodes placed in selected points [5]. These electrical potentials are recorded as ECG signal. Finally, a reversed process named repolarization occurs and recovery the excitable cells to resting state again from the discharged state, preparing for the next cycle.

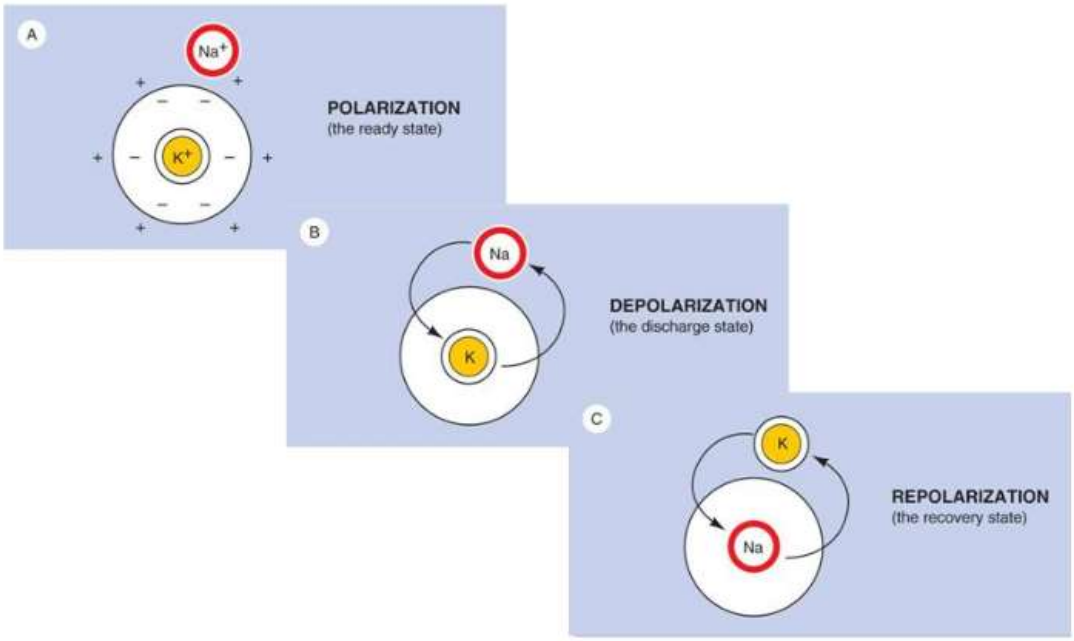

Figure 2. Depolarization and Repolarization processes [5]

1.4 LEADS IN ECG

An ECG lead is a pair of electrodes (positive (+) and negative (-)) that records a graphical result of the heart's electrical activity, particularly the fluctuation between different voltages. These electrodes measure the leaked impulses generated by the depolarisation process as difference in electrical potential between electrodes. The standard 12-lead ECG recording process uses total 10 leads attached to 10 different anatomical positions on human body, 4 leads for each limb and 6 leads across the chest as shown in Figure 3. These leads work simultaneously in the ECG recording process and give 12 views from frontal to horizontal angles of the heart.

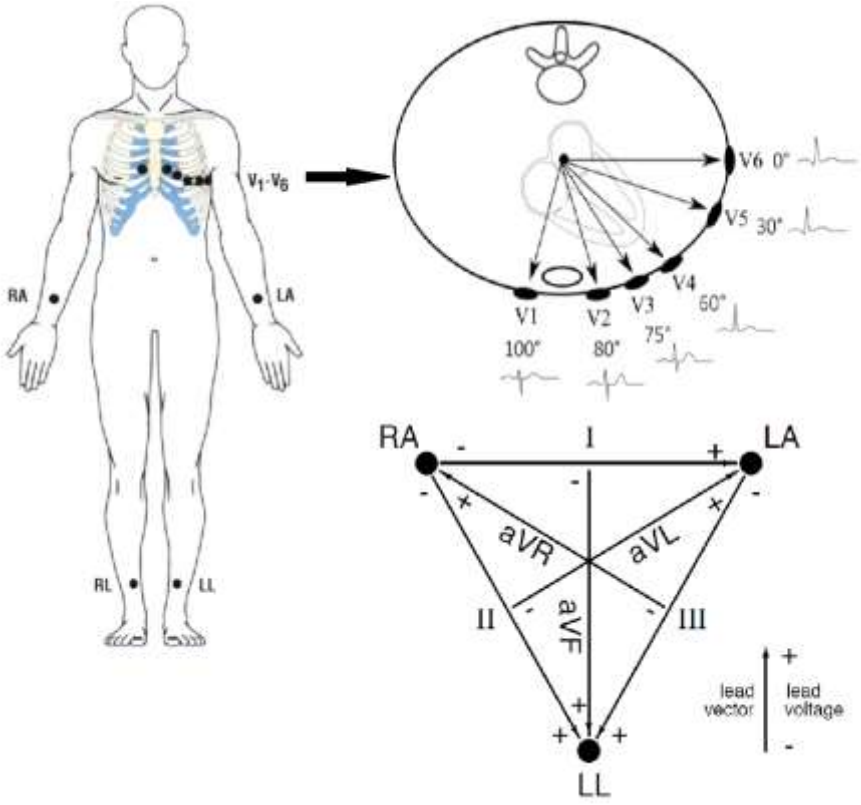

Figure 3. Leads' placement and Einthoven's triangle

There are two types of leads:

Bipolar: a pair of positive and negative electrodes used to measure the electrical potential between two selected positions. The positive pole is served as the point of view and the negative one is served as the zero reference.

Unipolar: also consists of a single positive electrode, but unlike the bipolar system, the unipolar utilize a combination of two other electrodes to form a zero reference.

Limb lead types deploy 4 limb-electrodes, right leg electrode is considered as the ground and the others form the Einthoven's triangle in Figure 3. Leads' placement and Einthoven's triangleFigure 3. Standard limb leads I, II and III are bipolar, they measure the potential difference between 2 of the 3 limbs in the triangle. On the other hand, augmented limb leads AVR, AVL and AVF are unipolar and use one electrode as positive pole and take the average of the two electrodes' input as the zero reference. For chest leads (precordial leads) from V1 to V6, the positive poles are corresponding electrodes, and the reference negative value is the

same for all chest leads and is calculated as an average of inputs from the 3 limb electrodes. Like the augmented limb leads, they are also unipolar but give the horizontal views of the heart rather than the frontal views displayed by the 6 limb leads. The characteristic of each lead is illustrated in Table 1. The technical recording mechanism of every lead based on depolarization and repolarization process. When the depolarization occurs toward a lead, a positive deflection is detected, and the depolarization away a lead results in negative deflection. Similarly, the reverse is true for repolarization. Therefore, graphical results from ECG are different from one another in waveforms, depending on the selected electrodes.

Category	**Lead**	**Type**	**Positive Electrode**	**Negative Electrode**	**Heart Surface viewed**
Standard limb lead	I	Bipolar	Left arm	Right arm	Lateral
	II	Bipolar	Left leg	Right arm	Inferior
	III	Bipolar	Left leg	Left arm	Inferior
Augmented limb leads	AVR	Unipolar	Right arm	Midway between left arm and left leg	None
	AVL	Unipolar	Left arm	Midway between right arm and left leg	Lateral
	AVF	Unipolar	Left leg	Midway between left arm and left leg	Inferior
Chest leads	V1	Unipolar	Right side of sternum, 4^{th} intercostal space	Heart	Interventricular septum

	V2	Unipolar	Left side of sternum, 4th intercostal space	Heart	Interventricular septum
	V3	Unipolar	Midway between V2 and V4	Heart	Anterior surface
	V4	Unipolar	Left midclavicular line, 5th intercostal space	Heart	Anterior surface
	V5	Unipolar	Left anterior axillary line, 4th intercostal space	Heart	Lateral surface
	V6	Unipolar	Left midaxillary line, 5th intercostal space	Heart	Lateral surface

Table 1. Characteristics of ECG leads [5]

According to S. Meek [4], from all the mentioned leads, lead II which gives good graphical results of the ECG recording, especially the P wave, is prolonged for more accurate assessment of heart rhythms.

1.5 ECG WAVEFORM

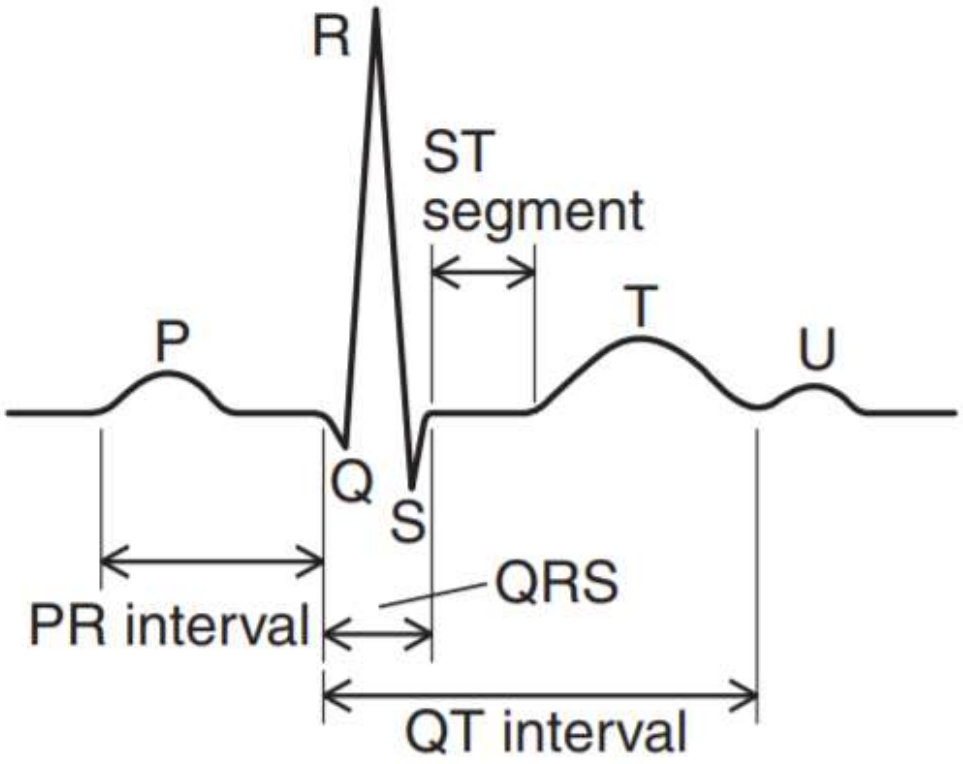

Figure 4. ECG waveform and its components [3]

Waves	**Amplitude (mV)**	**Duration (ms)**	**Presentation**
P wave	<= 0.3	<= 110	Initialization of the electrical impulse. Depolarization and contraction initialization of the left and right atria
PR interval		75-110	Conduction duration from the atria to the ventricles. It has a delay to ensure the atria fully bump blood to the ventricles.
QRS complex	> 0.5	< 120	The spread of the electrical

			impulse. Ventricular contraction and depolarization of the left and right ventricles. Indication of atria repolarization.
ST segment		<120	The interval between ventricular depolarization and repolarization
T wave	<= 0.5	<= 200	Ventricular repolarization
QT interval		380 ~40% R-R interval	Cycle duration of ventricular depolarization/repolarization
R-R interval		400-1200	One complete cardiac cycle

Table 2. Characteristics of ECG waveform components [3]

Characteristics and meaning of each waveform and segment of typical Lead II are shown in table 2 above. The QRS complex has the greatest voltage deflection due to the large volume of the ventricular tissue compared with the atria. Although has a highest amplitude, the duration of the QRS complex is smaller than that of T wave since the repolarization of the ventricles occurs more slowly than the depolarization.

These values may vary, depending on human characteristics, gender, or ages. However, the abnormality in waveform morphology results in cardiac arrhythmia.

1.6 ATRIAL FIBRILLATION

Atrial Fibrillation (AF) is a kind of heart rhythm that beats rapidly and irregularly compared with normal sinus (NS) rhythm [6]. It occurs when the atria chaotically beat and out of sync with the ventricles of the heart. Therefore, it often causes shortness of breath and bounding heartbeat called palpitation. Although the AF itself is not considered as life-threatening disorder, it can lead to other dangerous CVDs such as stroke and heart failure.

Chapter 2: Literature Review

In this chapter, related works on ECGs classifications are discussed. It presents steps to analyse ECG signals from pre-processing, features extraction and selection to machine learning classification.

2.1 SIGNAL PRE-PROCESSING

According to J. Rodrigues and et al [7], ECG is non-stationary in nature, corrupted by many noise sources when being recording in a noisy hospital environment. There are some types of noise that have the frequency band overlapping the frequency of interest in the ECG, hence affecting the characteristics of the recording. As shown in Figure 5 below, there 4 primary noise sources that highly pollute the human physiological recordings e.g., motion artifacts, baseline drift, muscle contraction, and power line noise [7]

For a better signal analysis and recognition ability of machine learning algorithms, signal preparation step is needed. Different types of filters are applied to smoothen the signal or remove noisy segments. Various well-known denoising techniques are discussed in this section.

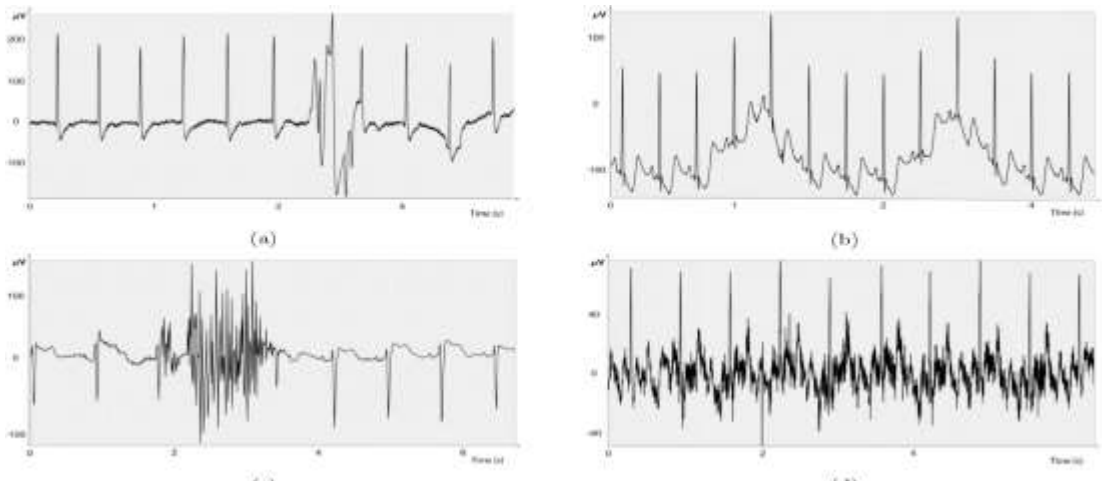

Figure 5. Typical noises in ECG signal: a) motion artifacts, b) baseline drift, c) muscle contraction, d) power line noise [7]

2.1.1 Motion artifact

Motion artifacts are technically transient baseline change with extremely high amplitude, much higher than normal inter-beats in the same recording. Normally, it

happens in short period of time. The presence of this artifact is due to movements of a human body that sketch skin locations attached with electrodes, altering the impedance around of the recording areas [7].

Figure 5.a illustrates an ECG signal suffered from motion artifacts. The contaminated segment can be removed by applying an adaptive filter [8].

2.1.2 Baseline drift noise

Baseline drift noise is the interference of low frequency under 0.5 Hz with the ECG recording, leading to a variation of the baseline. This type of noise is typically caused by random coughing or strong respiration period, resulting in variation of the heart. This noise can be clearly seen in chest leads' recordings. In addition, large movements of the human body are also responsible to the baseline drift that can be usually detected in the recordings of the 6 limb leads [7].

Figure 5.b shows how baseline drift noise impacts an ECG recording. Since the frequency band of the drift is no more than 0.5 Hz, a high pass filter with cutoff frequency of 0.5 Hz is applied to smoothen the noisy segment [8]. High-pass filter can attenuate/remove frequency component which is lower than the cutoff frequency. Therefore, high-pass filter with cutoff frequency of 0.5Hz will reduce baseline noise lower than 0.5Hz.

2.1.3 Muscle contraction noise

Muscle contraction noise is also known as Electromyogram (EMG) artifacts which has high frequency noise due to electrical activities of human muscle when it contracts [7]. The waveform contaminated with muscle contraction is characterized by unpredictable spikes of amplitude in ECG data.

Figure 5.c describes an ECG signal contaminated from EMG artifacts, causing problems for waveform detection ability. In order to remove or at least compress the noise suffered segment, a morphological filter is deployed [8].

2.1.4 Power line noise

Interference in power lines is a combination of fundamental frequency of 50-60 Hz and its harmonics that interfered the ECG recording process [7]. This type of noise is considered as the extraneous noise in bio-medical signal and is caused by improper grounding of ECG devices and interference from other neighbouring

electrical equipment. It graphically presents in the ECG result as a spike at 50/60 Hz (and its harmonics) with an amplitude up to half of that of the beat-to-beat ECG amplitude.

Figure 5.d gives information about power line noise affected an ECG recording. Since the noisy component has frequency band from 50 to 60 Hz, a notch (band-stop) filter can be used for the reduction of power line noise [8].

2.1.5 Wavelet Discrete Transform (WDT)

In addition to ordinary Fourier Transform (TF) based filtering techniques discussed above, Wavelet-based denoising method also gained interest of researchers recently[9, 10]. Wavelet-based filtering technique is also known as discrete wavelet transform (DWT) i.e., wavelet decomposition. It uses multiresolution analysis to analyse the entire signal with balanced time and frequency resolutions. Meaning that it can simultaneously locate signal features in time and frequency domain which is enormously helpful in inspecting instantaneous and time-varying signals, particular ECGs since it is non-station in nature and vary from people to people. A process of DWT can be performed as show in schematic diagram Figure 6

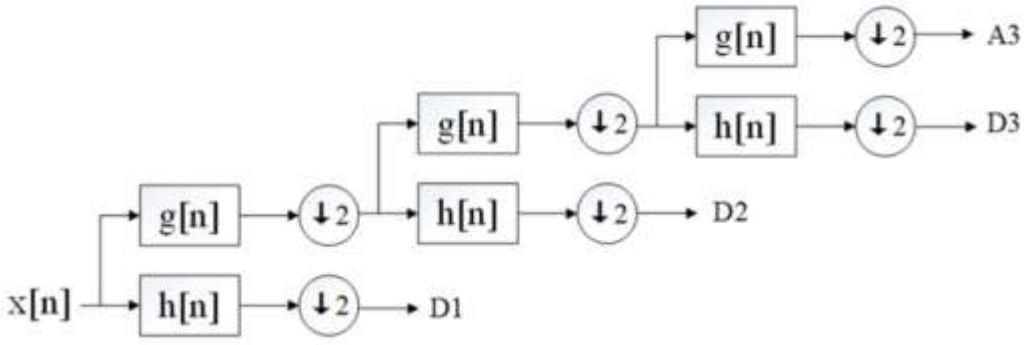

Figure 6. Wavelet Decomposition

The transformation is performed by passing the signal x[n] to a sequence of low-pass and high-pass filter [9, 10]. Then it is decomposed into detailed component D[n] which is high frequency/coefficient and approximate component A[n] which is low frequency/coefficient of different scales. Where n is the length of the signal, ↓ 2 is a down-sampling filter, g[n] and h[n] are low-pass and high-pass filters respectively. A[n] is the result after passing through low-pass filter g[n], and D[n] is the result after passing through high-pass filter h[n]. the down sampling process ensures the resolution of approximation and details are kept the same as the resolution of the

original signals. And these filters are built from mother wavelet functions. The more similar the mother wavelet and ECG signal are, the more accurate coefficient can be calculated. An example of mother wavelet is illustrated in Figure 7 below.

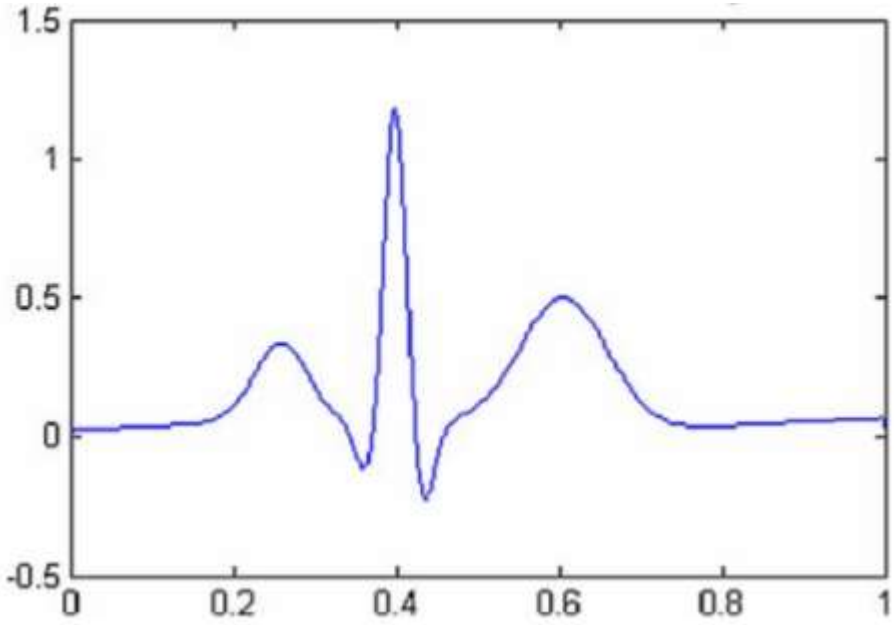

Figure 7. Daubechies (order 8) mother wavelet

After analysing signals into corresponding approximate and detailed components, we can threshold some curtain frequency band and only maintain the frequency bands in interest. Then we can produce the clean signal by combine the detailed and approximate components in typical ECG frequency band. We can denoise motion artifact, powerline noise, and baseline drift noise mention above by removing corresponding approximate and/or detailed frequency components in which these noise types occur. Usually, low-frequencies (approximates) are baseline wander, and some high-frequencies (details) are powerline noise.

An advantage of DWT over traditional Fourier based transform is that the frequency information can be in located in time-domain. In this way the noise that overlap the frequency in interest (e.g., 50-60Hz noise and R-wave have quite similar frequency band) can be represented as small coefficient and the important information is represented as larger coefficient. Therefore, we can threshold out the small coefficient and keep the greater one. As the result, the noise is moved without loss of crucial data.

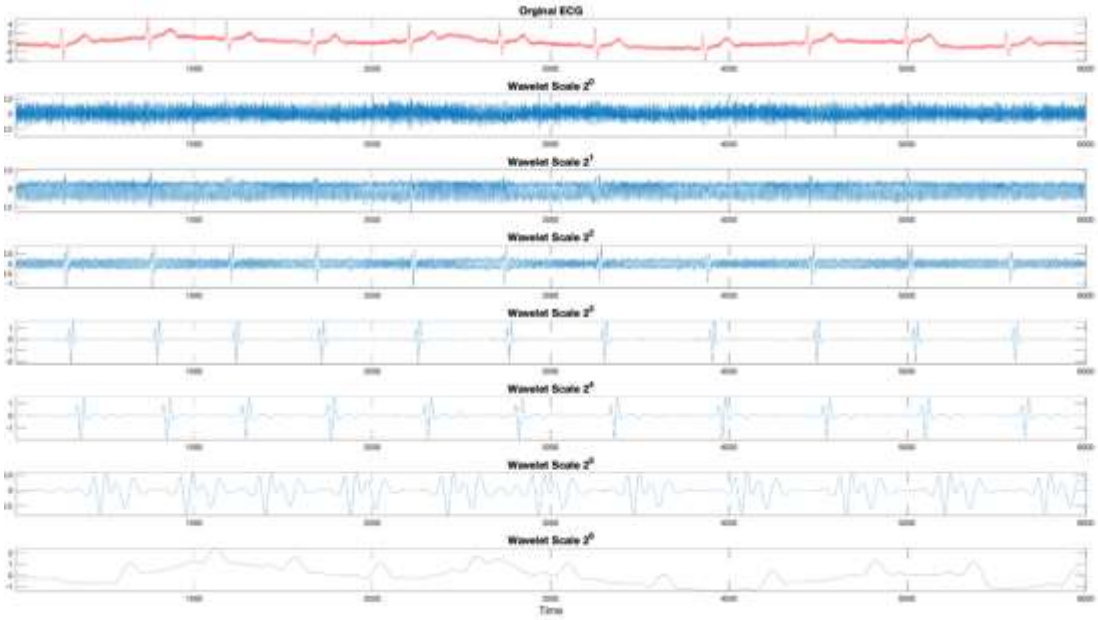

Figure 8. Example of wavelet decomposition for filtering purpose

As shown in Figure 8 [11], raw ECG recording (red) is decomposed into 7 different scales by Symlet 4 wavelet. Temporal and spectral information of the noisy signal is simultaneously analysed in approximate (a_n) and detailed coefficients (d_n). The last row is baseline wander, and we can clearly identify when it happens and its amplitude. In addition, the 50Hz powerline noise is also localized in time and frequency domain. This noise interferes throughout the signal, so it is showed during the second row. Hence, from this information, we can reconstruct the clean recording by combine other coefficients except the first and last rows which are high-frequency and low-frequency noise, respectively.

Research of B. N. Singh and A. K. Tiwari [10] showed that Daubechies mother wavelet of order 8 in Figure 7 is the most suitable wavelet basis for ECG denosing.

2.1.6 Moment of Velocity (MoV)

This denoising/transforming technique is a tool for signal analysis, and was first introduced by M. Dorraki et al [12]. It was inspired by the idea of instantaneous frequency (IF) of a signal and angular momentum of particle dynamics. Hilbert transform of a signal is the imaginary part of that signal (real part itself). Hence, the

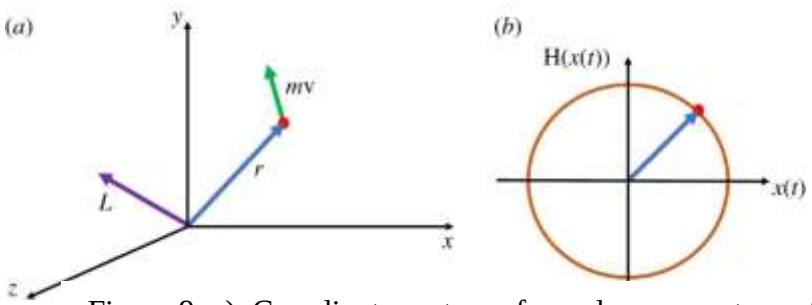

Figure 9. a). Coordinate system of angular momentum of a moving particle and b). Hilbert transformation of a signal. [12]

original signal and its Hilbert transformation are always orthogonal. Similarly, the MoV and its position vectors works in the same way. It is hard to imagine but it can be clearly explained by using mathematical equation below.
In Figure 9, the position vector is perpendicular to the angular momentum with the relationship:

$$L = r \times v \tag{2.1}$$

Where r and v are position vector and velocity vector of a particle respectively.
Applying cross-product formula:

$$\mathbf{L} = \begin{vmatrix} i & j & k \\ r_x & r_y & r_z \\ v_x & v_y & v_z \end{vmatrix} = (r_y v_z - r_z v_y)i + (r_z v_x - r_x v_z)j + (r_x v_y - r_y v_x)k \tag{2.2}$$

Where i, j, k are the standard unit vectors, and x, y, z are indices.
Assuming that the particle is moving in the x-y plan, so the angular momentum of the particle is perpendicular to this momentum vector (in the x-y plan). The equation above is reduced:

$$\mathbf{L}_z = r_x \frac{dr_y}{dt} - r_y \frac{dr_x}{dt} \tag{2.3}$$

Replacing r_x, r_y, and $\mathbf{L}_z$ in equation 2.3 by $x(t)$, $H(t)$ and MoV corresponding to input signal, Hilbert transform and MoV, respectively. We got the moment of velocity transformation.

$$moment\ of\ velocity = x(t)\frac{dH[x(t)]}{dt} - H[x(t)]\frac{dx(t)}{dt} \tag{2.4}$$

M. Dorraki et al [12] has shown that their approach for denoising/transforming technique was more robust with noise, and was better for R-peak detection compared to IF along. As shown in the Figure 10 below, the MoV of an ECG recording is quite similar with IF but more robust with noise, making it easier to detect R wave in ECGs.

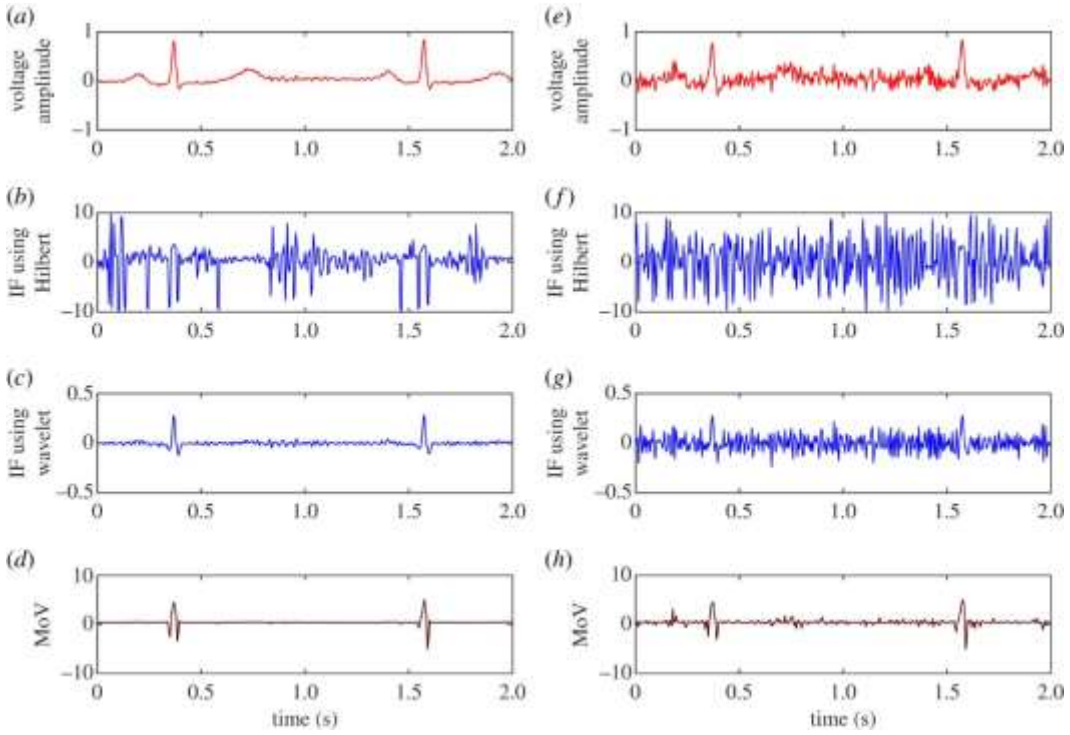

Figure 10.a). raw ECG signal, b-c). instantaneous frequency using Hilbert and wavelet transform respectively, d). movement of velocity transformation, e-h). White noise was added to effectiveness demonstration of Mov

2.2 FEATURE EXTRACTION AND SELECTION

After pre-processing stage, feature extraction is performed to produce useful information from the signal, downsize inputs for prediction models for better classification ability in terms of computational time and precision. However, not all features are worth extracting since there are some classes that have the same value for a particular feature and redundant features increase complexity of a classifier. Therefore, feature selection is needed. It finds the smallest number of features to get the required prediction rates [13].

A survey conducted by S. H. Jambukia and et al [14] showed that features from time domain, frequency domain or time-frequency domain are extracted. The information from one cardiac cycle or relation between successive cardiac cycles which helps distinguish between normalities and abnormalities is derived from analysing these domains.

2.2.1 Waveform detection

Morphological information of the ECG is used to extract precious features from all domains. Hence, waveform characteristics of the signal is the preliminary step to perform [14]. By identifying P and T waves as well as the QRS complex, the entire recording can be segmented into each cardiac cycle. Time series features and frequency components of every heartbeat and its relationship are then extracted. An experiment of Q. Zhao and et al [15] demonstrated Discrete Wavelet Transform (DWT) is outperformed other techniques for P-QRS-T detection task. With DWT the signal is decomposed into foundation building blocks with specific time localised, hence the waveforms, particularly the QRS complex is accurately identified.

2.2.2 Time domain features

In this domain, intervals of each P, Q, R, S and T peaks as well as their combinations are derived for analysing irregularity between heart beats [14]:

- Peaks: P, Q, R, S, T.
- Intervals (duration between peaks): RR, QR, QT, ST, QRS duration.
- Segments (duration between waveforms): ST, PR.

From these features, heart rate variability (HRV) of the entire signal is calculated [16]:

No.	Parameter	Unit	Description
1	SDNN	ms	Standard deviation of NN intervals
2	SDRR	ms	Standard deviation of RR intervals
3	SDANN	ms	Standard deviation of the average NN intervals for each 5 min segment of a 24 h HRV recording
4	SDNN index (SDNNI)	ms	Mean of the standard deviations

			of all the NN intervals for each 5 min segment of a 24 h HRV recording
5	pNN50	%	Percentage of successive RR intervals that differ by more than 50 ms
6	HR Max − HR Min	bpm	Average difference between the highest and lowest heart rates during each respiratory cycle
7	RMSSD	ms	Root mean square of successive RR interval differences
8	HRV triangular index		Integral of the density of the RR interval histogram divided by its height
9	TINN	ms	Baseline width of the RR interval histogram

Table 3: Parameters of HRV in time domain [16].

Note: these values estimated from inter-beat intervals (beat-to-beat) time elapsed between consecutive beats; NN intervals, inter-beat intervals from which artifacts have been removed; RR intervals, inter-beat intervals between all successive heartbeats.

2.2.3 Frequency domain features

Four frequency bands are identified: ultra-low-frequency (ULF), very-low frequency (VLF), low-frequency (VL), and high-frequency (HF) band. Then, measurements related to the distribution of these frequency bands are estimated [16]:

No.	Parameter	Unit	Description
1	ULF power	ms^2	Absolute power of the ultra-low-frequency band (≤0.003 Hz)
2	VLF power	ms^2	Absolute power of the very-low-frequency band (0.0033–0.04 Hz)
3	LF peak	Hz	Peak frequency of the low-frequency band (0.04–0.15 Hz)
4	LF power	ms^2	Absolute power of the low-frequency band (0.04–0.15 Hz)
5	LF norm	n.u	Relative power of the low-frequency band (0.04–0.15 Hz) in normalized units
6	LF norm	%	Relative power of the low-frequency band (0.04–0.15 Hz)
7	HF peak	Hz	Peak frequency of the high-frequency band (0.15–0.4 Hz)
8	HF power	ms^2	Absolute power of

			the high-frequency band (0.15–0.4 Hz)
9	HF norm	n.u	Relative power of the high-frequency band (0.15–0.4 Hz) in normalized units
10	LF/HF	%	Ratio of LF-to-HF power

Table 4: Parameters of HRV in frequency domain [16].

Power is calculated as a signal energy of a specific frequency-band, it can be represented as:

- Absolute power: the integral of all power values inside a frequency band.
- Relative power: this value of each band is calculated by divide the founded absolute power of the frequency range by the summation of all four frequency bands' absolute powers.

A result produced by C. Venkatesan and et al [17] showed that temporal features number 1, 2, 3, 5, 6, 7, 8, 9 and frequency features number 2, 4, 5, 8, 9, 10 were chosen in feature selection step for SVM technique and yielded the highest accuracy, reaching 96%, compare with other methods.

2.2.4 Time-frequency domain features

A combination of temporal and frequency features can be represented as spectrogram or scalogram by using short-time Fourier transform (STFT) and Continuous Wavelet Transform (CWT) respectively. As a result, 1-dimensional signal array is converted into 2-dimensional image array. The transferred input is then fed into prediction models that specialize in image processing to classify problematic signals from the healthy one. Convolutional Neural Network (CNN) and Convolutional Recurrent Neural Network (CRNN) proposed by M. Zihlmann and et al [18] proved its ability for ECG classification based on time-frequency domain.

Both networks got a high performance on the AF detection task, reaching overall accuracy of 90% and won the second prize of CinC Challenge 2017.

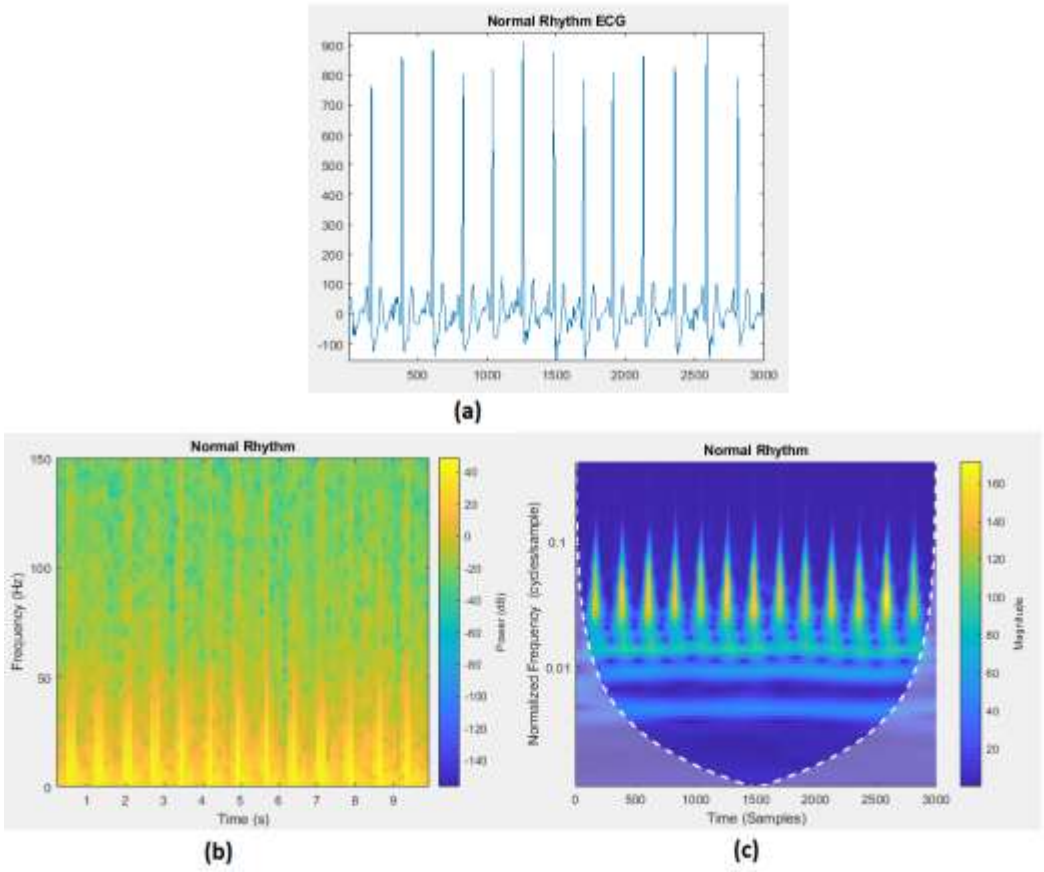

Figure 6: plots of a Normal Rhythm ECG signal from the PhysioNet 2017 Challenge: a) 1-D array ECG data, b) Converted Spectrogram 2-D image, c) Converted Scalogram 2-D image.

Figure 6 illustrates conversion from raw ECG recording to its corresponding spectrogram and scalogram images for image classification application. Although both approaches represent a signal in time-frequency domain, scalogram has better time and frequency resolution compared with spectrogram. Spectrogram has a dilemma between time interference and frequency resolution when optimizing a window size for STFT, however, it has a lower computational cost than that of CWT scalogram.

2.3 CLASSIFICATION

Analysing and distinguishing between arrhythmia and normal rhythm heart beats from biomedical signals may be challenging for doctors and clinicians when dealing with long or noisy recordings. Therefore, modern approaches based on

computational power of machines have been developed for more optimal ECG classification ability. Pattern recognition is the main principle behind these methods, machine learning techniques are deployed for this purpose.

Various machine learning algorithms have been proposed recently for biomedical signal prediction, Support Vector Machine is one of the most popular and effective solutions. It is usually used as a benchmark for other later techniques: Neural Network (NN) and its subclass: Convolutional Neural Network (CNN), Convolutional Recurrent Neural Network (CRNN). In this chapter, a traditional machine learning method - SVM and a deep learning method - CNN are discussed.

2.3.1 Support Vector Machine

SVM is supervised learning algorithm, meaning that it learns how to identify data by given labels. It is widely used for both pattern recognition and regression problems. The mechanism of SVM is finding a hyperplane that distinctively separates between classes of data in multi-dimensional space. The order of the dimension is the number of extracted features. For instance, SVM will draw a boundary line in 2-D space if 2 features are used. Similarly, if 3 or more features are fed to the SVM, the hyperplane is found in 3-D space and so on. It is abstractive to imagine how SVM work if more and more features are imputed, however, mathematical equations can clarify the ambiguity in interpretation.

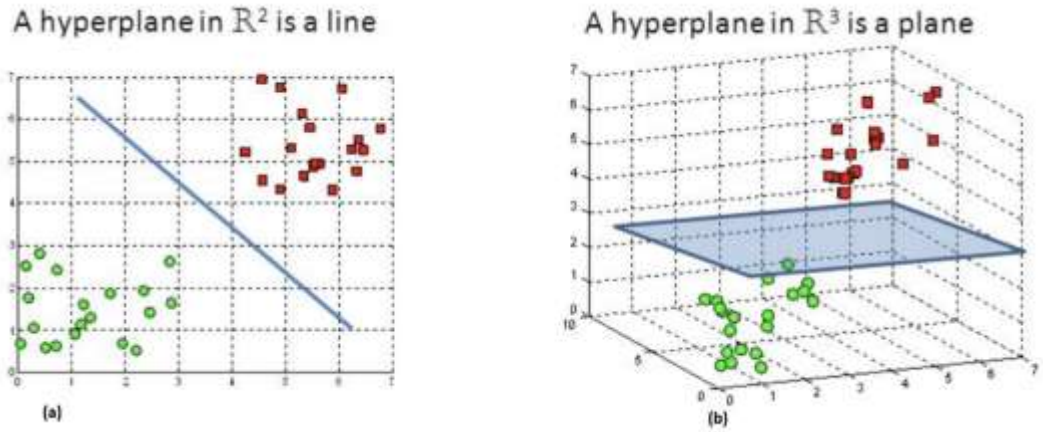

Figure 11. SVM classifying 2 classes in 2-D space (a) and 3-D space (b) [19]

Figure 11 describes how a hyperplane is drawn in multi-dimensions. Support vectors are the closest data points from each cluster to the hyperplane, hence they affect the position and orientation of the hyperplane. The goal is to maximize the

margin of the support vectors in order to have a clear separation between group of data points. Where the margin is the distance from the separating boundary to the support vectors.

2.3.2 Neural Network

Neural Networks like CNN and CRNN are efficient tools for pattern recognition problems, and hence they are deployed for biomedical signal processing [18]. NNs' architecture is inspired by human brain, and it simulates how human learn. Figure 12 below gives general information about a structure of a typical NN model.

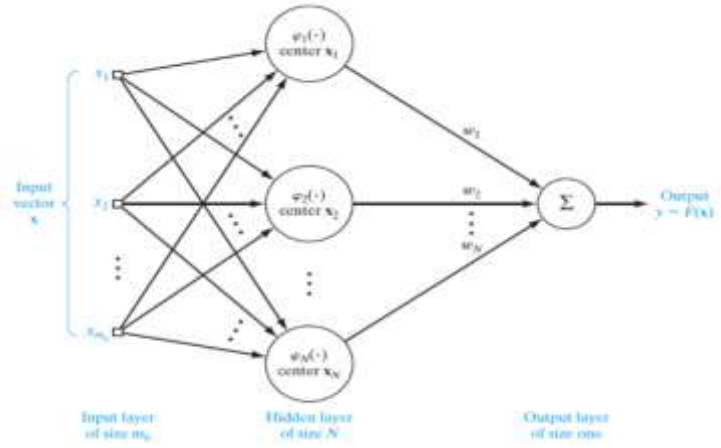

Figure 12. General structure of an Artificial Neural Network

Normally, the ANN has three types of layers ranging from input to hidden and output layers. Each layer consists of various nodes that represent for human neurons. These artificial neurons are inter-connected with neurons of other layers by channels, and there is no connection between nodes in the same layer. Characteristics of each layer types as follow:

- Input layer: where the raw input information is introduced to and then transferred to the hidden layers.
- Hidden layer(s): performs most of the computation required by the network, then it outputs results to the output layer.
- Output layer: makes the final prediction based on the processed data.

Each node is core processing unit and is linked with other nodes by channel, each channel is assigned with an adjustable weigh. The learning process of the network is repeated loops of forward propagation and back propagation:

- Forward propagation: the data from each node of the input layer is multiplied with its corresponding channel's weigh. Their sum is inputted to the node of the hidden layer. Nodes in this layer has their own bias value, together with the summation before, the total value is passed to the activation function

which is used to determined that whether the node is activated. Only the activated node is allowed to transmit data to nodes of the next hidden layer. This process continues until the data propagate to the output layer. In the output layer, the node with highest value is activated and produce the prediction.

- Back propagation: in this process, the released output is compared with the actual ground true label and prediction errors are estimated. Then the network reduces the errors by adjusting the weigh in each channel to have more accurate output in most of the cases.

2.3.2.1 Convolutional Neural Network

CNN is one type of NN that have an additional hidden layer called convolution layer. This layer uses filters (kernels) to extract features from spectrogram or scalogram. Image features can be morphological shape and edge of waveforms represented as power in 2-D time-frequency domain. Unlike other neural networks, CNN uses convolution operation instead of multiplication operation in the convolution layer to process and propagate data to the next layer.

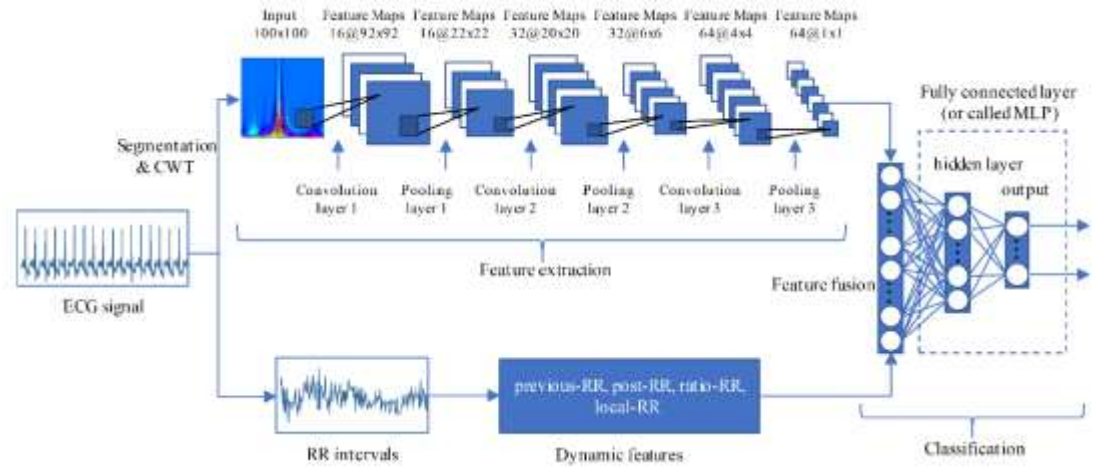

Figure 13. An example of CNN architecture using temporal features and scalogram images to classify ECG signal [20].

Figure 13 shows a process of classifying ECG signals by 2 sets of features proposed by Wand et al [20]. The proposed technique was CNN using convolutional and pooling layers to extract features from scalograms. Then, a fully connected layer was used to classify ECG into different classes.

2.4 VALIDATION

Normally, various prediction models are created during the training process and the best model with highest accuracy will be chosen. Therefore, validation which is used to compare different models is the final step needed to be done. In this step,

the entire data is split into training and testing set which is used for training and testing purposes. The model is trained on the training portion and give prediction for the rest testing set. Data separation is necessary because if the model is trained and tested on the same data, it will output prediction that it already saw before during the training process, causing “overfitting”. As a result, the model may perform well only on a particular dataset and may perform poorly if it is fed with a new one.

To avoid that bias, cross-validation is introduced to validate a model generally. A well-known method named k-fold cross-validation is used widely for its efficiency. This technique splits the entire data into k equal subsets and sets 1 subset as testing and the rest as training. This process occurs repeatedly until all subset is already set as testing set. The overall performance is evaluated by the average accuracy of all iterations. Another method is leave-one-out cross validation which is an extreme case of k-fold since it sets 1 observation as testing and the rest as training set. It is clearly a more robust technique compared with k-fold since every single example of the dataset is used for testing. However, it is not recommended for large dataset due to computational cost [24].

Chapter 3: Research Methodology

3.1 OVERVIEW

Firstly, the research on related works about ECG classification techniques must be conducted to have a brief overview of the project. This process is called literature review. It gives information about the knowledge needed for the project, emphasize previous achievement in the similar projects as well as current challenges in the field.

When the general idea about the ECG recognition is formed, steps to analyse and classify this biomedical signal is performed as follow:

1. Data Collection
2. Signal Pre-processing
3. Feature Extraction
4. Classification and Validation

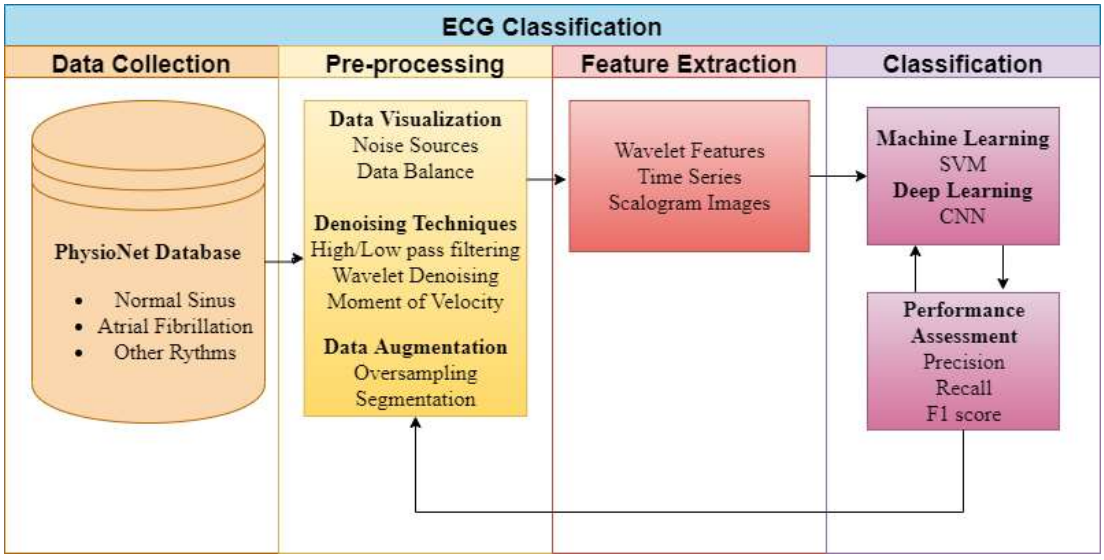

Figure 14. Proposed ECG classification method

As the Figure 14 illustrated, our research on cardiovascular diseases prediction is primary divided into four main steps, starting from Data Collection, Pre-processing, Feature Extraction, and ending with Classification.

Throughout these tasks, data are gathered and explored to check its integrity, balance, and manually identify its patterns if possible. Then some signal processing

techniques will be performed to clean up the data in case some recordings are suffered from noise sources. In addition, if one type of data outnumbers the others or there is a lack of a particular label compared to the others, data augmentation methods will be applied to make dataset balanced, hence classification algorithms can have enough data to improve its learning ability. Then, since raw input recordings of ECGs cannot really help to distinguish between different heart rhythms, a special process named feature extraction is applied to obtain meaningful information from the signals. Some useful features for pattern recognition were described in section 2.2. The next step is deploying machine learning algorithms to learn from and classify data. Each algorithm needs different type of input and will perform differently depending on its structure and the fed data. After a training process, a model is built based on its appropriate algorithm and used dataset, it is now able to give prediction on new recordings. The models are then evaluated by several criteria e.g., precision, recall, f1 score on a new input. There is a recurrent process from pre-processing and classification since the purpose is trying to increase the performance of classifier by experimenting wide variety of data denoising/transforming techniques as well as ML algorithm to find the best combination for AF detection. Details for each stage will clearly illustrated below.

3.2 DATA COLLECTION AND EXPERIMENT TOOLS

Data used in this project is collected from the PhysioNet 2017 Challenge [25]. There are 8528 recordings sampled at 300 Hz and have length between 10 to 60 seconds. However, most signals are 30-second long. The entire data are divided into 4 groups by experts:

- Normal Rhythm (N): 5050 recordings
- Atrial Fibrillation (A): 738 recordings
- Other Rhythm (O): 2456 recordings
- Noisy Recording (~): 284 recordings

We decided to use normal rhythm, atrial fibrillation, and other rhythm for ML algorithm since the noisy type was too noisy to analyse and classify.

MATLAB are chosen for machine learning algorithm implementation. In addition, ADAPT (Any Device Any Place and any Time) of the University of

Adelaide is used in case of large data or algorithm that required huge computational cost. It provides student remote access and utilize university's application.

3.3 PRE-PROCESSING

In this step, raw ECG data must first to be explored and understood. Visualization and statistical methods are used to examine the dataset. Possible issues e.g., too-short recordings or noisy segment, and patterns inside the dataset will be identified. In addition, data imbalance is a serious problem in ML classification tasks Then, according to J. Wang and et al [21], it is necessary to make sure all recordings have an acceptable length e.g., greater, or equal 10 seconds for each observation, since abnormality needs enough heart beats to be identified. As the dataset has signals vary from 10 to 60 seconds, we will perform signal segmentation to divide the longer recordings to smaller ones, in this way, the number of samples will be increased. Then visualizing some random signals of each label to assess the signal quality. Identify possible noises that interfered the ECGs.

For denoising techniques, since the previous projects about detecting problematic ECG recordings mainly used ordinary low-pass, high-pass, and bandpass filters and received high results, we proposed to use two other different methods which are wavelet denoising and moment of velocity as discussed in chapter 2.1.5 and 2.1.6 respectively. Both approaches were proven to effectively remove/reduce noise from signals. Therefore, we experimented whether the denoising methods have any impact on the classification ability of machine learning algorithms by using raw data, wavelet denoised data, and MoV transformed data. We applied all three types of datasets with each ML techniques and compare the results.

3.3.1 Wavelet Denoising implementation

From what we learnt in the literature review, if the mother wavelet has the similar morphological shape with the ECG, it will produce more reliable denoising process by providing more accurate approximation and detailed coefficient. Based on the result of B. N. Singh and A. K. Tiwari [10] we decided to use Daubechies mother wavelet of order 8 to decompose the signal into 8 different levels. Inspect and reconstruct the recordings without noise coefficient.

3.3.2 Moment of velocity implementation

For the MoV transformation, we were provided with a MoV tool developed by M. Dorraki [22]. With his agreement, we could use the tool to analyse ECGs and do not need to implemet it ourselves.

3.4 MACHINE LEARNING AND FEATURE EXTRACTION

Similar to denoising process, we also performed a range of machine learning algorithm with different type of data to find the best combination that led to the highest performance of detecting AF. Each ML requires corresponding input. For example, SVM models need one-dimensional array of useful features while CNN models are well suited for two-dimensional array data e.g., images. In addition, LSTM networks are designed to deal with sequence data. With those characteristics, we managed to apply ML techniques with appropriate input as follow:

3.4.1 Support Vector Machine

Since atrial fibrillation is all about chaotical and irregular beats of the cardio vascular system, we performed test to check the importance of variability in heart rhythm e.g., heart rete variability in the AF parttern recognition. Therefore, there were a comparision between two SVMs using different features set.

SVM with HRV features

By analysing the beat-to-beat intevals of consecutive heart beat, we performed feaeture-extraction to get precious information about the heart rate and how much it change over the selected period of time. From a study of F. Shaffer and J. P. Ginsberg [16], we implemented and calculated some features in time domain that we assumed useful for AF rhythm detection. Table 3 gave information about the extracted HVR features.

Feature	Meaning	Unit
Heart rate	Number of heart beats per minute	bpm
Mean interval	Mean value of beat-to-beat intervals	ms
SDNN	Standard deviation of beat-to-beat intervals	ms
SDSD	Standard deviation of difference of beat-to-beat intervals	ms

RMSSD	Root mean square of beat-to-beat intervals	ms
NN50	The number of intervals that greater than 50ms	ms
pNN50	The percentage of intervals that greater than 50ms	%
NN20	The number of intervals that greater than 20ms	ms
pNN20	The percentage of intervals that greater than 20ms	%
ShE	Shannon entropy of heart beats	du
	Total	10

Table 3. Features in HVR

SVM with multi-type features

To test the performance of only HRV information, we conduct another experiment using SVM but with much more complicated and detailed features. The features were described in Table 4 below.

Type	Features	Number
Time domain	SDNN, RMSSD, NNx	8
Frequency domain	LF power, HF power, LF/HF	8
Non-linear features	SampEn, ApEn, Poincaré plot, Recurrence Quantification Analysis	95
Signal Quality	bSQI, Isqi, Ksqi, rSQI	36
Morphological features	P-wave power, T-wave power, QT interval	22
	Total	169

Table 4. Multi-type Features in various domain

These features come from multi-dimension e.g., time and frequency domains, non-linear statistics, qualities of signal, and morphological shapes of waveforms in ECGs. Thanks to the MATLAB tool developed by F. Andreotti et al [23] which is available on Github [24], we modified the tool so that it could run on our data. His approach was proven to effectively classify the AF and get the average F1 score of 0.79 for distinguishing 4 classes including normal, atrial fibrillation, other rhythm, and noisy data also. He used the set of features with decision tree (DT) algorithm while our intention was deploying SVM for this purpose. The result of SVM using 169 features was used as a benchmark for testing the significance of HRV features in detecting AF as well as testing other deep learning approaches.

3.4.2 Long-Short Term Memory

Building on from RNNs, LSTMs are an advanced variation that can cope with long-term temporal dependencies which is the major problem of conventional RNNs.

Because human physiological signals are time-series and non-stationary in nature, LSTM maybe a best suit for solving pattern recognition problems in these biomedical signals, particularly ECGs. Hence, we examined the classification ability of LSTM networks with frequency features as suggested in an example using MATLAB [25]. The sequence of features includes instantaneous frequency and spectral entropy which were proven to successfully discern atrial fibrillation from normal rhythm.

3.4.3 Convolution Neural Network

In the last experiment, we attempted to use transfer learning method to classify our data. Transfer learning is an approach to reuse a well-trained model to train on other dataset to save training time and solve the problem of insufficient data. Gajendran et al [26] showed the effectiveness of utilizing available CNNs architectures to identify abnormality in heart rhythm. Their replaced some layers and fine-tune so that their modified CNNs could classify different heart patterns. Therefore, with that inspiration and limitation of data sample, we conduct a final experiment using transfer learning method. A process of transfer learning is described in Figure 15 below.

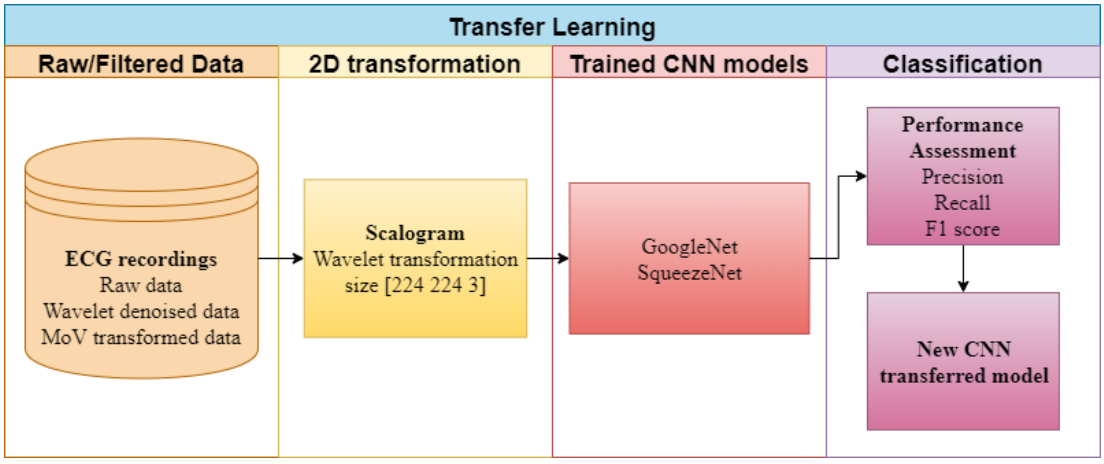

Figure 15. Transfer learning schematic diagram.

Our chosen pre-trained model was SqueezeNet which is the smallest CNN with demonstrated performance. Despite of its small size, it is also designed to be able to classify images in 1000 categories with the same level of accuracy of other deep CNNs [27]. According to Iandola et al [27], SqueezeNet was able to reach AlexNet accuracy level with 50x fewer parameters and 510x smaller model size. SqueezeNet

was chosen for this task since it is the smallest pre-trained deep CNN, hence it would be feasible to be deployed on memory-limited hardware.

3.5 PERFORMANCE ASSESSMENT

All the classifiers will use 5-fold validation to avoid overfitting. In addition, the number of each class in the dataset will be made balanced by oversampling and segmentation techniques. Regarding testing, there will be also a balanced dataset in the test set which contains equal number of recordings for individual class, and all models will be tested on the same test set or its transformation.

There some several criteria to assess the classifier performance including confusion matrix, precision, recall, f1 score, and receiver operating characteristics (ROC). Confusion matrix is a table used to evaluate how well the classifier can make prediction on the test set. It gives information about the ground true labels and the predictions that a model made on the actual data. An example of confusion matrix is given in Table 5 below.

N = 100	Predicted: A	Predicted: B
Actual: A	35	20
Actual: B	15	30

Table 5. Confusion matrix

For example, we are trying to detect class A from class B, the sum of columns of the first row is the actual total samples of class A while that of second row is total samples of class B. Moreover, the first and second column represents how many samples the model predicted as class A and class B respectively. For label A detection, some terms below are defined:

- True positive (TP): actual A and was recognized as A
- True negative (TN): true label is not A and was detected as not A
- False positive (FP): true label is not A and was detected as A
- False negative (FN): actual A and was detected as not A

Precision is the proportion of A-class predictions was actually correct.

$$precision = \frac{TP}{TP + FP}$$

Recall is the proportion of true A-class samples was actually identified.

$$recall = \frac{TP}{TP + FN}$$

F1 score is a measurement that conveys a balance between precision and recall.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

For real-world problems, using F1 score is more reliable than accuracy in terms of unbalanced dataset. For example, a sample set containing 90 samples of class A and 10 samples as class B. A classifier can easily get 90% accuracy if predict all 100 observations as A class. It is a high result; however, it would be a catastrophic if class B represents for heart diseases. In the case above, precision, recall and f1 score for class B are all 0 since it did not identify any B sample. Hence, using f1 score is much more reliable than accuracy.

3.6 PARTICIPANTS

I, Hien Long Nguyen, and Sonia Kleinig are main participants of the project "Can We Teach a Machine to be Cardiologist?". And since this project is suitable for remote students, we can work remotely from home on our own computer to prevent the spread of the COVID-19 pandemic. We work under the supervision of Professor Derek Abbott and co-supervisor Mohsen Dorraki.

Chapter 4: Result and Discussion

4.1 ONE-PAGE REVIEW

Since biomedical signal processing and machine learning are totally new for me and Sonia, we spent the first two weeks to gather information about these two fields. After getting a big picture about our project, the one-page review was written and developed to the literature review in chapter 2. The one-page review summarized background of ECG signals and machine learning application in diagnosing people health condition. The literature review was built on these initial concepts, it detailed the sequence of steps to process and analyse physiological signals, especially ECGs. More important, machine learning and deep learning approaches to diagnose ECGs were carefully described. Current limitation and challenges also documented in the review.

4.2 DATA EXPLORATION

Data used in this project is collected from the PhysioNet 2017 Challenge [25]. There are 8528 recordings sampled at 300 Hz and have length between 10 to 60 seconds. The entire data are divided into 4 groups by experts:

Rhythm	**Number of recordings**
Normal Rhythm (N)	5050
Atrial Fibrillation (A)	738
Other Rhythm (O)	2456
Noisy Recording (~)	284
Total	8528

Table 6. Ground truth labels of entire dataset

We decided to use normal rhythm, atrial fibrillation, and other rhythm for ML algorithm since the noisy type was too noisy to analyse and classify.

Since the length of each signal in the dataset varies from 3000 to 18000 samples, a histogram was plotted to examine the length distribution. As shown in

Figure 16 below, the majority of recordings is 9000-sample long e.g., 30 seconds in time

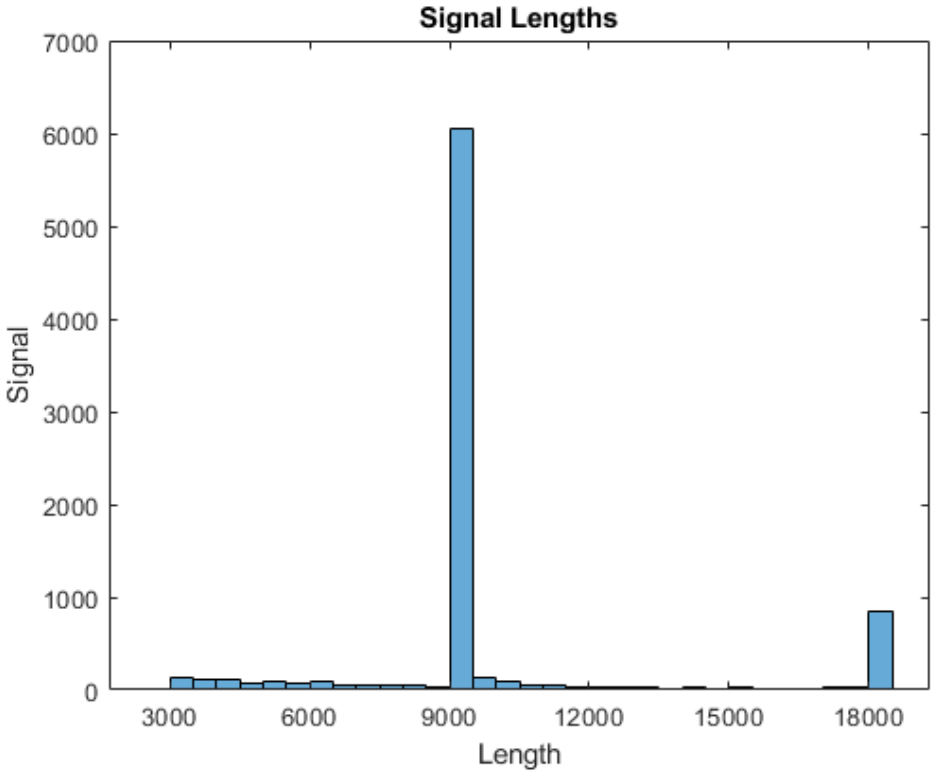

Figure 16. Histogram of signal length

According to J. Wang and et al [21], the smallest recording length to detect abnormality is 10 seconds. To guarantee the integrity of heart rate variability as well as time series signals, we chose the signal length of 15 seconds, corresponding to 4500 samples per observation. Hence, we removed disqualified signals and segmented longer recordings. After reduction and segmentation, the number of recordings per label were shown in Table 7 below.

Rhythm	**Number of recordings**
Normal Rhythm (N)	10327
Atrial Fibrillation (A)	1494
Other Rhythm (O)	5417
Total	17238

Table 7. Ground truth labels after data preparation

4.3 ECG ANALYZE AND NOISE DETECTION

4.3.1 Manually analyzing ECGs

At the first step, we manually analyze waveforms of ECG signals, detecting P, QRS complex, and T wave to visually recognize different pattern between normal and abnormal recordings. Some random ECGs were plotted and analyzed below.

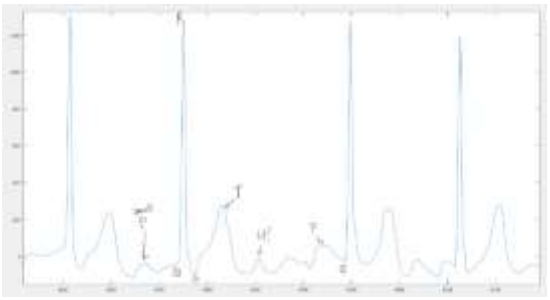

Figure 18. AF rhythm

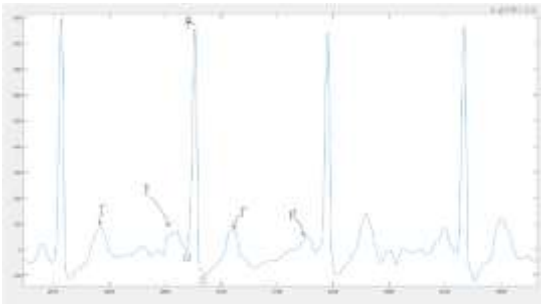

Figure 17. Normal rhythm

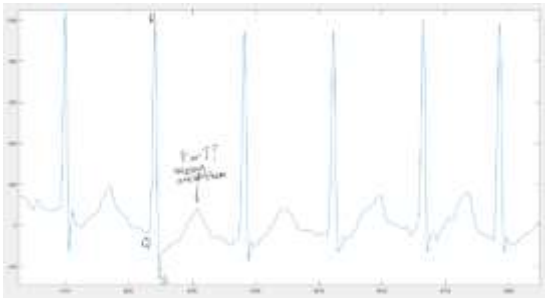

Figure 19. Other rhythm

It is noticeable that the major difference between AF and normal is the time between consecutive R-peaks. The beat-to-beat intervals of healthy recording in Figure 17 are relatively constant, and the features in waveforms are clearly shown in expected locations and amplitudes. On the other hand, the abnormality in AF rhythm is represented as irregularity of R-peak intervals as well as the absence of some waveform, particular the P wave. Electrical depolarization of the atria is represented in P waves, AF is caused when the atria beat out of sync with ventricles, hence P waves are usually small in amplitude or even flat in AF rhythm. This characteristic can be seen in Figure 18. In addition, other rhythm signal may hard to detect since it includes a wide variety of rhythm which is hard to discern with AF and normal signals. However, we noticed that the main pattern of other rhythm is lack of some

waveform e.g., T, Q, S or P wave, or negative QRS complex, or it is not lead II ECG recording.

4.3.2 Noise detection

By inspecting some random signals, we have realized various noise sources that polluted the dataset.

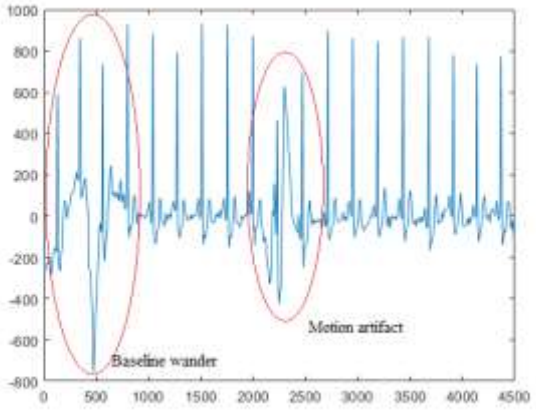

Figure 21. Baseline wander and motion artifact

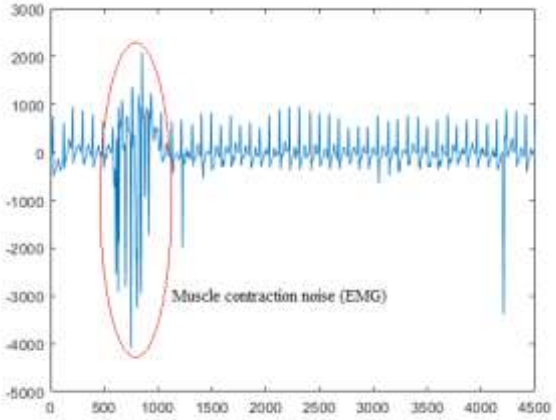

Figure 20. Muscle contraction noise (EMG)

Our dataset is mainly suffered from baseline wander, motion artifact and muscle contraction as shown in Figure 20 and Figure 21. As mentioned in chapter 2, baseline drift is a variation of baseline with frequency below 0.5 Hz and is caused by strong respiration or coughing that alter the heart beats. Motion artifact is quite similar to baseline noise, but it happens in a short period of time with much higher amplitude. Movements that sketch skin areas attached with electrodes lead to this type of artifact because it changes the impedance of recording locations. Regarding to EMG noise, this noise is cause by the electrical activities of our muscle when it contracts. It represents as a high frequency waveform with unpredicted spikes and large magnitude as shown in Figure 20. It maybe seems like motion artifact, but EMG fluctuates more chaotically.

4.4 PRE-PROCESSING

4.4.1 Denoising approaches

As proposed in the methodology section, we used wavelet denoising and moment of velocity transformation to clean the signal. The effectiveness of these to techniques is described in Figure 22 and Figure 23 below.

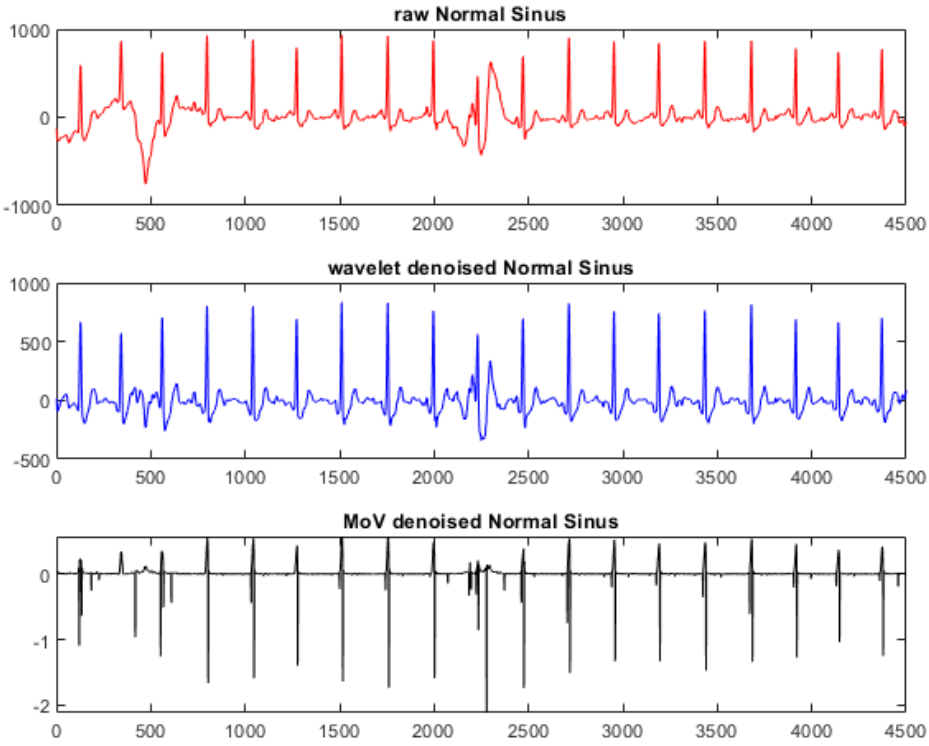

Figure 22. Effect of denoising techniques on baseline wander and motion artifacts

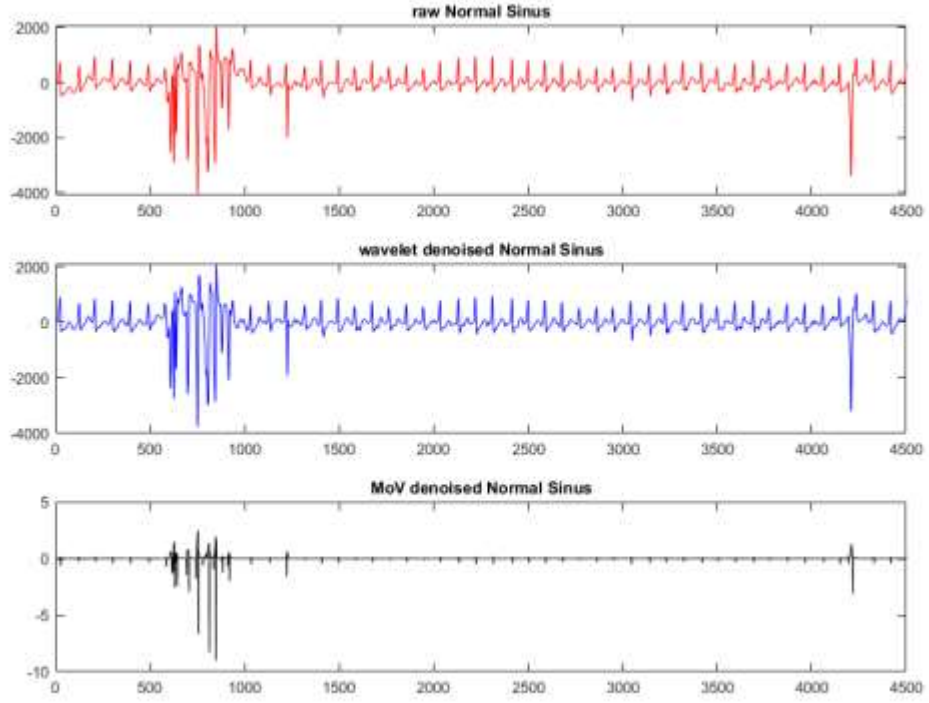

Figure 23. Effect of denoising techniques on contraction noise

For baseline wander and motion artifacts, the two proposed method seemed to effectively suppress the contaminated segments. While the wavelet-based approach retained the information in both time and frequency domain well, the MoV-based method seemed to flatten all the low frequency features, especially P and T wave, and only maintain the R-peak location. However, the amplitude of peaks in MoV denoised signal was small and varied from beat to beat, making it difficult to detect peaks.

In terms of contraction noise, both approaches did not work well on suppressing this type of noise. Since the contaminated segment had large fluctuation and had a quite similar frequency range with ECG waveform, it is not possible for DWT techniques to threshold EMG out when it has greater wavelet coefficient than that of ECG. More severely, MoV transformed recording could not detect any R-peaks except peaks of EMG noise, making it impossible to either detect intervals time series or frequency information of the signal. Therefore, muscle contraction noise is the main challenge when process our ECG dataset.

4.4.2 Peak detection

To extract features in time series e.g., heart rate and heart rate variabilities, we used an algorithm called Pan Tompkins which is available in MathWorks [28]. This algorithm is used for QRS detection, it localizes R peaks so that we can calculate duration between consecutive peaks.

For a small variation of baseline noise and motion artifacts, the algorithm worked well and effectively detect all the peaks. Even with raw data, it can still localize R peaks thanks to its adaptiveness and filtering system inside the function. However, it did not successfully define peaks on MoV transformed data.

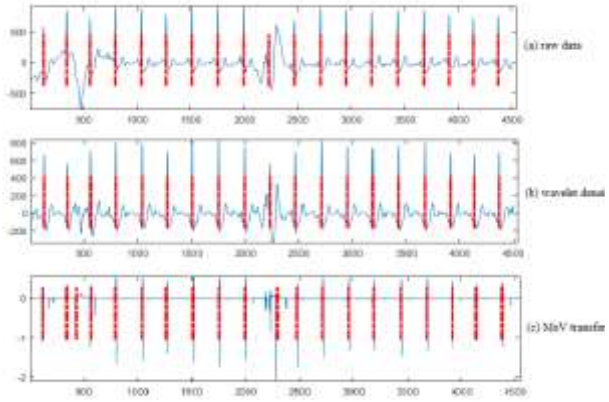

Figure 25. Peak detection on raw and denoised data (baseline wander and motion artifact)

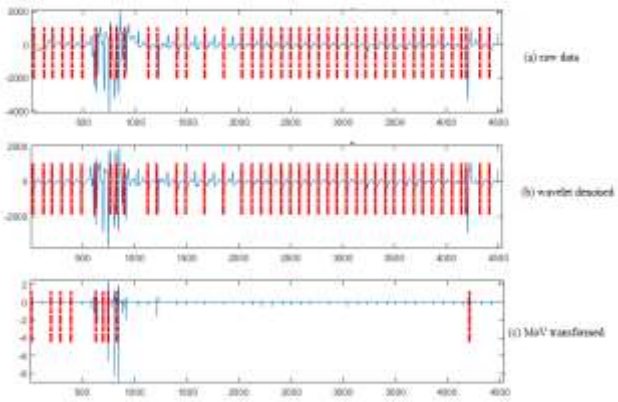

Figure 24. Peak detection on raw and denoised data (muscle contraction)

As shown in Figure 24 and Figure 25 above, while the Pan Tompkins approach can work on raw and wavelet filtered data, it missed some peak on the MoV graph due to

the small and varied amplitude. Particularly, with the impact of EMG noise, the detection function performed worst and missed a lot of peaks as well as wrong recognized noise as QRS complex on MoV filtered data.

4.4.3 Data augmentation

Thanks to the prior segmentation process, the total number of recordings increased double, from 8528 to 17238 signals. However, it was still unbalanced since the majority of observation comes from normal and other signals respectively. Therefore, we increased the sample size of AF rhythm by oversampling techniques. We duplicated the entire AF and other recordings until it reached the number of normal samples e.g., 10327. The quantity of each label is shown in Table 8.

Rhythm	**Number of recordings**
Normal Rhythm (N)	10327
Atrial Fibrillation (A)	10327
Other Rhythm (O)	10327
Total	30981

Table 8. Augmentation dataset

Rhythm	**Train**	**Validation**	**Test**
Normal Rhythm (N)	7435	1859	1033
Atrial Fibrillation (A)	7435	1859	1033
Other Rhythm (O)	7435	1859	1033

Table 9. Training, Validation and Testing set for deep learning

For deep learning with computational cost like LSTM and CNN, we decided to use hold-out validation to speed up the training process. Each label was equally divided into training, testing and validation set to avoid imbalanced data, and overfitting. 90% and 10% of dataset were used for training and testing purposes. In addition, 20% of the training set is used for validation during classifier learning process. The detailed quantity of each portion is shown in Table 9.

For traditional machine learning, 5-fold cross validation was utilized to avoid overfitting. The same training and testing set were utilized.

4.5 FEATURE EXTRACTION AND CLASSIFICATION

4.5.1 Support Vector Machine

As proposed in chapter 3, we conducted 2 experiments using SVM with different denoising methods and corresponding feature sets.

After extracting features, feature scaling was performed. Feature scaling was used to map feature values of data into a particular range so that the figures in greater numeric ranges do not dominate those in smaller number ranges [29]. For SVM, standardization was used to scale features.

$$z = \frac{x - \mu}{\sigma}$$

Where z is the standardized value, x is the original value, μ and σ are the mean standard deviation of feature vector.

Then, SVM models was fitted into data and fine-tune by trial and error. Since there is a huge number of hyperparameter combination, we chose to apply 6 popular kernel functions to SVMs. The other hyperparameters were set as follow.

Hyperparameter	**Value**
Kernel scale	Automatic
Box constraint level	1
Multiclass method	One-vs-one

Table 10. Default hyperparameters

To avoid overfitting, a 5-fold cross validation was used with 9294 observations (e.g., training and validation) for each class.

SVM using HRV features

As mentioned before in the proposed experiment, several features in time domain were extracted to examine the importance of heart rate variability in detecting AF. These features were introduced in Table 3.

Validation accuracy of each kernel function was documented in Table 9 below.

	Accuracy (%)					
	Linear	**Quadratic**	**Cubic**	**Fine Gaussian**	**Medium Gaussian**	**Coarse Gaussian**
Raw data	70.6	70.9	33.5	76.5	73.5	68.7
Wavelet denoised data	70.3	67.1	39.5	77	73.4	68.4

MoV denoised data	59.6	54.6	39.2	70.7	66.4	60.4

Table 11. SVMs using HRV features extracted from raw/filtered data

Via training and validation process, SVMs with Fine Gaussian kernel appeared to be the best models. The performance of these models on the hidden test set are described in the confusion matrix and ROC below

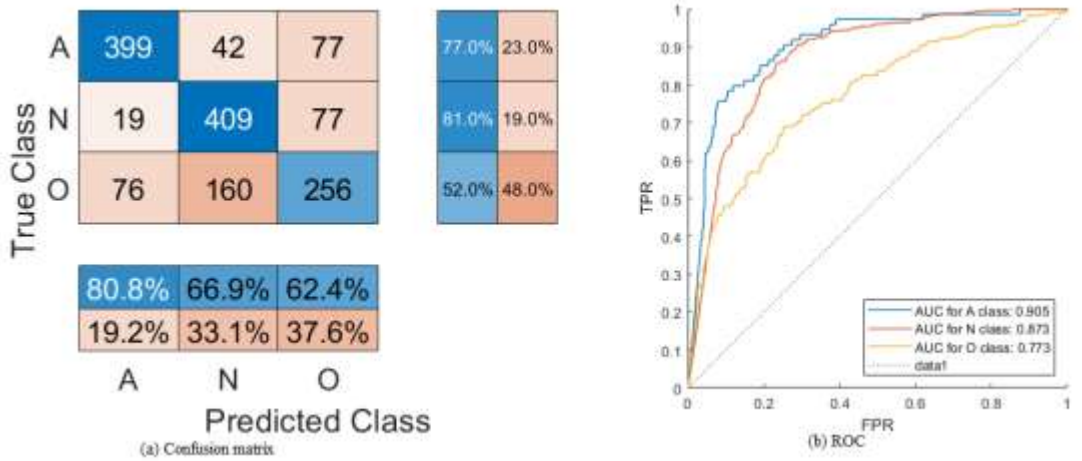

Figure 26. Confusion matrix and ROC of SVM using HVR features extracted from raw data

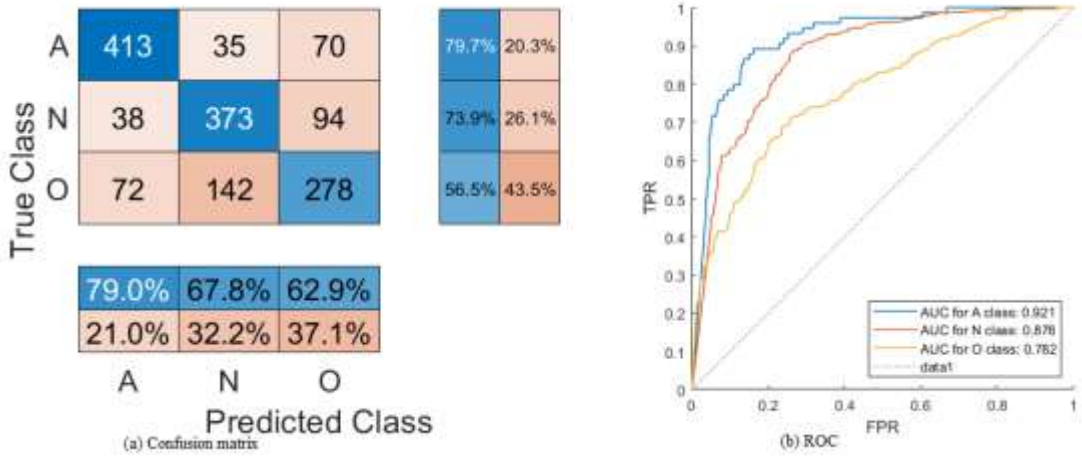

Figure 27. Confusion matrix and ROC of SVM using HVR features extracted from wavelet denoised data

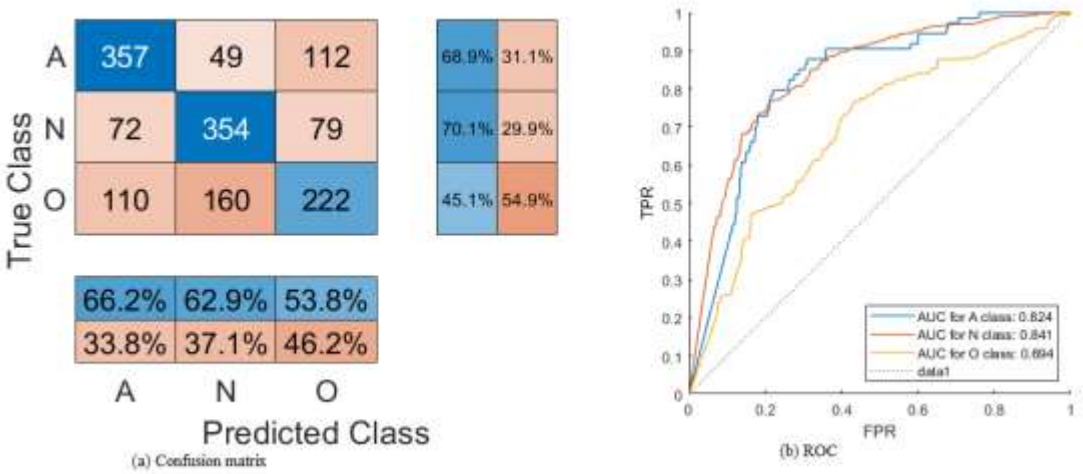

Figure 28. Confusion matrix and ROC of SVM using HVR features extracted from MoV denoised data

	Precision	**Recall**	**F1 score**
Raw data	0.808	0.77	0.788
Wavelet denoised data	0.79	0.797	0.793
MoV denoised data	0.662	0.689	0.675

Table 12. Precision, recall and F1 score for AF detection SVM with HVR features

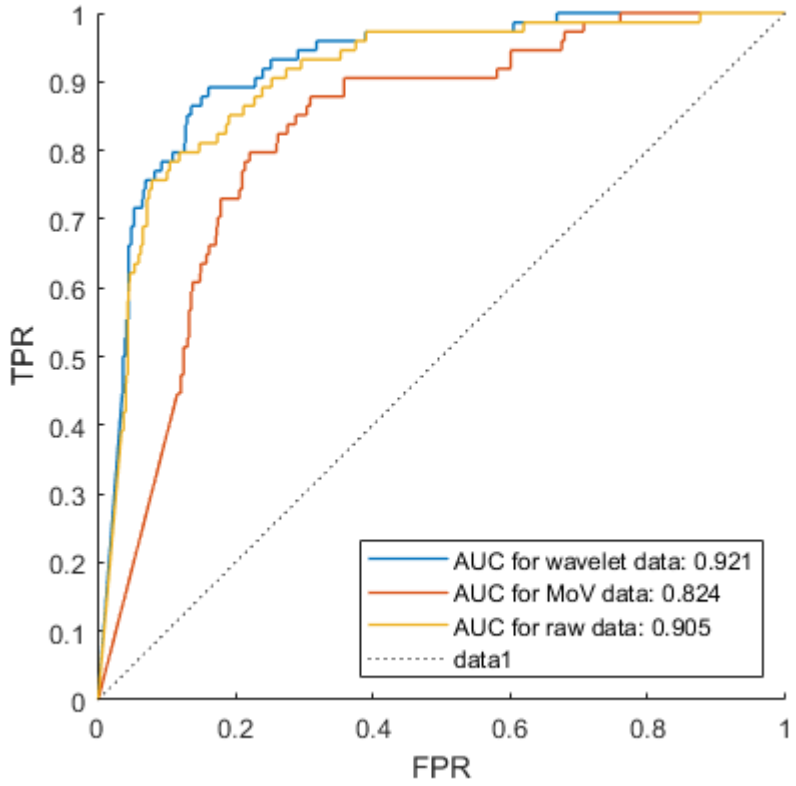

Figure 29. Robustness comparison of SVMs using HRV features extracted from raw/filtered data

SVM using multi-type features

	Accuracy (%)					
	Linear	**Quadratic**	**Cubic**	**Fine Gaussian**	**Medium Gaussian**	**Coarse Gaussian**
Raw data	77.2	81.9	83.3	74.6	80.2	74.1
Wavelet denoised data	78.6	83.5	85	78.9	82.1	75.8
MoV denoised data	70.4	74.4	74.6	67	74.3	67.9

Table 13. SVMs using multi-type features extracted from raw/filtered data

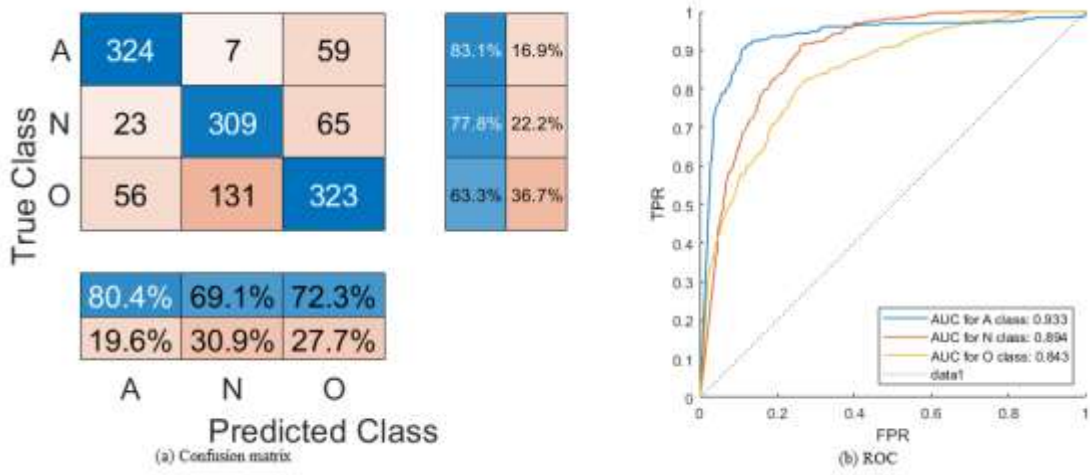

Figure 30. Confusion matrix and ROC of SVM using multi-type features extracted from raw data

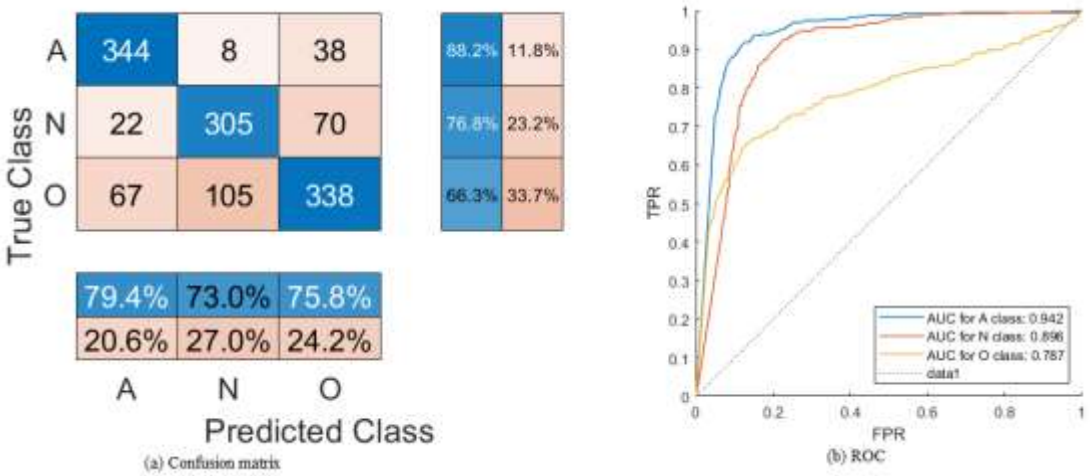

Figure 31. Confusion matrix and ROC of SVM using multi-type features extracted from wavelet denoised data

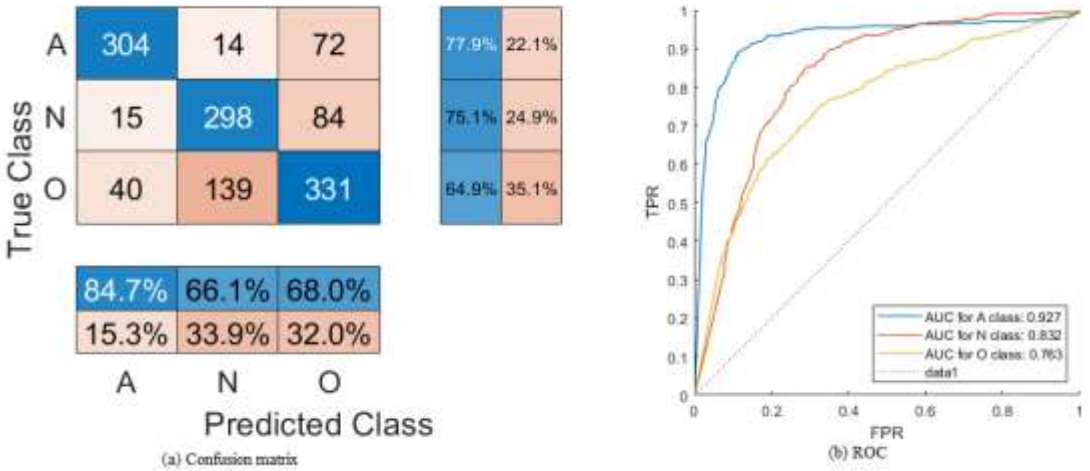

Figure 32. Confusion matrix and ROC of SVM using multi-type features extracted from MoV denoised data

	Precision	**Recall**	**F1 score**
Raw data	0.804	0.831	0.817
Wavelet denoised data	0.794	0.882	0.836
MoV denoised data	0.847	0.779	0.812

Table 14. Precision, recall and F1 score for AF detection SVM with multi-type features

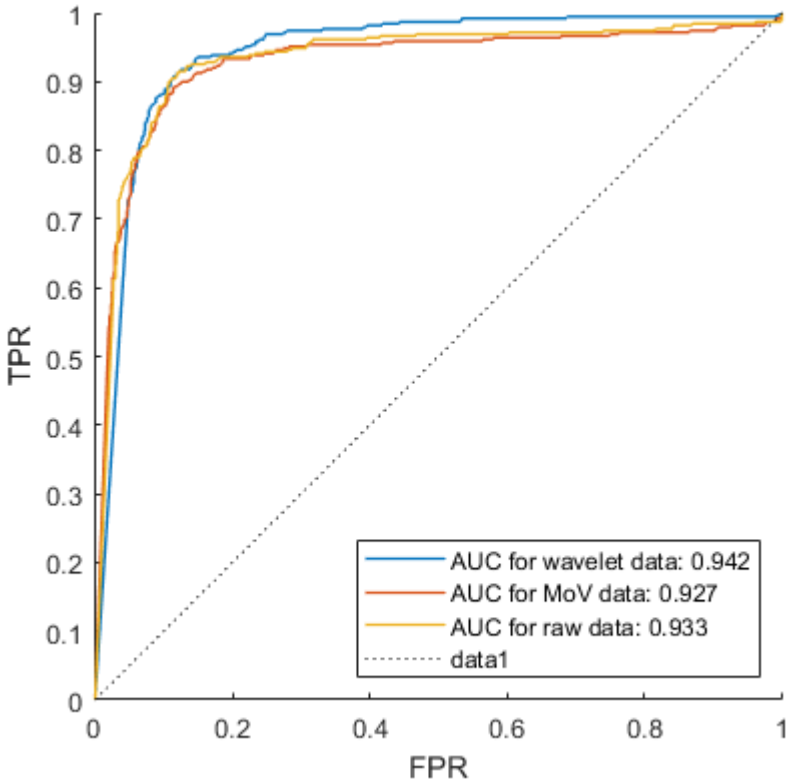

Figure 33. Robustness comparison of SVMs using multi-type features extracted from raw/filtered data

4.5.2 Long-Short Term Memory

Layer	Name	Specification
1	Sequence Input	Sequence input with 2 dimensions
2	BiLSTM	BiLSTM with 100 hidden units
3	Fully Connected	3 fully connected layer
4	Softmax	softmax
5	Classification Output	Crossentropyex

Table 15. LSTM network for ECG classification

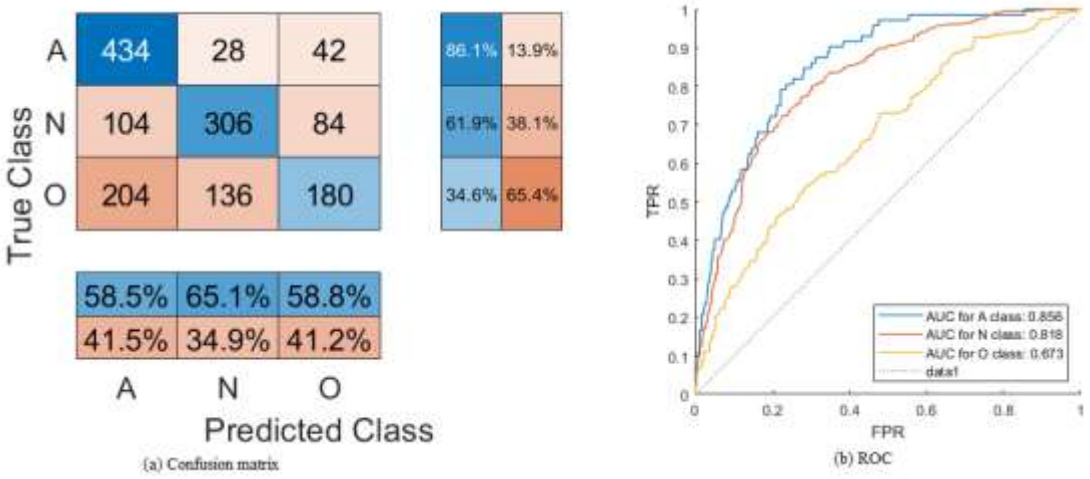

Figure 34. Confusion matrix and ROC of LSTM using frequency features from raw data

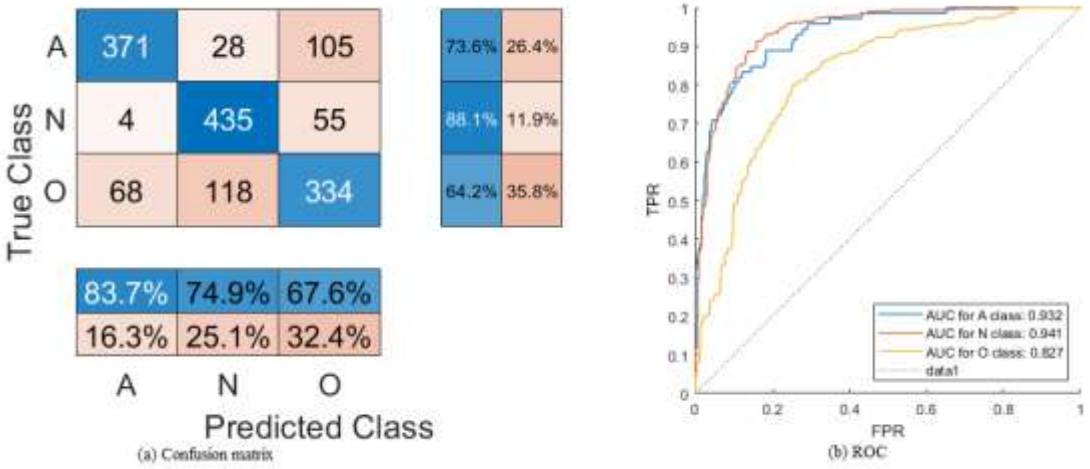

Figure 35. Confusion matrix and ROC of LSTM using frequency features from wavelet denoised data

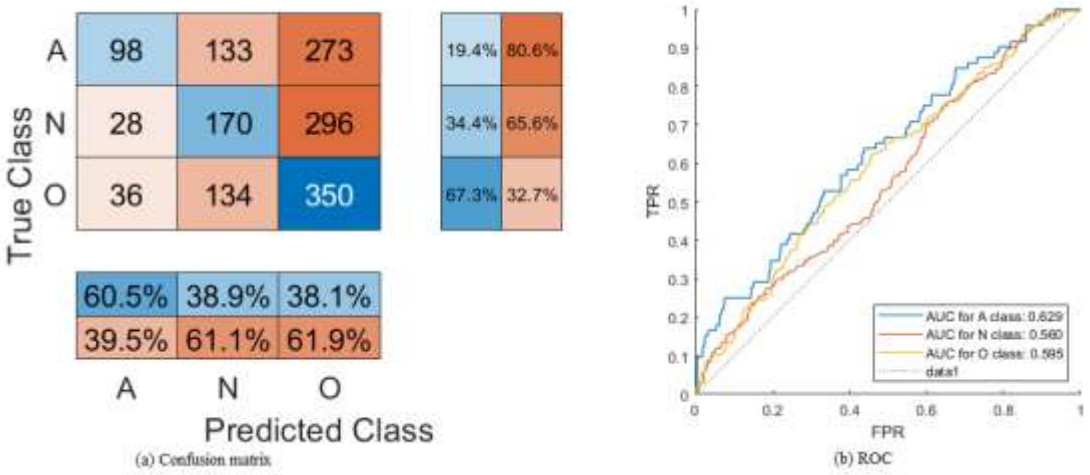

Figure 36. Confusion matrix and ROC of LSTM using frequency features from MoV denoised data

	Precision	**Recall**	**F1 score**

Raw data	0.585	0.861	0.7
Wavelet denoised data	0.837	0.736	0.783
MoV denoised data	0.605	0.194	0.294

Table 16. Precision, recall and F1 score for AF detection LSTM with frequency features

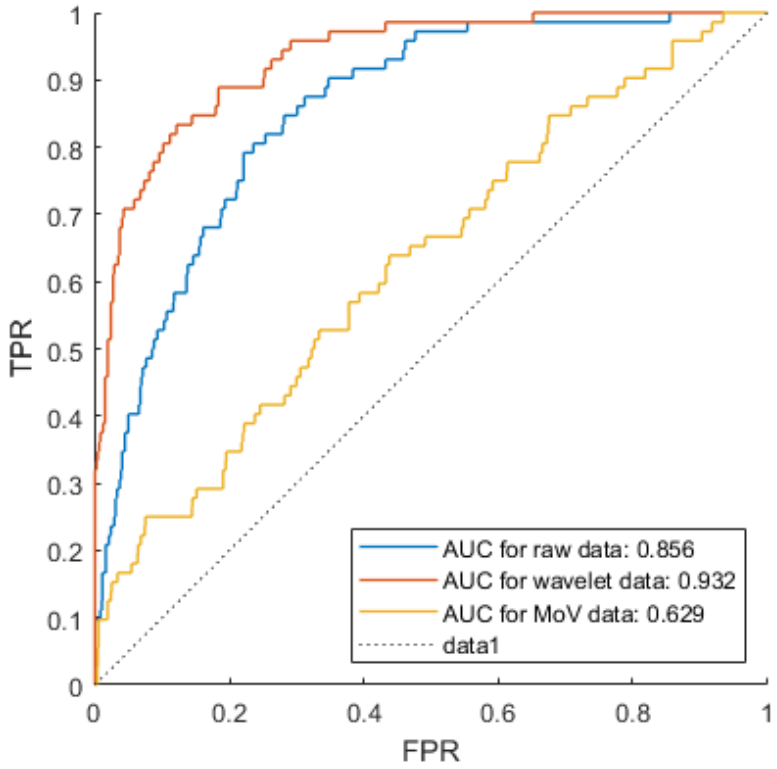

Figure 37. Robustness comparison of LSTMs using frequency features extracted from raw/filtered data

4.5.3 Transfer Learning

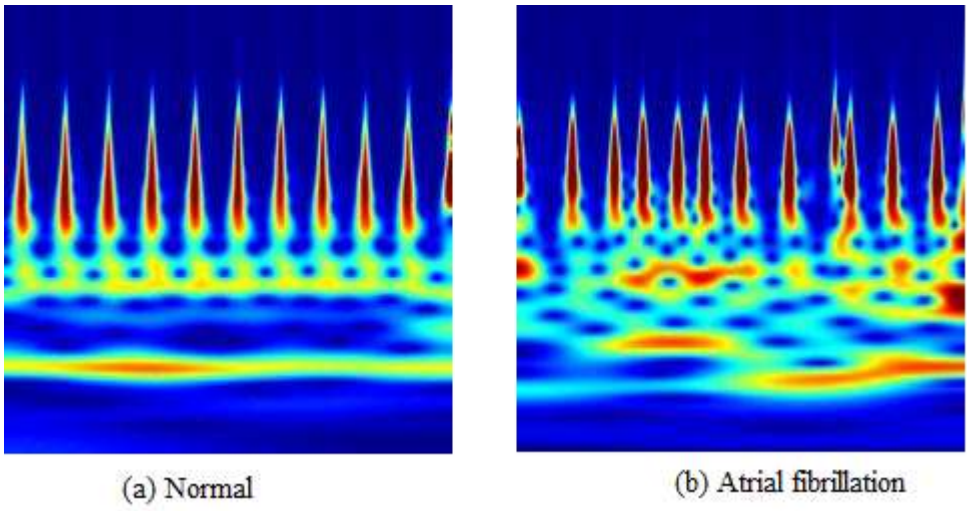

Figure 38. Spectrogram of Normal and AF rhythm

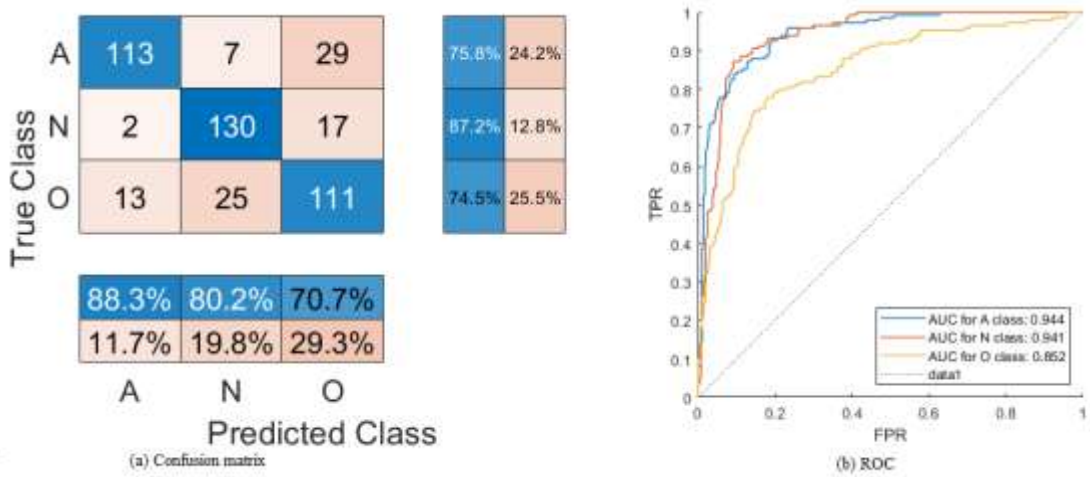

Figure 39. Confusion matrix and ROC of CNN using scalogram from raw data

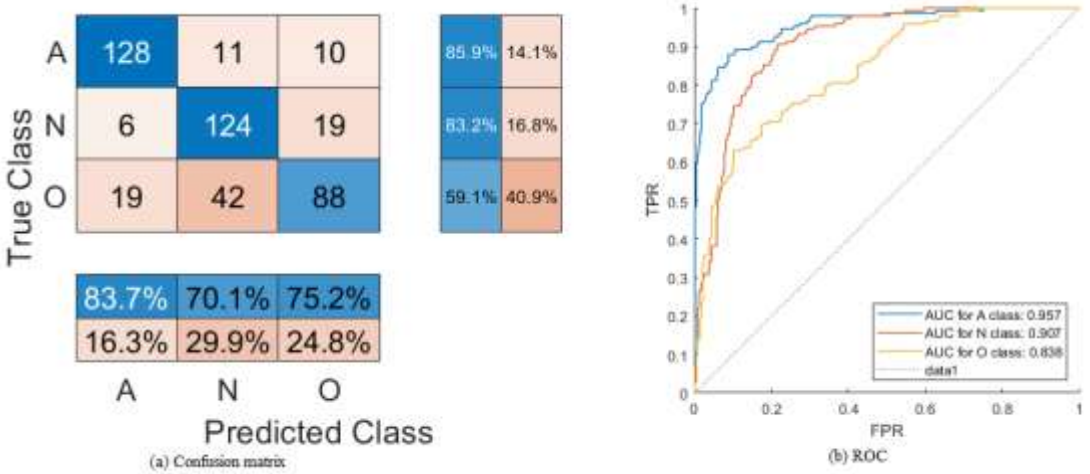

Figure 40. Confusion matrix and ROC of CNN using scalogram from wavelet denoised data

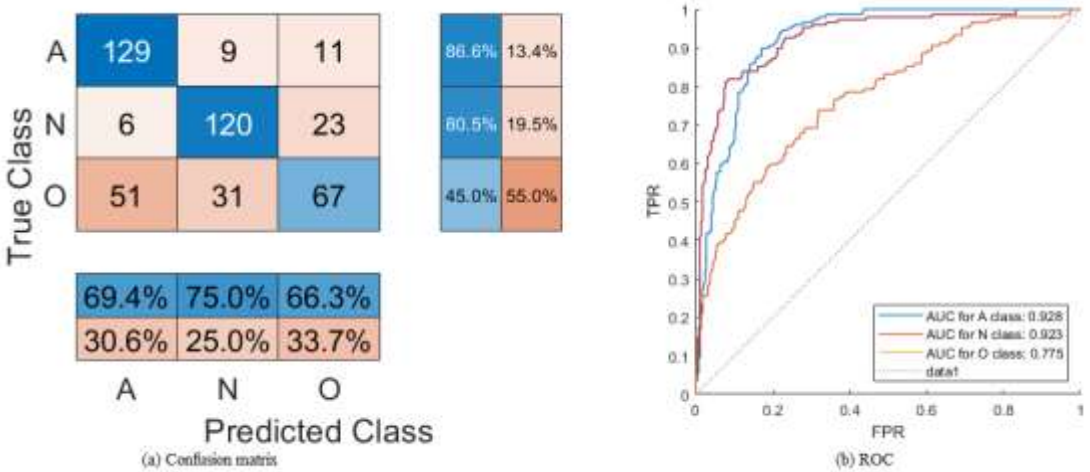

Figure 41. Confusion matrix and ROC of CNN using scalogram from MoV denoised data

	Precision	Recall	F1 score
Raw data	0.883	0.758	0.816
Wavelet denoised data	0.837	0.859	0.848
MoV denoised data	0.694	0.866	0.771

Table 17. Precision, recall and F1 score for AF detection using CNN

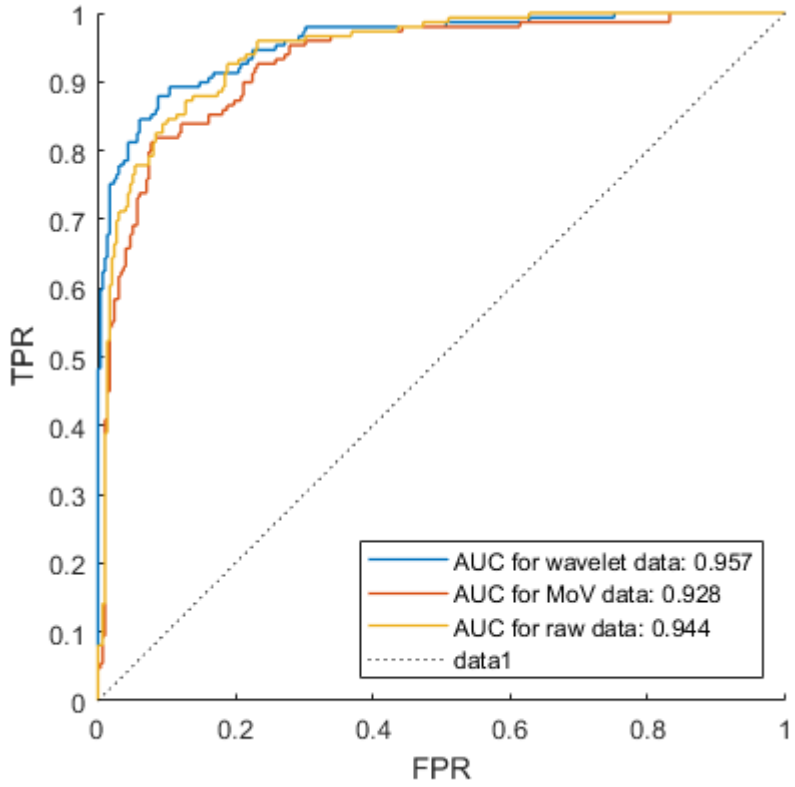

Figure 42. Robustness comparison of CNN using scalogram from raw/filtered data

4.6 COMPARISON BETWEEN DIFFERENT CLASSIFIERS AND EFFECT OF DENOISING TECHNIQUES

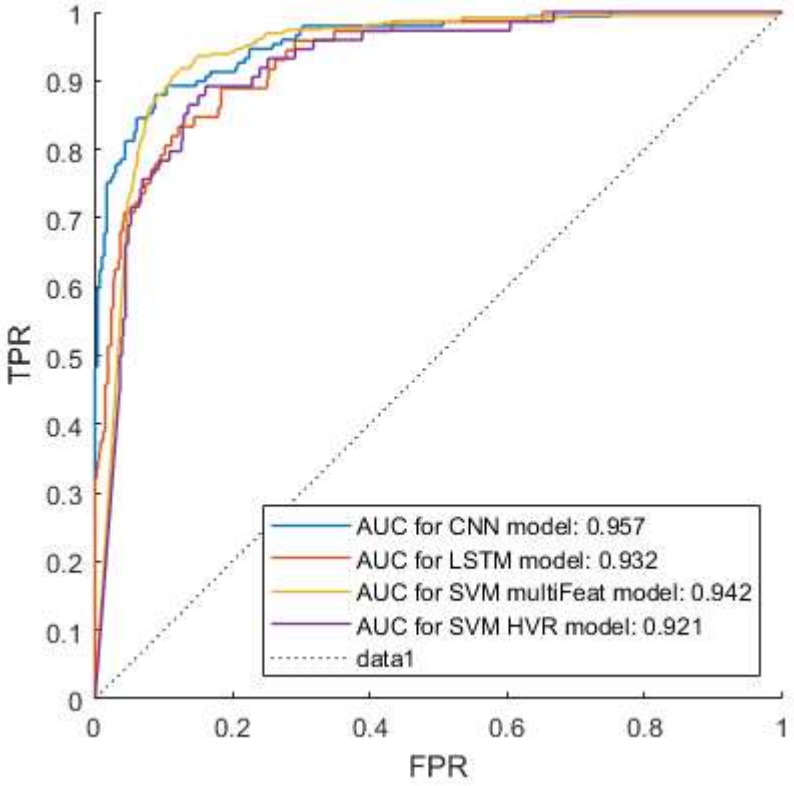

Figure 43. Robustness comparison between various classifiers on denoised data

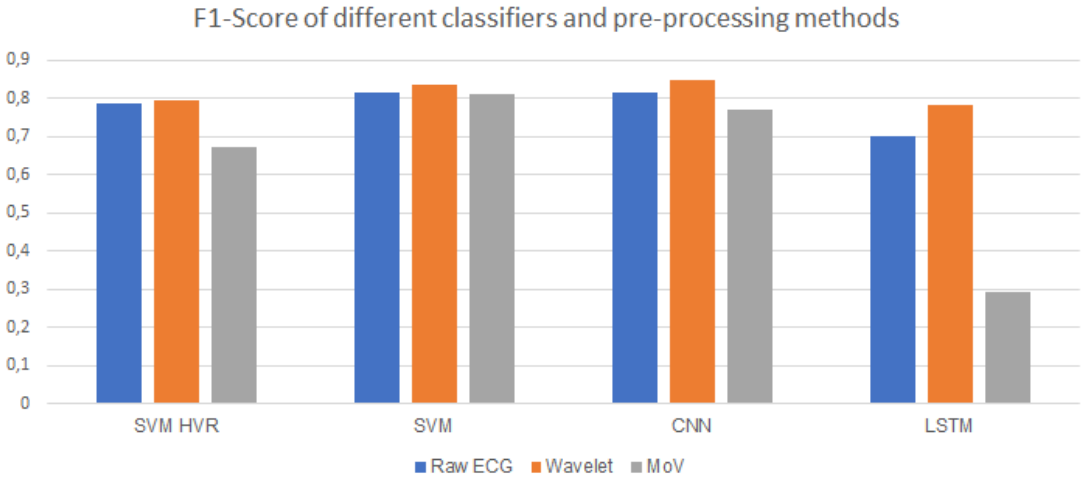

Figure 44. Comparison of results for each technique

The Figure 44 above is plotted based on F1-score of class A (e.g., atrial fibrillation). We decided to used F1-score of AF because it is crucial to recognize life-threatening disease from normal and other signals. Given early diagnoses, people can prevent premature death by early actions.

The results demonstrated that CNN and SVM using multi-type features outperformed other classification techniques, particularly when the wavelet filtering approach was applied. Similarly, LSTM model also performed well in case of wavelet denoised data was used. However, the input sequence of LSTM were instantaneous frequency and spectral entropy which were extremely sensitive to noise, leading to a low result in raw data. Moreover, MoV transformation remove all the low frequency features, especially P wave, which is useful for identifying AF, hence resulted in worst F1-score of 0.3.

In addition, SVM using HVR features ranked closely to SVM using multi-type features means that the most important attributes for distinguishing AF from other rhythms are heart rate variability (e.g., variation of heartbeat duration over time).

While pre-processing, especially denoising step was proven to be necessary for efficient classification, MoV-based filtering seemed to have no effect on increasing performance of classifiers. There are two reasons for this phenomenon. First, the R-peaks are small in amplitude and greatly vary with time, especially in case of muscle contraction noise when there is no peak was located as shown in Figure 24. Secondly, the Pan Tompkins algorithm used for identifying QRS complex is not optimal and adaptive enough to recognize small peaks.

The ROC and AUC shown in Figure 43 demonstrate that CNN and multi-feature SVM line closely to each other and significantly classify AF better than other investigated techniques do. ROC curves of CNN and multi-feature SVM line at very top-left corner and their AUCs are also greater than that of other models.

Chapter 5: Conclusion and Future Work

5.1 ACHIEVEMENT

Our findings can answer the question:” Can we teach a machine to be cardiologist?”. The response is yes. By performing a sequence of step from analysing and denoising signal, and deploying machine learning algorithms, we can enable a computer to diagnose human health conditions via their physiological recordings. In this way, we can help medical experts in reducing their workload, saving their time and effort for researching purposes or solving complicated problems that need human flexibility. With advanced computational power of computer, it is possible to analyse biomedical signals in way faster and human power.

In machine learning task, especially classification problems, pre-processing plays a primary role since it significantly improve the performance of classifiers. Our results with discrete wavelet transform have demonstrated that with outstanding F1-scores compared with raw dataset. However, an appropriate filtering approach needs to be well tested and take into consideration. Despite moment of velocity technique is designed for R-peak detection, it still has some limitation when dealing with real-world recordings which were highly suffered from environment noise or human artifacts.

Input features for machine learning can greatly influence the performance of model. In the case of LSTM using raw ECG as an input, the training process was extremely time-consuming with low accuracy, while using frequency features, the classification ability was sharply improved with less training time. Features in time domain, frequency domain or time-frequency domain can be used to effectively classify ECGs as our results shown. However, heart rate variability features were indicated to be the most important attribute for atrial fibrillation recognition since it ranked closely to multi-feature SVM and cutting-edge deep learning methods e.g., CNN.

Transfer learning can be utilized to save time and effort to build and train deep neural network from scratch. A well-demonstrated pre-train model can be deployed to adapt to new data and learn from it. SqueezeNet has shown the result of this approach as the top classifier among the investigated ML methods. With its small size, it is feasible to implement into limited-memory hardware for real time AF detection.

5.2 LIMITATION

Despite progress and accomplishment, there are several challenges we have faced during the project that negatively impacted the result.

5.2.1 Imbalanced data

The PhysioNet challenge 2017 has an imbalanced dataset since AF recordings are much less than Normal and Other signal. With the lack of material, ML algorithm tends to predict the outnumber class over other classes. At the start of the project, some attempts were performed to classify the original dataset (e.g., imbalanced) but the results were not good and hence were not shown.

5.2.2 Muscle contraction noise

As mentioned in chapter 4, EMG was the major challenge when analysing ECG. It affected the peak detection algorithm as well as polluted the frequency spectrum, indirectly impacted feature extraction process, and hence the prediction ability of classifiers.

5.2.3 Machine learning and optimal models

Since machine learning is a totally new field for us, it was initially difficult to self-learn and find suitable resources to practice algorithm. Therefore, our models were built based on available examples from MathWorks with modification and adaptation to our data. Although our developed models also got a quite high performance, it still cannot be compared with optimal models in the literature review. Hence, there is huge combination of hyperparameters, and we cannot cover all of the possibility that lead to the optimal models. Therefore, a deeper understanding about machine learning and neural networks as well as more fine-tuning will produce a more reliable model.

5.3 FUTURE WORK

For the future direction, we aim to apply other effective denoising techniques to cope with muscle contraction noise and other possible noise that may interfere during biomedical signal recordings. In case of imbalanced data, other advanced data augmentation e.g., adding noise, image rotation… will be implemented to overcome the lack of data sample.

More practical experience and deeper understanding about machine learning maybe required to increase the classification performance. Try other classifiers and build our own neural networks for better ECG pattern recognition. Find out more optimization methods for tuning hyperparameters.

Our developed models are currently able to distinguish AF from Normal and Other signal. It is quite general and simple to be applied in real application. Therefore, we are planning to extend the classification ability of our models so that it can detect a greater range of cardiovascular conditions.

Since our models are very small and use only 15-second length ECG to predict AF, it is promising to deploy these models into compact wearables and diagnose

heart problems in real time. With that aim, we will learn and apply some embedded techniques to run our classifiers on hardware. In addition, it will have adaptive function that characterize for individual user. Our goal is making a real-time prediction device that can help to monitor and diagnose people health, prevent them from premature death.

References

[1] H. Foundation. "Key Statistics: Cardiovascular Disease," 3, 2021; https://www.heartfoundation.org.au/activities-finding-or-opinion/key-stats-cardiovascular-disease.

[2] WHO. "Cardiovascular diseases (CVDs)," 3, 2021; https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[3] B. J. Aehlert, *ECGs made easy-E-Book*: Elsevier Health Sciences, 2015.

[4] S. Meek, and F. Morris, "Introduction. I—Leads, rate, rhythm, and cardiac axis," *Bmj*, vol. 324, no. 7334, pp. 415-418, 2002.

[5] F. Morris, W. J. Brady, and A. J. Camm, *ABC of clinical electrocardiography*: John Wiley & Sons, 2009.

[6] M. Clinic. "Atrial fibrillation," 3, 2021; https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624.

[7] J. Rodrigues, D. Belo, and H. Gamboa, "Noise detection on ECG based on agglomerative clustering of morphological features," *Computers in biology and medicine*, vol. 87, pp. 322-334, 2017.

[8] R. Kher, "Signal processing techniques for removing noise from ECG signals," *J. Biomed. Eng. Res*, vol. 3, pp. 1-9, 2019.

[9] D. Zhang, S. Wang, F. Li, J. Wang, A. K. Sangaiah, V. S. Sheng, and X. Ding, "An ECG signal de-noising approach based on wavelet energy and sub-band smoothing filter," *applied sciences*, vol. 9, no. 22, pp. 4968, 2019.

[10] B. N. Singh, and A. K. Tiwari, "Optimal selection of wavelet basis function applied to ECG signal denoising," *Digital signal processing*, vol. 16, no. 3, pp. 275-287, 2006.

[11] S. Talebi. "The Wavelet Transform," 5, 2021; https://towardsdatascience.com/the-wavelet-transform-e9cfa85d7b34.

[12] M. Dorraki, A. Fouladzadeh, A. Allison, B. Davis, and D. Abbott, "On moment of velocity for signal analysis," *Royal Society open science*, vol. 6, no. 3, pp. 182001, 2019.

[13] E. H. Houssein, M. Kilany, and A. E. Hassanien, "ECG signals classification: a review," *International Journal of Intelligent Engineering Informatics*, vol. 5, no. 4, pp. 376-396, 2017.

[14] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati, "Classification of ECG signals using machine learning techniques: A survey." pp. 714-721.

[15] Q. Zhao, and L. Zhang, "ECG feature extraction and classification using wavelet transform and support vector machines." pp. 1089-1092.

[16] F. Shaffer, and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, pp. 258, 2017.

[17] C. Venkatesan, P. Karthigaikumar, A. Paul, S. Satheeskumaran, and R. Kumar, "ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications," *IEEE Access*, vol. 6, pp. 9767-9773, 2018.

[18] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification." pp. 1-4.

[19] R. Gandhi. "Support Vector Machine — Introduction to Machine Learning Algorithms," 3, 2021; https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.
[20] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, and C. Ou, "Automatic ECG Classification Using Continuous Wavelet Transform and Convolutional Neural Network," *Entropy*, vol. 23, no. 1, pp. 119, 2021.
[21] J. Wang, P. Wang, and S. Wang, "Automated detection of atrial fibrillation in ECG signals based on wavelet packet transform and correlation function of random process," *Biomedical Signal Processing and Control*, vol. 55, pp. 101662, 2020.
[22] M. Dorraki. "Moment of Velocity," 9, 2021; https://github.com/Dorraki/Moment-of-Velocity/blob/master/MomentOfVelocityII.m.
[23] F. Andreotti, O. Carr, M. A. Pimentel, A. Mahdi, and M. De Vos, "Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ECG." pp. 1-4.
[24] F. Andreotti. 8, 2021; https://github.com/fernandoandreotti/cinc-challenge2017/tree/master/featurebased-approach.
[25] MATLAB. "Classify ECG Signals Using Long Short-Term Memory Networks," 7, 2021; https://au.mathworks.com/help/signal/ug/classify-ecg-signals-using-long-short-term-memory-networks.html.
[26] M. K. Gajendran, M. Z. Khan, and M. A. K. Khattak, "ECG Classification using Deep Transfer Learning." pp. 1-5.
[27] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," *arXiv preprint arXiv:1602.07360*, 2016.
[28] H. Sedghamiz. "Complete Pan Tompkins Implementation ECG QRS detector," 7, 2021; https://au.mathworks.com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementation-ecg-qrs-detector.
[29] A. A. Tokuç. "Why Feature Scaling in SVM?," 8, 2021; https://www.baeldung.com/cs/svm-feature-scaling.

Appendices

ONE-PAGE REVIEW
CAN WE TEACH A MACHINE TO BE CARDIOLOGIST?

Electrocardiogram (ECG) is a periodic signal which reflects the activity of the heart. A lot of valuable information is obtained from ECG for diagnosing normal and abnormal physiology of the heart. However, the ECG signal is challenging to analyse due to its unstable characteristic and can be varied from people to people [1]. There are three waves with a frequency of 1Hz in one cardiac cycle, representing for different functions of atria and ventricular parts of the heart [2]:

- P wave is produced by atrial depolarization and the duration is less than 120 ms.
- QRS complex wave has duration of 70-110 ms and it represent ventricular depolarization.
- T wave is produced by the repolarization of ventricles having duration greater than 300 ms.

Irregular waveforms of P, QRS or T will result in abnormal heart rhythms. Atrial Fibrillation (AF) is one kind of abnormalities that can be early detected by ECG techniques. It is characterized by chaotic electrical impulses in the atria, which leads to irregular heartbeats and can develop blood clots and stroke [3].

In order to obtain information in the ECG waveforms and use it for detect different heart problems, several methods using machine learning and deep learning is discussed: Support Vector Machine (SVM), artificial neural networks (ANNs), Adaboost, Naïve Bayes and Random Forest [2-7].

There are four steps for ECG classification, namely: pre-processing, feature extraction, feature selection and classification [7]

Preprocessing:

The ECG signal must first be filtered out of noise for more precise classification. This preprocessing stage avoids the overlap between ECG and motion artifacts as well as high frequency disturbances. Low-pass, high-pass, Butterworth band-pass filters and Wavelet denoising are used to eliminate the mentioned noise [1][4].

Feature Extraction and Selection:

QRS complex, P and T waveforms are identified prior to extract some necessary features to input to machine learning classification. Different signal features such as inter-beat mean value, standard deviation, heart rate variability, … are calculated using beat-to-beat intervals time series between the mentioned waves [4]. The feature selection process ordinarily is designed to provide a means for choosing the features which are best for optimised classification [1][7].

Classification:

The selected features are then inputted to an appropriate classifier. In recent years, SVM and ANN have gained attention by most researchers in ECG classification [7].

SVM is a supervised learning method used for classification and regression prediction tool to maximize the accuracy. It uses N - the number of features extracted to find hyperplane that separate classes of data points in the N –dimensional space [2][4].

ANN is a widespread and efficient tool for pattern recognition problems like heartbeat classification. It is designed to simulate the way the human brain analyses and processes information. In ANNs, there are three layers: input, hidden and output layer which it uses to learn and predict the desired output [5][7]. In terms of accuracy, the ANN is more accurate than SVM within the same data size [4].

Validation

Validation is a process that access the performance of the machine algorithm when making prediction on a testing data. For small dataset and the accurate estimation is more important than the computational cost, leave-one-out cross-validation is preferred [9]. Then a best model with highest accuracy will be selected.

References

[1] S. H. Jambukia ,et al., *Classification of ECG signals using machine learning techniques: A survey*, IEEE, 2015, Accessed: 16 March 2021, [Online] DOI: 10.1109/ICACEA.2015.7164783

[2] S. El-Khafif ,et al, *Artificial Neural Network-Based Automated ECG Signal Classifier*, IEEE, 2010, Accessed: 21 March 2021, [Online] DOI: 10.1109/ICEIE.2010.5559887

[3] M. Zabihi ,et al., *Detection of Atrial Fibrillation in ECG Hand-held Devices Using a Random Forest Classifier*, IEEE, 2017, Acessed: 19 March 2021, [Online] DOI: 10.22489/CinC.2017.069-336

[4] S. Celin, K. Vasanth, *ECG Signal Classification Using Various Machine Learning Techniques*, J Med Syst 42, 241 (2018), Accessed: 19 March 2021, [Online] https://doi.org/10.1007/s10916-018-1083-6
[5] Yaghouby F., Ayatollahi A. (2009) *An Arrhythmia Classification Method Based on Selected Features of Heart Rate Variability Signal and Support Vector Machine-Based Classifier*. In: Dössel O., Schlegel W.C. (eds) World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany. IFMBE Proceedings, vol 25/4. Springer, Berlin, Heidelberg, Accessed: 19 March 2021. [Online] https://doi.org/10.1007/978-3-642-03882-2_512
[6] S. Padmavathi., *Naïve Bayes Classifier for ECG abnormalities using Multivariate Maximal Time Series Motif*, ResearchGate, 2015, Accessed: 18 March 2021, [Online] DOI: 10.1016/j.procs.2015.03.201
[7] E. H. Houssein, et al., *ECG signals classification: a review*, ResearchGate, 2017, Accessed: 16 March 2021, [Online] DOI: 10.1504/IJIEI.2017.10008807
[8] S. Faziludeen, P. V. Sabiq, *ECG Beat Classification Using Wavelets and SVM*, IEEE, 2013, Accessed: 16 March 2021, [Online] DOI: 10.1109/CICT.2013.6558206
[9] J. Brownlee, *LOOCV for Evaluating Machine Learning Algorithms*, 27 July 2020, Accessed: 1 April 2021, [Online]: https://machinelearningmastery.com/loocv-for-evaluating-machine-learning-algorithms/