**BIO-SIGNAL PROCESSING REPORT**

DETECTION AND QUANTIFICATION OF MYOCARDIAL DAMAGE USING EKG SIGNAL

**INSTRUCTOR:**

**Dr. Le Quoc Trung**

**STUDENT:**

**Pham Khoi Nguyen**

**Table of contents**

[I. literature review](#_Toc421714432) 1

[1.1 MYOCARDIAL INFARCTION](#_Toc421714434) 1

[1.2 ELECTROCARDIOGRAM SIGNAL](#_Toc421714438) 2

[1.2.1 SIGNAL PHYSIOLOGY](#_Toc421714439) 3

[1.2.2 CHARACTERISTIC WAVES](#_Toc421714440) 4

[ii. METHODOLOGY](#_Toc421714448) 6

[2.1 DATABASE ACQUISITION](#_Toc421714449) 6

[2.2 SIGNAL PREPROCESSING](#_Toc421714451) 9

[2.2.1 BASELINE WANDER REMOVAL](#_Toc421714452) 11

[2.2.2 ESTIMATION AND REMOVAL OF NOISE AND MUSCLE ACTIVITIES](#_Toc421714453) 13

[2.3 ECG DELINEATION](#_Toc421714454) 15  
2.4 FEATURES EXTRACTION 21  
 2.4.1 MORPHOLOGICAL FEATURES 21  
 2.4.2 INTRA - BEAT DETRENDED FLUCTUATION 23  
2.5 SCORING SYSTEM 26

[iii. result and validation](#_Toc421714455) 28

[3.1 FEATURES EXTRACTION SUMMARY](#_Toc421714456) 28

[3.2 ACCURACY VALIDATION](#_Toc421714457) 30

[3.2.1 MORPHOLOGY FEATURES](#_Toc421714458) 30

[3.2.2 DFA FEATURES](#_Toc421714459) 31

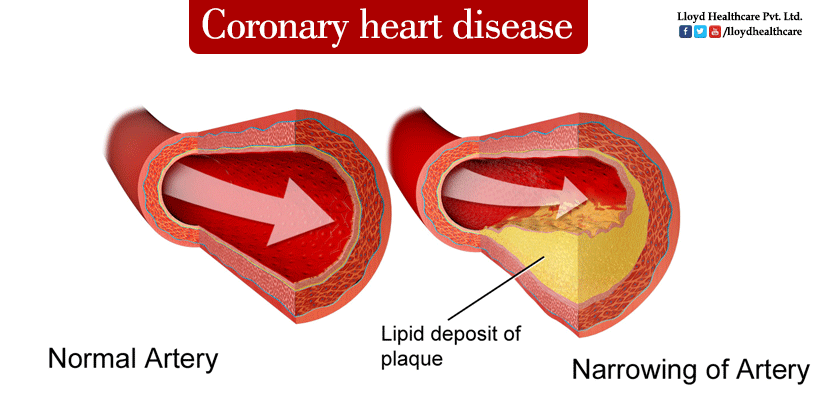
[iv. CONCLUSION AND DISCUSSION](#_Toc421714461) 35

[REFERENCES](#_Toc421714462) 36

LITERATURE REVIEW

### 1.1 MYOCARDIAL INFARCTION

Acute Myocardial Infarction (AMI), also known as Heart Attack, is a disease caused by insufficiency of blood supply to the heart’s tissue [33 - 39]. Generally, heart’s tissue is supported by a system of blood vessels. When these blood vessels suffer from Coronary Artery Disease (CAD) – the constriction of the artery that obstruct blood flow by the formation of fat and cholesterol beneath the vessel’s inner wall, some part of the heart does not receive enough blood supply. This phenomenon, if left untreated for a period of time, can eventually lead to cells death. The condition when some region of the heart died because of the derivation of blood supply and cannot function normally is called Myocardial Infarction.

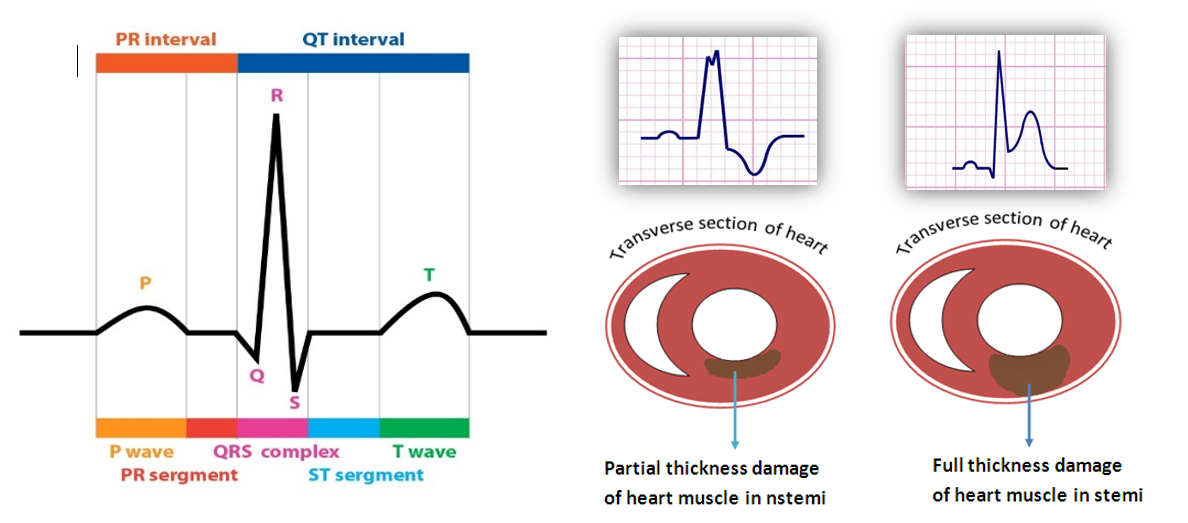


*Figure 1: The most common cause of cardiovascular disease is the formation of a lipid deposit that obstructs artery blood flow, causing an insufficiency of blood supply to the myocardium that eventually leads to cardiovascular damage.*

Acute Myocardial Infarction is the leading cause of death during patients’ hospitalization in the United State. It has been reported that each year in the America, 1.1 millions of people suffer from Myocardial Infarction and haft of them get an acute attack [4]. In Vietnam, the number of patient who suffered from AMI tends to increase drastically during the last 20 years (from 1980 to 2000): from 1980 to 1990 there was 108 patients, from 1991 to 1995 there was 92 patients and from 2000 to 2001 the number escalated quickly: more than 1.500 patients [4] and 17.4% (261 patients) died [4]. Most of these patients are elderly (>65 years old) whose biological characteristics of coronary artery make them more susceptible to AMI.

### ELECTROCARDIOGRAM SIGNAL

ECG, or Electrocardiogram, measures the electrical activity of the heart during each consecutive heart beats. The current clinical diagnostic technique using ECG analyses the shape of the waveform and calculate the magnitude, energy and entropy of the signal to deliver valuable information about the heart. For example, by focusing on some specific segments of the signal: P wave, T wave, the presence of Q wave and ST segment, detection of myocardial infarction, cardiac arrest and arrhythmia can be achieved.

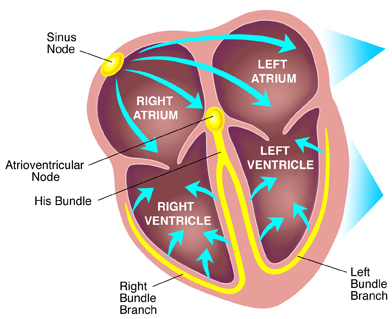


***Figure 7****: from left to right are a normal ECG waveform versus two altered ECG waveforms that correspond to different types of cardiovascular problem*

ECG measurement has been used as a standard procedure for approving patients to the Heart Disease Department in almost hospitals around the world because of its low expense, fast and reliable. In addition to that, technical aspects of the signal also elevate ECG as a wonderful approach to develop a small, light-weight device that is suitable for home care solution. To illustrate this point, many famous chipset companies such as Texas Instrument and National Instrument, is currently providing small, affordable ECG modules with very good signal quality. Most importantly, the greatest interesting feature of this signal remains in its medical prognostic value. The use of ECG to forecast the occurrence of heart attack is still an uncultured field but yet extremely profitable if it was discovered. The model can help elder people prevent the occurrence of AMI or make immediate responses to sudden cardiac attack.

### 1.2.1 SIGNAL PHYSIOLOGY

Figure 7 above represents a typical example of the electrical activity of the heart. ECG (Electrocardiogram) is a sequence of wave forms that manifests the dynamical activity of the heart during consecutive heart beats. Each part of the wave form corresponds to a specific function in a specific region of the heart.

****

***Figure 8****: Electrical conduction system of the heart*

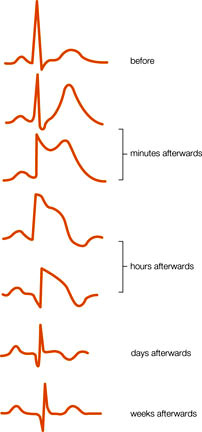
Our heart consists of 4 champers: 2 atrial receiving blood from the body and 2 ventricles that receive blood from the atrial and pump back to the body. The pumping activity of these champers is described as the peak value in the ECG wave form, where the first peak (P wave) represents for the depolarization and contraction of the atrial, and the second peak (R wave) represents for the depolarization and contraction of the ventricle. The final peak (T wave) manifests the repolarization, or relaxation, of the ventricle thus preparing the heart for the next heart beats.

When analyzing ECG signal, the combination interpretation of waves together with the interval between these waves can yield reliable information about the overall activity of the heart. Each normal heart beat consists of a P wave, a QRS complex and a T wave and the corresponding interval in between. In figure 1.2, the electrical current that flows through the heart in each consecutive heart beat is described, providing the physiology perspective of the intervals.

First an action potential is generated at the SA node, producing an electrical current that flows to the atrial. This electrical current stimulates the contraction of the atrial, resulting in the formation of P wave in the ECG signal. The current then flow through the atrial to the AV node, where it then travels to the left and right branch of the ventricle by conductive fibers known as left bundle branch and right bundle branch. The arrival current then stimulates the contraction on both the left and right ventricle, representing the R peaks in QRS complex. It is also important to mention that during this interval, the simultaneous contraction of the ventricle and relaxation of the atrial contribute to the complexity of the QRS pattern. After this point, ST segment represents for the time delay between ventricle contraction and relaxation, where all the ventricular tissues, after contraction, tend to bounce back to normal in order to be ready for the repolarization. This is the most important interval for spotting any damage to the ventricle tissues because the electrical current developed by any injury or inflammation to the ventricle appears clearly during this isoelectric process. Finally, the T wave occurs as the result of ventricle repolarization, or relaxation.

### CHARACTERISTIC WAVES

As previously mentioned, the ST segment is the most valuable part of the ECG waveform to identify any damage to the ventricle, where most AMI occur. Beside ST segment, other valuable waves that also need attention is the T wave and Q wave. Acute Myocardial Infarction manifestation in ECG signal is a sequence of changes in the T wave, ST segment and Q wave as described in figure 7. Firstly, the constricted blood vessels prevent or decreases blood flow to some specific region of the heart. Atrial tend to have larger coronary vessels, therefore it is less susceptible to blood derivation. However, the ventricle is supported by a complicated vessel system, some is really big (Bundle of His) and some is really small (Purkinje Fibers), therefore the restriction of blood supply is more viable, thus making the ventricle more susceptible to blood derivation. When the tissue is lack of blood supply, the Ischemic Event occurs, resulting in the injury or inflammation in some region of the heart. These injury generates addition electrical current that can be detected during isoelectric process of the heart: the ST segment. During this phase, T wave first becomes peaked and then ST changes occur. If the ST segment elevates, the myocardium is interpreted as having full thickness damage of the heart muscle (Figure 7b). If the ST segment depresses, then the traverse damage can be the cause (Figure 7c).



***Figure 9****: Dynamical changes of ECG waveform during the formation of AMI*

Finally, if the disease is left untreated for a long period of time, tissues death will eventually occur. The formation of a pathological Q wave (larger and more negative Q wave) also develops during this stage. This is in fact the final stage of Acute Myocardial Infarction.

In conclusion, if the ST segment are elevated representing tissue injury, the phenomenon is categorized as ST Segment Elevation Acute Myocardial Infarction (STEMI). If the ST are horizontally normal or depressed representing ischemic event or tissue injury, it is categorized as Non – ST Segment Elevation Myocardial Infarction (NSTEMI).

METHODOLOGY

**2.1 DATABASE ACQUISITION**

First, EKG physiological database dedicated for cardiovascular diseases detection is researched and documented. The target database needs to contain high quality, good resolution EKG signal with considerably long measuring time that is higher than couple of minutes. Therefore, the following databases used in this research all contain a decent signal acquisition system that is less subjected to noise, all come with 12 to 14 bits of data resolution, digitalized at 250 samples per second and the recording time ranges from 2 hours to 20 hours of continuous monitoring. In addition, it is useful that clinical diagnosis is also documented to provide method for validation.

**2.1.1 LONG-ST DATABASE**

The Long-Term ST Database is the most novel EKG database dedicated for development and quantification of ischemia and other types of cardiovascular diseases. It contains 86 lengthy ECG recordings of 80 human subjects, chosen to exhibit a variety of events of ST segment changes, including ischemic ST episodes, axis-related non-ischemic ST episodes, episodes of slow ST level drift, and episodes containing mixtures of these phenomena. The greatest advantage using this database is that it also provides disease description for each patients, however, many of its records are subjected to high noise level that reder them unsuable for the scope of this research.

The following records are carefully chosen from the database, with the aim to only quantify clear signal while still provide various cases of myocardial injury.

data\_path = 'C:\Nguyen Pham\MY THESIS\database\longst\';

recordings = [20011 20021 20031 20041 20051 20061 20071 20081 20091 20101 20111 20121 20131 20141 20151 20171 20181 20191 20201 20211 20221 20231 20241 20251 20261 20271 20272 20274 20461 20161 20361];

leads = ones(1,length(recordings));

**2.1.2 EUROPEAN DATABASE**

**2.1.3 ST CHANGES DATABASE**

**2.1.2 EUROPEAN DATABASE**

Beside Long - ST database, the European database is also very famous in the field of medical application for cardiovascular damage. It is intended to be used for algorithms evaluation of ST and T-wave changes. This database consists of 90 annotated excerpts of ambulatory ECG recordings from 79 subjects. Not only that, Myocardial ischemia was diagnosed or suspected for each subject, additional selection criteria were established in order to obtain a representative selection of ECG abnormalities in the database, including baseline ST segment displacement resulting from conditions such as hypertension, ventricular dyskinesia, and effects of medication. Each of these records lasts for 20 hours and comes with very good quality and resolution.

The following records will be used in this research:

data\_path = 'C:\Nguyen Pham\MY THESIS\database\euro\';

recordings = [103 104 105 106 112 113 118 121 129 133 136 139 154 161 162 163 170 105 108 112 115 123 129 133 147 154 104 112 118 122 154 161 612 801 808];

leads = [001 002 001 002 002 002 001 001 002 002 002 002 002 002 002 002 001 001 001 002 002 002 002 002 002 002 001 002 002 002 001 002 002 002 002];

**2.1.3 ST CHANGES DATABASE**

Finally, ST changes database is a compact and suplementation to the European database. This database includes 28 ECG recordings of varying lengths, most of which were recorded during exercise stress tests and most of which exhibit transient ST depression. Due to this practice, most of the records have very high noise level, therefore, a rejection criteria need to be developed to reject outputs generated from noisy and unsuable input. This section will be covered later on in this section.

Because each record only lasts for 2 hours, all of them will be included in this research.

data\_path = 'C:\Nguyen Pham\MY THESIS\database\stchange\';

recordings = 300:327;

leads = ones(1,28);

In conclusion, a total of 94 records will be used to train and validate the algorithm developed in this research. Each record has 12 or 14 bit data resolution, digitalized at 250 samples per second. The time range used for each records will be 1 hour long and a moving window of 10 seconds will be applied to calibrate the EKG features. The following code describes the whole process of reading EKG signal, signal preprocessing and baseline wander removal that will be covered in the next parts of this research. The research methodologies is also described in figure 15.

% filename: EURO\_TASK.m

for record = 1:length(recordings)

try

filename = ['e0' num2str(recordings(record))];

disp(filename);

full\_path = [data\_path filename '.hea'];

ECGw = ECGwrapper( 'recording\_name', full\_path);

% READ SIGANL AND ANNOTATION-------------------------

ann = ECGw.ECG\_annotations;

hea = ECGw.ECG\_header;

sig = ECGw.read\_signal(1,hea.nsamp);

sig1\_raw = sig(:,leads(record));

sig1\_raw = sig1\_raw(1:end);

% NORMALIZATION CODES--------------------------------

sig1\_raw = sig1\_raw - mean(sig1\_raw);

L = length(sig1\_raw);

Ex = 1/L \* sum(abs(sig1\_raw).^2);

sig1\_raw = sig1\_raw / Ex;

% BASELINE REMOVE USING Wavelet\_decompose------------

[approx, detail] = wavelet\_decompose(sig1\_raw, 8, 'db4');

sig1 = sig1\_raw - approx(:,8);

sig\_backup = sig1;

% NORMALIZA THE SIGNAL FROM 0 TO 1

sig1 = sig1 + abs(min(sig1));

sig1 = sig1 / max(sig1);

% GENERAL PARAMETERS---------------------------------

fs = hea.freq;

ts = 1/fs;

try

REPORT;

catch

disp('An error occured while calibrating this record');

failed\_records{end + 1} = filename;

end;

catch

disp(['record ' filename ' not found. Proceed to next one']);

end;

end;

**2.2 SIGNAL PREPROCESSING**

**2.2.1 SIGNAL NORMALIZATION**

The purpose of normalization is to scale down the signal into a specific range so that records from different databases become comparable to each other. There are various techniques for signal normalization. In this research, the signal is normaized against its net energy.

1. Firstly, the signal is substracted by its mean
2. Then net energy value is calculated as follow:
3. Devide the orginal signal with the net energy to get the scaled version
4. Then change the scale into from 0 to 1 with the equation

Matlab code presentation:

% NORMALIZATION CODES---------------------

sig1\_raw = sig1\_raw - mean(sig1\_raw);

L = length(sig1\_raw);

Ex = 1/L \* sum(abs(sig1\_raw).^2);

sig1\_raw = sig1\_raw / Ex;

% NORMALIZA THE SIGNAL FROM 0 TO 1

sig1 = sig1 + abs(min(sig1));

sig1 = sig1 / max(sig1);

Wavelet Filtering

Baseline wander removal

Normalization

Input EKG signal

Yes

P wave absence?

Atrium damage

Delineation

No

Q wave presence?

Yes

*Figure 15: research methodology of this paper*

Irreversible myocardial infarct

No

Yes

Myocardial ischemia

Tarchycardia

Heart rate

Fast?

No

No

Yes

Yes

STD in mirror leads?

Myocardial infarction

Slow?

Bradycardia

No

No

Down

STD direction?

Chaotic?

Yes

ST Deviation calculation

Arrythmia

Up

Yes

Peaked or inverted?

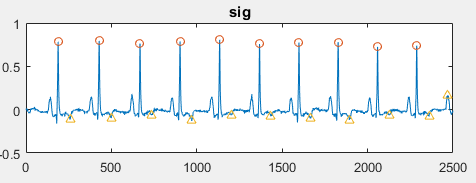
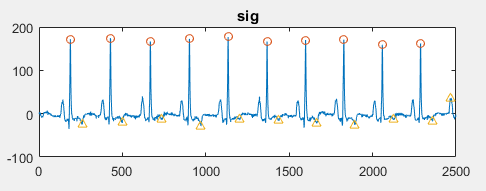
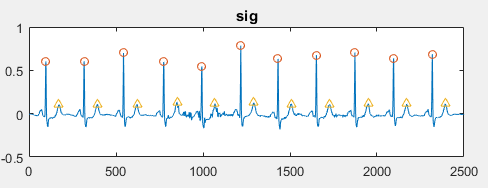
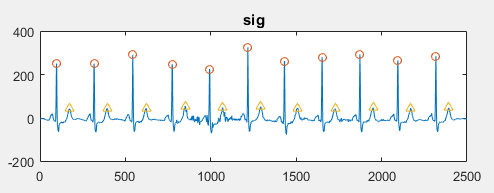
Myocardial infarction

Ischemia or posture changes

T wave height calculation

No symptoms

No

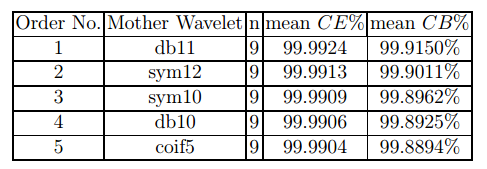
****

***Figure 16:*** *Result obtained after normalizing the signal. Two segment of data with different value ranges are scaled down into the range of from 0 to 1 to make comparison between them becomes feasible.*

**2.2.2 BASELINE WANDER REMOVAL**

During recent years, many researchers have constantly report the use of wavelet decomposition for removing low frequency components within biological signals. According to Khawala, the baseline wander of EKG signal can be accurately located at the 9th level approximation coefficient of Daubechies11 mother wavelet. In another study, the authors also suggest using Daubechies4 mother wavelet to perform up to 4th level of decomposition and use the approximation coefficient at level 4th to address the baseline wander. Baseline wander removal using wavelet decomposition is one of the most novel and state of the art in the field of EKG signal processing because it is not only effective, but also address the disadvantages corresponding to the conventional FIR and IIR filter design. It is very well known that after applying FIR or IIR filter into EKG signal, the morphology of ST segment will be altered, which will have critically negative effect on the analysis of the signal because ST segment is the most important feature for addressing many types of cardiovascular diseases.

The table below is extracted from Khawala’s paper, in which he performed baseline wander removal using many different types of mother wavelet. From his research, it is documented that the best technique for removing baseline wander is using Daubechies11 mother wavelet to perform signal decomposition. The approximation coefficient at the 11th level will be chosen to be the baseline wander and it will be subtracted from the original signal.



***Figure 17****: Result obtained from Khawala research. The best mother wavelet for removing baseline wander for EKG signal is Daubechies11 wavelet at decoposition level 11th*

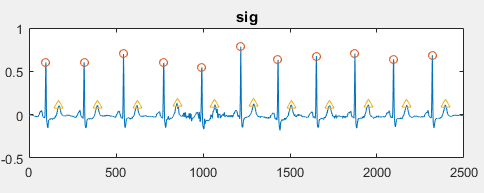
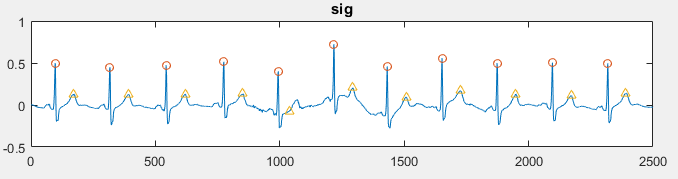
Matlab code presentation:

%-BASELINE REMOVAL-------------------------------------

[approx, detail] = wavelet\_decompose(seg, 11, 'db11');

seg = seg - approx(:,11);

baseline = approx(:,11);



***Figure 18****: Daubechies11 mother wavelet successfully remove the effect of baseline wander on the EKG signal*

**2.2.3 ESTIMATION AND REMOVAL OF NOISE AND MUSCLE ACTIVITIES**

Noise from the EKG signal comes from many different sources. Among these, muscle activities are the most prominent source of noise that greatly affect the quality of detection algorithm. EKG signal processing has been tremendously studied to determine the best frequency range that have critical effect in removing muscle activities while retaining the morphologies of the characteristic waves. It has been shown that appropriate frequency range that could be used for removing muscle activities is from 0.5 Hz to 50 Hz. In other researches, Discrete Wavelet Transform (DWT) is also very well known for its ability to remove noisy components from EKG signal that could represent for muscle activities. Further documentation about this novel technique can be found in Khawala’s research. After the process of noise removal, a rejection criteria will also be developed to remove the remaining extremely noisy segment because the technique can not absolutely remove all sources of noise. The following steps will be applied to obtain noise free signal that will be used for algorithm development.

1. First, a band-pass filter using FIR design technique, with a cutoff frequency ranges from 0.5 to 50 Hz, filter order of 24 and a rectangle moving window will be designed.
2. The filter is convoluted with the original signal to obtain the first level noise free signal.
3. The signal will then be decomposed using Discrete Wavelet Transform (DWT), with a Sym2 mother wavelet following 1 level of decomposition. The first level detail coefficient (CD) will be chosen to represent the noise from muscle activities.
4. Establishing a threshold for CD with the following criteria
5. The threshold can be chosen as fixed, with some information gathered prior to the signal itself. Under the assumption of white noise, the threshold with variance, as described in Khawala research, can be chosen as follow:
6. Since is often unknown in practice, it is estimated as the medican of the absolute deviation which avoids the influence of the outliner values.

Matlab code presentation:

%-SIGNAL PREPROCESSING---------------------------------------------

%-BASELINE REMOVAL-------------------------------------------------

[approx, detail] = wavelet\_decompose(seg, 11, 'db11');

seg = seg - approx(:,11);

baseline = approx(:,11);

%-NOISE ESTIMATION REMOVAL-----------------------------------------

[approx, detail] = wavelet\_decompose(seg, 1, 'sym2');

noise\_level = detail(:,1);

noise\_variance = 1.483 \* median(noise\_level);

noise\_threshold = noise\_variance \* sqrt(2 \* log(length(noise\_level)));

for hkn = 1:length(noise\_level)

if abs(noise\_level(hkn)) < noise\_threshold

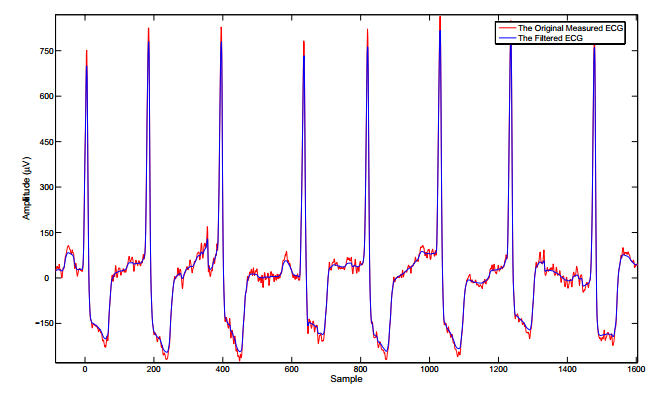
noise\_level(hkn) = 0;

end;

end;

%-DENOISING--------------------------------------------------------

seg = seg - noise\_level;



***Figure 19:*** *Comparison between the noisy signal and filtered signal obtained from the previous de-noising technique*

**2.3 EKG DELINEATION**

The purpose of EKG delineation is capturing the most prominent peaks and event of the characteristic waves, including P wave, QRS complex, R peak, ST segment onset and offset and T wave. The interest of this research is performing R peak detection, T wave detection and ST segment detection by specifying ST onset and offset.

**2.3.1 R PEAK DETECTION**

R peak detection of EKG signal has been the subject of interest for many decades. In other to develop analysis algorithm using EKG signal, it is extremely important that the detection of R peak need to be extremely precise. There are plenty of traditional and novel method for detection of R peak detection, including Pan Tompkins algorithm or Kaman filter, Neural Network and Wavelet based decomposition technique. In this research, a new detection algorithm will be described and the accuracy of the method will be compared with other methods in the final thesis.

1. First, ECG segment will be stored in another variable dedicated for transformation and filtering
2. The segment will be filtered using a band-pass filter, designed using FIR filter design, with a frequency range from 10 Hz to 25 Hz to capture the QRS complex while removing other waveforms
3. The whole segment will be shifted above 0 before squaring, so that the prominent peak obtained is only represented for the R peak because there is no minima that is located below the zero baseline that could becomes prominent after signal squaring. This step is the main different of this technique comparing to the Pan Tompkins algorithm.
4. Instead of squaring, the signal is powered by 12 to maximize the effect of the R peaks on the signal.
5. Next, local maxima will be calculated
6. A moving window of 200 ms will be applied to scan the whole signal. A value is chosen to be an R peak if it qualifies all of the following criteria:
   * It is a local maxima
   * It is larger than the threshold
   * It locates at least 300 ms behind the first highest peaks

With the threshold is chosen as the mean value of the segment in addition to its standard deviation.

1. Then, the location of these values are looked up within the original signal. The amplitude and location will then be stored in the corresponding variables for further analysis. Figure 20 describes the quality of this R peak and T peak detection algorithm.

Matlab code presentation:

%-READ SIGNAL----------------------------------------------------------

data\_length = span \* fs;

startpoint = (inputloop - 1) \* data\_length + 1;

endpoint = inputloop \* data\_length;

seg = sig1(startpoint:endpoint);

% GENERAL PARAMETERS-----------------------------------------------

QRS\_amps = [];

QRS\_locs = [];

QRS\_amps2 = [];

QRS\_locs2 = [];

T\_amps = [];

T\_locs = [];

%-QRS DETECTION----------------------------------------------------

%-Filter 10 - 25Hz to remove other waves---------------------------

filt = fir1(24, [10/(fs/2) 25/(fs/2)],'bandpass');

seg2 = conv(filt,seg);

seg2 = seg2(12:end,1);

%-BRING THE SIGNAL ABOVE 0-----------------------------------------

seg2 = seg2 + abs(min(seg2));

seg2 = seg2.^12;

thres\_mean = (mean(seg2) + QRS\_std\_thres \* std(seg2)) \* ones(1,length(seg2));

[pks, locs] = findpeaks(seg2,'MinPeakDistance',100);

for i = 1:length(locs)

if pks(i) > thres\_mean(1)

ind = locs(i);

QRS\_amps(end + 1) = seg(ind);

QRS\_locs(end + 1) = ind;

end;

end;

**2.3.2 T PEAK DETECTION**

This section represents a simple yet effective technique to locate the T peaks after performing R peak detection.

1. The first step is capturing the isoelectric baseline of the signal. The value of isoelectric baseline is chosen to be the value located at the middle of the RR interval
2. Next step is performing baseline transformation to each RR interval using this criteria

The purpose of this step is to consider only the part of the RR segment where the T peak can take place. The interval is chosen to be between 0.2 and 0.6 of the RR interval.

1. The segment is then subtracted with the isoelectric baseline and then powered by 12 to capture the maximum effect of peaks within this segment. The process of subtracting isoelectric baseline is crucial because T wave can both be traverse or inverse.
2. T wave is then detected as the highest value of this segment

Matlab code presentation:

%-T WAVE DETECTION-----------------------------------------------------

sig = seg;

%-ZEROING EACH BEAT INTERVAL-------------------------------------------

for hk = 1:length(QRS\_locs) - 1

data = sig(QRS\_locs(hk):QRS\_locs(hk + 1));

leng = floor((QRS\_locs(hk + 1) - QRS\_locs(hk)) / 10);

isoelec = seg(QRS\_locs(hk) + 5 \* leng);

data = data - isoelec;

data = abs(data);

data = data.^12;

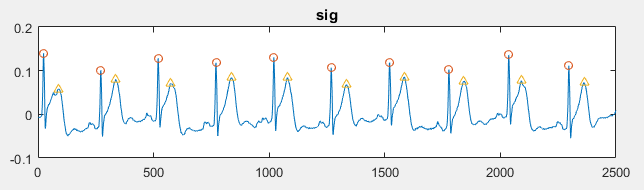
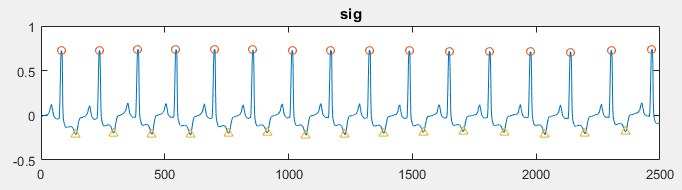
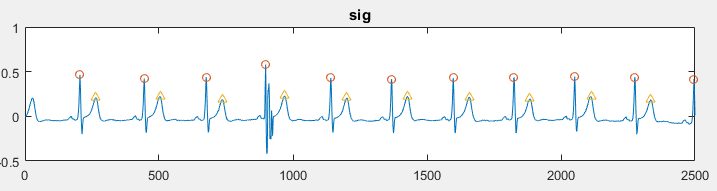
data = data(2 \* leng:6 \* leng);

[val, ind] = max(data);

T\_locs(end + 1) = QRS\_locs(hk) + 2 \* leng + ind;

T\_amps(end + 1) = seg(QRS\_locs(hk) + 2 \* leng + ind);

end;

****

***Figure 20:*** *Result obtained after performing EKG delineation. The location of R peaks and T peaks are correctly determined.*

**2.3.3 REJECTION CRITERIA**

In some cases the signal is extremely noisy that can not be used for algorithm development process. Therefore after the EKG delineation takes place, it is important to reject the segment with unwanted results. The segment will be rejected if one of the following criteria is met:

1. Any R peak and its corresponding T peak is located too far away (larger than 1000 ms) or too close (less than 25 ms)
2. Any consecutive R peaks that is located too far away (larger than 2000 ms) or too close (less than 100 ms)
3. Any consecutive T peaks that is located too far away (larger than 2000 ms) or too close (less than 100 ms)

Matlab code presentation:

% REJECTION CRITERIA-----------------------------------------------

for rjloop = 2:length(QRS\_locs)

condition = QRS\_locs(rjloop) - QRS\_locs(rjloop - 1);

if condition > 450

rejected = rejected + 1;

continue;

end;

end;

for rjloop = 2:length(T\_locs)

condition = T\_locs(rjloop) - T\_locs(rjloop - 1);

if condition > 450

rejected = rejected + 1;

continue;

end;

end;

for rjloop = 1:length(T\_locs)

condition = T\_locs(rjloop) - QRS\_locs(rjloop);

if condition > 250 || condition < 25

rejected = rejected + 1;

continue;

end;

end;

**2.3.4 ST SEGMENT DETECTION**

After detecting R peaks and T peaks, ST segment is described as the interval from 0.4 from 0.65 of the RT interval. ST segment is crucial for calibrating ST deviation and ST slope, which are the most important features that have been tremendously cultured in the field of myocardial injury analysis using EKG signal.

Matlab code presentation:

for km = beat\_start:beat\_end

if ~isnan(QRS\_locs(km)) && ~isnan(T\_locs(km))

leng = floor((T\_locs(km) - QRS\_locs(km))/4);

pheight = (seg(QRS\_locs(km) + floor(2.6 \* leng)) - seg(QRS\_locs(km) + floor(1.6 \* leng))) / seg(QRS\_locs(km)) \* 100;

width = floor(2.6 \* leng) - floor(1.6 \* leng);

STslope(end + 1) = pheight / width \* 10;

Tinv(end + 1) = seg(T\_locs(km));

ToR(end + 1) = abs(seg(T\_locs(km))) / abs(seg(QRS\_locs(km))) \* 100;

ST\_on\_locs(end + 1) = QRS\_locs(km) + floor(1.6 \* leng);

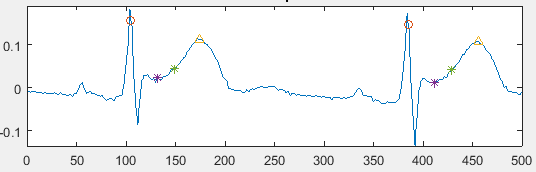
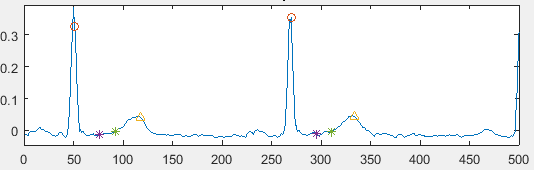
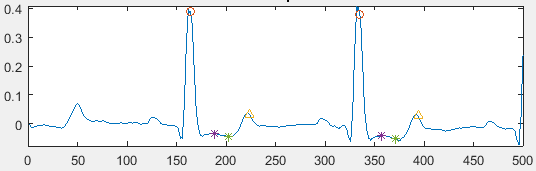
ST\_on\_amps(end + 1) = seg(QRS\_locs(km) + floor(1.6 \* leng));

ST\_off\_locs(end + 1) = QRS\_locs(km) + floor(2.6 \* leng);

ST\_off\_amps(end + 1) = seg(QRS\_locs(km) + floor(2.6 \* leng));

end;

end;



***Figure 21:*** *ST segment detection quality*

**2.4 FEATURES EXTRACTION**

Reminding that features extraction will be performed for each data segment lasting for 10 seconds. Firstly, the following code breaks down the entire signal into different segments, then features extraction is performed for each beat located within each segment. The feature values representing for each segment are taken as the mean of the consecutive value calibrated during each beat. The following code describes the process of calibrating number of loops for each record as well as looping the entire signal length.

Matlab code presentation:

fraction = 1/2;

% CALCULATE SOLOOP--------------------------------------

total\_length = length(sig1);

window\_length = fs \* span;

number\_of\_loop = floor(total\_length \* fraction / window\_length);

for soloop = 1:number\_of\_loop

do something …

end;

**2.4.1 MORPHOLOGICAL FEATURES**

In other to detect abnormalities within EKG signal that represents for myocardial infarction and ischemia, detection of ST deviation and ST slope need to be accomplished. Revise literature review that if there is a transient ST deviation and ST slope up warding, the patient is prescribed with ST elevation myocardial infarction. If ST deviation and ST slope are down warding, the patient is suspected with ST depression myocardial infarction if the mirror leads have transient symptoms of ST deviation and ST slope up warding, or prescribed with Ischemia if such abnormalities do not occur. In case of no transient ST segment elevation or depression, the direction of T wave is considered. T wave is essentially upward in most of EKG leads and often exhibit down warding characteristics or even becomes significantly peaked when ischemic episode occurs. However, recent researches have shown that T wave down warding can be a result from posture changes that does not necessarily indicate cardiovascular injury. Therefore, the clinical value of T wave direction and amplitude is not as sensitive as the ST slope and ST Deviation, but it is also necessary to be calculated.

Therefore, up to 4 morphological features will be calibrated within the scope of this research. These are ST Deviation, ST slope, T direction and T amplitude. The methodology is described below:

1. ST deviation is calculated as the area under the curve of ST segment as derived from section 2.3.4 and the isoelectric baseline. Then the value is normalized with the length of the segment to calculate the empirical height that represent the deviation from the isoelectric line.
2. ST slope is calibrated as the tan value between the line connecting the onset and offset of ST segment against the isoelectric line.
3. T wave amplitude is calculated as the different between the amplitude of T peak with the isoelectric line.
4. T amplitude score is calculated as the division of T amplitude against the R peak amplitude, then scaled to percentage value.

Matlab code presentation:

%-CALDULATE Stslope, Tinv and ToR---------------------------------------------

STslope = [];

for km = beat\_start:beat\_end

if ~isnan(QRS\_locs(km)) && ~isnan(T\_locs(km))

leng = floor((T\_locs(km) - QRS\_locs(km))/4);

pheight = (seg(QRS\_locs(km) + floor(2.6 \* leng)) - seg(QRS\_locs(km) + floor(1.6 \* leng))) / seg(QRS\_locs(km)) \* 100;

width = floor(2.6 \* leng) - floor(1.6 \* leng);

STslope(end + 1) = pheight / width \* 10;

Tinv(end + 1) = seg(T\_locs(km));

ToR(end + 1) = abs(seg(T\_locs(km))) / abs(seg(QRS\_locs(km))) \* 100;

ST\_on\_locs(end + 1) = QRS\_locs(km) + floor(1.6 \* leng);

ST\_on\_amps(end + 1) = seg(QRS\_locs(km) + floor(1.6 \* leng));

ST\_off\_locs(end + 1) = QRS\_locs(km) + floor(2.6 \* leng);

ST\_off\_amps(end + 1) = seg(QRS\_locs(km) + floor(2.6 \* leng));

end;

end;

%-CALDULATE STDeviation------------------------------------------------

for km = beat\_start:beat\_end

if ~isnan(QRS\_locs(km)) && ~isnan(T\_locs(km))

leng = floor((T\_locs(km) - QRS\_locs(km))/4);

pdata = seg((QRS\_locs(km) + floor(1.6 \* leng)):(QRS\_locs(km) + floor(2.6 \* leng)));

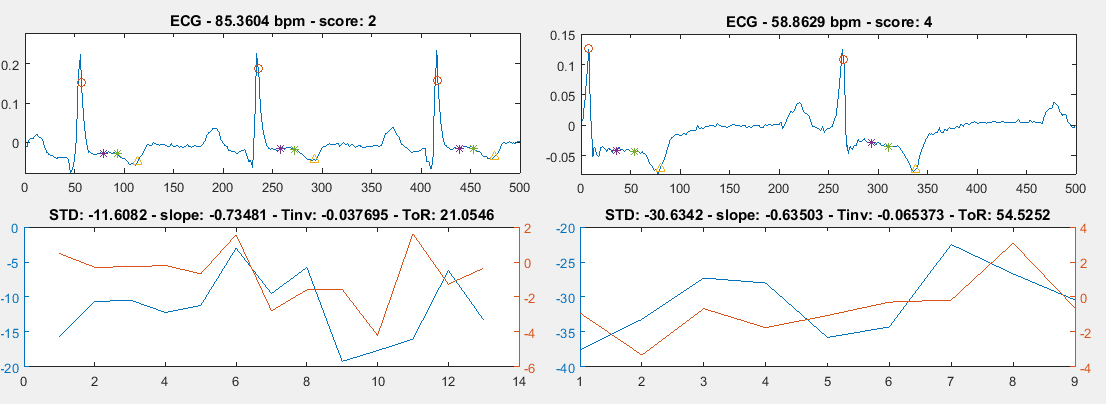
RRinterval = QRS\_locs(km + 1) - QRS\_locs(km);

iso = ones(length(pdata),1) \* seg(QRS\_locs(km) + floor(0.5 \* RRinterval));

STDeviation(end + 1) = (trapz(pdata) - trapz(iso)) / length(pdata) \* 100 \* 10;

end;

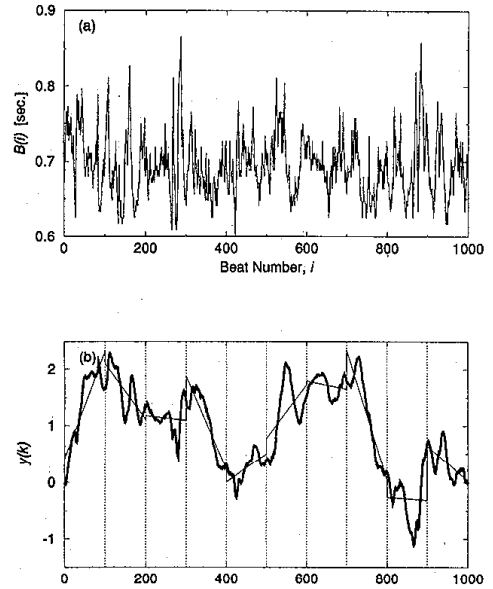
end;

****

***Figure 22:*** *Morphological features obtained after calculation. The consecutive values are presented as a graph and the mean value is displayed in the title of each graph.*

**2.4.2 INTRA - BEAT DE-TRENDED FLUCTUATION**

During many recent studies, it has been shown that De-trended Fluctuation Analysis had demonstrated its potential as a feature for detection of cardiovascular diseases. De-trended Fluctuation Analysis analyzes the chaotic behavior of a system without considering its trend. First, the segment is divided into different box of data. Then a polynomial function is fit into the data in each box, representing the trend. After that, the signal within each box is subtracted with its trend and the degree of fluctuation is calculated as:



***Figure 22****: The data is divided into several boxes and the trend within each box is computed*

The computation is repeated for all of the boxes to provide the relationship between the fluctuation and the box size. A linear relationship on the double log graph indicates the present of scaling and the value of scaling exponent is calculated as the slope of line relating to . According to many researchs, will be higher than 1 for patients who are suffering from coronary artery dieases, while staying mostly below 1 for healthy subjects.

In this research, each RR interval will be fetched into a matlab function to calculate the value of for each beat. Then the DFA feature is defined as the mean value of all consecutive values. The purpose of this practice is to quantify the sensitivity and specificity of DFA in detecting cardiovascular damage, finding the answer for the questions of what type of disease it represent for and how accurate it is in detecting of this disease.

Matlab code presentation:

%-CALCULATE DFA-----------------------------------------------

for i = beat\_start:beat\_end

data = seg(QRS\_locs(i):QRS\_locs(i+1));

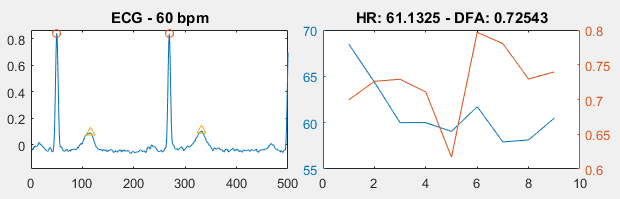
dfa = DetrendedFluctuation(data);

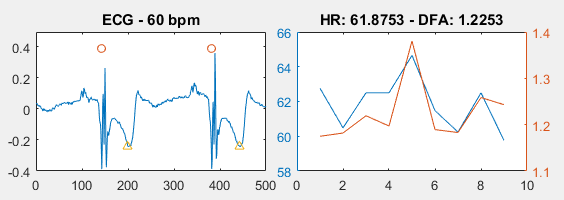
DFA(end + 1) = dfa;

end;

DFA = DFA';

RP\_DFA\_bin = [RP\_DFA\_bin; mean(DFA)];

****

****

***Figure 22:*** *DFA value obtained for each data window of length 10 seconds*

**2.5 SCORING SYSTEM**

Until this part, up to 5 features have been extracted for the analysis of Myocardial infarction and ischemia. Using these 5 features, a scoring system will be developed to address the type and level of injury for each data window. If one of the following criteria is met, the overall score is increased by 1.

1. Transient ST deviation value that is higher than a maximum threshold or lower than a minimum threshold
2. Transient ST slope up warding or down warding value that is higher than a maximum threshold or lower than a minimum threshold
3. Presence of T wave inversion or T wave peaked
4. Presence of DFA value greater than 1

Matab code presentation:

if mean\_STD > 20 || mean\_STD < -20

score = score + 1;

end;

if mean\_STS > 8 || mean\_STS < 0

score = score + 1;

end;

if mean\_Tinv < 0.02

score = score + 1;

end;

%if mean\_HR > 100 || mean\_HR < 50

% score = score + 1;

%end;

if mean\_DFA > 1

score = score + 1;

end;

The threshold for each criteria above is chosen intuitively during this state. In the future work, it will be compared to actual diagnosis from doctors and physician for better qualification. The process will be covered in the final thesis. The following code provides extra general clinical prescription for the data segment.

%-MAKING THE DIAGNOSIS-------------------------------------------------

if mean\_STD > 100 || mean\_STS > 11

diagnosis = 'Transient ST Elevate';

elseif mean\_STD < -40 && mean\_STS < -4

diagnosis = 'Transient ST Depress';

elseif mean\_Tinv < -0

diagnosis = 'T wave inverted';

elseif mean\_Tinv < 0.02

diagnosis = 'T wave absence';

elseif mean\_DFA > 1 && mean\_STD > 20

diagnosis = 'Minor positive STD';

elseif mean\_DFA > 1 && mean\_STD < -10

diagnosis = 'Minor negative STD';

elseif mean\_STD > 20

diagnosis = 'Minor positive STD without DFA';

elseif mean\_STD < -10

diagnosis = 'Minor negative STD without DFA';

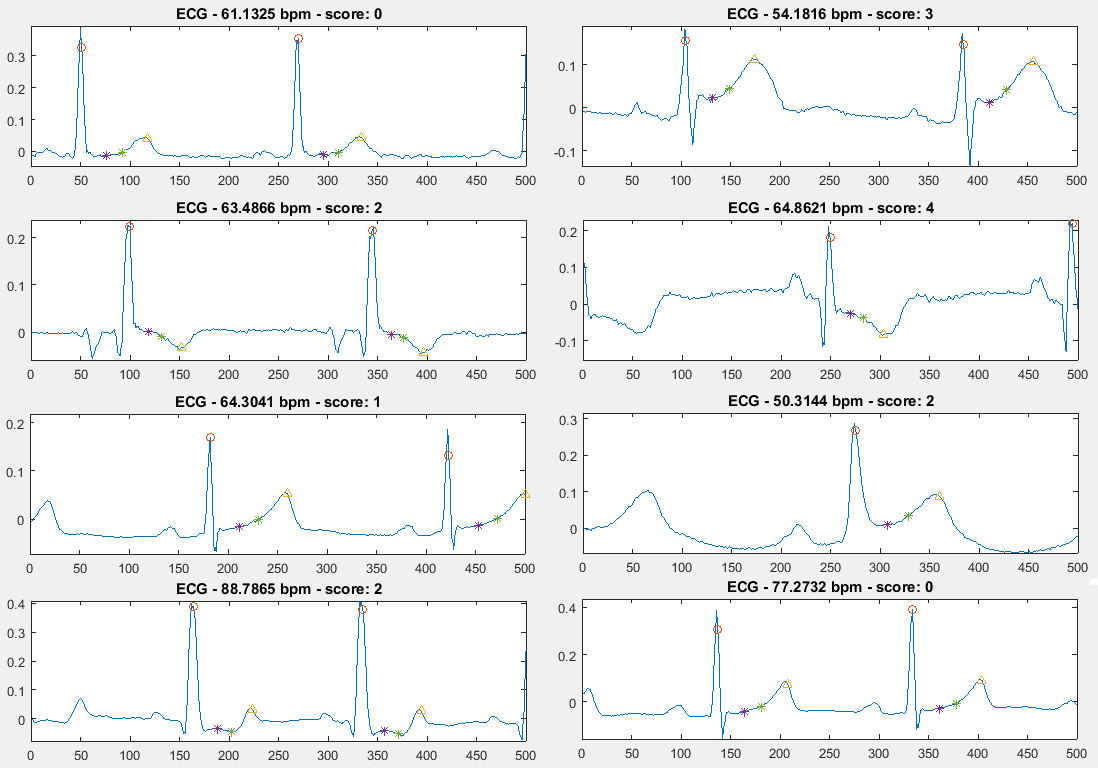
elseif mean\_DFA > 1

diagnosis = 'STD spotted by DFA without MF';

else

diagnosis = 'Normal ECG';

end;



***Figure 23****: Scoring system creates the risk score for each of the data segment*

RESULT AND VALIDATION

In this section, the results obtained from performing features extraction and the potential quality of the risk score system in diagnosing cardiovascular damage will be covered.

**3.1 FEATURES EXTRACTION SUMMARY**

After performing feature extraction for each window length of 10 seconds, a package of total 5 features are extracted: ST deviation, ST slope, T wave direction T wave amplitude and DFA score. The process is repeated for all 94 records described in section 2.1, each record contains the data that last for approximately 60 minutes. In total, a table of 31,000+ rows are automatically computed to provide the database for validation and further analysis. The table below summarize the descriptive statistics of this database:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SCORE** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Normal | 7894 | 24.9 | 24.9 | 24.9 |
| Caution | 7710 | 24.3 | 24.3 | 49.3 |
| Risk | 6796 | 21.5 | 21.5 | 70.7 |
| Danger 1 | 7152 | 22.6 | 22.6 | 93.3 |
| Danger 2 | 2119 | 6.7 | 6.7 | 100.0 |
| Total | 31671 | 100.0 | 100.0 |  |

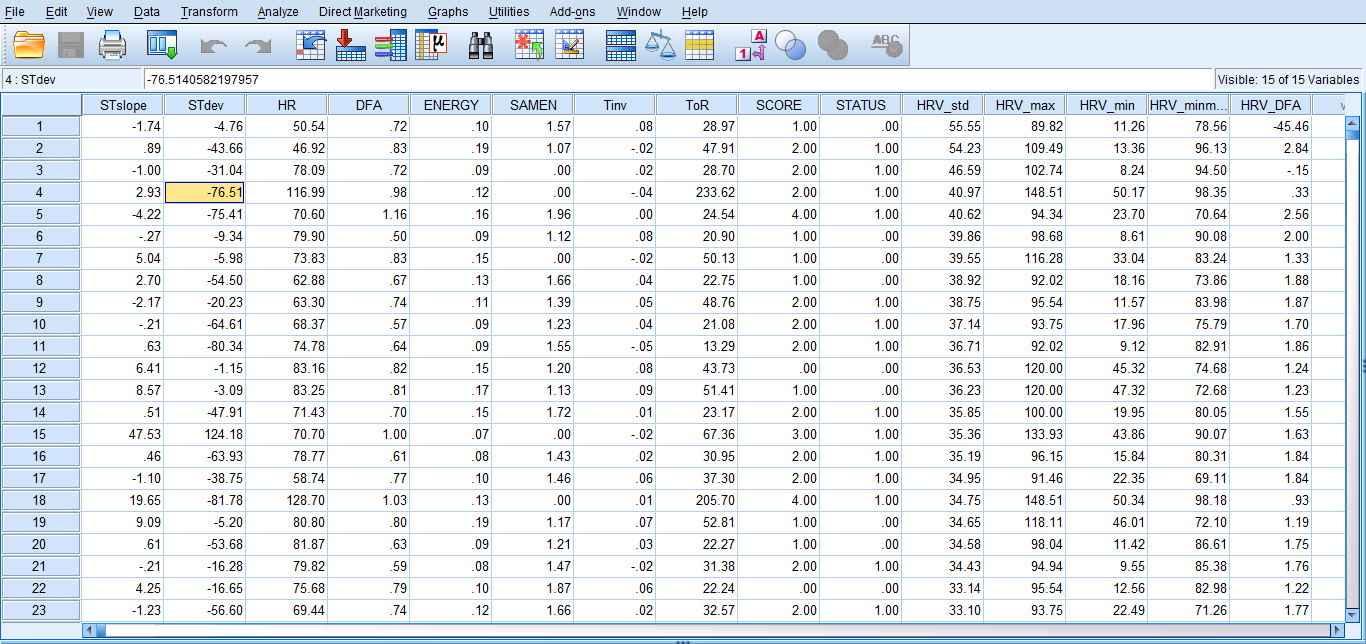
***Figure 24:*** *Descriptive statistic for the SCORE system, with score = 0 indicates Normal, score = 1 indicates Caution, score = 2 indicates Risk and score > 3 indicates Danger.*

As we can clearly see from the descriptive statistic table, the number of cases among the groups are approximately similar, each accompanies about 22% of the total cases. This is the result of careful and manual selection of different types of diseases observed within the first few minutes of the record. The final class (Danger 2), however, lacks the appropriate number of cases comparing to other classes. This class represent for the myocardial ischemia disease because it represents the highest possible score where ST deviation, ST slope downward and T wave inversion happen at the same time. The records representing for this disease were carefully chosen, however, during the period of 1 hour calculation, the EKG signal may have altered and the obtained is the increase in other classes while the number of cases for this classes is not met. Further effort will be made in order to equalize the number of cases within each classes for the further researches.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **STATUS** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Normal | 15604 | 49.3 | 49.3 | 49.3 |
| ST Deviation | 16067 | 50.7 | 50.7 | 100.0 |
| Total | 31671 | 100.0 | 100.0 |  |

***Figure 25:*** *Descriptive statistic for the STATUS variable, in which ‘’Normal’’ subject has the score of less than 2 and the ‘’ST Deviated’’ subject has the otherwise.*

It is always a good practice to create a database with appropriately equal number of cases between each class to avoid the effect of over-fitting and under-fitting.



***Figure 26:*** *The analysis is carried on SPSS statistics. The table above contains in total 31,761 rows of data*

**3.2 ACCURACY VALIDATION**

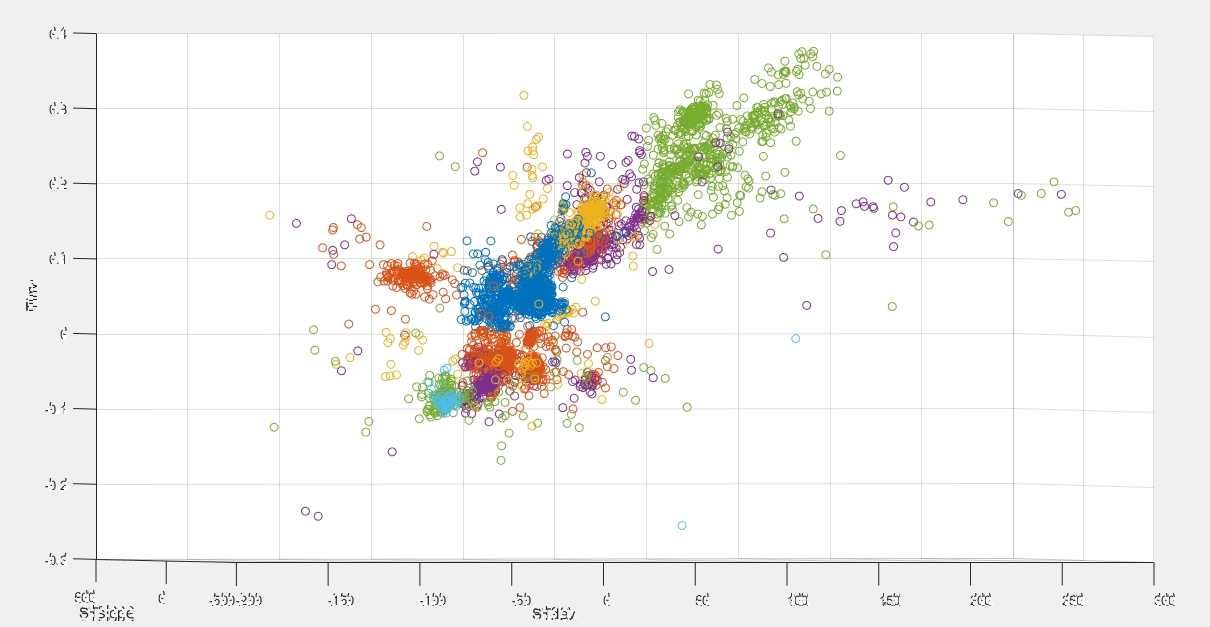
**3.2.1 MORPHOLOGICAL FEATURES**

Any abnormalities observed within a morphological feature is defined as the transcendence of this value over a defined threshold. At this state, the threshold is intuitively specified according to the author’s knowledge about EKG clinical interpretation. In future work, this process will be qualified by doctors and clinician for better quantification of the thresholds for each features. However, it is observable that the risk score value calculated from morphological features is currently appropriate in both describing and classifying cardiovascular diseases. The following table describes the theory of classification.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Score | Description | Disease types |
| Valid | Normal | 0 | Normal EKG | Healthy |
| Caution | 1 | Small ST deviation or T inversion | Postures changes or anxiety |
| Risk | 2 | Transient ST deviation with DFA confirmation or with T inversion | Suspected of myocardial injury or ischemia |
| Danger 1 | 3 | Transient ST deviation, ST slope with DFA confirmation | Diagnosis with ST elevation myocardial infarction |
| Danger 2 | 4 | Transient ST deviation, ST slope with DFA confirmation and T wave inversion | Diagnosis with ST depression myocardial infarction |

***Figure 27****: Table summary of disease classification with different risk scores*

Classification according to the above table successfully separate each cases in term of ST deviation, ST slope and T wave inversion. The figure below describes the distribution of all cases in term of ST deviation, ST slope and T wave inversion previously stated against its class.

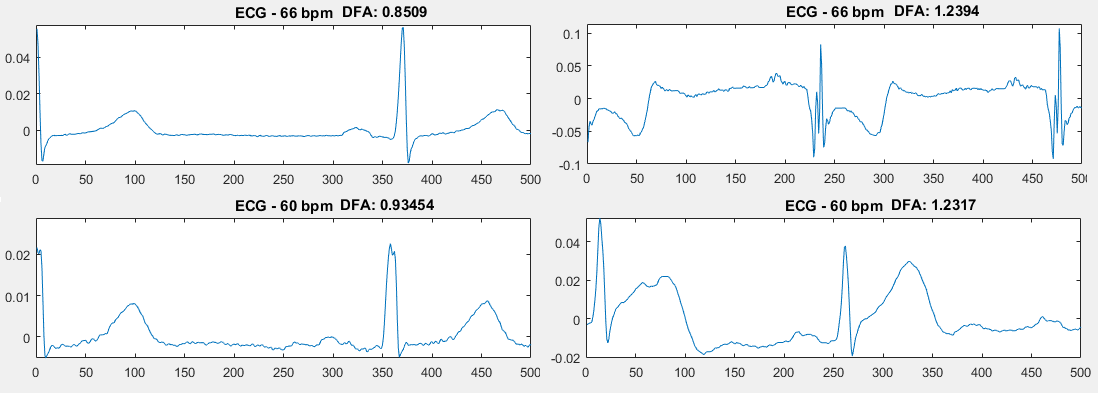


***Figure 28****: Quality of classification basing on risk score system*

From the above figure, it can be concluded that the classifier successfully clusters the group of transient ST deviation confirmed with DFA (green color) and the normal control (blue color). The other classes need further improvement.

**3.2.2 DFA FEATURES**

From the study, it is clearly observed that DFA score is significantly higher than 1 when transient ST elevation or depression occurs as seen in the following figures.



***Figure 29:*** *Quality of using DFA score as a diagnostic value for transient ST segment deviation*

In order to quantify the sensitivity and specificity for this theory, the following matlab code is applied into the database. For each data entry, if the DFA is greater than 1 and the score is higher than 3 indicating transient ST deviation, it is considered as a true calibration. The total number of cases of true calibration is divided for the total number of cases of DFA which is greater than 1 to obtain the sensitivity. The same technique is applied to calculate the specificity.

%-CALCULATE DFA SENSITIVITY-----------------------------------

for i = 1:length(aaaa)

if aaaa(i, 4) > 1 % DFA > 1

total\_DFA\_1 = total\_DFA\_1 + 1;

if aaaa(i, 10) > 0 % STATUS = 1

accurate\_DFA\_1 = accurate\_DFA\_1 + 1;

end;

end;

end;

sensitivity\_DFA = accurate\_DFA\_1 / total\_DFA\_1;

%-CALCULATE DFA SPECIFICITY-----------------------------------

for i = 1:length(aaaa)

if aaaa(i, 4) < 1 % DFA < 1

total\_DFA\_0 = total\_DFA\_0 + 1;

if aaaa(i, 10) < 1 % STATUS = 0

accurate\_DFA\_0 = accurate\_DFA\_0 + 1;

end;

end;

end;

specificity\_DFA = accurate\_DFA\_0 / total\_DFA\_0;

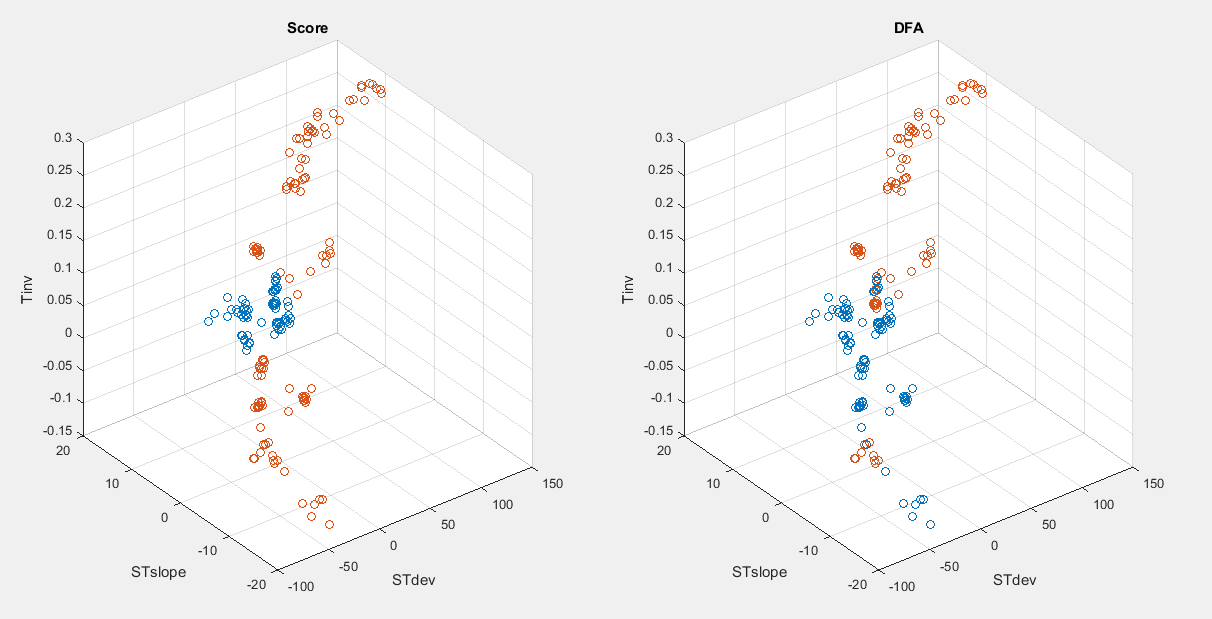
disp(['DFA sensitivity: ' num2str(sensitivity\_DFA)]);

disp(['DFA specificity: ' num2str(specificity\_DFA)]);

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Number of cases | Sensitivity | Specificity |
| DB | European | 12523 | 0.8594 | 0.6539 |
| Long ST | 15155 | 0.9166 | 0.7022 |
| ST changes | 4263 | 0.8108 | 0.5395 |

***Figure 30:*** *Sensitivity and Specificity of using DFA to detect transient ST deviation within different database systems*

From the table above, it is quite clear that DFA is a potentially good diagnostic value for detection of transient ST deviation that might represent for acute myocardial infarction or ischemia. The sensitivity is considerably high with the greatest value of 0.9166 and the smallest value achieved is also greater than 0.8. However, the sensitivity is not as diagnostically valuable as the sensitivity. Therefore, it is also very important to consider other morphological features when performing diagnosis for achieving better results. The sensitivity of DFA, in other hand, yields as a potentially better candidate for fast diagnosis of transient ST deviation than the traditional calibration for ST slope and ST deviation. The main reasons are that this method is faster to compute and less subjected to noises. Further analysis about this topic will be discussed later in the final thesis.



***Figure 31****: Scatter plot of 180 data entries computed from the European database.*

On the left side of the figure lies the scatter plot representing for patients with and without transient myocardial disease, with red indicates presence of disease and blue indicates absence of disease. On the right side is the scatter plot representing for those same cases but in term of DFA cluster. If the DFA is greater than 1, the point is red and if DFA is less than 1, the point is blue. The figure shows that every red point on the right graph is also a red point in the left graph, indicating high sensitivity. The computed sensitivity for the above calibration is 0.84337.

CONCLUSION

This research has covered the technique to develop an algorithm for quantification and detection of myocardial damage using EKG signal. The output is a risk score system that is capable of detecting ST segment abnormalities that manifest the presence of ST elevation and ST depression myocardial infarction. In this research, it has been found that for patients with transient ST deviation, the risk score is generally higher than 2. For patients with a risk score of 0, it is clearly shown that these patients have no symptoms of cardiovascular damage. Classification of disease types can be found in the table below as a summary for this research.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Score | Description | Disease types |
| Valid | Normal | 0 | Normal EKG | Healthy |
| Caution | 1 | Small ST deviation or T inversion | Postures changes or anxiety |
| Risk | 2 | Transient ST deviation with DFA confirmation or with T inversion | Suspected of myocardial injury or ischemia |
| Danger 1 | 3 | Transient ST deviation, ST slope with DFA confirmation | Diagnosis with ST elevation myocardial infarction |
| Danger 2 | 4 | Transient ST deviation, ST slope with DFA confirmation and T wave inversion | Diagnosis with ST depression myocardial infarction |

***Figure 27****: Table summary of disease classification with different risk scores*

Not only that, De-trended Fluctuation Analysis (DFA) demonstrates as a potentially important technique to detect transient ST deviation within EKG signal. It is found that for patients exhibiting transient ST deviation due to cardiovascular damage, the DFA value is higher than 1. The sensitivity of this theory is calculated and demonstrated good result. However, the specificity or the inverse statement is not high. Further improvement need to be made in other to obtain better result.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Number of cases | Sensitivity | Specificity |
| DB | European | 12523 | 0.8594 | 0.6539 |
| Long ST | 15155 | 0.9166 | 0.7022 |
| ST changes | 4263 | 0.8108 | 0.5395 |

***Figure 30:*** *Sensitivity and Specificity of using DFA to detect transient ST deviation within different database system**.*

REFERENCES

1. Chatfield, C., *Time-series forecasting*. 2000: CRC Press.

2. Shumway, R.H. and D.S. Stoffer, *Time series analysis and its applications*. 2013: Springer Science & Business Media.

3. Hamilton, J.D., *Time series analysis*. Vol. 2. 1994: Princeton university press Princeton.

4. Hamilton, J.D., *A new approach to the economic analysis of nonstationary time series and the business cycle.* Econometrica: Journal of the Econometric Society, 1989: p. 357-384.

5. Park, D.C., et al., *Electric load forecasting using an artificial neural network.* Power Systems, IEEE Transactions on, 1991. **6**(2): p. 442-449.

6. Taylor, J.W., P.E. McSharry, and R. Buizza, *Wind power density forecasting using ensemble predictions and time series models.* Energy Conversion, IEEE Transactions on, 2009. **24**(3): p. 775-782.

7. Reis, B.Y. and K.D. Mandl, *Time series modeling for syndromic surveillance.* BMC Medical Informatics and Decision Making, 2003. **3**(1): p. 2.

8. Soni, J., et al., *Predictive data mining for medical diagnosis: An overview of heart disease prediction.* International Journal of Computer Applications, 2011. **17**(8): p. 43-48.

9. Getzen, T., *Forecasting health expenditures: short, medium and long (long) term.* Journal of Health Care Finance, 2000. **26**(3): p. 56-72.

10. Kirkwood, B.R., *Essentials of medical statistics*. 1988: Blackwell Scientific Publications.

11. Knaus, W.A., et al., *The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults.* Chest Journal, 1991. **100**(6): p. 1619-1636.

12. Rünstler, G., et al., *Short‐term forecasting of GDP using large datasets: a pseudo real‐time forecast evaluation exercise.* Journal of forecasting, 2009. **28**(7): p. 595-611.

13. Armstrong, J.S., *Long-range forecasting*. 1985: Wiley New York ETC.

14. Cohen, M.A. and J.A. Taylor, *Short‐term cardiovascular oscillations in man: measuring and modelling the physiologies.* The Journal of physiology, 2002. **542**(3): p. 669-683.

15. Box, G., G. Jenkins, and G. Reinsel, *Time series analysis: Forecasting and control. 3rd Prentice Hall.* Englewood Cliffs, NJ, 1994.

16. Christini, D.J., et al., *Application of linear and nonlinear time series modeling to heart rate dynamics analysis.* Biomedical Engineering, IEEE Transactions on, 1995. **42**(4): p. 411-415.

17. Fan, J. and I. Gijbels, *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*. Vol. 66. 1996: CRC Press.

18. Tong, H., *Non-linear time series: a dynamical system approach*. 1990.

19. Mallat, S., G. Papanicolaou, and Z. Zhang, *Adaptive covariance estimation of locally stationary processes.* Annals of Statistics, 1998: p. 1-47.

20. Esteghamatian, M., et al., *Real time cardiac image registration during respiration: a time series prediction approach.* Journal of real-time image processing, 2013. **8**(2): p. 179-191.

21. Chung, D., et al. *Real-time registration by tracking for MR-guided cardiac interventions*. in *Medical Imaging*. 2006. International Society for Optics and Photonics.