ECG Signal Based Angina Pectoris Detection Using Statistical and Machine Learning Approach

Md. Merajul Islam, Md. Asadur Rahman

Abstract— Angina pectoris due to ischemia is very crucial to detect because, if it can be detected earlier, doctor can provide the patient proper medication to cure. Sometimes from a long ECG data it is tiresome to calculate and differentiate the normal and angina pectoris affected ECG peak to decide about the condition of the patient. In addition, the remote areas of lower and mid-lower income countries often face lack of experienced doctors or highly cost devices like eco-cardiogram, MRI to detect angina. In these regions, automatic identification of ECG features can be a fruitful solution. Therefore, only way is computerized and efficient ECG analyzing algorithm development that can be able to detect angina pectoris. In this work, an efficient algorithm is developed to detect angina pectoris from ECG signal. This algorithm consists of several steps to take decision on ECG signal. First of all this algorithm removes baseline wandering from ECG signal by baseline wandering path finding algorithm. After that it removes other noises from ECG signal by Gaussian weighted moving average window method. In this consequence, QRS complex was detected by very well-known method First and Second Derivative (FS2) algorithm and gradually other important points like S, J, K, and T were detected by possible range maxima-minima criterion. Besides, the isoelectric line of ECG signal is estimated and eventually the statistical features of J-K points of normal and abnormal ECG peaks are compared with that isoelectric line by the algorithm. Finally, this algorithm takes the decision whether the patient is suffering from angina or not. This algorithm is applied on MIT arrhythmia Database to detect angina pectoris. From the result provided by this algorithm, we have found 94% (average) accuracy which is noticeable. In addition with that the sensitivity and specificity of our proposed algorithm have also been calculated which are found 91% and 89%, respectively. Since the previous work is based on the single feature, it may prove inappropriate for all the time. Therefore with the help of multiple features machine learning based approach k-nearest neighbor (kNN) method has been deployed in this research work to make it more accurate and acceptable. Although kNN based prediction method also provide almost similar results found by the previous methodologies. Therefore our proposed approach for angina pectoris detection has been testified by both statistical and kNN based method. It is expected that the proposed algorithm will be helpful for computerized angina pectoris detection from ECG signals.

Index Terms— Electroencephalogram (EEG), Angina Pectoris, Ischemia, Peak Detection, S-T Segment, FS2 Algorithm, k-Nearest Neighbor (kNN), MIT Arrhythmia Database.

1 INTRODUCTION

Heart diseases are the leading causes to death in the world. The most common for heart disease is myocardial ischemia which is one of the most serious & prevalent heart diseases. According to the report of WHO 2015 [1], ischemic heart disease and stroke are the world’s biggest killers which leads to 15 million deaths (54%) in 2015. Myocardial ischemia is very common cardiac disease. Electrocardiography or ECG signal is a pattern of graphical representation of electrical activity of cardiac system and it is often used to measure the various cardiac diseases and abnormalities of the heart function. The graphical representation of ECG signal consists of P, QRS complex, T wave which is given in Fig. 1. In this figure, the most important and commonly claimed peaks are given with their standard waveshape by a cycle of ECG beat. In medical science, each waveshape presents unique significant information about the functionality of the heart. As a result, a number of risk predictions of the heart can be assessable from this ECG signal. These waves correspond to the fields induced by specific electric phenomenon on the cardiac surface. Atrial depolarization produces the P wave, ventricular depolarization produces QRS complex and ventricular repolarization produces T wave [2]. Angina pectoris, commonly known as angina, is severe chest pain due to ischemia of the heart muscle, generally due to obstruction or spasm of the coronary arteries. The medical science broadly classifies the angina pectoris by two kinds: i) stable angina and ii) unstable angina. Stable angina occurs during exertion, can be quickly relieved by resting or taking nitroglycerin, and lasts from three to twenty minutes. On the other hand, unstable angina increases the risk of a heart attack, occurs more frequently, lasts longer, is more severe, and may cause discomfort during rest or light exertion which may be a serious indicator of an impending heart attack.

The important and common feature of ischemia in the cellular level is the depolarization of the cellular resting membrane potential. As large the areas of the heart muscle become ischemic, its relaxation and contraction patterns are affected, which cause variations in the ST level and T-wave in ECG. As a result, early detection of heart disease through ECG signal has a vital significance in the medical science which is demanding challenge for the Biomedical Engineers. ECG could be a potential signal of heart that can be applied to find the computerized angina pectoris detection.

Md. Merajul Islam has just completed his M. Sc. Eng. degree program in the Department of Biomedical Engineering in Khulna University of Engineering & Technology (KUET), Khulna-9203, Bangladesh. E-mail: snigdoshokale@gmail.com

Md. Asadur Rahman is currently pursuing his PhD degree in the Department of Biomedical Engineering in Khulna University of Engineering & Technology (KUET), Khulna-9203, Bangladesh. E-mail: bmeasadur@gmail.com

(Both authors contributed equally regarding this research work)
People with an average age of 62 years, who have moderate to severe degrees of angina, have a 5-year survival rate of approximately 92% [3]. It is very crucial to detect ischemia because, the effects of myocardial ischemia can be almost reversible if it can be detected early enough [4]. Therefore, the importance of identification of angina pectoris is significant. There are various invasive and non-invasive modalities for detecting angina pectoris. In recent days [5] the researchers suggested various invasive costly modalities for detecting angina pectoris such as Angiography, Electro-beam CT, multi-detector row CT, Coronary Angiography, Cardiac MRI, Echocardiography, etc. These modalities are very costly which quite impossible for the low or lower-middle income country to offer this service for their remote area patients. The patients of the remote areas of those countries should have some alternative solution for this problem. Beyond the clinically applicable highly costly device, ECG could be an alternative solution for angina detection because due to ischemia ECG features are also significantly changed, mostly the S-T segments.

Though specificity and sensitivity for detecting cardiac ischemia of aforementioned modalities are comparatively high but these modalities are costly for patients in third world like Bangladesh. Globally, an estimated 54 million people have angina, 16 million of whom are from the South-East Asia region. These modalities are unavailable in most of the remote places for diagnosis of angina pectoris. Electrocardiography is the non-invasive modalities for detecting angina pectoris. It is comparatively cheap and available in remote places. In most of the cases it is observed that, doctors of diagnostic center cannot detect the angina pectoris by eyes from long ECG signal due to lack of experiences.

There are several research articles [6]-[11], those proposed methods to detect angina pectoris from ECG signal. These methods are mostly based on machine learning based [6]-[10] or wavelet based [11]. The research work proposed in [6] used PCA based feature extraction method to reduce the dimensionality reduction. This method shows lower classification accuracy. On the other hand, the research work in [7] cannot propose a real computer assisted algorithm for the real classification of ECG beats into angina affected and not affected classes. In [8], authors discussed about the classification optimization of the myocardial affected ECG signal regarding different promising classifiers. Although a deep learning based a complex artificial neural network has been proposed in [9], the accuracy level is lower instead it would be. On the other hand, an overall prediction on ECG signal to find out whether there is any abnormality exists or not has been proposed with the help of artificial neural network in [10]. This article avoids taking decision on beat based feature of ECG signal. A wavelet based feature extraction method proposed in [11] claiming itself an approximately prediction algorithm considering the overall features of ECG signal.

Eventually, from the wide literature review based on the previous works we can reach to an argument that these methods can detect whether the ECG of the patient is angina pectoris affected or not. The dissimilarity between short time angina and chronic angina cannot be differentiated by these methods. These methods cannot count how many affected ECG beats and how many normal ECG are existed in the ECG signal. But these are important for angina pectoris detection. To reach the each peak and their corresponding features should have to explore to take proper decision against the patient. Therefore, the previously mentioned methods can be proved false due to its assumption and there is no scope for a doctor to take decision on the result of the outcomes of these methods. In addition, these methods need to train with a huge number of normal and abnormal ECG data which is also a drawback of these proposed systems.

From this view point, there arises a scope to develop an algorithm to detect angina pectoris from ECG signal which will be able to diagnosis of angina pectoris. In this work, an algorithm based on statistical feature is proposed including a number of preprocessing steps like baseline wandering removal, power noise elimination, peak findings, isoelectric line estimation, comparing, etc. Based on a number of features of S-T segment of the each beat of the ECG signals, a predictive model by machine learning algorithm comparing suitable features of the signal beats can be implemented to count how many abnormal ECG beat are there in the testing signal.

The objective of this project is to develop an efficient algorithm that can be able to count the abnormal and normal ECG peaks which leads to decide whether the patient has angina or not. The specific objectives of this research work are summarized below:

- Collecting necessary ECG data related to angina pectoris and processing the raw signal to remove noises.
- Finding the different peaks for feature extraction
- Extracting Features for effective analysis and classification.
- Classifying whether the ECG data contains angina or not using efficient classifier.
- Investigating the performance and accuracy of the proposed system.

The rest of the paper is organized as follows: Section 2 presents the Literature Review and the Proposed Methodology has been given in Section 3. We have presented our experimental results with proper discussion in Section 4. The total research work is briefly concluded in Section 5 referred as Conclusions.
2 Literature Review

2.1 Mechanical and Electrical Properties of Heart

The heart is a four chambered pump. The top two chambers are called Atria, the bottom two are called Ventricles. They are separated from top to bottom by valves; the right and left sides are separated by a septum as given in Fig. 2. The heart is a muscle and functions primarily as a double-sided pump. The left side of the heart pumps oxygen rich blood to all parts of the body, while the right side of the heart pumps blood back to the lungs to pick up more oxygen. For this pump to operate, it needs a source of power or energy, in this case electricity. The heart consists of four chambers: two upper chambers called the left and the right atria and two lower chambers known as the right and left ventricles. Between these upper and lower chambers are valves which open and close to direct the flow of blood. The left ventricle performs the most work and is the strongest of the four chambers. It ejects blood into the aorta: the main pipeline which comes out of the heart supplying oxygenated blood to the entire body [13].

During a normal heartbeat, an electrical impulse originates in the right atrium where the main pacemaker of the heart is located. This natural pacemaker is called the sinoatrial node (SA node). The impulse then travels to the left atrium and down the interatrial septum to the atrioventricular node (AV node). This is the secondary pacemaker to the heart. The AV node slows down these impulses while they continue travelling down a common pathway branching off into the right and left bundle-branches and eventually to the ventricles. This cycle is known as normal sinus rhythm (NSR), which describes a well synchronized pumping action between the upper and the lower chambers producing the familiar (lub-dub) heart sounds [14].

2.2 Angina Pectoris

The heart muscle is a special type of muscle which has property of both voluntary and involuntary muscles. It contracts regularly and continuously to pump and circulate blood to the body and the lungs. The pumping action is caused by a flow of electricity through the heart which continuously repeats with a cyclic rhythm. If this electrical activity is disrupted due to any kind of disturbance, it affects in the rhythm of heart which is called arrhythmia. Arrhythmia can affect the ability of heart to pump properly. Therefore, an arrhythmia is an irregular heartbeat or loss of rhythm of the heart. During an arrhythmia the heartbeat may change regularity such as beat too fast, too slow or contract too early [15].

Ischemia or Myocardial ischemia is a very common disease for cardiac patient. Generally it happens due to the lack of sufficient blood flow in specific area of the heart named by contractile cells and this effect may lead to myocardial infarction. The most common cause of cardiac ischemia is plaque buildup in the arteries due to the long term effects of coronary artery disease. The development of plaque within the coronary artery that blocks more than 70% of the lumen of the vessel can cause symptoms of myocardial ischemia [16]. It is considered that this may be the first instance where the subject begins to experience effects of the suboptimal operation of the heart due to insufficient blood supply.

Myocardial ischemia, also called cardiac ischemia, can damage your heart muscle, reducing its ability to pump efficiently. A sudden, severe blockage of a coronary artery can lead to a heart attack. Myocardial ischemia might also cause serious abnormal heart rhythms [17]. In Fig. 3, a graphical representation is shown to clarify how the blood flow is blocked which causes insufficient oxygen flow.

Angina pectoris, commonly known as angina, is usually severe chest pain or discomfort due to coronary heart disease. It occurs when the heart muscle becomes unable to get as much blood as it actually needs for normal functioning [18]. This usually happens because one or more of the heart’s arteries become narrowed or blocked (as given in Fig. 3). This phenomenon is also called ischemia. The symptom of angina often occurs when the heart muscle needs more blood than it is getting. During the times of physical activities or strong emotions, generally heart rates are increased while demand of oxygen by the heart muscle is increased as well. Severely narrowed arteries usually fail to allow enough blood to reach the heart and due to this consequence a patient feel uncomfortable pressure, fullness, squeezing or pain in the center of the chest. It may also feel the discomfort in neck, jaw, shoulder, back or arm of a patient.

There are several different kinds of angina those are ex-
plained here. Generally, there are five different kinds of angina have been identified [20], with the two most common being stable angina and unstable angina.

**Stable Angina:** Stable angina occurs when the heart has to work harder than normal, during exercise, for example. It has a regular pattern, and if you already know that you have stable angina, you will be able to predict the pattern. Once you stop exercising, or take medication (usually nitroglycerin) the pain goes away, usually within a few minutes.

**Unstable Angina:** It is more serious, and may be a sign that a heart attack could happen soon. There is no predictable pattern to this kind of angina; it can just as easily occur during exercise as it can while you are resting. It should always be treated as an emergency. People with unstable angina are at increased risk for heart attacks, cardiac arrest, or severe cardiac arrhythmias (irregular heartbeat or abnormal heart rhythm).

**Less common kinds of angina:** It includes variant angina, microvascular angina, and atypical angina. **Variant angina** is also known as Prinzmetal’s angina. It often occurs while someone is resting (usually between midnight and 8:00 in the morning), and it has no predictable pattern—that is, it is not brought on by exercise or emotion. This kind of angina may cause severe pain, and is usually the result of a spasm in a coronary artery. **Microvascular angina**—sometimes referred to as Syndrome X—occurs when tiny vessels in the heart become narrow and stop functioning properly, even if the bigger arteries are not blocked by plaque. Usually it is treated with common angina medications. **Atypical angina** often doesn’t cause pain, but you may feel a vague discomfort in your chest, experience shortness of breath, feel tired or nauseous, have indigestion, or pain in your back or neck. Women are more likely than men to have feelings of vague chest discomfort [21].

### 2.3 Normal ECG waves and its features

Atrial and ventricular depolarization and repolarization are represented on the ECG as a series of waves: the P wave followed by the QRS complex and the T wave.

**P Wave:** The first deflection is the P wave associated with right and left atrial depolarization. Wave of atrial repolarization is invisible because of low amplitude. Normal P wave is no more than 2.5 mm (two-and-a-half 1-mm-divisions) tall and less than 120 ms (three 1-mm-divisions) in width in any lead. In sinus rhythm when the SA node is the pacemaker, the mean direction of atrial depolarization (the P wave axis) points downward and to the left, in the general direction of lead II within a coordinate between $15^\circ$ and $75^\circ$ and away from lead aVR.

**QRS Complex:** The QRS complex is the largest voltage deflection of approximately 10–20 mV but may vary in size depending on age, and gender. The voltage amplitude of QRS complex also gives information about the cardiac disease [22]. Duration of the QRS complex indicates the time for the ventricles to depolarize and may give information about conduction problems in the ventricles such as bundle branch block. Generally a normal ECG beat gives a QRS deflection as Fig. 4(a) but in the abnormal case ac ECG bit may give a QR deflection as Fig. 4(b).

**T Wave:** The T wave represents the current of rapid phase 3 ventricular repolarization (see diagram above). The polarity of this wave normally follows that of the main QRS deflection in any lead. The ventricles are electrically unstable during that period of repolarization extending from the peak of the T wave to its initial downslope as given in Fig. 5. A stimulus (e.g. a run away heart beat called a premature beat) falling on this vulnerable period has the potential to precipitate ventricular fibrillation: the so call R-on-T phenomenon.

**ST Segment:** Following the QRS complex is the ST segment, extending from where the QRS ends (irrespective of what the last wave in the complex is) to where the T wave begins. The junction between the end of the QRS and the beginning of the ST segment is called the J point indicated in Fig. 6. ST segment reflects the current flow associated with phase 2 of ventricular repolarization. Since there is no current flow during this plateau phase of repolarization, the ST segment is normally isoelectric with the baseline.
Therefore in the absence of any cardiac pathology, the end of depolarization and the beginning of repolarization are normally isoelectric; this region is called the ST segment. On the ECG signal, the ST segment is defined as the region between the ends of the S-wave, also called the J-point, and the beginning of the T-wave. ST Segment is shown in Fig. 6 and Fig. 7. The current standard of determining the ST segment measurement is by measuring the voltage difference between the value at a point 60-80 milliseconds after the J-point and the isoelectric baseline. The isoelectric baseline is either between the P- and Q-waves (the P-R interval) or in front of the P-wave (the T-P interval) which shown in Fig. 7.

2.4 ECG and Its Characteristics during Angina
The pain of angina generally begins beneath the sternum as a vague pain. It may proceed to a severe, crushing, or intense pain that may radiate to the left shoulder, down the arm, to the back, throat, or even the jaws and teeth. Typical angina is often associated with atherosclerosis; with the accompanying plaque formation in the coronary vessel, occludes blood flow to the myocardium. ECG changes associated with typical angina include ST depression. Two types of effects are found in ECG signals and those are ST elevation and ST depression which is shown in Fig. 8. In typical angina pectoris, ST segment is depressed 1 to 2-mm which is visualized in V5 and V6 lead. But this change is transient, after two hours ST segment it come to normal. In chronic angina pectoris ST segment is depressed as well as T wave is inverted.

The main characteristic of ischemia in the cellular level is the depolarization of the cellular resting membrane potential. As large areas of the heart muscle become ischemic, its relaxation and contraction patterns are affected, which cause variations in the ST level and T-wave in ECG due to the development of an injury current between the ischemic and non-ischemic regions of the heart. The ST level change episodes lasting several seconds or sometimes some minutes, is an important indication in the diagnosis of myocardial ischemia. Repeated ischemia may lead to tissue injury and changes in electrophysiological properties that may in turn predispose the heat to arrhythmias.

The effect of angina is strongly connected with the behavior of S-T segment of the ECG signal [23]. To prepare a demo ischemia in heart, exercise ECG is considered for diagnosis of angina because this process is more sensitive and specific than the resting ECG in detecting myocardial ischemia [24]. The ECG of the affected patients change associated with myocardial ischemia includes horizontal or down-sloping ST-segment either in depression or elevation. It is observed that this effect on ST segment occurs after 60-80ms of the end of the QRS complex. Especially when the behavior of ST segments is accompanied with chest pain it is suggested to angina. Angina pectoris occurred at a low workload during the early stages of exercise and it persists for more than 3 minutes after exercise [25]. Increasing the threshold of a positive test to $\geq 2\text{mm (0.2mV)}$ ST-depression, will increase specificity at the expense of sensitivity.

![Fig. 7. ST Segment Measurement](image)

![Fig. 8. A typical angina pectoris characteristic: (a) ST Elevation and (b) ST Depression](image)

<table>
<thead>
<tr>
<th>Peak Wave</th>
<th>Amplitude (mV)</th>
<th>Peak to Peak Interval</th>
<th>Parameter Duration (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>0.25</td>
<td>RR</td>
<td>0.12-0.20</td>
</tr>
<tr>
<td>R</td>
<td>1.60</td>
<td>QRS</td>
<td>0.09</td>
</tr>
<tr>
<td>Q</td>
<td>25% of R wave</td>
<td>QT</td>
<td>0.31-0.44</td>
</tr>
<tr>
<td>T</td>
<td>0.1-0.5</td>
<td>ST</td>
<td>0.05-0.15</td>
</tr>
</tbody>
</table>

3 Methodology
Since the goal of this paper is to identify the angina pectoris from ECG signal, an algorithm is developed that is able to meet the challenge by a number of signal processing steps. This algorithm is able to detect the normal and abnormal peaks of ECG signal based on the characteristics of ST segment. From the signal loading to the final step to decide the presence of angina is presented by the block diagram given in Fig. 9. The major steps given in this block diagram are discussed gradually in the following.

3.1 Baseline Wandering Removal
After signal loading first preprocessing step is to remove the baseline wander from the raw ECG signal. Baseline wander is removed from the ECG signal by BWPF (baseline wander path finding) algorithm described in [26] which is designed by the polynomial fitting technique with multiple segmentation of the signal. By this methodology, total signal cannot be considered for polynomial fitting. This is because with a very high degree of polynomial will also be unable to trace the path of wandering. Considering this fact, after loading the full length ECG signal, it is divided into several segments which help to
divide the wandered path into several limited ordered wandered ECG signal. The procedure to implement BWPF algorithm is presented by block diagram given in Figure 3.2.

In this algorithm, segmentation number is taken as user or operator input based on the wandering path presented in ECG signal. Polynomial fitting order is self-assessed in our proposed methodology based on the threshold value of fitting. The algorithm starts from order 1 to fit the ECG wandering path and compute its fitting error to compare the standard value given as threshold. If the error is greater than the threshold value the process is repeated with the order=previous_order+1 and will continue the process to achieve the error value less than or equal to the threshold value. After fitting all the segments by different order and different values, the wandering paths are subtracted from the segmented signals and add together to represent the previous signal as wander free. After that the signal can be simply filtered to prepare it for the further processing like heart rate calculation, peaks identification, ischemia detection, angina pectoris identification, etc.

The aforementioned algorithm finds the wandering path and finally this path is subtracted from the raw ECG signal which gives baseline corrected signal. After that the noise presented in ECG signal can be removed by the Gaussian weighted moving average filter. In this type of moving average filter the center value of the window is weighted generally two or three times than the other values for convolution.

Since moving average filter, using as FIR criteria, has a property of low pass filter. Besides there are several noises having the frequency level is below the cutoff frequency and nearer to zero those can be incorporated due to low pass filter. Sometimes, the QRS complex of ECG signal becomes broad and distorted for the activation pulse originating in the ventricle as well as not going the actual conduction path. Frequency domain analysis of Fourier Transform can be applied to discriminate the high-frequency and low-frequency ranges of ventricular rhythm, atrial rhythm, parasympathetic and sympathetic activity signals. For example, spectral analysis is the linear transform used to diagnose ventricular tachyarrhythmia. Power spectrum of individual QRS complex is found significant differences in the 0-20 Hz frequencies. The spectrum in VT has maximum amplitude at 4 Hz [27]. Spectral analysis is also used in case of analyzing the heart rate variability (HRV) signals. Power spectra in the 0.15-0.5 Hz ranges reflect respiratory sinus arrhythmias and cardiac vagal activity. On the other hand, baro-receptor control is mediated by vdagal and sympathetic systems in the 0.04-0.15 Hz ranges. In addition with that very low-frequency (~0.04 Hz) is related to thermoregulatory, vascular mechanisms and rennin-angio tension systems [28-29].

According to the previous discussion we can conclude that if we use low pass filter like Gaussian window moving average filter, we can remove high frequency noises from ECG signal normally. Nonetheless, if we clear out the all noises except ECG signal we should use bandpass IIR (infinite impulse response) filter. In that case our proposal to set the band for ECG signal filtering should be 0.5Hz-45Hz. It will remove all physiological and power line noises. Therefore, Gaussian windowed moving average filter has been discussed briefly.

### 3.2 Gaussian windowed moving average filter

In electronics and signal processing, a Gaussian filter is a filter whose impulse response is a Gaussian function (or an approximation to it). Gaussian filters have the properties of having no overshoot to a step function input while minimizing the rise and fall time. This behavior is closely connected to the fact that the Gaussian filter has the minimum possible group delay. It is considered that the ideal time domain filter, just as the sinc is the ideal frequency domain filter [30]. These properties are important in areas such as oscilloscopes [31] and digital telecommunication systems [32]. Since signal exhibits one-dimensional property, the one-dimensional Gaussian filter has an impulse response given by,

\[
g(x) = \frac{a}{\pi} e^{-a \cdot x^2} \tag{1}
\]

Here, \( a \) is a constant and \( x \) is the data vector. On the other hand, the frequency response from the previous relationship given by the relation given in (1), we can estimate from the Fourier transform and which can be equated as given by the following relationship,

\[
g(f) = e^{\left(\frac{\pi^2 f^2}{a}\right)} \tag{2}
\]
Here, \( f \) is the ordinary frequency. These equations can also be expressed with the standard deviation as parameter in both time and frequency domains, and the relationships are presented by the following two relations (3) and (4), respectively [33, 34, 35],

\[
g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (3)
\]

\[
\hat{g}(f) = e^{-\frac{f^2}{2\sigma^2}} \quad (4)
\]

By writing \( a \) as a function of \( \sigma \) with the two equations for \( g(x) \) and as a function of \( \sigma_f \) with the two equations for \( \hat{g}(f) \) it can be shown that the product of the standard deviation and the standard deviation in the frequency domain is given by,

\[
\sigma \sigma_f = \frac{1}{2\pi} \quad (5)
\]

Where, the standard deviations are expressed in their physical units, e.g. in the case of time and frequency in seconds and Hertz. It means the before and after of the ECG signal we have added 100 vectors as zero. The window length is considered as 15. The MATLAB command `gausswin()` returns us the convolutional parameters of 15 window size and the output filtered signal is achieved by the multiplication with the gaussian window values and corresponding addition (can be called as convolution) obeying the principle of moving average filtering.

### 3.3 QRS Complex Detection

The QRS complex duration is one of the major parameters in the field of the ECG signal analysis. QRS complex are the middle part between P and T wave. This parameter is defined as the time it takes for depolarization of the ventricles. Normal depolarization requires normal functioning of the right and left bundle branches and it varies from 0.04 to 0.09 seconds. Any block in either the right or left bundle branch delays depolarization of the ventricle due to the blocked bundle. In abnormal cases the QRS interval is 0.1 seconds or more. There is an intra-ventricular conduction delay when the QRS interval is between 0.1 to 0.12 seconds. For QRS complex detection we used the FS2 algorithm and amplitude threshold in this work. The FS2 algorithm is designed based on the first and second derivative of ECG signal which was developed by Balda [37] in 1977, and further modified by Ahlstrom and Tompkins [38] in 1983, for high speed R-peak detection of recorded ECG signals. The first and second derivatives act as high pass filters which attenuates the low frequencies. To implement the algorithm the absolute value is taken from the first derivative of the original signal as given in (6) and (7),

\[
Y_0(n) = |ECG(n+1) - ECG(n-1)| \quad (6)
\]

\[
Y_1(n) = |ECG(n+2) - 2 * ECG(n) + ECG(n-2)| \quad (7)
\]

The first derivative is then smoothed by a 4-point moving average low-pass filter as given in (8) and then added to the second derivative as (9). This reduces the P and T wave deflections. Since the P and T wave deflections are almost removed, the only peak found is QRS which gives us the location span of QRS peak. This result is implemented to the filtered ECG signal to find the QRS complex.

\[
Y_1(n) = [Y_0(n-1) + 2 * Y_0(n) + Y_0(n+1)] / 4 \quad (8)
\]

\[
Y_2(n) = [Y_1(n) + Y_1(n)] \quad (9)
\]

Finally, threshold is determined to find the R peak location from the signal. Generally, threshold = 0.8 * max(Y3(n)). A QRS candidate is found when the sum exceeds the primary threshold, \( QRS(n) = Y3(n) > \) primary threshold.

### 3.4 S - T Segment Estimation

S wave is the last wave in QRS complex. Since the QRS complex is already detected by the previous algorithm the S wave can be found through the minimum value in the range \( R_{loc}+5 \) to \( R_{loc}+50 \) [39]. On the other hand, T wave can be found between the range \( R_{loc}+25 \) to \( R_{loc}+100 \) [40]. The ST segment is the most important feature of the ECG for investigating angina pectoris. The ST interval is measured from the J point (the minimum point of S wave) to the end of the T wave. Basically, the highest point of T wave is also known as K point. The elevation and depression of ST segment together with the T-wave changes indicate the zone of ischemia around the applied lead. First we identified the J point and T-onset. Then average amplitude between J point and T-onset is calculated.

### 3.5 Detection of Isoelectric Level and Comparison with ST Segment for Angina Detection

A reference level is necessary to measure the elevation or depression of ST segment. The Isoelectric line (IEL), which is the flat parts of the ECG can be used as this reference level. The IEL usually place between the T and P waves or between the P wave and the QRS complex. The IEL is found from the flat portion of ECG characteristics points J and K. Then we compare the ST segment level with the isoelectric line level. If this segment is considerably underneath the isoelectric line then ST segment depression occurs which indicate that part of the patient’s myocardium is not receiving sufficient oxygen and results the myocardial ischemia. If the ST segment is raised above the isoelectric line it indicates myocardial infarction. More quantitatively, when ST level is greater than 0.08 mv underneath the isoelectric line and has slope of more than 65 degree with respect to the perpendicular axis of ECG, it is considered as ST depression [41]. On the contrary, ischemic is considered if ST segment is raised above the isoelectric line more than 0.08mv.

The algorithm is simulated in MATLAB environment using normal and affected patient’s ECG signal from MIT arrhythmia database [42] to assess its effectiveness. The ECG signals named by ecg100_1, ecg100_2, ecg105, ecg109, ecg111, ecg112, ecg113, ecg114, ecg115, and ecg119 in MIT-Arrhythmia database are tested. For example, the average isoelectric line and ST segment level for ecg100_2 are found as: \( st_{avg} = -0.4951 \) mv and \( iso_{avg} = -0.4060 \). The resultant subtracted value is -0.0891 mv which shows that ST segment is deviated below the isoelectric line and hence it is considered as ST depression or negative ST deviation.

### 3.6 k-Nearest Neighbor Classification

Since the previously explained method depends on only one
feature to distinguish the normal and abnormal ECG pattern which may be often proved unacceptable. It will be more accurate approach if we consider a number of significant features of the ST segment like slope, standard deviation, energy, skewness, etc. For multiple comparisons the machine learning approach is more suitable and acceptable approach. In this sense, kNN based method is better than the other method like linear discrimination analysis (LDA) and Bayesian algorithm (BA). In this project, we have used kNN based predictive model to find the total normal and abnormal ST segment in each ECG signal.

Nearest neighbor is a nonparametric estimation method that implements a refinement where the feature environment is high resolution in regions with dense training and low resolution in variance. The method works based on the following steps.

- Suppose, \( Z(z) \subset R^n \) is a hypersphere of volume, \( v \) and center, \( z \). Let us, consider that the training set \( S_k \) consists of \( N_k \) number of samples. Therefore, according to binomial distribution, the probability of including precisely \( n \) samples within the hypersphere, \( Z(z) \) can be equated by the following expectation.

\[
E[n] = N_k \int_{y \in Z(z)} p(y | w_k) \approx N_k v p(z | w_k)
\]

Here, \( w_k \) represent the number of class.

- The radius, \( z \) of the sphere \( Z \) is selected in way that \( Z \) contains exactly \( K \) samples. In this sense, the volume becomes the function of \( z \) and it can be presented as \( \nu(z) \). So, the estimation of density based on the number of \( K \) can be calculated as,

\[
\hat{p}(z | w_k) = \frac{K}{N_k \nu(z)}
\]

- The parameter, \( k \) can control the stability between the variance and bias. To classify any vector, \( z \) the radius of the sphere is selected such a way that the sphere contains exactly \( k \) samples occupied from \( S_k \). These samples are termed as the \( k \)-nearest neighbor. Let \( k_i \) indicates the number of samples originated with the class \( w_k \). The conditional density is estimated as the following relationship given in (12).

\[
\hat{p}(z | w_k) = \frac{k_i}{N_k \nu(z)}
\]

The following distances can be chosen to find the expected nearest neighbor: i) Minkowski Distance, ii) Manhattan Distance, iii) Euclidean Distance, or iv) Chebyshev Distance.

3.7 Performance Measurement Criteria

The performances of the classifiers were determined by the calculation of sensitivity, specificity, and total classification accuracy by the following relationships.

- Sensitivity (True positive rate) = TP/(TP+FN)
- Specificity (False positive rate) = TN/(TN+FN)
- Total classification accuracy = (TP+TN)/(TP+TN+FP+FN)

Where,

Total number of correctly classified positive patterns = TP,
Total number of actual positive patterns = TP+FN
Total number of correctly classified negative patterns = TN
Total number of actual negative patterns = TN + FP

Total number of correctly classified patterns = TN + TP
Total number of applied patterns = TN + FN + TP + FP.

4 RESULTS AND DISCUSSIONS

For the performance measurement, MIT-arrhythmia database (most standard ECG database for myocardial disease) is used in this work. The project is accomplished by designing an algorithm in MATLAB platform. The signal processing and analysis of ECG signal is done by the code written in MATLAB 2012b. In this algorithm, some predefined well known functions are used. On the other hand, a number of functions have been created as required to estimate different properties of ECG signal. Finally a complete algorithm is developed to check the angina existence in ECG signal. The performance of the proposed algorithm is measured by some metrics like accuracy, sensitivity, and specificity. The real condition and the results achieved by proposed work is also compared in this section.

4.1 The MIT-BH Arrhythmia Database

Academic and industry research groups have long studied irregularities of the heart cycle and identification of arrhythmias, and having access to databases of ECG recordings has been a useful tool for these activities. The MIT-BIH Arrhythmia Database, found on the PhysioNet web site, is a free-to-use database that has been available since 1980, and has been regularly used for the evaluation of arrhythmia detectors; as Moody [43] said, by having the availability of the databases, the general standard of commercial arrhythmia detectors has improved rapidly.

The database originated from the Arrhythmia Laboratory of the Beth Israel Hospital (BIH) for the intent of making them available to the research community. There are a total of 48 signals available and were taken from patients aged 23 to 89 years old. The recordings were made using one of nine DEL Mar Avionics model 445 two channel reel-to-reel Holter recorders, whilst the digitising used a Del Mar Avionics 660 playback unit. Each digital record lasts approximately 30 minutes and contains 650000 samples; the digitising rate was produced at 360 samples per second which accommodated a simple 60Hz notch filter implementation. The ECG lead or electrode placement varied with some patients, however in most cases they were recorded using the modified limb lead II (MLII).

Four of the 48 recording had pacemaker artifacts that were not accurately reproduced due to the pass band filter used during the analogue to digital conversion. These unwanted variations in signals have the ability to substantially affect the results obtained from peak detection algorithms. This has given researchers additional challenges to overcome when designing new algorithms.

4.2 Preprocessing Results of the ECG Signal

According to the previous description provided in chapter 3, at first the raw ECG signal is loaded. After loading the ECG signal, baseline wandering was removed from the signal. It is already mentioned that the baseline wander removing method was followed by BWPF algorithm. By this algorithm, the base-
line wander is sectioned by some suitable part of the total length of the signal. Then, from every part of the raw ECG signal, the baseline path is estimated by polynomial fitting. The polynomial fitted path is subtracted from the raw ECG signal. Eventually the raw baseline wandered signal turned into baseline wander removed signal or baseline wander free signal. The result of the BWPF algorithm on MIT-arrhythmia ECG signal is shown in Fig 10. From the figure, we came to know that the result removes the baseline wander from the raw ECG signal.

**Noise removal from the ECG signal:** The preprocessing part of angina pectoris detection is very important. Besides baseline wander removal and noise cancellation different point of ECG signal like P peak, QRS complex, S-T Segment, J &K point identification cannot be possible appropriately. Without these properties estimation, no one cannot achieve computer based angina pectoris detection algorithm. On this consequence, after baseline wander removal the physiological and power noises are removed from the signal. There are two types of noise cancellation choices in this case. One is lowpass filtering and another is bandpass filter. Since, baseline wander is removed by the advanced algorithm (BWPF), no need to use bandpass IIR filter. Therefore, on this condition noises are removed by Gaussian window based moving average filter. It is well known that the moving average filter has significant property of low pass filtering. In this project, Gaussian window is used so that the R peak would not be decreased significantly by the effect of moving average filter. The baseline wander removed signal with noise and noise free signals are shown in Fig. 11 where it is observed that the signal becomes quite noise free and smoothed.
4.3 QRS Complex Detection

Since QRS complex detection is very important role for this project, most widely used and sophisticated method FS2 algorithm is used. FS1 algorithm can be used as well but FS2 is more accurate. In this algorithm, 1st and 2nd derivation of original signal is used obeying the rule described by the relationships by the relations 6 to 9. For QRS complex detection amplitude threshold is also used. The proper concept can be achieved by the program code given in [44].

According to the FS2 algorithm every steps from filtered ECG signal to reach the QRS complex are calculated and presented by the Figure 4.3. From the figure, we can observe that the 1st and 2nd derivative effect of the signal and the variation formed by the FS2 algorithm. By the application of FS2 algorithm on ECG signal all other peaks like P and T became flat and on that point depending on some threshold value, QRS complex is detected.

Considering the threshold described in chapter 3, the Q, R, and S points are detected by our algorithm. The techniques are described by the code given in previous chapter. Fig. 12 depicts the corresponding QRS points found by the algorithm. In addition with that, the points are registered in original ECG signal given in Fig. 13.

4.4 S-T Segment Estimation and J-K Point Finding

The ST segment is the most important feature of the ECG for investigating angina pectoris. The ST interval is measured from the J point to the end of the T wave. The elevation and depression of ST segment together with the T-wave changes indicate the zone of Ischemia around the applied lead. S wave
is the last wave in QRS complex. Since the QRS complex is already detected by the previous algorithm the S wave can be found through the minimum value in the range Rloc+5 to Rloc+50 [11]. On the other hand, T wave can be found between the range Rloc+25 to Rloc+100 [11]. The ST segment is the most important feature of the ECG for investigating angina pectoris. The ST interval is measured from the J point (the minimum point of S wave) to the end of the T wave (which is also known as K point). The elevation and depression of ST segment together with the T-wave changes indicate the zone of ischemia around the applied lead. First we identified the J point and T-onset. Then average amplitude between J point and T-onset is calculated.

A similar method is developed for the detection of ST segment level. First we identified the J point and T-onset. Then average amplitude between J point and T-onset is calculated using MATLAB. Detected J and K point in ECG signal in given in Figure 4.5. In addition with that the corresponding S and T points are also identified by the algorithm given in Figure 4.6.

4.5 Performance Measurement of the Algorithm

Isoelectric line is the flat parts of the ECG, for example between the T and P waves or between the P wave and the QRS complex. A band selection method is developed for the detection of isoelectric level detection. Result obtained from QRS detection technique i.e., the magnitude of R-peak is also known as first maxima of the ECG signal. The last step in the algorithm is to compare the ST segment level with the isoelectric line level. If this segment is significantly below the isoelectric line it is called ST segment depression and it suggests that part of the subject’s myocardium is not getting enough oxygen (myocardial ischemia). This segment is also frequently elevated (ST segment elevation) above the isoelectric line in the early stages of a myocardial infarction.

When ST deviation is more than 0.08 mv below the isoelectric line and has an angle larger than 65 degree measured from vertical line, it is considered as negative ST deviation or ST depression [12]. The second rule classifies beat as ischemic if ST deviation is more than 0.08mv above the isoelectric line which is considered as ST elevation. Consider the results obtained from isoelectric line and ST segment level detection method which are: st_avg=0.4951 mv and iso_avg = -0.4060. When subtracted gives the result -0.0891 mv which shows that ST segment is deviated below the isoelectric line and hence it is considered as ST depression or negative ST deviation.

Using our proposed algorithm designed in MATLAB environment, normal and affected patient’s ECG signal from MIT Arrhythmia database are analyzed for evaluating the adeptness of the proposed algorithm. The results of different ECG signals of MIT Arrhythmia database by our proposed algorithm are presented gradually. Here, ECG signals named by ecg105, ecg112, ecg113, ecg114, ecg115, and ecg119 in MIT Arrhythmia database are tested.

![Detection of J and K point of ECG Signal](image1)

**Fig. 14. J and K point identification by the proposed method**

![Detection of S and T point of ECG Signal](image2)

**Fig. 14. J and K point identification by the proposed method**
The result is given in Fig. 16. All results found by the kNN method and found the abnormal ST segments are 50 in that signal where the actual is 51. Therefore the classification accuracy is 98%, specificity= 98%, and sensitivity 100%. The result is given in Fig. 16. All results found by the kNN classifiers are given in Table IV with the other information like accuracy, sensitivity, and the specificity of the proposed algorithms.

Table III: MIT Database ECG Signal and Their Angina Calculating Results by Proposed Method

<table>
<thead>
<tr>
<th>Data ID</th>
<th>Detected Normal</th>
<th>Detected Angina</th>
<th>Actual Angina</th>
<th>Accuracy %</th>
<th>Sensitivity%</th>
<th>Specificity%</th>
</tr>
</thead>
<tbody>
<tr>
<td>ecg100_1</td>
<td>12</td>
<td>52</td>
<td>51</td>
<td>96.8%</td>
<td>98.4%</td>
<td>-</td>
</tr>
<tr>
<td>ecg100_2</td>
<td>51</td>
<td>15</td>
<td>52</td>
<td>88.57%</td>
<td>96%</td>
<td>86.6%</td>
</tr>
<tr>
<td>ecg105</td>
<td>14</td>
<td>57</td>
<td>75</td>
<td>91.7%</td>
<td>82.3%</td>
<td>87%</td>
</tr>
<tr>
<td>ecg109</td>
<td>8</td>
<td>75</td>
<td>100</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>ecg111</td>
<td>56</td>
<td>5</td>
<td>5</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>ecg112</td>
<td>3</td>
<td>72</td>
<td>75</td>
<td>92.6%</td>
<td>-</td>
<td>96%</td>
</tr>
<tr>
<td>ecg113</td>
<td>11</td>
<td>37</td>
<td>30</td>
<td>85.4%</td>
<td>61%</td>
<td>100%</td>
</tr>
<tr>
<td>ecg114</td>
<td>25</td>
<td>21</td>
<td>21</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>ecg115</td>
<td>34</td>
<td>22</td>
<td>20</td>
<td>96.4%</td>
<td>94.4%</td>
<td>100%</td>
</tr>
<tr>
<td>ecg119</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>97.7%</td>
<td>100%</td>
<td>95.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Average=</td>
<td>94.9%</td>
<td>83.2%</td>
</tr>
</tbody>
</table>

4.6 kNN based Results Comparing Multiple Features

It is already stated that the only one feature cannot be criteria to take decision. Therefore, with comparing multiple features, kNN based classifier has been trained with normal and abnormal features of S-T segments of each beat of the ECG signal. At first we have collected 30 normal S-T segments and 50 abnormal S-T segments. From these segmented signals, four features (mean, slope, and standard) have been extracted and trained the kNN classifier as to predict the nature of normal and abnormal property of ECG S-T segment.

If we have a time series of N number of samples, we can define it by \( x(n) \). The mean of \( x(n) \), \( \bar{x} \) can then be calculated as,

\[
\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x(n) \tag{13}
\]

Similarly, the variance, \( \sigma^2 \) is calculated as,

\[
\sigma^2 = \frac{1}{N} \sum_{n=1}^{N} (x(n) - \bar{x})^2 \tag{14}
\]

On the other hand, the slope is calculated from the linear fit of the data points found in the given series \( x(n) \). If a set of data from the time series \( x(n) \) is fitted with the line \( y=mx+c \) according to the least square error method, then the slope of the series is \( m \). Here, \( c \) is the value of the intersecting point at \( x(0) \) in \( y \) axis. From the ECG signal, we have separated all the S-T segments and then the features have been extracted. Some normal and abnormal conditioned S-T segments were separately due to train the classifier. Here, in Table III the normal and abnormal features are given to understand the actual variation in the features space of the normal and normal ST segment of ECG signal. From the features given in Table III we have found that the ST segment of the abnormal ECG beat has the negative slope with a larger value, on the other hand normal ST segments often have the slopes with positive values. With the help of the feature given in the following table the kNN classifier has been trained so that it could be able to differentiate the normal and abnormal ECG beat due to the effect of angina pectoris.

After the training with given data set, a small set of testing data has been applied to check the networks response with respect to its different distance calculating procedures of the nearest neighbors. The results of the testing performance are given in Fig. 15. From the figure, we have found that the query points for the testing points and their distance calculating through Mikowski and Chebyshev distance have the same effect. In addition to that, by the kNN based trained network we have checked the classification accuracy, sensitivity, and specificity of all the ECG signals. These results are based on the property of ST segments of all the beats of a signal. The data 100_1 of Arrhythmia data base have been checked by the proposed method and found the abnormal ST segments are 50 in that signal where the actual is 51. Therefore the classification accuracy is 98%, specificity= 98%, and sensitivity 100%. The result is given in Fig. 16. All results found by the kNN classifiers are given in Table IV with the other information like accuracy, sensitivity, and the specificity of the proposed algorithms.
In this paper, we have proposed a method to detect angina pectoris from ECG signal. In this method additionally we have used baseline wandering correction that makes the method more correct in angina pectoris detection. From the results we get in average 94.9% accuracy, sensitivity 83.2%, and specificity 86.5%. In addition with the help of kNN classifier we have checked the classification accuracy of the proposed method based on four features. From the classification results, we have found that the proposed system can classify the normal and abnormal ECG beat based on the features of S-T segment with 96.48% classification accuracy, 96.5 % specificity and 97.2% sensitivity. From this point of view, we can conclude the results that the kNN classifier can classify the abnormal condition of ECG beats more accurately. We believe that this method will be helpful for computerized analysis of ECG data in the field of angina pectoris detection which will be supportive for the doctors in diagnosis as well.

Though MIT data base is used to calculate the results, for any ECG data with similar sampling frequency, this method will work appropriately. It can be a future direction to match the algorithm for different sampling rate considering a ratio based approach in peak detection problem. Eventually, if the sample matching algorithm can be implemented all types of ECG signal can be analyzed through this algorithm with its prior information about the sampling rate. It is also mentionable that there are nine different algorithms to find the peaks from ECG signal. Here, only FS2 algorithm has been used. Therefore, the other peak detection algorithms can be used to check the performance.

Additionally, other classifiers can also be applied to improve the performance of the algorithm. Someone can also design a MATLAB based graphic user interface (GUI) program using the proposed algorithms which can be more user friendly to detect angina pectoris from ECG signal.

ACKNOWLEDGMENT

The authors would like to thank Prof. Dr. A. B. Aowalad Hossain to guide the total research work because we believe without his direction this work would not be accomplished.

REFERENCES


