

DETECTION OF HEART DISEASES BASED ON EXTRACTED FEATURES

Asst. Lect. Sura Saleem Rasheed

General Directorate of Education in Wasit, Wasit, Iraq

Asst. Lect. Riyam Faisal Alwash

College of Medicine, University of Babylon, Babylon, Iraq

Article information:

Manuscript received: 21 Apr 2025; **Accepted:** 5 Jun 2025; **Published:** 19 Jun 2025

Abstract: Cardiac rhythms that deviate from the normal are known as dysrhythmias, heart murmurs; they are the outcome of any of a number of disorders that involve the heart's electrical conduction system. These abnormalities can progress to life-threatening disorders like Irregular atrial rhythm, ventricular arrhythmias, or sudden death of cardiac cause. Thus, the rapid and accurate diagnosis of such arrhythmias is important for early clinical intervention.

In this paper, we propose an inter figure approach for high-speed inter figure AI based on features extracted from The MIT-BIH Arrhythmia Database was used to classify whether a given heart rhythm was normal or abnormal. electrical heart activity. Preprocessing of data consisted of filtering of the signal and R-peak for R-point peak detection routine with the Pan-Tompkins algorithm, and of analysis features extraction from both time-domain (for example, average and standard deviation) and frequency-domain (for example, dominant frequency and spectral entropy).

A model using the extracted parameters was conducted by XGBoost, and strong data performance was observed: Achieved a precision rate of 96.20%, a true positive rate of 95.40%, a true negative rate of 97.00%, and an F1 metric value of 96.10%." To better understand the decision making of the model, a feature importance evaluation was also preformed, revealing that the most effective predictors were dominant frequency and spectral entropy. The outcomes validate the proposed strategy to be used in real-time automatic ECG tracking systems.

Keywords: Electrocardiogram (ECG), Arrhythmia Detection, Heartbeat Classification, Feature Extraction, XGBoost Classifier.

1. Introduction

Cardiovascular diseases (CVDs), including ischemic heart disease and stroke, still remain the leading cause of death globally, with an estimated 17.9 million deaths, accounting for nearly 32% of total global deaths, in 2019 [1]. These diseases disproportionately burden low- and middle- Low-resource countries. Timely detection of heart related problems is essential to improve the survival rate. the patient with improvement of unfavorable prognosis. Of these diagnostic methods, electrocardiography (ECG) is commonly preferred due to its cost effectiveness, non-invasiveness, and ability to detect a wide range of cardiac disorders [2].

Although it has been widely used in the clinics, the precise interpretation of ECG signals tend to be a difficult task for the medical practitioners. Studies have demonstrated high rates of misdiagnosis in serious disorders such as ventricular tachycardia, atrioventricular block, and myocardial infarction. Difficulties in diagnostics often result from an overlap in morphological properties of different cardiac pathologies, as well as from interpatient variety of ECG patterns because of, i.a., age and physiological differences or occurrence of comorbidities [2], [5].

In order to overcome these limitations, convolutional automated electrocardiogram analysis systems (CIEs) have been developed. However, current methods usually do not reach the desired level of diagnosis accuracy, interpretability and compliance with international guidelines, so their continued oversight by health professionals is required [6], [7]."

The main features of ECG include P wave, QRS complex and T segment. three major waveform elements of a standard ECG strip. These correspond to the electrical depolarization and repolarization of cardiac muscle in a sequence [13]. Any deviations in the morphology or timing of these waves may be suggestive of some underlying heart condition. The ECG signal as an ECG* depicting the electrical pattern of a single heart beat, was used in this study as the basis for waveform decomposition and off-line extraction of analytical parameters, outlined in Figure 1.

Recent developments in artificial intelligence have seen machine learning models applied to automate ECG classification with promising performance and robustness. Here, these methods were investigated for their ability to generate meaningful results in an interpretable manner, at lower computational cost, when combined with artificial features extracted from time and frequency domains after simple pre-processing. We used the XGBoost classifier to build an intelligent binary classification model that can determine between normal and abnormal cardiovascular rhythm using our extracted features for segmented ECG waveforms.

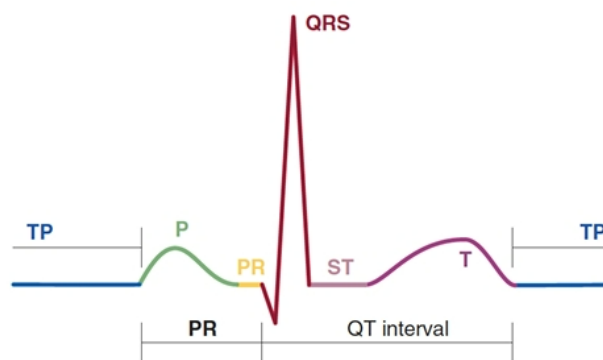


Figure 1: A representative electrocardiogram (ECG) trace highlighting The different parts of the typical wave form — namely, the P factor, the QRS complex, and the T element

To address the mentioned concerns, this work introduces a transparent, feature-driven classification model utilizing the XGBoost machine learning technique. By applying advanced feature extraction techniques to segmented ECG data—including both temporal and spectral attributes—this approach aims to deliver precise arrhythmia detection with clarity, interpretability, and optimized computational performance, thereby supporting early diagnosis and clinical decision-making.

2. Related work

Over the past years, many works Have worked on using computational learning models to detect/ classify heartbeats from ECG data. and arrhythmias detection - the MIT-BIH

Arrhythmia Database, thanks to its rich annotations about people and its clinical variety, is probably the most used public benchmark.

In addition to Kishore et al. and Wang et al. [8], proposed a classification model, which consists of a genetic algorithm for feature selection and a radial basis function neural network (RBFNN) to perform the mapping procedure. The classification accuracy, sensitivity, specificity of this method were 98.5%, 98.3% and 99%. Mezaini At the same time, Karthiga et al. [9] used a convolutional neural network (CNN) to obtain automatic feature abstraction from ECG signals (without the requirement of manual feature extraction) and achieved good classification accuracy overall. 91.92%. Similarly, Madan et al. As an example, [10] proposed a hybrid deep learning method in which the ECG signals were transformed into 2-dimensional images using CWT and then were classified by a CNN-LSTM model. A total accuracy of 97.3% was performed by their system with 97% sensitivity and 98% specificity. As mentioned above, it deals with the concept of embedding ECG data as images, so Ting et al. [11] extracted the STFT-based time-frequency mapping generated by the short-time Fourier transform (STFT) and classified it by the deep learning architecture Efficient-B0, and reached an accuracy of 97.3% and the same number of 89.55% of the sensitivity and specificity. In another study, Alfaras et al. presented a real-time method for categorizing ensemble of echo state networks (ESNs), which learned only informative features without any heuristics, with overall accuracy and sensitivity of 96.8% and 92.7% in their analysis [12]. Bhatia et al. Similarly, [13] proposed a hybrid CNN-LSTM model to classify heart beats without handcrafted features and showed good performance of inputted beats achieving 98.36% accuracy and 94.36% specificity.

Additionally, He et al. In [14] dual deep learning architecture was investigated based on merging residual convolutional modules with bi-directional LSTM (Bi-LSTM) networks to enhance the accuracy of heartbeat classification. Their model did not share precise performance statistics but concentrated on fortification and generalization over different arrhythmic types.

Notice that these studies illustrate a trade-off between interpretability and performance. Deep learning approaches, although also producing competitive results, are mostly black-boxes. In contrast, models using handcrafted features and interpretable algorithms like gradient boosting allow for an understanding of feature importance, thus providing greater explainability into diagnostic frameworks. This is the motive for the current work, in which the focus is on a feature-based heartbeat classification methodology based on the XGBoost algorithm, leading to optimal performance with respect to accuracy, interpretability, and clinical relevance.

While recent studies on ECG classification have demonstrated that deep learning models can achieve excellent results, the limited interpretability and high computational load of these models hamper their integration into practice. In contrast, classical machine learning methods with hand-crafted features provide a greater transparency albeit inferior performance in some cases. Hence in this work we try to fill this gap by proposing a highly accurate yet interpretable model using a set of manually chosen handcrafted features combined with the XGBoost algorithm, with a good balance between accuracy, efficiency, and precision.

3. Methodology of proposed methods

3.1 Dataset

Created in 1980, the MIT-BIH Arrhythmia Database was created in a joint project by MIT and Beth Israel Hospital and is highly useful for electrocardiogram (ECG) signal analysis

[15]. This dataset has since become a fundamental reference in the area, a vital reference for cardiac arrhythmia classification, heart disease detection [16], and some recent progress on AI-based research in ECG signal processing [16].

It contains forty-eight recordings of two-channel dynamic ECGs with their annotation files sampled at 360 samples/s for half an hour pipe signals from 25 men and 22 women [18]. The set consists of well-annotated data include 112,647 ground truth annotations [17]. The annotations are classified to beats and non-beats, among which beat annotations contain 15 subtypes, shown in Table 1. This database is very useful for testing and evaluation of types of QRS detection methods [15], because it includes a wide range of QRS waves, and subtypes which cover all types of ECG signals.

Before any of these algorithms can be implemented in a medical device, a rigorous assessment of the algorithm performance must be conducted; the MIT-BIH database is a crucial step in this testing. Mistakes at this evaluation stage can lead to diagnostic tools for use by health professionals being less accurate, which would then have a downstream effect on clinical decision-making and patient treatment outcomes. Hence, datasets such as MIT-BIH need to undergo stringent validation and verification in order to increase reliability and clinical applicability [16], [17] "apparatus for treatment and diagnosis

Table 1: MIT-BIH Arrhythmia Database Beat Classifications:

Annotation symbol	Beat type	Description
N	Normal beat	Normal heartbeat
L	Left bundle branch block beat	Left bundle branch block beat
R	Right bundle branch block beat	Right bundle branch block beat
A	Atrial premature beat	Premature atrial contraction
a	Aberrated atrial premature beat	Aberrated atrial premature contraction
J	Junctional premature beat	Premature junctional beat
S	Supraventricular premature beat	Supraventricular premature beat
V	Premature ventricular contraction	Premature ventricular contraction
r	R-on-T premature ventricular contraction	R-on-T premature ventricular beat
F	Fusion of ventricular and normal beat	Fusion of a ventricular and normal beat
e	Atrial escape beat	Atrial escape beat
j	Junctional escape beat	Junctional escape beat
n	Supraventricular escape beat	Supraventricular escape beat
E	Ventricular escape beat	Ventricular escape beat
Q/?	Unclassified beat	Unclassified beat

Core Takeaways

N (Physiological heartbeat): Indicates a regular cardiac cycle.

The others are signs and relations to rhythm disturbances such as early contractions and escape beats.

it as an important concept in ecg studies. What makes it unique is the structured classification system that it has, which contains five standardized arrhythmia groups [15]. Most of these studies utilize QRS detection methods as a primary building block, since they are crucial for arrhythmia detection, classification, and diagnostic purposes.

Since this database has been home to many QRS detection algorithms, testing and validating them has been done based on it. These algorithms based on the embedded beat

annotations on the ECG recordings which are used for training and evaluation purposes. These beats are used as a foundational data in learning-based approaches, as they enable steady and consistent pathways for benchmark performance model development across studies [16] [17]. An overview of the process of the present study is shown in Figure 2. The process comprises five primary stages — data acquisition, data preprocessing, feature extraction, classification, and model evaluation. All of the stages are essential to convert noise-free ECG signals to the form that can be used for efficient Division of cardiac rhythm types.

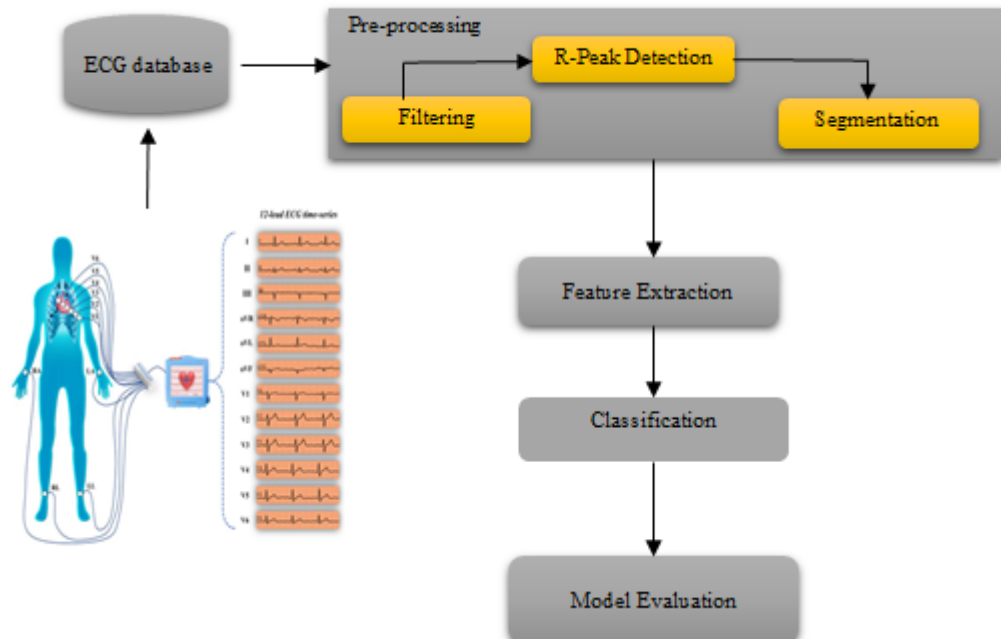


Figure 2: Developed System for the Heart Disease Diagnosis

3.2 Pre-Processing

One of the essential steps in ECG signal analysis is preprocessing, which serves as the preparation stage for raw ECG signals to enable them to be read and processed correctly in health monitoring systems. Electrocardiogram (ECG) signals are often prone to baselines and other types of signal corruptions such as power line variation and muscular electromotive activity. therefore this step is highly required to improve the quality of signal and retain clinically significant features.

This study focuses on the preprocessing stage consisting and consist of three basic operations: noise reduction using filters to eliminate unwanted signal noise, R-peak detection to highlight the critical cardiac events in the ECG recording and segmentation which implies breaking the intact ECG signal into smaller, single heartbeat segments for a better feature extraction and classification.

3.2.1 Filtering

The present study uses various digital filtering methods for noise cancellation of the original ECG signal to improve signal quality before feature extraction and classification. These filtering techniques were used to obtain the signals clean, smooth, and worth to apply analysis accurately also based on the MIT-BIH arrhythmia database

➤ High-Pass Filter (HPF)

ow-frequency noise removal, mainly caused by baseline wander as a result of respiration and motion of the electrodes, is performed using a high-pass filter since it can alter the

structural shape of the ECG waveform and interfere with the accurate identification of essential features like the P-wave and QRS complex. During ECG signal preprocessing, high-pass filters are typically configured with a cutoff frequency near 0.5 Hz [18]. The Butterworth filter is commonly favored due to its smooth frequency response and minimal phase shift distortion [19].

➤ Low-Pass Filter (LPF)

Apply a low-pass filter customized to eliminate high frequency disturbances — for the most part electromyographical artifacts and power line perturbations — which are usually presented at high frequencies, well above 50 Hz. A commonly used cutoff for ECG signal processing is 50 Hz, in order to eliminate power line noise and muscular artifacts. Smooth signal preservation is often performed using Butterworth, Chebyshev, or elliptic [18], [20] designed filters. Fig. The performance results of the high-pass and the low pass filters are presented in subfigures 3(a) and 3(b)

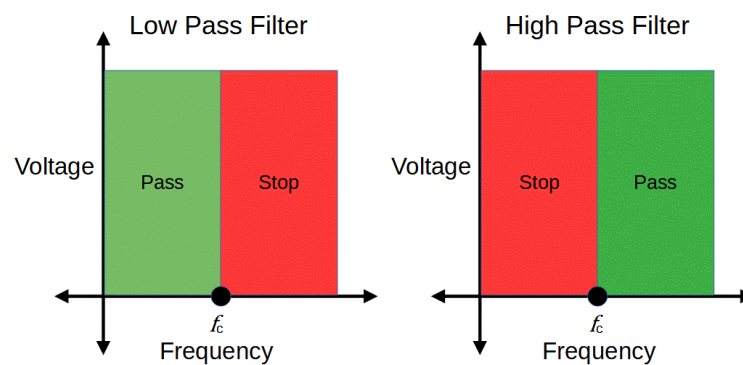


Figure 3: Comparison between high-pass and low-pass filters

➤ Notch Filter

The main job of the notch filter is to attenuate local frequency interferences -- particularly those generated by utility power at 50 or 60 Hz, depending on the local power-grid frequency. This unwanted signal, produced by electric devices, may markedly disturb ECG recordings. A notch filter only attenuates a certain frequency band (e.g., 49–51 Hz) while leaving the remainder of the signal unchanged. The Butterworth notch filters are often desired in application due to their stable and gentle attenuation responses [18]. Its operation can be explained with the aid of Figure 4.

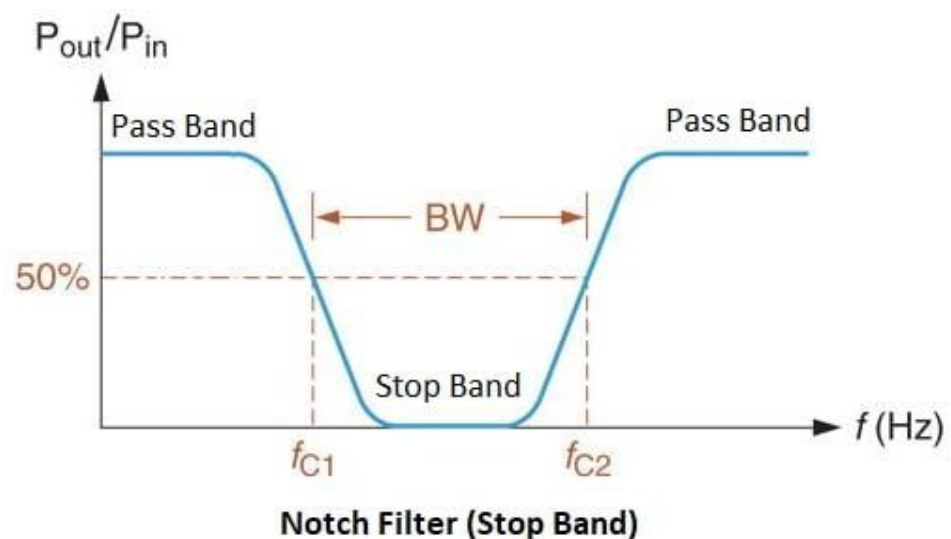


Figure 4: Notch Filter.

➤ Band-Pass Filter (BPF)

Band-pass filtering is crucial for extracting the specific frequency interval that encompasses the dominant energy of the QRS segment, by default it is selected between 5 Hz and 15 Hz. It assists in scrapping off of unimportant low- and high-frequency components of signals, yet keep important data about the heart. Due to lower distortion of signal and improvement of QRS [18],[21], Butterworth band-pass filters are preferable. Figure 5: Notch filter operation.

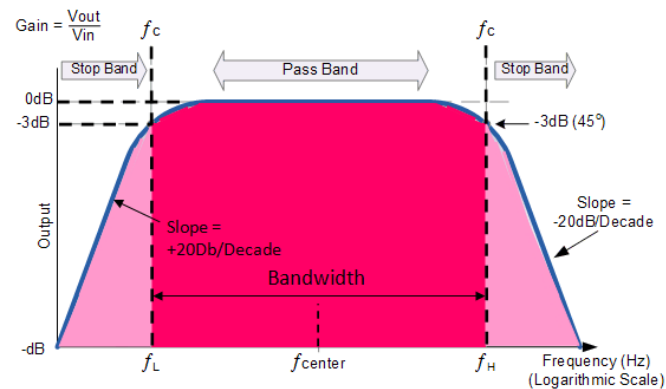


Figure 5: Band Pass Filter

3.2.2 R-Peak Detection

Temporal locations of R-wave peaks, that serve as important landmarks of ventricular depolarization—is a fundamental step in interpreting an electrocardiogram (ECG) signal. This procedure enables the division of the signal into discrete heartbeat segments, which are subsequently utilized for diagnostic assessment or classification. The process involves multiple stages: the signal undergoes filtering and amplification, and the detected peaks are evaluated against a defined threshold to confirm their authenticity as true cardiac events.

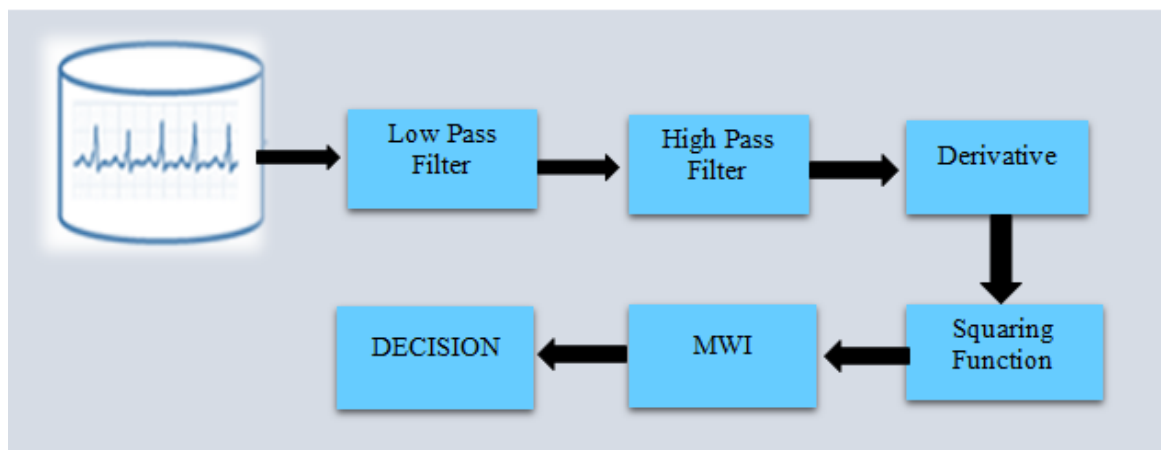
3.2.2.1 Pan-Tompkins Algorithm

For real-time applications, The Pan-Tompkins algorithm is still popularly used in QRS waveform separation complexes [22]. In this method, R-wave peaks are detected using parameters like amplitude, slope or width in a single processing window [22]. The algorithm is divided into two major parts: Signal conditioning and peak detection.

During the first pre-processing phase, the raw ECG signal has to be pre-processed to be appropriate for further processing. This phase involves noise filtering and signal enhancement through a smoothing process, aiming to highlight the slope of the QRS segment and diminish its width for a more accurate detection.

In the decision stage, dynamic The noise features are suppressed by setting threshold levels to distinguish true signal peaks. The system has a band-pass filter which is implemented with the sum of low pass and high pass filter stages. and their resulting derivatives, squaring, and moving window integration (WMI). This complete filtering–transformation–integration–thresholding–classification continuum is illustrated in the diagram Figure 6.

The sample rate for this approach is 360 Hz. A digital band-pass filter is used to reduce noise and motion artifacts. Moreover, in order to deal with fluctuations in



QRS morphology and cardiac rhythm, threshold values reflect automatically updated values in accordance with adaptive parameters. The steps taken in the decision making: in Figure 7, we present a representation of algorithm workflow including the process in details in the flowchart in Figure 7.

Figure 6: A block diagram of The Pan-Tompkins algorithm.

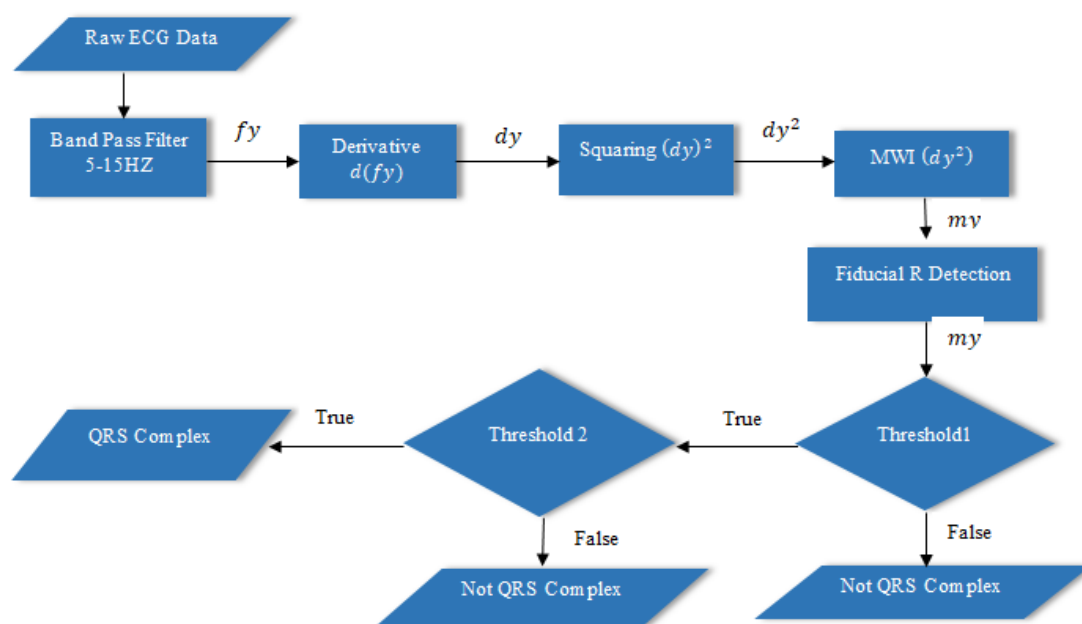


Figure 7: The Process Flow of the Pan-Tompkins algorithm

➤ Band-Pass Filter

The dual-edge frequency filter stage is used to eliminate undesired noise components. The passband is chosen to best preserve the QRS complex energy within a pre-defined frequency band. A LPF and HPF were connected in succession to exceed a target passband of about 5-15 Hz. High-frequency disturbances including electromyographic (EMG) activity, electric line interference, and T-wave noise are removed by the LPF, whereas a low-frequency signal content are preserved. On the other hand, the HPF suppresses Baselines fluctuation). The contrast between the..." unfiltered (a) and band-pass filtered (b) ECG signal is shown in Fig. 11. The resulting signal indeed shows less residual movement artifacts and more general signal clarity, as also illustrated in Fig.

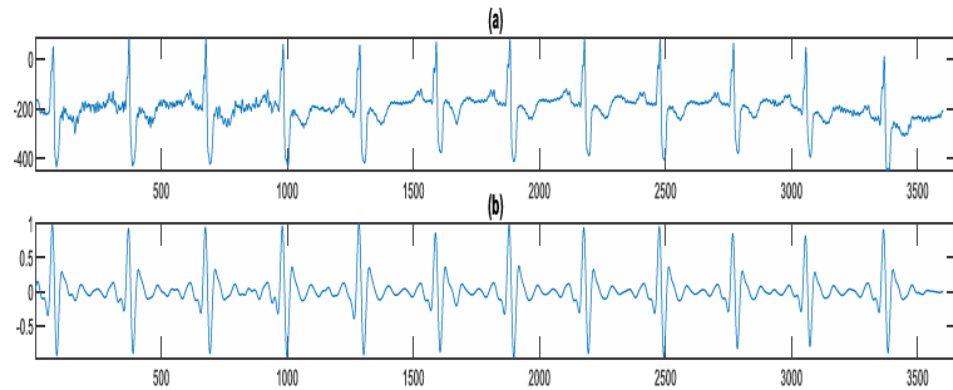


Figure 8: MIT Database (a) Raw Signal Data and (b) Filtered Data.

➤ Derivative

The electrocardiogram signal was differentiated after filtering to determine the slope properties. features which provide distinction between the QRS complex proper and adjacent waveforms. At this point low-frequency content like Reducing the significance of both P and T wave inflections, the preceding frequency content corresponding to the steeper gradients of The QRS-elements were signal-enhanced. Figure 9 shows as result of the disduction step, with particular attention to the better visualization of the waveform slope

QRS segments in the MIT dataset.

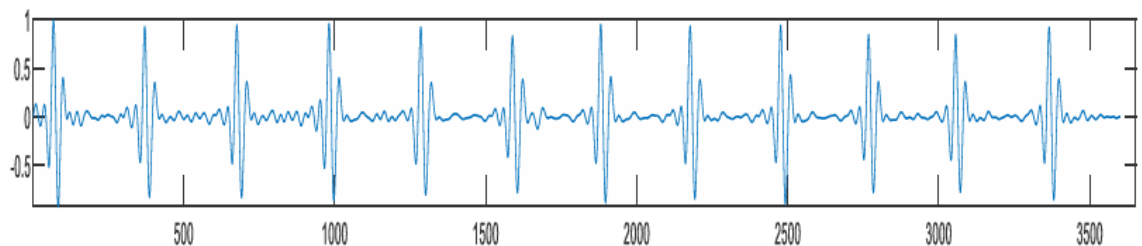


Figure 9: MIT Database (Derivative Output).

➤ Squaring Function

After differentiating, the signal was squared so that the negative and positive ranges were transformed into an equivalent positive range components of the waveform were positive. Figure 10 shows the squared signal. This process additionally accentuated the high-amplitude section associated with the QRS-complex. It will also help remove the effect of higher T-wave magnitudes which if not compensated may could lead to incidents of misclassification. The methodology is based on amplifying marked variations originating from QRS activity.

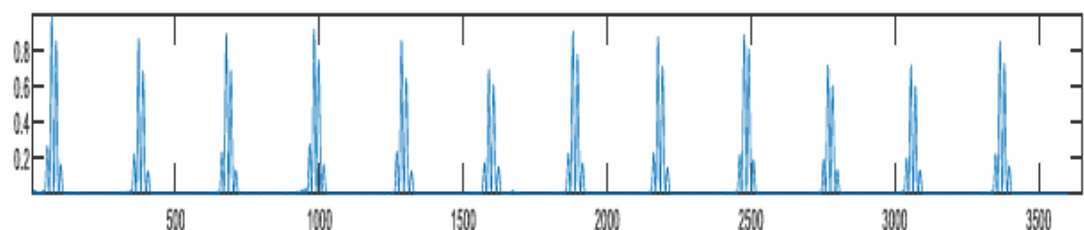


Figure 10: MIT Database (Squaring Output).

➤ Moving Window Integration (MWI)

MWI (Moving Window Integration) is used to find most meaningful characteristics of waveform and to compute the R-wave slope as part of the total processing. This part

depends largely on the selection of an integration window. For the delay to hit approximately the correct length of a QRS complex, the window must be long enough. In this instantiation, a 30-sample window (approximately 150 ms We used a sampling frequency of 200 Hz as described in [22]. The extracted R-wave slope is shown in the signal trace of Fig. 11.

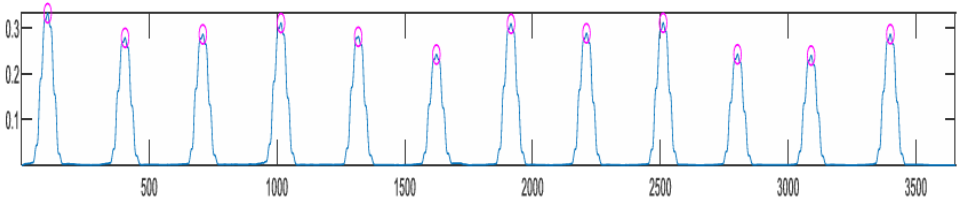


Figure 11: The Visualization of the MWI processing outputs on the MIT arrhythmia database file no 101

➤ **Decision**

Having pre-processed the signal, the next step in the analysis is a sign test. This part of the component compares the Moving Window Integration (MWI) output to pre-defined moving thresholds to confirm the existence of a QRS complex. Two individual adjustable threshold values are employed to keep the detection precise. The QRS sequences thus derived, from the MIT arrhythmia database, are displayed in Figure 12.

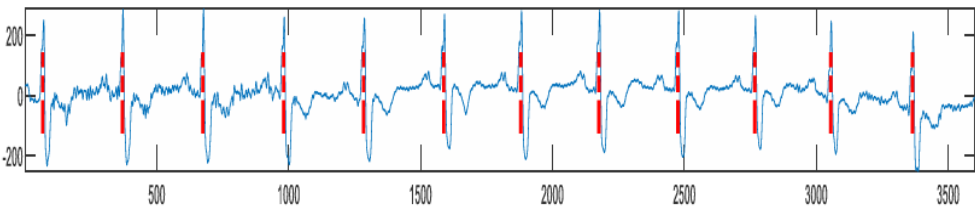


Figure 12: QRS Complex Recognition Based on MIT Data

The algorithm performance were evaluated on all 167 test ECG records of the MIT database dataset also showed high values for sensitivity and positive predictive (a) All records specially high positive predictivity and sensitivities were calculated for record 118, 100% sensitivity values and 99.96% positivitivy, and for record 119, 100% for both sensitivity and positive predictivity. An overview of experimental results is shown in Table 2. concerning measured results of the MIT-BIH test of the algorithm

Table 2: Performance of the algorithms on MIT data

Records	Se	+P (%)	Records	Se	+P (%)
100	100	100	201	96.64	100
101	99.89	99.73	202	99.53	100
102	100	100	203	97.11	99.25
103	99.95	100	205	99.74	100
104	98.70	98.57	207	98.98	99.68
105	99.46	98.27	208	98.65	99.83
106	99.85	100	209	100	99.97
107	99.25	100	210	98.45	99.85
108	99.77	83.27	212	100	100
109	99.80	100	213	99.94	100
111	99.48	100	214	99.82	99.96
112	100	100	215	99.91	100
113	100	100	217	99.82	99.91
114	99.95	100	219	100	100
115	100	100	220	100	100
116	99.05	99.87	221	99.84	100
117	100	100	222	99.07	99.07
118	100	99.96	223	99.54	100
119	100	100	228	99.66	99.56
121	99.89	100	230	100	100
122	100	100	231	100	100
123	99.80	100	232	100	99.94
124	99.32	100	233	99.90	100
200	99.85	99.85	234	99.89	100

3.2.3 Segmentation

A precise segmentation of the ECG signal at the, was the result of the R-peak recognition with the help of the Pan- Tompkins method end of the pre-processing stage. This method enabled each heart cycle to be separated individually, and every R peak was isolated and extracted "totaling 360 points on record. Maintaining a standard segment length in the" beat period at denoted time of each heart beat, variations in duration of the beat were reduced leading to better consistency and generality of the set.

3.2.4 Construction of the Beat Matrix

Extracted heartbeats were placed in an organized manner forming a 2D matrix, where each row was an individual heartbeat (360 points per beat). A time-axis corresponding to a particular time-ordered sample of all the beats was preserved for each column. An annotation guideline source data used in were the basis of as a annotation protocol for labeling, with each normal beat having a label of 0 and all abnormal beats having a label of 1.

A structured representation of the data is achieved by transforming raw ECG waveforms into a database where each row indicates a single heartbeat, enabling efficient extraction of features and subsequent classification. Such structure ensures that each beat is taken as an isolated instance, while respecting the natural temporal nature of the heartbeat cycle. This approach thus increases the reliability of downstream analytical tasks such as developing "Supervised learning models for arrhythmia detection. These models are summarized in Table 3 in terms of a 2-d matrix of heartbeats, labeled as such."

Table 3: Beat matrix with labels:

Beat index	Sample 1	Sample2		Sample360	Label
1	0.13	0.12		0.15	0
2	-0.10	-0.09		-0.12	1
3	0.08	0.09		0.12	1
4	0.11	0.14		0.19	1
5	-0.5	-0.04		-0.07	1
6	0.9	0.10		0.14	1
7	0.05	0.07		0.10	1
8	-0.12	-0.11		-0.14	1
9	-0.02	0.01		0.05	1
10	0.07	0.09		0.11	1
11	0.04	0.03		0.02	1
12	0.08	0.06		0.09	1
13	0.12	0.10		0.13	0
14	-0.13	-0.11		-0.15	1
15	0.05	0.07		0.08	1
16	0.11	0.10		0.12	1

3.3 Feature Extraction

After denoising and enhancement of raw ECG signals in the preprocessing phase, feature extraction is an important step. This is a basic step to convert the unstructured high-dimensional ECG waveform to a structured low-dimensional numeric representation which can retain maximal information embedded in the original waveform while discarding redundancy [23], [24]. Feature extraction is an essential step for types of heart beats, detection of irregular beats, and di- agnostic support in medicine by applying learning techniques.

- **Time-Domain Features** — Statistics describing the variations and distributions of amplitudes from a single beat. Such features enable a straightforward evaluation of morphology of the signal and are widely used in routine ECG analysis [25].
- **Spectral Domain Features** These features are computed from the Fast Fourier Transform (FFT), and describe the frequency content of the ECG, i.e., repetitive sequences in the rhythm and energy peaks in the signal corresponding to regular or most common rhythms, that may represent arrhythmic activity [26].

3.3.1 Time-Domain Features in ECG Signal Analysis

Time Domain Signal Characteristics — These features, which are obtained from the time domain, are widely used due to their intuitive explanation and the high capacity to capture the time-varying dynamic of the signal [23], [25].

These figures involving statistical and structural features, are derived directly from the ECG trace with no transforming the data using mathematical transforms such as the FFT or DWT. They are essential for investigating waveform shape, fluctuations and power distribution – vital for separating normal from abnormal heart beats. Essentials of time-domain features: Usefulness in arrhythmias detections Time-domain features are immensely significant features of arrhythmias because they recognize changes regarding the nature of the signals, such as dramatic oscillations, intensity differences, and distortion of the electrocardiograms (ECG) waveforms.

Types of Time-Domain Features in ECG Analysis

1. Mean (Arithmetic mean)

Mean: Helps to get an estimate of the nominal amplitude of the ECG beat which is a basic statistic that is frequently used to see if the beat is near the baseline or displaced. Any abnormal cardiac activity or baseline wander can indicate that the given value is different from the mean. Expected value.

Mathematical formula:

$$Mean = \frac{1}{N} \sum_{i=1}^N x_i$$

4 where with: In this relation, x_i denotes the size of ECG sample at position i and N is the casual number of samples.

2. Standard Deviation(SD)

SD – The Standard Deviation (SD) indicate how much the amplitude of the ECG signal fluctuates, or, at least, how wide the variation is. Higher SD values correspond to more variability, which is characteristic of an irregular heart activity, while smaller SD values are indicative of a more regular and stable heart rhythm.

Mathematical formula:

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - Mean)^2}$$

Mean is the Mean value of the ECG beat.

3. Range (Max - Min)

This attribute measures the whole wave amplitude range in a single ECG cycle by taking the difference between the maximum and the minimum of its absolute value. It is related to the contractile power of the heart—it is greater with stronger activity, and is lower in case of weaker contractions.

Mathematical formulation

$$Range = \max(x) - \min(x)$$

"with $\max(x)$ indicating the peak amplitude value of the ECG trace, while $\min(x)$ corresponds to the lowest amplitude within the same heartbeat segment."

4. Energy

To quantify energy in an ECG beat, it is recommended to calculate the total signal power by summing the squares of the amplitude values. This parameter is useful in differentiating between beats of different intensities, and is useful in the detection of abnormalities such as ischemia or ventricular fibrillation.

Mathematical formula:

$$Energy = \sum_{i=1}^N x_i^2$$

Where x_i represents the amplitude at sample i .

5. Root Mean Square(RMS)

The RMS reduces the noise by the meanquare) rectifier approach to obtain a stable representation of signal intensity that has reduced impact from baseline variability and improved accuracy for amplitude-based classifications.

Mathematical formula:

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$$

x_i is an amplitude of the ECG signal sampled at the i -th time

6. Skewness

This feature, skewness, measures the asymmetry of ECG waveforms values around their mean. Usually" a positive value of skewness indicates parties dependence and explanatoriness. thought of as a higher peak toward the positive direction, whereas negative skewness means the shapes of the waveforms flatter toward both peaks, and the negative goes Dominance. This enables the detection of irregular heart rhythms and waveform abnormalities.

Where

$$Skewness = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - Mean)^3}{SD^3}$$

Here, SD is the statistical standard deviation of the electrocardiographic waveform.

7. Kurtosis

Kurtosis is a measure of the peakedness or flatness of individual ECG waveforms and is useful for telling apart forms with various sharp or low peaks. High kurtosis stand for many peaks and low kurtosis is uniform distribution.

Mathematical formula::

In case you need the equations, here it is

$$Kurtosis = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - Mean)^4}{SD^4}$$

SD = the calculated SD of the amplitude of the ECG signal

3.3.2 Frequency-Domain Features in ECG Analysis

Characteristic frequencies are often system-wide and general used for ECG-based arrhythmia detection, providing spectral perspective in addition to time-domain characteristics of cardiac activity [24], [26].

Time domain features reflect the morphology of the waveforms in ECG analysis; however, they are less informative about spectral characteristics of the signal. The ECG signals are subjected to frequency domain analysis using Fast Fourier Transform (FFT) that converts transforming signals Changing the analysis of underlying periodic patterns and prominent spectral elements.

These types of features are particularly valuable for identifying arrhythmic behaviors that may time-domain representations can obscure this. Frequency-domain characteristics are extensively utilized in automated systems for diagnosing cardiac abnormalities, allowing algorithms to differentiate between regular abnormal heartbeat ontologies and diseases, like AFib, VT, respectively, and various other arrhythmias.

In the context of this research, the aim is to extract the most clinically significant feature from the frequency domain, which in turn can assist to improvise the accuracy and reliability in a usage for classification of heart rhythm segments.

The following section describes the main frequency-domain parameters estimated from the ECG signal.

1. Dominant Frequency (DF)

Domianent The DF, or dominant frequency, is the spectral content power of the highest power within the ECG time series. This is an important parameter when examining the rhythm of the heart, since changes in the dominant frequency, may be and may. be associated with different arrhythmic states, for example, AFib (atrium fibrillation). In normal sinus rhythm, the dominant frequency is more stable whilst in irregular heartbeats, the dominant frequency is more dispersed into the frequency domain. This dispersal is an indicator of arrhythmicity, a feature of many cardiac diseases. Mathematically, the dominant frequency is formulated as:

$$DF = \operatorname{argmax}|FFT(x_i)|$$

Where $FFT(x_i)$ and argmax are the amplitude of the Fourier-transformed ECG signal and identify the frequency where the maximal spectral power is found respectively.

2. Mean Frequency(MF)

This feature called Mean Frequency is a parameter which compute average frequency

weighted by spectral power for all frequency components of the ECG spectrum. This additionally mirrors the complete spectral allotment of the signal, making it appropriate in recognizing between typical and unusual cardiac rhythms. Conditions with high frequency such as ventricular tachycardia (rapid electrical activity in the heart) will yield a higher mean frequency, while conditions of bradycardia or other slow rhythm abnormalities will yield a lower mean frequency.

The basic formula of the mathematical structure for computing the mean frequency is as follows:

$$MF = \frac{\sum f_i P_i}{\sum P_i}$$

where f_i is the frequency of the i th spectral element, with P_i being the power at frequency f_i

3. Power Spectral Density (PSD)

In particular, the Power Spectral Density (PSD) describes how the ECG signal power is distributed among the various frequency components. This assists in visualizing the distribution of energy across a particular frequency band which is very instrumental in arrhythmic activity identification. For example, in low-frequency bands, a high PSD could indicate the presence of sinus arrhythmia or bradyarrhythmia and a higher PSD value at higher frequencies are evident in AFib and PVCs. Mathematically, PSD can be expressed as:

$$PSD(f) = \frac{|FFT(x_i)|^2}{N}$$

"N is the cardinality of the sample within the ECG cycle and $FFT(x_i)$ is the spectrum of the response corresponding to the i -th sample."

4. Total Spectral Energy (TSE)

Defined as the sum of all spectrum power values, TSE offers an estimate of the total amount of kinetic energy present within the frequency band (the full length of the band of frequencies an ECG heartbeat. Therefore, high levels of SPE may be linked to corresponds to the situation of ventricle fibrillation, when the heart's electrical activity is extremely irregular; low spectral energy correlates with ischemic states or weak myocardial contractions. Total Spectral Energy is:

$$TSE = \sum |FFT(x_i)|^2$$

where $FFT(x_i)$ refers to the frequency domain version of ECG signal resulting from Fourier Transform.

5. Spectral Entropy (SE)

Used to research randomness and complexity reflected in the electrocardiographic frequency spectrum. This capability specifies Power content is spread out in the different frequency components, as high entropy value implies that the signal energy is more equally spread across a wider frequency range, a trait of atrial fibrillation (AFib). On the other

hand, low spectral entropy signifies that the signal is more periodic and structured as in normal sinus rhythms. Mathematically, Consequently, the spectral entropy might be expressed as:

$$SE = - \sum P_i \log_2(p_i)$$

where p_i is the standardized spectral power for each frequency band

3.4 Classification by XGBoost

In this methodology, it uses a supervised learning to classify the extracted ECG heartbeats as normal and abnormal. Out of all of the classification techniques investigated, the eXtreme Gradient Boosting (XGBoost) algorithm was chosen for its strong generalization, excellent performance when working with structured and high-dimensional data, built-in regularization, and built-in feature importance analysis [27], [28].

XGBoost stands for eXtreme Gradient Boosting and it "XGBoost is an enhanced gradient boosting machine learning algorithm which is based on the principle of boosting. It incrementally builds trees each of which is a solution to the prediction errors of the tree immediately preceding it. Using second-order gradient computations, shrinkage techniques and randomly selected features in each iteration, the algorithm leads to both high computational efficiency and strong prediction performance. prediction accuracy, unlike any traditional boosting algorithms[27]. The use of random forest was driven by its nature of automatic feature evaluation, which is directly linked with the natural objective of this work, particularly to detect cardiac abnormalities based upon extracted features.

A balanced data set including 5,000 normal heartbeats and 5,000 abnormal (other than normal) heartbeats was constructed by random sampling in the model training stage, as to avoid the bias towards a specific arrhythmia type. Every heartbeat was characterized by a group of features composed "Features from both the time and frequency domain were used for modeling. The dataset was randomly divided into set for training and for testing. Training used default parameter values in the first place and then further tuning was achieved through cross-validation techniques to improve the parameter setting and model generalization ability.

3.4.1 Evaluation Metrics

For assessing how well the XGBoost `classifier can predict four commonly used evaluation metrics, i.e., Accuracy, Sensitivity (Recall), Specificity and F1-score, were used for rigorous assessment of the classifier. Such metrics offers a full assessment of the models ability to predict binary classification using imbalanced data, such as, distinguishing normal heartbeat patterns from irregular ones.

Let us define the classification results as follows:

- TP (True Positives): The total number of pathological beats that were correctly classified
- TN (True Negatives): cases in which normal cardiac signals were correctly classified.
- FP (False Positives): is the number of normal heartbeats without normal prediction result from the model prematurely.
- FN (False Negatives):" abnormal heart rhythms that are incorrectly recognized as normal. They then have the Sensitivity and Specificity metrics as the following equations.

- **Accuracy:**

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Sensitivity(Recall)**

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

- **Specificity**

$$Specificity = \frac{TN}{TN + FP}$$

- **F1-score**

$$2 \times \frac{Precision \times Recall}{Precision + Recall}$$

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

Both False Positives and False Negatives are considered by these evaluation measures into account. In the medical domain, misclassifications can translate into severe clinical implications, thus making interpretability critical [29].

The MATLAB software (R2022a) was utilized to implement the preprocessing, feature extraction, and the classification steps. MATLAB built-in signal processing toolboxes were used for Signal preprocessing, filtering and R-peak Detection using Pan-Tomkins algorithm. The feature extraction was performed by custom-written MATLAB scripts using MATLAB functions for statistical and spectral analysis. In the third stage, classification was made using the XGBoost model via the interface of python in MATLAB environment in which it uses a few external libraries. All the experiments have been performed on a PC with Intel core i7 CPU utilise and 16 GB RAM.

4.Results

4.1 Classification Results

We also performed a quantitative evaluation of the classifier using the performance measures previously defined. Table 4 shows the outcomes from testing the XGBoost model on the held-out portion of data. Table 4: Test Results on Test Data for XGBoost Model

Table 4: Performance of the XGBoost classifier.

Metric	Value(%)
Accuracy	96.20
sensitivity	95.40
specificity	97.0
F1-score	96.10

The result justifies XGBoost being an accurate and resilient technique for heartbeat classification with high sensibility in arrhythmia detection and high specificity to avoid false alarms [28], [29].

4.2 Features Importance Analysis

In addition to assessing the performance of overall classification, this study pays great attention to model interpretability by calculating the relative importances of the extracted features contributing to the decisions. One of the key advantages of the XGBoost algorithm is that it automatically computes the importance of the features during training [27], [29].

Feature importance was calculated in this study by using the Gain metric, which provides an indication of how much the presence of the feature improves decision-making by the model. much a feature contributes toward improving model accuracy across all decision trees. Gives meaningful information about the discriminative ability of single feature for ECG beat classification.

Figure 13: Important features in XGBoost model Identifying the most influential features for classification aided by Gini impurity: Frequency-domain descriptors like Dominant Frequency and Spectral Entropy had the best Gain scores, indicating that they strongly impact classification performance. Standard Deviation and Energy, time-domain features, were closely behind and were also equally important.

These outcomes in an evident manner suggest the advantage of utilizing a time-frequency combined feature representation that measures both structural and temporal dynamics of ECG signals. First, this result corroborates the central tenet of this study: that high classification performance is achievable via well-designed features without necessarily requiring complex deep learning models or direct ingestion of raw signals.

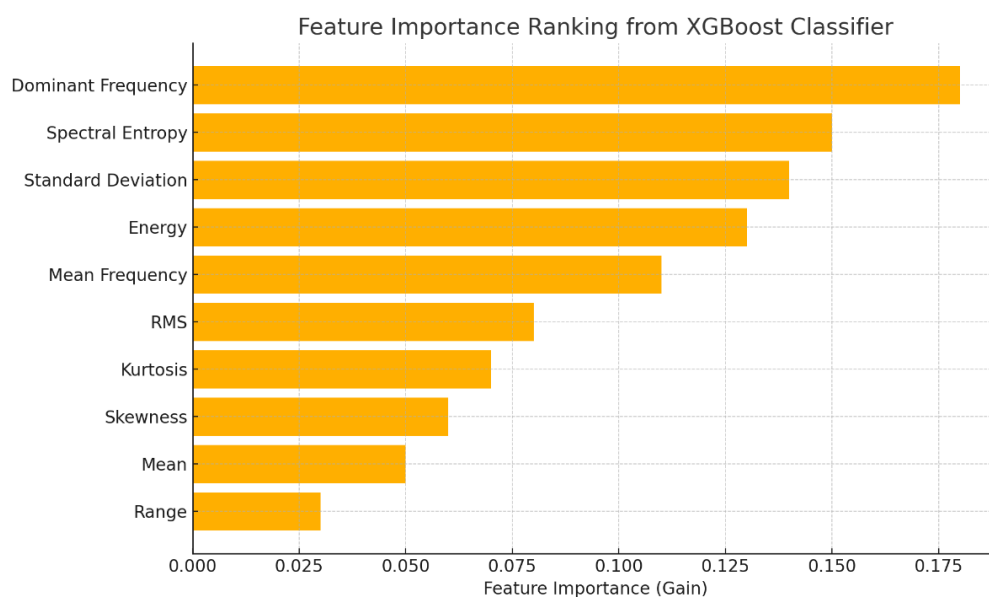


Figure 13: Feature Importance Ranking using XGBoost Classifier.

5. Discussion

This study proposes a fast and efficient approach towards automated detection and classification of heartbeats for ECG signal. using the handcrafted features and the XGBoost algorithm. The approach combines the temporal and the spectral features estimated from the segmented ECG cycles and utilizes a bagging-based classification scheme. The resulted model achieved with a reported metrics of 96.20% accuracy, 95.40% sensitivity, 97.00% specificity and F1 of 96.10%

These results further support the concept of good statistical features being a solid and interpretable replacement for large deep learning models, particularly when computation speed and model explainability is important. This fusion of temporal and spectral

descriptors allowed the model not only to capture morphological features reflected by the HRV signal but also metrics of cardiac rhythmicity which are both important information for arrhythmia detection.

Interestingly, for feature importance analysis, only a small number of features drove the model predictions. Dominant Frequency and Spectral Entropy were the most important frequency-domain-based features, whereas, for time-domain metrics, Standard Deviation and Energy were the two most influential ones. The study also demonstrates how this knowledge can aid in developing less invasive, clinically-relevant feature sets to guide new diagnostic paradigms or surveillance solutions in real time.

The model has also a balanced performance as its sensitivity and its specificity are high. Crucially, this provides the appropriate balance of true positive detection of arrhythmia with low false positive rate (FPR) for normal rhythm - a crucial requirement for clinical decision-support efforts to gain the trust of doctors, and to reduce notification fatigue.

Although the model achieves a classification accuracy of 96.20%, some new" works shows the better performance. [30], especially up to this point were studied: deep learning architectures are used. Nonetheless, the methodology presented in this paper achieves a good compromise between accuracy, interpretability and computational effort, which can facilitate its use in real-time diagnostic systems in resource limited environments.

Table5: To illustrate the efficiency of the proposed model, a comparative performance analysis with existing works is performed. 1-w 92.1 % Apple-f 30.6 % This section sheds light on how our model compares to other models, with the results summarized in Table 5.

Table 5: Comparison of classification performance with previous studies.

study	method	Accuracy %	Sensitivity %	Specificity%
Kishore et al. [8]	RBFNN & Genetic Algorithm	98.50	98.30	99.00
Karthiga et al. [9]	CNN	91.92		
Madan et al. [10]	CNN-LSTM & CWT	97.30	97.00	98.00
Ting et al. [11]	EfficientNet-B0 & STFT	97.30	89.55	89.55
Alfaras et al. [12]	Echo state Networks	96.80	92.70	
Bhatia et al. [13]	CNN-LSTM	98.36		94.36
He et al. [14]	Residual CNN & Bi-LSTM			
Proposed Study	XGBoost (handcrafted)	96.20	95.40	97.00

6. Conclusion

This work highlights that a hand-crafted feature based classifier can reliably identify arrhythmic signals in ECG signals. This is done by combining features extracted from both domains of the time scale and frequency. domain features using XGBoost algorithm, our model is able to achieve high accuracy as well as interpretation ability without the complexity of deep learning model. We find that the minimal set of the informative features is nough which means the distinguishing between normal and abnormal beats can be done with a few features including Dominant Frequency, Spectral Entropy, Standard Deviation, and Energy. This information is useful in developing lightweight, but still clinically relevant, diagnostic models.

To conclude, the method proposed exhibits a high-performance, interpretable, and computationally efficient arrhythmia classification methodology. Balancing performance

and interpretability makes it suitable to be transferred to be integrated into the real-time health care applications and wearable platforms for continuous monitoring of the heart activity.

Future Work and Challenges:

While the proposed approach performs strongly, future work has a number of challenges and opportunities. One challenge is to generalize the model to wider and more diverse range of patient cohorts, including patients with rare arrhythmia phenotypes. Furthermore, the current model is limited to binary classification and extending applications to multiclass classification for different arrhythmia types would greatly enhance its clinical applicability.

Further studies could investigate the addition of additional physiological signals (e.g., blood pressure, SATO2) to increase diagnostic accuracy. In addition, deploying the model to embedded systems or wear-able devices may still demand more optimization, with field experiment on real-time environments. Meeting these demands is crucial for enabling robust and clinically applicable automated ECG analysis systems.

References

1. World Health Organization, "Cardiovascular diseases (CVDs) Fact Sheet," 2019. [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
2. G. D. Clifford, F. Azuaje, and P. E. McSharry, *Advanced Methods and Tools for ECG Data Analysis*, Norwood, MA: Artech House, 2006.
3. J. Higuera, F. Roldán, and A. Vázquez, "Accuracy of ECG interpretation among physicians: a multi-level study," *Journal of Electrocardiology*, vol. 55, pp. 78–84, 2019.
4. U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated ECG diagnosis: challenges and future directions," *Computers in Biology and Medicine*, vol. 87, pp. 1–10, 2017.
5. M. Alghamdi, N. Elhadad, and S. M. Hammad, "Automatic detection of arrhythmia using ECG signals with machine learning," in *Proc. 2021 Int. Conf. on AI in Biomedicine*, pp. 123–130.
6. P. Kligfield et al., "Recommendations for the standardization and interpretation of the electrocardiogram," *Journal of the American College of Cardiology*, vol. 49, no. 10, pp. 1109–1127, 2007.
7. B. J. Drew, M. A. Califf, D. W. Funk, et al., "Limitations of computer-based ECG interpretation," *New England Journal of Medicine*, vol. 351, no. 4, pp. 287–289, 2006.
8. K. Kishore, B. V. Kumar, and R. S. Bhuvaneswaran, "Arrhythmia classification using radial basis function neural network with a novel feature selection algorithm," *Computers in Biology and Medicine*, vol. 103, pp. 179–191, 2018.
9. R. Karthiga and A. Kalaivani, "An efficient ECG arrhythmia classification using deep convolutional neural network," *Journal of Medical Systems*, vol. 43, no. 10, pp. 1–10, 2019.
10. A. Madan, R. Yadav, and P. K. Kalra, "ECG beat classification using hybrid deep learning model," *Biomedical Signal Processing and Control*, vol. 68, p. 102595, 2021.

11. L. Ting, Y. Zhang, and X. Li, "Arrhythmia detection using STFT-based spectrogram and EfficientNet," in *Proc. IEEE Int. Conf. on Biomedical Engineering and Informatics (BMEI)*, 2021, pp. 215–220.
12. M. Alfaras, C. Soriano, and R. Orts, "Real-time cardiac arrhythmia classification using Echo State Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 556–566, 2020.
13. S. Bhatia, R. Arora, and N. Gupta, "A hybrid CNN-LSTM model for arrhythmia classification," *Procedia Computer Science*, vol. 167, pp. 2410–2418, 2020.
14. J. He, Y. Wu, and H. Zheng, "Multi-scale residual BiLSTM network for ECG classification," *IEEE Access*, vol. 9, pp. 91115–91124, 2021.
15. G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, May–June 2001.
16. G. B. Moody and R. G. Mark, "The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it," in *Computers in Cardiology 1990*, pp. 185–188.
17. "MIT-BIH Arrhythmia Database v1.0.0," PhysioNet. [Online]. Available: <https://physionet.org/content/mitdb/1.0.0/>
18. M. Elgendi, "Fast QRS detection with an optimized knowledge-based method: Evaluation on 11 standard ECG databases," *PLoS ONE*, vol. 8, no. 9, p. e73557, 2013.
19. G. D. Clifford, F. Azuaje, and P. E. McSharry, *Advanced Methods and Tools for ECG Data Analysis*, Artech House, 2006.
20. M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sörnmo, "Clustering ECG complexes using Hermite functions and self-organizing maps," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 838–848, 2000.
21. J. S. Sahambi, S. N. Tandon, and R. K. P. Bhatt, "Using wavelet transforms for ECG characterization: An on-line digital signal processing system," *IEEE Engineering in Medicine and Biology Magazine*, vol. 16, no. 1, pp. 77–83, 1997.
22. J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, 1985.
23. U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81–90, 2017.
24. G. D. Clifford, F. Azuaje, and P. E. McSharry, *Advanced Methods and Tools for ECG Data Analysis*, Artech House, 2006.
25. M. M. Gul, R. Anwar, and M. S. Khan, "Classification of ECG signals using statistical and spectral features for cardiac arrhythmia detection," in *Proc. IEEE Int. Conf. on Engineering and Emerging Technologies*, 2020.
26. S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2016.

27. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
28. A. M. Al-Ameen, M. M. Islam, and M. R. Islam, "ECG classification using machine learning techniques with XGBoost," in *2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, pp. 219–224, IEEE, 2020.
29. S. Wang, L. Wu, and H. Sun, "Automatic classification of arrhythmias based on deep features of ECG signals using XGBoost," *Biomedical Signal Processing and Control*, vol. 57, p. 101819, 2020.
30. Y. Xue, T. Zhang, and Y. Yang, "An Automatic Arrhythmia Classification System Based on Deep Learning and High-Resolution ECG Signals," *IEEE Access*, vol. 7, pp. 171455–171464, 2019.