A Novel Framework for Grading of Heart Attack

T. M. Rajesh(✉), M. N. RenukaDevi, S. G. Shaila, and CauveryRaju

Department of Computer Science and Engineering, Dayananda Sagar University, Bangalore, India

{rajesh-cse,shaila-cse,Cauvery-cse}@dsu.edu.in

Abstract. In recent years, cardiovascular diseases have become common. Serious health problems arise in the human body as a result of an unhealthy lifestyle, the use of alcohol and tobacco, obesity, stress, and dietary changes. This has complicated surgeons' ability to diagnose heart failure at the right time. A heart attack occurs when the blood flow that brings oxygen to the heart muscle is severely condensed or cut off completely. ECG is a medical test that is used in the detection of heart attacks in patients. Extracting the essential features from ECG images is the most crucial task. The key features are extracted using connected component analysis, hierarchical centroid, Hough line transform, and height and width. Various techniques like Fast Fourier Transform, Discrete Fourier Transform, Decision Tree and Principal Component Analysis are used to predict heart failure. In this model, we are going to examine ECG signal images and detect whether the person is prone to heart attack or not. A comparative study of different models showed that the proposed work enhanced the previous accuracy score in predicting heart failure using FFT.

Keywords: Fast Fourier Transform · Discrete Fourier Transform · Decision Tree · Principal Component Analysis

1 Introduction

The heart is mainly responsible for pumping blood and circulating oxygen and nutrients all over the body. The heart is considered one of the most essential organs in the body. Any irregularity or abnormality may cause extreme changes or effects in other parts of the human body. Over the last decades, cardiovascular diseases (CVDs) have contributed to the death rate all over the globe. As per the records of the World Health Organization (WHO), an approximate 17.9 million individuals died due to CVDs in 2019, showing 32% of overall deaths [1]. Of these deaths, 85% were due to strokes and heart arrest. The symptoms of heart attack and stroke depend on family medical history, gender, age, and total health condition. The symptoms can arise rapidly and without notice. The signs of a heart attack may include sweating, breathlessness, chest pain, uneasiness, nausea, wooziness, fatigue. There are various indications that can happen during a heart attack, and signs can vary amongst women and men. Heart failure may happen due to smoking,

S. Tamane et al. (Eds.): ICAMIDA 2022, ACSR 105, pp. 320–339, 2023.
https://doi.org/10.2991/978-94-6463-136-4_28

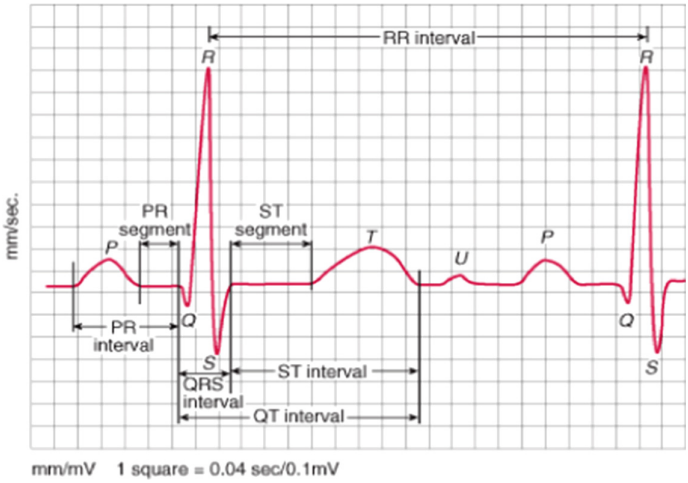


Fig. 1. Pictorial Representation of ECG Signal (Source: life in the fastlane)

no physical activity, alcohol, and a high intake of oily foods, which leads to hypertension. Aside from leading a healthy lifestyle and eating a nutritious diet, early detection can help prevent heart problems. An electrocardiogram can help (ECG) is used to diagnose different kinds of heart diseases. An ECG indicates the electrical movement of the heart muscle as it fluctuates with time. The ECG signal consists of a QRS complex, a T wave and a P wave. Figure 1 shows the representation of an ECG signal in the heart.

A typical ECG paper permits an estimated approximation of the heart rate from an ECG demo. Every second of time is characterized by 250 mm, i.e., 5 large squares along the parallel axis. If the number of big squares among every QRS complex is 2, the value of 2 indicates the heart rate is 150 per minute, the digit 3 signifies the heart rate is 100 per minute, and the number 5 denotes the heart rate is 60 hits per minute.

The proposed work involves three phases, such as preprocessing the input image, feature extraction, and classification to detect heart failure using machine learning algorithms. The first stage is to collect the ECG image, which is converted to a binary image. Noise is unwanted information that is present in the image. Noise removal is done using linear and non-linear filtering techniques. A non-linear filter such as the Median Filter replaces the pixel value with its median values instead of the mean of neighboring pixels. The purpose of this is to eliminate the salt and pepper noise in ECG images. Binary images are comprised of pixels that have only two values; 0 for black and 1 for white. Image resizing is done once the noise is eliminated from the ECG image. Resizing is carried out to increase or reduce the total number of pixels. In the feature extraction phase, significant features need to be extracted from the image. Techniques like Connected Component Analysis, Hierarchical Centroid, Hough Line Transforms, and Height and Width are used to extract relevant information. In CCA, it examines every pixel from leftward to rightward and top to bottom in order to recognize connected pixel regions, i.e., regions of neighboring pixels that distribute similar sets of intensity values. The Hough transform can be used to detect regular curves like circles, ellipses, and lines. To find height and width, vertically, height is calculated using the voltage of a given wave. Horizontally, width represents the time of the given wave.

The final phase is classification. Different techniques like Fast Fourier Transform (FFT), Discrete Fourier Transform, Principal Component Analysis (PCA) and decision trees are used to detect heart failure. FFT is used to take out the feature constituents such as PQRST signals from the ECG. FFT will use the Discrete Fourier Transform (DFT) to examine a signal. A supervised learning technique, the Decision Tree, is used for both regression and classification problems. A comparative study of various models showed that proposed work FFT enhanced the accuracy score in predicting heart failure.

2 Related Work

The ECG signals comprise different noises like electrode motion artefact noise, powerline interference, electromyographic (EMG) noise, and baseline wander. Baseline wander occurs due to patients' movement and improper electrodes. A cutoff frequency in the range of 0.5–0.6 Hz of a high-pass filter is used to eliminate it. Powerline interference occurs due to noise from the main supply. It overlays the low-frequency ECG waves like T and P waves. A notch filter with a cutoff frequency of 50–60 Hz can be used to detach the superimposition of waves. A high frequency noise of above 100 Hz is termed "EMG noise" and is eliminated by passing through a low-pass filter. Electrode motion is mainly due to stretching of skin artifacts. It can be suppressed by adaptive filters [2]. The authors used the dataset from the Kaggle platform, which is identified as the Heart Disease Dataset. The researchers considered 14 attributes that are essential to identifying heart disease. The following attributes are considered: RestBP, chestpain, sex, age, Thal, Target, OldPeak, slope, heart beat, Exang, cholestrol, Fast Blood Sugar, and CA. They used five models, such as Support Vector Machine, Nave Bayes, Random Forest, Decision Tree, and Logistic Regression. The accuracy of the decision tree was high, and Naive Bayes showed the lowest accuracy. In the previous work, different tools like RapidMiner, Matlab, and Weka were used. The author chose Rapid Miner. It is a data mining tool that provides functionalities to implement clustering and classification problems. The accuracy was improved; the decision tree and SVM gave better results [3]. The authors have used IOT Device pulse sensor to detect pulse rate along with other parameters such as current smoker, gender, fast blood pressure, age, cholesterol, resting blood pressure and multiple regression for prediction of heart problems. In multiple regressions, there will be one dependent variable and more than one independent variable. The pulse sensor is connected to Aurdino board and pulse rate is stored in dataset. After prediction, the status of the health is sent to the person via message [4]. The authors developed a system to detect whether a patient is susceptible to heart failure or not using a two-class boosted decision tree, which gives a probability of which the person is heart prone. They used a dataset of nearly 10k patients, which considers parameters like total minutes of exercise per week, average heart beats per minute, cholesterol, age, sex, body mass index, family background history, smoking habits from the past 5 years, and palpitations per day. If the probability is greater than or equal to 50%, the ECG data is passed to the CNN model. A support vector machine is used to identify the type of heart attack and achieved an accuracy of 84%, while an artificial neural network showed an accuracy of 88.30% [5].

The previous study involves the examination of ECG signals. The authors have considered length of the segment, distance between PQ, QR, RS, ST & PT, time interval

between heart beats. In addition to this, PQ & ST intervals are considered for identification of heart disease. Naive Bayes Classifier is trained with signals with and without heart disease. Accuracy was found to be 99.20% [6].

The proposed work aims at reducing overfitting problems faced by machine learning algorithms. Overfitting is defined as any algorithm/model that performs well in the training phase but not in testing, where the generalisation error is high compared to the testing error. The dataset used in the research was Cleveland, and 13 features were considered. The proposed system is the RSA-RF (Random search algorithm-Random forest) model for heart failure prediction. In the paper, RSA selects the finest subset of features [7].

In the paper, the author calculates the heart rate of the patient, and the database used is the MIT-BIH arrhythmia database. ECG signals are susceptible to various kinds of noise due to electrical instruments used, movement of the patient, etc. ECG signals contain noises like baseline wandering and electrode motion artefacts. The authors were successful in removing baseline wander using the wavelet approach. To estimate the heart rate of a patient, at higher amplitude, the QRS complex has a fluctuating frequency compared to additional waves of the signal. The proposed technique predicts heart rate using time-frequency analysis and has an accuracy of 97.54 percent. The authors were successful in classifying different types of arrhythmia [8].

The authors used the dataset from UCI machine learning. Various techniques like Naive Bays, Logistic Regression, Decision Trees, and Random Forest are used to predict heart problems. A confusion matrix is generated for all the machine learning models. Different metrics were considered for performance measures like accuracy, recall, F-score, specificity, and precision. Random Forest gave a better accuracy of 90.16% [9]. This is depicted in Table 1 and Fig. 2 below.

3 Proposed Methodology

A novel approach for grading of heart attack:

Our proposed method comprises four important stages. Pre-processing of the dataset is the first step. The next stage is feature extraction, and the final stage is classification, as shown in Fig. 3.

1. **Pre-processing:** In this step, we perform de-noising of the image and resize the given image as per the conditions.
2. **Feature extraction:** In this step, features are mined by following these steps.

    - Identifying the length and width of the ECG signal and also calculating the distance between the peaks in the ECG signal
    - We are comparing the methods using connected component analysis, hierarchal centroid, and the Hough line transform.

3. **Classification:** After training we will be classifying the given ECG graph to identify grading of heart attack. We are using Fast Fourier Transform, Discrete Fourier Transform, Principal component analysis and Decision tree.

Table 1. A Study on different techniques used by several researchers.

Sl no	Title	Methodology used	Accuracy (%)	Limitations/Future work
1	Implementation of Machine Learning Model to Predict Heart Failure Disease(2019)	Naive Bayes, Decision Tree, Random Forest, Logistic Regression, Support Vector Machine	Decision Tree with accuracy of 93.19.	The dataset has inadequate amount of patient's details.
2	A System to detect Heart Failure using Deep Learning Techniques(2019)	Support Vector Machine, ANN	Support Vector machine with an accuracy of 84 and artificial neural network showed 88.30 accuracy.	The work can be carried out to identify various heart diseases.
3	Identification and Classification of Heart Beat by Analyzing ECG Signal using Naive Bayes.(2019)	Naive Bayes.	99.20	Further work can include diagnoise more complicated level of the disease.
4	An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection(2019)	Random Search Algorithm + Random Forest	93.33	The accuracy of the algorithm can be improved by considering more significant features.
5	Analysis of Different Heart Rate Monitoring and Pre-Processing Techniques for ECG(2020)	Time frequency analysis	97.54	Author can further classify ECG signals into various classes of arrhythmia.
6	Heart Disease Prediction using Machine Learning(2022)	Naive Bayes, Logistic Regression, Decision Tree and Random Forest	Random Forest with an accuracy of 90.16	The work can be further improved by considering real data set for the study to get better result.

The stages involved are explained in Fig. 4:

1. Training

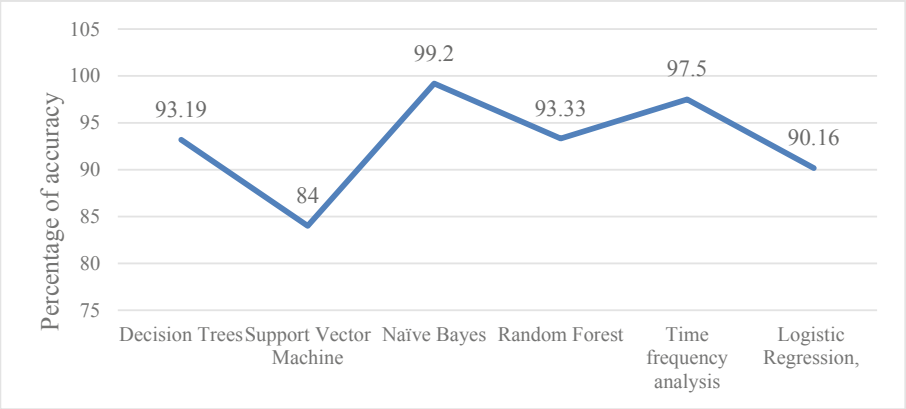


Fig. 2. Comparison of the accuracy of the different models

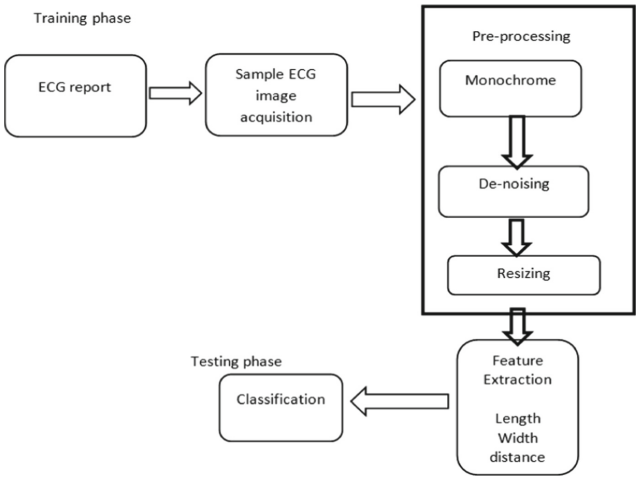


Fig. 3. Block diagram of proposed method

In this phase, all the input ECG images undergo a training process to model the features extracted, and they are kept in the database.

- Acquiring Images: In our study, we considered the ECG dataset from the Kaggle platform and real datasets from hospitals.
- Image preprocessing: image preprocessing is very essential to remove undesired noises and distortions in images, which improves the quality of the images for further processing.
- Feature Extraction: Extracting key features for further analysis is crucial for grading heart attacks.
- Training: Training the given set of input ECG images for the selected key features which are given as input to the classifier.

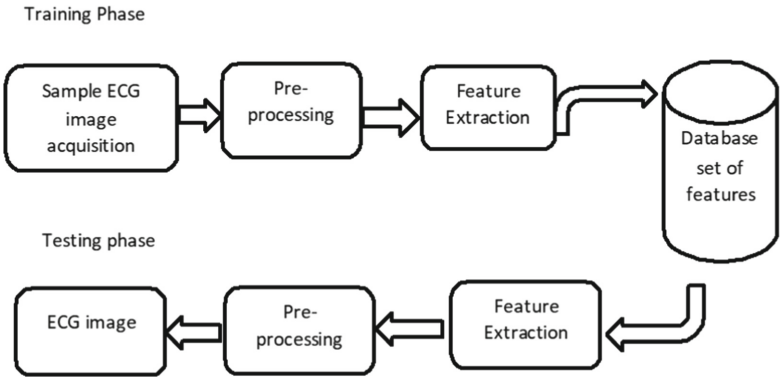


Fig. 4. Block diagram of Training and Testing phase

- Classification: in this step, the classifier detects a heart attack in the patient from given ECG images.

2. Pre-Processing of images.

- The preprocessing stage involves converting the given input ECG image to a monochrome image as shown in Fig. 4. A monochrome image refers to an image displaying different shades of a single color. Monochrome photos include black and white photography, which consists of varying shades of neutral grey. In research studies, monochrome images are used for artistic and aesthetic purposes. In Fig. 5, we convert the RGB ECG image to a monochrome image before processing, since this has more advantages.
- We need 24 bits to save the sole colour pixel of an RGB image, but only 8 bits are required to store a single pixel of the image when we convert RGB to monochrome. So, we require 33% less memory to store grayscale images than to store RGB images.
- Monochrome images are easier to work with compared to RGB images since for segmentation of images and operations related to morphology, it requires less time to work with greyscale images, which are single-layered compared to three-layered images, which are RGB. This is depicted in Fig. 6 below.
- Features of the ECG graph can easily be extracted using single-layered images [10].

4 Experimental Results

We have plotted ECG Monochrome image into graph as shown in Fig. 7. Then again it undergoes into processing called binarization where it removes the grid present in the image as indicated in Fig. 8.

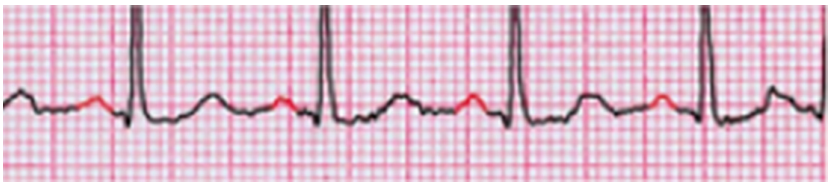


Fig. 5. Normal ECG image as input.

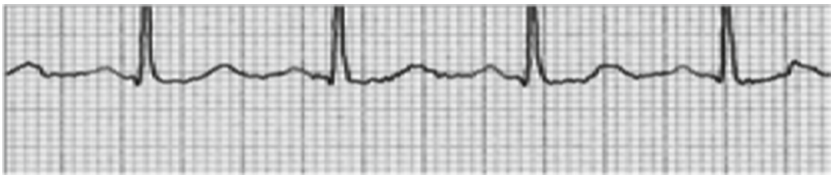


Fig. 6. Converted Monochrome image

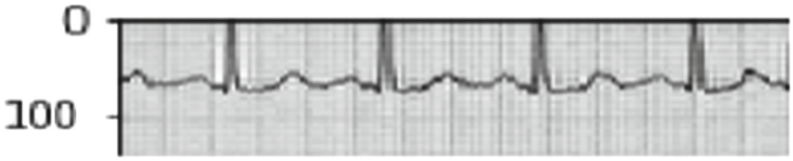


Fig. 7. Plotted in to graph

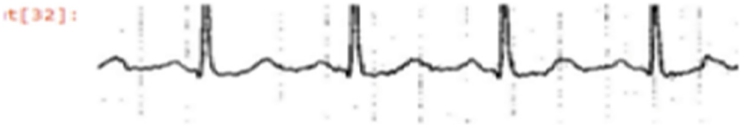


Fig. 8. Binarized image with noise

4.1 DE Noising

Noise removal is a fundamental problem in signal processing. Noise is unnecessary data in the images, which weakens our main results to a greater extent. Images are often degraded by noise in the acquisition stage. Image de-noising is used to remove the additive noise in the image and retain the most important key features required for further processing. Image de-noising techniques are very much required to prevent corruption from digital images of the ECG graph to predict the heart attack grade.

We are using a median filter for removing salt and pepper noises in ECG images, as depicted in Fig. 9. The median filter considers each pixel in the image in turn and looks at its close neighbours to decide whether or not it's descriptive of its surroundings. It replaces the pixel value with its median values, instead of the mean of neighbouring pixels [11].

Fig. 9. Noise removed image

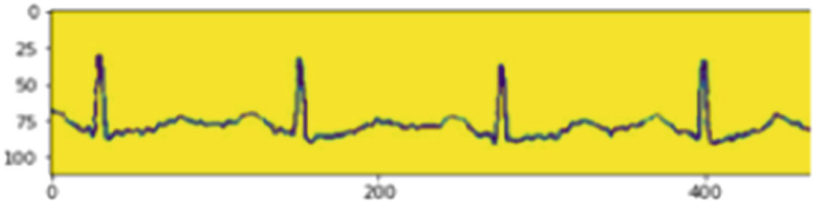


Fig. 10. Input image

Median filter is a nonlinear digital filtering technique to remove noise from the images. This type of noise reduction technique is distinctive pre-processing step to progress results for later processing of ECG images. If we are using 2D ECG images median filter for images can be developed as follows.

$\mathrm{M(k)} = \mathrm{medw(k)} = \mathrm{med}\{x_{-n}(\mathrm{k}), \ldots\ldots, x_{-1}(\mathrm{k}), x_0(\mathrm{k}), x_1(\mathrm{k}), \ldots\ldots\ldots x_n(\mathrm{k}$ [11].

Results:

From this we will crop the region of interest.

4.2 Resizing of the Images

Image resizing is required when we need to upsurge or reduce the entire amount of pixels. Zooming refers to upsurge the amount of pixels so that when you zoom an image we can see a more detailed image. Shrinking refers to reducing the number of pixels to recover the lost information in the image. It involves method of finding right pixels to be discarded.

In our study, we are resizing ECG images based on our processing steps. We perform image zooming when we want to increase quantity of pixels in the ECG images. It involves 2 steps.

- Formation of new pixel localities and assigning the grey level to those new localities.
- We consider the closest pixels in the original image to accomplish grey level assignment for any point in the overlay, and then assign its grey level to the new pixels in the grid.

Figure 10 shows the Input image of ECG graph and Fig. 11 depicts the resized zooming image.

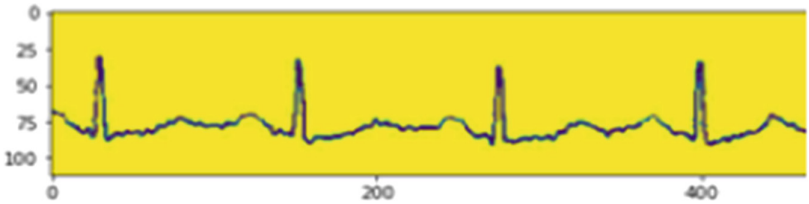


Fig. 11. Resized zooming image

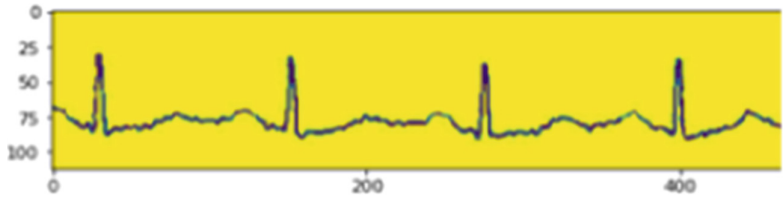


Fig. 12. Output of image shrinking

4.3 Image Shrinking

Shrinking of the image is done in similar manner as defined for zooming. Here we perform row and column deletion which is the corresponding process of pixel duplication. When we want to perform shrinking of image by one-half, we remove every row and column [12] as shown in Fig. 12.

4.4 Feature Extraction

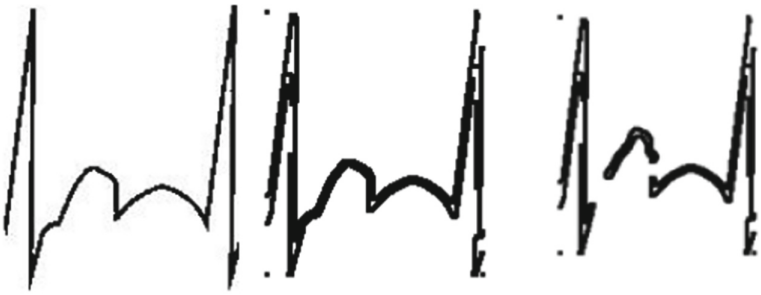
Features to be extracted from the ECG input images. We need to identify the key pixels which are required for feature extraction which consists of necessary significant information which is required for further processing. For feature extraction we have used cross connection of the following algorithms.

4.4.1 Connected Components Analysis

Labeling tests an ECG image and clusters its pixel in to components based on pixel connectivity. In connected components, all pixels are connected in some way and they share similar pixel intensity values with each other. Set of connected components partition an image in to segments. A pixel $a \in b$ is supposed to be connected to $c \in b$ if there is pathway from a to c containing completely of pixels of b. A component labelling algorithm calculates all the connected components in an ECG image and allocates a distinctive label to all points in the similar component [13].

4.5 Recursive Algorithm

Consider the case when the background pixels are 255 and the pixels in the region are 0.



Input Image **Output Image** **Output Image**

Fig. 13. Outputs of images obtained after applying the algorithms

1. Scan the ECG image for a region that isn't labelled (0) pixel and assign a new label to it.
2. A recursive label is applied to all of its nearby 0s.
3. Stop the recursive procedure when we come across unlabeled 0 pixels.
4. Return to step 1.

4.6 Sequential Algorithm

This process requires 2 iterations. It works with 2 rows of an image at a time.

1. Scan the image up and down and left to right.
2. If there is 0 pixel then,

    i. Copy the label only if one of its higher and left neigbour pixel has a label.
    ii. They copy the label if both have same label.
    iii. We need to enter the labels in the equivalent tables and copy the upper labels, if both pixels have dissimilar labels

3. Repeat step 2 if there are no more pixels further to considers.
4. If there are no pixels for further to consider, then repeat step 2
5. Find the last label for every corresponding set in the table
6. Test the entire image; change each label by the last label in the set.

Results

The Outputs of images obtained after applying the algorithms id depicted in Fig. 13 and Table 2 below.

4.7 Hierarchical Centroid

In hierarchical approach, cluster is defined by a data point, and joins remaining clusters at every stage. In this method, distance between two mean vectors of the clusters is nothing but distance between two clusters. Finally, we combine 2 clusters which have smallest centroids at each stage of the process.

Table 2. Comparative values for normal and abnormal node beat and connected components

Normal node	3	
Abnormal node	5	
Connected Components	3 for a. output image 5 for b. output image	
Co-ordinates	Normal node beat	Abnormal node beat
X1	1	1
X2	160	157
Y1	1	1
Y2	132	123

$X_1X_2.....X_N$ Interpretations from cluster 1
$Y_1Y_2.....Y_N$ Interpretations from cluster 2
$d(X, Y)$ distance between object with interpretations from cluster 1 and cluster 2

4.7.1 Centroid Method States

$d_{12} = d(\bar{X}, \bar{Y})$ Indicates finding distance between the 2 centroids and mean vector location for each cluster [14].

Results
If node1 = node2, then $d_{12} = d(0.\ 0\ )$
If node1 $\neq$ node3, then $d_{13} = d(0.\ 192\ )$

4.8 Hough-Line Transform

Using this technique, we may do feature extraction in image processing and use a selection strategy to find instances of objects that aren't perfect inside a specific class of forms. To isolate the features in the image, the Hough-line transform expects desired features to be provided in a precise form, and it is also used to detect regular curves such as circles and ellipses. This is depicted in Fig. 14 below.

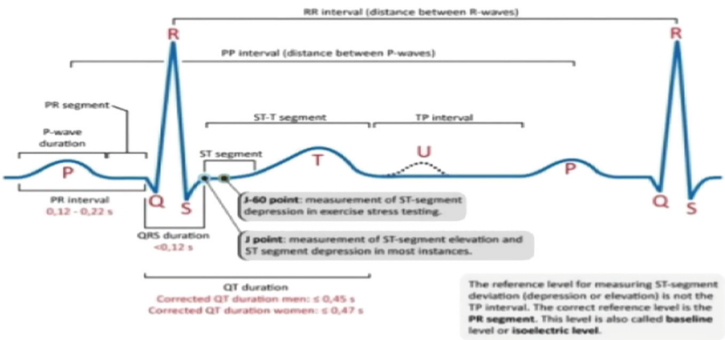


Fig. 14. Feature extraction based on Hough-line transform

Table 3. Comparative values for normal and abnormal beat

Normal	Abnormal
Node 1-length = 28	Length = 27
Max length = 85	Max length = 56
Width value = 175	Width value = 175
Depth = 6	Depth = 6
Height value = 240	Height value = 240

Let a be the vector carrying the analytic equation's parameters, and x be the vector containing the image's parameters. As shown in Eq. 1, f(x,a) is a line with an equation, also known as analytic expression.

$$\mathrm{S} = \mathrm{x}x_1\cos\theta + x_2\cos\theta \tag{1}$$

Then a = [sθ] and .x = [X_1X_2]

- To begin, quantify the parameter space inside the parameters a's boundaries. The number of parameters in the vector a determines the dimensionality n of the parameter space.
- By setting all values in the od structure to 0, we may create an n-dimensional accumulator array A(a) with the same structure as the quantized parameter space. If (x,a) = = 0,

  Increase all accumulator cells A(a).

$$\mathrm{A(a)} = \mathrm{A(a)} + \Delta\mathrm{A}$$

- Local maxima in the accumulator array A(a) correspond to curves f(x,a) in the original image [15].

  This is depicted in Table 3 below.

4.8.1 Height and Width

The height (amplitude), or voltage, of a specific wave or deflection is measured vertically on the ECG graph. The breadth of the ECG graph is measured horizontally, i.e. it records the time of the supplied wave. Calculating the height and width of a heart disease patient and comparing it to a normal ECG report can also help determine whether the heart attack is minor or serious. This is depicted in Table 4 below.

Table 4. Comparative values for normal and abnormal node height

Method	Normal node	Abnormal node
Node height	175	174
Node width	240	238

4.9 Classification

After training, we are using classifiers to classify the given input ECG image as heart attack detected or not.

The system strategy generally comprises of

1. Acquiring images
2. Pre-processing of images
3. Segmentation of images
4. Feature extraction
5. Training
6. Classification using FFT (Fast Fourier Transform), DFT (Discrete Fourier Transform), PCA (Principal Component Analysis) and Decision trees.

4.9.1 Fast Fourier Transform (FFT)

The feature points in ECG signals such as PQRST wave amplitude and wave function are mined using the Fast Fourier Transform. To remove lower order harmonics from a signal, it must be split into samples, and the input signal must be periodic, which is nothing more than the addition of different frequencies of sinusoidal signals [16].

FFT stands for Fast Fourier Transform, which is a technique for extracting relevant data from statistical aspects of an ECG signal. T 0 is the period of a periodic signal f(t) that can be seen using the Fourier series, as shown in Eq. 2.

$$\mathrm{f(t)} = A_0 + \frac{1}{2}\sum_{n=-\infty}^{\infty}(a_n - jb_n)e^{j2\pi nt/T_0} = \sum_{n=-\infty}^{\infty} a_n e^{j2\pi nt/T_0} \tag{2}$$

α_n Representing complex coefficients of the Fourier series can be represented in exponential form in Eq. 3 (Table 5).

$$\alpha_n = \frac{1}{T_0}\int_{\frac{T_0}{2}}^{\frac{T_0}{2}} f(t)e^{\frac{-j2\pi nt}{T_0}}\,\mathrm{dt} \tag{3}$$

$$n = 0, \pm 1, \pm 2$$

Table 5. Confusion matrix for FFT

	Normal (grade1)	Normal (grade2)	Abnormal (grade1)	Abnormal (grade2)
Normal (grade1)	76.37%	17.35%	6.28%	76.37%
Normal (grade2)	5.03%	71.54%	23.43%	5.03%
Abnormal (grade1)	6.69%	13.74%	79.57%	6.69%
Abnormal (grade2)	0	0	0	0

4.9.2 DFT (Discrete Fourier Transform)

The frequency spectrum of a signal is calculated using DFT. It investigates the information encoded in the sinusoidal components' frequency, phase, and amplitude. With the use of DFT, FFT analyses a signal [17]. According to algorithms, in continuous FT, Xj is a continuous function of xn.

The quantity of DFT of individual signals is equal to the quantity of DFT of a summation of signals.

DFT symmetry properties

- Xj ω in continuous FT is a continuous function of xn
- DFT of a summation of signals is equivalent to the quantity of DFT of individual signals.
- Symmetry properties of DFT

This is depicted in Table 6 below.

4.9.3 PCA (Principal Component Analysis)

Principle Component Analysis (PCA) is a widely utilised technique for reducing data dimensionality and mining essential feature vectors. All major fundamental features in PCA are referred to as principle components, and these principle components must satisfy the orthogonality criteria. The use of principle components to characterise the ST segment in an ECG signal is a more accurate and universal method [18].

The equation of PCA is based on hypothesis that the signal x is a zero mean random process considered by the correlation in Eq. 4.

$$R_x = E[XX^T] \tag{4}$$

The principal components of X are calculated by applying an orthonormal linear transformation.

$$\varphi = \{\varphi_1 \varphi_2 \varphi_3 \ldots\ldots\ldots \varphi_N\} \text{ to x} \tag{5}$$

Table 6. Confusion matrix for DFT

	Normal (grade1)	Normal (grade2)	Abnormal (grade1)	Abnormal (grade2)
Normal (grade1)	76.47%	17.45%	6.38%	76.47%
Normal (grade2)	5.13%	71.64%	23.63%	5.13%
Abnormal (grade1)	6.79%	13.84%	79.57%	6.89%
Abnormal (grade2)	0	0	0	0

W $= \varphi^T$X so the elements of the principle component vector. w $= \{w_1 w_2 w_3 \ldots\ldots\ldots w_N\}^T$

Becomes mutually uncorrelated. The first principle component is gained as scalar product $w_1 = \varphi_1^T$x where the vector φ_1 is chosen so that the variance of w_1 in Eq. 6

$$\mathrm{E}[w_1^2] = \mathrm{E}[\varphi_1^T XX^T \varphi_1] = \varphi_1^T R_x \varphi_1 \tag{6}$$

Is maximized subject to constraint that

$$\varphi_1^T \varphi_1 = 1 \tag{7}$$

The maximal variance can be obtained when φ_1 is selected as normalized eigen vector which corresponds to largest Eigenvalue of $R_x can$ be denoted by λ_1

Therefore, to get the entire set of N different principal components, the eigenvector equation for R_x needs to be resolved,

$$R_x \varphi = \varphi \boldsymbol{\lambda} \tag{8}$$

So $\boldsymbol{\lambda}$ signifies a diagonal matrix with the eigenvalues $\lambda_1,\ldots,\lambda_N$, the N $\times$ N sample correlation matrix, defined by

$$R_x = \frac{1}{M} XX^T \tag{9}$$

replaces R_x when the eigenvectors are designed, This is depicted in Table 7 below.

4.10 Decision Trees

Decision trees are a type of supervised learning technique that can be used to solve classification and regression issues at the same time. It's called a tree structured classifier because the core nodes represent dataset attributes, the branches represent decision rules, and the leaf nodes provide the final output. Each node represents a feature (attribute), each link (branch) represents a choice (rule), and each leaf represents a result (categorical or

Table 7. Confusion matrix for PCA

	Normal (grade1)	Normal (grade2)	Abnormal (grade1)	Abnormal (grade2)
Normal (grade1)	77.25%	18.12%	4.63%	0
Normal (grade2)	11.27%	70.64%	18.09%	0
Abnormal (grade1)	2.89%	21.22%	75.51%	0
Abnormal (grade2)	0	0	0	68.73%

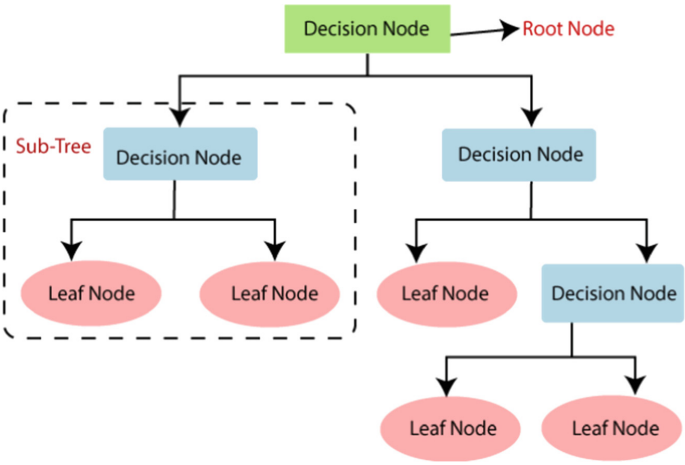


Fig. 15. Generic Structure of Decision Tree (Source: Machine Learning Approaches for Auto Insurance Big Data)

continous value). It's a graphical depiction for getting all feasible solutions to a problem choice based on certain supplied conditions. Decision trees resemble a tree structure in that they twitch at the root node, expanding extra branches and forming a tree-like structure. To create a tree, we can utilise the CART method (Classification and Regression Tree algorithm) [19, 20]. This is depicted in Fig 15.

A binary tree is used to represent the CART model, with each root node representing a single input variable x and a split point on that variable. The output variable y will be assigned to denote predictions at the tree's leaf node. This is depicted in Tables 8 and 9 below

Table 8. Confusion matrix for Decision trees

	Normal (grade1)	Normal (grade2)	Abnormal (grade3)	Abnormal (grade1)
Normal (grade1)	72.33%	20.21%	7.46%	0
Normal (grade2)	18.45%	65.79%	15.76%	0
Abnormal (grade1)	5.62%	22.87%	71.51%	0
Abnormal (grade2)	0	0	0	62.57%

Table 9. Algorithm Comparison for Grades of heart attack

Algorithm	**Accuracy %**	**Sensitivity%**	**Specificity%**
FFT	85	89.36	83.60
DFT	74.26	78.57	77.77
Decision Tree	79.57	84.69	72.40
PCA	87.58	99.46	89.80

5 Conclusion

In this work, the publicly available dataset, Montgomery and Shenzhen dataset were combined together. The process was divided into several steps first was pre- processing which included conversion RGB to gray images, resizing, de-noising the image. The next was feature extraction which included several methods like hierarchical centroid, connected component etc., to extract features like nodes height, width, edge, connectivity, gaps etc. The output of this was fed to the classification algorithms like FFT, DFT, Decision Tree, PCA and did a rigorous comparison analysis and we observe that PCA gave a better accuracy. The proposed model has the potential to be introduced into clinical settings as a helpful tool to aid the normal people in the reading of ECG heartbeat signals and to understand more about them.

References

1. World Health Organization related to CardioVascular Disease https://www.who.int/india/
2. Rahul Kher (2019), "Signal Processing Techniques for Removing Noise from ECG Signals". J Biomed Eng 1: 1–9,2019
3. Fahd Saleh Alotaibi, "Implementation of Machine Learning Model to Predict Heart Failure Disease" International Journal of Advanced Computer Science and Applications(IJACSA), 10(6), 2019

4. ShrutiJalapur, Vanishree Hatti, SnehaJingade, TahaseenPinjar, MadhuHippargi, 2019, Prediction and Detection of Heart Attack using Ai and ML Technology, International journal of engineering research & technology (ijert) rtesit – 2019 (volume 7 – issue 08)
5. ShubhangiKhade, AnaghaSubhedar, KunalChoudhary, Tushar Deshpande, Unmesh Kulkarni, "A System to detect Heart Failure using Deep Learning Techniques" International Research Journal of Engineering and Technology (IRJET)(volume 6-issue 06),2019.
6. A. Subashini, L. SaiRamesh and G. Raghuraman, "Identification and Classification of Heart Beat by Analyzing ECG Signal using Naive Bayes," 2019 Third International Conference on Inventive Systems and Control (ICISC), 2019, pp. 691–694, https://doi.org/10.1109/ICISC44355.2019.9036455.
7. A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," in *IEEE Access*, vol. 7, pp. 180235–180243, 2019
8. Rambhia and A. Naik, "Analysis of Different Heart Rate Monitoring and Pre-Processing Techniques for ECG," *2019 International Conference on Nascent Technologies in Engineering (ICNTE)*, 2019, pp. 1–5, https://doi.org/10.1109/ICNTE44896.2019.8946042.
9. ApurbRajdhan, Milan Sai, Avi Agarwal, Dundigalla Ravi, Dr. Poonam Ghuli, "Heart Disease Prediction using Machine Learning",International Journal of Engineering Research & Technology (IJERT),Vol. 9 Issue 04, April-2020
10. Hammad, Mohamed, Mina Ibrahim, and MohiyHadhoud. "A novel biometric based on ECG signals and images for human authentication." *Int. Arab J. Inf. Technol.* 13.6A (2016): 959–964.
11. Bhateja, Vikrant, et al. "A non-iterative adaptive median filter for image denoising." *2014 International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2014.
12. Hussain, Tarique, et al. "Zoom imaging for rapid aortic vessel wall imaging and cardiovascular risk assessment." *Journal of Magnetic Resonance Imaging* 34.2 (2011): 279–285.
13. Chai, Bing-Bing, Jozsef Vass, and Xinhua Zhuang. "Significance-linked connected component analysis for wavelet image coding." *IEEE Transactions on Image processing* 8.6 (1999): 774–784.
14. Geva, Amir B. "Hierarchical unsupervised fuzzy clustering." *IEEE transactions on fuzzy systems* 7.6 (1999): 723–733.
15. Loresco, Pocholo James M., and Aaron Don Africa. "ECG print-out features extraction using spatial-oriented image processing techniques." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 10.1–5 (2018): 15–20.
16. Tayel, Mazhar B., and Mohamed E. El-Bouridy. "ECG images classification using artificial neural network based on several feature extraction methods." *2008 International Conference on Computer Engineering & Systems*. IEEE, 2008.
17. Najarian, Kayvan, and Robert Splinter. *Biomedical signal and image processing*. Taylor & Francis, 2012.
18. Zhang, Gong, et al. "Automated detection of myocardial infarction using a gramian angular field and principal component analysis network." *IEEE Access* 7 (2019): 171570–171583.
19. Snegireva, Ekaterina, Grigory R. Khazankin, and Igor Mikheenko. "ECG printout interpretation system for clinical decision support." *2020 Cognitive Sciences, Genomics and Bioinformatics (CSGB)*. IEEE, 2020.
20. Diker, Aykut, et al. "Examination of the ECG signal classification technique DEA-ELM using deep convolutional neural network features." *Multimedia Tools and Applications* (2021)
21. Rajesh, T. M., Shaila, S. G., & Koppal, L. B. (2022). Tumor Detection Based on 3D Segmentation Using Region of Interest. In High-Performance Medical Image Processing (pp. 181–195). Apple Academic Press.

22. Shaila, S. G., VijayaLaxmi, I., Rajesh, T. M., Anusha, H. P., Pranami, C., Shahwar, A. K. S., & Sindhu, A. (2022). Analysis and Prediction of Breast Cancer using Multi-model Classification Approach. In Data Engineering and Intelligent Computing (pp. 109–118). Springer, Singapore.
23. Kulkarni, P., & Rajesh, T. M. (2022). A Multi-Model Framework for Grading of Human Emotion Using CNN and Computer Vision. International Journal of Computer Vision and Image Processing (IJCVIP), 12(1), 1–21.
24. Devi, R., & Raju, C. (2021, August). Detection of tumours from MRI scans using Segmentation techniques. In 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 1281–1288). IEEE.
25. Koppal, L. B., Rajesh, T. M., & Vedamurthy, K. B. (2021). A Novel Model for Disease Identification in Mango Plant Leaves Using Multimodal Conventional and Technological Approach. In Data Engineering and Intelligent Computing (pp. 131–140). Springer, Singapore.
26. Rajesh, T. M., Shaila, S. G., Koppal, L. B., & Renuka, N. B. (2021). An Efficient Framework to Bifurcate Healthy and Diseased Vegetables and Fruits Using Multimodal Approach. In Data Engineering and Intelligent Computing (pp. 463–479). Springer, Singapore.
27. Kulkarni, P., & Rajesh, T. M. (2021). Video Based Sub-Categorized Facial Emotion Detection Using LBP and Edge Computing. Rev. d'Intelligence Artif., 35(1), 55–61.
28. Rajesh, T. M., Dalawai, K., & Pradeep, N. (2020). Automatic data acquisition and spot disease identification system in plants pathology domain: agricultural intelligence system in plant pathology domain. In Modern Techniques for Agricultural Disease Management and Crop Yield Prediction (pp. 111–141). IGI Global.