IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol-5, Issue-4 E-ISSN: 2347-2693

Heart Disease Analysis Using Support Vector Machine and Sobel Edge Detection

K. Rajalakshmi[1*], K. Nirmala [2]

[1*]Department of Computer Science, Bharathiyar University, Coimbatore, India
[2]Department of Computer Science, Quaid-E- Millath Government College, Chennai, India

Corresponding Author: rajithiru14468@gmail.com, Mob: 9884287560

www.ijcseonline.org

Received: 26/Mar/2017, Revised: 04/Apr/2017, Accepted: 20/Apr/2017, Published: 30/Apr/2017

***Abstract*—** Diagnosis of heart disease is a challenging task which requires much knowledge and experience. The most traditional way for predicting heart disease are doctor's examinations or taking number of medical tests like as Heart MRI, ECG, Stress Test etc. Now a days, health care industry includes large amount of health care data, which is having hidden medical information. For providing a better and efficient result, novel techniques like Support Vector Machine (SVM) and Sobel Edge Detection has been proposed. This proposed technique provides better output for heart disease detection. The pre-processing step improves the image quality of heart disease MRI image. Increasing of image quality makes the process ease to find affected region. The region of interest techniques sharps the edges in scanned image. Region classification is being applied for isolating the abnormal and normal regions in the heart cells with SVM for identification of various types of abnormalities. The training process classifies the features and recognizes the affected region. The Eclipse IDE tool being used for analyzing the heart disease and several type of heart disease image dataset is being collected from various online sources and stored in a database.

***Index Terms*—** Heart Disease, Support Vector Machine (SVM), Water Shed Segmentation (WSS), Sobel Edge Detection (SED), ROI segmentation, Eclipse IDE, Heart MRI.

I. INTRODUCTION

Heart is playing a very important role in human body, in other word; life is significantly depending on perfect processing of heart. If heart process is faulty, whole body would be affected. There are several reasons for getting heart disease: narration of heart disease in family, cholesterol, smoking, obesity, lack of physical exercise, high blood pressure. Today, the main cause of casualty in the world is because of Heart disease and the general way of predicting by doctor's inspection or by number of medical test like as Heart MRI, Stress Test, and ECG [1].

The WHO (World Health Organization) has estimated around 12 million deaths in all over the world because of heart disease in one calendar. In the year of 2008, approx 17.3 million people lost their life because of this disease. The WHO has estimated that 23.6 million populations will be dead by the year of 2030 because of heart disease [2].

For giving a mileage to the innovative approach of heart disease prediction Sobel edge detection and Support vector machine with segmentation and edge detection will be a better technique for enhancing the approach towards heart disease detection. The filtration process and region classification will support the doctor to analyze the heart disease with new techniques. An Automatic system for diagnosis will beneficial for providing accurate and efficient detection of the disease and computerized information will add additional advantage for achieving the clinical test and making decision.

A step by step algorithm is proposed for segmentation purpose of Heart MRI images for identifying the abnormal heart segment. The proposed steps are utilized for better image quality [3], [4].

a) Normalized image: This image is also called as contrast stretching. It transforms the range of pixel intensity values.

b) Gray level normalized image: This normalization is based on the strength value of normalized and original images [5].

Firstly scanning of artifacts and labels and at last it obtains the image lacking other labels and artifacts by utilizing binary image and original Heart MRI images. The algorithm Sobel edge detection is utilized to every Heart MRI images for discover the edges in Heart, labels and artifacts after that the watershed segmentation is used to develop sharp edges in

original MRI of sobel filtered images. The algorithm for region classification is applied for differentiating the normal and abnormal regions in the heart cell. Lastly the stage of disease is recognized for each Heart MRI image and hence final result is provided [6].

The motivation of this paper is to analysis of heart disease using novel techniques. Using watershed segmentation and ROI segmentation the heart disease image is processed and using SVM classification technique the disease images classified into normal and abnormal stages.

II. RELATED WORK

The improvement in the decision quality of Heart disease diagnosis could be increase to predicting systems . The data mining technique present a way to have the information hidden into the data. Several experiments were performed on non linear and linear features of CMAR (Classification based on Multiple Association Rules) [7], C4.5 (Decision Tree) [8], Bayesian classifiers [4], SVM (Support Vector Machine) [5]. SVM overcome from the different classifiers. The difficulty in recognizing constrained association rules for the Heart disease prediction was considered by Carlos Ordonez [9].
The evaluated data set has included Heart disease medical witness people with quality for artery narrowing, risk factors and heart perfusion capacity. Three limitations were utilized to reduce the no of patterns. The first one requires the quality to show only one side of the rule. The second one isolate attributes into unexciting groups. The final constraint checks the no of attribute present in a rule. The experiment provides the constraints decrease the no of discovered rule extremely next to reduced number time. Two groups of rules predict the absence or presence of heart disease inside the 4 special arteries.

The Data mining technique could support the clinicians in prediction of several patients and in the variation of practices regularly Le Duff et al. [10] may be performed for every medical problem or medical procedure and it would be possible to construct decision tree quickly through data of a physician. The comparison of data mining analysis and traditional analysis demonstrate the role of the data mining method in arranging the variable and give the effect or consequence of data and variables on condition of study. The main disadvantage of the process was knowledge gained and the importance of collecting sufficient amount of data for creating an appropriate model.

In Huan Liu and Lei Yu [11] has given an important concept, major correlation and proposed a fast filtered technique that can distinguish the related features and redundancy between the applicable features without pair wise correlation investigation. The effectiveness and competence of their technique are established by comparison with different methods by using real world data of high dimensionality. The method to compare the classification method by model assessment validity and stability in variable section were given by J.Shreve et.al [12].
This study gives the efficient model for comparing performance of 6 classification technique by the utilization of Monte Carlo simulation and show that the variable selection process is essential in comparing methods to make sure, enhanced stability, least bias and optimized performance. Adithya Sundar et.al [13] talk about a example using data mining techniques, namely Naïve Bayes and WAC (weighted associative classifier). Srinivas et.al [14] briefly study the prospective use of arrangement based data mining techniques like as Decision tree, Rule based, Naïve Bayes, Artificial Neural Network and huge Vol. of healthcare data [15].

Jaya Rama Krishnaiah showed how data classification depends on the algorithm of supervised machine learning that result in accuracy and time acquire for building algorithm. The use of Tanagra tool for classifying the data and are calculated using entropy based partition technique and cross validation and its results are compared. The described algorithm discussed in this paper for the Heart disease dataset through Cleveland Heart Disease database, onto line repository of huge dataset. So, by the use of Tanagra tool the best result s are attained.

III. PROPOSED WORK

In the proposed work, the heart disease database is being generated with whole scanned images by digital system for detecting disease in every particular image. After the detection of Heart disease the pre-processing step is performed. The first step is the Normalization which is also called as the contrast stretching where it is used to vary the pixel intensity value range. Second step are the normalization of grey level image.

This normalization is processed through mapping the level of intensity of Heart MRI image in the range y_{min} and y_{max}. The Edge Detection is the final step. This step is for identifying the exaggerated area boundary. After receiving the pre-processed image it is transformed into binary image for quick reorganization of undesired objects like label and scanning artifacts extracted through original image of Heart MRI.

A. Image Preprocessing

The no of heart disease images gathered from healthcare institute. The MRI scan image of heart disease taken as input. In preprocess, the noise present in the image is removed to enhance the eminence of heart disease images, since there are huge amount of artifacts (e.g. speckles, attenuations, signal dropout and shadow) in heart disease images.

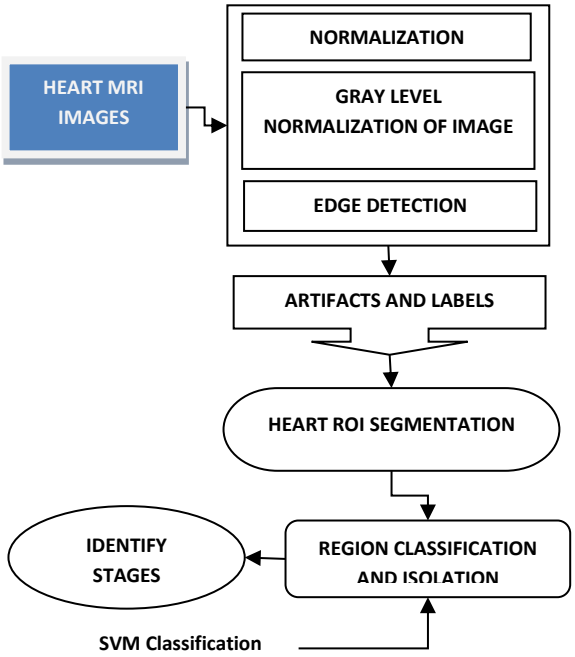

Figure 1: Overall Architecture

B. Gray Level Normalized Image

In normalization process, the novel image strength value is normalized. The Sobel edge detection algorithm is utilized to detect the edges of the heart disease MRI image. The region of the heart cell image classified as normal and abnormal by adjusting contrast of the pixel present in the image.

C. Contrast Enhancement

The poor contrast in MRI image decreases the quality of original image. The consequence of the imperfection has huge collision on contrast of the image. When poor contrast present in image the contrast enhancement method acting vital role. For this case every pixel gray level is extent to increase the contrast. The Contrast enhancements increase the images visualization.

D. ROI

The features can deliberate from the Region of Interest distinctiveness like as shape, size, borders smoothness and density etc. The feature liberty is very huge and composite due to the normal tIssue.s extensive diversity and abnormalities selection. In feature extraction procedure various redundant features should detached to increase the classifier performance.

E. Sobel Edge Detection

The Sobel operator mostly focused in the image processing technique, mainly in the algorithm of edge detection. This is standard to neighborhood of 3x3 matrixes. The orthogonal vector addition is taken as a central gradient evaluation. Each orthogonal vector is expected for directional derivative and this vector is multiplied for unit vector recognizing the derivative direction. The vector addition of these central gradients estimates sum of 8 directional derivative vectors. Thus, Cartesian grid point with its 8 neighbor value density as given below:

Table 1: Point and Its 8 Neighboring Values 1

a	b	c
d	e	f
g	h	i

The estimated directional derivative vector D, which were defined as density difference / distance to neighbor. The directional derivative estimate vector *D* was defined as density difference /distance to neighbor. The vector is certain like the direction of *D* is given by the unit vector to the nearer neighbor. Note, these neighbors are group into the antipodal pairs: (a,i), (b,h), (c,g), (f,d). Vector sum for this gradient estimate:

$$D = \frac{(c-g)}{s}.\frac{[1,1]}{s} + \frac{(a-i)}{s}.\frac{[-1,1]}{s} + (b-h),[0,1] + (f-d),[1,0] \quad (1)$$

Where, S = √2. This vector is obtained as

$$D = \frac{(c-g-a+i)}{2} + f - d \,.\, \frac{(c-g+a-i)}{2} + b - h$$

In the above equation, the given vector is multiplied by 2 as of replace the partition by 2. The corresponding formula is specified as follows

$$D' = 2D = [(c - g - a + i) + 2(f - d),(c - g + a - i) + 2(b - h)] \quad (2)$$

Table 2 And 3: Sobel Masks Of 3 X 3 Dimensions

-1	0	+1
-2	0	+2
-1	0	+1

Dx

+1	+2	+1
-2	0	+2
-1	0	+1

Dy

The Sobel operators is having a couple of 3x3 masks of convolution as shown below in figure single mask are simple and other are rotated by 90°. The given masks are drawn for

maximally respond at edges running horizontally and vertically with respect to pixel grid, 1 masks for every 2 perpendicular orientation.

F. Boosting

Boosting refers to common and effectual technique of generating very precise classifier through combining coarse and moderate imprecise regulations of thumb. It's based on surveillance that ruling many coarse thumb regulations can very easier than discovering single, accurate classifier. To initiate, we describe the algorithm for resulting the thumb rules called weak learner. Boosting algorithm repetitively calls weak learner, and every time it providing dissimilar distribution over training data (Adaboost). Every call produces weak classifier and combining all into single classifier so as to, hopefully, is greatly more accurate than any rules.

i) AdaBoost

AdaBoost is chronological algorithm that reduces the empirical higher bound categorization error by choosing weak classifiers with their weights. These are trailed one-by-one by everyone being selecting to maximally decrease the higher bound error. AdaBoost describes a weights distribution over data samples. These weights updated every time new weak classifier added like that samples misclassified by new weak classifiers prearranged more weight. Currently misclassified samples accentuated more while subsequent weak classifier selection at exponential rate empirical error congregates to zero.

G. Color layout descriptor

The Color layout descriptor (CLD) is very compressed and resolution-invariant color illustration for high-speed image recovery and competently signifies the colors spatial distribution. This feature is used for extensive diversity of content filtering, similarity-based retrieval and visualization. It's particularly valuable for spatial structure-based retrieval appliance. This descriptor is acquired by concerning DCT on 2-D array local delegate colors in Y or Cr or Cb color space. The main purpose of CLD are basically Image matching.

H. Edge Histogram Descriptor

The EHD (Edge Histogram Descriptor) separating image into 4×4 sub images and characterizes local edge distribution and representing every local distribution sub image by the histogram. The EHD contains local-edge histograms, construct it as very flexible. Creating histograms, sub images edges are classified into five category- horizontal, vertical, non directional and diagonal edges. Every sub image is separated into non overlap square blocks image with exacting size which depends on image resolution. Every blocks of image classified into five reveal edge categories or non edge block. The AdaBoost and SVM classifier output merged with Majority Base Classifier. This research has demonstrated that we require selecting the suitable classifiers for feature extraction to improve the precision of classified image.

I. SCD

An astonishing unexpected patient death with or without necessary cardiovascular difficulty and recognized or unrecognized history of cardiovascular disease is confidential as Sudden Cardiac Death (SCD). The SCD is the majority well-known coronary illness sign and responsible for cardiovascular death.

J. Support Vector Machine

SVM is one of the best known methods in pattern classification and image classification. It is designed to separate of a set of training images two different classes, (x1, y1), (x2, y2), ..., (xn, yn) where xi in Rd, d-dimensional attribute space, and yi in {-1,+1}, class label, with i = 1.n. SVM build the finest extrication hyper planes based on kernel function (K). All images, of which quality vector deceit on one part of hyper plane, are belong to class -1 and the others are belong to class +1.

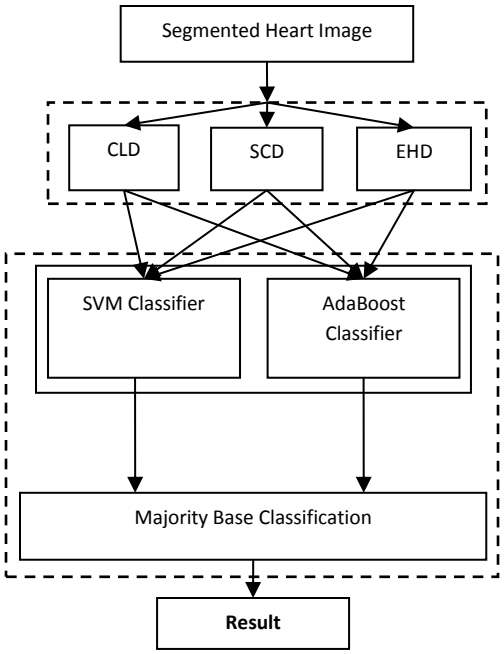

Figure 2: Support Vector Machine Workflow

K. Watershed Segmentation Algorithm

Watershed transforms Segmentation works correctly if we mark or identify foreground object with background locations. The primary segmentation's gradient magnitude is acquired by applying Sobel operator. This essential procedure is followed by Marker-controlled watershed segmentation as:

- Calculation of segmentation function. It is an image and their dark region is an object we try to segment.
- Calculation of foreground markers. It is a connected blob of pixel into every object.
- Calculation of background markers. It is a pixel which is not part of any object.
- Transform the segmentation function therefore it has only minima at background and foreground marker locations.
- Calculate watershed transform of modified segmentation function.

L. Classification

The output image of segmentation is function as input to the seeded region growing here it differentiates various segment of heart based on their intensity values. Each region has different intensity values. The ducts, glands, fatty tIssue. and lobules are having different intensity values; thus separation of regions depends upon their different intensity value. The region classification algorithm with abnormalities such as tIssue. around the heart is having higher intensity value other than the common heart cells. It is important to categorize all closed structures of abnormalities that are found as per its intensity value. In every segmented region the intensity of pixel allocation is not similar in every segmented region, but most pixels have similar intensity value. Hence, Arithmetic mode value is calculated to find the intensity value of pixel by original image and those pixels are changed in that region through calculated mode value.

1. SVM Algorithm

Pseudo Code: SVM
Require: A linear separable set *S*, learning rate $\eta \in \Re^+$
Step1: $a_i=0$; $x_i=0$; y=0;
Step2: $K = \max_{1 \le i \le l} ||x_j||$
Step3: while at least one mistake is made in the for loop do
Step4: for $j = 1,...,l$ do
Step5: if mi($<a_k,x_j>+x_k$)$\le$0then
Step6: $w_{k+1} = w_k + \eta y_j x_j$
Step7: $x_k + \eta y_j \Re^2$ (updating bias[1])
Step8: $k = k + 1$
Step9: end if
Step10: end for
Step11: end while
Step12: Return, a_k, x_k, where *k* is the number of mistakes

2. Segmented Region-SVM Algorithm

Step 1: Take the heart image as input
Input Image I = MRI image;
Step 2: preprocess the image
Preprocess= remove_ noise (I, filter type);
Step 3: Convert into gray level
grayimg = rgb2gray(I);
grayimg = 0.299r + 0.587g + 0.114b(I);
Step 4: Regulate the contrast of the image
Normalize = add (rgb);
Step 5: function [] = Roi()
imgData=rgb2gray(imread(image));
A,B]=find(imgData~=255);
Step 6: Find dimension of the image
dim = image[I];
Step 7: c = [1 2 1; 0 0 0; -1 -2 -1];
H = conv2(double(image),c, 'same');
V = conv2(double(image),c','same');
Ed = sqrt(H.*H + V.*V);
edgeImage = uint8((Ed > threshold) * 255);
Step 8: CLDRnd.RndLocate Locate = new
CLDRnd.RndLocate();
List<float[]> Descriptors =
Locate.extract(image,600);
Step 9: Find Histogram
bin=255;
Val= boundary (I,pixel);
H=hist(Val,0:bin);
Step 10: WatershedSegmentation
WatershedSeg_image = (H);
Step 11: Classification using svm
Size = (0xn);
Position = (0,1);
svm = input (size,position);
If(White_pixel >=1)
return disease

IV. RESULT AND DISCUSSION

The measurement characteristic for performing the proposed approach, the conducted some of research and extract the dataset which was prepared on the Java1.7 with Eclipse JDT Framework.

The heart disease dataset has been collected from the various local and online sources for MRI scanned images for several heart disease. The accuracy of the heart disease dataset is enhancing the heart disease analysis motive.

Attributes

There are several attributes have been considered over the heart cancer detection in last some years, here the

presented technique is producing some latest and strong attributes for heart disease prediction

1. History of heart disease in family,
2. Cholesterol,
3. Smoking,
4. Obesity,
5. Lack of physical exercise,
6. High blood pressure

A. *Performance of different Algorithm*

Table 4: Performance Of Different Algorithm

Algorithm used	Accuracy	Time (ms)
Sobel Edge Detection	52.64%	612ms
CMAR	45.6%	1000ms
Data Mining	52%	745ms

The Table 4 shows Performance of different algorithm. The three different algorithms accuracy calculated based on the time in Mille second.

B. *Performance Accuracy of Sobel Edge Detection:*

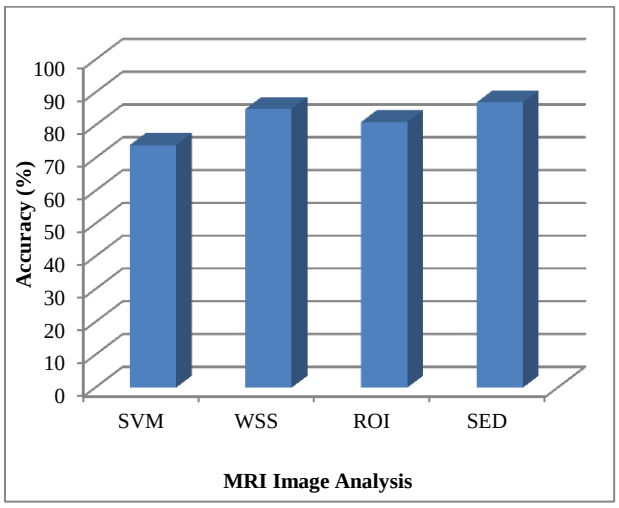

Figure 3: Performance Accuracy of Sobel Edge Detection

The above result in the Figure 3 shows the accuracy in performance graph, the Sobel Edge Detection is more accurate than existing approaches.

C. *Performance of Cross Validation Error Rate:*

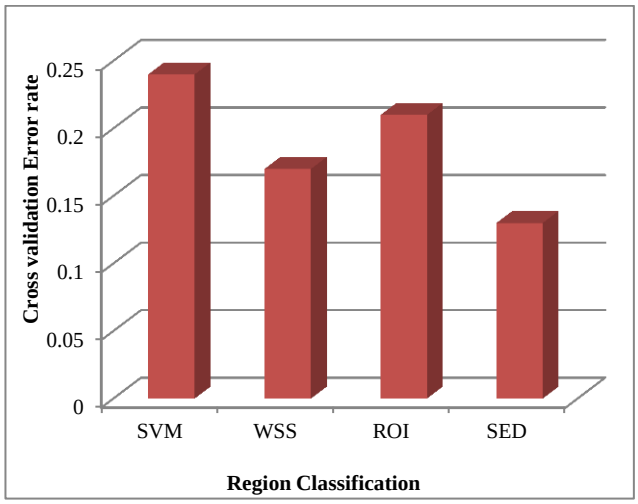

Figure 4: Performance of Cross Validation Error Rate

The above result in the Figure 4 shows the Cross Validation Error performance graph, Support Vector Machine (SVM) is more accurate than existing approaches.

D. *Graphical Representation of Accuracy for each method*

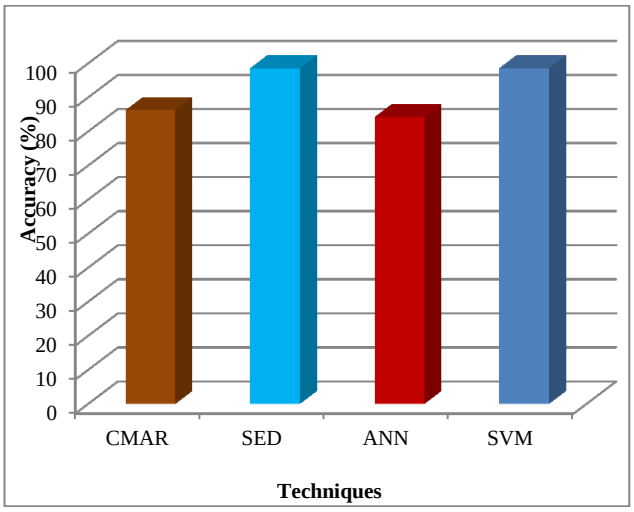

Figure 5: Graphical Representation of Accuracy for each method

In figure 5 the Region classification Techniques in Health Care and Heart Disease Prediction is taken by compare Sobel Edge Detection with CMAR and Data Mining from the existing.

According to the values the accuracy is calculated, from figures (2-3) represents the resultant values of above

classified dataset using Region Classification algorithms and it shows the highest accuracy and lowest computing among the two. It is logical from graph that compared on basis of performance and precision value, Error rate and finally the highest accuracy and again lowest computing time.

Sobel Edge Detection algorithm shows the superior performance compared to other algorithms. Figure 4 the Region classification Techniques in Health Care and Prediction of Heart Disease is taken by compare the above existing technique. The objective of this work is to examine the Heart MRI Classification based Sobel Edge Detection, for the prediction of heart disease.

E. ROI Segmentation

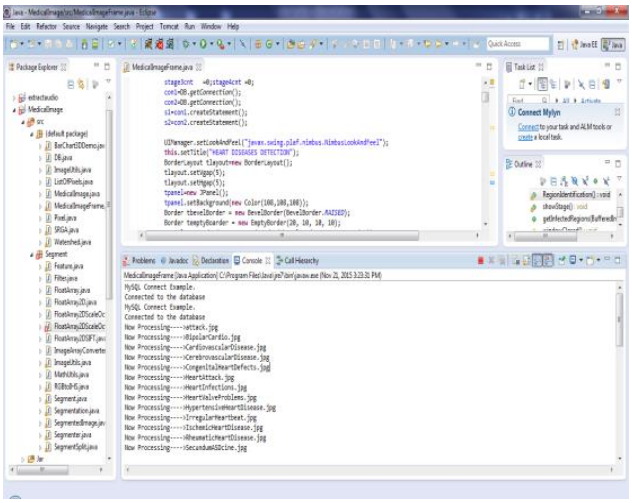

Figure 6: ROI Segmentation

The above figure 6 is representing the processing of input heart diseases image.

F. SOBEL Edge Detection

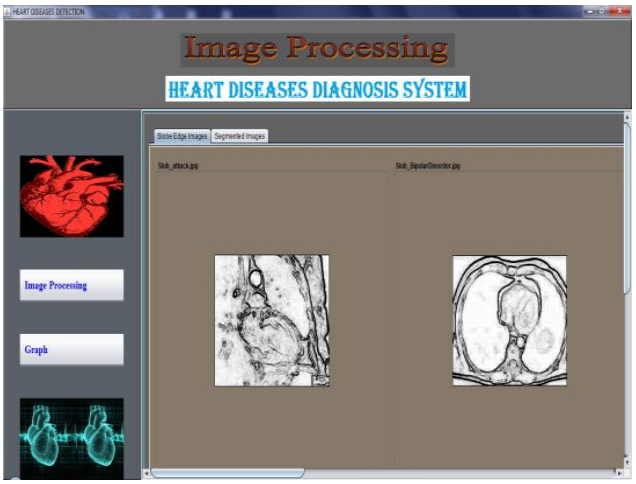

Figure 7: SOBEL Edge Detection

The above figure 7 is presenting the edge detection image of the heart by using Sobel Edge Detection technique.

G. ROI Segmentation

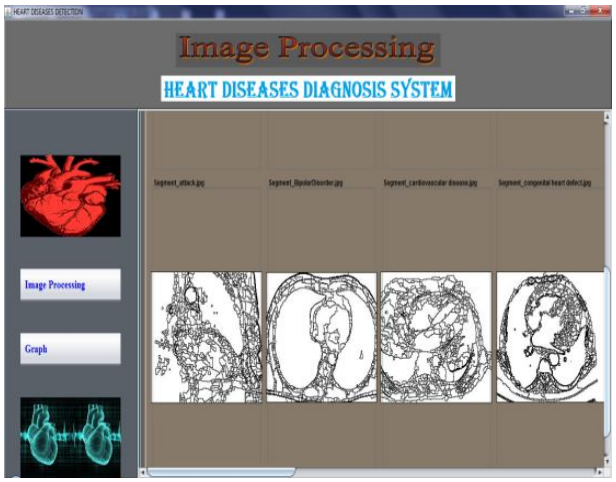

Figure 8: ROI Segmentation

The above figure 8 is presenting the segmented image of the heart disease image by using Region of Interest technique.

H. Infected Heart Image Detection

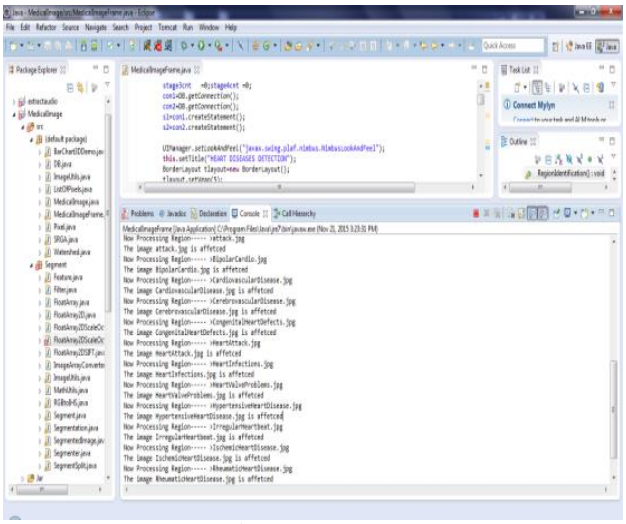

Figure 9: Infected Heart Image Detection

The above figure 9 is presenting the detection result of the heart disease.

I. Infected Areas

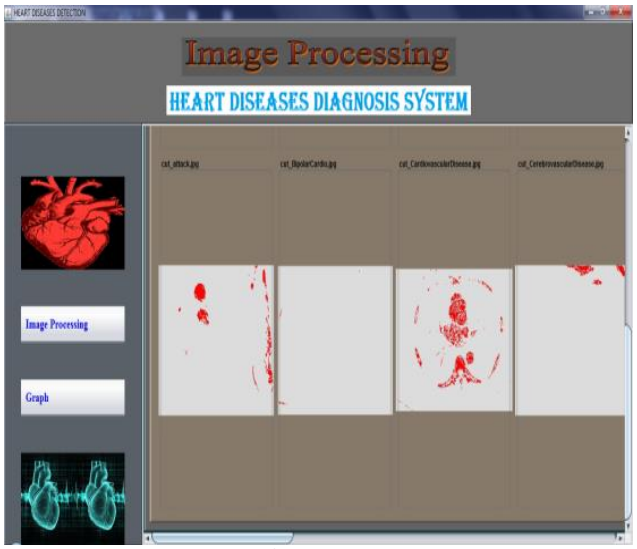

Figure 10: Infected Region Detection

The above figure 10 is presenting the infected region of the disease; the Highlighted area of the above image is affected with the disease and the region is identified.

J. Stage Detection

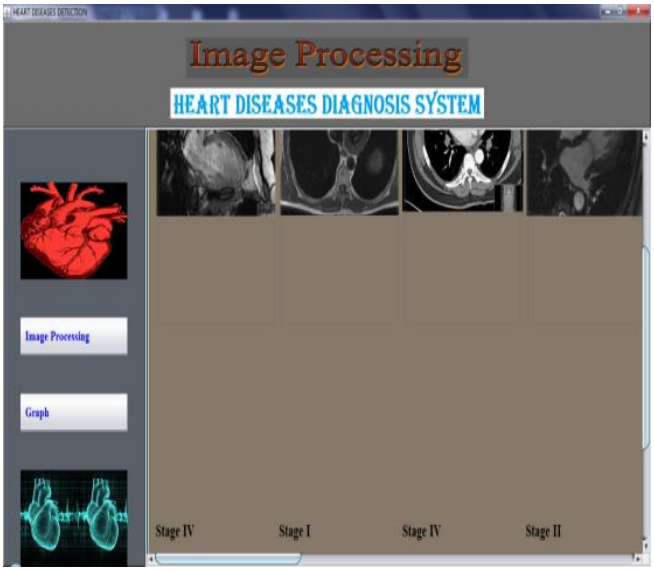

Figure 11: Stage Detection

The above figure 11 is presenting the disease stage.

V. CONCLUSION

The Sobel Edge technique is proposed to process the MRI images and altered into a binary images. Support Vector Machine (SVM) and the Sobel Edge Detection (SED) are employed to detect the heart disease and provide a better prediction for separating the abnormal and normal regions in Heart cell. Finally the stages of disease are detected for each Heart MRI images using performance steps. The accuracy of proposed system SVM and SED are 98.61% and 87% respectively. The future work might be carryout for identifying heart disease with several mechanisms like heart disease history in family, cholesterol, genes and etc.

REFERENCES

[1] Priyadharsini C, ASs Thanamani, "*An Overview of Knowledge Discovery Database and Data mining Techniques*", IJIRCCE, Vol.2, SIssue..1, pp.23-29, 2014.

[2] John Peter, Soma Sundaram, "*Study and development of novel feature selection framework for heart disease prediction*", International journal of scientific and research publications, Vol. 2, Issue.10, pp.1-7, 2012.

[3] JR Krishnaiah, D.V. Chandrasekhar, K. Ramchand, H Rao, "*Predicting The Heart Attack Symptoms Using Biomedical Data Mining Techniques*", The International journal of computer science & applications, Vol. 1, No. 3, pp.10-18, 2012.

[4] HG Lee, KY Noh, KH Ryu, "*Mining Biosignal Data: Coronary Artery Disease Diagnosis Using Linear And Nonlinear features of HRV*", LNAI 4819: Emerging technologies in knowledge discovery and data mining, USA, pp. 56-66, 2007.

[5] N. Cristianini, J. Shawe-taylor, "*An Introduction to support vector machines",* International Journal of scientific and research publications, Vol. 3, Issue. 6, pp.88-96, 2013.

[6] *M. Patel, A. Hasan , S.Kumar, "A Survey: Preventing Discovering Association Rules for Large Data Base",* International Journal of Scientific Research in Computer Science and Engineering, Vol.1, Issue.2, pp.30-32, 2013.

[7] Wenmin Li, Jiawei Han, Jian Pei, "*CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules"*, Natural Sciences and Engineering Research Council of Canada (grant NSERC-A3723), USA, pp.369-376, 2011.

[8] HC. Koh G. Tan, "*Data Mining applications in healthcare*", Journal of healthcare info. Management Vol.19, Issue. 2, pp.64-72, 2005.

[9] MA Jabbar, P. Chandrab, BL Deekshatuluc, "*Heart Disease Prediction System using Associative Classification and Genetic Algorithm*", Vol.1, Issue.1, pp.183-192, 2012.

[10] F. lemke, JA. Mueller, "*Medical Data Analysis Using Self-Organizing Data Mining Technologies - systems analysis modeling simulation*", Vol.43, Issue.10, pp.1399-1408, 2003.

[11] M K J Siddiqui , M Anand , P K Mehrotra , R Sarangi , N Mathur, "*Women With Benign and Malignant Breast Disease"*, Environmental research,vol.98, Issue.2, pp.250-257, 2005.

[12] J. Shreve, H. Schneider, O. Soysal, "*A Methodology for Comparing classification methods through the assessment of model stability and validity in variable selection*", Decision support systems, Elsevier Science Publishers, The Netherlands, Vol.52, Issue.1, pp.247-257, 2011.

[13] NA. Sundar, PP. Latha, MR Chandra, "*Performance Analysis Of Classification Data Mining Techniques Over Heart Disease Data Base*", International. Journal of Engineering Science & Advance Technology, Vol.2, Issue.3, pp. 470 - 478, 2012.

[14] K. Srinivas, BK. Rani, A. Govrdhan, "*Applications of data mining techniques in health care and prediction of heart attacks*", International journal on computer science and engineering, Vol. 02, Issue.2, pp. 250-25, 2010.

[15] V. Krishnaiah, G. Narsimha, NS. Chandra, "*Survey of Classification Techniques in Data Mining*", International Journal of Computer Sciences and Engineering, Vol.2, Issue.9, pp.65-74, 2014.

Authors Profile

K.Rajalakshmi pursed Bachelor of Mathematics from Madras University, Chennai in 2006 and Master of Computer Science from Anna University in 2008 and Master of Philosophy in Computer Science from Vinayaka Mission University in 2009.She is currently pursuing Ph.D in Computer Science at Bharathiar University, Coimbatore, India. She is working as Assistant Professor in Department of Computer Science. She has 5 years of teaching experience. Her main research work focuses on Data Mining Techniques.

Dr.K.Nirmala pursed Bachelor of Commerce from Valliammal College, University of Madras,chennai in 1992 and Master of Computer Application from Presidency College, University of Madras in 1996. She pursed PhD in Computer Science from NITTR , University of Madras, chennai in 2007.She was as a member of Board of Studies in Vels University. She has Provided number of Lectures on Computer Oriented Studies and guided more than six Students for M.Phil from Dravidian university and Guiding five students for PhD from Bharathiar University, Coimbatore,India.

