Agile Data Mining Approach for Medical Image Mining

Reetika Koli
computer science and engineering
Dev bhoomi group of institutuion
Dehradun, India
reetikakoli89@gmail.com

Ritu Pal
computer science and engineering
Dev bhoomi group of institutuion
Dehradun, India
ritu.pal26@gmail.com

Nisha chaube
computer science and engineering
Graphic era University
Dehradun, India
nisha.chaube20.gmail.com

Kavita Joshi
computer science and engineering
GRD college
Dehradun, India
kjkavita86@gmail.com

Ankit Maithani
computer science and engineering
Dev bhoomi group of institutuion
Dehradun, India
ankitmaithani2518@gmail.com

***Abstract*—As Record keeping has turned into an unavoidable action in our everyday life, Medical aspect of this isn't immaculate medical data in digitized structure for example images is gathered from medical health centers. We need a successful quickly developing application to store, procedure and break down such a gigantic measure of information. Information mining approaches comprises of a lot more systems to break down future prediction and decision making. In our proposed procedure, Ensuring the storage of image dataset we perform hybrid technique of data mining approached such as CBIR technique to extract the features for region of interest, based on similarity of attributes like color, texture, shape, attribute, and text domain concepts and then deal with the classification of images using SVM. The performance statistics shown based on the accuracy of the classified brain images.**

Keywords—Medical image mining, Content based image retrieval (CBIR), SVM support vector Machine), Hue saturation value (HSV), Picture archiving and communication system (PACS), National institute of health (NIH). Magnetic Resonance Imaging (MRI), positron emission tomography (PET), Computed tomography (CT).

I. INTRODUCTION

From past few years, Hospitals that are well-equipped, have started generating and collecting their own image data records throughout medical treatment and various health examination modalities such as MRI, PET, CT scan, X ray and ultrasound images. Hospitals have installed advance monitoring equipment to generate these image data. This huge data is digitized in medical areas which produces very voluminous collection of digitized datasets. Such a huge collection of medical image data is hard to analyse, classify access query images with high efficiency within reasonable time because the data sets contains several kinds of feature that creates problems to visualize the images correctly with high efficiency based on attribute while maintaining reasonable execution time and limited resources which are available.

The two important concerns with image mining is to mining huge collection of medical images, and combined data mining of large collections of images with associated alphanumeric data. The importance of mining for medical data to assist the physician to make the accurate decision by minimizing the diagnostic error this leads to increasing the quality of medical treatment by highlighting the efficient characteristics of medical data as well as analysis with respect to time. Image mining provides a framework that uses a raw format of images stored in the database, which cannot be processed directly, it must be processed them first to use in high level modeling [7].An image mining technique is considered as a good technique if it capable of easy and effective user interaction during retrieving the patterns and knowledge from the collection of huge image database. Useful information can be retrieved to the human users by analysing these images.

As a result, there is a requirement of images mining systems, which can review semantically relevant information by self-acting from large amount of image data. Medical image mining is now one of the rapidly growing areas of biomedical technology. It enables quantitative analysis and processing of medical images and microscopic data which contain varied information about the human anatomy as well as diseases. It has proven to be more powerful and effective for retrieving as well as discovering useful patterns which are not explored before and for new invention. Due to such capability of Medical Image Mining, world has witnessed peaceful growth of biotechnology.

By using standard analysis tools and user-interface, researchers can easily share research data and their analysis with suitable result, thereby enhancing their potential to research, diagnosis, monitor, and treat medical disorders globally at remote sites.

II. RELATED WORK

Today, Health care have huge collection of enormous dataset, which is designed by compliance and patient care record keeping regularity requirement while most of this dataset are stored in record file, but the new aim is toward fast digitization of these datasetMaintaining the Integrity of the Specifications.

Beyer et al. [16] presented increasing the large number of features to represent the image can create a problem. We have to keep the count of features as low as possible to improve the inequity accuracy. By using histogram Color distribution of image is expressed but to distinguish mammogram lesions as malignant and benign, shape features can be used, it leads to upraise the calibration complexity of the process.

Textural variations in image mammogram depict the discrepancy in density of tissue because Texture feature

describes the Properties of regularity, acuteness and smoothness.

Carson et.al. [18] Presented image representation which provides a transformation from raw pixel data to a small set of localized coherent regions in color and texture space.

Ji Zhang et al. [1] proposed various image mining research issues, frameworks used for image mining, current developments in image mining, state of the art techniques and systems. A structure for texture information of an image and accomplishment of higher retrieval efficiency than the shape features of an image is presented by Monika Sahu et al. [17].

Maria-Luiza Antonie et al.[19]Presented the working on the mammography by using neural network by segmenting the image and then apply the association rule mining and get ~70% accuracy.

Abdullah et al. [15] presented the working of SVM with lab techniques images under two categories, either normal or abnormal brain which refers to brain tumor. The determination of normal and abnormal brain image is based on symmetry which is exhibited in the axial and coronal images. SVM defines the decision boundaries in the training step and the method can also provide good generalization in high dimensional input spaces. Several researchers reported that feature selection is important for SVM.

And give average 60 to 67% accuracy in the classification of the classes of the brain images.

Dhanshree et al. [20] presented the conversion of unused data into useful one and classifies and predict the heart disease from unknown samples by using naïve Bayes data mining algorithm.

Barbora zaharadnikova et al. [14] presented the automated image analysis and consequent knowledge acquisition based on computer driven processing. By emphasize the fact of users about the need and utilize an incredible amount of picture streaming data coming from different private as well as commercial sites globally.

III. MATERIAL AND METHODS

In our prospective methodology we performed various data mining techniques on the brain medical image to mine the Patterns for classification or for new discoveries, treatment for the severe brain diseases and about the normalcy of the tissues in the brain. It will help to mine the pattern of the nervous system and spot the clotting of region of interest (ROI) easily. These approaches provide cheapest way to analyze the large medical dataset with higher accuracy. These are defined below

A. Medical Image Data

Image data is taken from the PACS (Picture achieving communication and system) of national institute of health (NIH).512*512 pixels has been chosen as input data after consultation with the neurologist. The input image data involved 50 patients (35 abnormal and 15 normal) with in the age of 25 to 40 years old. The abnormal brain images dataset consists of images of brain clotting, brain tumor, meningitis (inflammation in brain lining), swelling and injury. Selection of clearly seen of lateral ventricles input image data is the prime indicator at the very first stage. The lateral ventricles and arteries are situated in the cerebral hemisphere of the brain.

B. Pre-processing and Feature Extraction

In our proposal system aims to reduce noise, search and find medical image data were enhance with the image enhancement technique such as Median filter, Content based image retrieval (CBIR) using Matlab 2014 i.e. Noise reduction is a prototypical pre-processing step to remove the noise from huge dataset of images. Median filter is used to reduce the noise from the images or signal; it is a nonlinear digital filtering technique.

Visual content to retrieve images from large images databases as per the user's interest. CBIR is applicable because searches are dependent on annotation quality and provide integrity to huge data and for database for calculate similarity and dissimilarity of specific images by query based on their content. Annotation of images is done by manually selecting metadata tags, and keywords in large dataset may be insufficient, tedious, and laborious this may not capture the required keyword to explore the image. The term "content" indicates the visual features such as edge, color, shape, texture or any other information that can be derived from the image itself [5].

Color is one of the prestigious features this is most widely used visual features in CBIR for associate with the underlying objects in an image. It is vigorous to background complications, scaling, perspective, orientation and size of an image. It is obligatory to choose coherent features that are compatible to each other so as generate revamp retrieval performance and to combine chosen feature effectively without expended feature vector dimension. The CBIR system extricates the color and texture feature separately or in combination. Firstly system extracts the color feature on the basis of RGB query image then it transforms the image into HSV color image to generate the color feature vector.

C. Gabor filter

After processing feature extraction, the system extract texture feature on query brain MRI image Texture Features are extracted from the segmented region of interest (ROI) and organized into feature vectors. By applying Gabor filter is used to form a texture feature vector by extracted regular or irregular patterns from the segmented region of interest (ROI) and organized into feature vectors on query MRI image of image. Texture contains salient feature regarding underlying structural arrangement of the surfaces in a medical image. When a small area in an image has wide Variety of discrete tonal feature, the assertive property of that area is texture. It is used a prime feature for segmentation or classification in area of research in pattern recognition of medical images. There are variety of techniques has been used for measuring textual similarity but here.

Gabor wavelet transform has both the multi-orientation and multi-resolution properties and are flawless for calculating local spatial frequencies. Besides, it has been found to bear distortion resistance space for pattern

recognition tasks. Biological motivation is best modeled as a family of self-similar 2D Gabor wavelets. It has been used in many image analysis applications, and this report focus its applications on texture classification in bio medical field.

These filters are based on multichannel filtering, which imitate some attribute of the human visual system.

$$h(x,y)=\frac{1}{2\pi\sigma 2}\exp[-(\frac{x2-y2}{2\sigma 2})].\exp(j2\pi F(x\cos\theta+y\sin\theta)) \quad (1)$$

- The radial frequency (F) where the filter is centered in the frequency domain.
- The standard deviation (σ) of the Gaussian curve.
- The orientation (θ) is calculated for the directions of 0_0, 45_0, 90_0 and 135_0 and for the phase 6. Every input image is expressed by its uniqueness in feature vector.

The retrieval system binds these feature (color+ texture) vectors and calculate the similarity by using similarity metric L2 Euclidian distance between the combined feature vectors of query image and each of training image dataset.

IV. CLASSIFIACTION

The classification will work only when the number of rules extracted for each class is balanced. In other cases, a further tuning of the classification system is required. The attune of the classifier is mainly entitled by finding some flawless intervals of the confidence such as both the overall recognition rate and the recognition rate of normal and abnormal cases are at its maximum value.

SVM is a universal constructive learning procedure based on the statistical learning theory developed by Vapnik in 1996[15]. SVM is a high-performance classifier in several domains; it assists the benefits of high dimensional spaces are being used with expeditious evaluation of learned target function and accuracy on the image classification. This has been generally used in pattern recognition application due to its good generalization and computational efficiency performance.eg. To Classify, visualize and analyze application in financial as well as biomedical study.

The major goal of SVM is the configuration of flawless hyper plane is generally used for the problems of pattern classification and identification. It is easy to have a linear hyper plane between the classes by using Kernel trick. It is effective in high dimensional spaces where number of dimensions is greater than number of samples.

SVM needs to classify the given class of pattern correctly so that is can enlarge the margin that defines the efficiency of SVM algorithm. The efficiency in classifying pattern will improve based on the size of the margin i.e. greater the margin size more exactly it classifies the patterns and it is memory efficient because it uses a subset of training points in the decision function.

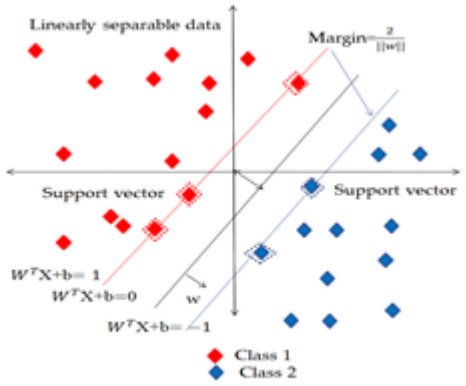

Fig. 1. SVM shows the linear hyper plane with classes

$$F(X) = \boldsymbol{\alpha} iyi\,(xi^T X) + b \quad (2)$$

When $W^T X + b = 0$ and c ($= W^T X + b$) = 0 defines the same plane. We have ability to choose the normalization of w.

Choose normalization such that:

$W^T X+b=+1$ for positive (3)

and

$W^T X+b=-1$ for negative (4)

However, this is used for the classification of the brain images, this will help to classify the normalcy of the image i.e. which is normal or abnormal.

V. PROPOSED METHODOLOGY

In Proposed methodology we are performing a classification of brain images to diagnose the tumor, clotting, injury normalcy of the brain. The classification is performed by applying some data mining algorithms i.e. CBIR based on content with SVM together. The analysis performed effectively to assist the physician to take accurate decision on the basis of better predictive modeling on brain image datasets. The flow chart in fig.1 defines the whole scenario of our work.

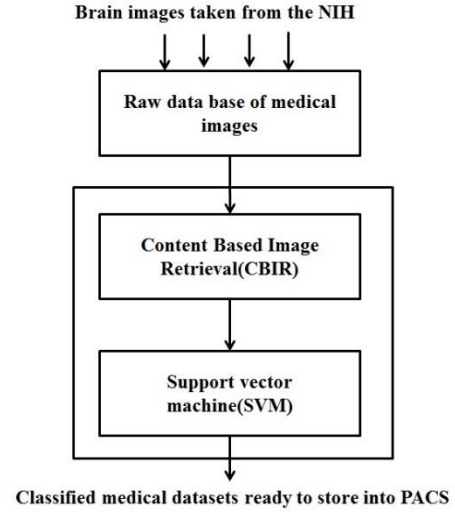

Fig. 2. Flow chart of proposed methodology

Above flow chart works for methodology to classify the large dataset of brain images.

- Brain image data set is taken by the National Health institute (NIH) and creates a raw database.
- Median filtration is performed for Pre-processing to spackle the noise in huge dataset.
- Initially Color feature is extracted to create color feature vector by using HSV.

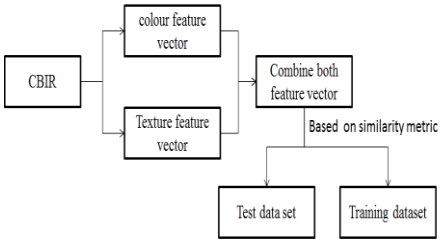

Fig. 3. Flow chart of CBIR working

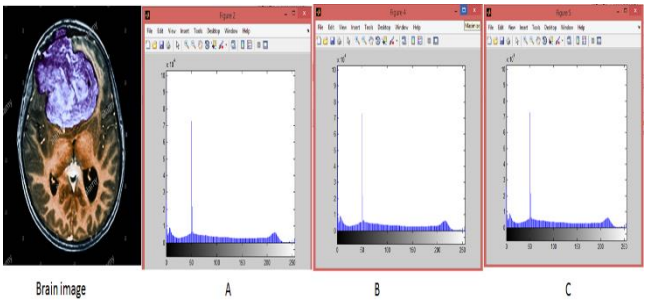

Fig. 4. shows the RGB histogram for the brain image data set one by one

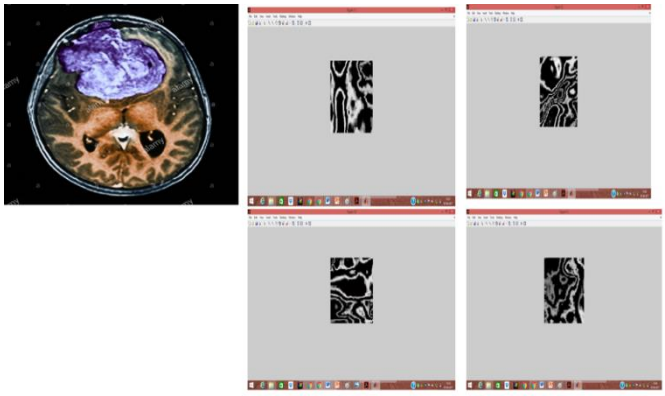

Fig. 5. shows the Extracted texture feature

- After color texture feature is extracted to create Texture feature vector by using gabor filter.
- Combine that feature vector and create test data set as well as Training data set.
- L2 Euclidian similarity distance metric is used to match the new and Query images in training data set from the test data set.
- By applying SVM we can easily classify the class of the image, for classifies query image and predict the disease for the unknown samples, then it will store as a new pattern contain extremely new information for new inventions or for future predictions.
- It gives clear result i.e. which side of the brain is affected from the tumor, injury, swelling and clotting. This comes under the ROI (Region of interest).
- On an average a classifier performs an extremely well as compared to other classifier on the medical image data set.
- Now the Data is classified into classes and ready to store into the PACS (Picture achieving and communication system) system for Diagnosis and research globally at remote sites.

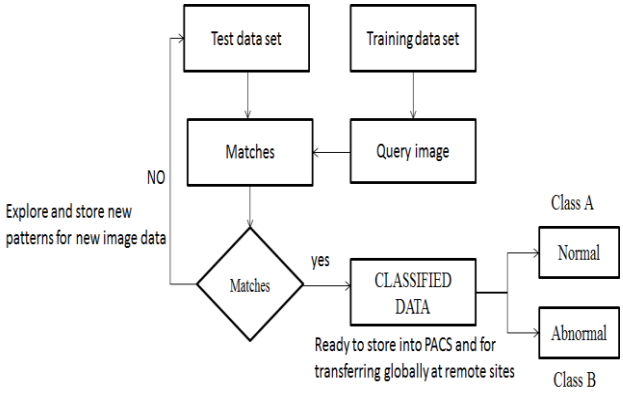

Fig. 6. Flow chart of SVM working

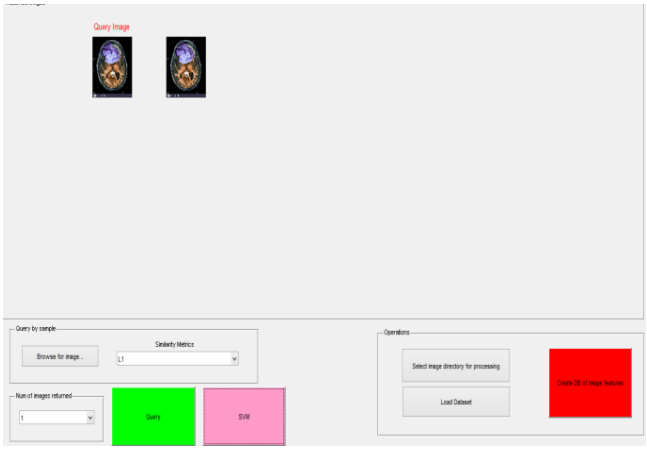

Fig. 7. SVM gives ~80% accuracy working with CBIR

VI. Results and Analysis

In this work, we have to measure overall accuracy as a performance evaluation metrics this is this is one of the essential problems in CBIR and SVM. We have used the most often evolution method namely precision and recall. In which recall value one just by retrieving all images and in similar manner the precision value kept in higher value by retrieving only few images or precision and recall should either be used combine or the number of images retrieved should be specified.

$$\text{Precision} = \frac{No.of\ relevant\ images\ retrieved}{Total\ no.of\ images} \quad (1)$$

$$\text{Recall} = \frac{No.of\ relevant\ images\ retrieved}{Total\ no.of\ relevant\ images\ in\ the\ databse} \quad (2)$$

```
Processing orientation 6

Code for <Variation in Phase Symetry> Starts
Taking Median for scale 1/4
Taking Median for scale 2/4
Taking Median for scale 3/4
Taking Median for scale 4/4
Code End for <Variation in Phase Symetry>
Processing orientation 1
Processing orientation 2
Processing orientation 3
Processing orientation 4
Processing orientation 5
Processing orientation 6

Code for <Variation in Phase Symetry> Starts
Taking Median for scale 1/4
Taking Median for scale 2/4
Taking Median for scale 3/4
Taking Median for scale 4/4
Code End for <Variation in Phase Symetry>
SVM (1-against-1):
accuracy = 81.05%
Confusion Matrix:
    38     3     7     0     0
     4    38     6     2     0
     8     3    35     4     0
     3     1     5    41     0
     0     0     1     0    49

Predicted Query Image Belongs to Class = 1
>>
```

Fig. 8. SVM gives ~80% accuracy working with CBIR

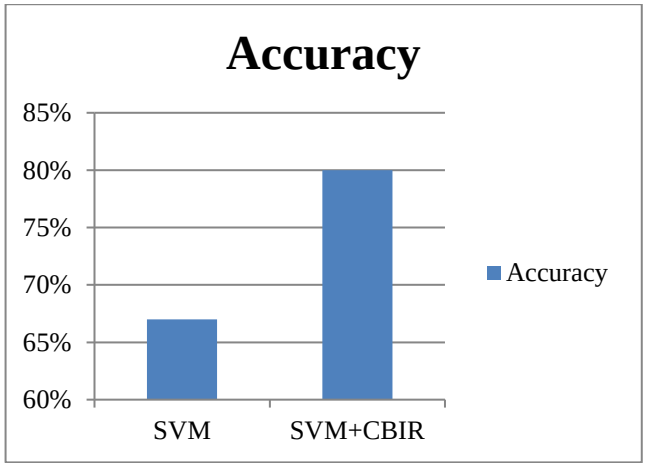

Fig. 9. shows the comparison of working SVM alone and in combination with CBIR

VII. Conclusion

Efficient classification of medical images i.e. brain images are computed which creates a better result for classifying data and for new inventions. To visualize, classifies, store, register and transfer of large collection of brain image dataset over remote sites i.e. MRI images, CT scan images, PET images and ultrasound images into the PACS system. The motivation of proposed work is to use hybrid technique in combination of CBIR and SVM data mining techniques is the effective techniques to reduce redundancy and effectively classify image datasets correctly. Enhancement in performance by improving the accuracy to ~80% and above based on texture and colour pattern applying on data by reduce noise and redundant data but without CBIR, SVM gives 67% accuracy, so the performance of prediction is good for both in term of subjective quality and objective quality with this novel approach.

In future this approach is applicable on the Medical videos to extract the content and to diagnose the disease. It may be applicable on the echocardiogram to predict the heart disease such as blockage in the vein of heart, shape of the heart, and rhythm (beat) of the heart and breast cancer.

References

[1] Ji Zhang, Wynne Hsu and Mong Li Lee ,Image Mining: Trends and Developments, In preceding Journal of intelligent information systems vol.19,pp.7—23,2002.

[2] Ordonez and Edward Omiecinski ,Image Mining: A New Approach For Data Mining,Carlos, Georgia Institute of Technology,1998.

[3] Mining: Foschi, Patricia G,et ol Feature Extraction for Image, Multimedia Information Systems ,pp 103—109,2002.

[4] A.Hema, E.Annasaro "Survey in need of Image mining techniques" et ol International journal of Advanced Research in Computer and Communication Engineering vol.2,Issue 2,February 2013.

[5] C.Kanimozhi, D.Bavya, S.Gokulavani " Image mining to image reterival" , International journal of Advanced Research in Computer Science and Software Engineering vol.5,Issue 3,March 2015.

[6] M.janani, Dr.R.Manicka Chezian,"A Survey On Content Based Image Reterival System" In International journal Of Advanced Research In Computer Engineering Technology Vol.1,Issue5,July 2012.

[7] D.S.Zhang and G.Lu, "Content Based Image Retrieval Using Texture Features", In Proc. of First IEEE Pacific-Rim Confrence on Multimedia (PCM) pp.393-395,Sydney,Australia,December2000 .

[8] A.Kannan,Dr.V.Mohan and Dr.N.Anbazhagan,"An Effective Method Of Image Retrieval Using Image Mining Techniques" In The

International Journal Of Multimedia Its Application (IJMA) Vol.2,No.4,November 2010.

[9] Priyanka B.Kamdi and Pravin Kulurkar,"Data Mining Approach For Image Retrieval in Multimodal Fusion Frequent Pattern Tree" In International Journal Of Advance Research In Computer Science and Management Studies Volume 3,Issue 7,July 2015.

[10] c Neethu, J., & Wilson, A. (2014, November). Retrieval of images using data mining techniques. In Contemporary Computing and Informatics (IC3I), 2014 International Conference on (pp. 204-208). IEEE.

[11] J.Zhang,W.Hsu, and M.L.Lee, "Image mining: Issue, frameworks and Techniques",In proceeding of the 2nd International Workshop on Multimedia Data Mining and the ACM SIGKDD confrence(MDM/KDD'01)(San Francisco),pp.13-20,August 2001.

[12] Deepika Kishor Nagthane, "Image Mining Techniques and Application " In preceding of International Journal Of Engineering Sciences and Research Technology, July 2013.

[13] Barbora Zahradnikova and Sona Duchovicova et al "Image Mining: Reviews and New Challenges", In preceding Intenational Journal of Advanced computer science and Application,pp.242—246,vol.6,no.7,year 2015.

[14] Mandavi, Roshani and Nagwanshi, Kapil Kumar "Content Based Image Retrieval and Classification Using Principal Component

[15] Analysis" In International Journal of Science and Research (IJSR) , Volume 4 Issue 6, June 2015 .

[16] Abdullah, Noramalina and Ngah "Image classification of brain MRI using support vector machine", In book title Imaging Systems and Techniques (IST), 2011 IEEE International Conference,pp.242—247,year 2011

[17] Beyer,K. et al. "When is nearest neighbor meaningful?' In Proc. Int. Conf. Database Theo. (ICDT), 1999 pp 217-235.

[18] Sahu Monika, Shrivastava Madhup, "Image Mining: A New Approach for data mining based on texture", in Proceedings of IEEE International Conference on Computer and Communication Technology,2012

[19] Carson, Chad, Serge Belongie, Hayit Greenspan, and Jitendra Malik. "Region-Based Image Querying", in Content-Based Access of Image and Video Libraries, 1997. Proceedings. IEEE Workshop on, pp. 42-49. IEEE, 1997.

[20] Maria-Luiza Antonie, Osmar R. Za¨ıane and Alexandru Coman "Application of Data Mining Techniques for Medical Image Classification" Proceedings of the Second International Conference on Multimedia Data Mining,pp.94—101,year 2001

[21] Dhanshree S Medhekar, Mayur P. Bote et al. "Heart disease Prediction System using Naïve Bayes"in preeceding of International Journal of Advanced Research in Computer Science and Communication Engineering,pp.6787—9,vol.3,no.5,year 2014 Avinash patil, G.B.Hangargi: Routing protocol in Mobile Ad-Hoc Network. ISSN: 2348-4748 Volume 2, Issue 5 (2015).