Classification of Brain MRI Using SVM and KNN Classifier

Vijay Wasule
Department of Electronics and Telecommunication Engineering,
Rajiv Gandhi Institute of Technology, Andheri, Mumbai
viju.wasule@gmail.com

Poonam Sonar
Department of Electronics and Telecommunication Engineering,
Rajiv Gandhi Institute of Technology, Andheri, Mumbai,
poonamsonar@gmail.com

***Abstract*— The classification of the brain MRI is an important task. In this paper, the automatic approach to the classification of brain tumor into malignant Vs. benign and low grade Vs. high grade glioma is present. This method employs GLCM technique to extract the texture features from images and stored as a feature vector. The extracted features were classified using supervised SVM and KNN algorithm. The proposed system is applied on the 251 images (85 malignant and 166 benign) of clinical database and 80 images (50 low grade glioma and 30 high grade glioma) of brats 2012 training database. The accuracy of the proposed system is 96% and 86% for SVM and KNN respectively for clinical database and 85% and 72.50% for SVM and KNN respectively for Brats database.**

Keywords*— *Benign, GLCM, high grade glioma, KNN, low grade glioma, malignant, MRI, RBF kernel, SVM.

I. Introduction

With the rapidly increasing population, cancer has the major global public health issue. Imaging plays very decisive role in analysis of diagnosis of patients having brain tumors. Tumors are undesirable groups of cells (tissue) which grows by uncontrolled cell division. Brain tumors are named depends on the cell type from which they grow. Brain tumors classified as a primary and secondary. Primary tumors are composed of cells just like those that belong to the organ or tissue where they start. A primary brain tumor initiate from cells in the brain. Malignant tumors grow rapidly and can extend over large surrounding tissues. Secondary tumors are formed of cells from another part of the body that has spread to one or more areas. Radiologists inspect MRI images based on the visual clarification to identify the presence of tumors. There might be a possibility when large volume of MRI to be analyzed, then there is a possibility of wrong diagnosis by radiologists because the sensitivity of the human eye decreases with the escalating number of cases, predominantly when only a small number of slices are affected. Hence there is an obligation for efficient automated systems for analysis and classification of medical images. The early diagnosis and right treatment of brain tumors are most important to fix damage to the brain or even unprotecting death. Accurate facts with the position of the tumor and its size are most important for accurate and effective treatment. Brain magnetic resonance imaging (MRI) is offered for the observation and segmentation of brain tumors. Magnetic resonance imaging (MRI) broadly used nowadays in hospital and clinics for medical diagnosis especially in brain imaging. As MRI has advantage of soft tissue contrast and is noninvasiveness. MRI does not use any ionizing radiation. MRI is a used in a brain imaging because it is non-radioactive, non-aggressive and pain free method. .

II. Literature review

H. B. Nandpuru et al. [1] present the method for brain tumor classification. The brains MRI are classified into the normal and cancerous using SVM a supervised machine learning technique. Firstly, the texture, symmetrical and gray features were extracted. The proposed classifier gives 84% accuracy

Janki Naik et al. [2] proposed classification of brain tumor in the brain MRI images using image mining technique. The MRI images preprocessed by the median filtering and features have been extracted using texture feature extraction technique. Decision Tree classification and interclass relationship in text classification to improve the efficiency than traditional mining method. The system used support vector machine (SVM) classifier which gives the 83% accuracy.

Classification of brain tissues in MRI using a hybrid approach of GA and SVM is proposed by Ahmed Kharrat et al. [3]. The features are extracted by spatial gray level dependence method called SGLDM. The proposed system gives a good accuracy about 85.22%.

Ali Reza Fallahi et al. [4] proposed morphological operations to the image and then extracted the features. Analysis result from MLPNN and SVM show these operations can improve classification results in symmetry and gray scale features but reduce results in texture features. Using SVM, the system gets better result than MLPNN and RBFNN. Because of brain symmetrical structure, symmetrical features have better accuracy and texture features have lower accuracy.

978-1-5090-4929-5©2017 IEEE

Mehadi Jafari et al. [5] introduced hybrid approach using genetic algorithm with SVM. Three sets of features were extracted and took as an input to the classifier. In the first set, statistical features like entropy, energy, mean, standard deviation, kurtosis, skewness, momentum, correlation. In second set, wavelet based feature were considered and in third set, frequency transformation based features were extracted. In this method genetic algorithm plays the role of feature reduction technique and SVM classifies the brain MRI as normal or abnormal. This approach improves the classifier accuracy up to 83.22%.

Another hybrid approach is introduced by El-Dahshan et al. [6]. The brain MRI is an input to the system; features were extracted by discrete wavelet transform, reduced by principal component analysis technique and classify by using feed forward back propagation artificial neural network (FF- BPNN) and KNN. Using these classifiers gives an accuracy of 99% on both training and testing datasets.

Hong Men, et al. [7] present two machine learning algorithm neural network and SVM for classification of brain MRI. They used two kind of support vector machine based on polynomial kernel and radial basis function for different parametric values. The result of this experiment indicates that the support vector machine method is superior to the neural network algorithm.

III. Proposed system

This system has four steps: Preprocessing, morphological filtering, feature extraction and classification.

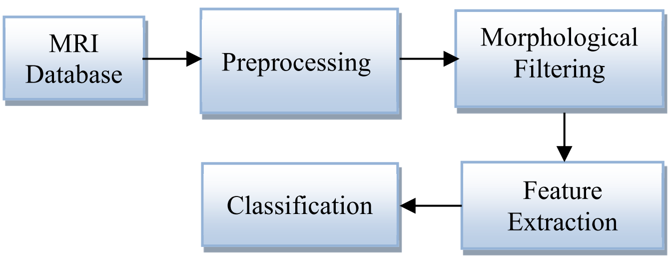


Fig. 1. Block diagram of proposed system

A. Database image

The MRI images for this approach are taken as a clinical database from Sahyandri hospital, Pune, the database contains malignant and benign tumor images and standard database images from a BRATS 2012 database, contain glioma low and high grade images.

B. Preprocessing

The input images are in RGB format. First RGB images are converted into grayscale. The captured medical images are noisy. Most of the medical images are suffered by Rician noise and pepper and salt noise. The preprocessing operations include median filter and power law transformation. Median filtering is a nonlinear filtering technique. It is important to remove the noise at an earlier stage to getting the accuracy at the last stages. Median filter is used to suppress the pepper and salt noise from the image while preserving the edges. The 3x3 mask is used to remove the noise by considering middle pixel of the mask as updating pixel.

Medical images are poor in contrast [8]. In the lower contrast image, neighboring pixels get merged into one another. To improve the contrast of the image, this system uses power law transformation [9]. The power law transformation (S) of the given image is given by

$$S = CT^{\gamma} \tag{1}$$

Where, C is input intensity
γ is the output intensity

C. Morphological filtering

Morphological operation is related to the external features of an image like shape, boundaries and skeleton. In this operation, small mask of different shapes are applied to the image called as structuring element. The structuring element is of different shape such as line, disk and diamond. In this approach disk shape kernel has been used.

The erosion of an image A by kernel element B is given by

$$A \ominus B = \{z \in E \mid B_{\approx} \subseteq A\} \tag{2}$$

The dilation of an image A by kernel element B is given by

$$A \oplus B = \bigcup_{b \in B} A_b \tag{3}$$

In the proposed system, the disk shaped kernel mask of size 5x5 is created. Firstly small object in the object are cleaned by applying erosion operation then the interested part of the image gain the original shape by dilation operation.

D. Feature Extraction

Feature extraction is a technique to represent images in the feature set of object of interest. The different types of features are texture, color, shape. MRI image can be better distinguished by texture features. GLCM is a widely used texture feature extraction technique [10]. Texture classification produces classified output of input images where each texture region is identified by different distances and different direction. For this approach of feature extraction single distance and four directions method is used to extract the GLCM feature. The GLCM extract different features tabulated in TABLE I

978-1-5090-4929-5©2017 IEEE

TABLE I. GLCM FEATURES

Sr. No	Images / Features	1	2	3	4	5	…..	247	248	249	250	251
1	**Contrast**	102.487	187.846	109.8215	146.2739	194.6974	…..	87.68984	94.49233	74.26616	91.18161	85.15297
2	**Homogeneity**	0.77690	0.77208	0.772047	0.775845	0.788496	…..	0.835514	0.822537	0.829509	0.8301	0.831823
3	**Correlation**	0.00657	0.00674	0.007634	0.007463	0.005922	…..	0.010938	0.011135	0.010026	0.009275	0.010897
.	….	……	…..	……	……	…..	…..	….	…..	……	……	…..
.	….	…..	…..	…..	……	…..	…..	….	….	…..	…..	…..
12	**Difference Entropy**	1.23500	1.23008	1.283363	1.237382	1.128355	…..	1.003298	1.100886	1.068918	1.034651	1.015931
13	**INV**	0.80742	0.80283	0.803938	0.806966	0.818463	…..	0.863887	0.848649	0.853229	0.858257	0.860274
14	**INN**	0.99057	0.9891	0.99052	0.989766	0.989384	…..	0.993668	0.992796	0.993409	0.99327	0.993756

E. Classification

In this proposed work, brain MRI is classified by two machine learning algorithms: SVM and KNN.

1. SVM

Support vector machine is a flawless method to find out the hyperplane between two different particular classes in high dimensional feature space which can be used as a classification. Support vector machine is a supervised machine learning algorithm [11]. Supervised learning method processed through two steps: Training and Testing. In the training phase, two type of databases considered the 251 (85 malignant and 166 benign) MRI clinical database images and 80 (50 low grade glioma and 30 high grade glioma) standard MRI images is considered for training and 100 (50 malignant and 50 benign) images of clinical database and 40 (25 low grade glioma and 15 high grade glioma) MRI images for testing respectively.

The SVM classifier classifies the image using linear function as

$$f(x) = W^T X + b \tag{4}$$

Where, X is the training samples
W is the weight assigned
b is bias or offset

SVM classified into two type linear and Non-linear classification. The linear SVM classifier is worthwhile to nonlinear classifier to map the input pattern into higher dimensional feature space. The data which can be linearly separable can be examine using hyperplane and the data which is linearly non-separable those data are examine methodically with kernel function like higher order polynomial. SVM classification algorithm is based on different kernel methods i.e. Radial basic function (RBF), linear and quadratic kernel function. The RBF kernel, is apply on two samples **x** and **x'**, which indicate as feature vectors in some input space and it can be defined as,

$$K(x,x') = \exp\left(\frac{||x-x'||^2}{2\sigma^2}\right) \tag{5}$$

The value of kernel function is decreases in proportion to distance and ranges between zero (in the limit) and one (when x = x').

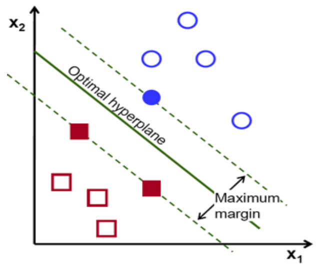


Fig. 2. Optimal hyperplane margin

2. KNN

KNN is a simple and robust classification technique. In this classifier, the testing feature vector is classified by finding the k nearest training neighbor vector. The distance between the training and testing vector is calculated by different distance metric measurement technique such as Euclidean distance, cityblock, chebychev, Minkowski, Mahalanobis, cosine, correlation, Spearman, hamming, Jaccord etc. In this method, Euclidean, cityblock, cosine, correlation distances are measured between training and testing data vector. The Euclidean distance between training and testing vector is given by

$$d(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \tag{6}$$

The label of the smallest distance feature vector is conveying to the testing vector.

978-1-5090-4929-5©2017 IEEE

As feature extracted for training and testing set of images, we get different dimension in some space and these value of extracted feature take as an observation for this characteristic to be its coordinate in that dimension, so set of points in a space. We can then consider the similarity of two different points to the distance between them in a space under some suitable metric.

The way in which the applied algorithm decide which of the points from the training set are enough similar to the point considered when choosing the class to estimate for a new observation is to pick the k closest data points to the new observation, and to take the most common class among these. In this way the k Nearest Neighbor algorithm performed.

The K- nearest neighbor algorithm is as follow [12]:

1. A positive integer value k is defined, along with the new sample.
2. Select the k values in our database which are closest to the new testing sample.
3. We find out the most similar classification of these entries.
4. This is the classification we give to the new sample using the value of k.
5. If the satisfactory results not obtained, changed value of k till correct results not obtained

IV. Result and Discussion

In the proposed work, the brain MRI is processed through four main steps: Image preprocessing, Morphological filtering, feature extraction and supervised classification. The results of each step are shown below.

Case I: Results of malignant brain MRI

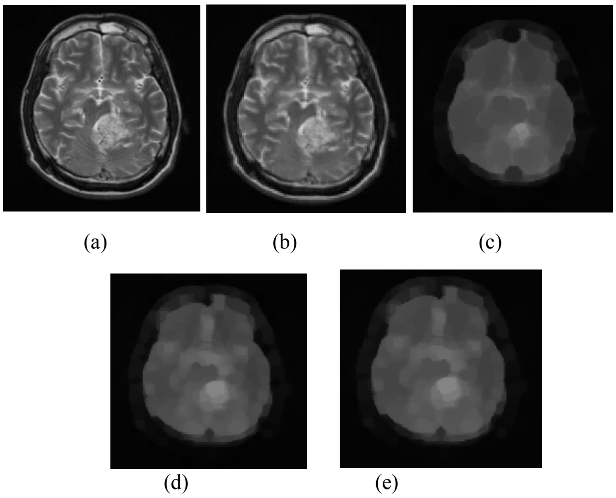


Fig. 3. Preprocessing and morphological operation result of malignant brain MRI (a) Database Image (b) Median filter output (c) Erosion output (d) Dilation Output (e) Power law transformation output

Case II: Results of benign brain MRI

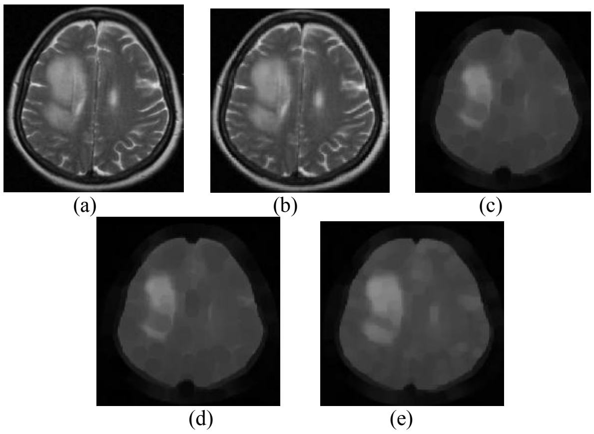


Fig. 4. Preprocessing and morphological operation result of Benign brain MRI (a) Database Image (b) Median filter output (c) Erosion output (d) Dilation Output (e) Power law transformation output

TABLE II. PERFORMANCE ANALYSIS OF SVM CLASSIFIER ON CLINICAL DATABASE

Parameters \ Kernels	RBF	Linear	Quadratic
TP	23	19	25
TN	25	15	22
FP	0	10	3
FN	2	6	0
Sensitivity	92%	76%	100%
Specificity	100%	60%	88%
Accuracy	96%	68%	94%

TABLE III. PERFORMANCE ANALYSIS OF KNN CLASSIFIER ON CLINICAL DATABASE

Parameters \ Distance	Euclidean	Cityblock	Cosine	Correlation
TP	12	12	21	20
TN	25	25	22	22
FP	0	0	3	3
FN	13	13	4	5
Sensitivity	48%	48%	84%	80%
Specificity	100%	100%	88%	88%
Accuracy	74%	74%	86%	84%

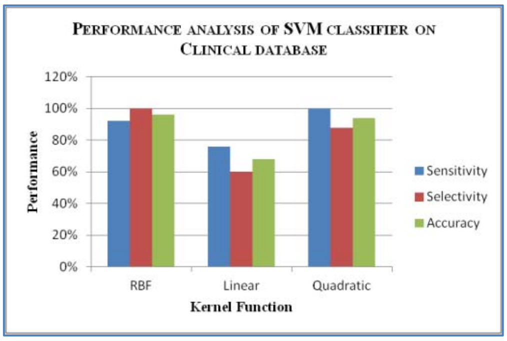


Fig. 5. Comparative analysis of SVM classifier on clinical database

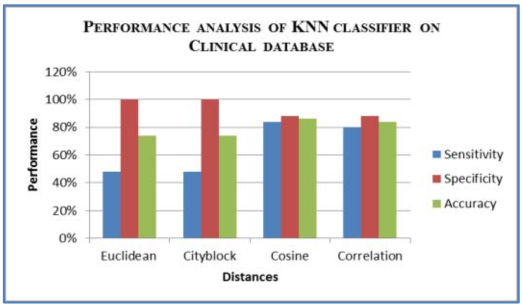
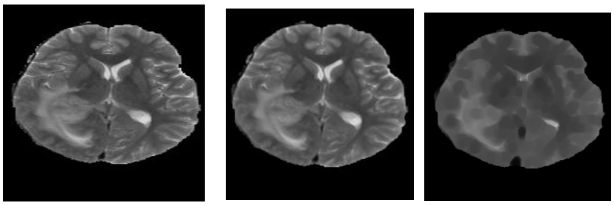
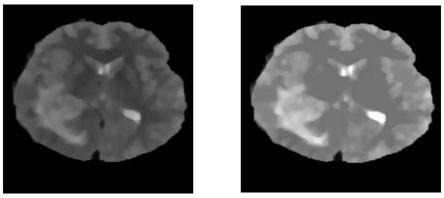


Fig. 6. Comparative analysis of KNN classifier on clinical database

Case III: Results of Low grade glioma images



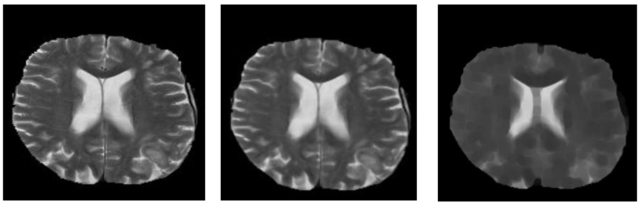
(a) (b) (c)



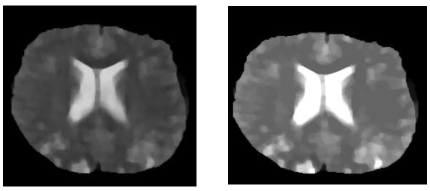
(d) (e)

Fig. 7. Preprocessing and morphological operation result of low grade glioma brain MRI (a) Database Image (b) Median filter output (c) Erosion output (d) Dilation Output (e) Power law transformation output

Case III: Results of High grade glioma images



(a) (b) (c)



(d) (e)

Fig. 8. Preprocessing and morphological operation result of high grade glioma brain MRI (a) Database Image (b) Median filter output (c) Erosion output (d) Dilation Output (e) Power law transformation output

The performance of the Classifiers is calculates based on the three performance parameter; Sensitivity, Specificity and accuracy of the system. The formulae for the metrics are

$$Sensitivity = \frac{TP}{TP+FN} * 100\% \quad (6)$$

$$Specificity = \frac{TN}{TN+FP} * 100\% \quad (7)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (8)$$

Where,

TP = Malignant image is detect as malignant
TN = Benign image is detected as benign
FP = Benign image is detected as Malignant
FN = Malignant image is detected as benign

The comparative analysis of different methods and classifiers are tabulates as follow

TABLE IV. PERFORMANCE ANALYSIS OF SVM CLASSIFIER ON BRATS DATABASE

Parameters \ Kernels	RBF	Linear	Quadratic
TP	19	18	22
TN	15	15	11
FP	0	0	4
FN	6	7	3
Sensitivity	76%	72%	88%
Specificity	100%	100%	73.33%
Accuracy	85%	82.50%	82.50%

TABLE V. PERFORMANCE ANALYSIS OF KNN CLASSIFIER ON BRATS DATABASE

Parameters \ Distance	Euclidean	Cityblock	Cosine	Correlation
TP	18	22	22	18
TN	6	7	6	6
FP	7	3	3	7
FN	9	8	9	9
Sensitivity	67%	73%	71%	67%
Specificity	46%	70.00%	67%	46%
Accuracy	60%	72.50%	70%	60%

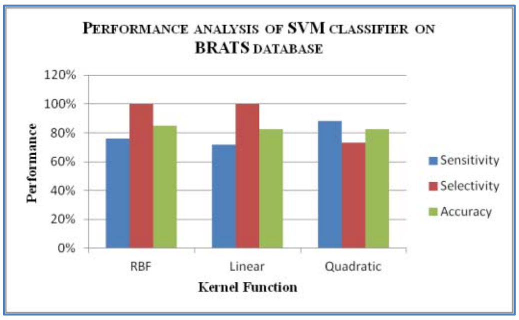


Fig. 9. Comparative analysis of SVM classifier on Brats database

978-1-5090-4929-5©2017 IEEE

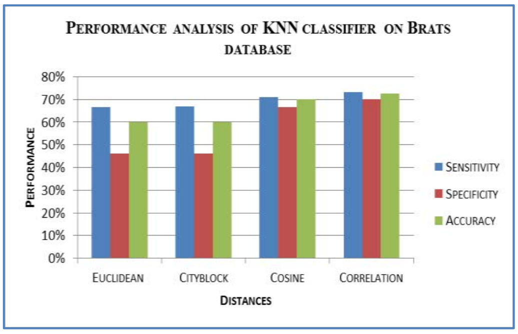


Fig. 10. Comaparative analysis of KNN classifier on Brats database

The performance of the proposed system has been compared with existing method described by V. Vani. The existing system used DWT for feature extraction and features were classified by SVM and KNN algorithms. Comparison of the proposed system with the system proposed by [15] is shown in TABLE VI and graphically represented in Fig. 12.

TABLE VI. COMPARISION OF PROPOSED SYSTEM WITH EXISTING SYSTEM ON BRATS DATABASE

Methods	Precision (%)	Recall (%)	F measure (%)
Proposed Method (SVM)	100	76	86.36
Proposed Method (KNN)	88	73.33	79.99
SVM [15]	66.67	66.67	66.52
KNN [15]	84.31	88.89	83.08

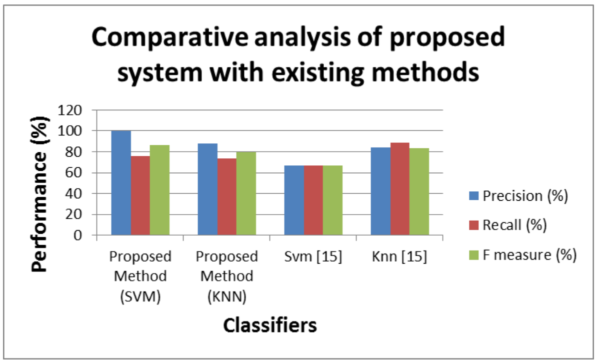


Fig. 11. Comparative analysis of proposed system with existing methods

From comparative analysis it is observed that the performance of the proposed system is better than the existing method.

V. Conclusion

This system is implemented to classify the brain MRI images into malignant and benign type using supervised SVM and KNN classifiers. The proposed system is applied on the 251 images (85 malignant and 166 benign) of clinical database and 80 images (50 low grade glioma and 30 high grade glioma) of brats 2012 training database. The accuracy of the proposed system is 96% and 86% for SVM and KNN respectively for clinical database and 85% and 72.50% for SVM and KNN respectively for Brats database. From the results of proposed system, it is concluded that the accuracy of the SVM classifier is greater than the KNN classifier. It is also found that as we increases the number of training images the performance of SVM classifier increases.

In future, the accuracy of the proposed system can be increased by using the hybrid SVMKNN classifier.

REFERENCES

[1] Hari Babu Nandpuru, Dr. S. S. Salankar, Prof. V. R. Bora, "MRI Brain Cancer Classification Using Support Vector Machine", 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science, 978-1-4799-2526-1.

[2] Janki Naik, Prof. Sagar Patel, "Tumor Detection and Classification using Decision Tree in Brain MRI", Tumor Detection and Classification using Decision Tree in Brain MRI, ISSN: 2321-9939.

[3] Ahmed KHARRAT, Karim GASMI, Mohamed BEN MESSAOUD, Nacera BENAMRANE and Mohamed ABID, "A Hybrid Approach for Automatic Classification of Brain MRI Using Genetic Algorithm and Support Vector Machine", Leonardo Journal of Sciences, Issue 17, July-December 2010, pp. 71-82, ISSN 1583-0233.

[4] Ali Reza Fallahi, Mohammad Pooyan, Hojat Mohammad nejad, "Application of Morphological Operations in Human Brain CT Image with SVM", IEEE journal, 978-1-4244-2902-8/09.

[5] Mehadi jafari, raza, "A hybrid approach for automatic tumor detection of brain MRI using support vector machine and genetic algorithm", global journal of science, engineering and technology ISSN 2322-2441, issue 3, 2012, pp 1-8.

[6] El-Sayed Ahmed El-Dahshan, Tamer Hosny, Abdel-Badeeh M. Salem "Hybrid intelligent techniques for MRI brain image classification" ELSEVIER Digital Signal Processing, vol. 20, pp-433–441, 2010.

[7] H. Men, Y. Gao, Y. Wu and X. Li, "Study on Classification Method Based on Support Vector Machine," 2009 First International Workshop on Education Technology and Computer Science, Wuhan, Hubei, 2009, pp. 369-373..

[8] Hanan Saleh S. Ahmed and Md Jan Nordin,. "Improving Diagnostic Viewing of Medical Images using Enhancement Algorithms", Journal of Computer Science 7 (12): 1831-1838, 2011.

[9] T.Romen Singh, O.Imocha Singh , Kh. Manglem Singh , Tejmani Sinam and Th. Rupachandra Singh, "Image Enhancement by Adaptive Power-Law Transformations", *Bahria University Journal of Information & Communication Technology Vol. 3, Issue 1, December 2010.*

[10] Fritz Albregtsen, "Statistical Texture Measures Computed from Gray Level Coocurrence Matrices", Image Processing Laboratory Department of Informatics University of Oslo November 5, 2008.

[11] V. Vapnik, "The Nature of Statistical Learning Theory", Springer, N.Y., 1995. ISBN 0-387-94559-8.

[12] Shivangi Govil, Nupur Chugh, "Prediction of Defects in Software Using K-Nearest Neighbour Algorithm for Cost Reduction", International Journal of Advanced Research in Computer Science & Technology (IJARCST 2015), Vol. 3, Issue 2 (Apr. - Jun. 2015).

[13] http://www2.imm.dtu.dk/projects/BRATS2012/data.html

[14] http://mipav.cit.nih.gov

[15] V.Vani, M. Kalaiselvi Geetha, "Automatic Tumor Classification of Brain MRI Images using DWT Features", International Journal of Advanced Research in Computer Science, Volume 7, No. 5, September-October 2016, pp. 93-99.