Efficient Detection of Brain Tumor from MRIs Using K-Means Segmentation and Normalized Histogram

Garima Singh,
Electrical Department, School of Engineering
Gautam Buddha University, Greater Noida, U.P
gsing300@gmail.com

Dr. M.A. Ansari
Electrical Department, School of Engineering
Gautam Buddha University, Greater Noida, U.P
mahmadiitr@gmail.com

***Abstract*- Magnetic resonance imaging (MRI) is a technique which is used for the evaluation of the brain tumor in medical science. In this paper, a methodology to study and classify the image de-noising filters such as Median filter, Adaptive filter, Averaging filter, Un-sharp masking filter and Gaussian filter is used to remove the additive noises present in the MRI images i.e. Gaussian, Salt & pepper noise and speckle noise. The de-noising performance of all the considered strategies is compared using PSNR and MSE. A novel idea is proposed for successful identification of the brain tumor using normalized histogram and segmentation using K-means clustering algorithm. Efficient classification of the MRIs is done using Naïve Bayes Classifier and Support Vector Machine (SVM) so as to provide accurate prediction and classification.**

***Keywords*— Brain tumor; Magnetic Resonance Imaging (MRI); Median filter; Normalized Histogram; K-means segmentation; PSNR; MSE.**

I. INTRODUCTION

Brain tumor has been one of the major causes of increase in mortality rate in the world. Brain tumor identification using Magnetic resonance imaging (MRI) is a widely used technique for diagnosis [1]. MRI provides significant information about the anatomy of the brain which is important for the detection of the tumor [2].

In computer vision, Brain tumor segmentation is a process of dividing or segmenting an image into regions by dividing the neighborhood pixels in the image based on some pre-defined features or properties of the pixels [4].

II. PROPOSED ALGORITHM

In this paper, a novel technique has been used which includes Normalization of Histogram and K-means Segmentation. First, input image is pre-processed in order to remove the unwanted signals or noise from it. The histogram of the pre-processed image is normalized and classification of MRI is done. Finally, the image is segmented using K-means algorithm in order to take out the tumor from the MRI. For implementation MATLAB is used.

III. NOISE REMOVAL

Before applying the algorithm to the image it is very important to process the data in order to remove the undesirable parts from the MRI [3].

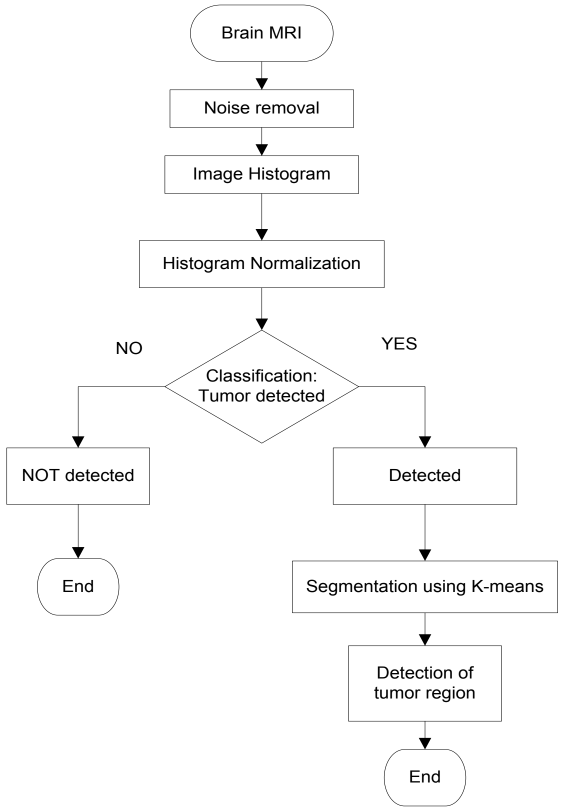


Fig. 1: Block diagram of the Proposed Algorithm

Step 1: The MRI image is converted into the binary form, standard deviation of the binary image is calculated which is used as the threshold value. It is calculated using the given formula where F [a, b] is the grayscale image.

$$S_d = \sqrt{\frac{1}{t-1}\sum_{a,b\in I}(F[a,b]-F_{avg})^2} \quad (1)$$

$$S_d = \sqrt{\frac{\sum_{a,b\in I}F^2[a,b]-tF_{avg}{}^2}{t-1}} \quad (2)$$

Where t is the total intensity value and F_{avg} is the mean of the pixel intensity of the image.

Step 2: The morphological "open" operation is applied to remove the artefacts. The result of this step is a binary image.

978-1-4673-6984-8/16/$31.00 © 2016 IEEE

Step 3: To obtain the tumor portion without any artefact, the original image and the binary image obtained in the step 2 are multiplied together.

A. Median Filter

Median filter is a digital and non-linear filter. It removes the unwanted signals or noise from the image. It sorts all the values of the pixels from the neighborhood in an order from low to high. Then it calculates the median [8, 9].

Median filter offers an advantage as it preserves the edges and does not create unrealistic pixel values when the pixels at the edges are under consideration.

Median filter has the best performance against the impulse noises [3].

B. Adaptive filter

An Adaptive filtering is a linear filter. It removes impulse noise, speckle noise and works well for images with abrupt intensity changes [3].

C. Averaging filter

Averaging filter is one of the simplest and easy to implement filter which is used for smoothing the images. It is a linear and low-pass filter. This filter eliminates the values of the pixels which do not contribute to the representation of its surroundings. It works as a convolution filter based on kernels [3]. Larger is the size of the kernel, more is the smoothing of the image.

D. Gaussian filter

This filter is used to blur the images and can be called a smoothing operator. It removes the fine details which are inherently present in the image. Its impulse response is a Gaussian function which defines the probability distribution of the noise. It is effective for the removal of Gaussian noise. It is a non-uniform, linear and low pass filter which uses a Gaussian function with a given standard deviation [3].

Gaussian function in 2-D form:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{3}$$

Here, σ is the standard deviation.

The intensity of smoothing can be controlled by the factor σ. For high smoothing, larger value of σ is used.

E. Un-sharp masking filter

The un-sharp masking filter is basically a sharpening operator which is used to enhance the edges by subtracting the un-sharp version of the image from the input image [3].

$$g(x,y) = p(x,y) - p_{smooth}(x,y) \tag{4}$$

Here, $p_{smooth}(x,y)$ is the smoothed image, $p(x,y)$ is the input image.

IV. Normalized Histogram

The histogram of an image is a graphical depiction of the number of pixels in the image as a function of their intensity.

Step 1: The captured MRI image is converted into gray-scale image [range 0-255].

Step 2: The random variable 'X' is defined as the pixel intensity. 'X' takes the values from 0 to 255.

The obtained histogram is normalized so that sum of all the probability must be equal to 1.

$$\sum_{x=0}^{255} p(x) = 1 \tag{5}$$

Where, p(x) is the probability of X which is the pixel intensity [0 to 255].

Probability of the defined random variable greater than 150 is calculated using the equation given below:

$$\beta = P(X > 150) * 100 \tag{6}$$

$$\beta = \left(\sum_{x=150}^{256} p(x)\right) * 100 \tag{7}$$

The obtained β is passed through the decision module in order to identify and classify the image into non-tumor and tumor class.

To find out the optimal threshold, the mean of all computed $'\beta'$ values is calculated, say $'\gamma'$. The probability of random variable (X > 150) for test image is computed and decision is made on the comparison between the calculated probability and $'\gamma'$ value.

$$\gamma = [(\sum_{k=1}^{100} \frac{\alpha_k}{100})] \tag{8}$$

The calculated value of $\gamma = 8.9523$, it is considered as the threshold value. Classification of MRI into two classes i.e. TUMOR and NON-TUMOR is done by comparing γ and β.

V. Classification

Classification is used to analyze the image properties and classify the data/information into classes which are dissimilar in nature. It consists of two sections: training section and testing section [3].

A. Naïve Bayes Classifier

Naive Bayes classifiers are profoundly versatile. This classifier is computationally faster and can handle the information efficiently [12].

B. Support vector machine

SVM is a binary classifier. It divides the data points into two classes and maximizes the separation margin between the classes [11, 13].

VI. K-MEANS SEGMENTATION

Segmentation of image is the procedure to partition or divide the image into different regions or segments. The objective of this technique is to represent the image into a more significant and purposeful form which is easier to investigate [5, 6].

K-means segmentation technique is unsupervised in nature. It segments a group of data points into k number of clusters [7, 10].

Step 1: k clusters or groups and cluster center is selected randomly by the user for each group.

Step 2: Euclidean distance is calculated between the cluster center and every pixel of the given image using the formula given below:

$$d = || P(x, y) - c_k || \tag{9}$$

Here, d is the Euclidean distance.

$$d(x,y) = \sum_{i=1}^{p} |x_i - y_i| \tag{10}$$

Step 3: Each pixel is assigned to a cluster with the closest center depending upon the Euclidean distance.

Step 4: The position of the cluster center is re-calculated using the relation given below:

$$c_k = \frac{1}{k} \sum_{y \in c_k} \sum_{x \in c_k} P(x,y) \tag{11}$$

Where c_k is the new cluster center.

Step 5: The process is repeated until it converges.

Step 6: Cluster pixels are then reshaped into an image.

In Fig. 3, (a) image filtered using median filter, (b) image filtered using adaptive filter, (c) image filtered using averaging filter, (d) image filtered using un-sharp masking filter, (e) image filtered using Gaussian filter with σ=2, (f) image filtered using Gaussian filter with σ=4.

(a) Median filter has performed very well. Noise is reduced in the filtered image and the edges are preserved. The quality of the filtered image is good.
(b) Adaptive filter has shown satisfactory result. Noise is reduced but the edges are not clear.
(c) Averaging filter has shown satisfactory result by reducing the noise but edges are not clear.
(d) Median filter has performed very well. Noise is reduced in the filtered image and the edges are preserved. The quality of the filtered image is good.
(e) Adaptive filter has shown satisfactory result. Noise is reduced but the edges are not clear.

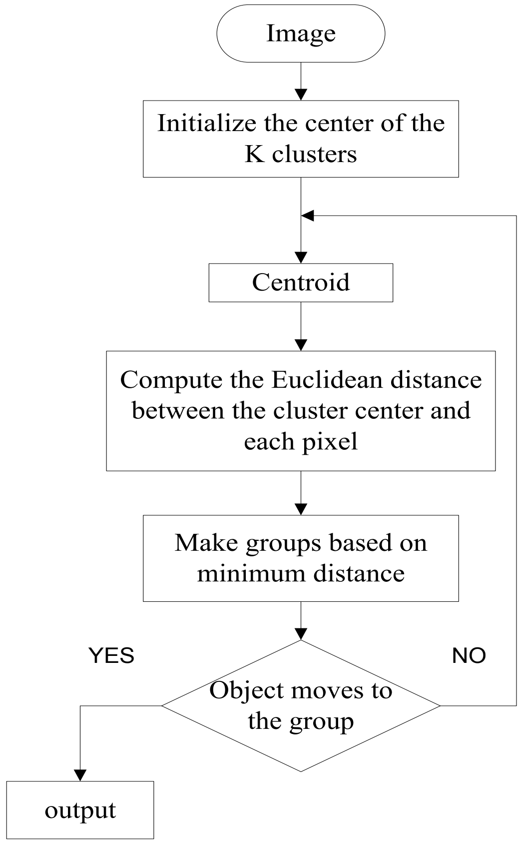
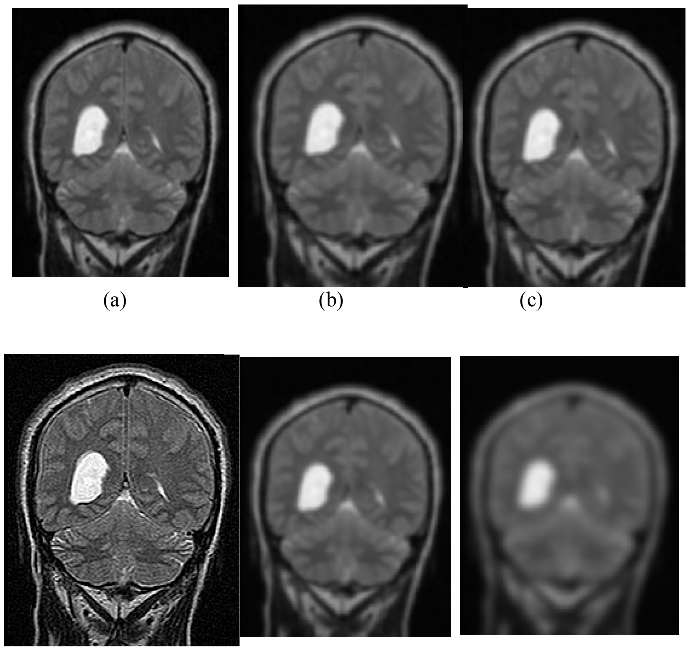


Fig. 2: Block diagram of the K-means clustering

VII. RESULTS AND DISCUSSION

A. Pre-processing results

To study the proposed technique, MRI data with brain tumor has been evaluated.



(a) (b) (c)

(d) (e) (f)

Fig. 3: Pre-processing results using different filters

(f) Averaging filter has shown satisfactory result by reducing the noise but edges are not clear.

(g) Un-sharp masking filter has shown satisfactory result. Desired reduction of noise is not done. Edges and boundaries have become more clear and sharp.

(h) Gaussian filter with standard deviation σ=2 has given satisfactory result. The image is blurred and smoothing is done to an undesirable extent.

(i) Gaussian filter with σ=4 has given poor result. The filtered image is blurred to a very high extent.

TABLE I. PSNR AND MSE VALUES OF FILTERED IMAGES

FILTERS	PSNR	MSE
Median filter	78.7316	0.0039
Adaptive filter	74.8801	0.0040
Averaging filter	68.8675	0.0038
Gaussian filter	68.0017	0.0041
Un-sharped masking filter	65.4431	0.0071

The table 1 shows Peak Signal to Noise Ratio and Mean Square Error values of different filters.

Median filter has given the highest PSNR and lowest MSE, therefore, the image filtered by it is used for further classification and segmentation.

B. Classification results

In this paper, Naive Bayes classifier and SVM are used for classification. A dataset of 110 brain images (MRI) is taken from Yatharth Hospital, Noida [14]. Histogram of the input image is normalization and the values of γ and β are calculated. The calculated value of γ is 8.9523, it is considered as the threshold value. The images are classified into 'Tumor MRI' and 'Non-Tumor MRI' by comparing the values β with γ.

$\beta \geq \gamma$; Tumor detected

$\beta < \gamma$; Tumor not detected

TABLE II. RESULTS OF THE NORMALIZED HISTOGRAM

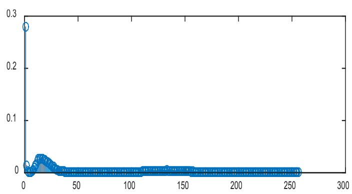
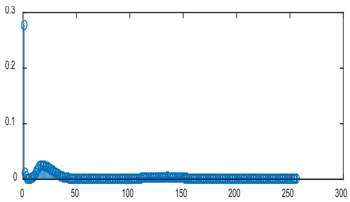
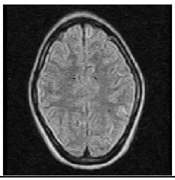
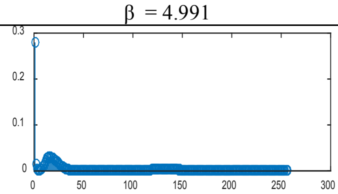
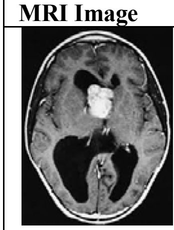
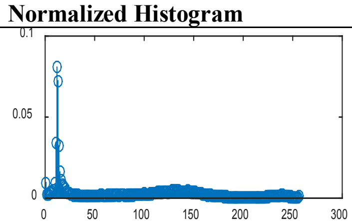
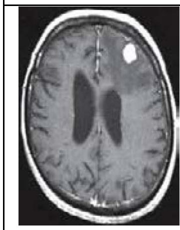
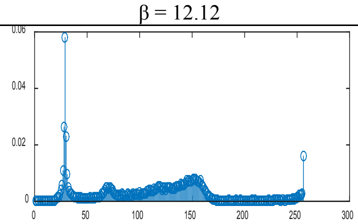
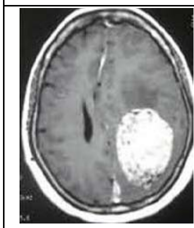
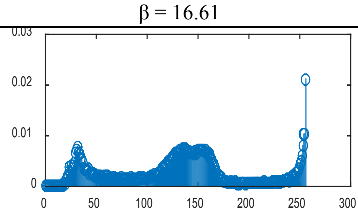
MRI Image	Normalized Histogram	Decision
	 β = 6.484	Tumor not detected
	 β = 4.991	Tumor not detected
	 β = 4.887	Tumor not detected

TABLE IV. RESULTS OF THE NORMALIZED HISTOGRAM

MRI Image	Normalized Histogram	Decision
	 β = 12.12	Tumor detected
	 β = 16.61	Tumor detected
	 β = 26.80	Tumor detected

The table 2 and 3 show the normalized histograms of MRIs and the value of β for the images.

Images with brain tumor are classified into one group and non-tumor brain MRIs are classified into another group. The efficiency of both the classifiers is computed.

Percentage efficiency of classifiers:

Naive Bayes Classifier = 87.23%

SVM Classifier = 91.49%

It can be concluded by comparing the percentage efficiency of classification that SVM is much better than Naive Bayes classifier. After classification, the image in which tumor was detected is segmented using K-means clustering algorithm.

C. Segmentation results

The results of K-means segmentation for tumor images:

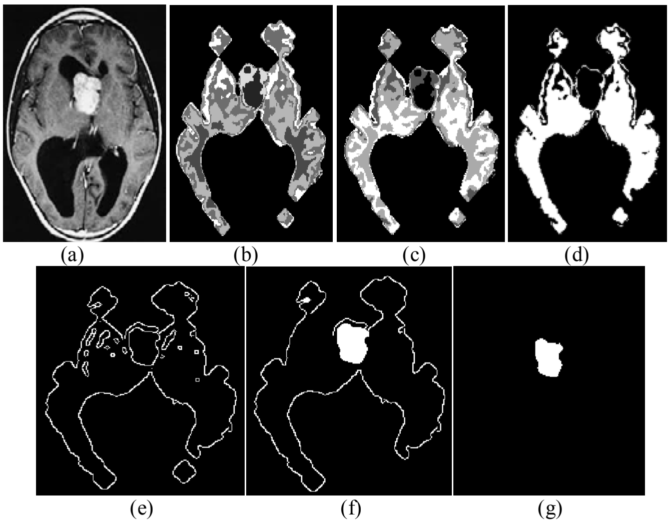


Fig. 4: Clustering results using k-means algorithm for image MRI1

In fig. 4, the image is converted from RGB to grayscale. Thresholding is done using standard deviation to convert image (a) into binary image. The structuring element used here is 'diamond'. Morphological steps are applied to the image (a).

Morphological Opening is applied to image (a) after converting it into binary image. Opening operation erodes the foreground pixels which the structuring element can not reach when it is slided onto the image (a).

Median filter [5*5] is applied to remove the noise. Initially, value of clusters is set randomly, since, K-means is an unsupervised learning technique. Initial value of K is taken as 8. Image (b) is the first cluster. Then, K is taken as 4 which gives image (c) as the second cluster. Finally, K is taken as 2 which gives image (d) as the third cluster.

Morphological operation Dilation is applied on the image. Structuring element used here is 'line' with length=4, angle= 40° and length=2, angle= 90°, in order to increase the thickness of the noisy pixels, image (e) is obtained.

Canny edge detection technique is applied on image (e). Opening operation is used again to remove the noisy pixels. In the image (f), the edges of the tumor area are detected. Erosion is applied on the image (f) and unwanted pixels are removed, leaving behind the tumor region. In image (g), the accurate tumor region is detected.

In fig.7, Median filter, Morphological closing, dilation is applied on image. Canny edge detection technique is applied on image (e). Erosion is applied on image (f) but the tumor region can not be detected. The algorithm does not give desired result in fig. 7. After applying morphological Opening, Dilation and Erosion, K-means technique is unable to define the tumor edges and thus, unable to detect the brain tumor in the given MRI. From the above result, it can be concluded that the proposed algorithm does not work efficiently for all the cases.

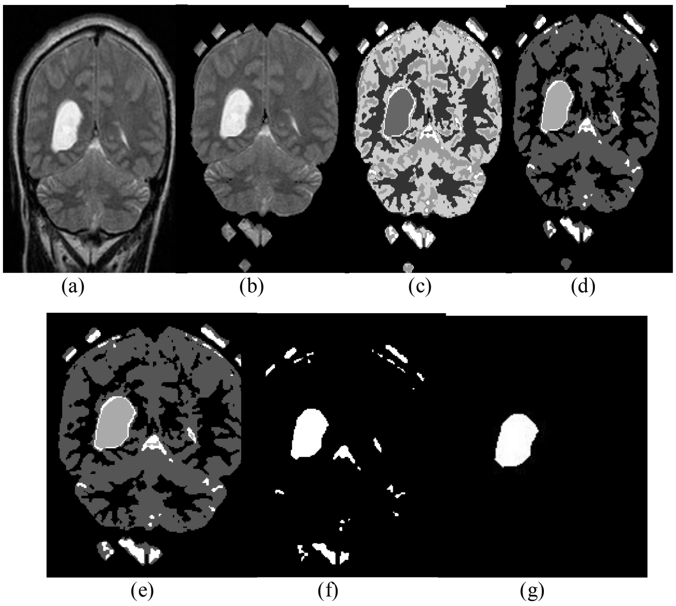


Fig. 5: K-means clustering results of MRI2

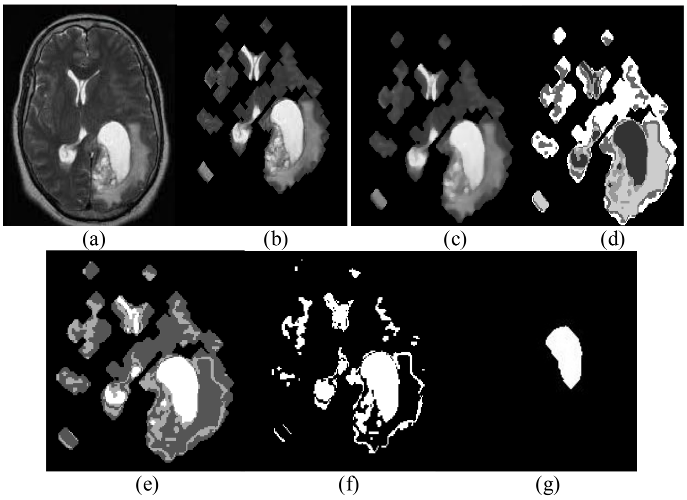


Fig. 6: K-means clustering results of MRI3

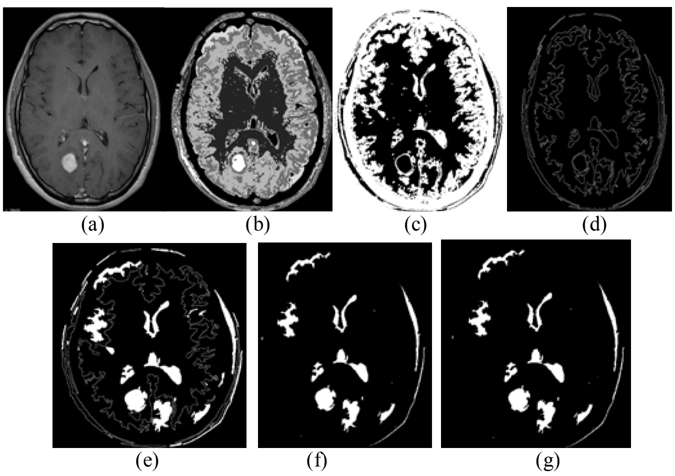


Fig. 7: K-means clustering results of MRI4

VIII. Conclusion

Brain tumor is a standout amongst the well-known brain diseases, so recognition and segmentation of the tumor is imperative in medical analysis. A study on pre-processing, segmentation and classification of brain MRI is presented. The brain tumor from MRI is segmented using k-clustering algorithm. Comparative analysis of Median, Adaptive, Averaging, Gaussian and Un-sharp masking filters is done on the basis of PSNR and MSE.

On the basis of PSNR, Median filter works best for noise removal (PSNR= 78.7316) and by calculating MSE, Averaging filter has given the best result (MSE=0.0038). The images were classified into 'tumor image' and 'non-tumor image' after histogram normalization using Naïve Bayes classifier and SVM. Efficiency of SVM = 91.49% and Naïve Bayes = 87.23%. It is concluded that SVM has given better efficiency than Naive Bayes classifier.

The proposed method has some limitations because in some tumor images, the results were not satisfactory, the detection of tumor was not accurate. The algorithm could not find out the precise or accurate boundary of the tumor region. There is a scope of improvement in the algorithm for better detection of the tumor.

In the future, improvement in the proposed algorithm can be done by working on the limitations, the quality of the output images can be improved by using better morphological operations.

References

[1] E. El-Dahshan, H. Mohsen, K. Revett and A. Salem, "Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm", *Expert Systems with Applications*, vol. 41, no. 11, pp. 5526-5545, 2014.

[2] N. Nabizadeh and M. Kubat, "Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features", *Computers & Electrical Engineering*, vol. 45, pp. 286-301, 2015.

[3] R. Gonzalez, R. Woods and B. Masters, "Digital Image Processing, Third Edition", *J. Biomed. Opt.*, vol. 14, no. 2, p. 029901, 2009.

[4] K. Mardia and T. Hainsworth, "A spatial thresholding method for image segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 6, pp. 919-927, 1988.

[5] K. Qureshi and S. Hussain, "A Comparative Study of Parallelization Strategies for Fractal Image Compression on a Cluster of Workstations", Int. J. Comput. Methods, vol. 05, no. 03, pp. 463-482, 2008.

[6] M. Wu, C. Lin and C. Chang, "Novel image copy detection with rotating tolerance", Journal of Systems and Software, vol. 80, no. 7, pp. 1057-1069, 2007.

[7] H. Rajini.N and B. Bhavani.R, "Automatic MR Brain Tumor Detection using Possibilistic C-Means and K-Means Clustering with Color Segmentation", International Journal of Computer Applications, vol. 56, no. 2, pp. 11-17, 2012.

[8] B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation",J. Electron. Imaging, vol. 13, no. 1, p. 146, 2004.

[9] M. Waseem Khan, "A Survey: Image Segmentation Techniques", IJET, pp. 89-93, 2014.

[10] S. Madhukumar and N. Santhiyakumari, "Evaluation of k-Means and fuzzy C-means segmentation on MR images of brain", The Egyptian Journal of Radiology and Nuclear Medicine, vol. 46, no. 2, pp. 475-479, 2015.

[11] J. Luts, A. Heerschap, J. Suykens and S. Van Huffel, "A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection", Artificial Intelligence in Medicine, vol. 40, no. 2, pp. 87-102, 2007.

[12] P. R., Y. Sharma and M. Kshirasagar, "Performance Analysis of Intrusion Detection Systems Implemented using Hybrid Machine Learning Techniques", International Journal of Computer Applications, vol. 133, no. 8, pp. 35-38, 2016.

[13] J. Luts, A. Heerschap, J. Suykens and S. Van Huffel, "A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection",Artificial Intelligence in Medicine, vol. 40, no. 2, pp. 87-102, 2007.

[14] Yatharth Hospital, Noida, http://www.yatharthhealthcare.com/