Segmentation of Tumours from Brain Magnetic Resonance Images using Gain Ratio Based Fuzzy C-Means algorithm

Swetha Sridharan, J. Angel Arul Jothi, V. Mary Anita Rajam

Department of Computer Science and Engineering

College of Engineering, Guindy, Anna University,

Chennai- 600025, India

Abstract— *Image segmentation is a challenging step in medical image processing. It deals with separating the area of interest from the images. Among the popular methods of segmentation, clustering is an important one which clusters the image pixels based on the similarity of pixels based on the centroid values. But clustering faces a major problem of inhomogeneous clusters. In this paper we propose a new segmentation algorithm called Gain Ratio based Fuzzy C-Means algorithm (GRBFCM) which overcomes the problem of inhomogeneous clusters. This GRBFCM algorithm exploits the concepts of information theory in combination with fuzzy logic. The proposed algorithm is applied to brain Magnetic Resonance Image (MRI) to segment tumor region. The segmentation accuracy is evaluated using parameters like Dice Coefficient (DC) and Jaccard Similarity (JS). We found that the proposed algorithm has better performance.*

Keywords— *Image Segmentation, GRBFCM, Dice Co efficient, Jaccard Similarity*

I. INTRODUCTION

The study and analysis of large and complex MR Image datasets is emerging as a challenging task for physicians as they must extract critical information manually. This becomes time-consuming and is easily prone to errors. Hence computerized methods are adapted for disease diagnosis and testing. This type of automatic process enables to provide efficient treatment to the patients [6].

The first step in the image processing technique involves the acquisition of images. This paper uses brain images acquired using the Magnetic Resonance Imaging technique which helps the physicians to observe the internal organs and locate any abnormalities if present [7]. The next step is to study these images which involve performing various operations on them such as segmentation and classification. Segmentation is usually referred to as the division of an image into meaningful structures. Classification is a supervised process of categorizing an image into predefined classes.

The proposed work in this paper can be summarised as follows. The Brain MR images acquired are pre-processed to remove unwanted noise from them. The next step involves extraction of texture features from the images using the Local Binary Pattern (LBP) technique. The features that are extracted using LBP are fed to the classifier and the classification process is carried out using the Support vector machine (SVM) classifier. It determines if the images belong to the normal or the abnormal class. In the next step, the abnormal images alone are considered and the Gain Ratio Based Fuzzy C-Means algorithm (GRBFCM) is applied to segment the tumour region from these images.

The GRBFCM algorithm is based on the concepts of information theory. The algorithm is carried out in two phases. At the end of the second phase we get the output of segmentation based on the GRBFCM algorithm.

This paper is further organised as follows. The section II deals with the literature review. Section III explains the proposed method in detail. Section IV gives the details on the data set used and the ground truth images. The Section V explains the experimental setup. Section VI deals with the evaluation parameters. Section VII provides the results of the experiment and their discussions. The Section VIII provides the conclusion about the proposed work.

II. LITERATURE REVIEW

There are a lot of algorithms which have been already proposed for brain tumour segmentation. The following paragraphs discuss these algorithms.

Eman Abdel-Maksoud et al. [1] have proposed a hybrid clustering technique to segment the tumour region from MR images. According to this technique, the number of iterations of Fuzzy C-Means is reduced by initializing the right cluster centers. This also minimizes execution time. Though the system achieves accuracy, it does not deal with intensity issues.

Hakeem Aejaz Aslam et al. [2] have proposed a system for segmentation based on pillar K-Means algorithm. This method includes a new mechanism for grouping the elements of high resolution images in order to improve accuracy and reduce the computation time. The system uses K-Means for image segmentation optimized by Pillar algorithm. The Pillar algorithm considers the placement of pillars. The pillars should be located as far from each other to resist the pressure distribution of a roof. This algorithm applies the same concept to the centroids in the K-Means clustering. This algorithm is able to optimize the K-means clustering for image segmentation in the aspects of accuracy and computation time. But the algorithm suffers from intensity inhomogeneity issues.

Arun K Gupta et al. [3] have described Knowledge Based Brain Tumour Segmentation system (KBBTS) using histogram interpretations for predicting brain tumour area from trans-axial Magnetic Resonance Imaging (MRI). This system showed significant improvements over traditional threshold-based tumour segmentation methods. Although KBBTS is not designed to work in real time, it serves as potential research advancement for real time brain tumour segmentation using computer-aided systems with high performance.

Roshan G. Selkar et al. [4] have designed a system for brain tumour detection which helps in finding the exact size, shape, boundary extraction and location of tumour. The system consists of three stages to detect and segment a brain tumour. An efficient algorithm is proposed for tumour detection based on segmentation and morphological operators. Firstly quality of scanned image is enhanced and then morphological operators are applied to detect the tumour in the scanned image. After that edge detection operator is applied for boundary extraction and to find the size of the tumour.

Recently, Asifullah Khan et al. [5] have proposed a method for tumour segmentation using Fuzzy C-Means based on Information Gain called IGFCM. The method works in two phases with the first phase involving Fuzzy C-Means membership. The second phase involves calculating the information gain for various classes and updating the membership as a new sorted membership function to segment the tumour region.

III. PROPOSED METHODOLOGY

This section provides a detailed description of the proposed algorithm. The proposed work is explained using the Fig. 1 which depicts the work flow.
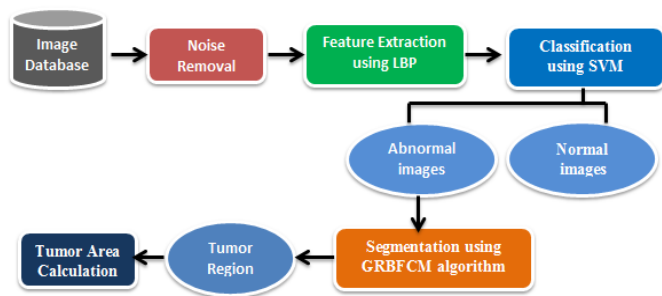

Fig. 1 Block diagram for the proposed method

The image database contains the images of human brain captured using the MRI technique. This project uses T1 weighted MR images of 1mm thickness. It is estimated that the MR images in the database contain 3% noise. The database contains both normal (non-tumour) and abnormal (tumour) images of the brain. These images from the database are classified as tumour and non-tumour images using the Support vector machine classifier. Only those images which contain tumour are given as input to the Gain Ratio Based Fuzzy C-Means (GRBFCM) segmentation algorithm. The entire process of classification and segmentation is explained in the following subsections.

A. Noise Removal

The brain MRI is prone to random noise due to many external factors like transmission noise, noise due to machinery wear and tear etc. Hence, in this work pre-processing is done to remove undesirable signals from the image using median filter. Median filtering is a nonlinear operation often used in image processing to reduce random noise. Median filtering is able to remove the noise and outliers without reducing the sharpness of the input image. The medfilt2() function of MATLAB is used to remove the noise [17]. It performs median filtering over the image matrix in two dimensions. Each output pixel contains the median value in the neighbourhood around the corresponding pixel in the input image. The medfilt2 pads the image with zeros on the edges. The output of the filter is the de-noised image which can be used for further processing.

B. Feature Extraction

Image representation requires enormous amount of data which occupies large memory and time. In order to overcome this, the features are extracted from the image that contains relevant information about an image. This work focuses on extracting the texture based features that provides higher order description of an image and includes information about the spatial distribution of gray tones. The texture features defines the homogeneity between regions of an image.

In this work, the Local Binary Pattern (LBP) technique is applied to the pre-processed image and then texture features like contrast, entropy, correlation, homogeneity and energy [15] are extracted. LBP is a simple implementation method which binarises each pixel of the image. It starts by dividing the image into windows of required dimension. This technique considers the center pixel of the window and follows a circular path to compare the center pixel with one of its neighbours, say p_i. If the value of the center pixel is greater than its neighbour we rewrite the value of the center pixel as 1 otherwise 0. The pixel p_i is compared with its neighbour p_{i+1} which appears in the circular path traced. Then, the value of the pixel p_i is replaced with a 1 if it is greater than p_{i+1} or 0 if it has a lesser value. By proceeding in this fashion the entire image is binarized and we get the LBP image [13]. From the LBP image the texture features are extracted which forms the feature vector.

Let P_d represent the distance between two pixels i and j, N_g be the total number of gray levels in the entire image, μ_i and σ_i represent the mean and standard deviations of an ith pixel in an image respectively. Then, the definitions of the features extracted in this work are as follows.

1) Energy: It is also known as angular second moment or uniformity which measures the homogeneity in the image given by the formula in the equation (1)

$$Energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_d^2(i,j) \qquad (1)$$

2) Entropy: Entropy measures the loss of information or message in a transmitted signal. It is given by the formula in the equation (2).

$$Entropy = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_d(i,j) \log(P_d(i,j)) \quad (2)$$

3) Contrast: Contrast measures the quantity of local changes in an image. It reflects the sensitivity of the textures in relation to changes in the intensity. It is given by the formula in the equation (3).

$$Contrast = \sum_{n=0}^{N_g-1} n^2 \sum_{|i-j|=n} P_d(i,j) \quad (3)$$

4) Correlation: This feature measures how correlated a pixel is to its neighborhood. It is the measure of gray tone linear dependencies in the image. The formula for correlation is explained in the equation (4).

$$Correlation = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu_i) P_d(i,j)}{\sigma_i \sigma_j} \quad (4)$$

5) Homogeneity: Homogeneity measures the similarity of pixels. It becomes large if local textures only have minimal changes. It is given by the formula in the equation (5).

$$Homogeneity = \sum_{i=1}^{N_r} \sum_{j=1}^{N_r} \frac{P_d(i,j)}{1+|i-j|} \quad (5)$$

The feature vector thus obtained is provided as input to the classifier.

C. Classification using Support Vector Machine

SVM classifier can be used to classify both linear and non-linear data. It deals with the problem of classifying data by transforming the original training data into a higher dimension using a non-linear mapping. The method searches for a hyperplane within this new dimension that separates the data effectively which is called as the maximal margin hyperplane. The maximum marginal hyper plane is one which provides maximum separation between the classes. SVM finds the hyper plane using the Support vectors and margins. The equation of the hyperplane is

$$w * x + b = 0 \quad (6)$$

In the equation (6), w is the weight vector, x is the training vector and b is the bias. The bias b can be replaced by an additional weight term w0 and the equation can be rewritten as equation (7).

$$w_o + w_1 x_1 + w_2 x_2 = 0 \quad (7)$$

Thus any pixel value above the separating hyperplane satisfies the equation (8)

$$w_o + w_1 x_1 + w_2 x_2 > 0 \quad (8)$$

Similarly, any pixel value below the separating hyperplane satisfies the equation (9)

$$w_o + w_1 x_1 + w_2 x_2 < 0 \quad (9)$$

SVM involves two phases namely training and testing. The SVM training takes as input the feature vectors of the training sample images [12]. It chooses an appropriate margin between the classes and uses an optimal hyperplane. It trains using the features and classifies the image as abnormal or normal class. The SVM trained system is supplied with the testing image that is to be classified. The system extracts the features from the image and checks the data against the trained database.

After this process, the SVM classifier returns the class of the testing image [14] [16].

The output of this module is to predict whether the images are normal or if they contain tumour. Only those images which contain the tumour or in other words only those images classified as abnormal images are further fed as input to the segmentation algorithm in the next module.

D. Segmentation using Gain Ratio Based Fuzzy C-Means(GRBFCM)

This work proposes GRBFCM algorithm to segment the tumour from the brain MR Images. The proposed algorithm is an improvisation of the work done by Astifullah et.al. They have proposed the Information gain based Fuzzy C-means (IGFCM) algorithm for clustering of the plaque in the carotid artery images. The proposed algorithm works in a similar pattern like IGFCM except that instead of the information gain values, it uses the gain ratio. This is due to the fact that the IGFCM suffers from the problem of favouritism. In other words, it is biased towards the class of pixels having a majority of the values for class labels. GRBFCM helps to overcome this disadvantage of the IGFCM. It also provides normalization to the information gain values using the split info.

The GRBFCM algorithm works in two phases. In the first phase, the Fuzzy C-Means (FCM) algorithm is initiated. As a first step, the FCM algorithm initialises the cluster centers randomly. In the next step, FCM algorithm assigns membership to each pixel in the image corresponding to each cluster center. This assignment is done using the distance measure between the cluster center and the pixel on the basis of distance between the cluster center and the data point. The membership function returns a larger value for a pixel with respect to a cluster if the pixel is closer to the cluster. The input image is segmented by FCM. The output or the membership function is used as input to the next phase. In the second phase, for each pixel in the FCM-segmented image, entropy and gain ratio are computed. The proposed GRBFCM algorithm updates the fuzzy membership values of a pixel by swapping. The order of swapping of the membership values is based on the Gain Ratio values obtained. The cluster centroids of the pixel are updated. Finally, the FCM iteration is performed using the membership values and the updated cluster centroids. This process continues until the difference between cluster centroid values in consecutive iterations reduces below a pre-determined threshold (stopping criterion). The de-fuzzification of the membership function u_{jk} yields the final image segmented by GRBFCM algorithm.

Let C be the total no of clusters (here C=4), N be the total number of pixels in the image and n_i be the number of pixels belonging to i, where $i \in C$.

The membership function of the Fuzzy C-Means algorithm is represented by u_{jk} and the new updated membership function is denoted by u_{jk}^{new}. The cluster centroids are denoted as v_{kn}. Then, the proposed

GRBFCM algorithm proceeds as follows. The values i and j denote the clusters, where $i, j \in C$.

Step 1: The FCM algorithm is initiated using the membership function u_{jk} and the centroid values v_{kn}.

Step 2: Calculate the probability of each class

$$p_i = \frac{n_i}{N} \qquad (10)$$

In equation (10), n_i represents the number of pixels in the image which belong to class i and N represents the total number of pixels in the image.

Step 3: Calculate the entropy of each class using

$$entropy(i) = -\sum_i p_i \log_2 p_i \qquad (11)$$

In equation (11), p_i is the probability of 'i' th cluster in a certain neighbourhood.

Step 4: Calculate the entropy between classes where entropy (*i,j*) for each two-class combination is calculated using the equation (12)

$$entropy(i, j) = -\sum_i p_i \log_2 p_i - \sum_j p_j \log_2 p_j \qquad (12)$$

Step 5: The gain for each class is calculated using the formula

$$Gain(i) = entropy(i) - entropy(i, j) \qquad (13)$$

In equation (13) the gain value is calculated as the difference between the entropy of i [th] class and the inter class entropy between i and j.

Step 6: The split info for each class is computed using the formula

$$SplitInfo(i, j) = -\sum_{i \in [1,C]} \sum_{j \in [1,C]} \frac{n_{(i,j)}}{N} \log_2 \frac{n_{(i,j)}}{N} \qquad (14)$$

In equation (14), $n_{(i,j)}$ represents the number of pixels belonging to classes i, j.

Step 7: Now, from this the gain ratio is calculated according to equation (15)

$$GainRatio(i) = \frac{Gain(i)}{SplitInfo(i, j)} \qquad (15)$$

Step 8: The membership function vector is sorted in descending order based on Gain Ratio values.

Step 9: Update the membership function using following expression in equation (16)

$$u_{jk}^{new} = u_{rank(GR,GR(i)),k}^{sorted} \forall (j,k) \in [1,C] \qquad (16)$$

where GR is the sets of gain ratio values for all classes and Rank (GR, i) returns the rank of number i within a set of numbers GR. u_{jk}^{new} returns the new membership function of the GRBFCM algorithm.

Step 10: Use the new membership function u_{jk}^{new} and the cluster centroids v_{kn} to perform FCM iteration again.

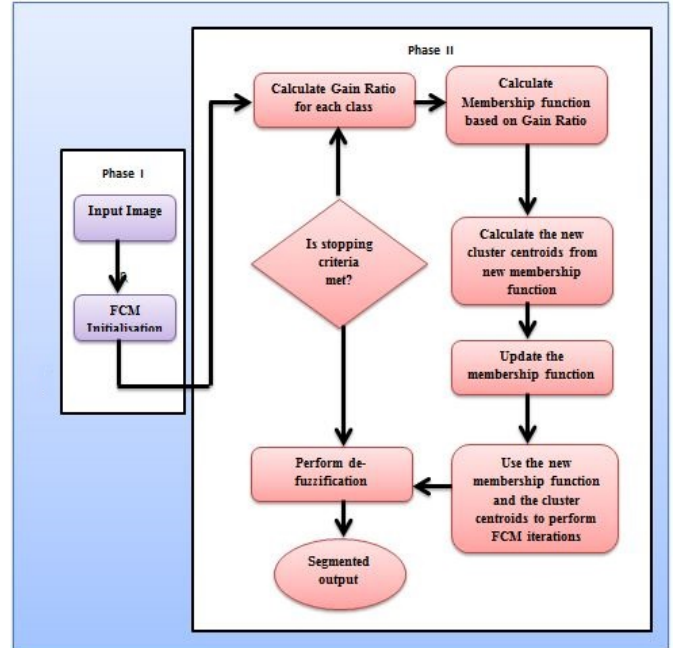**Step 11**: Repeat all the steps until the stopping criterion is met.

Fig. 2 GRBFCM algorithm in two phases.

The Fig.2 describes the complete GRBFCM algorithm as a two phased process. The output of segmentation is 4 cluster regions. From the segmented clusters tumour region is identified. The tumour area is calculated from the segmented output using the formula,

$$TumorSize = \sqrt{P} * 0.264 \qquad (17)$$

In equation (17), P is the number of pixels identified as tumour across an image of height H and width W.

IV. DATASET AND GROUND TRUTH

This project utilizes the Brain MR Images of around 60 patients collected manually. The images are TI - weighted MRI with around 5 % noise factor. The normal MR images refer to those images which do not contain tumour and abnormal images refer to those which contain tumour in them. For the classifier, the training process is carried out using the set of training images present in the dataset and for testing we provide a new set of images. For segmentation only those images which are classified as abnormal are considered. The corresponding ground truth images are images obtained by manual segmentation.

V. EXPERIMENTAL SETUP

The experiment is carried out using the given dataset in the MATLAB environment. The local binary pattern is used to obtain the features from an image using a 8x8 neighbourhood. The SVM classifier is trained using the

LBP feature vector obtained and stored as a .mat file. The test images are provided to the Support Vector Machine classifier which classifies the images as either normal or abnormal. Those images that are classified as abnormal are segmented using the GRBFCM algorithm. Next, the entire process is carried out using the Probabilistic Neural Network (PNN) classifier [8] for the same set of images. The classifiers efficiency is evaluated using the parameters accuracy, sensitivity, specificity. In case of segmentation, the proposed segmentation algorithm is compared with the Rough Fuzzy C- Means algorithm [9] and the K-Means algorithm. The three segmentation algorithms are compared and evaluated using the parameters namely, Dice Coefficient and Jaccard Similarity.

VI. EVALUATION

The parameters accuracy, sensitivity and specificity are used for the evaluation of classifiers performance. The parameters Dice coefficient and Jaccard Similarity are used for evaluating the segmentation accuracy. The following paragraphs give the definitions and the formulae for each parameter.

A. Accuracy

Accuracy of a system can be referred to as the system's ability to produce the actual results for segmentation when the input is provided. A system with higher accuracy is reliable than a system with lower accuracy. Accuracy is given by the formula in equation (18).

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad (18)$$

where, TP- True and positive TN- True and negative FP- False and positive FN – False and negative. True Positive (TP) in equation (19) indicates that patient suffered diseases and result was positive; True Negative (TN) in equation (20) indicates that patient did not suffer disease and result was negative; False Positive (FP) in equation (21) indicates that patient did not suffer disease and result was positive; False Negative (FN) in equation (22) indicates that patient suffered disease and result was negative.

$$TP = \frac{\text{Number of images having tumour}}{\text{Total number of images}} \qquad (19)$$

$$TN = \frac{\text{Number of images not having tumour}}{\text{Total number of images}} \qquad (20)$$

$$FP = \frac{\begin{array}{c}\text{Number of images not having tumour}\\\text{and detected positive}\end{array}}{\text{Total number of images}} \qquad (21)$$

$$FN = \frac{\begin{array}{c}\text{Number of images not having tumour}\\\text{and detected negative}\end{array}}{\text{Total number of images}} \qquad (22)$$

B. Sensitivity and specificity

Sensitivity and Specificity are the two prime performance parameters used in medical statistics. They are denoted by the formulae in equation (23) and equation (24) respectively.

Sensitivity denotes the value which is the number of images which were correctly identified with a tumour.

$$Sensitivity = \frac{(TP)}{(TP + FN)} \qquad (23)$$

Specificity denotes the value which is the number of images which were correctly identified as not having a tumour.

$$Specificity = \frac{(TN)}{(TP + FN)} \qquad (24)$$

C. Dice coefficient

Dice coefficient, is a statistic used for comparing the similarity of two shapes. It measures the accuracy of segmentation between the output of the segmentation algorithm and the ground truth images. It is defined as the ratio between twice the intersection and union of the region obtained by automated segmentation X and ground truth region [11]. The formula for Dice Coefficient is given in equation (25)

$$S = \frac{2|X \cap Y|}{|X| + |Y|} \qquad (25)$$

D. Jaccard Similarity

The Jaccard Similarity index is a statistic used for comparing the similarity and difference of sample sets. The Jaccard coefficient measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample set

$$JS = \frac{A \cap B}{A \cup B} \qquad (26)$$

In the equation (26) A represents the auto-segmented image B represents the ground truth image and J represents the Jaccard similarity value. If A and B are both empty, then it is defined as JS (A, B) = 1. Jaccard similarity always takes a value between 0 and 1 [10].

VII. RESULTS AND DISCUSSIONS

This section provides a description on the results obtained from the project with the help of tables and graphs. Table 1 tabulates the segmentation accuracy for various algorithms namely K-Means, RFCM, GRBFCM algorithms with SVM classifier which are evaluated using the parameters Dice Coefficient and Jaccard Similarity. The corresponding graph is plotted as shown in Fig.3.

I. SEGMENTATION ACCURACY

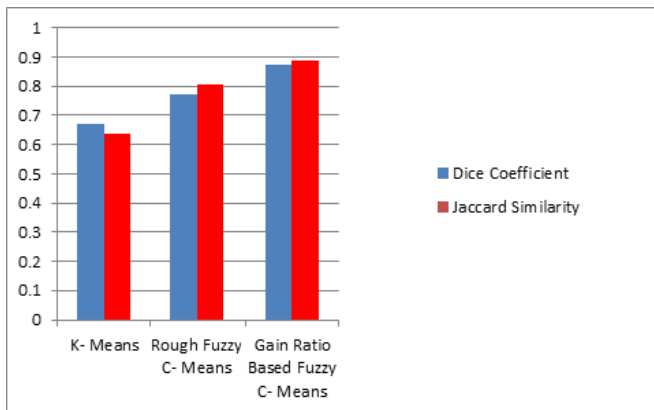| Name of the method	Dice Coefficient	Jaccard Similarity
K- Means	0.6710	0.6356
Rough Fuzzy C- Means	0.7722	0.8077
Gain Ratio Based Fuzzy C- Means	0.8735	0.8886

Fig. 3 Graph depicting segmentation accuracy

Table 2 tabulates performance of the classifier SVM against PNN classifier for the parameters accuracy, sensitivity and specificity. The corresponding graph is plotted as shown in Fig.4

II. CLASSIFIER PERFORMANCE

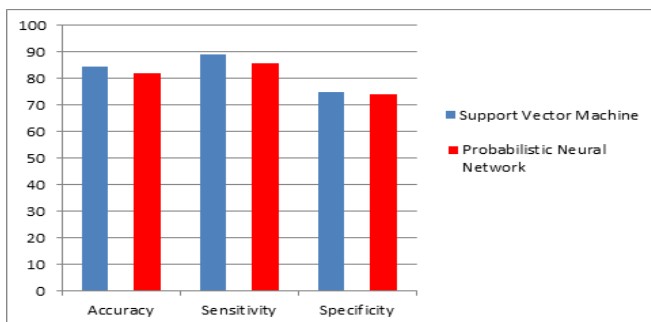| Classifier	Accuracy	Sensitivity	Specificity
Support Vector Machine	84.6	88.8	75
Probabilistic Neural Network	81.8	85.7	74

Fig. 4 Graph depicting classifier performance.

The proposed GRBFCM algorithm helps to improve the segmentation output results by introducing a new membership function which produces normalized output values. It also overcomes the favouritism problem of the previously proposed algorithms. The algorithm produces homogeneous clusters even in the presence of noise. The above evaluated results show that SVM classifier performed better than the PNN classifier the in terms of performance. This is due to the fact that the PNN classifier requires more training time when compared to the SVM classifier.

VIII. CONCLUSION

The paper proposes a new algorithm called the Gain ratio based Fuzzy C-Means algorithm. This work addresses the problem of clustering inhomogeneity and the problem of favouritism of Information Gain. In this GRBFCM algorithm, the image is pre-processed initially to remove noise from them. The texture features are extracted using Local Binary Pattern and is fed as input to the SVM classifier. Those images reported as abnormal by the classifier are fed as input to the GRBFCM algorithm. The algorithm works in two phases to alter a new membership function which is provided as input to the FCM algorithm. GRBFCM produces the segmented output from which the tumour region can be identified. The algorithm is evaluated which proves that this algorithm performs better than other traditional algorithms in terms of quality of segmentation.

REFERENCES

[1] Eman Abdel-Maksoud, Mohammed Elmogy, Rashid Al-Awadi, "Brain tumour segmentation based on a hybrid clustering technique," *Egyptian Infomatics Journal*, vol. 16, pp 71-81, 2015.

[2] Hakeem Aejaz Aslam, Mohammed Imtiaz Ali Ahsan, Tirumala Ramashri, "A New Approach to Image Segmentation for Brain Tumour detection using Pillar K-means Algorithm," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, pp. 1429-1436, 2013.

[3] Arun K Gupta, Kotikalapudi Raviteja, Maya D Bhat, Chandrajit Prasad, "Knowledge Based Brain Tumour Segmentation Graphical User Interface," *International Journal of Engineering and Advanced Technology*, vol. 3, pp. 361-366, 2013.

[4] Roshan G. Selkar, M. N. Thakare, "Brain Tumour Detection And Segmentation By Using Thresholding And Watershed Algorithm", *IJAICT*, vol. 1, pp. 321-324, 2014.

[5] Asifullah Khan, M. Aksam Iftikhar, Asmatullah Chaudhry, Mehdi Hassan, "Robust information gain based fuzzy c-means clustering and classification of carotid artery ultrasound images," *Elsevier transaction on Computer Methods and Programs in Biomedicine*, vol. 113, pp 593-609, 2014.

[6] Chih-Chin Lai, Chuan-Yu Chang, "A hierarchical evolutionary algorithm for automatic medical image segmentation", *Elsevier transaction on Expert Systems with Applications,* vol.36, pp. 248-259, 2009.

[7] Ed-Edily Mohd. Azhari, Muhd. Mudzakkir Mohd. Hatta, Zaw Zaw Htike, Shoon Lei Win, "Brain Tumour Detection and Localization in Magnetic Resonance Imaging," *International Journal of Information Technology Convergence and Services*, vol. 4, pp 1-11, 2014.

[8] Alexei N. Skourikhine, Bernd R. Schlei, Lakshman Prasad, "Neural Network for Image Segmentation", in *Proc. Conference for Applications and Science of Neural Networks, Fuzzy Systems and Evolutionary Computation*, vol. 4120, pp. 28-35, 2000.

[9] Pradipta Maji, Sankar K. Pal, "RFCM: A Hybrid Clustering Algorithm Using Rough and Fuzzy Sets," *Fundamental Informatics*, vol.80, pp. 475-496, 2007.

[10] Wikipedia page on Jaccard Index [Online]. Available: https://en.wikipedia.org/wiki/Jaccard_index/

[11] Wikipedia page on Dice_Coefficient [Online]. Available: https://en.wikipedia.org/wiki/Dice_coefficient

[12] V.P.Gladis Pushpa Rathi and S.Palani, "Brain Tumour MRI Image Classification With Feature Selection And Extraction Using Linear Discriminant Analysis," *International Journal of Information Sciences and Techniques*, vol. 2, pp. 131-146, 2012.

[13] Athira Krishnan, Sreekumar K, "A Survey on Image Segmentation and Feature Extraction Methods for Acute Myelogenous Leukemia Detection in Blood Microscopic Images", *International Journal of Computer Science and Information Technologies*, vol.5, 2014.

[14] Vladimir Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines," *Machine Learning*, vol. 46, pp. 389-482, 2002.

[15] A. Gebejes and R. Huertas, "Texture Characterization based on Grey-Level Co-occurrence Matrix," *Proc. Conference of Informatics and Management Sciences*, pp. 25 - 29, 2013.

[16] Jiawei Han, Micheline Kamber and Jian Pei, Data Mining Concepts and Techniques, 3rd ed, Morgan Kaufman, Ed. Waltham, USA: Elseiver, 2012.

[17] Mathworks page on Median Filter 2 [Online]. Available: http://in.mathworks.com/help/images/ref/medfilt2.html.