# Detection of tumor in brain MRI using fuzzy feature selection and support vector machine

**Conference Paper** · September 2016

DOI: 10.1109/ICACCI.2016.7732331



**2 authors:**

Amiya Halder
St. Thomas College of Engineering and Technology


Oyendrila Dobe
Michigan State University

# Detection of Tumor in Brain MRI Using Fuzzy Feature Selection and Support Vector Machine

Amiya Halder
CSE Department
St. Thomas' College of Engineering and Technology
Kidderpore, Kolkata-700023
West Bengal, India
Email: amiya.halder77@gmail.com

Oyendrila Dobe
CSE Department
St. Thomas' College of Engineering and Technology
Kidderpore, Kolkata-700023
West Bengal, India
Email: oyendrila.dobe@gmail.com

***Abstract*—This paper proposes a technique to categorize a brain MRI as normal, in the absence of a brain tumor or as abnormal in the presence of one. Proposed method is divided into two steps. First, a set of feature is generated for accurately differentiating between a normal and abnormal MR scan images. Then, these features are reduced using fuzzy c-means (FCM) algorithm. Further, a Support Vector Machine (SVM) is used to classify the scan images into two groups, namely, tumor-free and tumor affected. The proposed method aims to produce higher specificity and sensitivity than the previous methods.**

**Keywords:-**Image Segmentation, MR Images, SVM.

## I. Introduction

A brain tumor is an ensemble of abnormal and uncontrolled growing cells. They are basically of two types- primary tumors originate in the brain itself; whereas secondary ones are those that spread to the brain as result of metastasis. Primary tumors usually develop from brain cells, nerve cells, glands or the membranes that surround the brain (meninges). In adults, gliomas and meningiomas are the most common brain tumors. Primary tumors can be either benign or malignant. Lung, breast, kidney, skin cancers can metastasize to the brain. Secondary tumors are always malignant as benign tumors do not spread from one part of the body to another.

According to National Brain Tumor Society of US' statistics there are 688096 people are living with brain tumor in US alone. 550042 tumors are benign whereas 138054 tumors are malignant. 68470 people are estimated to receive diagnoses for primary brain tumor this year around 13770 people will die from brain cancer this year. Given the above statistics, it can be fairly proved that brain tumor is one of the biggest killers in present medical field. We know that most tumors can be identified with naked eyes by diagnostics. However, like in every manual case, a certain high chance of error is prevalent here. Hence it is obvious that an efficient automated brain tumor algorithm will be of helpful. This paper, henceforth suggests such an algorithm which has a higher specificity and sensitivity than previous algorithms.

In the field of medical image analysis, various research efforts have been assisting in diagnosis and clinical studies [1]. There are different imaging techniques available like magnetic resonance imaging (MRI), ultrasonography (USG), the computed tomography (CT) scan, positron emission tomography (PET) scan and several others. Amongst them, MRI is the best way to see inside a human body without opening them. Several algorithms have been proposed to detect the presence of brain tumor [2]-[17] using K-means clustering [2], an ANN approach [3], Statistical Pattern Recognition techniques [4], [5], Particle Swarm Optimization algorithm [6], Rule based approach [7], Neural Networks [9], unsupervised clustering [8], Bayesian classifier [11], Genetic Algorithm [12] and etc. All of the above techniques do not classify accurately or do not properly detect the tumor from MR images or size of the database used by them was extremely small for proper calculation of efficiency of the algorithm. Hence the need arises for a better segmentation and classification process with higher efficiency.

## II. Proposed Method

The method proposed here includes features extraction from brain MRIs [13], feature set dimensionality reduction using fuzzy c-means algorithm and finally, training a support vector mechine with radial basis function [14] to differentiate the MRI database into two classes namely, tumor present and tumor free. The components mentioned above are henceforth defined.

### *A. Data Description*

Datasets, comprising of both normal and abnormal images, have been created. Different databases have been used for the purpose of reducing feature set, training and testing SVM. The MR scan images have been used from the IXI-dataset [20], a research project by the Imperial College London. It has been collected from three different hospitals in London. MR images of weighted T1 and T2 types have been used to check the effectiveness of the proposed method.

### *B. Normlization*

The MRI images have been initially converted to gray scale with intensity values ranging from 0-255. Further a gray covariance matrix is generated.

### C. Feature sets

A lot of time is needed to compare between images, because a large amount of data or memory is used for represent an image. In order to reduce the amount of data, memory and time, we extract certain features from the image. Several feature selection and classification techniques available [18], [19]. The values of the extracted features, contain the information of an image that are relevant to us. These are further used as an input to the classifier for image classification and segmentation. We consider f(x, y) is a two-dimensional function of a gray scale image, $r(i)$ is the intensity level of an image, $G_n$ is the total number of gray levels in the entire image and $p(i)$ is the probability density. The following features have been extracted from the images to accurately differentiate between the two classes of images:

1) Image entropy: It is defined as the uncertainty of a random variable.

$$entr = \sum_{r=0}^{G_n-1} \sum_{s=0}^{G_n-1} (p(r,s) * log_2(p(r,s)+1)). \quad (1)$$

2) Image contrast: It measures the deviation of the current pixel from its neighboring pixels.

$$const = \sum_{r=0}^{G_n-1} \sum_{s=0}^{G_n-1} (r-s)^2 * p(r). \quad (2)$$

3) Image energy: It refers to the sum of squared intensities.

$$ener = \sum_{r=0}^{G_n-1} \sum_{s=0}^{G_n-1} p(r-s)^2. \quad (3)$$

4) Homogeneity: It measures the proximity of the distribution of elements in the GLCM to the diagonal elements in the GLCM.

$$entr = \sum_{r=0}^{G_n-1} \sum_{s=0}^{G_n-1} \frac{p(r,s)}{(1+(r-s)^2)}. \quad (4)$$

5) Pixel correlation: It measures the relation between the current pixel and its neighbor pixels.

$$cor = \frac{1}{a_w a_v} \sum_{r=0}^{G_n-1} \sum_{s=0}^{G_n-1} r * s * p(r,s)^2 - m_w m_v. \quad (5)$$

Where $a_w$,$a_v$ are standard deviations and $m_w$,$m_v$ are means of $p(r)$,$p(s)$.

6) Variance: It measures the gray tone variance.

$$var = \sum_{r=0}^{G_n-1} (r-m_w) * p(r). \quad (6)$$

where $m_w$ is mean of pixel value $p(r)$.

7) Standard deviation: Measures the deviation of values from the mean value.

$$std = \sqrt{\sum_{r=0}^{G_n-1} (r-m_w) * p(i)}. \quad (7)$$

8)

$$sumvariance(sv) = \sum_{r=0}^{2(G_n-1)} (r-se)^2 * p_{w+v}(r). \quad (8)$$

9)

$$sumaverage(u) = \sum_{r=0}^{2(G_n-1)} r * p_{w+v}(r). \quad (9)$$

10)

$$sumentropy(se) = \sum_{r=0}^{G_n-1} p_{w+v}(r) * log(p_{w+v}(r)). \quad (10)$$

11)

$$inertia(in) = \sum_{r=0}^{G_n-1} \sum_{s=0}^{G_n-1} (r-s)^2 * p(r,s). \quad (11)$$

12) Kurtosis:It measures the flatness of the histogram.

$$kurtosis(kurt) = a^{-4} \sum_{r=0}^{G_n-1} ((r-m_w)^4 * p(r)) - 3. \quad (12)$$

### D. Fuzzy c-means Feature Set Reduction

The above used features correctly define an image and distinguish between them. However there are redundancies which when removed can further simplify our algorithm. Fuzzy c-means algorithm attempts to fragment all given points into specified number of clusters. It is better than K-means clustering since it has an additional membership function for additional accuracy achievement. The converging of the membership function to 0 or 1 ensures creation of crisp partitions in the data. Hence a fuzzy c-means algorithm has been used to decrease the dimensionality of the feature set. For each of the above feature, a threshold value has been set by selecting a number of random normal brain MR image and calculating the feature value for that image. An average of all the values for each feature has been considered as the feature threshold value. Fig.1 shows the overview of this process.

### E. Data Classification Using Support Vector Machine

Support vector machine is a supervised learning method basically applied for classification of data. SVM is based on the concept of decision planes that define decision boundaries. A decision plane or hyperplane is one that separates between a set of objects having different class memberships. In the proposed method, a non-linear radial basis function has been used. Here the classification is done by casting the problem into a higher dimensional space, where separation becomes more likely. Any hyperplane can be depicted as,

$$\overrightarrow{\chi}.\overrightarrow{\delta} - k = 0 \quad (13)$$

Where $\overrightarrow{\chi}$ is the vector normal to the decision plane and is the offset, from the origin along the normal to the hyperplane. If the training data are linearly separable, two parallel hyperplanes can be selected that separate the two classes of data, so

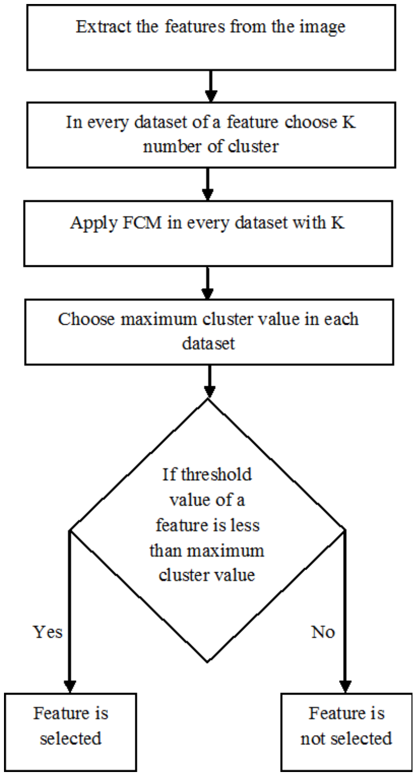


Fig. 1. Block diagram of the feature selection method.

that the distance between them is as large as possible. The region bounded by these two hyperplane is called the "margin", and the maximum-margin hyperplane is the hyperplane that lies halfway between them. These hyperplane can be given as,

$$\overrightarrow{\chi}.\overrightarrow{\delta} - k = 1 \tag{14}$$

and

$$\overrightarrow{\chi}.\overrightarrow{\delta} - k = -1 \tag{15}$$

However, here we have used a radial basis kernel in the SVM. Radial basis functions(RBF) are a set of functions which form a hypothetical basis for input patterns for expansion to the hidden space. The most widely used RBF is Gaussian functions given below,

$$O(r) = exp(\frac{-r^2}{2\sigma^2}) \tag{16}$$

for some $\sigma > 0$ and $r \epsilon R$. In SVM, the support vectors are basically training samples which the algorithm extracts to find the optimal plane. Margin of separation is the distance between closest data point and the separating hyperplane.

## III. Experimental Results

An approximate data set was made of about 100 MRI images in order to apply FCM for feature selection. From the set of 12 features, 9 were finally selected, eliminating the redundant features namely, correlation, sum average and kurtosis. A new dataset, of 100 images, was created of normal

TABLE I
Observation results.

| Session | Normal image | Identified correctly | Abnormal image | Identified correctly |
|---|---|---|---|---|
| Training | 65 | 65 | 41 | 41 |
| Testing | 54 | 54 | 41 | 39 |

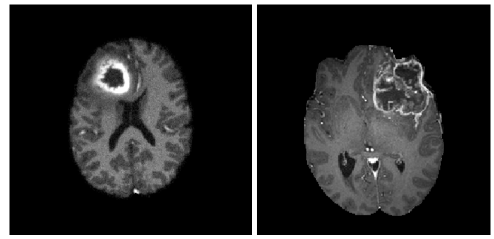

Fig. 2. Tumor affected MR images.

and abnormal (tumor present) images for training the support vector machine. Finally with a dataset of 95 images a test was run which produced the given results. Different sized of images ( $512 \times 512, 640 \times 640, 1500 \times 1845, 586 \times 586, 1000 \times 1100$) has been used to determine universality of the algorithm. All program is implemented using MATLAB. Efficiency or accuracy of the classifiers for each analysis method is analyzed by error rate. This error rate is defined by the terms normal and abnormal right and normal and abnormal wrong as follows:

1) Abnormal Right (AR): The test gives positive result if tumor is present.
2) Normal Right (NR): The test gives negative result if tumor is absent.
3) Normal Wrong (NW): The test gives positive but tumor is absent.
4) Abnormal Wrong (AW): The test gives negative but tumor is present.

Sensitivity, specificity and accuracy are used to describe the clinical efficiency of the classification and segmentation algorithm.

$$Sensitivity = \frac{AR}{AR + AW} * 100 \tag{17}$$

$$Specificity = \frac{NR}{NR + AR} * 100 \tag{18}$$

$$Accuracy = \frac{AR + NR}{AR + NR + NW + AW} * 100 \tag{19}$$

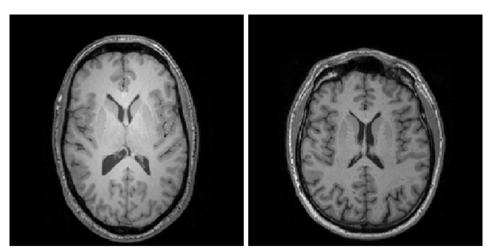

Fig. 3. Tumor free brain MR images.

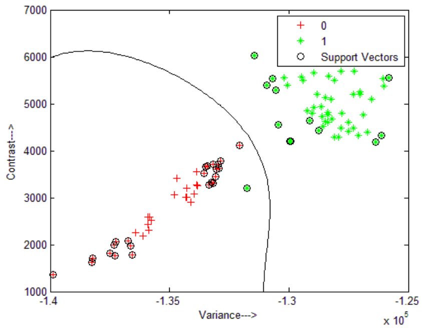


Fig. 4. Plot to show training dataset in SVM being classified into two different classes, separated by a hyperplane generated by a RBF kernel.

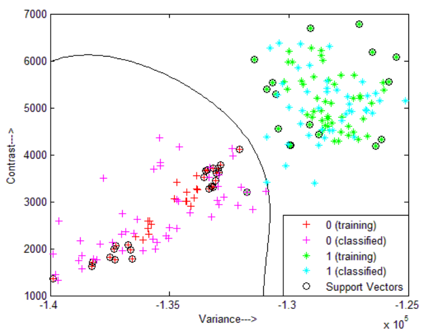


Fig. 5. Plot to show both training and classified dataset in SVM being partitioned into two different classes, separated by a hyperplane generated by a RBF kernel.

The result of the proposed method (shown in TABLE I) has been compared with detection process using Fuzzy c-means, K-means [2], Bayesian classifier [11], Genetic Algorithm (GA) [17], SVM (RBF kernel)[14], [15], [16] and Neural Network [9]. The proposed algorithm shows significant increase in efficiency with respect to the above algorithms. Table II shows the accuracy measure of the above mentioned methods. Fig.2 and Fig.3 show the sample images of abnormal and normal MR scan images, respectively. Fig.4 shows the SVM plot for training data only and Fig.5 shows the SVM plot for both training as well as classified data.

## IV. Conclusion

In this paper, we have presented an efficient detection algorithm to detect tumor in MRIs using FCM based Support Vector Machine. This proposed algorithm is established to give encouraging results than the other existing brain tumor detection algorithm based on segmentation or classification algorithms.

TABLE II
ACCURACY OF THE DIFFERENT METHOD.

| Different method | Accuracy |
|---|---|
| FCM clustering | 85% |
| K-means clustering | 87% |
| Classification approach (Bayes classifier) | 89% |
| GA Approach | 93% |
| SVM method | 92.71% |
| Neural Network approach | 96.33% |
| Proposed Method | 97.89% |

## V. Acknowledgment

We would like to thank Principal and Director of St. Thomas College of Engineering and Technology (STCET), Kolkata for their cooperation and support.

## References

[1] J. S. Duncan, N. Ayache, Medical Image Analysis-Progress Over two decade and challenges ahead, *IEEE Trans on PAMI* , vol 22, pp. 85-106, 2000.
[2] M. Prastawa, E. Bullitt, S. Ho and G. Gerig, Robust Estimation for Brain Tumor Segmentation, *Medical Image Computing and Computer-Assisted Intervention* pp. 530-537, (2003).
[3] A. Zamre, T. Shah, H. Thadhani, D. Bangar, Detection and classification of brain tumors in MRI images, *International Journal of Science, Engineering and Technology*, vol 2, pp. 1-7, 2014.
[4] K. Held, E.R. Kops, B.J. Krause, W.M. Wells 3rd, R. Kikinis, H.W. Muller-Gartner, Markov random field segmentation of brain MR images, *IEEE Transaction Medical Imaging*, vol 16, pp 878-886, 1997.
[5] Y. Zhang, M. Brady, S. Smith, Segmentation of Brain MR Images through hidden Markov random field model and the expectation maximization algorithm, *IEEE Transactions on Medical Imaging*, vol 20, pp. 45-57, 2001.
[6] N. Forghani, M. Forouzanfar and E. Forouzanfar, MRI fuzzy segmentation of brain tissue using IFCM algorithm with Particle Swarm Optimization, *22nd international symposium on Computer and information sciences*, pp 1-4, 2007.
[7] M. Milan, L. Sven, P. Damir, A rule based approach to stroke lesion analysis from CT brain Images, *2nd International Symposium on Image and Signal Processing and Analysis*, pp. 219-223, 2001.
[8] T. Hau Lee, M. Faizal, A. Fauzi and R. Komiya, Segmentation of CT Brain Images using unsupervised clusterings, *Journal of Visualization*, vol 12, pp. 31-38, 2009.
[9] W. H. Ibrahim, A. Osman and Y. I. Mohamed, MRI brain image classification using Neural Networks, *International Conference on Computing, Electrical and Electronics Engineering*, pp. 253-258, 2013.
[10] M. C. Clark, L. O. Hall, D. B. Goldgof, R. Velthuzien, F. R. Murtagh and M.S. Silbiger, Automatic tumor segmentation using knowledge based techniques, *IEEE Transactions on Medical Imaging*, vol 17, pp. 187-192, 1998.
[11] A. Farzan, A. Rahman Ramli, S. Mashohor and R. Mahmud, Fuzzy modeling of brain tissues in Bayesian segmentation of brain MR images, *2010 IEEE EMBS Conference o Biomedical Engineering and Sciences*, pp. 77-80, 2010.
[12] R. Ganesan, R. Radhakrishnan, Segmentation of Computed Tomography Brain Images using genetic algorithm, *International Journal of Soft computing*, vol 4, pp. 157-161, 2009.
[13] D. Selvaraj, R. Dhanasekaran, A Review on Tissue Segmentation and Feature Extraction of MRI Brain images, *International Journal of Computer Science and Engineering Technology*, vol 4, pp. 1313-1332, 2013.
[14] M. Parviainen, Radial Basis Function (RBF) and Support Vector Machines (SVM) networks, 2012.
[15] D. C. Shubhangi, P. S. Hiremath, Support vector machine (SVM) classifier for brain tumor detection, *International Conference on Advances in Computing, Communication and Control*, pp. 444-448, 2009.
[16] S. Rajeswari, K. T. Jeyaselvi, Support Vector Machine Classification For MRI Images, *International Journal of Electronics and Computer Science Engineering*, pp. 1534-1539, 2012.
[17] A. Halder, A. Pradhan, S. K. Dutta and P. Bhattacharya, Extraction from MRI images using Dynamic Genetic Algorithm based Image Segmentation and Morphological Operation, *International Conference on Communication and Signal Processing*, pp. 1541-1545, 2016.
[18] J. Singha and R. H. Laskar, Hand gesture recognition using two-level speed normalization, feature selection and classifier fusion, *Multimedia Systems* , pp. 1-16, 2016.

[19] V. B. Semwal, K. Mondal and G. C. Nandi, Robust and accurate feature selection for humanoid push recovery and classification: deep learning approach, *Neural Computing and Applications*, pp. 1-10.
[20] http://brain-development.org/ixi-dataset/