Brain Tumor Classification using Adaptive Neuro-Fuzzy Inference System from MRI

Sudipta Roy[1], Shayak Sadhu[1], Samir Kumar Bandyopadhyay[2],
Debnath Bhattacharyya[3] and Tai-Hoon Kim[4]

[1]Department of Computer Science and Engineering,
Academy of Technology, Adisaptagram, India
[2]Department of Computer Science and Engineering,
University of Calcutta, Kolkata, India
[3]Department of Computer Science and Engineering,
Vignan's Institute of Information Technology,
Visakhapatnam-530049, India
[4]Department of Convergence Security,
Sungshin Women's University,
249-1, Dongseon-dong 3-ga, Seoul, 136-742, Korea
[1]sudiptaroy01@yahoo.com, shayakchemistry@gmail.com, [2]skb1@vsnl.com,
[3]debnathb@gmail.com, [4]taihoonn@daum.net

Abstract

Detecting correct type of brain tumor is a crucial task for diagnosis and curing the tumor. Identifying the correct type of brain tumor can provide a fast and effective way to plan the diagnosis of tumor. The proposed system provides a fast and efficient way to identify the correct type of tumor and classify it to the respective class label. Our proposed system is comprised of multiple stages. In the first stage MRI image is taken as input and is normalized. The second stage includes extraction of feature vectors from the image which results in reducing redundancy of data and will serve as the input to the classifier. The classifier takes each tuple of feature extracted vector to produce classified output. Performance analysis shows that our proposed methodology has performed very efficiently and accurately. In our work we demonstrate the application of Fuzzy Inference System (FIS) based classifier known as Adaptive Neuro Fuzzy Inference System (ANFIS) to successfully classify the input tuples in comparison to other two selected classifiers namely: Artificial Neural Network with Backpropagation Learning Model and K-Nearest Neighbors.

Keywords: MRI, Feature Extraction, Classification, Brain tumor, Fuzzy Inference System, Adaptive Classifier, texture analysis

1. Introduction

Brain tumor occurs when there is abnormal growth of some brain cells. Tumors originate in brain are called primary brain tumors and others that spread to the brain from other parts of the body are termed as secondary brain tumors. Primary brain tumors may be malignant or benign while secondary tumors are always malignant. Both the types of tumor pose a threat to life and can cause permanent disability even after treatment. We know from the anatomy of brain structure that the space inside the skull is limited and is filled up by the brain cells and cerebrospinal fluid. If there is an extension of brain tissues, the pressure inside the brain will increase. This consequently causes many health problems such as edema. Brain tumors forms the second most cause for cancer related deaths in children and adults. The Central Brain Tumor Registry of the United States

(CBTRUS) has reported that more than 70,000 primary brain tumors were diagnosed with primary brain tumors among which 4,600 were children between the ages 0-19.

A tumor can be graded in to several stages in accordance to the analysis of abnormality of the tumor cells and tissues. This grading gives us the acute probability of tumor growth in size and its spreading. Tumor grade can be determined using biopsy. It is to be noted that grading of tumor is not same as cancer stages. Very accurate diagnosis of brain tumors was very difficult due to high complexity and large variance having more than 120 types. Treatment of brain tumor in early stages is a challenging task owing to the variation in size shape and location and can only be performed by trained professional neuro radiologist. In the past several research works have been done for improving the detection, diagnosis and treatment of tumors in early stages. The extensive use of computer technology in medical decision making has been used in various medical fields. We need to study the features of the tumor and make decisions accordingly from medical imaging tests of the test subject. In this work we use MR imaging as it is non-invasive and produces no known biological hazards. MRI scans is capable of producing images with high knowledge content. There is several variation of MRI such as: T1 weighted, T2 weighted, FLAIR, *etc*. In our work we have used T2 weighted MRI as it displays damaged tissues very vibrantly. This imaging technique can correctly display the size, variation and location along an axis coordinate. Researchers have used this imaging for various research purposes.

Classification can be described as a process in which ideas and objects are analyzed and grouped into various class labels depending on the trend of the input data. Using this model we can group the unlabelled input data into grouped class labels. This grouping helps us in correct decision making consecutively improving the accuracy of classification. This method is useful in predicting a class label for any arbitrary tuple from the knowledge database. Classification is another important part used in our work. Classification involves two important steps: training phase and testing phase. The training phase involves the building of the classification data. Mathematical model (such as neural network or decision tree) is trained such that each set of inputs correctly gives us the resultant outputs. This training is needed to be done accurately in order to produce the most appropriate result during the testing phase. This training can be supervised or unsupervised. In case of supervised learning we train the model knowing the output class label for a particular tuple. In case of unsupervised training the output class label is not known. In testing phase we take the tuples with unknown class labels as input to get an appropriate class label as output.

There are many mathematical models that have been used for classification. Classifiers like decision tree (ID3, C4.5, CART or RIPPER), Bayes classifier, Neural Network, Support Vector Machine (SVM) and RBFN (Radial Basis Function Network) are some of the popular and most commonly known classifiers that have been used to classify various types of data. In the classification step, we need to build up a database with sample tuples corresponding to class labels. In this work we try to concentrate on supervised learning model. Supervised classifiers are capable of exerting maximum accuracy on the learning step but fails when we input a slice with a new type of tumor. This is one of the greatest disadvantages in this type of learning model. In our work we consider the following five types of brain tumor: 1) Glioma, 2) Meningioma, 3) Metastatic adenocarcinoma, 4) Metastatic bronchogenic carcinoma, 5) Sarcoma. We have concentrated on those five types of brain tumors as in this work we aim towards solving this problem of classification of brain tumor and its performance measurement. As there are more than 120 types of tumors that were reported but we have selected the five common types of brain tumors for the study. We have constructed a database which contains images associated with its corresponding class label. Images from this database were used in testing and training the classifiers.

This research study is arranged as follows: Section 2 includes the related works done in this area. Section 3 gives the description of the dataset being used while Section 4 explains the detailed procedure being used. Section 5 discusses the performance analysis and results and Section 6 is reserved for the conclusion of our work.

2. Review Work

Many researchers have attempted to correctly classify brain tumors into their appropriate types. An attempt to classify tumors using Gabor wavelet analysis had been made by Yi-hui Liu *et al.*, [1]. In their work they have used texture based analysis based on Gabor wavelets to improve the accuracy of classification. They have used Gabor filters in the feature extraction steps and support vector machine based classifier to classify the tumor. Wavelet analysis has also been used by El-Sayed *et al.*, [2] in which they have proposed a hybrid technique. In this technique they have used discrete wavelet transform on the MRI slices to extract the features and then minimized using principal component analysis. In their work thy have used two types of classifiers: feed forward back propagation neural network and K-nearest neighbors. Using these models they have achieved a maximum accuracy of 98.6% [2]. Both of the aforesaid works focuses on the features to be extracted from the input image. Feature selection and extraction have been demonstrated by S. Palani [3]. In this work they have made use of linear discriminant analysis and principal component analysis to select the most striking features (feature vector) among all the available features. They are used to reduce the number of features used and increase the accuracy of classification. Some hybrid models such as support vector machine recursive feature elimination (SVM-RFE) have been used by Evangelia I. Zacharaki *et al.*, [4]. They have used ranking based criterion which tests the discriminative power of each distinct feature. SVM-RFE has been used to find a subset of features that produces optimal performance for the classifier. This algorithm takes the help of ranking of features followed by backward sequential selection method which helps in removal of one feature at a time. Segmentation of region of interest (ROI) before feature extraction has been done in some of the works. Qiang Wang et al [5] demonstrate the process of segmentation of ROI and feature extraction from them. They have used the idea of fuzzy connectedness to find the region of interest from the given slice and then extract features to classify them using any standard classifier. EI papageevgious [6] *et. al.*, in their work proposed a fuzzy cognitive map (FCM) to find the grade value of tumor. They used the soft computing method of fuzzy cognitive maps to represent and model expert's knowledge FCM grading model achieved a diagnostic output accuracy of 90.26% & 93.22 % of brain tumors of low grade and high grade respectively. They proposed the technique only for characterization and accurate determination of grade. Noor Elaiza[7] in their work proposed an implementation of evaluation method known as image moseying in evaluating the MRI brain abnormalities segmentation study. 57 mosaic images are formed by cutting various shapes and size of abnormalities and pasting it onto normal brain tissue. Some methods like ANFIS, FCM are used to segment the mosaic images formed. Statistical analysis method of receiver operating characteristic (ROC) was used to calculate the accuracy. S.Gowri [8], in their work proposed detection of tumor growth by advanced diameter technique using MRI data. To find the volume of brain tumor they proposed diameter and graph based methods. The result shows tumor growth and volume. Lawrence *et. al.*, [9] proposed a system that automatically segments and labels tumor in MRI of the human brain. They proposed a system which integrates knowledge based techniques with multispectral analysis. The results of the system generally correspond well to ground truth, both on a per state basis and more importantly in tracking total volume during treatment over time. Khan IbleKharuddin[10], in their work suggested an enhanced implementation of artificial neural network algorithm to perform segmentation

of brain MRI data learning vector quantization and is used for segmentation. Their result suggests excellent brain tissue segmentation.

In this paper a new and improved method is implemented by FIS based adaptive network known as ANFIS. Compared to the previous work suggested in the literature discussed above, high accuracy is achieved for feature selection and extraction. The major drawback of classification has been drawn from above literature survey is that if the techniques are accurate; the time requirement is high and vice-versa due to two key reasons: (1) there is a large number of tumor types which differ greatly in size, shape, location, tissue composition and tissue homogeneity. In some cases, their border with normal tissues cannot be very well defined on images; therefore, they are even difficult for radiology experts to delineate. (2) The consequence of the phenomenon of partial volume effect, where one voxel may belong to multiple tissue types, in addition to noise of MRI automatic tumor type identification and analysis are still a difficult problem.

3. Proposed Methodology

The proposed methodology is composed of multiple stages as illustrated in Figure 1. Initially we have chosen tumor detection [11] methodology from MRI slices of brain. Then they are normalized to an acceptable range before being fed to feature extraction process. This process contains two steps: artefact removal [12] and noise reduction. Artefact removal is done to remove segments in images that in not needed such as information like date printed on the slice. Noise reduction is accomplished by using filters [12] in order to remove unwanted distortion in the image such as Gaussian noise and salt and pepper noise on the image. Next this normalized image is now fed to the next step to reduce redundancy of information using feature extraction. In the classification step, the model is first trained using training dataset obtained from the image database which also defines the class labels being used. After the classification model is trained, it is used to classify the testing dataset into appropriate classes that will help us in correct medical decision making and diagnosis of brain tumor. After getting the predicted output we compare them with practical values to get the performance measurement of the model being used. The detailed implementation of the proposed methodology can be given by the following subsections.

3.1. Input MRI Dataset

As discussed in section 1 we have used T2 weighted MRI slices that are helpful in identifying bulges and distinction among the brain tissues. This variety of MR imaging as input dataset to our proposed methodology. The images have been collected from Med Harvard [14]. In our work we have selected the five most commonly occurring type of brain tumor as aforementioned in Section 1. The experiments have been conducted on a dataset selected at random and its result is used in performance measure in later steps.
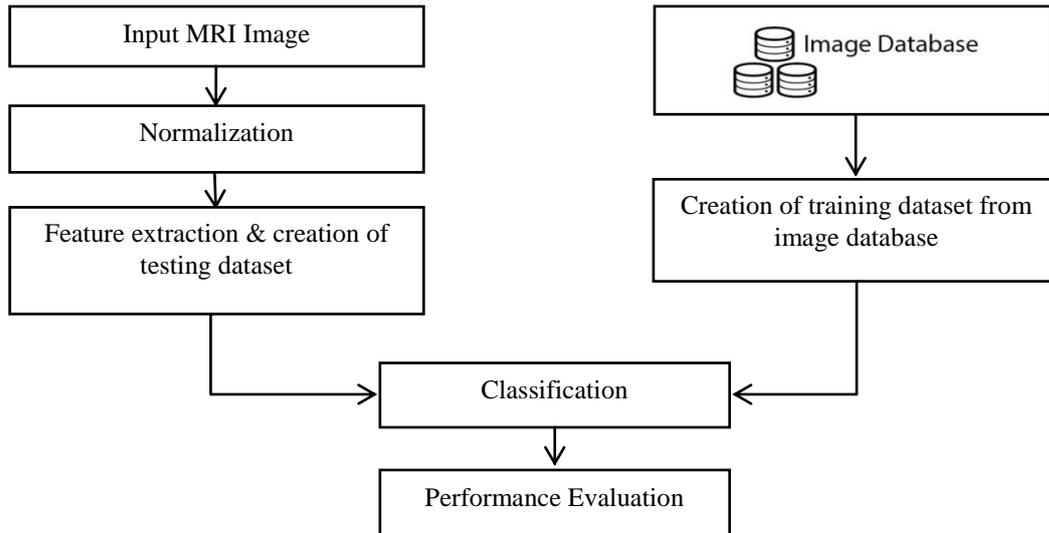
Figure 1. Total Workflow of the Proposed Methodology

3.2. Feature Extraction

In image processing, feature extraction can be defined as a type of dimensionality reduction which can effectively show interesting patterns in the image in form of a vector (feature vector). Useful features are extracted from the image and are used in the next step of classification of brain tumor. In our proposed methodology we intend to use intensity based and texture based features [15]. In our work we first order histogram based features is used for feature extraction. Histogram of an image represents the concise statistical information contained in the image. Now let us consider that $f(x, y)$ is the function that represents the intensity level for each pixel (x, y) in the image, where $x = 1, 2, \cdots, A$ and $y = 1, 2, \cdots, B$. Grey-level histogram calculation involves each individual pixel. Probability density for each occurring pixel intensity level $0, 1, \cdots, N - 1$ is calculated dividing them with $h(i)$ by total number of pixels. This can be represented as:

$$p(i) = \frac{h(i)}{AB} \quad for \ i = 0, 1, \cdots, N - 1 \quad (1)$$

$$h(i) = \sum_{x=1}^{A} \sum_{y=1}^{B} \delta(f(x, y), i) \quad (2)$$

, where $h(i)$ is the intensity level histogram function for the whole image and for each intensity level i. Here we take $\delta(i, j)$ as the Kronecker delta function that can be given as:

$$\delta(i, j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (3)$$

This histogram can be easily calculated and its shape gives us useful information about the characteristics of the image. Various useful parameters can be derived from the derived histogram to describe the statistical information about the image. These parameters are taken to be parameters for feature extraction. They are described below:

- *Mean* is the average value of intensity of the image and can be given as:

$$\mu = \sum_{i=0}^{N-1} i \cdot p(i) \qquad (4)$$

- *Variance* is the intensity variation around mean and can be represented by:

$$\sigma^2 = \sum_{i=0}^{N-1} (i - \mu)^2 \, p(i) \qquad (5)$$

- *Skewness* gives us the measure of the amount of symmetry of the histogram around mean. It can be given by:

$$\mu_3 = \sigma^{-3} \sum_{i=0}^{N-1} (i - \mu)^3 \, p(i) \qquad (6)$$

- *Kurtosis* measures the flatness in the histogram and is given by:

$$\mu_4 = \sigma^{-4} \sum_{i=0}^{N-1} (i - \mu)^4 \, p(i) - 3 \qquad (7)$$

- *Entropy* represents the uniformity of the histogram and is given by:

$$H = \sum_{i=0}^{N-1} p(i) \log_2 [p(i)] \qquad (8)$$

- *Energy* represents the mean of squared value of the pixel intensity and can be given as:

$$E = \sum_{i=0}^{N-1} [p(i)]^2 \qquad (9)$$

All the mentioned parameters give us the information extracted from the local image histograms and can be used for texture segmentation. The normalization as given in previous step results in better texture discrimination accuracy. One of the major advantages of using these parameters is that they are simple, but they are not able to completely characterize texture. To solve this we use the definition of joint probability distributions of pixel pairs. Using this definition we construct second order histogram known as grey-level co-occurrence matrix $h_{d\theta}(i, j)$. We divide this matrix by the total number of neighboring pixels $R(d, \theta)$ in the image, the resulting image becomes the joint probability $p_{d\theta}(i, j)$ for two pixels with distance d between them and along the direction i and j. In this case we consider the value for $d = 1, 2$ and $\theta = 0°, 45°, 90°, 135°$ are normally used. For a given image with intensity function $f(x, y)$ and N discrete intensity values, we can construct the matrix $h_{d\theta}(i, j)$ and defining the parameters i and j as:

$$f(x_1, y_1) = i \qquad \text{and} \qquad f(x_2, y_2) = j \qquad (10)$$

where
$$(x_2, y_2) = (x_1, y_1) + (d \cos \theta, d \sin \theta) \qquad (11)$$

This results in a matrix dimension equal to the number of intensity levels and for each distance d and orientation θ. Thus the co-occurrence matrix contains N^2 elements that can be considered as a reduced set of features. Some of the parameters that can be derived from the matrix can be given as follows, where we consider μ_x, μ_y and σ_x, σ_y as the mean and standard deviation derived from this matrix. The parameters are given as:

- Angular second moment (energy):

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} [\, p\,(i,\, j)]^{\,2} \tag{12}$$

- Correlation:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{ijp\,(i,\, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{13}$$

- Inertia (Contrast):

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-\,j)^{\,2}\, p\,(i,\, j) \tag{14}$$

- Absolute value:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left|i-\,j\right| p\,(i,\, j) \tag{15}$$

- Inverse Difference:

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{p\,(i,\, j)}{1+(i-\,j)^{\,2}} \tag{16}$$

- Entropy:

$$H = -\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p\,(i,\, j) \log_2 [\, p\,(i,\, j)] \tag{17}$$

Using all aforesaid parameters we have created a tuple (feature vector) which can give us enough number of features that can help in classification of the input image into the predetermined class label. Using the aforesaid intensity and texture based features we can describe the features and use them to reduce the redundancy of information in the images. Thus we can consider the above described parameters as attributes that can successfully define the information or pattern content in the image.

3.3. Classification

As given in figure 1 the classification step occurs in two consecutive steps: learning phase and testing phase. Learning phase also known as training phase, is the first step in classification. In this step we intend to build a model that can successfully classify a dataset. So, we need to teach the model the class labels and data variations in this training phase. We use the pre-defined dataset from the database (as illustrated in figure 1) to train the model. After this training has been done we use the testing dataset generated from the input image from feature extraction to predict its class label.

In our work ANFIS is treated as the primary classifier and compare its performance with Artificial Neural Network (ANN) with back propagation learning and K-Nearest Neighbors (K-NN) models [16-17]. The ANFIS uses fuzzy rules and fuzzy reasoning that is based on fuzzy set theory. It can also be said that fuzzy stands out to be a form of multivalued logic (that is many logical levels between 0 and 1). Using the fuzzy rules and reasoning we now try to build the fuzzy inference system. Adaptive network is a network structure that consists of a number of nodes that are connected through directed weighted links. In this subsection we discuss the implementation details of ANIS to predict the output class label for any input tuple.

Fuzzy inference system (FIS) is based upon the concepts of fuzzy set theory, fuzzy if-then rules and fuzzy reasoning. In fuzzy set theory we consider a set where there are many logical levels between two extremities of binary logic. In case of fuzzy if-then rules (also known as fuzzy rules) which take the following form:

$$if \quad x \ is \ A \ then \quad y \ is \ B \qquad\qquad (18)$$

This is a general form of fuzzy rules where A and B are linguistic values which is defined by fuzzy sets on the universe of discourse of X and Y respectively. In this case we may say that the statement " $x \ is \ A$ " is called the antecedent or premise and the statement " $y \ is \ B$ " is called the consequence or conclusion. This rule can also be abbreviated as $A \rightarrow B$. In this context this relation defines a relation between two variables x and y in which the fuzzy rule defined as a binary relation R on the product space $X \times Y$. Fuzzy set theory is used to describe these expressions on the product space. These fuzzy rules are first defined to be used in fuzzy reasoning. Fuzzy reasoning (also called approximate reasoning) tends to make a conclusion from the set of fuzzy rules and known facts. The basic structure of FIS is composed of three basic components: rule base, database (or dictionary) and a reasoning mechanism. The rule base contains a selection of valid fuzzy rules. The database contains the definition of membership functions that are used in fuzzy rules. The reasoning mechanism is responsible to derive a reasonable output or conclusion based on the rules and given facts. The basic fuzzy inference system is capable of taking either fuzzy input or crisp input (viewed as fuzzy singletons), but it always produces fuzzy sets as outputs. But in our case we need crisp output and therefore de-fuzzification is done. In this process we extract the crisp value that can best represent the fuzzy set. A typical Sugeno fuzzy [18] model can assume the following form of fuzzy rules:

$$if \quad x \ is \ A \ and \quad y \ is \ B \ then \quad z = f(x,y) \qquad\qquad (19)$$

, where A and B are fuzzy rules in the antecedent and $z = f(x,y)$ is a crisp function in the conclusion and is dependent on the input variables x and y or any function that can correctly describe the output. If $f(x,y)$ represents a first-order polynomial then the resulting FIS is called first order Sugeno fuzzy model. Similarly, if f is constant then the FIS results in zero-order Sugeno fuzzy model. The zero-order Sugeno fuzzy model gives a smooth function of its inputs variables as the output as long as the membership functions overlaps. In our adaptive network, each node in the network represents a processing element and is interconnected by weighted links. The input output behaviour in this network is determined by a collection of modifiable parameters. In this context of adaptive network we consider a feedforward network i.e. the inputs fare fed in the first node and the output is obtained in the last layer of the network. For simplicity, we consider a FIS with two input x and y, and one output z. Now let us consider the following two fuzzy rules for this model.

$$Rule \quad 1 : If \quad x \ is \ A_1 \ and \quad y \ is \ B_1, \ then \quad z = f_1(x,y) \qquad\qquad (20)$$

$$Rule \quad 2 : If \quad x \ is \ A_2 \ and \quad y \ is \ B_2, \ then \quad z = f_2(x,y) \qquad\qquad (21)$$

Using these two rules we now build an adaptive network that can properly reflect these rules when a mapping is done from input to output product space.
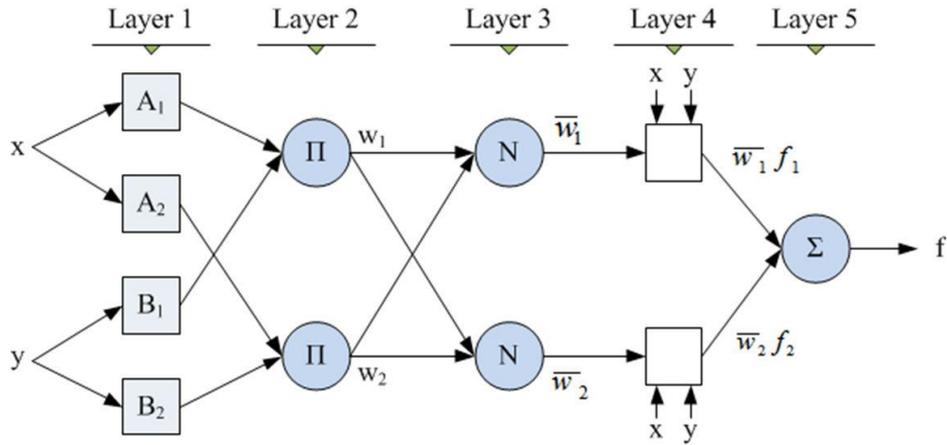
Figure 2. ANFIS Architecture Equivalent to Fuzzy Inference System

The adaptive network equivalent to this fuzzy model illustrated in figure 2 where we consider that each node in a particular layer performs the same function. Each ith node in the a particular layer l takes an input from the previous layer and produces an output $O_{l,i}$. This mapping for each individual layer described as follows:

• For layer 1 we take $l = 1$ and output as $O_{1,i}$ for each ith node in this layer. Every node in this layer is considered to be an adaptive node where the node function can be described as:

$$O_{1,i} = \begin{cases} \mu_{A_i}(x) & \text{for } i = 1,2 \\ \mu_{B_{i-2}}(y) & \text{for } i = 3,4 \end{cases} \tag{22}$$

where x or y is the input variable to node i and each node is assigned a linguistic label A_i or B_{i-2} to it. Here $\mu_A(x)$ denotes the membership function for A and can be any valid parameterized function that depends on a parameter set. This parameter set can also be called as premise parameters.

• In the next layer (Layer 2) we consider a fixed node label denoted by Π for which the output for the incoming signal can be defined by:

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y) \qquad \text{for } i = 1,2 \tag{23}$$

This output represents the firing strength of a rule.

• In layer 3 we denote a fixed node label for every node as N. In this layer we calculate the normalized firing strengths for every i[th] node and can be calculated as the ratio between the firing strength of that layer to the sum of all firing rules in that layer as:

$$O_{3,i} = \overline{w}_i = \frac{w_i}{\sum\limits_{i=1}^{2} w_i} \tag{24}$$

• In layer 4 we compute the output considering the parameter set embedded in the membership function. The parameters in this set are referred as consequent parameters. For every i[th] node the node function can be given as:

$$O_{4,i} = \overline{w}_i f_i(x,y) \tag{25}$$

• In the last layer (layer 5) each of the node is assigned with a fixed node label Σ. Each node computes the summation of all incoming signals as outputs as:

$$O_{5,i} = \sum_i \overline{w}_i f_i(x, y) = \frac{\sum_i w_i f_i(x, y)}{\sum_i w_i} \tag{26}$$

The learning algorithm uses a combination of the least-squares and back-propagation gradient descent methods to train the implemented model from the training dataset. It also checks the training procedure that over fitting of the training dataset does not occur. In this model it has been observes that if a particular tuple does not match with the established fuzzy rules defined in the model, the classifier tries to give a result that is outside the solution set. This is helpful in determining a new type of tumor. As fuzzy results obtained from the classifier cannot directly define the type of class label but it represents a numerical value that is within the range of solution set. Thus to convert this value into a crisp value we try to use defuzzification. There are many defuzzification techniques but in our case we use membership functions or fizzy rules to get a crisp output class label.

4. Results and Discussion

After building the models as specified in proposed methodology we take an arbitrary slice as input to test the effectiveness of classifying the slice to a class label as specified by the training dataset. The proposed methodology have been implemented and tested in MATLAB software (version R2013a). The training dataset is generated from image database [14]. In our case we consider only five class labels of brain tumors and the model is appropriately trained using the training database. After training it is tested against the testing dataset generated from the input slice. The predicted class is compared to the actual class label to measure the efficiency of the classifier. The testing of this methodology was carried on a personal computer with an AMD-A10-5750M processor with 2.50 GHz clock speed and 8 GB RAM.

The number of inputs can be varied according to the user and for our case we have used total 320 sample input slices (20 input slices for one type as been shown in figure 3). Each of the individual passed through the normalization and feature extraction processes. In this processes the images are normalized to the desired quality and the feature vector containing 13 elements is extracted from the slice. Each of the feature vector forms an input tuple to the classifier. This vector is generated for each of the 20 input slices shown in figure 3 has been illustrated by table 1.

Each of the generated feature vector from the normalized grayscale image are calculated from 1st order histogram based features and features from Gray Level Co-occurrence Matrix (GLCM). These values are based on the variation intensity values among the pixels in the selected slide. Each of the functions derives a single valued element in the feature extraction matrix for each of the 20 slices that is represented in table 1.

Each tuple contains its class label which denotes its tumor type which is used to build IF-THEN rules to generate the FIS. An adaptive network is then built depending on the training tuples and the FIS. ANFIS is this trained adaptive network that is now to be used to generate the class label for the input slices. For the testing phase we enter each of the tuples that can then give us a predicted class output. The output received in this case is fuzzy output. As class labels are crisp values we need to defuzzify the output. This defuzzified output is then compared to the actual class label to get the performance measurement of this classifier.

In the testing phase we have used ANFIS as the primary classifier and the other two as secondary ones to do a performance measurement among the classifiers. In the event of classification we have selected five class labels for each type of tumor. The class labels mentioned here were taken accordingly as: Class 1 = Glioma, Class 2 = Meningioma,

Class 3 = Metastatic adenocarcinoma, Class 4 = Metastatic bronchogenic carcinoma, Class 5 = Sarcoma. As we cannot use the actual names while using classification, we actually use the class labels which are numbers is the actual output of the classifier. The classifier gives fuzzy floating point numbers as output which is defuzzified to get the crisp actual class labeled values.

Figure 3. 20 input slices passed through the normalization and feature extraction processes. After classification we found that slices I1-I4 are Type 1 (Glioma); slicesI5-I8 are Type 2 (Meningioma); slices I9-I12 are Type 3 (Metastatic adenocarcinoma); slices I13-I16 are Type 4 (Metastatic bronchogenic carcinoma); slicesI17-I20 are Type 5 (Sarcoma)

Each tuple contains its class label which denotes its tumor type which is used to build IF-THEN rules to generate the FIS. An adaptive network is then built depending on the training tuples and the FIS. ANFIS is this trained adaptive network that is now to be used to generate the class label for the input slices. For the testing phase we enter each of the tuples that can then give us a predicted class output. The output received in this case is fuzzy output. As class labels are crisp values we need to defuzzify the output. This defuzzified output is then compared to the actual class label to get the performance measurement of this classifier.

Table 1. Feature Vector from the Normalized Grayscale MR Image

Mean	Variance	Skewness	Kurtosis	Energy	Entropy	Contrast	Correlation	Energy	Homogeneity	Inverse difference	Absolute value
15.5399	446.433	0.000108	1.27E-05	0.2816	3.6817	0.3334	0.3878	0.4131	0.8726	56710.6	17910
14.7607	427.681	0.000126	1.51E-05	0.2933	3.5751	0.3190	0.3833	0.4379	0.8774	57036.6	17210
15.5426	447.411	0.000108	1.27E-05	0.2810	3.6939	0.3415	0.3862	0.4071	0.8697	56512.8	18328
15.3294	434.847	0.000116	1.4E-05	0.2762	3.7014	0.3547	0.3508	0.4112	0.8659	56247.2	18914
15.9980	466.920	0.0001	1.18E-05	0.2718	3.8210	0.3618	0.3884	0.3971	0.8663	56246	18994
16.2995	455.317	9.72E-05	1.17E-05	0.2540	3.8918	0.3499	0.3785	0.3871	0.8652	56218.6	18910
16.8535	461.050	9.06E-05	1.1E-05	0.2365	3.9722	0.3697	0.3640	0.3731	0.8592	55806.6	19812
15.2010	463.274	0.000113	1.32E-05	0.2981	3.6796	0.3610	0.3920	0.4227	0.8712	56534.6	18504
13.6257	406.182	0.00015	1.85E-05	0.3333	3.2856	0.2931	0.3803	0.4756	0.8861	57632.2	15936
13.9531	413.331	0.000143	1.75E-05	0.3219	3.4029	0.2984	0.3755	0.4668	0.8831	57439.6	16314
13.1538	400.748	0.000161	2.00E-05	0.3529	3.2298	0.2885	0.3749	0.4930	0.8896	57854	15516
14.2234	417.177	0.000137	1.65E-05	0.3123	3.4526	0.3019	0.3595	0.4589	0.8800	57245.4	16676
13.6782	420.489	0.000142	1.67E-05	0.3506	3.2687	0.2558	0.4260	0.4907	0.8970	58395.4	14258
13.5103	417.582	0.000144	1.69E-05	0.3583	3.1849	0.2565	0.4179	0.4908	0.8972	58406	14248
13.2064	412.893	0.000151	1.78E-05	0.3698	3.1134	0.2443	0.4349	0.5019	0.9011	58674.6	13668
12.3658	388.257	0.000181	2.26E-05	0.3857	3.0153	0.2395	0.4074	0.5274	0.9029	58796.6	13412
12.9241	397.207	0.000167	2.1E-05	0.3573	3.2094	0.2881	0.3883	0.4918	0.8902	57893.2	15446
13.5875	407.496	0.000151	1.85E-05	0.3342	3.3352	0.2936	0.3933	0.4727	0.8867	57670.4	15878
14.2193	425.415	0.000137	1.64E-05	0.3174	3.4692	0.3158	0.3685	0.4617	0.8804	57227	16858
14.9834	443.361	0.000118	1.38E-05	0.3040	3.5664	0.3164	0.3804	0.4429	0.8786	57115.6	17050

In the testing phase we have used ANFIS as the primary classifier and the other two as secondary ones to do a performance measurement among the classifiers. In the event of classification we have selected five class labels for each type of tumor. The class labels mentioned here were taken accordingly as: Class 1 = Glioma, Class 2 = Meningioma, Class 3 = Metastatic adenocarcinoma, Class 4 = Metastatic bronchogenic carcinoma, Class 5 = Sarcoma. As we cannot use the actual names while using classification, we actually use the class labels which are numbers is the actual output of the classifier. The classifier gives fuzzy floating point numbers as output which is defuzzified to get the crisp actual class labeled values.

Performance measurement is done by comparing the actual class labels and the predicted class labels for each of the individual classifiers. This comparison is done to evaluate performance of each model. The performances are measured using confusion matrix, classification accuracy, error rate, and kappa statistic to measure [11] the improvement of the proposed system over the other two models. The other two models are selected based on their efficiency to classify the input slices. The output confusion matrix as illustrated in figure 4, figure 5, and figure 6 gives us the detailed output and describes the conflicting classified output with the actual output. In our case we have selected few performance measure statistics to differentiate the performance among the classifiers.

The confusion matrix can define the comparison of classified output to the actual output resulting in 2D matrix. This matrix as illustrates in figure 4, figure 5, and figure 6 gives us the classification accuracy and the elements in the matrix define different situations that can occur while classification. The matrix shows that all the diagonal elements show the number of correctly classified tuples while rest of the elements defines the tuples that were incorrectly classified. Kappa Index and Jacard index [11] denote the accurateness of the methods, whereas accuracy denotes the correct detection of the method.

Figure 4. Confusion Matrix for ANFIS

Figure 5. Confusion Matrix for MLP

Using this matrix we may derive various statistics such as kappa statistic that can be helpful in stating the superiority of ANFIS in comparison to other two classifiers. This can be illustrated in figure 7.

We observe from table 1 that the classifiers namely MLP and K-NN gives us nearly same results but ANFIS results in higher accuracy. The classification accuracy is 5% more for ANFIS in comparison to MLP and K-NN. As we have use low number of input dataset it may be noted that we are getting high classification accuracy for each of the classifiers but it is the same if we take large number of images as input.

ANFIS uses both neural network and fuzzy logic which gives an improvement over using traditional neural network. As we have seen in the classification results in figure 7 considering all the classifiers, we get an error rate as low as 5% in comparison to other two as 10%. With such a low error rate ANFIS proves its superiority even when the available training dataset is small. Similarly, the Kappa and Jaccard index is higher for ANFIS as observed from figure 7. In our case we have used hybrid learning that tends to

follow the data pattern. In figure 5 we have shown the suitability of this classifier over the others. The plot of the bar chart shows the suitability of ANFIS over others.
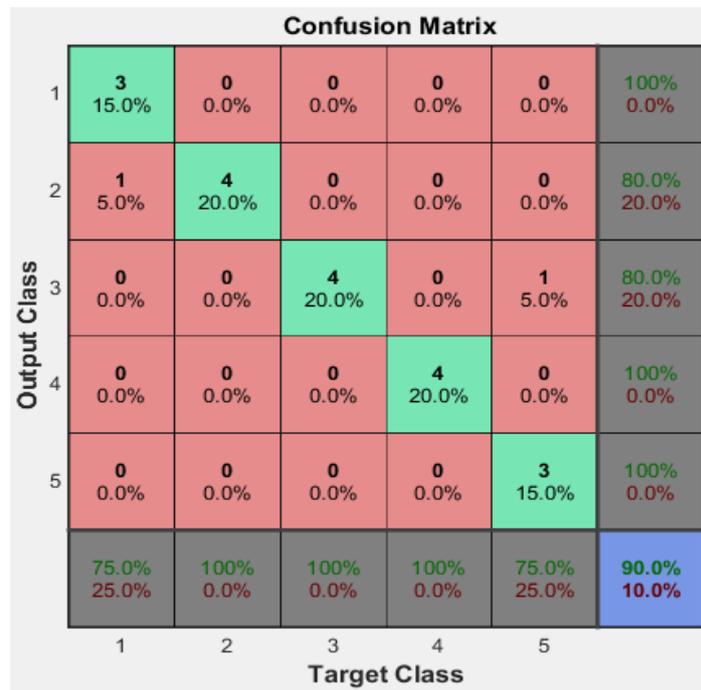
Figure 6. Confusion matrix for K-NN
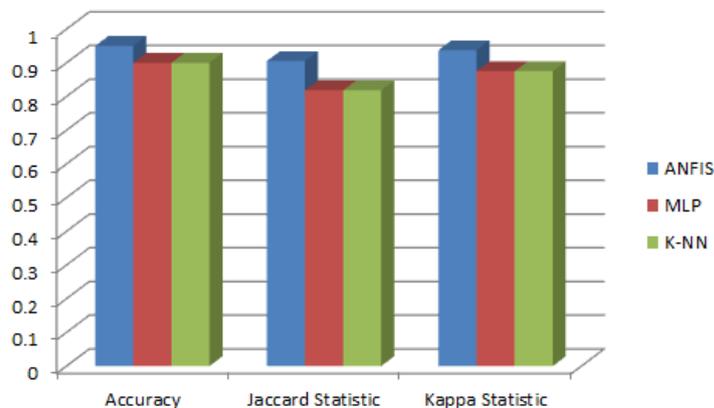
Figure 7: Statistical Comparison among the Classifiers

5. Conclusion

Our method is an improved classifier with ANFIS for brain tumor tissue characterization has been explored. The classifier obtained 98.25% accuracy on Harvard benchmark dataset [14], for both contrast and non-contrast images. The significance of feature sub-selection was revealed. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data, one of the major problems stems from the number of variables is involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and also generalizes successfully to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while describing the data with sufficient accuracy.

Feature selection is the technique of selecting a subset of relevant features for building robust learning models by removing most irrelevant and redundant features from the data, feature selection helps improve the performance of learning models by: i) Alleviating the effect of the curse of dimensionality, ii) Speeding up learning process, iii) Enhancing generalization capability iv) Improving model interpretability. Nevertheless, surrender of this stage, leading to huge dimensionality in feature space and ill conditioning of classifier performance. Automation of a model for computing an estimate of the type of tumor are verified by a radiologist, and a simultaneous measure of the quality of each phase is required to readily assess the automated image classification and segmentation algorithm performance. The brain and tumor tissue identification provides a better perceptive of the spatial relationship; thereby lend assistance to the adage of pre-operative treatment planning. However, more real time extensive training studies are required to further substantiate these effects to further validate the performance of this computer analysis methodology.

Acknowledgments

References

[1] Y. H. Liu, M. Muftah, T. Das, L. Bai, K. Robson and D. Auer, "Classification of MR Tumor Images Based on Gabor Wavelet Analysis", Journal of Medical and Biological Engineering, vol. 32 no. 1, **(2011)**, pp. 22-28.

[2] A. E. Sayed, E. Dahshan, M. A. Badeeh, Salem and T. H. Younis, "A hybrid technique for automatic MRI brain images classification", Studia Univ, Babes Bolyai, Informatica, vol. LIV, **(2009)**.

[3] V. P. Gladis, P. Rathi and S. Palani, "Brain Tumor MRI Image Classification with Feature Selection and Extraction Using Linear Discriminant Analysis", JIST, vol. 2, no. 4, **(2012)**.

[4] E. I. Zacharaki, S. Wang, S. Chawla, D. S. Yoo, R. Wolf, E. R. Melhem and C. Davatzikos, "MRI-based classification of brain tumor type and grade using SVM-RFE", IEEE International Symposium on Biomedical Imaging, **(2009)**, pp. 1035-1038.

[5] Q. Wang, E. K. Liacouras, E. Miranda, U. S. Kanamalla and V. Megalooikonomou, "Classification of brain tumors using MRI and MRS data", Proceedings of the SPIE, **(2007)**, vol. 6514.

[6] E. I. Papageorgiou and P. P. Spyridonos, "Brain Tumor characterization using the soft computing technique of fuzzy cognitive maps", Applied soft computing, vol. 8, **(2008)**, pp. 820-828.

[7] S. Ibrahim and N. E. A. Khalid, "Image Mosaicing for evalustion of MRI Brain Tissue abnormalities segmentation study", International Journal of Biology and Biomedical Engineering, vol. 5 no. 4, **(2011)**, pp. 181-189.

[8] S. Karpagam and S Gowri, "Detection of tumor growth by advanced diameter technique using MRI data", Proc. The world congress of Engineering, WEC, London, U.K., vol.1, **(2011)**.

[9] M. C. Clark and L. O. Hall , "Automatic tumor segmentation using Knowledge based techniques", IEEE transactions on medical imaging, vol. 17, no. 2, **(1998)**.

[10] C. A. Parra and K. Iftekharuddin, "Automated brain data segmentation and Pattern recognition using ANN", Proceedings of CIRAS, **(2003)**.

[11] S. Roy, P. Ghosh and S. K. Bandyopadhyay, "Segmentation and Contour Extraction of Cerebral Hemorrhage from MRI of Brain by Gamma Transformation Approach", FICTA, Proceedings of Springer, AISC, vol. 328, **(2014)**.

[12] S. Roy, K. Chatterjee, I. K. Maitra and S. K. Bandyopadhyay, "Artefact Removal from MRI of Brain Image", IRJES, vol. 2, no. 3, **(2013)**, pp. 24-30.

[13] H. G. Moreno, P. G. Jiménez, S. L. Arroyo, R. L. Sastre and S. M. Bascón, "A "Salt and Pepper" Noise Reduction Scheme for Digital Images Based on Support Vector Machines Classification and Regression", The Scientific World Journal, vol. 2014, **(2014)**.

[14] K. A. Johnson and J. Alex Becker, "The whole brain atlas", http://www.med.harvard.edu/aanlib/home.html, 2014

[15] Andrzej & Michal, "Texture Analysis Methods – A Review", Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels, **(1998)**.

[16] K. D. Kharat, P. P. Kulkarni and M. B. Nagori, "Brain Tumor Classification Using Neural Network Based Methods", International Journal of Computer Science and Informatics ISSN (PRINT): 2231 – 5292, vol. 1, iss. 4, **(2012)**.

[17] N. V. Chavan, B. D. Jadhav and P. M. Patil, "Detection and Classification of Brain Tumors", International Journal of Computer Applications, vol. 112, no. 8, **(2015)**.

[18] K. Mehran, "Takagi-Sugeno Fuzzy Modeling for Process Control", Industrial Automation, Robotics and Artificial Intelligence, School of Electrical, Electronic and Computer Engineering, **(2008)**, pp. 1-21.