International Journal of Technical Research and Applications e-ISSN:2320-8163,
www.ijtra.com Volume 2, Issue 4 (July-Aug 2014), PP. 26-31

BRAIN CANCER CLASSIFICATION USING BACK PROPAGATION NEURAL NETWORK AND PRINCIPLE COMPONENT ANALYSIS

Ganesh Ram Nayak[1], Mr. Toran Verma[2]
[1]M.Tech. (CSE), [2]Sr. Associate Professor,
Rungta College of Engg. & Technology,
Bhilai, Chhattisgarh, India – 490021.

***Abstract*— Classification of Brain Cancer is implemented by using Back Propagation Neural network and Principle Component Analysis, Magnetic Resonance Imaging of brain cancer affected patients are taken for classification of brain cancer. Image processing techniques are used for processing the MRI images which are image preprocessing, image segmentation and feature extraction is used. We extract the Texture feature of segmented image by using Gray Level Co-occurrence Matrix (GLCM). Steps involve for brain cancer classification are taking the MRI images, remove the noise by using image pre-processing, applying the segmentation method which isolate the tumor region from rest part of the MRI image by setting the pixel value 1 to tumor region and 0 to rest of the region, after this feature extraction technique has been applied for extracting texture feature and feature are stored in knowledge based, this features are used for classification of new MRI images taken for testing by comparing the feature of new images with stored features. We implemented three classifiers to classify the brain cancer, first classifier is back propagation neural network which perform classification in two phase which are training phase and testing phase, second classifier is the combination of PCA and BPNN means by using PCA to reduce the dimensionality of feature matrix and by using BPNN to classify the brain cancer, third classifier is Principle Component Analysis which reduce the dimensionality of dataset and perform classification. And finally compare the performance of that classifiers.**

***Key words*— Brain Cancer; MRI; Segmentation; Gray Level Co-occurrence Matrix; Principle Component Analysis; Back Propagation Neural Network.**

I.INTRODUCTION

The term Brain tumor is any mass that results from abnormal growths of cells in the brain. It may affect any person at almost any age. Brain tumor effects may not be the same for each person, and they may even change from one treatment session to the next. Brain tumors can have a variety of shapes and sizes; it can appear at any location and in different image intensities. Brain tumors can be benign or malignant [4].

Magnetic Resonance Imaging (MRI) has become a widely used method of high quality Medical imaging, especially in brain imaging where MRI's soft tissue contrast and noninvasiveness is a clear advantage. MRI provides an unparalleled view inside the human body. The level of detail we can see is extraordinary compared with any other imaging modality. Reliable and fast detection and classification of brain cancer is of major technical and economic importance for the doctors. Common practices based on specialized technicians are slow, have low responsibility and possess a degree of subjectivity which is hard to quantify [1] [7]. The advantage of magnetic resonance imaging (MRI) over other diagnostic imaging modalities is its high spatial resolution and excellent discrimination of soft tissues. MRI provides rich information about anatomical structure, enabling quantitative pathological or clinical studies [5].MRI is also a safe and valuable adjunct to the clinical examination of the knee and an aid to efficient preoperative planning. It is the most commonly used imaging modality in the evaluation of the knee joint [10].

A lot of research efforts have been directed towards the field of medical image analysis with the aim to assist in diagnosis and clinical studies. The medical images are obtained from different imaging systems such as MRI scan, CT scan and Ultra sound B scan. The computerized tomography has been found to be the most reliable method for early detection of tumors because this modality is the mostly used in radio therapy planning for two main reasons. The first reason is that scanner images contain anatomical information which offers the possibility to plan the direction and the entry points of radio therapy rays which have to target only the tumor region and to avoid other organs. The second reason is that CT scan images are obtained using rays, which is same principle as radio therapy. This is very important because the intensity of radio therapy rays have been computed from the scanned image [3]. The Medical Image for Brain Cancer Classification can be Magnetic Resonance Imaging (MRI) [1], or it can be Computed Tomography (CT) Scan [3].

The system uses many image processing techniques and classifiers which are explain in next section, and the classification result of brain cancer is shown in result section.

II.METHOD

The system is built by using many image processing techniques and classifiers, MRI images of brain cancer affected patients are taken as input and system give the class of that input MRI images. We organized a MRI images into five different classes in which some images from each classes are used for training the network and remaining images are for testing. To do so we used an image processing techniques which are image preprocessing, image segmentation and feature extraction. Finally classifiers are implemented for classifying the brain cancer.

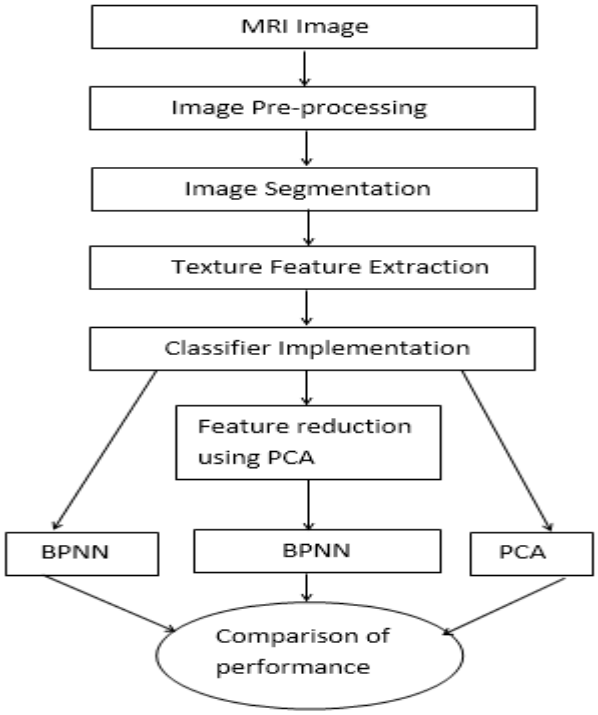

Figure 1.1 – Block diagram of the system

The block diagram of developed system is shown in the figure 1.1, which works in two phases, first phase is Training phase and second phase is Testing phase for two classifiers- BPNN and PCA with BPNN, in first phase it takes the MRI images of brain cancer affected patients and perform the image processing techniques shown in the figure 1.1 and feature are stored in the knowledge base and it is used for training the neural network, and in second phase testing MRI images are taken as a input and perform the same operation we performed before on training images and features of testing images are compared with stored image in knowledge base. The third classifier is Principle Component Analysis which takes features of both training and testing images to classify the brain cancer.

III.STAGES IN CLASSIFICATION

The stages of classification are shown in the figure 1.1, and detail of each stage is given below –

A. MRI Image

We used the MRI images of brain cancer affected patients, we takes the MRI images of five different diseases of brain cancer. The five types of MRI image are – Astrocytoma, Glioma, Meningioma, Metastasis bronchogenic carcinoma and Sarcoma. Each of these disease are organize to class means class I to class V.

B. Image Pre-processing

Brain images are noisy, inconsistent and incomplete, thus preprocessing phase is needed to improve the image quality and make the segmentation results more accurate [3].

For image preprocessing we used a Median filter, Median filtering is similar to using an averaging filter, in that each output pixel is set to an average of the pixel values in the neighborhood of the corresponding input pixel. However, with median filtering, the value of an output pixel is determined by the median of the neighborhood pixels, rather than the mean. The median is much less sensitive than the mean to extreme values (called outliers). Median filtering is therefore better able to remove these outliers without reducing the sharpness of the image. The medfilt2 function implements median filtering.

C. Image Segmentation

Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s).

By using image segmentation we isolate the tumor region from rest of the image. We applied the Optimum global Thresholding using Otsu method for image segmentation, this method computed the threshold value, each pixel's intensity of the image is compare with the threshold value, if the pixel's intensity is greater than threshold than pixel value is set to 1 otherwise set 0 and finally we get a segmented image.

D. Texture Feature Extraction

The work involves extraction of the important features for image recognition. The features extracted give the property of the texture, and are stored in knowledge base. [1]. the extracted features are compare with the unknown sample means the testing image for classification.

We used a Gray Level Co-occurrence matrix for texture feature extraction. Gray level co-occurrence matrix (GLCM) was firstly introduced by Haralick. A gray-level co-occurrence matrix (GLCM) is essentially a two-dimensional histogram. The GLCM method considers the spatial relationship between pixels of different gray levels. The method calculates a GLCM by calculating how often a pixel with a certain intensity i occurs in relation with another pixel j at a certain distance d and orientation Θ. A co-occurrence matrix is specified by the relative frequencies P (i, j, d, Θ). A co-occurrence matrix is therefore a function of distance d, angle Θ and gray scales i and j. [6] [8] [9].

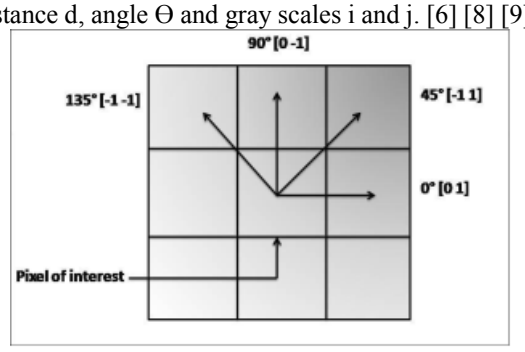

Figure 3.4.1 - Direction for generation of GLCM

We extracted the following texture feature which are given below –

1) Contrast:-

Contrast returns a measure of the intensity contrast between a pixel and its neighbor over the entire image [6][9].

Contrast = $\sum_{i=1}^{k}\sum_{j=1}^{k}(i-j)^2 p_{ij}$

Range = [0 (size(GLCM,1)-1)^2]

Contrast is 0 for constant image.

2) Correlation:-

Returns a measure of how correlated a pixel is to its neighbor over the whole image. [6][9].

Range = [-1 1]

Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN for a constant image.

Correlation = $\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{(i-m_r)(j-m_c)p_{ij}}{\sigma_r\sigma_c}$

$$\sigma_r \neq 0, \sigma_c \neq 0$$

Where m_r and m_c are mean computed along rows and column respectively, σ_r and σ_c are in form of standard deviations computed along rows and column respectively.

3) Angular Second Moment (Uniformity or Energy):-

Angular Second Moment is also known as Uniformity or Energy. It is the sum of squares of entries in the GLCM Angular Second Moment measures the image homogeneity. Angular Second Moment is high when image has very good homogeneity or when pixels are very similar [6][9].

Energy = $\sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij}^{\ 2}$

Rang= [0,1]

Energy is 1 for constant image.

4) Inverse Difference Moment (Homogeneity):-

Inverse Difference Moment (IDM) is the local homogeneity. It is high when local gray level is uniform and inverse GLCM is high. It return a value that measure the closeness of the distribution of element in the G to the diagonal of G.[6][9]

Rang = [0,1]

Homogeneity is 1 for diagonal G.

Homogeneity = $\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{p_{ij}}{1+|i-j|}$

5) Entropy:-

Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image. Entropy is defined as

Entropy = $-\sum_{i=1}^{k}\sum_{j=1}^{k} p_{ij}\log_2 p_{ij}$

Where k is the rows (or column) dimension of square matrix G probability is the ij-th element of G/n, where n is equal to sum of the element of G, and G is referred simply as co=occurrence matrix. [6][9].

E. Classifier Implementation

Classifiers are used to classify the brain cancer. We implemented three classifiers two classify the brain cancer which are given below –

1) Back Propagation Neural Network

Back propagation is a supervised learning method. In supervised learning, each input vector needs a corresponding target vector. Input vector and target vector are presented in training of the network. The output vector (i.e. actual output) which is result of the network is compared with the target output vector then an error signal is generated by the network. This error signal is used for adjustment of weights until the actual output matches the target output. Algorithm stages for BPN are initialization of weights, feed forward, back propagation of Error and updating of weights and biases [6] [2].

After the weights are adjusted on the training set, their value is fixed and the ANN's are used to classify unknown input images. The generalized delta rule involves minimizing an error term defined as [1]

$$E_p = \frac{1}{2}\sum_j (t_{pj} - o_{pj})^2$$

In this equation, the index p corresponds to one input vector, and the vectors tp and op are the target and observed output vectors corresponding to the input vector p, respectively [1].

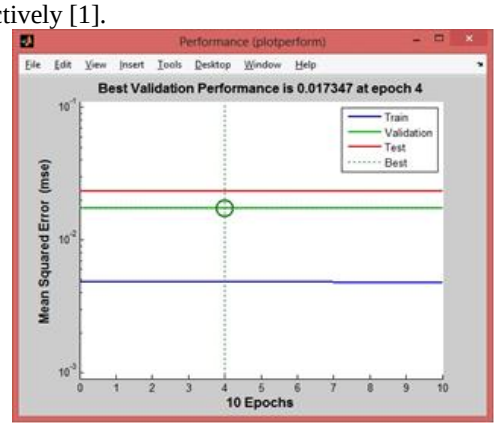

Figure 3.5.1 – Performance graph of BPNN

2) PCA with Back Propagation Neural Network

In this classifier we reduce the no of features by using Principle Component Analysis. We takes a 15 samples of MRI images for training the neural network, feature is extracted using gray level co-occurrence matrix and we got a 113 features value of each images and features of all images are reduce by using Principle Component Analysis and we got a reduce data set. Set procedure is applied for testing MRI images and classification is done by using Back Propagation Neural network explained in previous section.

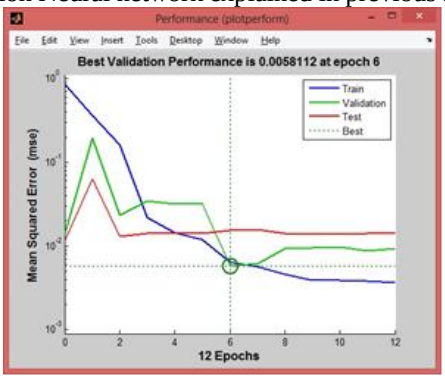

Figure 3.5.2 – Performance graph of PCA with BPNN

3) Principle Component Analysis

Here Principle Component Analysis is used as a classifier, in which it derives the new feature set from large extracted feature using GLCM. After deriving the new feature it applies a Euclidian distance for brain cancer classification.

The mathematical background required for PCA is given below-

a) Mean
b) Standard Deviation
c) Variance
d) Covariance
e) Covariance Matrix
f) Eigen Vector and Eigen value

a) Mean

Notice the symbol $\bar{X}$ (said "X bar") to indicate the mean of the set. All this formula says is "Add up all the numbers and then divide by how many there are" [11].

$$\bar{X} = \frac{\sum_{i=1}^{n} Xi}{n}$$

b) Standard Deviation

The Standard Deviation (SD) of a data set is a measure of how spread out the data is. How do we calculate it? The English definition of the SD is: "The average distance from the mean of the data set to a point". The way to calculate it is to compute the squares of the distance from each data point to the mean of the set, add them all up, divide by, (n-1) and take the positive square root. As a formula [11]

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}}$$

c) Variance

Variance is another measure of the spread of data in a data set. In fact it is almost identical to the standard deviation. The formula is this:[11]

$$var(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{(n-1)}$$

d) Co – variance

Standard deviation and variance only operate on 1 dimension, so that you could only calculate the standard deviation for each dimension of the data set independently of the other dimensions. However, it is useful to have a similar measure to find out how much the dimensions vary from the mean with respect to each other. Covariance is such a measure. Covariance is always measured between 2 dimensions. If you calculate the covariance between one dimension and itself, you get the variance. So, if you had a 3-dimensional data set (x, y, z), then you could measure the covariance between the x and y dimensions, the x and z dimensions, and the y and z dimensions. Measuring the covariance between x and x, or y and y , or z and z would give you the variance of the < , = and > dimensions respectively. The formula for covariance is very similar to the formula for variance. The formula for variance could also be written like this:[11]

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

e) Covariance Matrix

Covariance is always measured between 2 dimensions. If we have a data set with more than 2 dimensions, there is more than one covariance measurement that can be calculated. For three dimensional (x, y, z) data set calculate cov(x,y), cov(x,z) and cov(y,z) In fact for n dimensional data set, we can calculate total covariance values.[11]

$$\frac{n!}{(n-2)! * 2}$$

$$C = \begin{matrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{matrix}$$

f) Eigenvector and eigenvalue

As you know, you can multiply two matrices together, provided they are compatible sizes. Eigenvectors are a special case of this. Consider the two multiplications between a matrix and a vector in In the first example, the resulting vector is not an integer multiple of the original vector, whereas in the second example, the example is exactly 4 times the vector we began with. Why is this? Well, the vector is a vector in 2 dimensional space The other matrix, the square one, can be thought of as a transformation matrix. If you multiply this matrix on the left of a vector, the answer is another vector that is transformed from its original position. What properties do these eigenvectors have? You should first know that eigenvectors can only be found for square matrices. And, not every square matrix has eigenvectors.[11]

$$\begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

STEPS TO BE FOLLOWED IN PCA [11]

1. Get some data
2. Subtract the mean
3. Calculate the covariance matrix
4. Calculate the eigenvectors and eigenvalues of the covariance matrix
5. Choosing components and forming a feature vector
6. Deriving the new data set
7. Calculate the Euclidian distance between feature of unknown MRI image and All stored Feature, and find out minimum distance and according to these distance unknown samples are classified.

IV.RESULTS

The developed system classifies the brain cancer of MRI images of brain cancer affected patients. We implemented three classifiers, each classifiers have their own efficiency to classify the brain cancer. We takes the five types of brain cancer which are organize into five different classes from class I to class V. We isolate the tumor region from rest of the MRI image by using segmentation, which shows the tumor region of MRI images which is shown in the figure below –

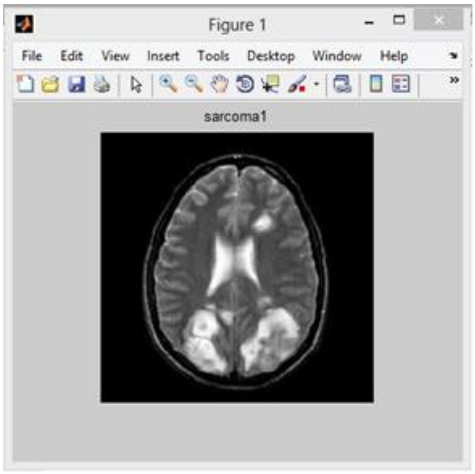

Figure 4.1 – Original MRI image

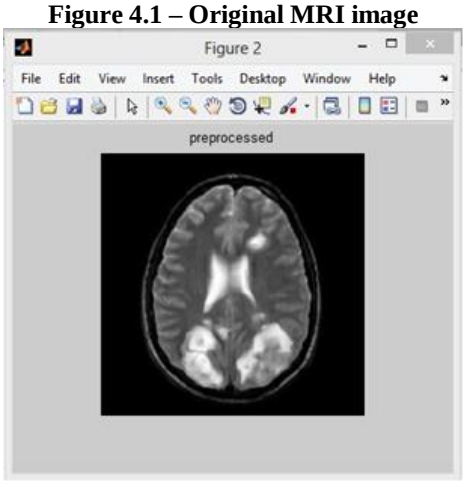

Figure 4.2 – Preprocessed MRI image

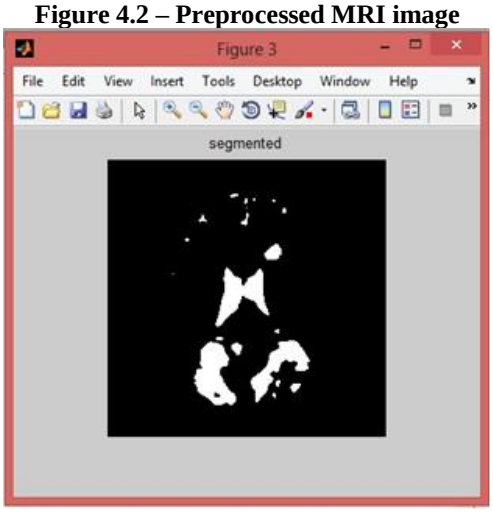

Figure 4.3 – Segmented MRI image

Figure 4.3 shows the tumor region of MRI image which image is used for feature extraction.

We developed Graphical User Interface for user friendly operation on the system, the GUI contain all the component of developed system means feature extraction of both training images and testing images, dedicated button of all three classifiers, training of neural network and training of PCA with neural network. The performance of three classifiers are shown in the table 4.1, which is given below –

Number of Testing sample	No. of Testing images successful classified by classifiers		
	BPNN	PCA with BPNN	PCA
5	4	3	5

Table 4.1 – Performance of Classifiers

V. Conclusion & Scope of Further Work

The brain cancer classification system has been designed by using Back propagation neural network and Principle Component Analysis which uses image processing techniques like image preprocessing, image segmentation and feature extraction using Gray Level Co-occurrence matrix. The system uses three classifiers for classification of brain cancer which are Back Propagation Neural Network, PCA with BPNN and Principle Component Analysis. The classification performance of three classifiers are shown in the table 4.1, which shows that performance of Principle Component Analysis is better as compare to two other classifiers for classification of brain cancer.

This system classify a few type of brain cancer, the main aim of this system is to compare the performance of classifier which are used in this system. The system can be implemented which classify all type of brain cancer by using appropriate classifier for each type of cancer. The scope of the system can further be improved by using other types (e.g. PET, MRS, CTS) of Images.

References

[1] Dipali M. Joshi, Dr. N. K. Rana and V. M. Misra. 2010. "Classification Of Brain Cancer Using Artificial Neural Network", 2nd International Conference on Electronic Computer Technology, pp. 112-116.

[2] Kailash D.Kharat, Pradyumna P.Kulkarni, M.B.Nagori.2012. "Brain Tumor Classification Using Neural Network Based Method", International Journal of Computer Science and Informatics, vol. – 1, pp. – 85-90.

[3] A. Padma, R. Sukanesh. 2011. "Automatic Classification and Segmentation of Brain Tumor in CT images Using Optimal Dominant Gray Level Run Length Texture Feature", International Journal of Advanced Computer Science and Applications,Vol. 2, No. 10, pp. – 53-59.

[4] Ahmed Kharrat, Karim Gasmi, Mohamed BEN. 2010. "A Hybrid Approach for Automatic Classification of Brain MRI Using Genetic Algorithm and Support Vector Machine", Leonardo Journal of Sciences, pp. 71-82.

[5] W. M. Wells, W. E. L. Grimson, R. Kikinis, F. A. Jolesz. 1996. "Adaptive Segmentation of MRI Data", IEEE Transaction on Medical Imaging, Vol. – 15, No. – 4, pp.429-442.

[6] Shweta Jain, Shubha Mishra. 2013. "ANN Approach Based on Back Propagation Network and Probabilistic Neural Network to Classify Brain Cancer", International Journal of Innovative Technology and Exploring Engineering, Vol. – 3, Issue – 3, pp. – 101-105.

[7] Matthew C. Clark, Lawrence O. Hall, Dmitry B. Goldgof, Robert Velthuzen, F. Reed Murtagh, Martin S. Silbiger. 1998. "Automatic Tumor Segmentation using Knowledge Based Techniques". IEEE Transaction on Medical Imaging, Vol. – 17, No. – 2, pp. – 187-201.

[8] Minakshi Sharma, Dr. Sourabh Mukharjee. 2012. "Brain Tumor Segmentation using hybrid Genetic Algorithm and Artificial Neural Network Fuzzy Inference System", International Journal of Fuzzy Logic System, Vol. – 2, No. – 4, pp. – 31-42.

[9] P. Mohanaiah, P. Sathyanarayana, L. GuruKumar. 2013. "Image Texture Feature Extraction Using GLCM Approach".

International Journal of Scientific and Research Publications, Vol. – 3, Issue – 5, pp. – 1-5.

[10] Mustafa Z. Mahmoud, Maram A. Fagiri, Atallh Fowzan Al-Motrfi, Abdelmoneim Sulieman. 2013. "Magnetic Resonance Imaging Findings in Knee Joint Pain at King Saud Medical City". International Journal of Science and Research, Vol. – 2, Issue – 6, pp. – 4-7.

[11] Varada S.Kolge, Prof.K.V.Kulhalli. "PCA and PNN assisted automated brain tumor classification", IOSR Journal of Electronic and Communication Engineering, pp. - 19-23.

Author Profile

Ganesh Ram Nayak, completed B.E. Information Technology from Shri Shankaracharya College of Engineering and Technology, Bhilai (C.G) in 2012 and Now he is pursuing M.Tech in Computer Science & Engineering from Rungta College of Engineering and Technology, Bhilai (C.G).

Mr. Toran Verma working as Associate Professor in Department of Computer Science & Engineering at Rungta College of Engineering and Technology, Bhilai (C.G).