Identification of Brain Tumor using Probabilistic Neural Networks and K-Means Segmentation

Rakshita R S[1], Sanjana P S[2], Suditi Hoysal[3], Dr. Kiran Y.C[4]

[1]Rakshita R S, Student,
[2]Sanjana P S, Student
[3]Suditi Hoysal, Student

[4]Kiran Y.C, Professor, Dept. of ISE, BNMIT, Bangalore, India

---

Abstract - *'Identification of Brain Tumor using Probabistic Neural Networks and K-means Segmentation' is a model built for automated classification of brain tumor in the input Magnetic Resonance Imaging (MRI) image. Using the MRI image of the brain as input, the proposed system aims to efficiently and accurately classify images into tumor and no tumor. The model implements various processes like pre-processing, decomposition, feature extraction and segmentation to achieve the objectives. A combination of Discrete Wavelet Transform and Gray Level Co-Occurrence Matrix is used for optimal feature Extraction. This is followed by classifying the images using Probabilistic Neural Network (PNN) which uses weighted values to classify and therefore result in higher accuracy. Accurate classification of MRI is necessary for medical treatment. Brain tumors have been seen in all ages of the human species and therefore steps taken to detect these tumors is also become more vital in this era of the health industry. Identifying these tumors will help medical professionals to identify and detect the brain tumor in a faster and more efficient way.*

Key Words: Gaussian Blur, Mean Filter, DWT, GLCM, PNN

1. INTRODUCTION

A brain tumor is the spread of unusual or uneven cells inside the brain and are of two types, namely noncancerous or benign and cancerous or malignant tumors. They can either begin in the brain (which is called essential), or in distinct parts/organs of the body and outspread to the brain (called the secondary). The care for cancer depends on size, location and type of the tumor. Their presence can cause pressure to build up inside the skull and can cause damage to the brain and in some cases prove to be fatal.

Presently, Brain anatomies are recorded and diagnosed by the Magnetic Resonance Imaging technology. But the MRI systems are not built to process these images, but exist only for imaging and therefore cannot give results on areas depending on such processing of images. Tumors in the brain are a major cause of fatality rates globally and the related deformities' can cause major transformations in life. Since the skull has constrained area inside, pressure within the skull increases as they develop, and cause Edam, which is decrease in the flow of blood and removal of the consequent degeneration of the healthy tissue.

MR imaging is an efficient and a cutting-edge imaging technique used to create top-quality and clear images of the organs of the human body. They are mostly used for the ministrations of brain tumors and other parts of the body. From these images that show the high amount of details of an image, detailed pathological data can be derived to study the development of human brains and find abnormalities. At present times there are many methods for categorizing MRI images, like neural networks, Discrete Cosine Transform(DCT), thresholding and variation segmentation. It uses a robust magnetic field and radio waves to produce images of the organs and tissues present in the human body.

Identification of the brain tumor is executed though a systematic flow of processes such as pre-processing, feature extraction, classification and segmentation. Pre-processing is performed to remove noise and unnecessary information. Feature extraction helps to extract the most important attributes such as contrast, correlation, homogeneity, energy, asm and dissimilarity. Classification is performed through Probabilistic neural network(PNN), a neural network wherein connections between the nodes do not form a cycle, which classifies images based on extracted features to provide a result on the type of tumor. Lastly, segmentation using K-means is performed to separate the tumor mass (if any) from the brain mass.

2. LITERATURE SURVEY

The literature survey on the identification of brain tumor using MRI scans employs a few common methodologies but this paper focuses on a different approach to classification using the Probabilistic Neural Network(PNN) algorithm.

The following papers mentioned henceforth mentions methodologies unique to them.

A survey paper on image analysis using biologically affected BWT and SVM is implemented by four main steps: pre-processing, skull stripping, segmentation and morphological operations. The unique aspect pertaining to this paper is the skull stripping process which is used for removing all non-brain tissues in the brain images such as skin and fat.

For extraction of the infected brain area, the image is first transformed into a binary image and in order to discard the white element, a destruction operation is applied. Finally, the destructed part and the original MRI image are split into two identical sectors and the black element section is considered as an MRI image mask. This is done in order to select prominent features. However, the accuracy that results from this system is a limitation [1].

The paper on "Image Segmentation for Detection of Benign and Malignant Tumour Using DCT and DWT Combination" proposes image histogram stretching and equalisation for improved accuracy. Here the image is clustered using 8x8 mass and Discrete Cosine Transform (DCT) is computed on the 8x8 smaller images. Inverse DCT is then applied to produce the primary MRI that has reduced optimal data. The resulted image is then binarized using Ostus Algorithm. Principal Component Analysis (PCA) is applied to the wavelet units. Support Vector Machine is then used to categorize the tumour. A large dataset of MRI images is then used to verify the results and determine the accuracy [2].

Hang Chen in the paper "Brain Tumour Segmentation Survival Prediction with Deep Learning" has proposed that because the MRI intensity value is reliant on the protocol and also the type of scanner employed, intensity normalisation is made use of to decrease the bias in imaging. Random Gaussian noise and random flipping is applied on the training set to reduce over-fitting. Categories such as first order statistics which include mean, median, tenth percentile, etcetera categorize the grey level intensity. Shape features which includes surface area, volume, etcetera categorize the shape of the tumor area. Another class is the texture attributes such as 22 Co-occurrence matrix features, 14 grey level dependence matrix features categorize the texture. Categorizing these features can be time consuming [3].

"Identification of Brain Tumours Using Ensemble Classifiers" proposed by B Vijay Kumar and Parasuraman Kumar states that using ensemble classifier which is a combination of neural networks automates the classification of brain tumour. The performance of the model on a new data has a negative impact as overfitting occurs as the model learns the detail and noise in the training data at an earlier stage [4].

Yousif Ahmed has proposed in the paper "Detection of Brain Tumour Using A Combination Of Fuzzy C-Means and Thresholding" three main steps for detection which includes pre-processing, segmentation of the image and contour representation. For segmentation Fuzzy C-Mean Clustering and Thresholding methods are used. The contour map is generated with algorithms such as Roberts, Prewitt, Sobel and LoG and Canny Detector. The LoG and Canny detector poses a limitation due to its long waiting time and complex compution [5].

3. METHODOLOGY

In the proposed method, detecting and classifying MRI images having a brain tumor using DWT and PNN method is done. Extracting the brain tumor region for analysis uses the K-Means segmentation. Fig -1 presents the architecture design of the proposed system. The system has five modules which are Pre-Processing, Decomposition, Feature Extraction, Classification and Segmentation. The project aims to output the result 'Tumor' for tumor affected MRI image and 'No Tumor' for MRI image that does not have tumor. In order to get to the result, preprocessing is done on the input image. The preprocessing module is employed to discard any unwanted noise from the image. This process is an important step as it affects the accuracy of the result. Feature Extraction is another important module. This step aims to extract essential characteristics from the image based on which the classification is done. For feature extraction, DWT is first implemented. DWT or Discrete Wavelet Transform is a process of decomposition yields the approximation description of the image which is in turn used for extracting optimal features. For optimal texture feature extraction, the GLCM matrix is computed and the feature vector is obtained.

Feature vector for a set of training dataset is obtained and trained using the Probabilistic Neural Network (PNN). A MRI image is then tested to classify as Tumor or No Tumor. A tumor affected image is finally used for segmentation where in the tumor affected region is extracted. Extraction is done using the K-Means Clustering.

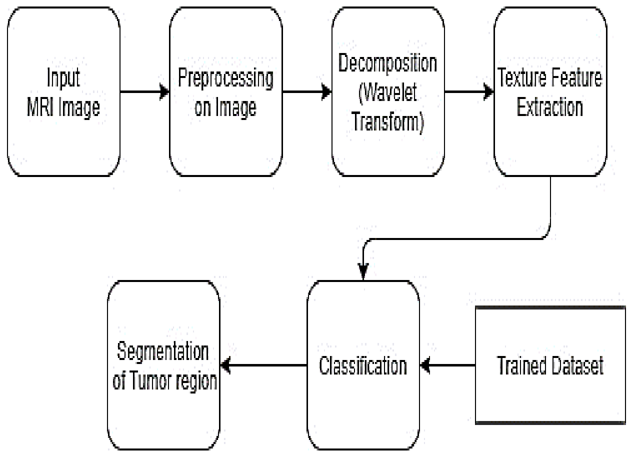


Fig -1: Architecture design

3.1 Pre-processing

The function of the image pre-processing module is to upgrade or enhance the data of the image and eliminate unwanted disturbances or interferences. Fig -2 shows the major steps involved in pre-processing which is Gray Conversion, Filtering and Binarization.

RCB to Grayscale Conversion: Here, the image may comprise of shades of gray. A 'gray' shading is one in which the red, green and blue components have comparative force in RGB space. These images should be changed over into grayscale which range from 0 to 255 pixel where range 0 characterizes the unadulterated dark shading and range 255 characterizes unadulterated white shading.

Noise Removal using Filtering: Filtering is a method utilized for removing out the noise present inside an image. During the transformation of an image from RGB to gray a type of disturbance creeps into the image. Interruptions like Gaussian and Poisson noise impure the MRI image. In the proposed system, algorithms used for filtering is the 'Gaussian Filter' whose processed image is then sent to the algorithm 'Mean Filter'.

Binarization: Image binarization is a procedure of taking a grayscale image and changing over it to high contrast, basically reducing the data contained inside the image from 256 shades of gray to 2: high contrast (black and white), a binary image.

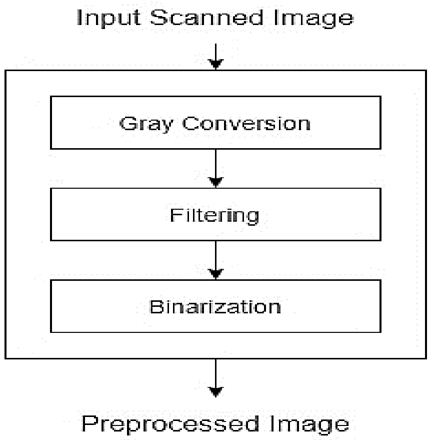


Fig -2: Pre-processing steps

3.2 Feature Extraction

Feature Extraction is a simple step that describes how an image is shaped into a group of attributes or features. Optimal features are drawn out from the original image for classification. The process of extraction of features or characteristics is one of the most pivotal and crucial process for classifying of an image.

In order to do this, the Discrete Wavelet Transform (DWT) is implemented for breakdown of the image into distinct bands. DWT is a wavelet transform used for analyzing an image and extracting feature. The process yields 4 sub-bands that describes the image along the vertical, horizontal, diagonal and approximation details. The approximation sub-band image data yields major information of the image and hence is selected for the construction of the GLCM matrix.

The Gray Co-Occurrence Matrix (GLCM) is a analytical procedure of examining texture of an image by taking into account the spatial relation of the image pixels. It is one of the feature removal process for acquiring the optimal texture features of the image after decomposition. Once the GLCM matrix is computed for an image, GLCM features are calculated for that image which yield a feature vector. The features considered are Contrast, correlation, Homogeneity, Dissimilarity, ASM and Energy. Having obtained the feature vector, it is proceeded for classification. Fig -3 depicts the process of feature extraction.

Table -1: Feature Definition and Formula

FEATURE	DEFINITION	FORMULA
contrast	Distinction between highest and lowest valued pixel intensity in an image.	$\sum_{i,j} \lvert i-j \rvert^2 p(i,j)$
correlation	An Estimation of how connected a pixel is to its neighbor in the entire image.	$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)\, p(i,j)}{\sigma_i \sigma_j}$
energy	Measure of information or localized change of the image.	$\sum_{i,j} p(i,j)^2$
homogeneity	Measure of closeness of the scattering of components to the GLCM diagonal.	$\sum_{i,j} \frac{p(i,j)}{1+\lvert i-j \rvert}$
asm	Returns the square of energy	$\sum_{i,j=0}^{levels-1} P_{i,j}^2$
dissimilarity	Returns the difference between pixels	$\sum_{i,j=0}^{levels-1} P_{i,j} \lvert i-j \rvert$

The above Table -1 defines the meaning of each of the texture features along with its formula.

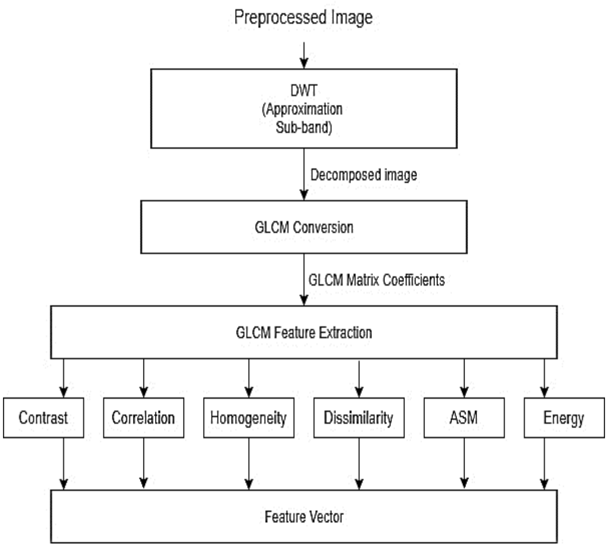


Fig -3: Feature Extraction process

3.3 Classification using PNN

The resultant feature vector is computed for a set of training dataset. These values are stored in the form of a csv file with the last column storing the target value for that row. The target field has values 0 for no tumor and 1 for tumor. This csv file is then given as input into the Probabilistic Neural Network (PNN) for classification. The PNN is made up of four layers namely, the input, pattern, summation and the output or the decision layer. The vector values go through each of these layers several times and computes a weight for the input.

The test image is given as an input into such a trained network based on which the MRI image classification is performed. The result of the classification is a message depicting 'Tumor' for the presence of tumor or 'No Tumor' for the absence of tumor in the MRI test image.

3.4 Segmentation using K-Means Clustering

Segmentation is done to extract the tumor region from the tumor affected MRI input image, separating the normal tissue segments from affected tumor tissues. Extraction of the tumor region is done by implementing the K-Means Clustering Algorithm. Using K-Means, clusters of tissues of similar intensity is formed yielding four different brain clusters. The objective of K-means is to partition the data points into different clusters which do not overlap, where each point belongs to only one group. K-means gives best result when data set is different. From this, the tumor affected cluster can be analyzed for further processing.

Fig -4 below explains the data flow diagram for the proposed system. The dataset consisting of MRI images is split into training and testing dataset. The training dataset is made used in order to get a trained model. Firstly, the MRI image to be tested is taken and pre-processing is applied to discard any disturbances present in the image. Two filters are used in order to do so which are the Gaussian Filter and Mean Filter. The Preprocessed image is then subjected to decomposition by the Discrete Wavelet Transform.

The DWT process results into four sub-bands that describes the image along the vertical, horizontal, diagonal and approximation details. The approximation image data is the given as input to the GLCM algorithm which computes a matrix used for feature extraction. The GLCM feature extraction process aims to extricate essential textural attributes of the image. The feature vector containing the texture feature values is subjected to a trained network.

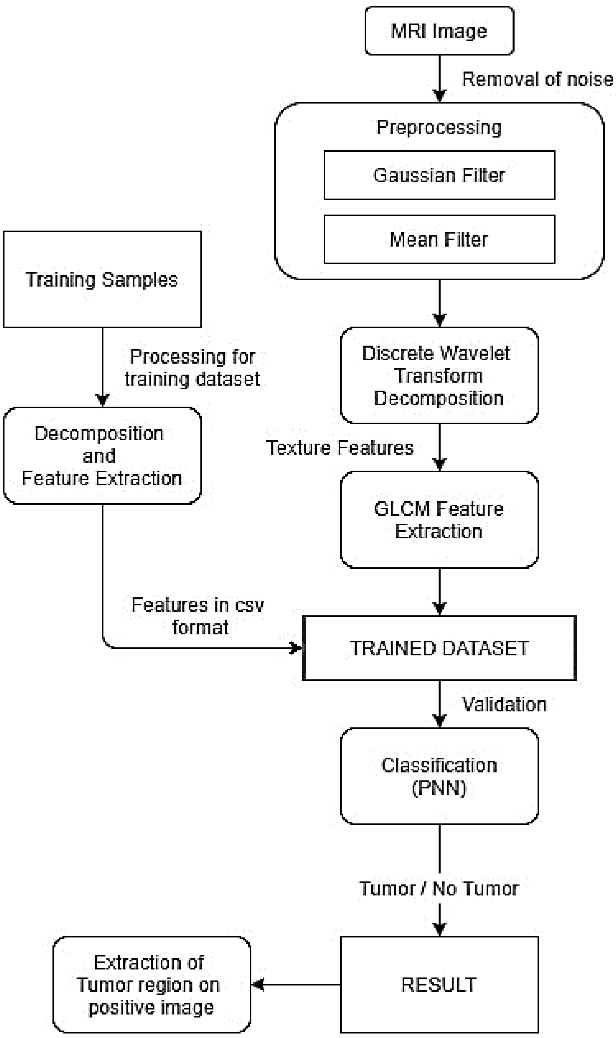


Fig -4: Data flow diagram

The trained network is a csv file containing the features vector values of the training sample. The training samples are similarly processed through preprocessing, decomposition and feature extraction yielding the csv file for training.

Probabilistic Neural Network (PNN) is executed in order to train the csv file consisting of the feature vector values. PNN is a classification algorithm that uses weights and network layers to train the dataset and validate the image. The result of the classification is a message depicting 'Tumor' for the presence of tumor in the MRI test image or 'No Tumor' for the absence of tumor in the image. The MRI image tested positive is finally given as input to a segmentation process that takes out the tumor affected section from the MRI image.

The K-Means Clustering algorithm is used for the extraction process. The removed tumor can hence be used for analysis.

4. RESULTS AND DISCUSSIONS

The following shows the step by step output of each process that lead to the accurate detection or identification of the brain tumor. The four main processes include the Pre-processing, Decomposition, Feature Extraction and Segmentation.

Fig -5 shows the original input MRI image selected for detection of brain tumor. The input MRI image is transformed into gray scale for processing.

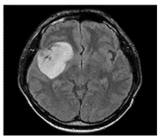


Fig -5: Original MRI Image

Fig -6 displays the snapshot of the preprocessing done by Gaussian Filter. This process removes any Gaussian noise and disturbances present in the original input image resulting into a better accuracy rate.

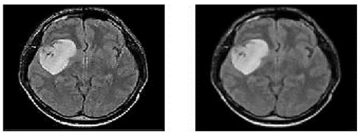


Fig -6: (a) Original MRI image (b) Gaussian filtered image

Fig -7 shows the output of the preprocessing done by the Mean Filter on the output of the Gaussian filtered image. Here, two level of preprocessing is carried out in order to remove any Poisson noise present in the image.

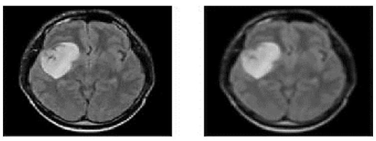


Fig -7: (a) Gaussian Filtered image (b) Mean filtered image

After preprocessing, a process of decomposition is done in order to get fine scale analysis of the image for feature extraction. Fig -8 and Fig -9 shows the sub bands obtained from level 1 and level 2 decomposition by DWT. Optimal texture attributes are then drawn out from the decomposed output image using the GLCM matrix. Some of the attributes extracted are contrast, correlation, homogeneity, energy, asm and dissimilarity.

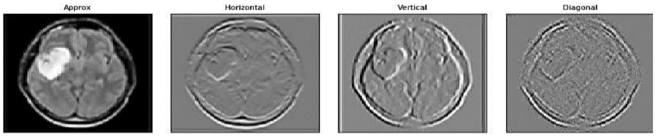


Fig -8: 1st level DWT Decomposition

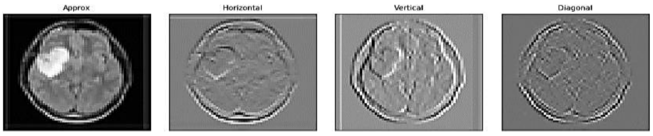


Fig -9: 2nd level DWT Decomposition

Fig -10 shows the segmented regions of the brain. This is executed using the k-means clustering initializing the value of k to 4

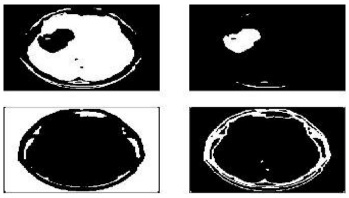
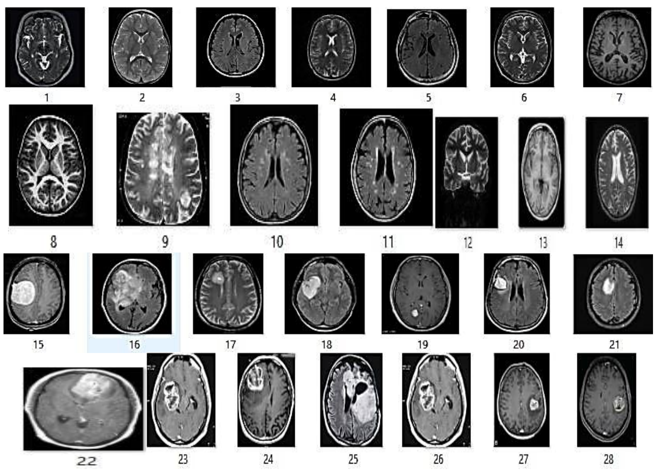


Fig -10: K-Means Clusters by Segmentation

Fig -11 displays the picture of some of the MRI images dataset used for the proposed system. The number under each image indicates the serial number of the MRI images. The dataset contains around 235 MRI images.

These images are divided into training and testing dataset for classification. The split ratio used was 0.1. The dataset is taken from the Kaggle repository. It containing the feature vectors of the MRI images are split in the ratio 0.1 with a random_state of 45. As a result, the number of training sets are 210 and the number of testing sets are 24. Having trained the model, we obtained an accuracy score of ***0.9166.***



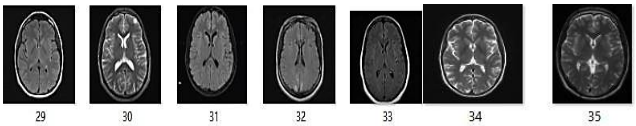


Fig -11: Dataset

5. CONCLUSIONS AND FUTURE WORK

The main objective of the system "Identification of Brain Tumor" is to identify and classify any input MRI scan into tumor affected and not tumor affected. Computerized and exact classification of MRI brain images is critical for clinical examination or analysis and understanding, helping medical professionals to identify and detect tumor in a faster and more efficient way. The system automates the brain image classification for early stage detection with the use of neural network classifier and extraction done by implementing the k-means clustering. Pre-processing, one of the most important module is implemented efficiently to produce accurate results for classification. It aims to remove any noise or disturbances present in the input Magnetic Resonance Imaging (MRI) image. Two-level decomposition is also implemented to get fine scale details for feature extraction. Texture features namely contrast, correlation, homogeneity, asm, dissimilarity and energy values are computed based on which the classification is done. A good number of MRI images are used for training the model by the Probabilistic Neural Network (PNN). The accuracy of the developed system is 91.6%. Testing on any individual image gave accurate results and a good response time.

In future, the tumor detected can further be processed for identifying the type of tumor present in the input image namely, benign or malignant. Having done the classification, the system can be further processed to give information on the direction of propagation of tumor. Direction of propagation can be an aspect that will help understand where the cancer cells will spread next. This can greatly help patients in better and faster recovery and help medical professionals provide better care. The area of the extracted tumor region can also be calculated that can help future analyses of the tumor present. The exact tumor location inside the brain can also be calculated for better diagnosis and treatment of the patient.

REFERENCES

[1] Nilesh Bhaskarrao Bahadure, Arun Kumar Ray, Har Pal Thethi "Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM", International Journal of Biomedical Imaging, Volume 2017, Article ID 9749108, March 2017.

[2] Santosh Virendra Vishwakarma, Balaji.G. Hogade "Image Segmentation for Detection of Benign and Malignant Tumors by Using DCT and DWT Combination", International Journal on Future Revolution in Computer Science & Communication Engineering, Volume: 4 Issue: 4, ISSN: 2454-4248, April 2018.

[3] Li Sun, Songtao Zhang, Hang Chen, Lin Luo "Brain Tumor Segmentation and Survival Prediction using Multimodal MRI Scans with Deep Learning", School of Innovation and Entrepreneurship, Southern University of Science and Technology, Shenzhen, China, College of Engineering, Peking University, Beijing, China, Volume 13, August 2019.

[4] N. Varuna Shree, T. N. R. Kumar "Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network", Brain Informatics, 5:23–30, January 2018.

[5] Parasuraman Kumar, B. VijayKumar "Brain Tumor MRI Segmentation and Classification Using Ensemble Classifier", International Journal of Recent Technology and Engineering, ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.

[6] Neha Rani, Sharda Vashisth "Brain Tumor Detection and Classification with Feed Forward Back-Prop Neural Network", International Journal of Computer Applications, Volume 146 – No.12, July 2016.

[7] Yousif Ahmed Hamad, Konstantin Vasilievich Simonov, Mohammad B. Naeem "Detection of Brain Tumor in MRI Images, Using a Combination of Fuzzy C-Means and Thresholding", International Journal of Advanced Pervasive and Ubiquitous Computing, Volume 11, Issue 1, March 2019.

[8] Md Shahariar Alam, Md Mahbubur Rahman, Mohammad Amazad Hossain, Md Khairul Islam, Kazi Mowdud Ahmed, Khandaker Takdir Ahmed, Bikash Chandra Singh, Md Sipon Miah "Automatic Human Brain Tumor Detection in MRI Image Using Template-Based K Means and Improved Fuzzy C Means Clustering Algorithm" Big data and cognitive computing, doi:10.3390/bdcc3020027, April 2019.

[9] Ahmed Kharrat, Mohamed Ben Messaoud, Nacéra Benamrane, Mohamed Abid "Detection of Brain Tumor in Medical Images", International Conference on Signals, Circuits and Systems, 978-1-4244-4398-7/09, March 2009.

[10] Swapnil R. Telrandhe, Amit Pimpalkar, Ankita Kendhe "Detection of Brain Tumor from MRI images by using Segmentation & SVM", World Conference on Futuristic Trends in Research and Innovation for Social Welfare, 978-1-4673-9214-3/16, March 2016.