Detection and classification of brain tumor using machine learning approaches

Pushpa B R*, Flemin Louies
Department of Computer Science, Amrita School of Arts and Sciences, Amrita Vishwa vidyapeetham, Mysuru, Karnataka, India

**Article History:**
Received on: 05.03.2019
Revised on: 20.06.2019
Accepted on: 25.06.2019

**Keywords:**
Masking, SVM, DWT, GLCM, Morphological operation

**ABSTRACT**
This paper proposes a method where a framework is constructed to detect and classify the tumor type. Over a period of years, many researchers have been researched and proposed methods in this space. We have proposed a method that is capable of analyzing the heterogeneous data and classifies tumor type. MRI images have been considered for this project since it gives the clear structure of the brain, without any surgery it scans and gives the structure of the brain this helps in further processing in the detection of the tumor. Human prediction in classifying the tumor from the MRI leads to misclassification. This motivates our project to construct the algorithm to predict the tumor. Machine learning plays a key role in predicting tumor. In this proposed paper, we have constructed a framework for detecting the brain tumor and classifying its type. The approach goes under pre-processing to filter and smooth the image. The segmentation is carried out by using morphological operation followed by masking, which increases the accuracy in the classification step. The multiple feature extraction methods are utilized to extract the feature from the masked image, and for classification, the kernel SVM is used.

*Corresponding Author
Name: Pushpa B R
Phone: +91-9738805472
Email: Preeths1@gmail.com

ISSN: 0975-7538
DOI: [https://doi.org/10.26452/ijrps.v10i3.1442](https://doi.org/10.26452/ijrps.v10i3.1442)

**INTRODUCTION**
The brain tumor is a massive or abnormal cell growth in a brain region. This can spread to any person among different ages. A brain tumor can be classified into different types; these tumors can either be Malignant (cancerous) or Benign (non-cancerous). The tumor that initially affects the brain are termed as the primary tumors whereas the tumors present in the other part of the body that might spread to the brain are known as the secondary tumors or metastatic tumors. It is estimated that in the United States in the year 2018 the statics states that approximately around 78,990 newer victims of primary tumors, (Benign)non-malignant tumors and other distinct tumors of the central nervous system are diagnosed. These include about 23,829 primary malignant tumors and 55,151 (Benign)non-malignant tumors. According to the analysis, India is the country which deals with more number of diagnoses as shown in Figure 1 (malignant) tumor is the cancerous one if it’s not spotted and treated in its premature grade; the outcome will be the death of a person. The observation from Human in predicting the tumor may mislead due to the noise and distortions found in the image. This motivates our work in constructing the algorithm to predict the tumor. This paper includes the method of detecting the tumor region and classifying them either normal, malignant or benign. The (MRI) image magnetic resonance image is an expertise technique that gives a rich standard image of the anthropoid. This
helps in medical recognition. The normal brain typically consists of three various tissues, namely white, grey matter and thirdly cerebrospinal fluid. The aim of tumor detection is to specify the tumor location, namely active tissue or necrotic tumor tissue. (Logeswari and Karnan, 2010). This is carried out by detecting the abnormal region comparing with the normal healthy tissues. Various methods are available in the field of medical image processing such as image enhancement, histogram, segmentation, morphological operation, feature extraction, and so on. Many segmentation and classification techniques have been used in this field to identify the tumor region. Automatic segmentation based on (CNN) Convolutional Neural Networks (Pereira et al., 2016). Fuzzy C and intelligent improvement mechanism, such like (GA) Genetic Algorithm and (PSO) Swarm Optimization approach to spot the tumor in terms of Pixel Error Rate and Position Error Rate (Gopal and Karnan, 2010). Many methods are applied on the images to perform the task such as pre-processing the given noisy image, image segmentation to segment the region, feature extraction and image classification to classify. Different data mining algorithms such as SVM, k-means, k-nearest neighbor, FCM, a neural network used for this purpose (Parmar and Gondaliya, 2018). Our proposed method first goes with the process filtering followed by segmentation and then classification.

This following session in this paper describes the methods and results of the existing research of the same domain, detection and classification of brain tumor (session II) it describes about the method proposed for the segmentation and classification (session III), and finally, it provides the results for the proposed method.

RELATED WORKS

Over the years, many specialists from different background worked and still working in the domain image processing dealing with the cancerous disease such as brain tumor, kidney tumor etc. They have used the number of methods to obtain the best results. Some of their works are described below.

Joseph et al. (2014) stated that Image processing segmentation plays a major role as it helps in the withdrawal of lesion affected regions in MRI images. Several mechanisms are adopted and applied to diagnosis the brain tumor. To segment the tumor region, the method adapts K-means cluster algorithmic rule along with morphological operation that removes the fault clustered area, and this may inevitably be shaped when segmentation. (Wu et al., 2007) Proposed the segmentation method that is applied on (RGB) colored image. They have used the K-means segmentation algorithm to route the affected region in the MRI image. The preliminary objective of this technique is to transform a grayscale image into an RGB image. Further, by applying the Histogram and K-Means clustering, the tumor will be extracted and separated from the base image. (Dahab et al., 2012) proposes a modified segmentation strategy, and this is applied on the MRI images to detect the defected part in the brain. PNN works accordingly to the LVQ technique on the images. This technique is used for the detection and classification of brain tumor using a neural network. The modification relies on (ROIs) regions of interest. Here, the features are withdrawn through feature extraction mechanism and then normalized that will help further in the classification of tumor type. (Mustaqeem et al., 2012) Proposed that the noise in a given MRI will be removed using Linear filters. Gaussian filters use gaussian laws to remove the false detection of tumor region in the images. It followed by applying the thresholding, watershed and Morphological algorithm to perform the segmentation. The results obtained is increased in the accuracy level when compared to other models. (Kumar et al., 2017) proposed an approach to detect brain tumor using MRI Images with the help of discrete wavelet transform (DWT) used to extract features and then applying Genetic algorithm (SVM) to classify the tumor type from the given MRI. (Parmar and Gondaliya, 2018) Stated that the diagnosis of the lesion cell region is complicated procedures in health care. The algorithm gives the steps for pre-processing, feature extraction, image segmentation, and image classification. Various algorithms are utilized like k-means, FCM, SVM, k-nearest neighbor, a neural network used for this purpose. (Sudharanani et al., 2015) Proposed methods like Re-sampling, Histogram, Distance Matrix, K-NN Algorithm. First, the graph specifies a full range of pixels which are divided among the specific image. The KNN method is used for identification and classification of the brain tumor, which is done by training of K. The Manhattan matrix has been utilized, and it calculates the distance for the proposed work. (Singh and Ansari, 2016a) have contrived a completely unique strategy for brain tumor identification, which involves histogram normalization and k-means segmentation. In this work, the input image MRI is pre-proposed and enhanced using filters like Median filter, Averaging filter, Adaptive filter, Gaussian channels and Un-sharp covering channel. The k-means segmentation method is used to segment these regions, which gives the specified number of clusters and using the SVM classifier, the tumor is classified into Benign or
malignant type.

From the above Literature Survey, it is seen that different methods have been proposed for Segmentation and Classification. According to the results provided, while comparing the classification of Benign and Malignant brain tumors, the rate of accuracy increases for Benign Classification whereas it decreases for Malignant Classification. In all the methods that are proposed, the SVM method can be termed the best for Classification. The data sets that are collected in the papers are mostly from Brats and TCIA.

**PROPOSED METHOD**

The brain tumor is cancerous or maybe non-cancerous mass or abnormal cell growth in the brain. Abnormal cell growth in the brain results in the brain tumor and affect a person’s life. The early and accurate detection of such disease can help the patient in medical healing. Imaging is an important side of bioscience is to picturize the diagnosed structures or shape of the human body, which helps in medical diagnosis. This project is divided into two main parts. The first part deals with the detection of the tumor from MRI images, and the second part contains the process of classification of tumor type (Benign, Malignant or Normal). The given input MRI image will undergo into the number of stages, which are pre-processing, segmentation and classification that contains the median filter, morphological operation, masking, feature extraction and SVM classification. The model that we have proposed is able to detect the affected region (tumor). The affected area will be separated using the morphological operation, which this operation separates the affected and normal region from the given MRI image. Figure 5 shows,

**Pre-processing**

The real-time images that have not processed may contain some noise or distortion, and hence, the image has to be enhanced before applying the algorithm to get better performance further (Rani and Vasudev, 2018). The goal of the pre-processing is to improve the image quality that helps for further processing. Pre-processing helps in reducing the complexity and increases accuracy. There are many methods out there in order to clear the noise and to obtain the smooth image we have considered the median filter and Gaussian high pass filter approach.

**Median filter**

The use of the Median Filter is to either denoise or extract the noise from a noisy image. One such example is salt and pepper image. It is a nonlinear filtering method that effectually reduces unwanted noise while preserving the boundary edges. (Apoorva et al., 2016). This filter operates on a single entry by moving one after another, by replacing the median of the window entry. If the entries are the odd number, it is simple to define the median. The median filter gives the best result in smoothing the image. (Sudharani et al., 2015) achieved the highest accuracy using the median filter. Figure 2 (a) shows,

\[
f(a, b) = \text{median}\{g(i, j)\}
\]

\[
(i, j) \in I_{a, b}
\]

**Gaussian high pass filter**

**Figure 1: Brain tumor Diagnosed Result 2019.**

**Figure 5: Proposed Algorithm**
Sharpening of the image results in giving better boundary edges. Using different high pass filters, we can easily achieve the sharper image. Once the image noise has been removed, sharpening the image is carried out in order to get the clear edges that help for post-processing. (Singh and Ansari, 2016b) Gaussian high pass filter is used for this process. Gaussian high pass filter gives a better accurate finer result while saving the right frequency and discarding the unwanted signals. Figure 2 (b) shows

**Morphological Operations**

We proposed a morphological operation to separate the tumor and the normal region. Morphological operations depend on the pixels value, not on the numerical value. It is used to detect the structure of the element (Havaei et al., 2017). The structure and size of the element will depend on the value zero or one that is represented in the matrix form. The binary or high contract pictures mostly have two types of
shading that are white and black. Commonly white represents the foreground image, and black represents the background of the image. Numerically it is denoted as ‘0’ represents black and ‘1 or 255’ used to represent the white region. Figure 3 and Figure 4 shows

Masking works by assigning some of the pixels in the image to zero (black background) while the rest (kernel) is non-zero (Prastawa, 2004). Once the morphology dilation has been successful, we have taken two input data (images) for masking. First, we have taken the image from the output image of the morphological operation. Secondly, the pre-processed MRI image considered and on successful subtraction from the first and second image, a masked image is obtained. Thus, the masked image further helps in classification and gives accurate accuracy. Figure 4 shows

Feature extraction
The feature extraction helps in decision making (Zulpe and Pawar, 2012). The proposed method has been implemented using multi-feature extraction, where multiple feature extraction methods are combined together and extracted. The features that are extracted are symmetrical, texture and grayscale. The classifier classifies on the basis of these features. After the masking process is done, the masked image further undergoes multiple extraction methods like DWT (discrete wavelet transform) techniques. Now the features which are important are withdrawn from the masked image, (Kumar et al., 2017). Then these features are combined together and normalized. DFT is used as a tool for analyzing the signal that decomposes it to a different frequency (time to frequency domain), but the drawback is it rejects the time information signal. This can be overcome by using DWT (Sudharrani et al., 2015). GLCM is one of the popular methods to extract the textural features. The gray level co-occurrence matrix is a technique used to withdraw the second-order features. In the GLCM matrix, the number of rows and the number of columns is equal to the number of gray level attributes. In our method, the features extracted from the GLCM are IDM, Entropy, correlation, Energy, Homogeneity, and Entropy (Minz and Mahobiya, 2017). Based on these multiple features, the classifier SVM classifies the type of tumor. Figure 9 shows,

Total of 13 features are extracted for our proposed method some of them are described below.

Mean: - The average value.

\[ x = \frac{1}{y} \sum_{j=1}^{y} a \]

Variance: - It is a total square distance distribution from the mean value, divided by distribution number.

\[ s = \frac{1}{y-1} \sum_{j=1}^{y} (y_i - x)^2 \]

Energy: - It is the equal measure between the pixels.

\[ E = \sum_{x, y} s(x, y)^2 \]
Correlation: - It is defined as the measurement of the relation among the window pixel.

\[ c = \sum_{a, b} \frac{(a, b)p(a, b) - \mu_i \mu_j}{\sigma_i \sigma_j} \]

Contrast: - The difference in the range of luminous is defined as a contrast.

\[ w = \sum_{a, b} |a - b|^2 s(a, b) \]

Homogeneity: - It is the rate of closeness of distributed elements in GLCM.

\[ H = \sum_{a, b} \frac{1}{1+(a-b)^2}p(a, b) \]

**PCA feature reduction**

Immoderate features increase in size (need for more memory), increases cost and also computational
Figure 10: Classification of Benign Tumor

Figure 11: Classification of Malignant Tumor
time. (Ratan et al., 2009). Sometimes, these features complicate the classification process, to overcome this problem, feature reduction is needed.

PCA is the tool which analyses the features and reduces the dimensionality of features. This is done by considering the most important variable first. The important variables are given the first priority. PCA orders and takes all the important feature and gives a new data set.

**Classification (SVM)**

Support vector machine

Introduction to support vector machine (SVM) is a benchmark in the area of machine learning. The supremacy of SVM is it gives better accuracy. (Zhang and Wu, 2012) SVM classifies into two classes based on machine learning. SVM classifies by raising a hyperplane between two classes. Tables 1 and 2 shows, this phases some problem when the different data types are placed at a different location to overcome this problem kernel technique is used (Corso et al., 2008). SVM will classify based on two groups a) linear, b) non-linear: Figures 6 and 7 shows,

**Linear**: - It is the uncomplicated one, the training sets are linearly separated. It is mathematically represented as below.

\[ f(k) = w^T k + b \]

The training samples are, the function keeps, \( f(k) \geq 0 \) for \( a_i = +1 \), and \( f(k) < 0 \) for \( a_i = -1 \), in a simple way, the test set of two different groups are separated by a hyperplane

\[ f(k) = w^T k + b = 0 \]

where \( w \) is weight, \( k \) is input test set, and \( b \) is the bias. \( f \)

The given training data set here it consists of two groups; the objective is to maximize the separating between the two groups + and − by a hyperplane.

**Non-linear**: - It is possible to easily separate the two classes in the linear separable just by distinguishing between the hyperplane (Machhale et al., 2015). But when training data are at different locations, the linear separable cannot be used. This can be overcome by non-linear separable such that different kernels are analysed. It will take the input and map into higher space. It is mathematically represented as below.

\[ f(k) = w^T \phi(k) + b \]

Such that \( a_i \) should be either +1 or -1, to optimize this problem we have to distinguish between two hyperplanes resulting into the prevention of data converging to the margin, this can be achieved by using mathematical formula as shown below.

\[ \text{minimum}_{w, b} \| w \| \]

\[ e, f, a_i(w, k_i) \geq 1, i = 1 \ldots I \]

Inactive condition, the \( \| w \| \) is changed to

\[ \text{minimum}_{w, b} \frac{1}{2} \| w \|^2 \]

\[ e, f, a_i(w, k_i) \geq 1, i = 1 \ldots I \]

**Data set**

Since our proposed method is dealing with classifying the tumor from a heterogeneous dataset. We have collected the data from two data warehouse (TCIA) and (BraTS). BraTS and TCIA datasets are considered as the standard dataset. And has been cited by many kinds of research working on the disease brain tumor. We have considered T1 and T2 MRI images from BraTS, since it gives better images then PET, a total of 290 HGG and 200 LGG images are collected and from TCIA around 1000 images were collected. From both the data warehouse, we have randomly selected the images for our proposed method. Figure 8 shows,

<table>
<thead>
<tr>
<th>Table 1: Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHODS</td>
</tr>
<tr>
<td>DWT+KNN</td>
</tr>
<tr>
<td>DWT+GLCM+KSVM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: SVM function classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
</tr>
<tr>
<td>Quadratic</td>
</tr>
<tr>
<td>RBF</td>
</tr>
</tbody>
</table>

**Experimental results**

We have taken an image of size (200 X 200), a total of 60 images were used for the training in which 20 are benign, 30 are malignant, and 10 are normal brain MRI images. Some features are shown in Figure 7. The experiment was carried out under the platform of Matlab 2015a with 4GB ram, 2.0 GHz processor with the operating system 8.1. The proposed algorithm performed well with time complexity. We could able to classify the tumor from heterogeneous data through our trained data set. Our proposed algorithm carried out in three stages. First pre-processing, Clearing the maximum noise out of the image followed by stage two segmenting the image to extract tumor and stage three classifying the tumor as benign, malignant or normal. For the proposed framework, we have achieved an accuracy of 99% for malignant and 99.6% for benign. We have also compared with another classifier shown in below table, and for the result, SVM gives the best accuracy. Figures 10 and 11 and Figure 12 shows,
Total number of Malignant images Classified = 100
Total number of false classification = 1
Total number of true classification = 99
Accuracy = 99/100 = 0.99 = 99%
Total number of Benign images Classified = 250
Total number of false classification = 1
Total number of true classification = 249
Accuracy = 249/250 = 0.996 = 99.6%

CONCLUSION

As per the literature survey, all the implemented model shows the good accuracy in classifying the benign tumor the accuracy drops while classifying the malignant tumor. In our proposed method, a better accuracy is obtained in classifying the malignant tumor (accuracy of 99%) comparing to the other existing system. The proposed method is limited that classifies the tumor as benign, malignant or normal, in future enhancement we are going to implement the algorithm that can also be able to classify the grade of the tumor (Garde I, Garde II, Garde IV or Garde V).

REFERENCES


Gopal, N. N., Karnan, M. 2010. Diagnose brain tumor through MRI using image processing clustering algorithms such as Fuzzy C Means along with intelligent optimization techniques. IEEE International Conference on Computational Intelligence and Computing Research, pages 1–4.


Singh, G., Ansari, M. A. 2016a. Efficient detection of brain tumor from MRIs using K-means segmentation and normalized histogram. 1st India International Conference on Information Processing


