Brain Tumor Classification using Conventional Machine Learning Algorithms

AHAMAD F.[1] ,SINGH K.[2], PANDEY P.[3], MISHRA A.R.[4], SEWAK A.[5]*

[1]School of Computer and System Sciences, Jawaharlal Nehru University
New Delhi, Delhi (India)-110067
[2,4,5]Department of Computer Science and Engineering, Rajkiya Engineering College
Sonbhadra, Uttar Pradesh-231206, India
[3]Department of Electronics Engineering, Rajkiya Engineering College
Sonbhadra, Uttar Pradesh-231206, India
(*) kalpanasingh30@gmail.com

Abstract

Brain tumors and associated nervous system cancers are one of the largest reasons behind the death of human beings. Timely identification of a tumor can help save a person's life. The detection of brain tumors is time-consuming and also needs a specialized radiologist for lesion detection. Machine Learning techniques can assist in the detection of brain tumors. These algorithms are critical in correctly predicting the presence of tumors in human beings. In this context, this paper uses the K-means algorithm to segment the brain MRI and then extracts features from these segmented MRIs for the purpose of detecting the brain tumor. The features are extracted by using the Gray Level Co-occurrence Matrix (GLCM). The features extracted are fed into the classifiers Support Vector Machine (SVM) and Decision Tree (DT) to segregate the tumorous and non-tumorous MRIs. Our proposed approach performs better than state-of-the-art methods in terms of classification accuracy.

Keywords*brain tumor, gray level co-occurrence matrix, K-means, machine learning, magnetic resonance imaging*

DOI Number: 1010.14704/nq.2021.19.3.NQ21031 **NeuroQuantology2021;19(3): 78-86**

1. Introduction

Standard Magnetic Resonance Imaging (MRI) for the purpose of detecting brain tumors is time-consuming, and analyzing these images by radiologists is prone to error in most circumstances, making automated image classification essential [1]. Analyzing the images in general and medical scans, in particular, is a procedure of many steps which mainly include (a) gathering raw or processed data (b) applying pre-processing techniques to reduce bias (c) data processing, and then classifying the data, and (d) visualizing the results. Several systems for the prediction of brain tumors have been proposed, but none have been tested in the field. This is mainly due to the issues of generalizability, or an inability of machine learning algorithms to understand or explain the eventual result. However, machine learning algorithms can be used to overcome the issues of prognosis of brain tumors timely and thus can assist the radiologist in providing an extra opinion in predicting tumors.

For the examination and therapy of an intracranial tumor, a range of Image-Processingmethods and techniques preferably have been used. The affected patch of brain tissue is extracted from MRIs using the method of segmentation and classification, which is a crucial step in the computer vision technique. The task of segmenting the affected tumor region is critical for cancer diagnosis. For tumor segmentation and classification, a variety of semi-automatic and automatic methods and methodologies, and approaches are used [2]. Image textures, local histograms, etc are all aspects of MRIsthat have been used in brain tumor segmentation analysis and classification. In tumor segmentation analyses, Machine-Learning(ML) approaches such as "Support Vector Machines (SVMs)" are often utilized for pattern identification. "Deep-Learning" based approaches and methodologies are rapidly gaining traction in brain tumor segmentation research owing to their high performance in image processing fields such as object recognition, image semantic segmentation, and classification [3]. Most of these techniques need a lot of data to generalize well and thus are practically not applicable in the case of medical diagnosis. Also, medical data is imbalanced and thus machine learning and deep learning models are biased towards the classes with more frequent data. Looking at these issues this paper proposes a lightweight machine learning model to predict brain tumors from a small dataset of 253 MRI images. The main contribution of this paper is as follows:

a) The dataset is segmented by the K-means algorithm
b) In the second step gray-level co-occurrence matrix is accustomed to extracting second order attributes from the segmented images.
c) The features extracted are used for classification. The popular classifying techniques employed are Support Vector Machines (SVM)andDecision Tree (DT).
d) Classification accuracy is used for evaluating the classification model.

The following sections or parts of this paper are defined as a) In section 2 we review the literature in the field and in section 3 the methodology is presented. In section 4 the classification algorithms are explained following which the experimental results are presented.

2. Background Study

In paper [1], Data, Image segmentation processing, and View DIV is used to diagnose brain tumors by employing deep learning methods. In paper [5], the modified level set method segments MRI images, and then GLCM along with Gabor and moment invariant features are used to classify MRI images from the dataset. In [6], the authors have used texture information along with the structural information of different brain image slice sequences to predict tumors. the segmentation process is carried out with the help of the "global thresholding method”. Then the extracted region of the tumor is fed into the classification model. The classification model employed in this research is convolutional neural networks (CNNs) which classify tumorous and non-tumorous MRIs. The authors in the paper [7], have segmented the MRI images by using the Fuzzy-Thresholding method. The method extracts the tumor region from the images. Before the segmentation saliency model is employed to detect the area which is pathologically significant. The method has shown good results after being compared with "adaptively regularized kernel-based fuzzy C- Means clustering, Mean Shift and Fuzzy C-Means clustering with Level Set Method". The authors of the paper [8] have used a non-parametric algorithm known as K-nearest Neighbor (k-NN)for classification. The algorithm makes use of the distance function based on Euclidean distance and the number of neighbor points decides the classification which is determined by the value of ‘k’. The algorithm calculates the distance, finds the nearest neighbor, and then votes for the class [9]. In the study carried out in [10], the authors have proposed a Bayesian model for the segmentation of images. This unsupervised segmentation of images makes use of spatially varying finite mixture models and the non-local means framework. In [11], the authors have used Gray level co occurrence matrix for feature extraction and before that MRI scans were segmented by Kernel based Fuzzy C means clustering algorithm. They were able to obtain an accuracy of 91.4% with the classification. In [12] authors have made use of CNN classifier along with the features extracted from GLCM features. The authors in [13], used CNN for extracting features and Kernel ELM was used as a classifier to classify the MRI images.

3. Methodology

The methodology of this work is presented in four steps. Firstly, the MRI images are segmented by employing the K-means algorithm and in the next step, the GLCM matrix is computed from the segmented images. This GLCM matrix is basically used to extract second-order attributes which are ultimately used for the classification process. The classification of the MRI slices is accomplished with the help of SVM and DTs. Figure 3. shows the proposed algorithm for the categorization of brain tumors.

3.1 Dataset Description

The dataset for this work was downloaded from Kaggle [3]. This dataset contains a total of two hundred and fifty-three (253) MRI slices out of which one hundred fifty-five slices are tumorous and ninety-eight are non-tumorous. Figure 1. shows the sample slices from the dataset.

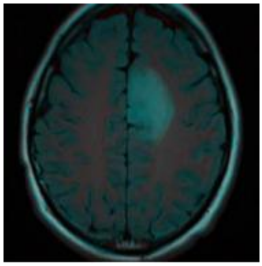
(a) Shows the nontumorous MRI slices from the dataset

(b)Shows the tumorous MRI slices from the dataset
Figure 1. shows MRI samples from the dataset

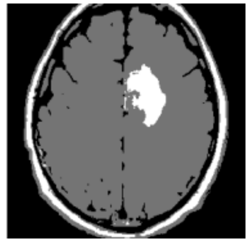
3.2 Segmentation

Image segmentation is a procedure that divides an image into distinct parts using criteria such as pixel threshold values. For biomedical image segmentation, researchers use a variety of methods such as thresholding,
clustering, and others [4]. The K-Means Clustering based algorithm technique is popularly used to segment MRI images in our work. The K-means clustering method divides a data set into K unique, non-overlapping groups in a straightforward and elegant way. To use K-means clustering, we first define the desired number of clusters, K, which was set as two in our experiment, and then the K-means algorithm allocates each observation to one of the K clusters. Figure 2. shows the MRI slices after the k-means segmentation algorithm.

eISSN 1303-5150



a



c

Figure 2 (a) shows original MRI image and (b) shows the K means Segmented image

3.3 Feature Extraction

Feature extraction is one of the most important stages of our proposed method to analyze the MRI data. We utilize the Gray-level Co-occurrencematrix for the purpose of extracting second-order texture features from the segmented MRI slices.

Gray-level Co-occurrence Matrix

A basic matrix of co-occurring pixel values, also known as a Gray-Level Co-Occurrence Matrix (GLCM), the dispersal of co-occurring image pixels (grayscale values, or colors) at a given offset. It's a texture analysis method that's employed in a huge number of applications, including medical picture analysis. In our experiment, we have extracted second-order texture values from the GLCM matrix computed on segmented MRI images. A total of four-second order features viz. entropy, homogeneity, correlation, and energy are extracted. These features are defined as follows.

a) $Energy = \sum_{i,j}^{N} P(i,j)^2$ (1)

b) $Homogeniety = \sum_i \sum_j \frac{P_{ij}}{1+|i-j|}$ (2)

c) $Correlation = \sum_{i,j}^{N} \frac{(i-\mu_i)((j-\mu_j)*P(i,j)}{\sigma_i \sigma_j}$ (3)

d) $Energy = \sum_{i,j}^{N} P(i,j)^2$ (4)

4. Classification

The features extracted by using the GLCM matrix are fed into two classification algorithms SVM and DT. A Guided Common Learning Technique, SVM, may be used to solve both groupings as well as regression tasks. However, it is for the most part utilized in ML, especially for Categorization purposes. The leading purpose of the SVM algorithm is to find the decision boundary or optimum line for categorizing n-dimensional space into classes. The main purpose behind this is to readily replace the additional data points in the proper

eISSN 1303-5150

classification in the future. A hyperplane may use as an optimal choice boundary. By undertaking a greedy search to determine the best split points inside a tree, DT learning uses a divide and conquer technique. This dividing procedure is then repeated in a top-down, recursive fashion until all, or the majority, of the entries, have been categorized under specified class labels.

5. Experimental Results

The features extracted through GLCM are fed into the classifiers SVM and DT. The parameter accuracy, Precision, Recall and F1-Score given from eq. 5-8, are used to measure the performance of the algorithm. Precision answers what percentage of positive identifications were correct. Whereas Recall defines the proportion of actually positive cases identified correctly.To completely assess a model's efficacy, we must look at both precision and recall. Regrettably, accuracy and recall are sometimes at odds. In other words, increasing accuracy usually decreases recall, and vice versa. In this case a third parameter call fi- score is used. Figure 3. shows the confusion matrix for both of these classifiers. The experimental results for accuracy, precision, recall and F1-score are mentioned in table 1.

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN} \qquad (5)$$

$$Precision = \frac{TP}{(TP + FP)} \qquad (6)$$

$$Recall \frac{TP}{(TP + FN)} \qquad (7)$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (8)$$

	0	1
0	136	19
1	9	89

a

	0	1
0	143	12
1	7	90

b

Figure 3. shows the confusion matrix a) Support Vector Machine (SVM) and b) Decision Tree (DT)

Table 1. Shows the accuracy, recall, precision and F1-score of the classifiers

Support Vector Machine				Decision Tree			
Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
88.93	**93.79**	**90.06**	**91.88**	**92.09**	**95.33**	**92.25**	**93.76**

From the confusion matrix it is evident that SVM classifier was able to classify 225 MRI images out of 253 and only 28 MRIs were wrongly classified. Whereas DT classifier was able to classify a total of 233 MRI out of 253. In terms of accuracy the DT classifier beats state-of-art models. The comparisons with the state of art models is depicted in Figure 5.

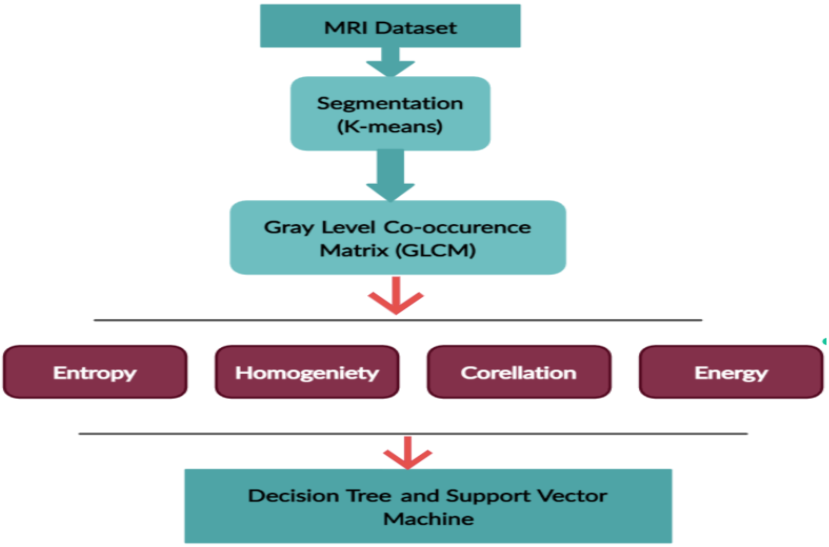


Figure 4. shows the architecture of the proposed algorithm

Results specifically accuracy of SVM and DT is 88.93% and 92.09% respectively. It can be directly inferenced from the confusion matrix in figure 3. Thus, the proposed algorithm performs very well and has a very less computation overhead. The algorithm can be used in clinical applications due to its simplicity, less computation overhead, and high classification accuracy.

6. Comparison with state-of-art methods

In this section, the proposed method is compared with the state-art-methods. It can be seen that the proposed method which uses K-means segmented MRI images with DT classifier performs better in terms of accuracy metric. The figure 5. Shows such comparisons with the already existing models in the field.

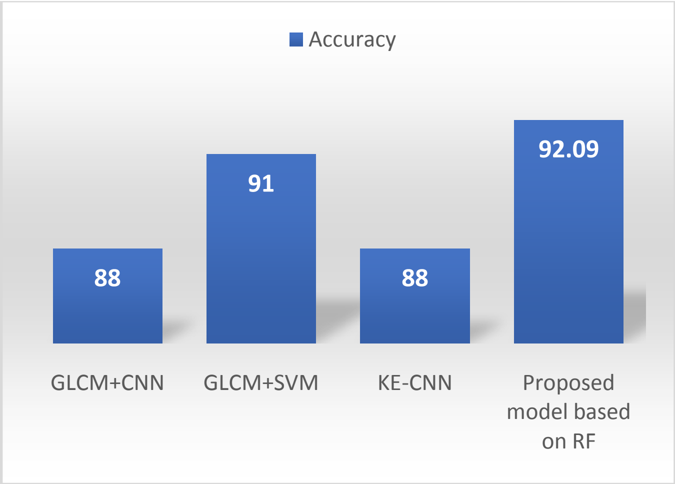


Figure 5. Comparison of the proposed method with the state of art methods in terms of classification accuracy.

7. Conclusion

One of the deadliest illnesses popularly known as brain or intracranial tumor, and it must be discovered as soon as possible. Specialized doctors are uncommon in developing nations like India, and brain tumor prognosis is time intensive, necessitating the use of skilled radiologists. We provide a second opinion for clinicians to consider when predicting tumors in this study report. The suggested technique is based on K-means segmentation of brain MRIs and feature extraction using the GLCM matrix. On binary classification, the method achieves a score of 92.09 percent accuracy. Using the suggested strategy in conjunction with other feature extraction techniques, the work may be extended in the future.

References

[1] Sharma, K., Kaur, A., & Gujral, S. (2014). A review of various brain tumor detection techniques in brain MRI images. IOSR Journal of Engineering (IOSRJEN), 4(05), 06-12.

[2] Devunooru, S., Alsadoon, A., Chandana, P. W. C., & Beg, A. (2021). Deep learning neural networks for medical image segmentation of brain tumors for diagnosis: a recent review and taxonomy. Journal of Ambient Intelligence and Humanized Computing, 12(1), 455-483.

[3] Nadeem, M. W., Ghamdi, M. A. A., Hussain, M., Khan, M. A., Khan, K. M., Almotiri, S. H., & Butt, S. A. (2020). Brain tumor analysis empowered with deep learning: A review, taxonomy, and future challenges. Brain sciences, 10(2), 118.

[4] Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., ... &Jambawalikar, S. R. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629.

[5] Amarapur, B. (2020). Computer-aided diagnosis applied to MRI images of brain tumor using cognition-based modified level set and optimized ANN classifier. Multimedia Tools and Applications, 79(5), 3571-3599.

[6] Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network

[7] Amin, J., Sharif, M., Gul, N., Yasmin, M., & Shad, S. A. (2020). Brain tumor classification is based on

DWT fusion of MRI sequences using a convolutional neural network. Pattern Recognition Letters, 129, 115-122.

[8] Yuan Z., Wang Y. (2012). Research on K Nearest Neighbor Non-parametric Regression Algorithm Based on KD-Tree and Clustering Analysis, Fourth International Conference on Computational and Information Sciences, 12997114

[9] Razzak, M. I., Naz, S., &Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps, 323-350.

[10] Singh, D., & Kaur, K. (2012). Classification of abnormalities in brain MRI images using GLCM, PCA, and SVM. International Journal of Engineering and Advanced Technology (IJEAT), 1(6), 243-248.

[11] Widhiarso, W., Yohannes, Y., &Prakarsah, C. (2018). Brain tumor classification using gray level co-occurrence matrix and convolutional neural network. *IJEIS (Indonesian Journal of Electronics and Instrumentation Systems)*, *8*(2), 179-190.

[12] Pashaei, A., Sajedi, H., &Jazayeri, N. (2018, October). Brain tumor classification via convolutional neural network and extreme learning machines. In *2018 8th International conference on computer and knowledge engineering (ICCKE)* (pp. 314-319). IEEE.

[13] Widhiarso, W., Yohannes, Y., &Prakarsah, C. (2018). Brain tumor classification using gray level co-occurrence matrix and convolutional neural network. *IJEIS (Indonesian Journal of Electronics and Instrumentation Systems)*, *8*(2), 179-190.