

Brain Tumor Detection by mining functional MRI Images

Neeru Kapur and Natasha Nargotra

Shree Guru Gobind Singh Tricentary University, Gurugram, Haryana, India

Corresponding author email: Neeru_fmhs@sgtuniversity.org

ABSTRACT

Medical image analysis is the most emerging and challenging field nowadays, among which brain tumor detection and classification is the most horrible and devastating type of cancer. Brain tumor classification helps in diagnosing the tumor at an early stage. Though numerous tumor detection methods and classification have been proposed, enhanced tumor detection and classification is still a challenging task because brain tumor images possess high diversity in boundaries as well as tumor appearance. This paper proposes a quick method for extracting metabolite values from graphs. Brain tumors are detected using metabolites such as N-acetylaspartate (NAA), choline, and Creatine (Cr₂). The tumor type is decided by taking Choline/N-acetylaspartate ratio. During the clustering process, each metabolite is assigned a weight. This clustering strategy has a precision of up to 88 percent. The suggested method is built on decision tree algorithms, which outperform clustering algorithms. Instead of storing fMRI (functional MRI) images the proposed theory stores values of metabolites in a dataset which improves the image processing tasks as well as memory requirements.

KEY WORDS: CHOLINE, CREATINE (CR), CREATINE (CR₂), FUNCTIONAL MAGNETIC RESONANCE IMAGING (FMRI), N-ACETYLASPARTATE (NAA).

INTRODUCTION

Functional magnetic resonance imaging (fMRI) detects changes in blood flow to assess brain function. This method is about the fact that neuronal activity as well as cerebral blood flow are inextricably linked. When a part of the brain is used, blood supply to that part of the brain improves. The detailed information related to brain tumor cellular structure, anatomy is given by Functional Magnetic Resonance Imaging (fMRI) which makes it an important tool for diagnosis. fMRI is often used in research work, and it also has some applications in medical field. It is utilized in other procedures of brain physiology like NIRS and EEG.

Innovative methods which largely use biomarkers rather than the blood-oxygen-level-dependent (BOLD) signal improves both spatial and time resolution. The devices such as lie detectors which works on the principle of fMRI techniques

are developed commercially by some of the companies, but the research is not efficient enough for the widespread commercialization. There are four types of metabolites used in this technique namely Creatine (Cr), Choline (Cho), N-acetylaspartate (NAA) (V. Megalooikonomou et. al. (2000)). There is a massive amount of data contained in hospital databases of brain tumor patients. These data are presented in the form of fMRI images (N. Lachiche et. al. (2005)). This dataset is being mined in order to improve understanding. Metabolite values are derived from fMRI images using the proposed technique. After that, the extracted values are saved in a dataset (N. Sivaram et. al. (2010)). When compared to fMRI files, the brain tumor dataset needs less memory. The produced dataset is used for various clustering and classification procedures.

Research Questions

1. How Brain Tumor is Detected Using Mining fMRI Images?

Literature Survey: "Data mining in brain imaging" research paper by Vasileios Megalooikonomou et al. presents data mining techniques that are deployed in the analysis of brain tumor images. The structural and functional are two types of brain imaging data which are considered in this method. This paper introduces statistical methods that help in the discovery

Biosc Biotech Res Comm P-ISSN: 0974-6455 E-ISSN: 2321-4007



Identifiers and Pagination

Year: 2021 Vol: 14 No (6) Special Issue

Pages: 375-379

This is an open access article under Creative

Commons License Attribution 4.0 Intl (CC-BY).

DOI: <http://dx.doi.org/10.21786/bbrc/14.7.79>

Article Information

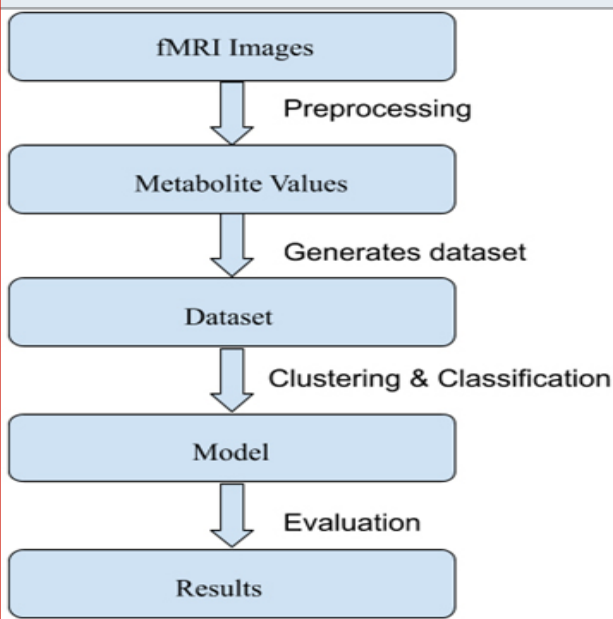
Received: 11th May 2021

Accepted after revision: 18th July 2021

of important patterns as well as association between clinical data and brain images. The various examples of applications to the real brain data is discussed. The various data mining issues, including method verification or validation is also discussed. “Neuronal Clustering of Brain fMRI Images” research paper by Nicolas Lachiche et al. shows a function of neuronal clustering technique in which it highlights active areas due to a relevant distance between the fMRI image sequences. This technique has been integrated with an interactive environment for enhancement in brain fMRI. The results obtained from a dataset validate the selected approach and open window for the further developments in this field.

“Applicability of Clustering and Classification Algorithms for Recruitment Data Mining” research paper by N. Sivaram et al. presents the knowledge acquisition techniques to extract domain knowledge. “The Evaluation of Preprocessing Choices in Single-Subject BOLD fMRI Using NPAIRS Performance Metrics” research paper by Stephen LaConte et al. suggest a new approach for simulation-based receiver operational characteristic analysis for evaluating fMRI data analysis methodologies. It employs the rapidly evolving nonparametric prediction, impact, activation, and reproducibility resampling framework to create a cross-validation-based model.

Figure 1: Block diagram of proposed system.



Geoffrey I. Webb presents a new grafting algorithm that considers a set of data on behalf of each leaf of the initial decision tree. This proposed new technique retains the error reduction power of the original grafting algorithm. It reduces computing time and the complexity of the inferred tree. The above mentioned studies on detecting brain tumor using various techniques can achieve accuracy lying between 49-85% which is not sufficient enough to be implemented in medical applications. So, the present study proposes a supervised learning algorithms that are more reliable and have higher accuracy as compared to the unsupervised

learning algorithms to detect brain tumor.

METHODOLOGY

Functional MRI (f-MRI) images are given as an input in this proposed system. Following that, values are derived from the fMRI graph and stored in a dataset. The dataset is then used in grouping and clustering. The full working of the proposed system is shown with the help of block diagram in below (Figure 1).

1. Preprocessing: The metabolite values present in the fMRI images are used to detect the type of brain tumor. The graph scanning procedure is utilized to find values considering the graph (V. Della-Maggiore (2002), S. LaConte (2003)). Graph scanning algorithm implements following steps:

The algorithm for graph scanning detects the peak points of graphs as well as records their values of coordinate. The distance between X-axis and the highest point is determined. Measured distance is later multiplied by the appropriate consumer scale (s) to obtain a metabolite value that is very similar to the original value. Figure 2 shows how the extracted values are saved in an excel sheet.

- Locate the X-axis of the MRI image (X-axis is the same for all MRI images).
- Metabolite [0] = Location of 2 on X-axis since NAA metabolite always occurs at 2 PPM.
- Metabolite [1] = Location of 3 on X-axis since Cr metabolite always occurs at 3 PPM.
- Metabolite [2] = Location of peak next to Cr.
- Metabolite [3] = Location of peak less than and near to 4PPM.
- Take suitable scale value (s) from the user.
- For all m in Metabolite
 - a) Scan along the Y-axis until peak point is detected on the MRI graph line.
 - b) Compute distance between the peak point and X axis.
 - c) $Temp = distance / 115$.
 - d) $Result [m] = Temp \times s$.
- End.
- Store the Result array.
- End.

Figure 2: Screenshot of data stored in Excel sheet.

	A	B	C	D	E	F
1	CHO/NAA	CHO	NAA	Cr2	Cr	Results
2	0.52	11	18.7	4.62	11.22	Benign
3	0.68	13.59	22.71	7.58	13.17	Mild
4	1017	2.9	4.37	3.64	2.77	Infection
5	1.42	15	8.84	5.45	6.48	Malignant
6	0.75	8.02	12.82	6.14	10.92	Mild
7	0.72	10.32	11.97	6.32	8.32	Mild
8						

2. Clustering and Classification Algorithms: The following algorithm has been considered for identifying brain tumors:

2.1. K-means using algorithm: Order of instances and initial centroid selection are the two factors on which

Accuracy of the K-means algorithm depends. Z-score ranking method is used to overcome this drawback to reorder the instances as well as in selecting initial centroids. The same algorithm is insufficient for detecting brain tumors since basic K-means and Z-score provide equal value to whole attributes. As a result, each metabolite is assigned a weight with the given equation 1.

$$W = \frac{1}{\sigma^2} \tag{1}$$

Where,

σ is standard deviation of attribute,
 W is attribute's weight.

Standard deviation is measured using equation 2:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \tag{2}$$

Mean value (μ) is measured using equation 3:

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \tag{3}$$

For Z-score calculation measured weights are used to determine each attribute's value. Each attribute is given an arbitrary weight based on the requirements. The following formula provides the mathematical Z-score rating formula as well as weight:

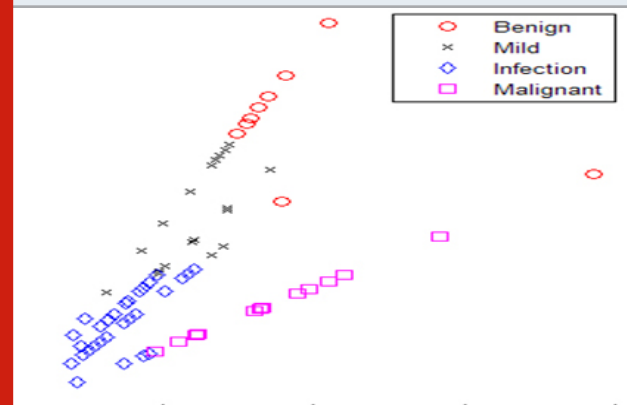
$$Z_{score} = \frac{(x_i - \mu) * W_i}{\sigma} \tag{4}$$

The data is rated using the above Z-score formula based on the value assigned to the listed attribute. The proposed algorithm combines Z-score ranking system along with weighted K-means. The gap formula as well as Z-score formula have been redefined as:

$$d = \sqrt{(x1 - x2)^2 * W_x + (y1 - y2)^2 * W_y} \tag{5}$$

The proposed algorithm i.e. simple K-means as well as weighted K-means is shown in Figure 3 and Figure 4 respectively.

Figure 4: Weighted K means



2.2. Functional Trees Algorithm: Functional tree algorithm works in coordination with both nominal attributes as well as integer that have values which are actually missing (J. Gama et. al. (2004)). In the current analysis, this algorithm was used to investigate the consequences of using combined attributes at decision nodes, leaf nodes, or both. A basic architecture for multi-vibrate tree learning is implemented for using functional nodes in different locations, which blends decision tree with linear structure using constructive induction. This system yields the decision tree, which is capable of employing decision nodes as well as multi-vibrate checks, as well as leaf nodes that render projections using linear functions.

2.3. J48graft Algorithm: The grafted C4.5 decision trees are generated in the J48graft algorithm. The advantage of tree grafting technique is that it increases predictive accuracy of a classifier (G. I. Webb et. al. (1999)). In this research this algorithm has been utilized in decision trees. It reclassifies regions where there is miss-classified data or no training data, which results in giving less prediction error. It determines the better cuts of current leaf regions and then branches out to create new leaves along with classifications which may vary from the first. During this method, a more complex tree is formed; however, only branching omits classification errors inside the data which has been correctly categorized previously, ensuring that the new tree eliminates errors even more.

RESULTS AND DISCUSSION

All of the algorithms described in this paper are applied to brain tumor datasets created by the graph scanning process. The dataset contains 80 cases, 22 of which are benign, 23 of which are mild, 20 of which are malignant, and 15 of which are infectious. The efficiency of weighted K-means using algorithm of Z-score rating is compared to

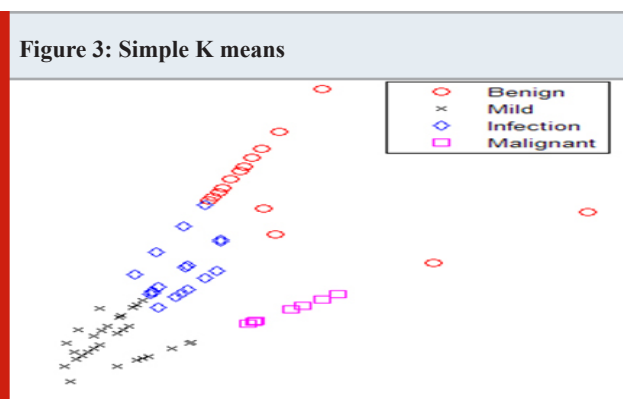


Figure 3: Simple K means

weighted basic K-means, K-means, as well as K-means having Z-score. It was found that weighted K-means having Z-score outperforms every other clustering algorithm. Figure 6 shows the clustering graphs of several other forms of K-means. X-axis represents Cho, while Y-axis represents NAA. The clusters produced by proposed algorithm outperform others. Using Cho Vs NAA various categories of K-means are plotted and is shown in the below Figure 5 and 6.

Figure 5: K means with Z score.

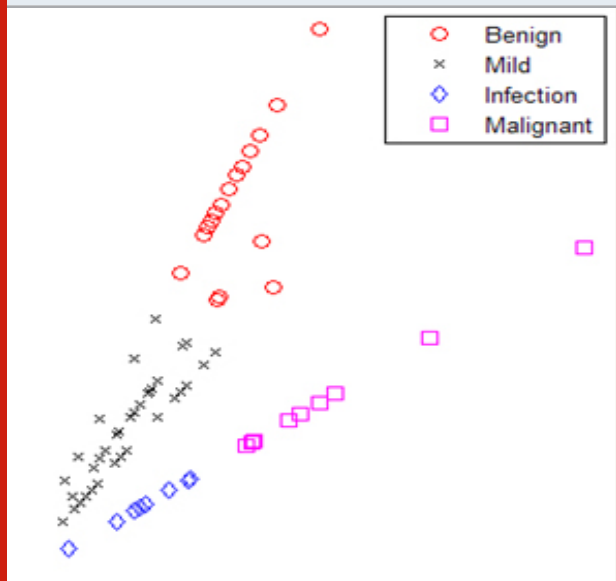
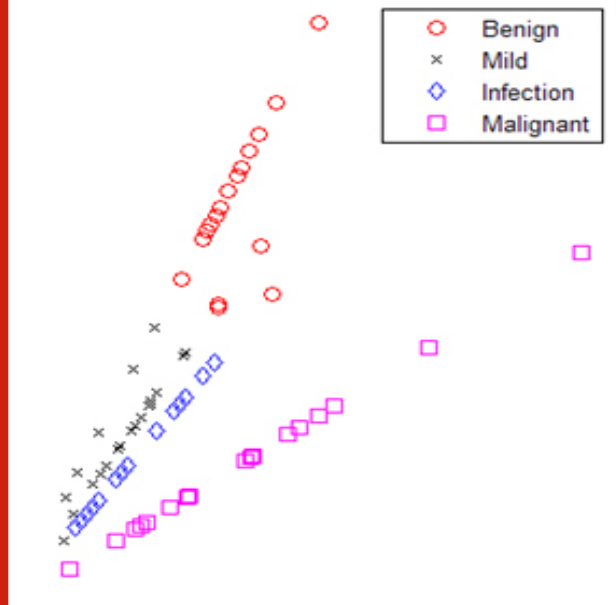


Figure 6: The Layout of Weighted K-means having Z score.



The consistency of all algorithms is measured in relation to evolving data instances. As seen in the graph below, the proposed weighted K-means with Z-score outperforms

Figure 7: The Accuracy of weighted K-means, simple K-means, K-means having Z-Scores well as weighted K-means having Z-score correspondingly.

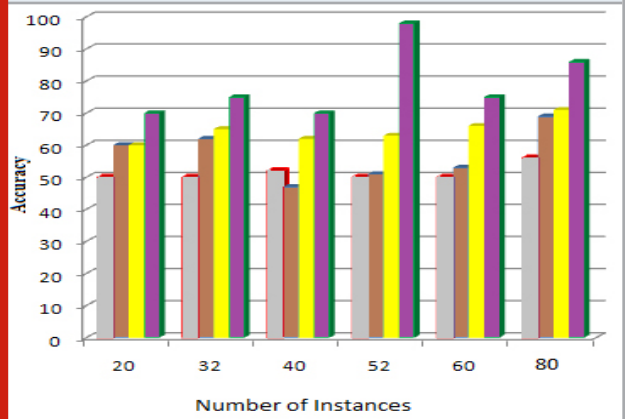
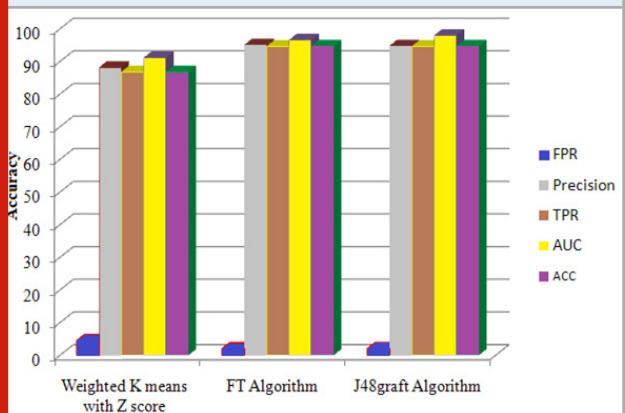


Figure 8: Evaluation of numerous algorithms.



the other strategies. The weighted K-means having Z-score algorithm's precision is equivalent to that of other algorithms. Figure 8 depicts a graph for Precision, True Positive Rate, False Positive Rate, Area Under Curve and Accuracy for 80 cases.

Table 1. Comparison of algorithms on the basis of metric evaluation.

Algorithm	FPR	Precision	TPR	AUC	ACC
Weighted K-means with Z-score	0.05	0.892	0.879	0.922	0.86.85
FT	0.02	0.961	0.957	0.976	94.74
J48graft	0.02	0.958	0.957	0.99	94.74

The J48graft as well as FT algorithms have up to 95.72 percent accuracy. The above algorithms can be subjected to a variety of success assessment metrics. Weighted K-means and Z-score ranking have drawbacks as opposed to classification algorithms in that they have lower Precision,

AUC, TPR, ACC, and higher FPR. Despite the fact that J48graft and FT have the same FPR, TPR as well as ACC, J48graft is considered the better algorithm due to its higher AUC. The different measurement metrics are compared and illustrated in the table below (Table 1).

CONCLUSION

The most severe and violent brain tumor is Glioma that leads to almost 50% patients' death within two years. The comparison of performance of various clustering as well as classification algorithms is done in this paper. The identification of brain tumors using various variants of K-means will achieve precision ranging from 49 to 85 percent. This level of precision is insufficient for use in medical applications. Decision tree algorithms are much superior to other algorithms in terms of performance and reliability for medical applications. As a result, this paper concludes that supervised learning algorithms are more accurate and have better precision in detecting brain tumors than unsupervised learning algorithms. The algorithm performed in this paper has achieved 95.82% accuracy and 0.97 AUC with the help of J48graft algorithm.

REFERENCES

Della-Maggiore, V., Chau, W., Peres-Neto, P.R. and McIntosh, A.R., 2002. An empirical comparison of SPM preprocessing parameters to the analysis of fMRI data.

Neuroimage, 17(1), pp.19-28.

Ericson, K. and Pallickara, S., 2013. On the performance of high dimensional data clustering and classification algorithms. *Future Generation Computer Systems*, 29(4), pp.1024-1034.

Gama, J., 2004. Functional trees. *Machine learning*, 55(3), pp.219-250.

N. Lachiche, J. Kommet, J. Korczak, and A. Braud, "Neuronal clustering of brain fMRI images," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2005.

LaConte, S., Anderson, J., Muley, S., Ashe, J., Frutiger, S., Rehm, K., Hansen, L.K., Yacoub, E., Hu, X., Rottenberg, D. and Strother, S., 2003. The evaluation of preprocessing choices in single-subject BOLD fMRI using NPAIRS performance metrics. *NeuroImage*, 18(1), pp.10-27.

Megalooikonomou, V., Ford, J., Shen, L., Makedon, F. and Saykin, A., 2000. Data mining in brain imaging. *Statistical Methods in Medical Research*, 9(4), pp.359-394.

Sivaram, N. and Ramar, K., 2010. Applicability of clustering and classification algorithms for recruitment data mining. *International Journal of Computer Applications*, 4(5), pp.23-28.

Webb, G.I., 1999, July. Decision tree grafting from the all-tests-but-one partition. In *Ijcai (Vol. 2, pp. 702-707)*.