Brain Tumour Disease Pattern Identification from Metabolites in Magnetic Resonance Spectroscopy Graph using Data Mining Techniques

Meghana Nagori
Computer Science and Engineering Department
Government Engineering College
Railway Station Road,Aurangabad,Maharashtra, India

Madhuri S. Joshi, PhD
Computer Science and Engineering Department
Jawaharlal Nehru Engineering College
Cidco,Aurangabad,Maharashtra,India

ABSTRACT
One of the significant applications of image classification is the medical field in which the abnormal brain tumor images are categorized prior to treatment planning. Accurate identification of the type of the brain abnormality is highly essential since the treatment planning is different for all the brain abnormalities. Any false detection may lead to a wrong treatment which ultimately leads to fatal results. By employing the Magnetic Resonance Spectroscopy (MRS) graph and thereby extracting the values of the metabolites from the graph one can classify the tumor based on the values of metabolites. The aim of this research is to identify brain tumour disease pattern from MRS images to perform differential diagnosis. The authors have employed the use of the Naïve –Bayes and J48 classifier for identification of the disease pattern from the three metabolite ratios.

General Terms
Pattern Recognition

Keywords
MRS, Metabolites, Brain tumour, Naïve-Bayes, Confusion Matrix, Cross-Validation, J48

1. INTRODUCTION
MRS is an important technique widely used by all radiologists as one of the method for diagnosing and evaluating common brain tumours.MRS can show abnormal findings in nearly 100% of brain tumors and in some instances it can be used for differential diagnosis of brain tumours [1][2].

The amount of metabolite is calculated with respect to the area under its spectrosgraphic peak.If the area is larger it indicates that the substance present is more.[3]

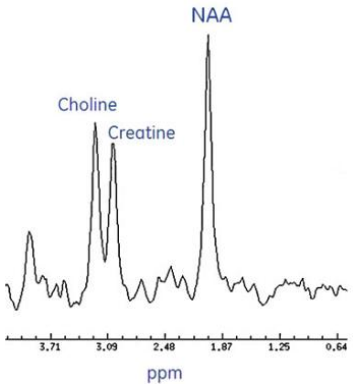

Figure 1: **Sample metabolites used for tumor classification**

1.1 MR Imaging Evaluation
In each patient ,axial T1-weighted,T2-weighted FSE, and fast fluid attenuated inversion recovery (FLAIR) images were obtained. After intravenous administration of gadolinium contrast-enhanced T1 weighted SE sequences were obtained in axial, coronal and sagittal planes. Multiecho multiplanar sequences through the brain were carried out both before and after administration of IV contrast. Multi-voxel spectroscopic data were obtained in all patients. As data was collected from multiple sources the scanning parameters for some images differed and for T2 sagittals and coronals and T1, T2, GRE and FLAIR were obtained. Diffusion DW images and single-voxel spectroscopy (35/144 TE) were obtained for some instances.

2. MATERIALS
A study population comprising of one hundred and twenty three patients collected from four different hospitals are classified into the three tumour classes (Benign, Malignant and Infection) using metabolites present in MRS graph. They were further evaluated for diagnosing the type of tumour disease Pattern .Metabolic peaks used in the differentiation of the different tumour types are shown in fig 1.The prominent metabolites used are NAA at 2.02 ppm, Cho at 3.22 ppm, Cr at 3.01 ppm. Metabolite ratios of NAA/Cho, NAA/Cr, Cho/Cr and NAA/Cho are calculated for classifying the tumour into the three classes defined as benign, malignant and infectious diseases.

3. PROPOSED SYSTEM

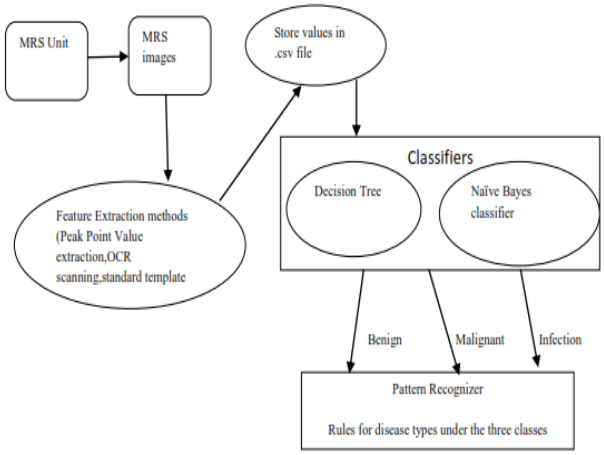

Figure 2:Proposed Architecture of the system

As shown in fig.2 the MRS images are collected from various hospitals. As storage of data in the form of images consumes space hence by applying different techniques as discussed in the next section the relevant metabolites are extracted from the images and stored in any CSV file. After that either CSV file itself or a .arff file as in our case is created and it is given as input to the various classification algorithms. The results are evaluated by different performance metrics like sensitivity and specificity. The result is the number of patients identified and classified into the appropriate three classes.

4. TECHNIQUES

4.1 J48

J48 algorithm steps [4]

1. Assume attributes are numerical.
2. Now Tree is constructed in a top-down recursive manner
3. At start, all the training examples are at the root
4. Examples are partitioned recursively based on selected attributes
5. Attributes are selected on the basis of an impurity function (e.g., information gain)

$$Entropy = \sum - p_i \log_2 p_i \quad (1)$$

The algorithm stops when any of the following conditions for stopping partitioning are met:-

1. All examples for a given node belong to the same class.
2. There are no remaining attributes for further partitioning majority class is the leaf
3. There are no examples left

4.2 Naive Bayes Classification

Bayesian Probability theory is used to judge the truth of the hypothesis of the given data.[5] The truth of the Hypothesis is carried out by checking some combination of values of attributes which classifies the instance to some class .This is carried out using Bayes rule:-

$$P(H/D) = P(D/H) * P(H) / P(D) \quad (2)$$

Here D is instance to check

H is hypothesis

And term P (D/H) is likelihood function.

This is a statistical method of learning and it reflects what we have learned about the validity of the hypothesis by considering the data.

5. EXPERIMENTATION AND RESULT DISCUSSIONS

The values of the metabolite ratios such as NAA/Cr,NAA/Cho and Cho/Cr play an important role in brain tumour classification. [6][7]As shown in table normal and abnormal values used in classification are depicted.All research in the field indicates that the metabolite NAA decreases in almost all brain tumours and Cho increases which thereby leads in increased value for the ratio Cho/NAA.

Table 1: Metabolite ratios

Metabolite ratio	Normal	Abnormal
NAA/Cho	1.6	<1.2
NAA/Cr	2.0	<1.6
Cho/Cr	1.2	>1.5

Table 2: **Range of values used in classifying tumor types and grades.**

Cho	Naa	Cr2	Cr	Naa/ Cho	Cho /Cr	Naa /cr	Result
14	9.84	5.46	6.46	.70	2.16	1.52	Malign ant
10	19.7	4.61	11.2 2	1.97	.891	1.75	Benign
3.9	3.37	3.63	2.78	.86	1.4	1.2	Infectio n

In data classification, the classifier is evaluated by a confusion matrix. The model performance is assessed by using the metrics true positive rate (sensitivity), false positive rate, specificity, precision and f-measure [8]. The words sensitivity and specificity have their original screening tests for diseases. Sensitivity is defined as the probability that the test says a person has the disease when in fact they do have the disease.

Sensitivity=TP/(TP+FN) (3)

Specificity is defined as the probability that the test says a person does not have the disease when in fact they are disease free.

Specificity=TN/(TN+FP) (4)

Table 3: **Training Dataset Mean values for all instances**

Classes	Mean Cho/Naa	Mean Cho/Cr	Mean Naa/Cr
Benign	0.565	0.857	1.5
Malignant	3.172	1.574	0.529
Infection	18.702	1.687	0.103

Table 4: **Test Dataset Mean values for all instances**

Classes	Mean Cho/Naa	Mean Cho/Cr	Mean Naa/Cr
Benign	0.494	0.883	1.506
Malignant	3.412	2.344	0.744
Infection	19.817	2.293	0.133

Table 5: **Sample of rules for malignant class disease**

If NAA/CHO <.63 and NAA/Cr<0.78 then tumor disease type is malignant_ perilesional meloma
If NAA/CHO <.63 and CHO/Cr>=1.31 and NAA/Cr <0.78 then tumor disease type is malignant_ metastatic deposits
If NAA/CHO<0.24 and CHO/Cr<0.4 then tumor disease type is malignant_ischemic_pathology

Table 6:**Sample of rules for benign disease**

If NAA/CHO>=1.74 and CHO/Cr<0.83 then tumor disease type is benign_tuberculoma
If NAA/CHO>=1.35 then tumor disease type is benign_encephallitic_pathology
If NAA/CHO<0.24 and CHO/Cr>=0.4 then tumor disease type is benign_epidermoid

Table 7: **Sample of rules for benign disease**

If NAA/CHO<.11 and CHO/Cr<1.05 then tumor disease type is infection_ischemic_pathology
If NAA/CHO<.11 and CHO/Cr>=1.05 then tumor disease type is infection_tuberculous_granulomatrus

5.1 Receiver Operating Characteristic curve (or ROC curve.):

Receiver operating curve is used to plot true positive rate against false positive rate. As seen in the figures 3-6 as the curve is closer to the left hand border it indicates that our tests for the three classes are more accurate.

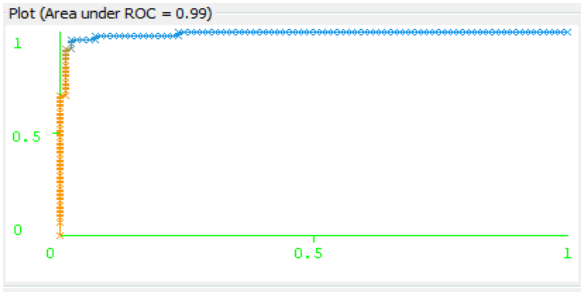

Figure 3: **ROC for malignant Class by Naïve Bayes Classifier**

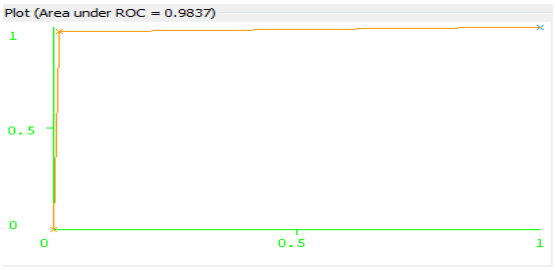

Figure 4: ROC for malignant Class by j48 Classifier

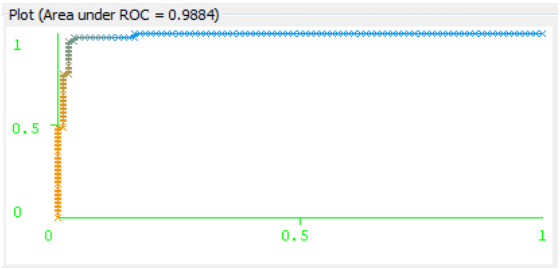

Figure 5: **ROC for benign Class by Naïve Bayes Classifier**

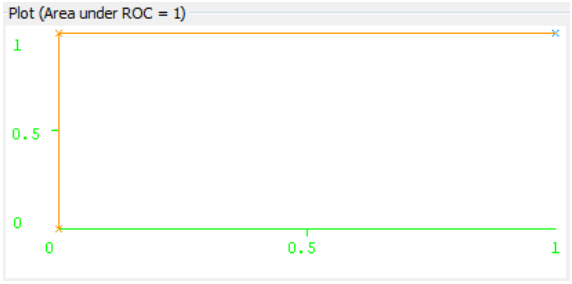

Figure 6: ROC for benign Class by j48 Classifier

5.2 K-fold cross validation

Here the dataset was divided into 10 subsets and the holdout method was repeated 10 times.In each iteration one of the subset was employed as test data and the remaining k-1 subsets were used for training purposes.This leads in calculation of average error across all trials.The advantage of this method is that every instance is used as both test and train data resulting in reducing the variance as k increases.[10]

Table 5: **Sample of instances depicted of Malignant, Benign and Infection type for 10 * Cross Validation**

inst#	actual	predicted	error	Prediction
1	2:Benign	2:Benign		1
2	2:Benign	2:Benign		0.97
3	2:Benign	2:Benign		0.865
4	2:Benign	2:Benign		1
5	2:Benign	2:Benign		0.968
1	2:Benign	2:Benign		0.976
2	2:Benign	2:Benign		0.955
3	2:Benign	2:Benign		1
4	2:Benign	2:Benign		0.959
5	2:Benign	2:Benign		0.959
1	2:Benign	2:Benign		1
2	2:Benign	2:Benign		0.994
3	2:Benign	2:Benign		0.963
4	2:Benign	2:Benign		0.966
5	2:Benign	2:Benign		0.872
1	2:Benign	2:Benign		0.976
2	2:Benign	2:Benign		0.955

3	2:Benign	2:Benign		1
4	2:Benign	2:Benign		0.959
5	2:Benign	2:Benign		0.959

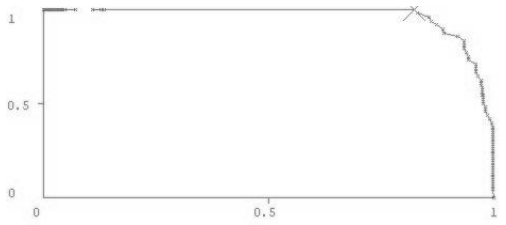

Figure 7: **ROC Curve for malignant class with accurracy=0.8254 obtained by 10 fold cross validation for Naïve Bayes algorithm**

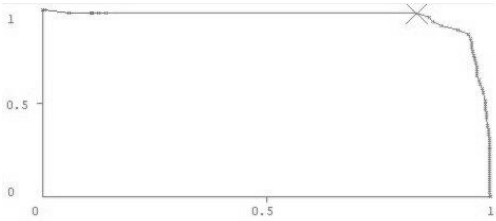

Figure 8: **ROC Curve for benign class with accurracy=0.8369 obtained by 10 fold cross validation for Naïve Bayes algorithm**

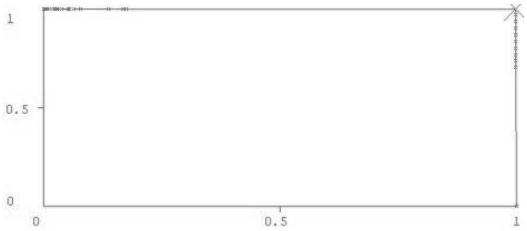

Figure 9: ROC Curve for infection class with accurracy=1.0 obtained by 10 fold cross validation for Naïve Bayes algorithm

6. CONCLUSION

The approach of J48 was used to generate rules for the tumor disease pattern. The rules were statistically verified by Naïve Bayes Algorithm. Furthermore the entire results were validated using 10 fold cross-validation. Both the results gave accuracy from 85-95% and 100% for infection class which was additionally incorporated to malignant and benign classes. This study was able to narrow down the range of values for the metabolite ratios in determining of tumor disease pattern. The results are promising and leaves much scope for exploration of not only similarly other brain tumor disease patterns but that these techniques can be used for other medical diagnostic applications too. This was a maiden attempt by authors in narrowing down the values for metabolite ratios that can enable them in generating rules for diseases in each category.This identification is based only on values of metabolite ratios which the authors have extracted by using optical character recognition techniques. Also in this research only three prominent metabolites were taken under consideration namely NAA, Cho and Cr. As this is an initial exploration into identification of disease patterns from spectroscopic peaks further research into this area can take into consideration other metabolites like lipids and glutamine etc for identifying other patterns as there are more than hundred plus type of brain tumors.

7. REFERENCES

[1] Christian Plathow, Wolfgang A. Weber, "Tumor Cell Metabolism Imaging", *The Journal of Nuclear Medicine*, Vol. 49, No. 6, June 2008

[2] S. Chawla,S. Wang,R.L. Wolf,J.H. Woo,J. Wang,D.M. O'Rourke,K.D. Judy,M.S. Grady,E.R. Melhem,H. Poptani, "Arterial Spin-Labeling and MR Spectroscopy in the Differentiation of Gliomas", *AJNR Am J Neuroradiol*, Oct 2007

[3] Alessandro Alimenti, Jacqueline Delavelle,François Lazeyras, Hasan Yilmaz , Pierre-Yves Dietrich, "Magnetic Resonance Spectroscopy in the Progression of Gliomas" , *European neurology ,Neurosurgery, Geneva University* Hospital, Geneva , Switzerland

[4] Fernando Gonz, Navarro and Llu, A. Belanche-Mu, "Using Machine Learning Techniques to Explore H-MRS data of Brain Tumors",8th Mexican IEEE international conference on artificial intelligence, 2009,pp 134-139

[5] L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, C. Majo, A. Moreno-Torres, M. Van Der Graaf, F.A. Howe, A.R. Tate, C. Aru, S. Van Huffel, "Brain tumor classification based on longecho proton MRS signals", *Artificial Intelligence in Medicine,* Vol. 31, pp.73-89, 2004

[6] Weibei Dou, Jean-Marc Constans, "Glial Tumors: Quantification and Segmentation from MRI and MRS",Diagnostic Techniques and Surgical Management of Brain Tumours,2011,pp 93-116

[7] Simonetti A.W., Melssen W.J., Szabo de Edelenyi F., van Asten J.J.A., Heerschap A., and Buydens L.M.C., "Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification", *NMR in Biomedicine*,Vol.18, pp.34–43, 2005

[8] Garcia S. and Herrera F., "Statistical comparisons of classifiers over multiple data sets for all pairwise comparisons", *Journal of Machine LearningResearch*, vol.9. Issue 2, , pp677–694, 2008

[9] Garcia-Gomez J.M., Luts J., Julia-Sape M., Krooshof P., Tortajada S., Vicente Robledo J., Melssen W., Fuster-Garcia E., Olier I., Postma G., Monleon D.,Moreno-Torres A., Pujol J., Candiota A.-P., Martinez-Bisbal M.C., Suykens J.A.K., Buydens L.M.C., Celda B., Van Huffel S., Arus C., and Robles M."Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy", *Magnetic Resonance Materials in Physics, Biology and Medicine*, Vol.22, pp.5–18, 2009

[10] Galanaud D., Nicoli F., Chinot O., Confort-Gouny S., Figarella-Branger D., Roche P., Fuentes S., Fur Y.L., Ranjeva J.-P., and Cozzone P.J., "Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy", *Magnetic Resonance in Medicine,*Vol 55, pp.1236–1245, 2006