CLASSIFICATION OF BRAIN TUMORS BY MEANS OF HIGH RESOLUTION MAGIC ANGLE SPINNING SPECTRA

V. Van Belle[1], J-B. Poullet[1], D. Monleon[2], B. Celda[2], M. Martinez-Bisbal[2] and S. Van Huffel[1]
[1]Katholieke Universiteit Leuven, Department of Electrical Engineering (ESAT), Division SCD, Belgium
[2]Universidad de València, Química-Física, Spain

Abstract

One of the techniques to classify brain tumors is to use nuclear magnetic resonance (NMR) signals [3][8][11]. As a consequence of interactions between different spins, not all interesting metabolites can be obtained in the spectrum. In this paper we investigate whether the use of High Resolution Magic Angle Spinning Spectra (HRMAS) can achieve a good performance in the classification of different brain tumors. Therefore a tree of classifiers to distinguish four tumor types is created.

1 Advantages of MRS

The principle of HRMAS is based on the principle of magnetic resonance spectroscopy (MRS) [3][6][8][11]. HRMAS signals have two main advantages. The signals are measured from a rotating biopsy aligned at an angle of 54.7° relatively to the external field, which results in the cancellation of mechanisms leading to peak widening. The second advantage is the possibility to lower the sampling frequency, resulting in a higher resolution.

2 Methods

The different methods explored in this research involve different preprocessing steps, feature extraction, feature selection, classification and testing methods. The paper is structured as follows. This section describes the different methods. Section 3 gives an overview of the obtained results. In section 4 the different experiments are summarized. Section 5 displays the results.

2.1 Preprocessing

The size of the measured Free Induction Decay (FID) signal makes calculations difficult. Because of the exponential decaying property of this kind of signals, only the first 2048 samples can be used without loosing much information contained in the signals. The calculated spectra suffer from an offset which has to be removed prior to further calculations. As the spectra are sensitive to inter-subject variability, they have to be normalized. Three possible normalizations are investigated: normalization by dividing the spectra by the energy, by the surface underneath the water signal or by the amount of creatine in the sample. The large overlap between contributions of different metabolites, caused by larger molecules, makes smaller interesting peaks invisible. To make the latter visible, contributions of lipids are reduced in two ways. The first method removes the first 40 samples of the time signal and so reduces contributions of molecules with a large damping. The second method makes use of the modeling of the signal by a sum of damped exponentials [9], and removes those components with the largest damping. Finally, the signal-to-noise ratio is improved by apodization. Since the earlier samples contain a larger informative signal and the noise level is assumed to be constant, the signal-to-noise ratio is improved by applying a larger weight to the first samples.

2.2 Feature Selection

In this research classification is performed using three kinds of inputs. Firstly the spectra as a whole (2048 samples) are used. To reduce computation time, features are extracted by Principle Component Analysis (PCA) [4] or by means of peak integration [10]. In contrast to PCA, peak integration calculates interesting measures of the concentrations of metabolites present in the biopsy, thereby providing chemically relevant features.

In order to achieve a good classification performance, those features with the highest discrimination ability have to be selected. The standard procedure to decide upon the statistical relevance of features is an ANOVA test. Since the assumptions of this test can not be assured, nonparametric tests – Wilcoxon rank-sum test and Kruskal-Wallis test – are used. Alternatively, the Fisher criterion can be applied.

2.3 Classification

As indicated in the previous section, classification is performed by using the spectra as a whole or using features. The performance of five different classification methods is investigated. If the performance of a linear decision plane is not inferior to that of a more complex decision plane, the former should be preferred. Linear Discriminant Analysis (LDA) [5] is the only investigated method creating a linear separating plane in the input space. Similarly to LDA, support vector machines (SVM) [13] and

least squares support vector machines (LS-SVM) [12] are supervised methods. The latter two project the input space into a higher dimensional feature space. A linear separating plane is created in the feature space, representing a non-linear decision plane in the input space (figure 1).

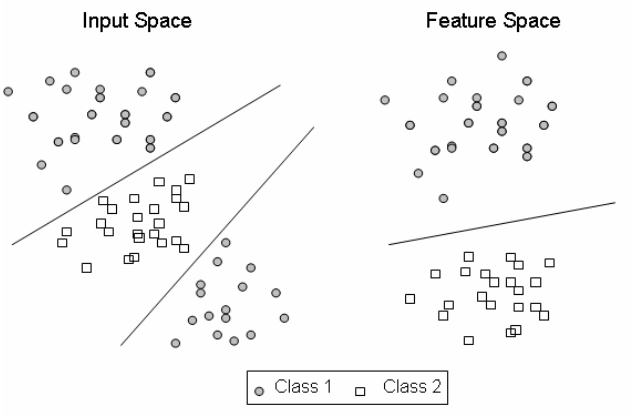


Figure 1:

In the input space the two classes are not linearly separable. By projecting the input space to a higher dimensional feature space the classes become separable.

As opposed to the former three methods, K-means clustering [1] and Fuzzy c-means clustering [2] are unsupervised methods. The difference between those methods is a weight term. K-means clustering assigns a class label to each of the given samples, whereas fuzzy c-means clustering adds a weight, giving a measure for being a member of each of the clusters. A disadvantage is their nondeterministic behavior.

2.4 Validation

In order to find an optimal value for the hyper-parameters a validation set is needed. As the limited amount of data makes it impossible to separate the dataset into a training and validation set, a leave-one-out cross-validation procedure is used.

3 Description of the data

The data used in this paper are part of the database of the EU project e-TUMOUR [7] and are provided by the UVEG-group in Valencia. The dataset contains HRMAS signals of 47 patients with a brain tumor. These patients are split up into four tumor groups: glioblastomas (17 persons), meningiomas (11 persons), metastases (8 persons) and gliomas (grade II and III: 11 persons). Because of the lack of data, each of these groups has a large variety of tumors. The class of gliomas contains astrocytomas of different grades as well as oligodendrogliomas and a schwannoma, rendering classification difficult.

4 Experiments

4.1 Classifier type

The goal of this research is to find the best classification procedure to assign samples to one of the four different classes. In that respect, different tree architectures are investigated. The binary tree is formed by three binary classifiers discriminating between two tumor groups. Figure 2 shows one binary tree. The first classifier has to discriminate glioblastomas and metastases on the one side and meningiomas and gliomas on the other side. The second classifier has to distinguish between glioblastomas and metastases. The third classifier has to decide whether the tumor is a meningioma or a glioma. The ternary tree contains one binary classifier, classifying for example between glioblastomas on the one hand and all three other tumor types on the other hand. The second classifier of the tree is ternary and has to make a distinction between three classes. The quarternary tree uses only one classifier discriminating all tumor types.

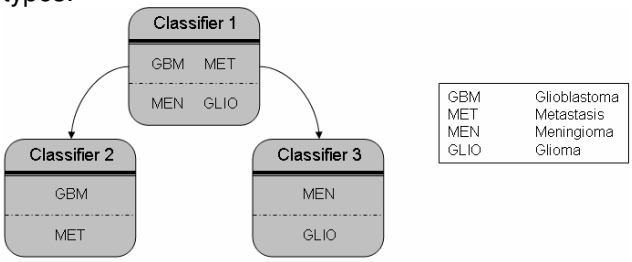


Figure 2: Binary classification tree.

To select the best performing tree, peak integration is used in combination with nonparametric feature selection. The study reveals that the binary tree achieves the best performance, namely 78.3% or 10 misclassifications on 46 tumors[1]. The ternary tree, with a performance of 69.2% is better than the quaternary one, with 20 misclassifications.

4.2 Feature selection

Since the spectra contain all of the information present in the signals, it could be expected that the better performance would be achieved when using the whole spectra. Unfortunately, this can become very time consuming. The performance of classification by means of PCA or peak integration is compared with classification based on the whole spectrum. The spectra are normalized by division by

[1] Normalization is done by dividing by the surface underneath the water signal. Since this measurement is unavailable for one person, there are only 46 samples for this study.

the surface underneath the water signal and the offset is reduced. Features are selected by means of nonparametric tests. The classifier using the whole spectrum achieves a performance of 78.7%, in contrast to 78.3% for peak integration and 70.2% for PCA. The reason for the increasing misclassification rate after PCA is illustrated in figure 3. Using PCA prior to classification can have major consequences. It is possible that the data are perfectly separable before PCA is applied, but afterwards the structure of the two classes can be completely lost.

In this research, it is observed that feature selection does not reveal any statistically significant differences in discrimination ability among the different tumor classes.

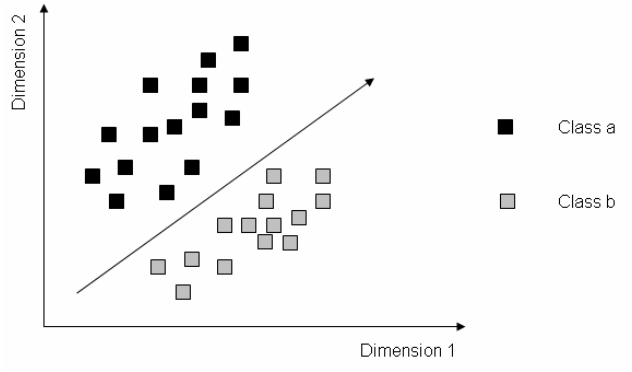


Figure 3:

Two separable classes can be inseparable after PCA!

4.3 Classifion method

This section describes the influence of the used classification method on the number of misclassifications. Methods creating a non-linear separating plane in the input space nearly always generate better classifiers. LS-SVM is preferred to SVM, in terms of performance and computation time. Nonsupervised methods have an extremely high number of misclassifications. A second disadvantage of this kind of methods is their nondeterministic behavior. It is clear that for medical applications a deterministic approach is preferred.

5 Results

Since a binary decision tree is used to classify data into four different classes, three trees can be built. The performances are summarized in tables 1 to 3. LS-SVM is the best classification method for all illustrated classifiers. For the classification of glioblastomas versus meningiomas or metastases, SVM gives the same performance as LS-SVM. LDA performs as well as LS-SVM when differentiating between gliomas and metastases or meningiomas. Classifiers using peak integration to create features achieve a better performance than classifiers using the whole spectrum. Only when discriminating between meningiomas and metastases, both classifiers are equally performing. Research could not show any relation between performance of the classifiers and oher preprocessing steps. Therefore, the different preprocessing steps are not mentioned.

As illustrated in table 1, the first classifier of the first tree has to make a distinction between glioblastomas and meningiomas on one side and metastases and gliomas on the other side. Five samples are wrongly classified, resulting in a performance of 89.13%. The following classifiers have to decide between a glioblastoma or a meningioma or between a metastasis or a glioma. No mistakes are made. In total, the first tree reaches a performance of 89.13%[2]. The second tree is analogous with a performance of 89.36%. The first classifier of the last tree assigns four tumors to the wrong class. In this case however, the second classifier makes 2 mistakes. Since one of these tumors is already misclassified by the first classifier, this does not count as a misclassification for the tree as a whole. The performance of the third tree is equal to that of the second one.

It should be noted that the given performances are calculated on the validation set. Due to the limited size of the dataset and a delay in the acquisition of new data, the performance on the test set could not yet be measured.

Table 1:

Performance of the first binary decision tree: 89.13%

Classifier	Performance (%)	Wrongly classified samples
(GBM, MEN) ↔ (MET, GLIO)	89.13	17, 24, 25, 34, 35
(GBM) ↔ (MEN)	100	/
(MET) ↔ (GLIO)	100	/

Table 2:

Performance of the second binary decision tree: 89.36%

Classifier	Performance (%)	Wrongly classified samples
(GBM, MET) ↔ (MEN, GLIO)	89.36	5, 18, 20, 21, 22
(GBM) ↔ (MET)	100	/
(MEN) ↔ (GLIO)	100	/

[2] Normalization is done by dividing by the surface underneath the water signal. Since this measurement is unavailable for one person, there are only 46 samples for this study.

Table 3:

Performance of the third binary decision tree: 89.36%

Classifier	Performance (%)	Wrongly classified samples
(GBM, GLIO) ↔ (MEN, MET)	91.49	22, 32, 46, 47
(GBM) ↔ (GLIO)	92.86	4, 47
(MEN) ↔ (MET)	100	/

Although the goal of the study is to discriminate upon four classes, from a clinical view it can be interesting to discriminate between two classes. Table 4 shows a higher performance when binary classification is considered. Discriminating glioblastomas from gliomas seems to be a problem. A possible explanation lies in the existence of two kinds of glioblastomas: primary and secondary. Since the latter resemble gliomas, this can cause classification problems. Because no data about the status of the glioblastomas is inserted in the database, this assumption can neither be confirmed, nor denied.

Table 4: Performance of binary classifiers.

Classifier	Performance (%)	Wrongly classified samples
(GBM) ↔ (MET)	100	/
(GBM) ↔ (MEN)	100	/
(GBM) ↔ (GLIO)	92.86	4, 47
(MEN) ↔ (MET)	100	/
(MEN) ↔ (GLIO)	100	/
(MET) ↔ (GLIO)	100	/

6 Conclusions

In summary, the results show a preference for the use of binary classifiers, peak integration and LS-SVMs. A few questions remain unanswered. No decisions could be made about the importance of the feature selection criterion, the reduction of the contribution of the lipids to the spectra, the kind of normalization and the impact of apodization. Further investigation on a larger dataset is necessary to find answers on the remaining questions.

Acknowledgments

Dr. Sabine Van Huffel is a full professor at the Katholieke Universiteit Leuven, Belgium. The research was supported by Research Council KUL: GOA-AMBioRICS, CoE EF/05/006 Optimization in Engineering, IDO 05/010 EEG-fMRI, several PhD/postdoc & fellow grants; by Flemish Government: FWO: PhD/postdoc grants, projects, G.0407.02 (support vector machines), G.0360.05 (EEG, Epileptic), G.0519.06 (Noninvasive brain oxygenation), FWO-G.0321.06 (Tensors/Spectral Analysis), G.0341.07 (Data fusion), research communities (ICCoS, ANMMM); and IWT: PhD Grants; by Belgian Federal Science Policy Office IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling'); by EU: BIOPATTERN (FP6-2002-IST 508803), ETUMOUR (FP6-2002-LIFESCIHEALTH 503094), Healthagents (IST–2004–27214, FAST (FP6-MC-RTN-035801) and by ESA: Cardiovascular Control (Prodex-8 C90242)

References

[1] Anderberg M., 1973. *Cluster Analysis for applications*. Academic Press, Inc., New York.

[2] Bezdek J., 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press, New York.

[3] Devos A., 2005. *Quantification and classification of magnetic resonance spectroscopy data and applications to brain tumor recognition.* PhD thesis, Faculty of Engineering, K.U.Leuven (Leuven, Belgium)

[4] Duda, R., Hart, P., Stork, D., 2001. *Pattern Classification,* 2nd ed. John Wiley& Sons.

[5] Fisher R., 1936. "The use of multiple measurements in taxonomic problems." *Annals of Eugenics,* vol. 7, pp.179–188.

[6] Gadian D., 1995. *NMR and its applications to living systems*. Oxford University Press.

[7] http://www.etumour.net/

[8] Howe F. and Opstad K., 2003. "^{1}H MR spectroscopy of brain tumours and masses." *NMR Biomed*, vol. 16, pp. 123-131.

[9] Laudadio T., Mastronardi N., Vanhamme L., Van Hecke P. and Van Huffel S., 2002. "Improved Lanczos algorithms for blackbox MRS data quantitation." *Journal of Magnetic Resonance*, vol.157, pp.292-297.

[10] Meyer, R., Fisher, M., Nelson, S., Brown, T., 1988. "Evaluation of manual methods for integration of in vivo phosphorus NMR spectra." *NMR Biomed,* vol. 1, pp. 131-135.

[11] Rijpkema M., Schuuring J., van der Meulen Y., van der Graaf M., Bernsen H., Boerman R., van der Kogel A. and Heerschap A., 2003. "Characterisation of oligodendrogliomas using short echo time ^{1}H MR spectroscopic imaging." *NMR Biomed*, vol. 16, pp.12-18.

[12] Suykens J., Van Gestel T., De Brabanter J., De Moor B. and Vandewalle J., 2002. *Least squares support vector machines.* World Scientific Publishing.

[13] Vapnik V., 1995. *The Nature of Statistical Learning Theory.* Springer, New York.