TUMOURCLASSIFIER, A JAVA TOOL FOR FAST DEVELOPMENT AND IMPLEMENTATION OF MRS- BASED CLASSIFIERS.

S. Ortega-Martorell[1,2], I. Olier[1,2], M. Julia-Sape[1,3], and C. Arus[1,3]

[1]Bioquimica i Biologia Molecular, Universitat Autonoma de Barcelona, Cerdanyola del Valles, Barcelona, Spain, [2]Computing Languages and Systems, Technical University of Catalonia, Barcelona, Barcelona, Spain, [3]CIBER-BBN, Universitat Autonoma de Barcelona, Cerdanyola del Valles, Barcelona, Spain

1. INTRODUCTION

Currently available methods for classifying MRS data rely on either commercial (SPSS, SAS), non commercial (R) or home-made programs running over Matlab. Today, multiplatform, easy to use software programs for fast and robust classification of MRS data are scarce.

2. PURPOSE

TumourClassifier is a java software solution that uses statistical machine learning techniques to design and implement classifiers based on *in-vivo* SV ^{1}H-MRS.

3. METHODS

TumourClassifier has been developed in java, using well-known multiplatform libraries to carry out specific tasks, such as i) weka[1], used for selecting and extracting features; ii) javastat[2], used with modifications, to apply Linear Discriminant Analysis (LDA), iii) statgraphics[3], to generate graphs used in classifier results evaluation.

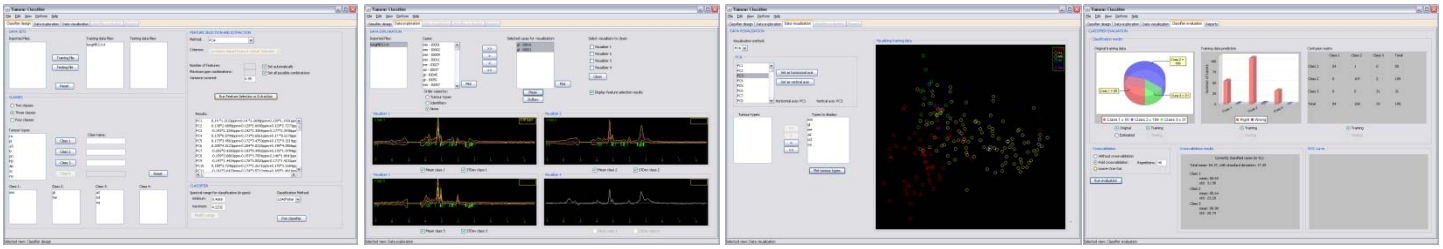
At the moment, *TumourClassifier* implements LDA as the technique of choice to separate two, three or four classes, depending on user needs. *TumourClassifier* is composed by the following modules: classifier design, data exploration, data visualization, classifier evaluation and reports.

Classifier design (figure 1a): Tunes the desired inputs for designing the classifier, such as the training datasets, the definition of tumor classes and the selection of relevant features. Three methods, Sequential Forward, Sequential Backward and Principal Components Analysis (PCA) have been implemented for selecting or extracting relevant features. The resulting features or combination of features are used as classifier inputs.

Data exploration (figure 1b): Allows displaying spectral data, mean and standard deviation for analyzing spectra type population structure and visually comparing spectra.

Data visualization (figure 1c): It is an up to 2D-latent space visualizer of PCA or LDA results.

Classifier evaluation (figure 1d): An essential part of the life cycle of the classifier development is its validation. *TumourClassifier* contains several methods for performing this such as: a) Confusion matrix, b) ROC (Receiver Operating Characteristic) curve, c) Graphs showing the statistics of well predicted cases, using test sets, and d) Cross-validation results with mean and standard deviation of correctly classified cases, in general and by group. The application also allows generating reports with the results obtained.



a) Classifier design tab b) Data exploration tab c) Data visualization tab d) Data evaluation tab

Figure. 1: Main tabs of *TumourClassifier*

Steps for creating a classifier

The process involves the following sequential steps:

1st) Importing data sets for training, 2nd) Indicating the number of groups to separate, 3rd) Selecting the tumor type/s corresponding to each group to be classified, 4th) Selecting an option for obtaining relevant features, 5th) running the method with all this information to create the classifier, 6th) Visualizing the results obtained, 7th) Evaluating the classifier.

Evaluation with real data: SV ^{1}H-MRS data from INTERPRET[4] at short and long TE were used to test the modules implemented. A three-class classifier at short and another at long TE were produced as in [5], for the following brain tumor classes: low-grade meningioma, aggressive and low-grade glial.

4. RESULTS

The computing time needed to develop a classifier is less than one minute in a 3GHz CPU and 2GB RAM personal computer.

Short TE: 10 features were selected (1.250ppm, 1.308ppm, 1.346ppm, 2.247ppm, 2.305ppm, 2.363ppm, 3.053ppm, 3.552ppm, 3.647ppm, 3.820ppm). The "leave-one-out" cross-validation results were: low-grade meningioma 89.65%, aggressive 90.32%, low-grade glial 82%, for a total mean of 88.88% with a standard deviation of 31.42%. Without cross-validation the results were: low-grade meningioma 91.37%, aggressive 91.12%, low-grade glial 85.71%, for a total mean of 89.4%.

Long TE: 10 features were selected (1.135ppm, 1.173ppm, 1.250ppm, 1.499ppm, 1.519ppm, 2.132ppm, 2.343ppm, 3.034ppm, 3.053ppm, 3.763ppm). The "leave-one-out" cross-validation results were: low-grade meningioma 81.81%, aggressive 81.65%, low-grade glial 90%, for a total mean of 82.98% with a standard deviation of 37.57%. Without cross-validation the results were: low-grade meningioma 85.45%, aggressive 82.56%, low-grade glial 93.54%, for a total mean of 87.19%.

The results obtained compare well with previous non-automated analysis of the same dataset [5, 6].

5. CONCLUSION

TumourClassifier is a simple and robust tool for helping spectroscopists to analyze and classify *in-vivo* SV ^{1}H-MRS, minimizing the user learning process curve, and allowing straightforward display and evaluation of the obtained results.

---

[1] http://www.cs.waikato.ac.nz/ml/weka/. [2] http://javastat.stat.wvu.edu/. [3] http://www.statgraphics.net/ . [4] Julià-Sapé et al. MAGMA. (2006) 19:22-33

[5] Tate et al. NMR Biomed. (2006) 19:411-34. [6] García-Gómez JM et al. NMR Biomed. (2008) [Epub ahead of print]