Feature Selection in Proton Magnetic Resonance Spectroscopy Dada of Brain Tumors

Lluís A. Belanche-Muñoz and Félix Fernando González-Navarro

Universitat Politècnica de Catalunya

Eight International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics

Special session: CI for health at the edge: oncology, degenerative diseases and the ICU (CIHE)

Outline

Is a non-invasive technique that provides information about the biochemical profile of brain tissue

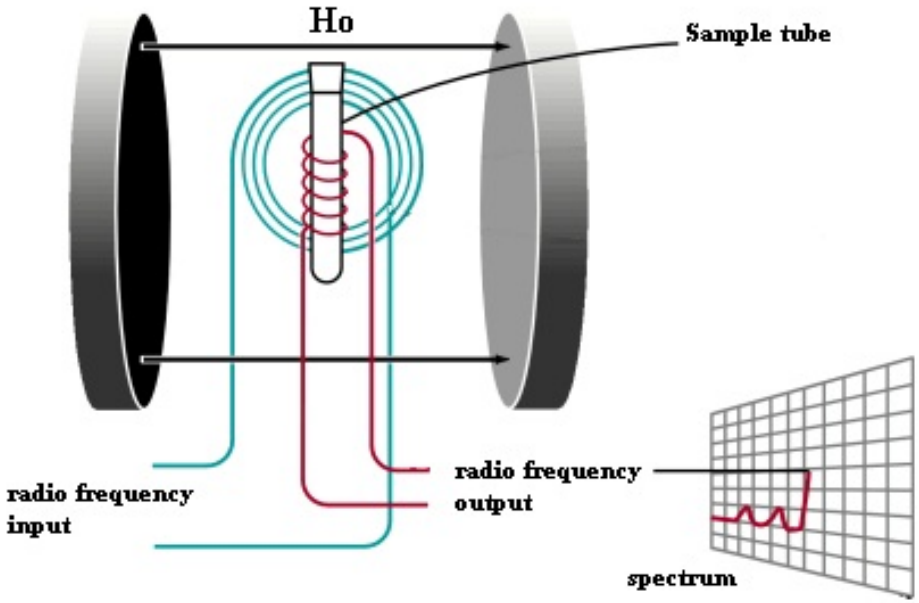


Figure 1: [1]H-MRS Fundamentals. Image source:

`http://www.chem.ucalgary.ca/courses/351/Carey/Carey.html`

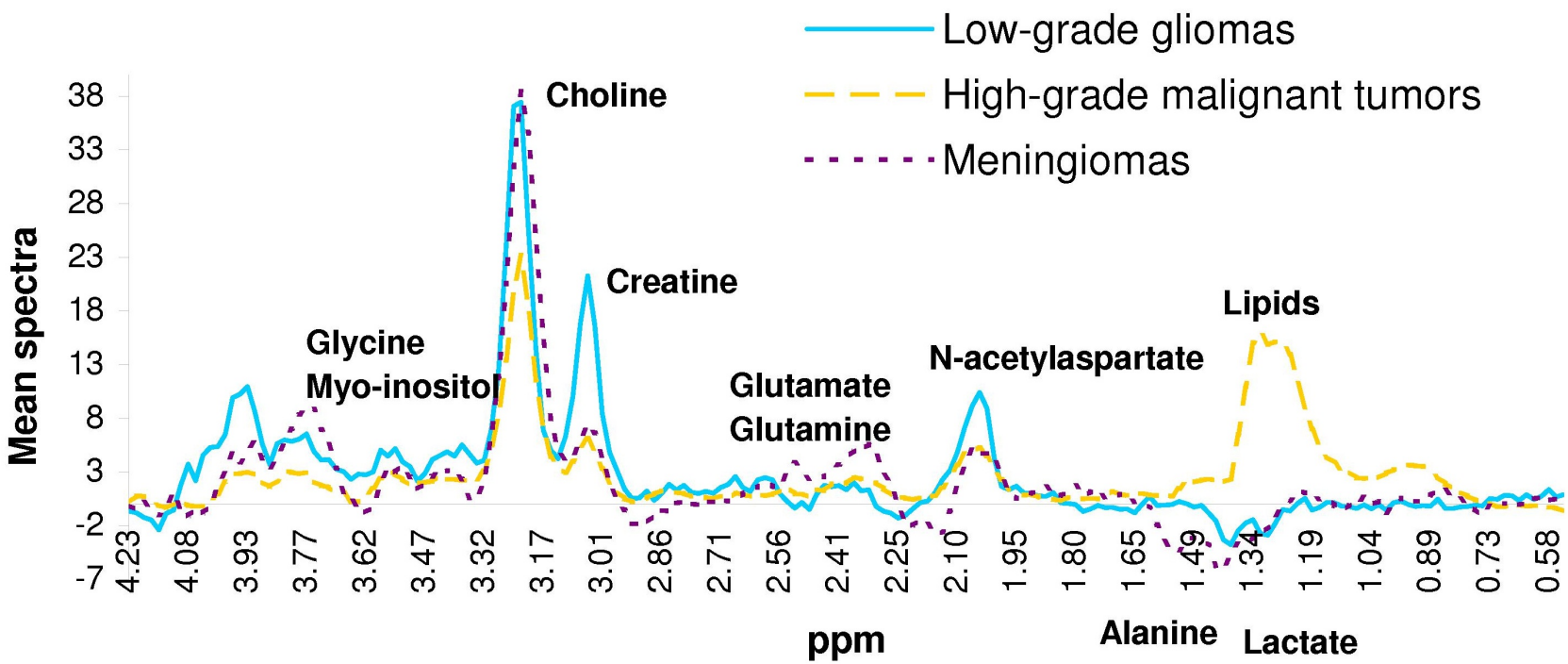


Figure 2: [1]H-MRS example.

Metabolites

- Meningiomas NAA $= \emptyset$, Choline $= \uparrow$
- Metastases Creatine $= \downarrow$
- High–Grade tumors Lipids $= \uparrow$

The Class-Separability Feature Selection Algorithm (CSFS)

1. Calculate distance value for each metabolite.

	Metabolites		
	1	...	m
Class 1	$\bar{X}_{1,1}$	$\bar{X}_{1,...}$	$\bar{X}_{1,m}$
Class 2	$\bar{X}_{2,1}$	$\bar{X}_{2,...}$	$\bar{X}_{2,m}$
Class 3	$\bar{X}_{3,1}$	$\bar{X}_{3,...}$	$\bar{X}_{3,m}$

The Class-Separability Feature Selection Algorithm (cont.)

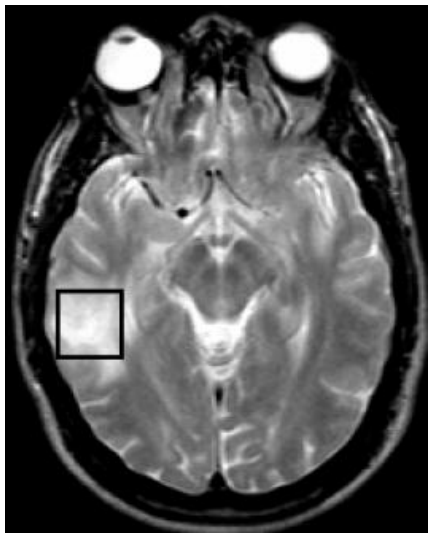
2. Separability degree for each metabolite.

$$DS_i = \sum_{j \neq k} |\bar{x}_{j,i} - \bar{x}_{k,i}|$$

3. Sort DS_i in descending order.

4. DS_i feeds a Forward-Backward Search.

5. All computations are bootstrap averages.



Single voxel ^{1}H-MR spectra acquired *in vivo* from brain tumor patients

LTE (PRESS 135–144 ms): 195 cases with 55 meningiomas
78 glioblastomas, 31 metastases, 20 astrocytomas grade II
6 oligoastrocytomas grade II and 5 oligodendrogliomas grade II

STE (PRESS 30–32 ms): 217 cases with 58 meningiomas
86 glioblastomas, 38 metastases, 22 astrocytomas grade II
6 oligoastrocytomas grade II, and 7 oligodendrogliomas grade II

LSTE: the merged LET and SET data, with 195 cases

195 frequency intensity values, from 4.21 ppm $\rightarrow$ 0.51 ppm

International Network for Pattern Recognition of Tumors Using Magnetic Resonance *INTERPRET*

Three Super-classes

1. High-grade Gliomas.
2. Low-grade Gliomas.
3. Meningiomas.

- The LTE, STE and LSTE data. The three super-classes.
- The median metric is used.
- Forward-Backward selection in wrapper mode
- Trained Six classifiers: NN, LDC, QDC, LR, ISVM, rSVM by 10x10 cross validation on the original data sets.

	NN	LDC	QDC	LR	lSVM	rSVM
LTE	[11] 89.70	[15] 93.51	[9] 89.09	[7] 91.48	[9] 91.82	[10] **93.88**
STE	[8] 92.21	[16] 93.34	[6] 87.40	[8] 90.48	[12] 93.14	[8] **94.48**
LSTE	[17] 96.14	[26] **98.27**	[8] 92.83	[7] 92.07	[14] 94.83	[12] 94.77

Table 1: CSFS Feature selection results and final performance. The number in square brackets is the final Best Spectral Subset (BSS) size. The right number is the averaged 10x10 CV accuracy in the original (continuous) ^{1}H-MRS datasets.

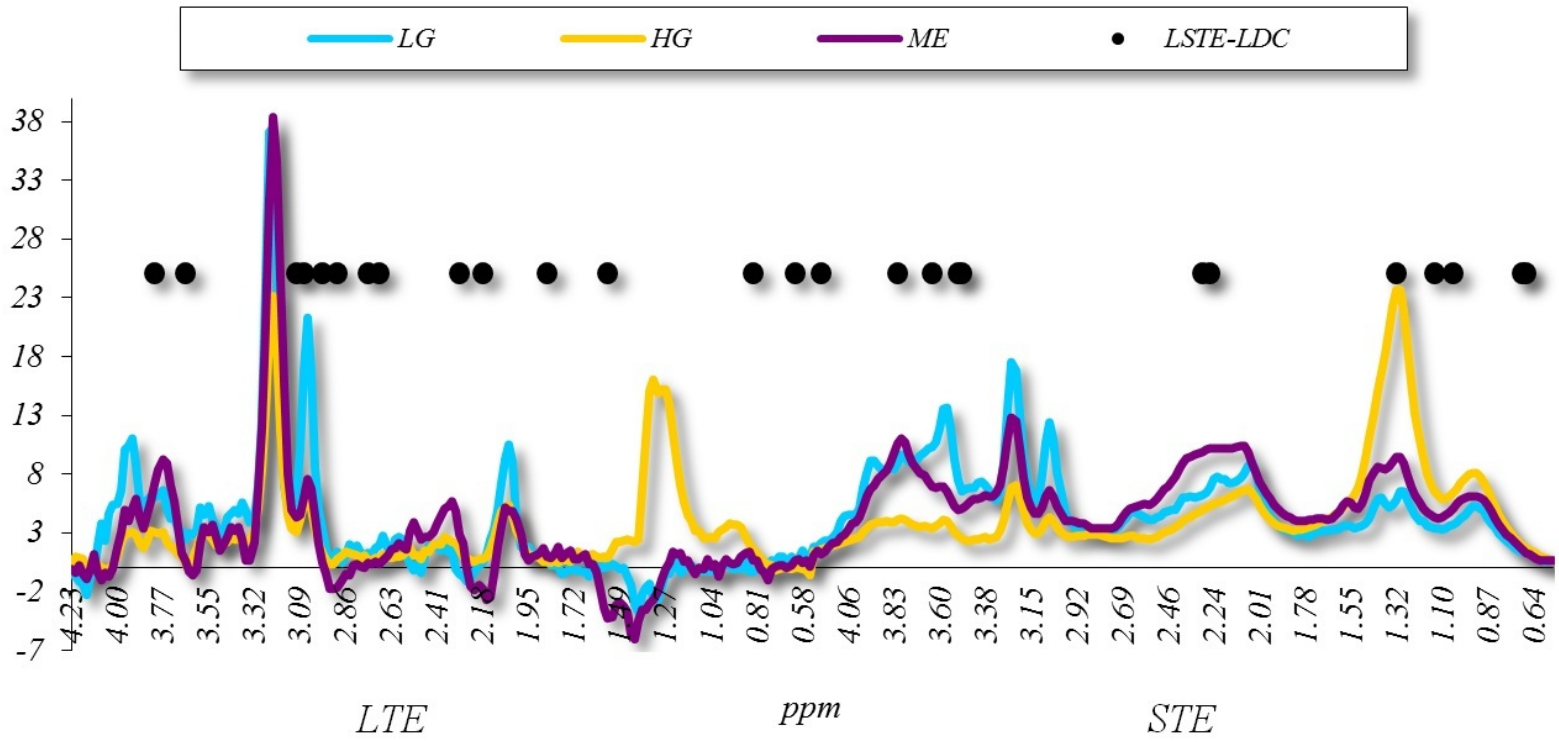


Figure 3: Best Spectral Subset from LSTE-LDC model as positioned in the whole spectrum.

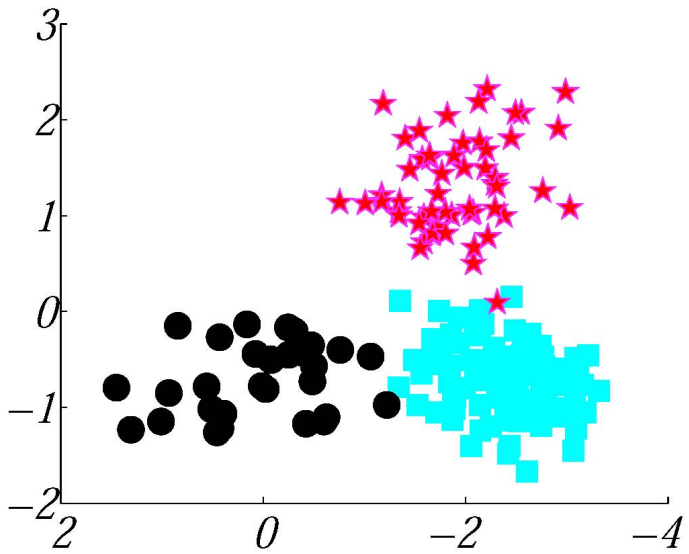


Figure 4: Projection of the dataset (using the final *selected* feature subset of the best LSTE-LDC model) onto the first two eigenvectors of the scatter matrices as coordinate system. Circles represent low-grade *gliomas*; filled squares high-grade *malignant tumors* and stars *meningiomas*.

- The proposed algorithm takes advantage of the differential presence of certain metabolites.
- It also accounts for possible redundancies.
- A competitive model w.r.t. other solutions in the literature.

- The “Best” Subset: It is extremely unlikely that it is found by a search process.
- This study confirmed that the LSTE combination renders better subsets in classification accuracy.
- Linear models are among the best suited for the task.

- Most of the identified metabolites are positively defined by the medical literature.
- Some concordances are found with successful recent machine learning works.