Automatic relevance source determination in human brain tumors using Bayesian NMF

Sandra Ortega-Martorell
Department of Mathematics and Statistics
Liverpool John Moores University
Liverpool, United Kingdom
S.OrtegaMartorell@ljmu.ac.uk

Carles Arús
Departament de Bioquímica i Biologia Molecular
Universitat Autònoma de Barcelona
Cerdanyola del Vallès, Spain
carles.arus@uab.es

Ivan Olier
Manchester Institute of Biotechnology
The University of Manchester
Manchester, United Kingdom
ivan.olier@manchester.ac.uk

Paulo Lisboa
Department of Mathematics and Statistics
Liverpool John Moores University
Liverpool, United Kingdom
P.J.Lisboa@ljmu.ac.uk

Margarida Julià-Sapé
Networking Research Center on Bioengineering,
Biomaterials and Nanomedicine (CIBER-BBN)
Universitat Autònoma de Barcelona
Cerdanyola del Vallès, Spain
margarita.julia@uab.cat

***Abstract*—The clinical management of brain tumors is very sensitive; thus, their non-invasive characterization is often preferred. Non-negative Matrix Factorization techniques have been successfully applied in the context of neuro-oncology to extract the underlying source signals that explain different tissue tumor types, for which knowing the number of sources to calculate was always required. In the current study we estimate the number of relevant sources for a set of discrimination problems involving brain tumors and normal brain. For this, we propose to start by calculating a high number of sources using Bayesian NMF and automatically discarding the irrelevant ones during the iterative process of matrices decomposition, hence obtaining a reduced range of interpretable solutions. The real data used in this study come from a widely tested human brain tumor database. Simulated data that resembled the real data was also generated to validate the hypothesis against ground truth. The results obtained suggest that the proposed approach is able to provide a small range of meaningful solutions to the problem of source extraction in human brain tumors.**

Keywords—non-negative matrix factorization; Bayesian NMF; brain tumors; ideal number of sources

I. INTRODUCTION

Brain tumors have a relatively low incidence amongst humans as compared to other more widespread cancer pathologies. Their clinical management is sensitive and difficult, though: the physical location of the tumor makes its direct removal a complex clinical procedure that entails a non-negligible risk of causing cognitive impairment. This also limits the availability of biopsy samples, whose histopathological analysis is the gold standard for tumor diagnosis and prognosis [1], [2]. As a result, the medical expert is often forced to rely on non-invasive indirect measurements of the tumor characteristics and growth.

In current radiological practice, these data measurements require technologies that belong to the modalities of either imaging or spectroscopy (or combinations of both) [3]–[5]. In this study, we approached this problem by using Non-negative Matrix Factorization (NMF) [6], [7], a group of unsupervised techniques in which a data matrix is approximately factorized into (usually) two matrices, namely the sources and the mixing matrix. Different variants of NMF have previously been applied in the context of neuro-oncology to distinguish normal from abnormal masses [8]–[11], and between different tumor types [12]–[14]. To different extents, they all have succeeded on identifying the underlying source signals, for which it was always necessary to know in advance the number of sources to calculate.

Over the last decade, full Bayesian approaches to modeling have become predominant in statistical machine learning. The most important advantage of a Bayesian approach is that the model complexity is explicitly incorporated into the optimization function. Thus, as excess complexity will be automatically penalized, the risk of overfitting the data is controlled.

Most NMF variants can be explained as constrained Bayesian models, whose non-negative factorizing matrices are estimated using maximum likelihood or maximum a posteriori under some assumptions on the distribution of the data and the factors. Bayesian modeling therefore provides not only an estimate of the factors, but also an estimate of their marginal posterior density, which is valuable for interpreting the factorization, computing uncertainty estimates, etc. From the different approaches that can be found in the literature, we will consider in the current study the Bayesian inference for NMF models [15].

The aim of the current study is not only to extract the underlying source signals using a Bayesian approach to NMF, but also to address one of the open questions in the use of NMF techniques in the context of neuro-oncology, which is estimating the number of sources that provides a meaningful characterization of the problem at hand. For this, we propose to start by calculating a high number of sources and automatically discarding the irrelevant ones during the iterative process of matrices decomposition of NMF, obtaining a reduced range of interpretable solutions in human brain tumors.

II. Materials

A. Real data

The empirical data used in this study were extracted from an international multi-center database [16] compiled by the INTERPRET European research project [17]. Class labeling was performed according to the World Health Organization (WHO) system for diagnosing brain tumors by histopathological analysis of a biopsy sample.

The data consist of single-voxel proton MR spectra (SV-^{1}H-MRS) acquired from brain tumor patients at a magnetic field intensity of 1.5T and with parameter settings at short echo time, 20-32 ms (STE). The acquired spectra comprise measurements from 22 tumor masses labeled astrocytoma grade II (A2), 86 glioblastomas (GL), 58 meningiomas (MM) and 22 normal brain controls (NO). Raw data were processed as described in [18]. A total of 195 clinically-relevant frequency intensity values measured in parts per million (ppm), a dimensionless unit of measurement, were sampled from each spectrum in the [4.24-0.50] ppm interval. Figure 1 shows the mean spectra of the analyzed tumor and tissue types across spectral ranges, which follow acceptable clinical practice.

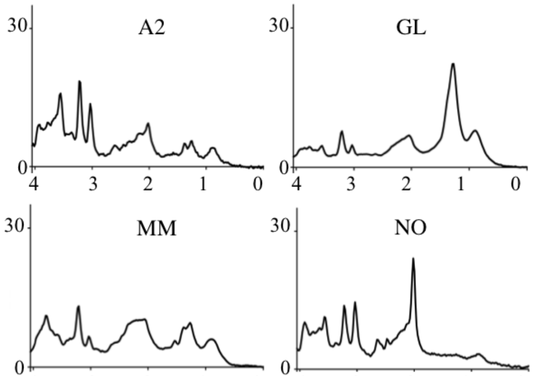

Fig. 1. Mean spectra of the unit length normalized (UL2) tumor and tissue types identified by their labels as described in the Materials section. Frequencies in the horizontal axes measured in ppm; magnitudes in the vertical axes in arbitrary units.

The label A2 indicates low-grade (grade II on a scale I-IV of the WHO) glial tumors, which grow by infiltrating normal brain tissue. This class of tumor masses may evolve to become highly malignant, WHO grade IV tumors, indicated by the label GL. Grade IV tumors usually have a necrotic pattern where infiltrating tissue has died through lack of blood perfusion leaving behind strong lipid signals that are most evident when obtaining MRS data at STE. However, not all GL have this necrotic pattern and some of them retain a spectral pattern, which is visually not too different from their low-grade glial counterparts, the A2. These cases might be considered as class outliers [19]. The label MM indicates low-grade meningiomas (WHO grade I and II) and they have a completely different origin, namely cells in the tissue that envelops the brain, called the meninges. Their spectral pattern is easy to recognize at STE, without necrosis, and different from the glial, metastatic or normal pattern.

B. Simulated data

The simulated data used in this study was modeled from samples extracted from the INTERPRET database (explained before). The selected cases were I0104, I0096, I0174, and I1474; which correspond to A2, GL, MM, and NO respectively. These cases were considered then the true sources, and were multiplied by a set of mixing matrices to form the simulated datasets. Two types of mixing matrices were randomly generated, slightly (20% variability) and highly (35% variability) mixed, to test these two levels of mixing of the sources. We created then 50 datasets per discrimination problem and level of mixing, and added Gaussian noise to all them resembling the typical height of the noise in this type of data (signal to noise ratio, SNR=66), as reported in [20].

III. Methods

As mentioned in the introduction, different approaches to Bayesian models of NMF can be found in the literature, including Bayesian NMF [21], in which a Markov chain Monte Carlo (MCMC) method is derived for estimating the posterior density, based on a Gibbs sampling procedure; Bayesian spectral decomposition (BSD) [22], which uses an atomic point-mass prior and MCMC methods to sample the solution space; Bayesian non-negative source separation [23] that incorporates a hybrid Gibbs-Metropolis-Hastings sampling procedure; and Bayesian inference for NMF models [15], which minimizes a Kullback-Leibler (KL) divergence with a hierarchical generative model consisting of an observation and a prior component, in which a variational Bayes algorithm and a Gibbs sampler are used for inference. In the current study, we focus our attention in the variational Bayes implementation of the latter approach.

A. Bayesian inference for Non-negative Matrix Factorization

Standard NMF methods [6], [7] decompose the data matrix **X** into two non-negative matrices **S** (the sources) and **A** (the mixing matrix). The differences between **X** and **SA** is given by the different cost functions used for measuring the divergence between them. In the particular variant of Bayesian NMF used in this study, proposed in [15], the author uses the terms templates (**T**) and excitation matrix (**V**) to define the model, in which $\mathbf{X} \approx \mathbf{TV}$. The joint probability distribution for the model is given by:

$$P(\mathbf{X},\mathbf{T},\mathbf{V}\,|\mathbf{\Theta}) = P(\mathbf{X}|\mathbf{T},\mathbf{V})\; P(\mathbf{T}|\,\mathbf{\Theta}^t)\; P(\mathbf{V}|\,\mathbf{\Theta}^v) \qquad (1)$$

Where $P(\mathbf{X}|\mathbf{T},\mathbf{V})$ is the likelihood, which is defined by a Poisson distribution since the model is minimizing the KL divergence between $\mathbf{X}$ and $\mathbf{TV}$. $P(\mathbf{T}|\mathbf{\Theta}^t)$ and $P(\mathbf{V}|\mathbf{\Theta}^v)$ are the model priors and are defined as Gamma distributions to enforce real positive values. $\mathbf{\Theta}^t$ and $\mathbf{\Theta}^t$ are the hyperparameters of the prior distributions over $\mathbf{T}$ and $\mathbf{V}$, respectively. In this study we use the variational variant proposed in [15], that defines a lower bound function over the evidence. For more details in the formulation, please refer to [15].

B. Relevant sources determination

To determine the relevant sources, we take advantage of the ability of the model to favor sparse representations by controlling the hyperparameters of the priors. The Gamma distributions Ga(x; a, b/a) that define the priors have shape a and scale b/a. The benefit of this representation is that we can control the sparsity of the model. For small values of a, most of the coefficients will be very close to zero and only very few will be dominating, hence enforcing sparsity. The parameter b is adapted to give the expected magnitude of each component.

By using a greedy strategy during the iterative process of matrices decomposition of NMF, we propose: i) to discard sources where the corresponding columns in the mixing matrix are zero (or a very small value) in all of their entries (which indicates that these sources are irrelevant or meaningless); and ii) where two sources are highly correlated (>0.98) between them (which suggests that both of them are representing the same kind of information), to discard one of these sources.

C. Experimental settings

Experiments were carried out for three different brain tumor diagnostic problems using simulated data and real world data from MRS acquired at STE. In each of these classification problems, we attempted to discriminate between two or three tumor types and healthy tissue, namely A2, GL, NO; A2, MM, NO; and A2, GL, MM, NO. All parameters used for the generation of the simulated data mirror the real data as closely as possible. The aim of using simulated data is to be able to test the proposed approach against known ground truth, that is, to evaluate to what extent the proposed method is able to estimate the correct number of sources.

From a data analysis point of view, the choice of these specific problems is meant to assess the ability of the proposed method to calculate a meaningful and small range of sources in problems that involve: i) infiltrating tumors, high-grade malignant tumors, and normal brain (A2, GL, NO); ii) infiltrating and non-infiltrating tumors, and normal brain (A2, MM, NO); and iii) infiltrating tumors, high-grade malignant tumors, non-infiltrating tumors, and normal brain (A2, GL, MM, NO).

As mentioned before, the Bayesian NMF variant used in this study formulates the decomposition of the dataset in terms of template and excitation matrices. These two matrices are interpreted in the current study as the sources and mixing matrix, respectively. The values of the hyperparameters were tied to a common Gamma distribution for each prior of the model, to reduce model complexity; and the values of the *shape* hyperparameters were chosen to be small, to encourage sparsity.

For the simulated data, we made 100 tests for each of the 50 datasets generated per discrimination problem and level of mixing, for a total of 10,000 tests. In each test, we set the method to calculate the sources starting from k=10 (k: number of sources), and to discard the irrelevant ones according to the criteria explained before. For the real data, we also made 100 tests for each discrimination problem and started from k=10.

For each discrimination problem (and level of mixing in the case of the simulated data) we determined the solution that approximates best the dataset for each value of k. For this, we quantified the accuracy of data reconstruction using the normalized root mean squared error (nRMSE) between the original data matrix and the reconstructed data.

IV. Results

Figure 2 groups the results obtained from the tests with simulated data. These histograms show the distributions of the solutions obtained per final number of sources, after discarding the irrelevant ones, for each discrimination problem and level of mixing.

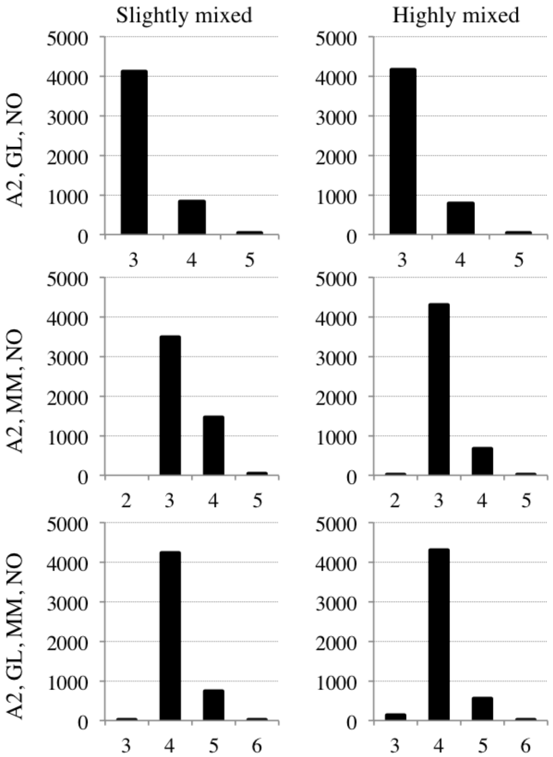

Fig. 2. Histograms showing the distribution of solutions per final number of sources corresponding to the simulated data, for each discrimination problem and level of mixing. Vertical axes represent the number of tests, while horizontal axes the number of final sources.

Similarly, Figure 3 shows the distribution of solutions per final number of sources when using the real data from the INTERPRET database.

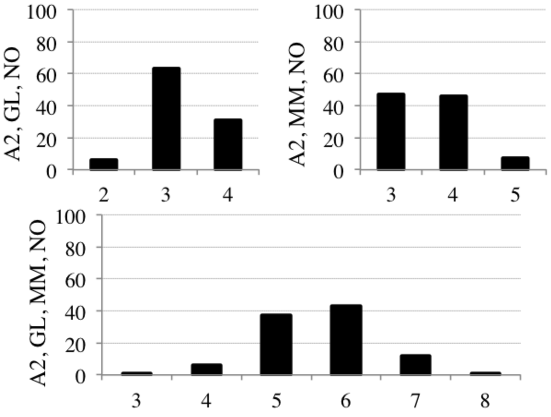

Fig. 3. Histograms showing the distribution of solutions per final number of sources corresponding to the real data, for the three discrimination problems. Axes are represented as in Figure 2.

Figures 4-9 show the set of sources obtained for the two most likely values of *k* (according to the histograms, see Figure 3) in the three discrimination problems studied with real data.

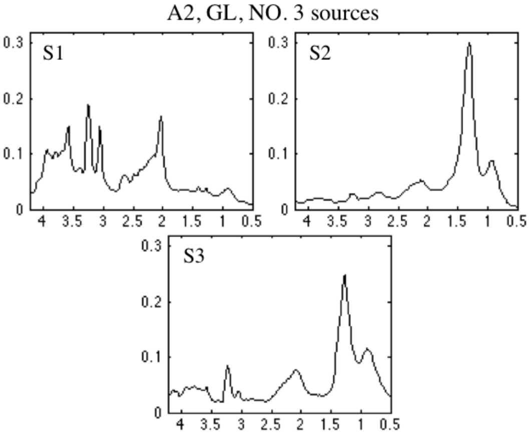

Fig. 4. A2, GL, NO. Sources from the k=3 solution that approximates best the dataset using real data. Axes are represented as in Figure 1.

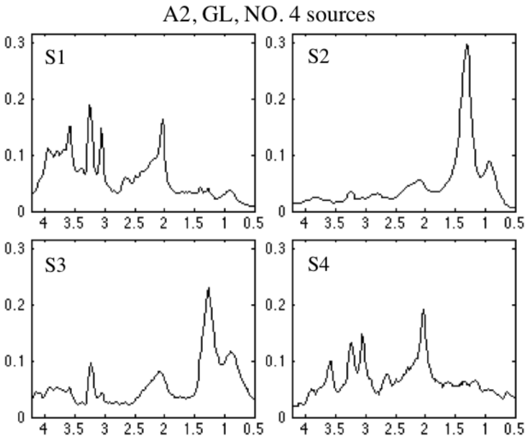

Fig. 5. A2, GL, NO. Sources from the k=4 solution that approximates best the dataset using real data. Axes are represented as in Figure 1.

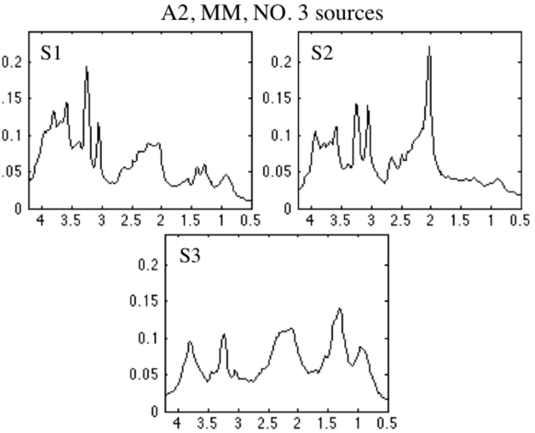

Fig. 6. A2, MM, NO. Sources from the k=3 solution that approximates best the dataset, using real data. Axes are represented as in Figure 1.

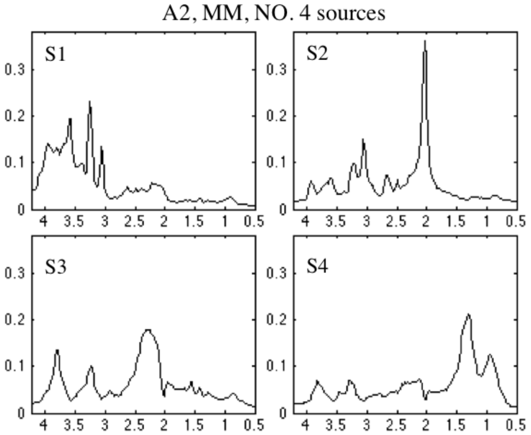

Fig. 7. A2, MM, NO. Sources from the k=4 solution that approximates best the dataset using real data. Axes are represented as in Figure 1.

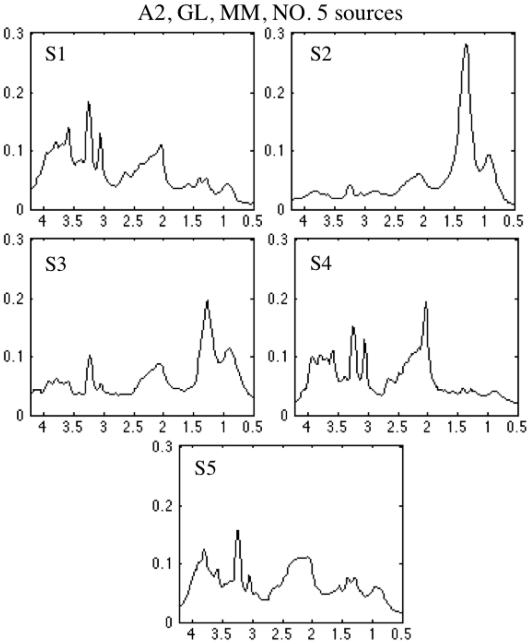

Fig. 8. A2, GL, MM, NO. Sources from the k=5 solution that approximates best the dataset using real data. Axes are represented as in Figure 1.

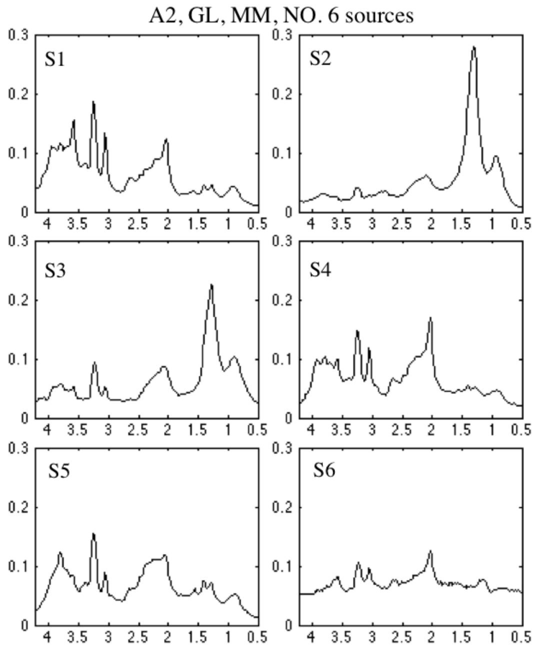

Fig. 9. A2, GL, MM, NO. Sources from the k=6 solution that approximates best the dataset using real data. Axes are represented as in Figure 1.

V. Discussion

From the results with simulated data presented in Figure 2 we can see that, empirically, the most likely solution for each discrimination problem (according to the number of sources with highest value in each histogram) matches the number of true sources from the corresponding dataset. That is, the solutions for the discrimination problems A2, GL, NO and A2, MM, NO, are more likely to have 3 sources; and the solution for A2, GL, MM, NO, is more likely to have 4 sources. Also, the quality of the sources obtained was as good as it can be with a standard version of NMF (see previous studies with standard NMF in [12]), and they resemble the representative examples (true sources) used to generate the datasets (the sources were not shown here for the lack of space). Hence, the simulated data provided us with the opportunity to corroborate that the proposed approach is able to retrieve the original, true sources.

In the case of the real data, we know from previous studies (e.g. in [12]) that the number of underlying sources is not necessarily the same as the number of tissue types involved in each classification problem. Also, depending on the particular question to address, we may find different meaningful solutions that involve the same tissue types. Therefore, for discrimination problems involving GL we expected to obtain two sources that describe this tumor type (when using the real data), as we know that there are cases of this type with or without necrotic lipids. For this reason, when the goal is to discriminate between tissue types, the expected solution for a problem like A2, GL, NO would have 4 sources. In Figure 3 we can see that the two most likely solutions for this problem are for k=3 and k=4. The solution for k=3 (Figure 4) is mainly separating the necrosis (S2 and S3) from non-necrosis (S1); and the solution for k=4 is the one that provides the separation from the tissue types involved (S2 and S3 explaining the GL, S1 the A2, and S4 the NO). The solution for k=2, not shown here, is interpreted similarly to the solution for k=3, as the sources look like S1 and S2.

In the discrimination problem A2, MM, NO, the number of solutions for k=3 and k=4 are very close (Figure 3), indicating that both models are very likely to occur. The k=3 solution (Figure 6) provides sources that resemble the tissue types involved (S1: A2, S2: NO, S3: MM); and the k=4 solution (Figure 7) is fairly similar, but in the latter the MM is being represented by two sources (S3 and S4), indicating that source S3 intend to represent grade I MM, while source S4 would represent grade II (atypical meningioma) contribution, known to display some mobile lipid content at short echo time. The analysis of the results for the discrimination problem A2, GL, MM, NO is similar to the previous two. The most likely solutions (according to the histogram in Figure 3) are those having 5 or 6 sources (Figures 8 and 9, respectively), and both have a sensible interpretation in terms of spectroscopy. For 5 sources S2 and S3 would originate from GL, S1 from A2, S4 from NO, and S5 from MM. Furthermore, for 6 sources, the additional S6 would suggest the contribution of an additional normal brain source. In this respect, the k=5 solution would seem optimal for the discrimination problem tackled in Figures 8 and 9.

VI. Conclusions

In this study we propose to take advantage of the ability of Bayesian NMF to favor sparse representation, to discard irrelevant sources during the iterative process of the training. This addresses one of the open questions in NMF, which is determining the ideal number of sources, for the particular problem of source extraction in brain tumors.

The obtained results show that the proposed approach is able to provide a small range of meaningful solutions to the problem of source extraction in human brain tumors, specifically in discrimination problems that involve infiltrating tumors, high-grade malignant tumors, non-infiltrating tumors, and normal brain.

Acknowledgment

This work was partly funded by MINECO grant MARESCAN (SAF 2011-23870) and Centro de Investigación Biomédica en Red – Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), an initiative of the Instituto de Salud Carlos III (Spain) co-funded by EU FEDER funds. S. Ortega-Martorell is funded by the European Union under the 7th FP, “Marie Curie” FP7-PEOPLE-2012-IEF.

References

[1] P. L. Lee and R. G. Gonzalez, “Magnetic resonance spectroscopy of brain tumors,” Curr. Opin. Oncol., vol. 12, no. 3, pp. 199–204, 2000.

[2] M. Julià-Sapé, D. Acosta, C. Majós, A. Moreno-Torres, P. Wesseling, J. J. Acebes, J. R. Griffiths, and C. Arús, “Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database,” J. Neurosurg., vol. 105, no. 1, pp. 6–14, 2006.

[3] N. A. Sibtain, F. A. Howe, and D. E. Saunders, “The clinical value of proton magnetic resonance spectroscopy in adult brain tumours,” Clin. Radiol., vol. 62, no. 2, pp. 109–119, 2007.

[4] J. Luts, A. Heerschap, J. A. K. Suykens, and S. Van Huffel, “A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection,” Artif. Intell. Med., vol. 40, no. 2, pp. 87–102, 2007.

[5] P. J. G. Lisboa, A. Vellido, R. Tagliaferri, F. Napolitano, M. Ceccarelli, J. D. Martin-Guerrero, and E. Biganzoli, “Data Mining in Cancer Research,” IEEE Comput. Intell. Mag., vol. 5, no. 1, pp. 14–18, 2010.

[6] P. Paatero and U. Tapper, “Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values,” Environmetrics, vol. 5, no. 2, pp. 111–126, 1994.

[7] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” Nature, vol. 401, no. 6755, pp. 788–791, 1999.

[8] Y. Su, S. B. Thakur, K. Sasan, S. Du, P. Sajda, W. Huang, and L. C. Parra, “Spectrum separation resolves partial-volume effect of MRSI as demonstrated on brain tumor scans,” NMR Biomed., vol. 21, no. 10, pp. 1030–1042, 2008.

[9] P. Sajda, S. Du, T. R. Brown, R. Stoyanova, D. C. Shungu, M. Xiangling, L. C. Parra, and X. M, “Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain,” IEEE Trans. Med. Imaging, vol. 23, no. 12, pp. 1453–1465, 2004.

[10] S. Du, X. Mao, P. Sajda, and D. C. Shungu, “Automated tissue segmentation and blind recovery of 1H MRS imaging spectral patterns of normal and diseased human brain,” NMR Biomed., vol. 21, no. 1, pp. 33–41, 2008.

[11] S. Ortega-Martorell, P. J. G. Lisboa, R. V. Simões, M. Pumarola, M. Julià-Sapé, and C. Arús, “Convex Non-Negative Matrix Factorization for brain tumor delimitation from MRSI data,” PLoS One, vol. 7, no. 10, p. e47824, 2012.

[12] S. Ortega-Martorell, P. J. G. Lisboa, A. Vellido, M. Julià-Sapé, and C. Arús, “Non-negative Matrix Factorisation methods for the spectral decomposition of MRS data from human brain tumours,” BMC Bioinformatics, vol. 13, no. 38, 2012.

[13] S. Ortega-Martorell, H. Ruiz, A. Vellido, I. Olier, E. Romero, M. Julià-Sapé, J. D. Martín, I. H. Jarman, C. Arús, and P. J. G. Lisboa, “A Novel Semi-Supervised Methodology for Extracting Tumor Type-Specific MRS Sources in Human Brain Data,” PLoS One, vol. 8, no. 12, p. e83773, 2013.

[14] A. Vilamala, P. J. G. Lisboa, S. Ortega-Martorell, and A. Vellido, “Discriminant Convex Non-negative Matrix Factorization for the classification of human brain tumours,” Pattern Recognit. Lett., vol. 34, no. 14, pp. 1734–1747, 2013.

[15] A. T. Cemgil, “Bayesian Inference for Nonnegative Matrix Factorisation Models,” Comput. Intell. Neurosci., vol. 2009, p. Article ID 785152, 2009.

[16] M. Julià-Sapé, D. Acosta, M. Mier, C. Arús, D. Watson, and T. Interpret Consortium, “A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients,” Magn. Reson. Mater. Phy., vol. 19, no. 1, pp. 22–33, 2006.

[17] INTERPRET project (2000/01/01 - 2002/12/31), “International Network for Pattern Recognition of Tumours Using Magnetic Resonance. IST-1999-10310.” Funded under 5th FWP 1.1.2.-1.2.2, p. http://azizu.uab.es/interpret.

[18] A. R. Tate, J. Underwood, D. M. Acosta, M. Julià-Sapé, C. Majós, A. Moreno-Torres, F. A. Howe, M. van der Graaf, V. Lefournier, M. M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. Luc Bosson, M. E. Cabañas, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. R. Wilkins, B. A. Bell, C. Rémy, A. Heerschap, D. Watson, J. R. Griffiths, and C. Arús, “Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra,” NMR Biomed., vol. 19, no. 4, pp. 411–434, 2006.

[19] A. Vellido, E. Romero, F. F. González-Navarro, L. A. Belanche-Muñoz, M. Julià-Sapé, and C. Arús, “Outlier exploration and diagnostic classification of a multi-centre 1H-MRS brain tumour database,” Neurocomputing, vol. 72, no. 13–15, pp. 3085–3097, 2009.

[20] M. van der Graaf, M. Julià-Sapé, F. A. Howe, A. Ziegler, C. Majós, A. Moreno-Torres, M. Rijpkema, D. Acosta, K. S. Opstad, Y. M. van der Meulen, C. Arús, and A. Heerschap, “MRS quality assessment in a multicentre study on MRS-based classification of brain tumours,” NMR Biomed., vol. 21, no. 2, pp. 148–58, 2008.

[21] M. Schmidt, O. Winther, and L. Hansen, “Bayesian Non-negative Matrix Factorization,” in Proceedings 8th International Conference ICA, 2009, vol. 5441, pp. 540–547.

[22] M. F. Ochs, R. S. Stoyanova, F. Arias-Mendoza, and T. R. Brown, “A New Method for Spectral Decomposition Using a Bilinear Bayesian Approach,” J. Magn. Reson., vol. 137, no. 1, pp. 161 – 176, 1999.

[23] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret, “Separation of Non-Negative Mixture of Non-Negative Sources Using a Bayesian Approach and MCMC Sampling,” IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4133 –4145, 2006.