ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

A machine learning pipeline for supporting differentiation of glioblastomas from single brain metastases

Victor Mocioiu[1,5] , Nuno M. Pedrosa de Barros[2], Sandra Ortega-Martorell[3,5], Johannes Slotboom[2], Urspeter Knecht[2], Carles Arús[1,5], Alfredo Vellido[4,5] and Margarida Julià-Sapé[5,1]1

1- Universitat Autònoma de Barcelona- Departament de Bioquímica i Biologia Molecular Cerdanyola del Vallès, Barcelona 08193, Spain
2- University Hospital Bern - SCAN, Freiburgstrasse 4, 3010 Bern, Switzerland
3- School of Computing and Mathematical Sciences, Liverpool John Moores University, Liverpool, UK
4- Departament de Ciències de la Computació, Universitat Politècnica de Catalunya, Edifici Omega, Campus Nord, 08034, Barcelona, Spain
5- Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina CIBER-BBN, Cerdanyola del Vallès, Barcelona, Spain

Abstract. Machine learning has provided, over the last decades, tools for knowledge extraction in complex medical domains. Most of these tools, though, are *ad hoc* solutions and lack the systematic approach that would be required to become mainstream in medical practice. In this brief paper, we define a machine learning-based analysis *pipeline* for helping in a difficult problem in the field of neuro-oncology, namely the discrimination of brain glioblastomas from single brain metastases. This pipeline involves source extraction using k-Means-initialized Convex Non-negative Matrix Factorization and a collection of classifiers, including Logistic Regression, Linear Discriminant Analysis, AdaBoost, and Random Forests.

1 Introduction

Machine learning (ML) has become, over the last few decades, a provider tool for knowledge extraction in complex medical domains. In few domains is this truer than in oncology [1]. Most of these tools, though, are *ad hoc* solutions that lack the systematic approach that would be required to be accepted as routine medical practice.

This brief paper defines one such ML-based analysis pipeline for a challenging problem in the field of neuro-oncology, namely that of differentiating a glioblastoma (GBM) from a single brain metastasis (MET) mass. This is a critical problem because of the different clinical approaches required for optimal treatment. GBMs are known to have an infiltrative nature and may benefit from a supra-maximal resection volume

[1] This work was funded by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° ITN-GA-2012-316679

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

[2], in which the required resection region is larger than the abnormal area seen on Magnetic Resonance Imaging (MRI). On the other hand, METs tend to be more spatially circumscribed and effective tumour resection can be performed with narrower margins. Unfortunately, due to their radiological resemblance (necrosis, enhancement of the tumor periphery and edema), MRI does not provide sufficient discriminative power and, more often than not, a definitive diagnosis requires post-surgical histopathological analysis. However, a non-invasive procedure is preferred because of the possible morbidity caused by the biopsy or because the tumor is located in a critical area of the brain. Previous studies [3-6] have alternatively used a different magnetic resonance modality, namely Magnetic Resonance Spectroscopy (MRS). MRS provides metabolic (biochemical) information about the investigated tissue; it can be single-voxel (SV), where information is read from one volume of interest or multi-voxel (MV), using a grid/matrix of many contiguous SVs (actually, SV-like spectral vectors, spv). However, there are also various difficulties when dealing with this kind of data, several pertaining to the discriminatory problem itself such as small sample size and class imbalance, and others to the intrinsic nature of MRS, such as high dimensionality. Previous studies have tackled these issues by employing dimensionality reduction techniques [4] and/or robust learners [3].

The proposed pipeline for this difficult clinical task involves the combined use of Convex Non-negative Matrix Factorization (cNMF) for MRS source extraction and a collection of statistical and ML classifiers, including Logistic Regression (LR) in different variants, Linear Discriminant Analysis (LDA), AdaBoost, and Random Forests (RF). We report on the performance of the pipeline and argue that, if validated, our methodology could also serve for improved resection strategies. We aim to set this as an example of employing currently available techniques to build a system that works outside the "comfort zone" of neatly curated databases. This is because, beyond the data issues outlined above, data acquisition in this field may often be suboptimal for circumstantial reasons (e.g., a tumor located close to the skull) or due to technical errors. The remainder of this paper is structured as follows: Section 2 describes the dataset and the pipeline construction, Section 3 provides an assessment of the pipeline performance, and finally Section 4 provides conclusions and future research directions.

2 Materials and methods

2.1 Materials

The analyzed data were acquired at 1.5T and long echo time (LTE, 135 ms) at Inselspital, Bern, Switzerland. The dataset comprises 48 MV grids (a total of 8,720 spv), 32 of which came from GBM patients (6,442 spv) and the remainder 16, from MET patients (2,278 spv). Echo time is an important parameter (can be either short/STE or long) as it determines the types of metabolites that can be observed in

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

the spectrum. LTE was chosen for this particular study because lactate, which can be regarded as a marker of infiltration [7], is less obstructed by the lipid peak than in STE. Furthermore, it is also known that LTE acquisitions are less prone to baseline distortions. All data were preprocessed as in [3]. Only the physiologically relevant MRS frequency interval between [4.20-0.50] parts per million (ppm) was used, thus giving us a vector of 195 intensity values for the analysis of each spectrum.

2.2 **Machine learning pipeline description**

The aim of the proposed pipeline is to i) predict whether a new case is a GBM or a MET and ii) provide a metabolic/nosologic map[2] that can be of practical use to the radiologist/neurosurgeon. Given the reduced number of available cases, a leave-one-out (LOO) cross-validation loop was used for the evaluation of results.

cNMF [8] is first applied to the training set, extracting three basis vectors, henceforth called sources. This choice is based on the hypothesis that cases will mainly consist of three tissue types – normal brain, necrotic tumoral core and infiltrative tissue (mostly in the GBM). cNMF was initialized using k-means, which was in turn initialized using the k-means++ algorithm [9]. This was followed by normalization of each row of the mixing matrix; then, using the normalized mixing matrix, three grids were built for each patient. Each element of a grid represents the contribution of the respective source to an individual voxel, i.e. the 195 intensity values of the corresponding spectrum. Every grid was then transformed into a binary image, via Otsu thresholding [10]. From the untransformed grids we extract mean, standard deviation, skewness and kurtosis; from the binary images we extract the perimeter to area ratio, eccentricity, filled area, major axis length, minor axis length, solidity and orientation.. This results in 33 features per patient (11×3 sources).

Using these transformed data, five classifiers, including linear and nonlinear types were built: LR, LR with a stratified 3-fold cross-validation loop to choose the regularization value (LRCV), LDA, AdaBoost, and a RF. LDA was used as a baseline to gauge to what extent the classes were linearly separable. The penalty norm used for LRCV was ℓ^2. Fifty decision trees were used as base estimators for AdaBoost, while ten estimators were used for RF; bootstrap samples of the training set were used to build the trees and the maximum number of features considered when looking for the best split was the square root of the total number of features. All nodes were expanded until all leaves were pure. For the remaining test patient, the three previously extracted sources were kept and non-negative least squares was used to estimate the corresponding mixing matrix. This procedure was repeated for feature extraction and feed the result to the classifiers. ROC curves were obtained for all classifiers and optimal thresholds were chosen according to Youden's index (YI) [11].

[2] Graphical representation of the spatial distribution of a given pathology.

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

3 Results

The cNMF source extraction results, reported in Figure 1, were very stable (that is, regardless of the fold, the extracted sources remained similar). The source representing normal brain (top row, right) was exactly the same for all folds; the necrotic source (top row, left) showed slight variation only in the lipids peak ([1.28:0.9] ppm); the infiltrative source (bottom row) also exhibited slight variation for the inverted lactate peak around 1.25 ppm.

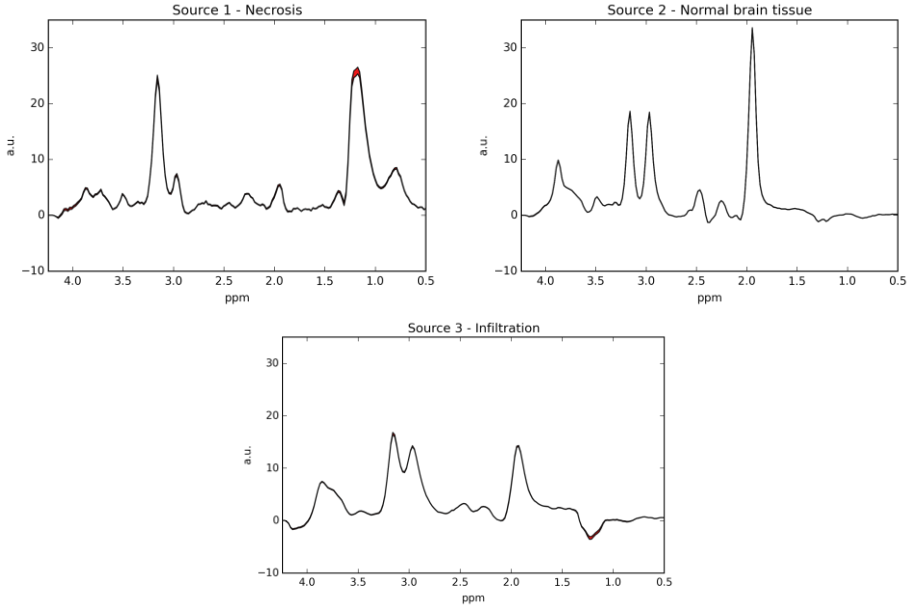

Fig. 1: Average amplitude (n = 48) of the three extracted sources: top row left) necrosis, right) normal brain tissue; bottom row) infiltration. The (rather small) standard deviation across all folds is represented as an envelope of the solid line.

Table 1 displays several classification metrics, including accuracy, sensitivity, specificity, mean F1-score and balanced error rate (BER) for the optimal threshold, for each individual classifier. The best threshold was found to be 0.5 for LDA and LR and 0.6 for LRCV, AdaBoost, and RF. The values for YI can also be found in Table1. ROC curves and AUCs for all classifiers can be found in Figure 2 (GBM was taken as the positive class).

4 Discussion and conclusion

From the previous results, it is worth highlighting that the RF performs quite consistently across all metrics and far better than any of the alternative classifiers,

ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2016, i6doc.com publ., ISBN 978-287587027-8.
Available from http://www.i6doc.com/en/.

coming first or at worse second-best (specificity). It does so at the price of unbalanced sensitivity/specificity, as GBM cases are far better classified than MET.

In fact, an interesting thing to note is that our pipeline has more difficulty in correctly classifying METs. This could be because of class imbalance, or maybe because the pipeline focuses on infiltration and it is known that some subtypes of MET can also present an infiltration pattern [12]. Besides diagnostic classification, the first stage of the proposed pipeline could also be used as a guide for surgery planning, by creating an RGB image from the values in the cNMF mixing matrix, which could be used as a map descriptor for tumor infiltration.

A previous study by Wijnen *et al*. [6] that also used MV MRS data on the same problem, reported similar results (AUC = 0.91) with an LDA classifier that uses three peak ratios as features. However, their methodology implies manual voxel selection based on the co-registered MRI, whereas ours is independent of the MRI. Furthermore their sample size is smaller (15 GBM patients and 15 MET patients) and with fewer total spv per case (270 for GBM and 195 for MET), thus perhaps leading to overly optimistic results.

Classifier (thr./YI)	Accuracy	Sensitivity	Specificity	F_1 Score	BER
LR (0.5/0.28)	0.70	0.84	0.43	0.76	0.33
LRCV (0.6/0.43)	0.79	0.93	0.5	0.78	0.20
RF (0.6/0.71)	**0.89**	**0.96**	0.75	**0.89**	**0.09**
LDA (0.5/0.40)	0.70	0.71	0.68	0.71	0.31
AdaBoost (0.6/0.59)	0.79	0.78	**0.81**	0.80	0.22

Table1: Optimal (according to the threshold selected by YI) performance metrics for the investigated classifiers. Best results highlighted in bold.

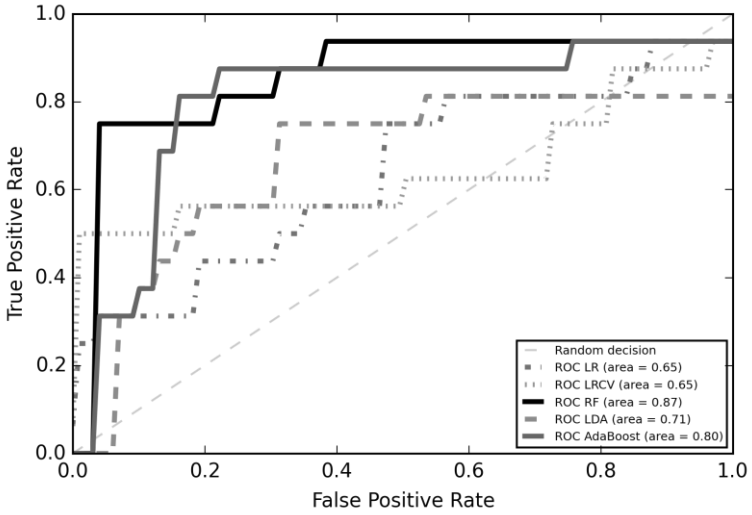

Fig. 2: ROC curves for the investigated classifiers.

The reported analyses have the limitation of using data from a single clinical center; they include, however far more spectra than any similar study. All in all, the proposed ML-based pipeline uses well-established methods to provide support in the challenging task of differentiation between GBM and MET, potentially improving the performance of a neuroradiologist in this task. It has the advantage of being completely automatic in the sense that the analysis does not require direct model manipulation from a neuroradiologist.

References

[1] A. Vellido, E. Biganzoli and P. J. Lisboa, Machine learning in cancer research: implications for personalised medicine. In M. Verleysen, editor, *proceedings of the 16th European Symposium on Artificial Neural Networks* (ESANN 2008), d-side pub., pages 55-64, 2008.

[2] A. Giese, T. Kucinski, U. Knopp, *et al.*, Pattern of recurrence following local chemotherapy with biodegradable carmustine (BCNU) implants in patients with glioblastoma, *Journal of Neuro-Oncology*, 66:351-360, 2004.

[3] A. Vellido, E. Romero, M. Julià-Sapé, *et al.*, Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel 1H MRS, *NMR in Biomedicine*, 25:819-828, 2012.

[4] S. Ortega-Martorell, H. Ruiz, A. Vellido, I. Olier, E. Romero, M. Julià-Sapé, J. D. Martín, I. H. Jarman, C. Arús, P.J.G. Lisboa, A novel semi-supervised methodology for extracting tumor type-specific MRS sources in human brain data, *PLoS ONE*:e83773, 2013.

[5] K. S. Opstad, M. M. Murphy, P. R. Wilkins, B. A. Bell, J. R. Griffiths and F. A. Howe, Differentiation of metastases from high-grade gliomas using short echo time ^{1}H spectroscopy, *Journal of Magnetic Resonance Imaging*, 20:187-192, 2004.

[6] J. P. Wijnen, A. J. S. Idema, M. Stawicki, *et al.*, Quantitative short echo time ^{1}H MRSI of the peripheral edematous region of human brain tumors in the differentiation between glioblastoma, metastasis, and meningioma, *Journal of Magnetic Resonance Imaging*, 36:1072-1082, 2012.

[7] F. Hirschhaeuser, U. G. Sattler and W. Mueller-Klieser, Lactate: a metabolic key player in cancer, *Cancer Research*, 71:6921-6925, 2011.

[8] D. D. Lee and H. S. Seung, Algorithms for non-negative matrix factorization, *Advances in Neural Information Processing Systems*, vol.13, pages 556-562, 2001.

[9] D. Arthur and S. Vassilvitskii, k-means++: The advantages of careful seeding. In proceedings of the *18th annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, pages 1027-1035.

[10] N. Otsu, A threshold selection method from gray-level histograms, *Automatica*, 11:23-27, 1975.

[11] W. J. Youden, Index for rating diagnostic tests, *Cancer*, 3:32-35, 1950.

[12] B. G. Baumert, I. Rutten, C. Dehing-Oberije, *et al.*, A pathology-based substrate for target definition in radiosurgery of brain metastases, *International Journal of Radiation Oncology Biology Physics*, 66:187-194, 2006.