Interpreting response to TMZ therapy in murine GL261 glioblastoma by combining Radiomics, Convex-NMF and feature selection in MRI/MRSI data analysis

Luis Miguel Núñez
Centro de Investigación Biomédica en Red Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)
Cerdanyola del Vallès, Spain
nunez.vivero96@gmail.com

Margarida Julià-Sapé
CIBER-BBN and *Dept. de Bioquímica i Biologia Molecular*
Universitat Autònoma de Barcelona and
Institut de Biotecnologia i Biomedicina
Cerdanyola del Vallès, Spain
margarita.julia@uab.cat

Enrique Romero
Dept. of Computer Science, IDEAI Research Center
Universitat Politècnica de Catalunya
Barcelona, Spain
eromero@cs.upc.edu

Carles Arús
Dept. de Bioquímica i Biologia Molecular
Universitat Autònoma de Barcelona and
CIBER-BBN and
Institut de Biotecnologia i Biomedicina
Cerdanyola del Vallès, Spain
carles.arus@uab.cat

Alfredo Vellido
Dept. of Computer Science, IDEAI Research Center
Universitat Politècnica de Catalunya and
CIBER-BBN
Barcelona, Spain
avellido@cs.upc.edu

Ana Paula Candiota
CIBER-BBN and
Dept. de Bioquímica i Biologia Molecular
Universitat Autònoma de Barcelona and
Institut de Biotecnologia i Biomedicina
Cerdanyola del Vallès, Spain
anapaula.candiota@uab.cat

***Abstract*—Machine learning (ML) methods have shown great potential for the analysis of data involved in medical decisions. However, for these methods to be incorpored in the medical pipeline, they must be made interpretable not only to the data analyst, but also to the medical expert. In this work, we have applied a combination of feature transformation, selection and classification using ML and statistical methods to differentiate between control (untreated) and Temozolomide (TMZ)-treated tumour tissue from a glioblastoma (brain tumour) murine model. As input, we have used T2 weighted magnetic resonance images (MRI) and spectroscopic imaging (MRSI). Radiomics features have been extracted from the MRI dataset, while convex Non-negative Matrix Factorization (Convex-NMF) was used to extract sources from the MRSI dataset. Exhaustive feature selection has revealed parsimonious feature subsets that facilitate the expert interpretation of results while retaining a high discriminatory ability.**

***Index Terms*—radiomics, machine learning, preclinical glioblastoma, therapy follow-up, feature selection, convex-NMF**

I. Introduction

Medical decision making is a well-honed combination of human expertise and available medical evidence. Due to technological and scientific advance, though, medicine is swiftly becoming a data-centered discipline. From a point-of-care viewpoint, this is exemplified by the standardization of the use of electronic health and medical records and by the intensive monitorization of the patient.

This *datafication* process risks unbalancing the subtle combination described above by adding more weight to the data-based medical evidence part, but also becomes a unique opportunity for quantitative data analysis based on statistics and ML. It has been argued that ML and similar approaches should only be used as an expert companion in the human expert decision making process. This is because, beyond providing enhanced accuracy, medical decision support systems must comply with existing guidelines, medical ethics and current legislation [1].

It has been claimed that for ML and similar methods to be accepted, legally compliant and put to work in the medical domain, they must be interpretable for the medical expert,

so that the expert can explain, at least to some extent, the *reasoning* behind the algorithm's decision [2].

In this paper, we implemented an analytical pipeline for a preclinical problem that involves distinguishing between treated and control (untreated) mice bearing GL261 glioblastoma (GB) tumours in a noninvasive way using magnetic resonance data, both in the modalities of imaging (MRI) and spectroscopic imaging (MRSI). The ultimate goal is making that distinction interpretable for the radiology expert in the hope that such interpretation reveals something new about the problem itself. Previous research [3]–[5] proved that the metabolomics pattern provided by MRSI could distinguish between treated/responding and control murine GL261 GB, but no attempt was made to investigate whether using MRI data could improve results. Accordingly, the data available in this study included both MRI and MRSI inputs. The pipeline includes two different data transformations: the regions of interest (RoI) of the images (see section II) were quantified using Radiomics [6], while the MRSI spectra were transformed using Convex-NMF [7]. Transformed data subsequently underwent feature selection, in the hope of finding an as parsimonious as possible selection of transformed features that both accurately distinguishes between treated tumour and control and remains interpretable for the radiologist.

II. Materials and methods

A. Data

The data used in this paper were acquired from 30 mice, in which tumours were induced by stereotactical injection of GL261 GB cells. Each individual receives a unique code (CXXX, being XXX a sequential number). Mice were divided into control subjects (untreated, $n = 13$) and TMZ-treated mice ($n = 17$) (see [3] for details on TMZ administration route and schedule). Therapy started eleven days after tumour generation. Tumour volume evolution was followed by MRI, and MRSI was acquired at chosen time points. Mice were euthanized and histopathological validation performed at endpoint or chosen datapoints, depending on the case. All studies were approved by the local ethics committee (*Comissió d'Ètica en l'Experimentació Animal i Humana* (CEEAH)), according to the regional and state legislations.

The data included both MRI and MRSI information. For this work, the available MRI was a $T2$ weighed ($T2w$) acquisition of the mouse brain, producing an image of 256×256 pixels. This was the basis for choosing the most suitable zone for MRSI acquisition. The whole MRSI grid is composed by a set of 32×32 voxels, that can be treated as *pixels* in the analytical pipeline, over the MRI image where each element of the grid is associated with one spectrum (Fig. 1). A smaller grid of 10×10 was defined as the Volume of Interest (VoI) to optimize and acquire the MRSI data, covering mostly the tumour/peritumoral zone, and was used as the input for this part of the study. Before applying the designed analytical pipeline, raw spectra were processed with 3D Interactive Chemical Shift Imaging software (version 1.9.11, Columbia University) and normalized with Dynamic MRSI Processing Module (DMPM, GABRMN-*Universitat Autònoma de Barcelona* [8]), while MRI was used without any further transformation. In summary, for each of the 30 mice, we had a 256×256 MRI image that matched the MRSI acquisition slice, and 100 spectra distributed in a grid of 10×10 covering most of the tumoral zone.

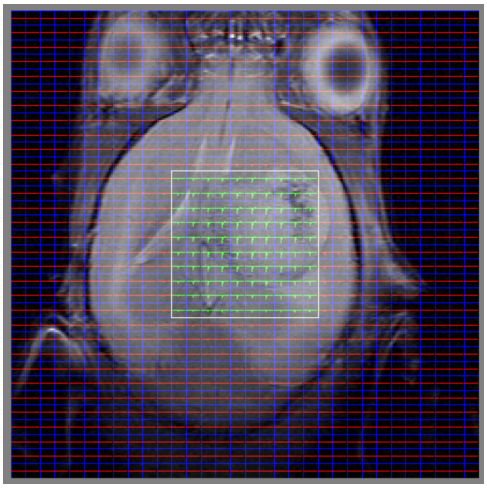

Fig. 1. Example of $T2w$ MRI image and the 32×32 grid superimposed over it from case C234. Green spectra belong to the VoI (yellow square) of 10×10 enclosing data used in this work. The rest of spectra are in red. Blue lines separate all grid pixels.

B. Radiomics

Radiomics is an emerging technique used in radiology that aims to extract quantitative information features from medical images that potentially bear descriptive and predictive capabilities [9]. This type of feature extraction or transformation is able to predict characteristics of different type of diseases and has already been applied to GB for survival prediction and patient stratification [10]. Radiomic features are calculated over a whole Region of Interest (RoI) that can be in 2D or 3D [9].

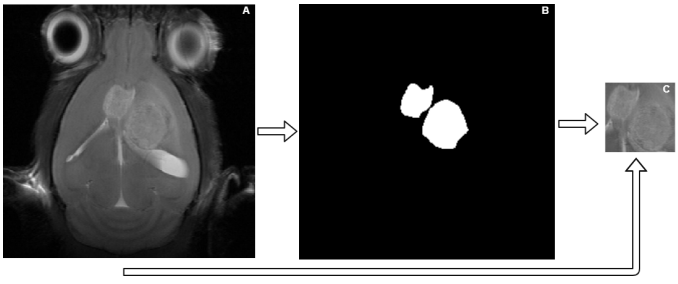

Fig. 2. Example of RoI extraction in a mouse (case C179). A: Raw image of the $T2w$ MRI. B: Mask of the manually performed tumour segmentation, used for the calculation of the Minkowski functions. C: Smallest rectangle of the MRI image that contains the whole mask of image B, used for texture matrices and features calculation.

In this study, information to distinguish whether mice belong to the control or TMZ-treated groups comes mostly from the tumor tissue, reducing the region used to extract

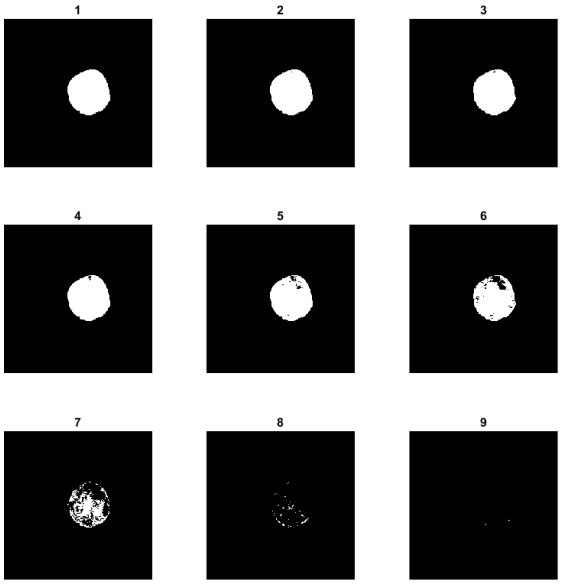

Fig. 3. Example of a 9 levels Minkowski threshold over a tumor mask (C234).

the Radiomics metrics and thus avoiding the inclusion of peritumoral/normal tissue. To achieve that, a manual segmentation of the tumor was performed (See Fig. 2) with expert supervision. Two types of Radiomic features were extracted to amplify the available mineable information: texture features and Minkowski functions.

First, 42 texture features were obtained using a Radiomics MATLAB toolbox [11] that extracts them from four texture matrices (See Table I), namely: Gray Level CoOcurrence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM) and Neighbouring Gray Tone Difference Matrix (NGTDM). In order to obtain texture matrices, an initial rectangular matrix is required as input. Since the segmented tumor does not have this shape, an additional step of automatic selection of the smallest rectangle that contains the whole tumor region was applied.

TABLE I
FEATURES EXTRACTED FROM EACH TEXTURE MATRIX

GLCM	Energy, Contrast, Entropy, homogeneity, Correlation, SumAverage, Variance, Dissimilarity, Autocorrelation
GLRLM	SRE, LRE, GLN, GLNN, RLN, RLNN, RP, LGRE, HGRE, SRLGE, SRHGE, LRLGE, LRHGE, GLV,RLV
GLSZM	SZE, LZE, GLN, ZSN, ZP, LGZE, HGZE, SZLGE, SZHGE, LZLGE, LZHGE, GLV, ZSV
NGTDM	Coarseness, Contrast, Busyness, Complexity, Strength

Second, Minkowski functions were obtained directly from the segmented tumor mask. They are morphological and structural descriptors of image heterogeneity that can characterize the tumor [12], [13]. A number *N* of levels is selected, and *N* auxiliary binary images are created with *N* thresholds that are equally spaced along the range of intensities of the mask (See example in Fig. 3). The area, perimeter and Euler parameter of the binarized object are then calculated, resulting in $3N$ features. From now on, these features will be referred to as Area#, Perimeter# or Euler#, being # the number of the threshold level associated to that feature. This transformation was performed using a specific MATLAB toolbox [14] and a range of different values of N was explored, settling for a value of 16 and, therefore, generating 48 features. The whole Radiomics set (texture + Minkowski functions) thus consisted of 90 features.

Data transformation using Radiomics still yields a large number of features and, accordingly, there would be risk of data overfitting [6]. This risk may be even higher in settings with a low cases-to-features ratio, such as the one described in this work. One way to decrease such risk is through dimensionality reduction in the form of feature selection, as described in the following sections.

C. Convex-NMF

Although Radiomics and Minkowski functions are suitable approaches for tumour characterization, they are not directly applicable to MRSI. This is because MRSI, although acquired in a spatial way, is a group of spectral vectors with metabolomic information from which texture matrices cannot be extracted. These spectra are complex signals and it can be assumed that they are the result of the combination of different sources (*paradigmatic* spectra). Accordingly, methods of blind source separation are suitable to extract those sources as a method of data transformation.

In this study, we used Convex-NMF [7], an unsupervised method for matrix factorization that extracts individual sources from a signal that results from a combination of those sources through a mixing matrix. This variant of NMF allows non-negative components in both source and mixing matrices.

NMF and Convex-NMF are not new to the neuro-oncology domain, where they have been used to differentiate abnormal masses [15]; to distinguish non-tumoral, responding tumoral and non responding tumoral tissue in glioblastoma through source extraction in a semi-supervised way [16]; for tumour type classification [17]; or for spectral data quality control [18], to name just a few applications.

Here, as in the Radiomics-based approach, only information extracted from tumor regions were considered and only spectra that represent completely the tumor region were selected based on the manual segmentation of the tumor (see an example in Fig. 4). This reduces the initial 3,000 available spectra to a subset of 1,313.

Convex-NMF was therefore applied to the 1,313 spectra to extract 20 source vectors. These sources can be qualitatively analyzed by radiology experts in terms of different metabolite contribution. On the other hand, the mixing matrix may inform of how these sources are combined to create each spectrum. Source weights were used as features in the subsequent models.

D. Feature selection

High data dimensionality increases the risk of model overfitting, while irrelevant and redundant features increase the

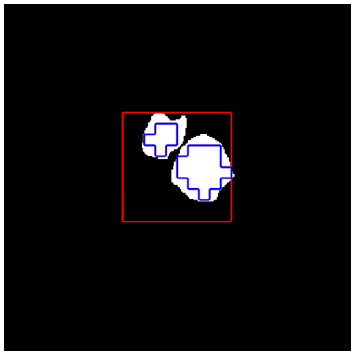

Fig. 4. Representation of selected spectra (blue contour) from the 10×10 grid (red) on the basis that they consist completely of tumor tissue from case C179.

search space and make the pattern detection difficult. This problem can be solved either with an exponential increase of the number of samples or with dimensionality reduction methods. Feature selection is an important step to extract more consistent knowledge from our data that are relevant to the problem, discarding irrelevant or redundant features [19].

Although the choice of the optimal feature selection algorithm is data-dependent, wrapper feature selection methods usually obtain good results, even though at the expense of a high computational cost. In the wrapper scheme, the accuracy of the classifier is used to gauge the relevance of a subset of features, unlike in the filter scheme, where the classifier is built only after the features are already selected. In the embedded scheme, the internal structure of the classifier allows to assign the importance of every feature. These three schemes can be applied with different search algorithms (exponential, sequential, etc). See [20] for further details.

In the present study, feature selection was used in two distinct situations. In MRI analysis, where there is a low cases-to-features ratio (30-90), feature selection must be performed in combination with a not-too-complex classifier to ensure model generalization and reliability. This is different from the source extraction analysis in MRSI. In this case, even if the data dimensionality does not force us to apply feature selection, it should lead us to a better quantitative and qualitative understanding of which sources are most responsible for the differentiation of control and treated cases.

Similar feature selection approaches were applied to both pipelines. First, a univariate t-test allowed to establish a rough rank of their singular importance in the separation of control and treated cases, following a filter scheme [21]. In a second experiment, we applied a recursive feature elimination (RFE) search scheme which starts with the whole set of features and removes one feature (the one considered less relevant) at every step. For the MRSI data, a standard wrapper RFE was applied. For the MRI data, a wrapper RFE did not allow to distinguish the relevance of the features (most accuracies were equal in the first steps) and, therefore, an embedded RFE with linear classifiers was performed instead, and over the selection of the embedded method in another step a wrapper feature selection is applied aiming to reduce even more the size of the subset. In this scheme, the relevance of a feature was computed as follows. Suppose that we have trained K linear classifiers with different data partitions (as in a K-fold cross validation), so that every classifier can be expressed as $f_j(x) = g\left(\sum_{i=1}^{N} \omega_{ij} x_i\right)$, where N is the number of features and g is some monotonic function (such as the linear function or the logistic function for the logistic regression classifier). Then, the relevance of a feature i can be expressed as the mean of $\{|\omega_{ij}|\}_{j=1}^{K}$. This idea is based on the hypothesis that irrelevant features produce smaller variations in the output values than relevant ones. Hence, a natural way to compute the relevance of a feature i in the trained model $f_j(x)$ is to compute the absolute value of the derivative of $f_j(x)$ with respect to x_i, which is $\left|g'\left(\sum_{i=1}^{N} \omega_{ij} x_i\right) \omega_{ij}\right|$. If we only want to compare the relevance between two features, we can ignore the term $g'\left(\sum_{i=1}^{N} \omega_{ij} x_i\right)$.

III. Experiments

In a first stage of the analysis, the relevance of MRI and MRSI features was ranked according to a univariate t-test and preliminary classified using LDA, Logistic Regression, SVM and KNN models, using a grid search for finding the optimal parameters. In the MRI pipeline, results for subsets of features ordered according to the relevance ranking varying from 1 to 30 are presented (Table II); that is, the remaining features are directly discarded. In the MRSI pipeline, all 20 sources are considered in the ranking. Results are reported through Area Under the ROC Curve (AUC) values in a leave-one-out cross validation procedure. In a second stage, the multivariate and more exhaustive RFE approach described in section II-D was implemented.

TABLE II
Ranking of radiomics features according to a t-test.

1	GLCMEnergy	11	Perimeter10	21	GLCMHomo
2	GLCMEntropy	12	Perimeter6	22	GLRLMRLN
3	GLRLMGLN	13	Euler7	23	GLSZMZP
4	Perimeter9	14	GLRLMSRHGE	24	GLRLMRP
5	GLSZMGLN	15	GTDMContrast	25	Perimeter4
6	Perimeter8	16	Euler6	26	Euler2
7	GLCMVariance	17	Perimeter5	27	GLSZMSZHGE
8	GLRLMRLV	18	Euler12	28	GLRLMHGRE
9	Perimeter7	19	Euler5	29	GLCMAutoCorr
10	GLRLMGLNN	20	GLRLMSRE	30	GLSZMHGZE

A. Results

AUC classification values for different classifiers and for different feature subsets following the t-test ranking are shown in Tables III and IV and in Figs. 5 and 6 for MRI and MRSI, respectively.

In MRI studies, the performance of classifiers varied both with the number of features selected and the classifier used. It is worth noting that performance of linear and non-linear

TABLE III
AUC FOR SUBSETS OF RADIOMICS FEATURES ACCORDING TO t-TEST RANKING SELECTION

NFeatures	LDA	Log Regression	SVM	KNN
1	0.85	0.83	0.85	0.80
3	0.83	0.82	0.83	0.82
3	0.81	0.80	0.82	0.82
4	0.87	0.81	0.90	0.85
5	0.83	0.81	0.89	0.83
6	0.83	0.78	0.89	0.86
7	0.79	0.73	0.88	0.84
8	0.91	0.76	0.90	0.86
9	0.90	0.81	0.89	0.84
10	0.89	0.80	0.86	0.86
11	0.84	0.80	0.85	0.85
12	0.89	0.84	0.85	0.87
13	0.86	0.92	0.84	0.86
14	0.83	0.83	0.89	0.90
15	0.79	0.84	0.90	0.89
16	0.76	0.76	0.89	0.90
17	0.70	0.83	0.89	0.90
18	0.70	0.75	0.89	0.92
19	0.65	0.70	0.88	0.92
20	0.68	0.69	0.91	0.90
21	0.58	0.67	0.92	0.91
22	0.64	0.71	0.92	0.90
23	0.58	0.68	0.92	0.90
24	0.57	0.63	0.93	0.91
25	0.47	0.56	0.93	0.90
26	0.46	0.56	0.93	0.91
27	0.48	0.62	0.93	0.90
28	0.59	0.59	0.94	0.90
29	0.57	0.44	0.94	0.90
30	0.63	0.56	0.94	0.90

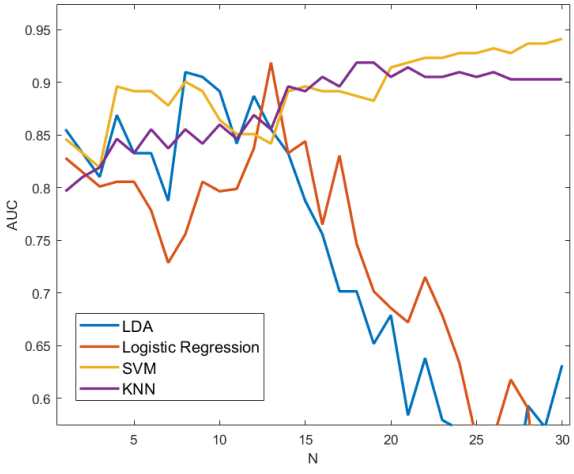

Fig. 5. Representation of different classifiers performance (AUC) depending on the number of Radiomics features selected from the t-test ranking.

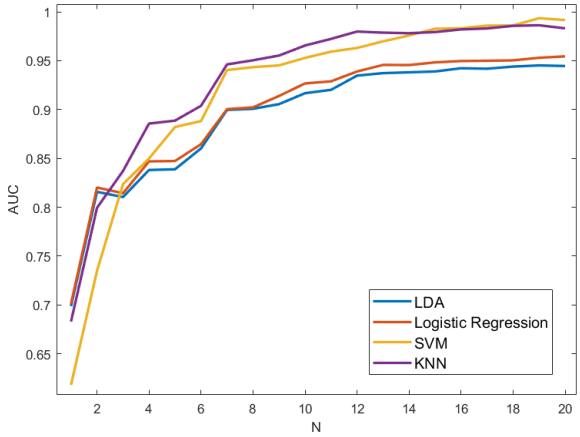

Fig. 6. Representation of different classifiers performance (AUC) depending on the number of Source features selected from the t-test ranking.

TABLE IV
AUC FOR SUBSETS OF CONVEX-NMF FEATURES ACCORDING TO t-TEST RANKING SELECTION

Nº Features	LDA	Log Regression	SVM	KNN
1	0.70	0.70	0.62	0.68
2	0.82	0.82	0.73	0.80
3	0.81	0.81	0.82	0.84
4	0.84	0.85	0.85	0.89
5	0.84	0.85	0.88	0.89
6	0.86	0.86	0.89	0.90
7	0.90	0.90	0.94	0.95
8	0.90	0.90	0.94	0.95
9	0.91	0.91	0.95	0.96
10	0.92	0.93	0.95	0.97
11	0.92	0.93	0.96	0.97
12	0.93	0.94	0.96	0.98
13	0.94	0.95	0.97	0.98
14	0.94	0.95	0.98	0.98
15	0.94	0.95	0.98	0.98
16	0.94	0.95	0.98	0.98
17	0.94	0.95	0.99	0.98
18	0.94	0.95	0.99	0.99
19	0.95	0.95	0.99	0.99
20	0.94	0.95	0.99	0.98

classifiers varied in a noticeable manner with the increase of the number of included features. The performance of linear methods such as LDA and Logistic regression deteriorates when the number of features increases beyond 10-15. On the other hand, the non-linear ones, SVM and KNN, improve their solution until reaching a maximum AUC value of 0.94.

In the MRSI analysis, all classifiers improve as the number of sources in the subset increases, with performance stalling at ca. 12 sources, even if the peak of performance is reached when employing 19 sources (0.95 AUC in linear classifiers and 0.99 in non-linear ones).

These preliminary results clearly indicate that we can dis-

criminate the classes with great accuracy. In the second experimental stage, we investigated the multivariate embedded-wrapper feature selection approach, in which the ultimate target is finding the most parsimonious selection of features retaining an optimal discrimination accuracy between control and treated cases. For the sake of brevity, only the leave-one-out results of the best performing models are presented. Feature selection was performed with KNN, Logistic Regression and SVM with polynomic kernel.

For the MRI radiomics data, KNN results were comparatively poor and are not presented here. Both Logistic Regression and SVM achieved 96.67% accuracy. Importantly, these results were obtained in the SVM for a polynomial of degree one, that is, a linear model. Logistic Regression with embedded feature selection retained a 96.67% accuracy (0.9729 AUC) with 9 features, namely Perimeter10, Euler1, Euler16, GLRLMSRE, GLRLMRLN, GLRLMSRHGE, GLRLMRLV, GLSZMSZLGE and GLSZMLZHGE and a 93.33% (0.9593 AUC) with just 7 of them: Perimeter10, Euler1, GLRLMSRE, GLRLMRLN, GLRLMSRHGE, GLRLMRLV, GLSZMLZHGE. Similar results (96.67% accuracy, 0.9683 AUC) were obtained with an alternative selection of 9: Area16, Perimeter10, Euler1, GLCMEnergy, GLRLMSRE, GLRLMSRHGE, GLRLMRLV, GLSZMSZLGE, GLSZMLZHGE, and 7 of them (93.33% accuracy, 0.9502 AUC): Perimeter10, Euler1, GLCMEnergy, GLRLMSRE, GLRLMSRHGE, GLRLMRLV, GLSZMLZHGE. In comparison, the performance of the linear SVM deteriorates quickly when removing features.

In a second-level process, wrapper feature selection was used to further reduce the dataset dimensionality, retaining the 96.67% accuracy (0.9638 AUC) with two subsets of features of size 6: Euler16, GLRLMRLN, GLRLMSRHGE, GLRLMRLV, GLSZMSZLGE, GLSZMLZHGE, or even only 5 (0.9638 AUC): Euler1, GLRLMSRE, GLRLMSRHGE, GLRLMRLV, GLSZMLZHGE. This hints to the existence of a resilient core of just 3 features: GLRLMSRHGE, GLRLMRLV, GLSZMLZHGE. Note that a 96.67% accuracy means that a single mouse has been misclassified. It happens to be always the same one: case C529, which the models consistently predict to be TMZ-treated when, in fact, it happens to be a control.

For the MRSI sources extracted with Convex-NMF, the first experiments hinted that non-linear classifiers performed better than linear ones. KNN, Logistic Regression, SVM with polynomic and Gaussian kernels were used. Given the data availability, 10-fold cross-validation was used this time. The SVMs clearly outperformed the other methods and only their results are presented. A polynomial SVM of 4^{th} degree and a Gaussian SVM both achieved a 97.72% accuracy (with AUC values of 0.9966 and 0.9943, respectively). A wrapper feature selection then increases the accuracy to 98.09% (0.9972 AUC) with 14 sources (shown in Fig. 7) using the polynomial kernel, and 98.10% (0.9973 AUC) with 17 sources using the Gaussian.

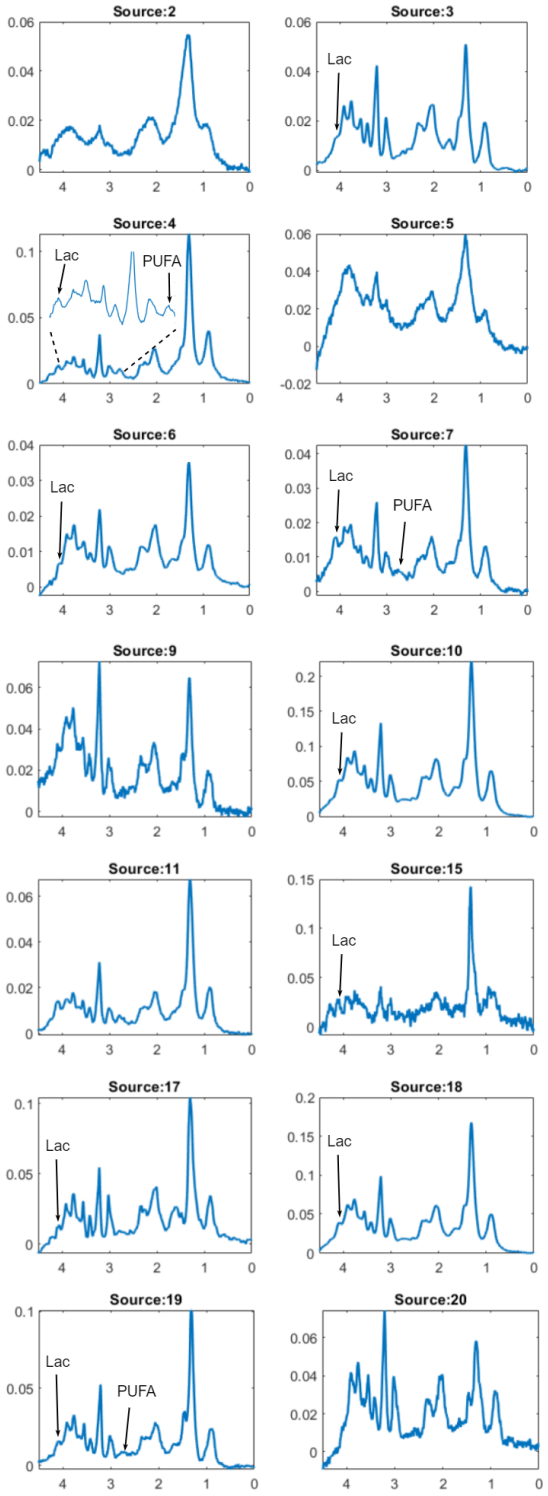

Fig. 7. Convex-NMF sources selected by RFE. Y axes: arbitrary units; X frequency axes: (parts per million, or ppm). Text inside plots, Lac: lactate, $PUFA$: Polyunsaturated Fatty Acids of some of the sources shown. In source 4, an expansion of the 4.1-2.8 ppm zone is shown for a better observation of the highlighted metabolites. See *Discussion* for more details

B. Discussion

Radiomics study

The study of tissue morphology is used in oncology to detect the presence of disease and also to follow treatment. The qualitative neuroimaging analysis is usually hampered by a lack of reproducible and objective measures, although approaches such as the VASARI features have been attempted [22]. In this sense, quantitative assessment of tissue morphology in an automated pipeline could be of great interest. MRI-based studies using radiomics are currently found to predict survival/progression, or follow response to therapy [6], [12], [23], [24].

Malignant brain tumours usually have an heterogeneous appearance in MRI, due to presence of necrosis and/or haemorraghic foci. Moreover, the architecture of the tissue can change upon treatment [28], [29]; at the histopathological level, transient response to TMZ, detected in our preclinical GB model under different protocols, leads to the appearance of giant cells, decrease of the proliferative rate and increase of acellular spaces [4]. Altogether, this may have a clear impact on MRI and its associated features.

Amongst the Radiomics features consistently selected for distinguishing treated from control tumours, we found GLRLM and GLSZM. The former indicates the coarseness of a texture, while the latter quantifies gray level zones in an image (the number of connected voxels sharing the same gray level intensity). Both measure, to a certain extent, aspects of tissue homogeneity, which may change upon therapy (Fig. 8). Moreover, changes in tumour properties such as ischemia, angioedema and avascular necrosis, might be more obvious on $T2w$ MRI, reinforcing the potential of this approach alone for a radiomics-based therapy response analysis. On the other hand, Minkowski functions have been described as useful in analyzing heterogeneity of peritumoral hyperintensity of glioblastoma, demostrating prognostic value in predicting survival [25].

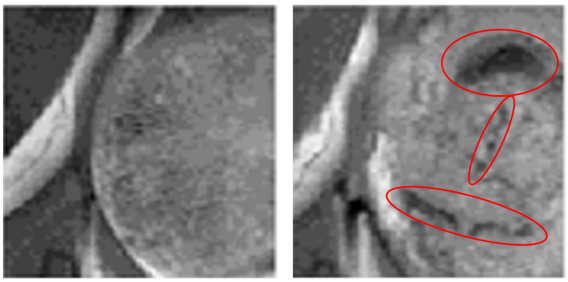

Fig. 8. Examples of control (left, mouse C583) and treated, transiently responding to TMZ according to histopathological parameters (right, mouse C574) murine GL261 GB tumours. Note the appearance of hypointense zones (red circles) in T2w MRI from the treated mouse, noticeably different from the more homogeneous appearance observed in the control case.

Recent results [5] suggest that treated GL261 tumours responding to TMZ treatment present significantly higher immune system elements within the tumour region. The local effects of these cells against the GB mass, may also cause local changes that could be spotted by MRI. This is supported by results described by others [26], in which radiomics features such as GLRLM correlate with histopathological features such as Ki67 in high grade glioma. Recent work [27] also described the value of GLRLM and GLSZM for evaluating response to therapy in GB, being able to distinguish pseudoprogression from true progression. Regarding case C529, consistently misclassified in our radiomics-based approach, its histopathological data suggest a slowly proliferating tumour with an apoptotic rate more similar to the average in treated cases [3]. This could partially explain the misclassification for this case, because such parameters have an impact in local tissue characteristics and, accordingly, in MRI.

One of the challenges still hampering proper comparison between different radiomic-based studies is the lack of standardization [30]. Due to the well known contrast-enhancement characteristics of GB tumours, some authors emphasize the potential of contrast-based imaging modalities [24], but since contrast agents are not recommended for some patients (e.g. with kidney diseases), a reliable method based on $T2w$ MRI would be of great interest for its translational potential.

MRSI study

MRSI analysis was centered in the tumour zone. Accordingly, the sources in Fig. 7 show basically the GL261 GB characteristic spectral pattern, such as high choline-to-creatine ratio, low N-acetyl containing compounds signal and high mobile lipids content (see [3] and [4]). Most of the extracted sources had good spectral quality, except by some borderline homogeneity (sources 2 and 5) and signal-to-noise ratio (source 15). This could be either due the presence of local haemorrhage or presence of large acellular spaces, not unexpected in this type of tumours, even more upon TMZ treatment. Sources shown in figure 7 represent the features mostly explaining differences between control and treated tumours. We have previously described [3] that TMZ-treated/responding cases presented differential spectral features with respect to control cases, especially polyunsaturated fatty acids (PUFA) at 2.8 ppm. Further differences were seen in lactate (Lac) at 1.3 and 4.1 ppm, mobile lipids (ML) at 0.9 ppm and myo-inositol/glycine (mI-Gly) at 3.55 ppm. The presence of PUFA is expected in treated-responding cases, indicating therapy-related metabolic changes such as apoptosis induced by treatment, as already described elsewhere [31]. Sources 4, 7 and 11 show noticeable signals compatible with PUFA. Sources 7 and 11 also show prominent signal from Lac at 4.1 ppm, clearly lower in sources 3 and 6. Since the approach and the number of sources extracted is different from the ones described in [3], direct comparison is not straightforward, but essentially sources 4 and 7 have an overall spectral pattern compatible with responding tumours such as described by Delgado *et al* [3]; this agrees with the transient response expected to take place with this preclinical model under such therapeutic approaches. On the other hand, sources 3 and 20 show mostly features compatible with control/untreated tumours.

IV. Conclusions

The accuracy achieved with the 6 to 9 radiomics features, based only in $T2w$ MRI was comparable with the accuracy obtained with MRSI using 14 to 17 sources. Features extracted from $T2w$ MRI are probably reflecting substantial changes in GB tumour tissue structure upon treatment (changes in perfusion, apoptosis, cell death, acellular spaces, presence of immune system elements). This radiomics-based approach could bear a great potential either in preclinical and for translational clinical studies, since $T2w$ MRI can be acquired at any available clinical scanner, is a fast acquisition (part of the pipeline of patients during treatment) and does not require any contrast administration. MRSI acquisitions are not widely implemented in all clinical centers/scanners and still need a consistent biochemical knowledge or sophisticated pattern recognition approaches in order to interpret output. The combination of data transformations and feature selection coupled with ML-based classification has allowed a first approximation to the characterization of the differences between treated and untreated cases, thus opening the doors of interpretability mediated by radiology experts in a problem that would have been almost untractable from direct data inspection.

Acknowledgments

L.M.N., A.V., C.A. and A.P.C. received EC funding from the ATTRACT project, under Grant Agreement 777222. A.V. and E.R. acknowledge funding from Spanish MINECO TIN2016-79576-R research project. MRI/MRSI preclinical data were acquired at Unit 25 of NANBIOSIS ICTS (www.nanbiosis.es).

References

[1] A. Vellido, "Societal issues concerning the application of artificial intelligence in medicine", Kidney Dis, vol. 5, pp. 11-17, 2019.

[2] A. Vellido, "The importance of interpretability and visualization in Machine Learning for applications in medicine and health care", Neural Comput. Appl., in press.

[3] T. Delgado-Goni, S. Ortega-Martorell, M.Ciezka, J. Olier, A.P. Candiota et al, "MRSI-based molecular imaging of therapy response to temozolomide in preclinical glioblastoma using source analysis", NMR Biomed., vol. 29(6), pp. 732-43, Jun 2016.

[4] N. Arias-Ramos, L. Ferrer-Font, S. Lope-Piedrafita, V. Mocioiu, M. Julià-Sapé et al, "Metabolomics of therapy response in preclinical glioblastoma: a multi-slice MRSI-based volumetric analysis for noninvasive assessment of Temozolomide treatment", Metabolites, vol. 7(2), pii: E20, 2017 May.

[5] S. Wu, P. Calero-Pérez, L. Villamañan, N. Arias-Ramos, M. Pumarola et al, "Anti-tumour immune response in GL261 glioblastoma generated by Temozolomide immune-enhancing metronomic schedule monitored with MRSI-based nosological images", NMR Biomed., 2020,in press.

[6] S. Narang, M. Lehrer, D. Yang, J. Lee, and A. Rao, "Radiomics in glioblastoma: current status, challenges and potential opportunities", Transl. Cancer Res., vol. 5(4), pp.383-397, 2016.

[7] C.H. Ding, T. Li, and M.I. Jordan, "Convex and semi-nonnegative matrix factorizations", IEEE T. Pattern Anal., vol. 32(1), pp.45-55, 2008.

[8] DMPM toolbox available at: "http://gabrmn.uab.es/dmpm", last accessed January 28th, 2020.

[9] V. Kumar, Y. Gu, S. Basu, A. Berglund, S.A. Eschirch et al, "Radiomics: the process and the challenges", Magn. Reson. Imaging, vol. 3(9), pp.1234–1248, 2012.

[10] P. Kickingereder, S. Burth, A. Wick, M. Gotz, O. Eidel et al, "Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models", Radiology, vol. 280(3),pp. 880-889, 2016.

[11] Radiomics Toolbox available at: "https://github.com/mvallieres/radiomics", last accessed January 28th, 2020.

[12] T.J. Larkin, H.C. Canuto, M.I. Kettunen, T.C. Booth, D.E. Hu et al, "Analysis of image heterogeneity using 2D Minkowski functionals detects tumor responses to treatment", Magn. Reson. Med., vol. 71, pp. 402-410, 2014.

[13] T. Matttfeldt, D. Meschenmoser, U. Pantle and V. Schmidt, "Characterization of mammary gland tissue using joint estimators of Minkowski functionals", Image Anal Stereol, vol. 26, pp. 13-22, 2007.

[14] Minkowski functions Toolbox available at: "https://es.mathworks.com/matlabcentral/fileexchange/33690-geometric-measures-in-2d-3d-images", last accessed January 28th, 2020.

[15] Y. Su,SB. Thakur ,K. Sasan ,S. Du and P. Sajda,"Spectrum separation resolves partial-volume effect of MRSI as demonstrated on brain tumor scans", NMR Biomed., vol 21, pp. 1030–1042, 2008.

[16] S. Ortega-Martorell, P.J.G. Lisboa , A. Vellido, RV. Simões and M. Pumarola, "Convex non-Negative matrix factorization for brain tumor delimitation from MRSI data", PLoS ONE, vol. 7 (10), 2012.

[17] 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(14), pp. 1734–1747, 2013.

[18] Y. Hernández-Villegas, S. Ortega-Martorell, C. Arús, A. Vellido and M. Julià-Sapé, "Extraction of artefactual MRS patterns from a large database using non-negative matrix factorization", NMR Biomed., in press.

[19] A. Vellido, E. Romero, M. Julià-Sapé, C. Majós, A. Moreno-Torres, and C. Arús, "Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel proton MRS", NMR Biomed., vol. 25(6), pp. 819–828, 2012.

[20] L. Carlos-Molina, Ll. Belanche and A. Nebot, "Feature selection algorithms: a survey and experimental evaluation", IEEE International Conference on Data Mining, 2002.

[21] C.L. König, M.I. Cárdenas, J. Giraldo, R. Alquezar and A. Vellido, "Label noise in subtype discrimination of class C G-protein coupled receptors: A systematic approach to the analysis of classification errors", BMC Bioinformatics, vol. 16(1),314, 2015.

[22] VASARI research project wiki. Cancer Imaging Archive. Available at: "wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project", (Updated 2015). Last accessed January 28th, 2020.

[23] T.C. Booth, T.J. Larkin, Y. Yuan, M.I. Kettunen, S.N. Dawson et al, "Analysis of heterogeneity in T2-weighted MR images can differentiate pseudoprogression from progression in glioblastoma", PLoS One, vol. 12(5), e0176528, 2017.

[24] N. Elshafeey, A. Kotrotsou, A. Hassan, N. Elshafei, I. Hassan et al, "Multicenter study demonstrates radiomic features derived from magnetic resonance perfusion images identify pseudoprogression in glioblastoma", Nat Commun., vol. 10,3170, 2019.

[25] Y. Choi, K.J. Ahn, Y. Nam, J. Jang, N.Y. Shin et al, "Analysis of peritumoral hyperintensity on pre-operative T2-weighted MR images in glioblastoma: Additive prognostic value of Minkowski functionals", PLoS One, vol. 14(5), e0217785, 2019.

[26] Li. Jing, S. Liu, Y. Qin, Y. Zhang, N. Wang et al, "High-order radiomics features based on T2 FLAIR MRI predict multiple glioma immunohistochemical features: A more precise and personalized gliomas management", PLoS One, vol. 15(1), e0227703, 2020.

[27] M.D. Patel, J. Zhan, K. Natarajan, R. Flintham, N. Davies et al, "Radiomic evaluation of treatment response in patients with glioblastoma: a preliminary study" ECR 2019, Poster C-2003.

[28] B.J. McCullough and J.W. Henson, "Issues in Response assessment of brain tumor chemotherapy". In: Handbook of brain tumor chemotherapy, molecular therapeutics, and immunotherapy (Ed: H.B. Newton, 2nd Edition), Academic Press, pp 715-727, 2018.

[29] L. Ferrer-Font, N. Arias-Ramos, S. Lope-Piedrafita, M. Julià-Sapé, M. Pumarola et al, "Metronomic treatment in immunocompetent preclinical GL261 glioblastoma: effects of cyclophosphamide and temozolomide", NMR Biomed., vol. 30(9), e3748, 2017 .

[30] A. Chaddad, M.J. Kucharczyk, P. Daniel, S. Sabri, J.J. Bertrand et al, "Radiomics in Glioblastoma: Current Status and Challenges Facing Clinical Implementation", Front Oncol., vol. 9,374, 2019.

[31] JM. Hakumäki, H. Poptani, A.M. Sandmair, S. Ylä-Herttuala and R.A. Kauppinen, "1H MRS detects polyunsaturated fatty acid accumulation during gene therapy of glioma: implications for the in vivo detection of apoptosis", Nat Med., vol. 5(11), pp. 1323-1327, 1999.