Classification of Brain Tumour ^{1}H MR Spectra: Extracting Features by Metabolite Quantification or Nonlinear Manifold Learning?

Guang Yang, Felix Raschke, Thomas R. Barrick, and Franklyn A. Howe
Division of Clinical Sciences, St. George's, University of London, London, UK.

ABSTRACT

Proton magnetic resonance spectroscopy (^{1}H MRS) provides non-invasive information on brain tumour biochemistry. Many studies have shown that ^{1}H MRS can be used in an objective decision support system, which gives additional diagnosis and prognostic information to the data obtained using conventional radiological modalities. Fully automatic analyses of ^{1}H MRS have been previously applied and can be separated into two types: (i) model dependent signal quantification followed by pattern recognition (PR), or (ii) model independent PR methods. However, there is not yet a consensus as to the best techniques of MRS post-processing or feature extraction to be used for optimum classification. In this study, we analysed the single-voxel MRS acquisitions of 74 patients with histologically diagnosed brain tumours. Our classification results show that the model independent nonlinear manifold learning method can produce superior results to those of using model dependent metabolite quantification.

1. INTRODUCTION

Single-voxel ^{1}H MR Spectroscopy (SV MRS) has a role in radiological diagnosis and prognosis of brain tumours by providing additional biochemical information [1]. Nevertheless, the manual interpretation and analysis of large amount of MRS data sets requires special expertise for radiologists and is time-consuming. Pattern recognition (PR) methods allow construction of an objective decision support system to assist radiologists and neuroscientists in brain tumour diagnostic and prognostic evaluation.

Various signal quantification methods have been investigated, including LCModel (LCM) [2], AMARES [3], QUEST [4] and TARQUIN [5]. Essentially, these methods minimise an object function (i.e., the squared distance between the acquired MRS data and a basis function), which is modelled according to prior knowledge of metabolic profiles of a typical MR spectrum. In addition, the use of peak fitting can provide a *data reduction* before pattern recognition [6]. This, in turn, increases the accuracy to represent the metabolic profile of the tumour. Furthermore, the fitting process combines correlated data points in the spectra, which originate from the same metabolite signal. Therefore, the signal-to-noise (SNR) ratio of the metabolic information also increases. In particular, the LCM method fits the data with a linear combination of real high SNR metabolite spectra, which are obtained by the appropriate acquisition sequences [2]. Nevertheless, there are potential drawbacks of these signal quantification methods including using inaccurate or incorrect prior metabolic information and systematic errors in fitting overlapping peaks. Under such circumstances, the original unfitted data may provide more accurate and consistent metabolic information.

There is also a considerable amount of literature on MRS classification using model independent PR, i.e., PR on the spectra without model fitting, including principal component analysis (PCA) [6], independent component analysis (ICA) [7], Non-negative matrix factorization (NMF) [8], linear discriminant analysis (LDA) [6], neural networks (NNW) [9] and others. Dimensionality reduction (DR), which aims to reveal compact and informative representations of the observed data, is of paramount importance and prerequisite in robust PR. DR refers to the mathematical mapping of high dimensional data into an eloquent representation of the intrinsic lower dimensionality using either linear or nonlinear projection methods. Thus, the extracted features in lower dimensionality are obtained without a substantial loss in class discriminative information, while preserving the geometry of the data or object relationships as much as possible. In addition, by using DR and its manifold unfolding ability, it is feasible to integrate multidimensional data for visualisation and interpretation. In this study, we advocate to use a nonlinear manifold learning method, i.e., Laplacian Eigenmaps (LE) [10], to perform the DR. Compared to conventional DR methods such as PCA and ICA, LE do not rely on linearity assumptions and instead can often identify embedded data in complicated manifolds of high-dimensional data that linear methods fail to detect [10].

The purpose of this study was to investigate whether a nonlinear manifold learning analysis of the original unfitted SV tumour spectra would outperform the metabolite quantification method as a feature selection tool for the classification of SV MRS of brain tumours. For the SV MRS dataset, we used short-echo spectra as they provide more biochemical signals than the long-echo spectra [6]. In addition, for the PR, we employed a neural network based classifier (NNW in Figure 1), which is a powerful tool that has been used in MRS classification [9]. To the best of our knowledge, this is the first investigation to compare the feature extraction and classification of SV MRS using nonlinear manifold learning based DR and LCM quantification, which are two automated techniques to assist brain tumour diagnosis from ^{1}H MRS data.

2. MATERIALS AND METHODS

2.1. Data Acquisition and Spectral Processing

We acquired MR data using a GE Signa Horizon 1.5T MR system (GE Healthcare, Milwaukee, WI, USA) equipped with 22mT/m gradients and a quadrature head coil. In addition, all the SV MRS data were acquired at short Echo Time (TE) using the GE developed point-resolved spectroscopic sequence (PRESS) protocol (Repetition Time (TR) = 2000ms, Echo time (TE) = 30ms, 2048 data points with 2500Hz bandwidth). An expert panel (including spectroscopists, pathologists and radiologists) validated the brain tissue types included in this study based on histopathological diagnosis of the central nervous system (CNS) tumours according to WHO criteria [11]. Individual voxels were placed to encompass predominantly viable tumour tissue as much as possible and avoid areas of pure necrosis.

Supported by CRUK C1459/A13303 and EU LSHC-CT-2004-503094.

The raw time-domain data were loaded and processed using our in-house software including apodisation in the time domain using a half Hann window followed by a fast Fourier transform and automatic phase correction according to Chen et al. [12]. Subsequently, each spectrum was referenced to both NAA at 2ppm (within the search region 1.80-2.20ppm) and Cho at 3.21ppm (within the search region 3.12-3.30ppm) for chemical shift alignment. In addition, each spectrum was truncated to the chemical shift range of 4.0 to 0.2ppm containing 198 data points representing the majority of metabolic information. All spectra analysed using PCA or LE were normalised using a L2-norm.

2.2. Patient Subjects

Seventy-four adult brain tumour patients in this study were recruited and scanned at St George's, University of London as part of the INTERPRET and eTUMOUR projects. In total, we obtained 74 SV MRS spectra in three groups (Figure 2), including 24 grade II (GII LGG) tumours (2 oligodendroglioma, 3 oligoastrocytoma, 3 fibrillary astrocytoma, 4 gemistocytic astrocytoma and 12 diffuse astrocytoma), 31 grade IV (glioblastoma multiforme, GIV GBM) and 19 meningioma (MNG). All study participants gave written informed consent in accordance with local ethics procedures. Biopsy or resected tumour tissue samples were available to confirm the tumour diagnosis histologically.

2.3. Extracting Features Using LCModel

We used LCModel (version 6.3-0L) to process all the original in-vivo SV MRS data with an appropriate LCM basis set. A number of different metabolites, lipids and macromolecules were quantified as shown in Figure 3. In addition, the LCM performed the quantification using an unsuppressed water signal as a reference, assumed to represent 41.7M, and we made no corrections for T_1 or T_2 relaxation time effects.

The final LCM concentration dataset contains 26 quantified metabolites (e.g., Cho, Cr, and Ins) and 7 combined metabolites (e.g., Lip13a+Lip13b) giving a total of 33 features for each spectral analysis. Subsequently, we applied the Kruskal-Wallis test, which is an extension of the nonparametric Wilcoxon rank sum test for two or more groups of data [13], to select the most significant features of LCM quantification. By setting the significance level to $p < 1 \times 10^{-5}$ the six most significant biochemical markers are selected for further classification (Figure 4). These are Cr, MM14+Lip13a+Lip13b (lipid and macromolecule), Ins, GSH, Glu, and Lip13a+Lip13b. Furthermore, LCM outputs a uncertainty estimation for each metabolite concentration using Cramér-Rao lower bounds (CRLB). We used a threshold for the CRLB that concentration estimates with CRLB larger than 25% were set to zero. In our experiments, we compared feature selection with and without CRLB consideration as shown by the boxplot in Figure 4.

2.4. Extracting Features Using Dimension Reduction

As aforementioned, in this study we advocate a nonlinear manifold learning based dimension reduction method, i.e., Laplacian Eigenmaps, to select metabolite features. LE DR is based on the assumption that the data points lie on a smooth manifold, which is a hyper-surface in high-dimensional space. LE aims to compute a low-dimensional representation of the data that can preserve local neighbourhood information and reflects the geometric structure of the manifold. Therefore, a weighted adjacency graph needs to be constructed and each node of the graph represents a single spectrum in the data set. Firstly, the connectivity of these nodes is defined by the *k*-nearest neighbour method (*k*-NN). Secondly, the weights of the connected edges are computed by a heat kernel that is

$$W_{ij} = \begin{cases} \exp\left(-\frac{\|s_i^* - s_j^*\|}{\sigma^2}\right), & \text{if } s_i^* \text{ and } s_j^* \text{ are connected;} \\ 0 & \text{otherwise.} \end{cases},$$

in which s_i^* and $s_j^* \in \mathbb{R}^M$ are the normalised original SV MRS. In addition, the Laplacian matrix is defined as $L = D - W$, in which W is the weighting matrix and $D_{ii} = \sum_j W_{ij}$ is the corresponding degree matrix. Based on standard spectral graph theory, a reasonable mapping is given by a matrix $Y \in \mathbb{R}^{M \times N}$, where $Y = (y_1, y_2, \ldots, y_n)$ maps the weighted adjacency graph to a low-dimensional space where the connected nodes remain close together. This is given by,

$$\arg\min_{y} \sum_{ij} \|y_i - y_j\|^2 W_{ij},$$

It can be proven that this is equivalent to solving a generalised eigenvalue problem $Ly = \lambda Dy$ [10]. In this study we make no extra optimisation step with respect to the two parameters, *k,* (in the *k*-NN algorithm) and σ (in the heat kernel function). The value of *k* was set according to the data size, (i.e. for 74 SV MRS then $k = 5 < 10\%$ of the data size in order to preserve the local neighbourhood properties) and σ was set to the median distance of all data points from one another in the high dimensional manifold (i.e. $\sigma = 0.48$ in this study). Furthermore, we compared the results of DR using LE to those obtained from conventional PCA. To ensure results were comparable to LCM quantification we used six LE and six PCA components (eigenvectors).

2.5. Data Classification

Pattern recognition was performed using a NNW based classifier, which consists of simple elements operating in parallel where the elements are inspired by biological nervous systems [14]. A series of logistic regression models (LRMs) stacked on top of each other with the final layer being another LRM can form a feed-forward neural network, a.k.a., multi-layer perceptron (MLP) classifier [15]. Thus, we use the logistic function as the transfer function. The architecture of the MLP network were defined using a cross-validation method such that the number of hidden neurons is 18 according to the best cross-validation accuracy using the features derived from the original SV MRS without any DR. For a fair comparison, we applied the same number of hidden layers for other experiments on extracted features obtained by PCA or LE DR and the LCM quantification.

2.6. Performance Measure

The classification performance was evaluated by (i) Leave-One-Out (LOO) cross validation, which is an unbiased predictor and is capable of creating sufficient training data for studies with small sample size [16]; (ii) the area under the receiver operating characteristic (ROC) [17] curve (AUC); (iii) the classification accuracy, sensitivity and specificity at the operating point on the ROC curve using the NNW classifier.

3. RESULTS

Table 1 gives the classification results using original unfitted SV MRS with PCA or LE DR. The most remarkable result to emerge from the classification is that that the LE DR produced the best cross validation accuracy (93.2% using LE vs. 85.1% using PCA). Also, higher AUC were obtained using LE DR (0.952) than using PCA (0.871). In addition, classifications using both DR methods have high sensitivity in recognising MNG (100%). However, PCA has lower sensitivity in recognising GII tumours (<70% sensitivity compared to 90.6% sensitivity of using LE DR). Nevertheless, PCA obtained slightly higher sensitivity (96.7%) to recognise GIV

tumours compared to 92% sensitivity using LE DR. Furthermore, both DR methods gained high specificity in separating GII class (>95%). Using PCA, both GIV and MNG have lower specificity (87.9% for GIV and only 64.7% for MNG) compared to LE DR (GIV 92% and MNG 85%).

Quantitative classification results for metabolite features selected by LCM fitting are shown in Table 2. Experiments were performed on features both without (LCM+FS) and with consideration of CRLB (LCMc+FS). The accuracy obtained from LCM fitted ^{1}H spectra are comparable with results obtained using original ^{1}H spectra with PCA. By considering the CRLB, the accuracy increased from 85.1% to 87.8%, and the AUC increased from 0.832 to 0.913. In addition, we achieved similar sensitivity results for separating MNG (>90%) compared to DR using PCA or LE DR (cf. Table 1). Although sensitivity for classification of GII tumours (67.7% and 78.9% without and with CRLB consideration) was lower than using the original spectra with PCA or LE DR, the sensitivity of detection of GIV tumour (100% and 94.7% without and with CRLB) was comparable. Again, LCM feature selection resulted in similar specificities for the three tumour types as found by using the original spectra with PCA and LE DR. The specificities of GII were 100% without and with CRLB, equivalent to using the original spectra with LE DR. Additionally, we obtained much lower specificity in recognising MNG tumours without CRLB consideration (58.8%) compared to with CRLB (85%) and features derived from LCM produced lower GIV specificities (88.9%) than using the original spectra with LE DR (92%). Moreover, using PCA or LE DR on LCM fitted spectra prior to NNW classification degraded the accuracy (86.5% and 81.1%).

4. DISCUSSION AND CONCLUSION

Our experiments corroborate with previous results, e.g., [6] [18]. We obtained a comparable overall accuracy using LOO cross-validation for GII, GIV and MNG. Using the original ^{1}H spectra with LE DR we obtained 93.2% accuracy and the best classification on LCM fitted features resulted in 87.8% accuracy using the CRLB consideration. Our previous research [18] obtained 90% overall accuracy using the same database; however, our previous study employed the Kennard and Stone algorithm to select the training and testing set, which may introduce bias into the classification framework and influence the accuracy. Furthermore, the AUC is a widely used measure of performance of supervised classification rules in biomedical applications and provides a single-number discrimination measure across all possible ranges of cut-points on the ROC curve. The classification on the original ^{1}H spectra with LE DR resulted in the best AUC. Interestingly, combining LCM with DR directly could not improve the classification accuracy (Table 2). LCM+PCA obtained comparable accuracy as using LCM with CRLB. But LCM+LE degraded the accuracy that may be attributed to the fact that both PCA and LCM are linear methods, which can lose information useful for classification. Therefore, for NNW preceded by PCA there is little difference whether using original unfitted spectra or adding the prior-knowledge from LCM fitting. However, for LE DR there is a large difference, with LE performing best on the original unfitted ^{1}H spectra and better than either of using the PCA (cf. Table 1 and Table 2).

The reliability of the ground truth (i.e. the classification of each spectrum according to the histopathological diagnosis) can affect the classification accuracy of our study. Therefore, we perform a qualitative ground truth analysis after we obtained the classification results. For example, Figure 2 shows the ground truth of all three tumour types using median spectra with 25% and 75% quartile ranges. Most spectra have common metabolite features for the different tumour types; however, there are some exceptions, e.g., an example GII (Figure 3) shows high Lipid peaks that resemble a high grade tumour, i.e., GBM. This can be attributed to the fact that the ground truth is obtained via local biopsy (in mm^3). In contrast, we acquired SV MRS data from a larger volume of interest across the entire tumour (in cm^3). Therefore, some discrepancy between ground truth and SV MRS characteristics are inevitable. Figure 5 shows FLAIR and T_2-weighted images of this GII case. Although diagnosed as GII the presentation is diffusive and may contain tissue that is transforming into higher grade tumour that may not have been sampled by biopsy. Thus, consideration of "ground truth uncertainty" requires exploration in future studies.

In conclusion, our study compared a novel nonlinear manifold learning approach to the conventional PCA and LCM fitting methods to extract metabolites features. Further classification demonstrated that our LE DR outperformed conventional PCA and LCM fitting. Future work will apply this LE DR and pattern recognition framework to more tumour types and take the uncertainty of the histopathological diagnosis into account.

5. REFERENCES

[1] F. A. Howe et al., "1{H} {MR} spectroscopy of brain tumours and masses," *NMR in Biomedicine*, vol. 16, pp. 123–131, 2003.

[2] S. W. Provencher et al., "Estimation of metabolite concentrations from localized in vivo proton NMR spectra," *Magnetic Resonance in Medicine*, vol. 30, no. 6, pp. 672–679, 1993.

[3] L. Vanhamme et al., "Improved Method for Accurate and Efficient Quantification of MRS Data with Use of Prior Knowledge," *Journal of Magn Reson*, vol. 129, pp. 35–43, 1997.

[4] H. Ratiney et al., "Time-domain semi-parametric estimation based on a metabolite basis set," *NMR in Biomedicine*, vol. 18, no. 1, pp. 1–13, 2005.

[5] M. Wilson et al., "A constrained least-squares approach to the automated quantitation of in vivo ^{1}H magnetic resonance spectroscopy data.," *Magnetic resonance in medicine*, vol. 65, no. 1, pp. 1–12, Jan. 2011.

[6] K. S. Opstad et al., "Linear discriminant analysis of brain tumour 1{H} {MR} spectra: a comparison of classification using whole spectra versus metabolite quantification," *NMR in Biomedicine*, vol. 20, no. 8, pp. 763–770, 2007.

[7] A. J. Wright et al., "Pattern recognition of MRSI data shows regions of glioma growth that agree with DTI markers of brain tumor infiltration.," *Magnetic resonance in medicine*, vol. 62, no. 6, pp. 1646–51, Dec. 2009.

[8] S. Ortega-Martorell et al., "Non-negative matrix factorisation methods for the spectral decomposition of MRS data from human brain tumours.," *BMC bioinformatics*, vol. 13, no. 1, p. 38, 2012.

[9] D. F. Brougham et al., "Artificial neural networks for classification in metabolomic studies of whole cells using 1H nuclear magnetic resonance.," *Journal of biomedicine & biotechnology*, vol. 2011, p. 158094, Jan. 2011.

[10] M. Belkin et al., "Laplacian Eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003.

[11] P. Kleihues et al., "The New WHO Classification of Brain Tumours," *Brain Pathology*, vol. 3, no. 3, pp. 255–268, 1993.

[12] L. Chen et al., "An efficient algorithm for automatic phase correction of NMR spectra based on entropy minimization," *Journal of Magnetic Resonance*, vol. 158, no. 1–2, pp. 164–168, Sep. 2002.

[13] M. Hollander et al., *Nonparametric Statistical Methods*, 2nd Ed. Wiley-Blackwell, 1999.

[14] M. T. Hagan et al., *Neural network design*. PWS Pub. Co., 1996.

[15] K. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012, pp. 16–19, 82–89, 101–103, 496–503, 563–564.

[16] B. Efron et al., "Estimating the error rate of a prediction rule: improvement on cross-validation," *Journal of the American Statistical Association*, vol. 78, no. 382, pp. 316–331, 1983.

[17] C. Metz, "ROC methodology in radiologic imaging," *Investigative radiology*, vol. 21, pp. 720–733, 1986.

[18] F. Raschke et al., "Classification of single-voxel 1H spectra of brain tumours using LCModel.," *NMR in biomedicine*, vol. 25, no. 2, pp. 322–31, Feb. 2012.

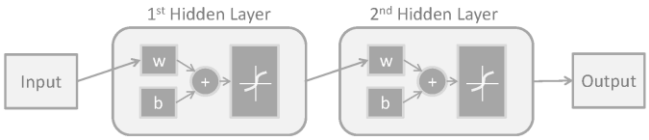

Figure 1: Schematic architecture of a NNW with two hidden layers.

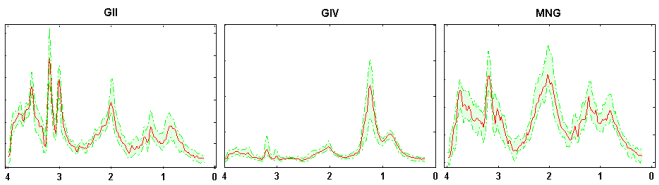

Figure 2: Ground truth of 74 brain tumour SV MRS (24 GII, 31 GIV and 19 MNG). The median (solid red curve) and 25% and 75% quartile (dashed green curves) spectra are shown for each tumour class.

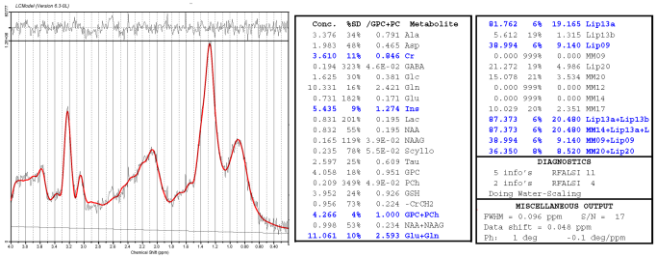

Figure 3: The LCM fitting (left panel) for a grade II case with all estimated metabolite concentration (middle and right panels).

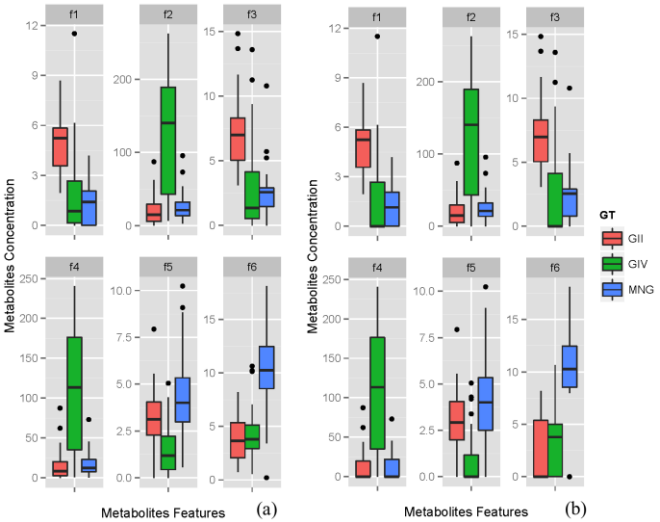

Figure 4: Boxplot of the 6 extracted features using metabolite quantification by LCM. (a) Use all the extracted features regardless the CRLB. (b) A CRLB less than or equal to 25% is considered to provide reliable estimation of the metabolites whereas features with larger than 25% CRLB were set to zero. Features are: f1—Cr; f2—MM14+Lip13a+Lip13b; f3—Ins; f4—GSH; f5—Glu; and f6—Lip13a+Lip13b.

Table 1: NNW classification results using the original ^{1}H spectra with PCA or LE DR.

Results of Original Unfitted Spectra Classification								
	Accu.	Sensitivity			Specificity			AUC
		GII	GIV	MNG	GII	GIV	MNG	
PCA	85.1	69.7	96.7	100.0	95.8	87.9	64.7	0.871
LE	93.2	90.6	92.0	100.0	100.0	92.0	85.0	0.952

Table 2: NNW classification results using the quantified data (LCM with feature selection (FS), LCM+PCA and LCM+LE).

Results of LCM Fitted Spectra Classification								
	Accu.	Sensitivity			Specificity			AUC
		GII	GIV	MNG	GII	GIV	MNG	
LCM+FS	85.1	67.7	100.0	90.9	100.0	88.9	58.8	0.832
LCMc+FS	87.8	78.9	94.7	100.0	100.0	75.0	85.0	0.913
LCM+PCA	86.5	75.8	96.2	93.3	100.0	89.3	66.7	0.917
LCM+LE	81.1	73.3	94.7	90.0	97.1	81.8	50.0	0.871

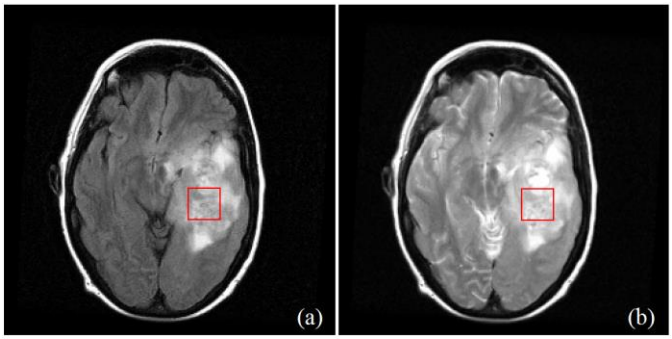

Figure 5: FLAIR image (a) and T_2-weighted image (b) of a Grade II case. The tumour and surrounding oedema regions are widespread. The red box shows the placement of the MRS voxel.