

**Manifold Learning in MR Spectroscopy using
Nonlinear Dimensionality Reduction and Unsupervised Clustering**

Guang Yang, Felix Raschke, Thomas R. Barrick, and Franklyn A. Howe

Neurosciences Research Centre, Cardio-vascular and Cell Sciences Institute,

St. George's University of London, London, UK.

Main Body Word Count: 4293

Abstract Word Count: 200

Running Head: Manifold Learning in MRS using Nonlinear DR and Unsupervised Clustering

Corresponding author: Guang Yang

Email: gyang@sgul.ac.uk

Phone: (0044) 020 8725 5120; Fax: (0044) 020 8725 2950

Address: Room 31b, Neurosciences Research Centre, Cardio-vascular and Cell Sciences

Institute, St George's, University of London, Cranmer Terrace, London, SW17 0RE.

KEYWORDS: Pattern recognition; Dimensionality reduction; Laplacian Eigenmaps; Magnetic resonance spectroscopy; Magnetic resonance spectroscopic imaging.

ABSTRACT

PURPOSE: To investigate whether non-linear dimensionality reduction improves unsupervised classification of ^1H MRS brain tumour data compared to a linear method.

METHODS: In-vivo single voxel ^1H MRS (55 patients) and ^1H MRSI (29 patients) data were acquired from histopathologically diagnosed gliomas. Data reduction by Laplacian eigenmaps (LE) or independent component analysis (ICA) was followed by k-means clustering or agglomerative hierarchical clustering (AHC) for unsupervised learning to assess tumour grade and for tissue type segmentation of MRSI data.

RESULTS: An accuracy of 93% in classification of glioma Grade-II and Grade-IV, with 100% accuracy in distinguishing tumour and normal spectra, was obtained by LE with unsupervised clustering, but not with the combination of k-means and ICA. With ^1H MRSI data LE provided a more linear distribution of data for cluster analysis and better cluster stability than ICA. LE combined with k-means or AHC provided 91% accuracy for classifying tumour grade and 100% accuracy for identifying normal tissue voxels. Colour-coded visualisation of normal brain, tumour core and infiltration regions was achieved with LE combined with AHC.

CONCLUSION: The LE method is promising for unsupervised clustering to separate brain and tumour tissue with automated colour-coding for visualisation of ^1H MRSI data after cluster analysis.

ABBREVIATIONS: PCA: principal component analysis; ICA: independent component analysis; PR: pattern recognition; DR: Dimensionality reduction; LE: Laplacian Eigenmaps; MRS: magnetic resonance spectroscopy; MRSI: magnetic resonance spectroscopic imaging.

INTRODUCTION

Gliomas are the most common type of primary brain tumour and accurate diagnosis of tumour grade is necessary for appropriate treatment and prognosis. Histopathological characterisation from a surgical biopsy carries risk (1,2) and can be problematic due to tumour heterogeneity (1,3,4). Conventional MRI provides limited accuracy in defining tumour heterogeneity and infiltration boundaries and ¹H magnetic resonance spectroscopic imaging (MRSI) is being investigated for its clinical potential to aid non-invasive diagnosis (5). The manual interpretation and analysis of large multivoxel ¹H MRSI datasets is difficult and time-consuming hence automated pattern recognition (PR) methods are being investigated for tumour classification (6–9). PR of MRSI data has shown promise in objectively delineating tumour boundaries and defining infiltrative growth patterns in gliomas (10,11) and nosologic images (3,12) used to summarise the presence of various tissue types and lesions in a single colour coded image overlay on to conventional MRI.

Prior to PR, dimensionality reduction (DR) is needed, which provides a mathematical mapping of high dimensional data into a lower dimensional representation. In particular, lower dimensional representations ideally should be obtained without substantial loss in discriminative information, while maximally preserving overall data geometry. Consequently, points which are in close proximity in the original high-dimensional space are mapped to adjacent locations in the low-dimensional space (13). PCA and ICA are well-known linear DR techniques that have been applied to ¹H MRS data, but there are assumptions in their application. Firstly, PCA will not extract the true variance from a non-Gaussian distribution of data. Secondly, ICA requires input data to be linear, independent and non-Gaussian. ICA can outperform PCA as the underlying probabilistic model is more powerful and is sensitive to higher-order statistics allowing superior

1
2
3 identification of data concentrations in n-dimensional space (14). In contrast, nonlinear DR
4
5 techniques (15–20) can unfold high-dimensional data that provide class separation where linear
6
7 methods often fail.
8
9

10 In this study, we focus on the application of the nonlinear Laplacian Eigenmaps (LE) method for
11
12 DR of ^1H MRS brain tumour data with an aim to aid diagnosis and determine regions of potential
13
14 tumour infiltration. LE-DR was chosen due to its simplicity of implementation and its
15
16 robustness, to outliers and noise (20). We use unsupervised learning based k-means and
17
18 agglomerative hierarchical clustering (AHC) methods to form 3 clusters (nominally representing
19
20 normal tissue, potential infiltration and tumour core) determined from the eigenvalues of the DR
21
22 output. SVS and MRSI results obtained by LE-DR are compared with those obtained using
23
24 PCA+ICA. We investigate the extent to which the non-linear LE-DR method provides improved
25
26 classification for generation of colour-coded maps of tumour core, infiltration and normal tissue
27
28 compared to conventional linear DR.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHODS

All MR data were acquired using a GE Signa Horizon 1.5T MR system (GE Healthcare, Milwaukee, WI, USA) equipped with 22 mT/m gradients and a quadrature head coil. Study participants gave written informed consent in accordance with local ethics procedures. Biopsy or resected tumour tissue samples were available to confirm histological tumour diagnosis providing a “ground truth” classification.

Single Voxel MRS: Subjects and Data Acquisition

SV MRS (SVS) data were acquired from adult brain tumour patients at St George's, University of London as part of the INTERPRET and eTUMOUR EU projects (data are web-accessible by permission for download from <http://solaria.uab.es/eTumour/>). In total N = 55 tumour MR spectra were obtained. This included 24 Grade-II tumours (2 oligodendrogliomas, 3 oligoastrocytomas, 3 fibrillary astrocytomas, 4 gemistocytic astrocytomas and 12 diffuse astrocytomas) and 31 Grade-IV (glioblastoma). To create a dataset of normal brain single voxel spectra, 79 spectra were extracted from MRSI data of 3 healthy volunteers, which had compatible TR/TE parameters (full acquisition details as defined in the next Section) and these data were interpolated to the same spectral resolution as the SVS tumour data. The final dataset comprised N = 134 individual spectra (Table 1).

All SVS tumour data were acquired at short Echo Time (TE) using the GE developed point-resolved spectroscopic sequence protocol (Repetition Time (TR) = 2000ms, TE = 30ms, 2048 data points with 2500Hz bandwidth). Voxels for MRS were positioned according to the clinical MRI of each patient, to predominantly include viable tumour tissue and avoid areas of pure necrosis. Each spectrum was referenced to both NAA at 2ppm (search region 2.20ppm –

1.80ppm) and Cho at 3.21ppm (search region 3.30ppm – 3.12ppm) for chemical shift alignment. In addition, each spectrum was truncated to 4.0ppm – 0.2ppm containing $M = 198$ data points representing the majority of metabolic information (Figure 1(a)).

MRSI Chemical Shift Imaging (CSI): Subjects and Data Acquisition

Our thirty ^1H MRSI data were obtained from 29 patients with glial tumours: Grade-II (2 oligoastrocytomas, 2 gemistocytic astrocytomas and 6 diffuse astrocytomas); Grade-III (1 anaplastic oligoastrocytoma and 6 anaplastic astrocytomas) and Grade-IV (1 gliosarcoma and 12 glioblastomas). For one patient diagnosed with an anaplastic astrocytoma post-surgical resection MRSI data were acquired 15 weeks after the initial scan, and both pre- and post-surgery scans were included in the study.

MRSI data were acquired using the GE PROBE-SI protocol ($TR = 2000\text{ms}$, $TE = 30\text{ms}$, 512 data points and 1000Hz bandwidth) with outer volume suppression and a 16×16 phase-encoded matrix. A 15mm thick axial plane was selected through the centre of the tumour mass with a selection volume for MRSI that included the tumour bulk and surrounding tissue while avoiding the skull. MRSI data were pre-processed by zero-filling prior to 2D Fourier transformation to produce a 32×32 matrix of voxels with in-plane spatial resolution of 6.875mm. In addition, zero-filling was applied to each free induction decay signal in the Fourier domain, interpolating the original 512 data points to 2048. This provided high-resolution spectra while preserving relative metabolite signal intensities.

A total of $N = 2097$ voxels were extracted from the MRSI data, but 132 voxels (6%) were excluded due to poor water suppression and misalignment (i.e., signals unable to be referenced as aforementioned) leaving $N = 1965$ voxels remaining (Table 1). Spectral data were limited

(4.0ppm – 0.2ppm) and the phased real part used for further analysis. Each MRSI spectrum consisted of $M = 498$ data points (Figure 1(h)).

Pattern Recognition

Dimension Reduction

An in-house Matlab program (Matlab 2013a, Mathworks, Natick, USA) was developed for processing, analysis, and display for the MR spectra. All SVS or MRSI spectra were stored as a 2D matrix, $S \in \mathbb{R}^{M \times N}$, where $S = (s_1, s_2, \dots, s_N)$ contains each spectrum as a column vector $s_n \in \mathbb{R}^M$. The column vectors were normalised,

$$1) \quad s_n^* = \frac{s_n}{\sqrt{\sum_m s_{nm}^2}} \quad \forall n \in \{1, 2, \dots, N\}$$

to form the matrix $S^* \in \mathbb{R}^{M \times N}$, in which N is the number of the spectra and M is the number of the data points of each spectrum.

We have implemented the LE method (20), which assumes that data points lie on a smooth manifold that is a hyper-surface in high-dimensional space. In particular, LE computes a low-dimensional representation of the data that preserves local neighbourhood information and reflects the geometric structure of the manifold. Subsequently, a weighted adjacency graph G with N nodes was constructed, in which each node represents a single spectrum in the dataset. To achieve this, the k -nearest neighbours method (k -NN) was applied to determine connectivity between the nodes according to the Euclidean distance in \mathbb{R}^N . Weights of connected edges was defined using a heat kernel,

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

The Laplacian matrix, $L = D - W$, was constructed in which W is the adjacency matrix defined in Equation 2) with corresponding degree matrix, $D_{ii} = \sum_j W_{ij}$. Based on standard spectral graph theory the matrix transformation, $Y \in \mathbb{R}^{M \times N}$ where $Y = (y_1, y_2, \dots, y_n)$ maps the weighted adjacency graph G to a low-dimensional space where connected nodes remain close together. This map is given by,

$$3) \quad \arg \min_y \sum_{ij} \|y_i - y_j\|^2 W_{ij},$$

and provides heavy W_{ij} penalties if neighbouring nodes s_i^* and s_j^* are mapped to be far apart. Minimisation of Equation 3) attempts to ensure that if s_i^* and s_j^* are close then the mappings y_i and y_j are similar. Furthermore, as $D_{ii} = \sum_j W_{ij}$, W_{ij} is symmetric and $L = D - W$, we can derive the following,

$$\begin{aligned} 4) \quad \arg \min_y \sum_{ij} \|y_i - y_j\|^2 W_{ij} &= \arg \min_y \sum_{ij} (y_i^2 + y_j^2 - 2y_i y_j) W_{ij} \\ &= \arg \min_y (\sum_i y_i^2 D_{ii} + \sum_j y_j^2 D_{jj} - 2 \sum_{ij} y_i y_j W_{ij}) \\ &= \arg \min_y (2y^T L y) \equiv \arg \min_y \text{Tr}(Y^T L Y); \end{aligned}$$

therefore, the minimisation problem may be simplified to

$$5) \quad \arg \min_y \text{Tr}(Y^T L Y) \text{ subject to } Y^T D Y = I$$

L is positive semi-definite thus the vector y that minimises the objective function is given by the minimum eigenvalue solution to the generalised eigenvalue problem (with respect to eigenvalues, λ) as follows,

$$6) \quad L y = \lambda D y.$$

The constraint $Y^T D Y = I$ prevents collapse onto a subspace of dimension less than $N - 1$ with the solution provided by the eigenvectors with the smallest non-zero eigenvalues (20).

For our LE-DR with further unsupervised clustering framework, we have no separate training dataset to optimise the two parameters (k in the k -NN algorithm and σ in the heat kernel function) in LE-DR. Recent studies have shown that $k \approx \frac{1}{10}N$ is suitable for various applications (21,22) and that, σ can be set according to the mean distance between graph neighbours that is widely used in various nonlinear DR methods (23–25). Therefore, we set $k = 12$ and $\sigma = 0.45$ for $N = 134$ SVS, and $k = 200$ and $\sigma = 0.52$ for $N = 1965$ MRSI.

Our comparison for the LE-DR method is ICA, a linear blind source separation technique used to identify individual signals from mixtures (26), which in our study may resemble spectral patterns of the tissue-subtypes in each MRS voxel (11). We employed the ‘fastICA’ package (<http://research.ics.aalto.fi/ica/fastica>), which implements a fast fixed-point algorithm in Matlab (27). Spectra were first normalised and mean-centred and then PCA applied prior to ICA to limit the number of ICs generated and to provide “whitening” of the data. Pre-processing the data using PCA enhances ICA performance by discarding small trailing eigenvalues and reduces computational complexity for the fastICA iterative algorithm by minimising pairwise dependencies (28). In addition, PCA also de-correlates the input mixing data, and the remaining higher-order dependencies can be separated by ICA. Whitening ensures that all dimensions are treated equally a priori before the application of ICA. Three PCs were then derived as this is the expected number of stable components in the glioma MRS dataset (11).

Unsupervised Learning Based Clustering

k-means is one of the most popular iterative-descent clustering methods (29). In our implementation the sum of absolute differences was used as the dissimilarity measurement, with each centroid representing the component-wise mean of the points in that cluster. Cluster number

was set to $\tilde{k} = 3$ under the assumption of the existence of three tissue classes representing normal brain, infiltrative tumour (Grade-II for the SVS analysis) and high-grade/necrotic tissue (Grade-IV for the SVS analysis).

For AHC the Euclidean distance was used as a pairwise dissimilarity measure. A bottom-up paradigm was used to recursively merge pairs of clusters into a single cluster at each level (30). Pairs were merged based on the smallest inter-group dissimilarity.

Data Analysis

SVS Data Analysis

PCA+ICA and LE were applied to the SVS dataset and each spectrum was assigned a spectral classification label according to their ranked eigenvalues. Mean eigenvector spectra and variance were then calculated for identification of the tissue types represented by each spectral class.

k-means and AHC methods were used to segment the 3D eigenspace for each DR method. Quantitative validation was provided by mean silhouette values (31). For each k-means or hierarchical cluster, the cluster label was defined according to the metabolic characteristics of the mean cluster spectrum. Clustering performance was evaluated by accuracy, sensitivity, specificity and balanced error rate (BER) (32) computed with respect to “ground truth” histopathological diagnosis for each patient.

MRSI Data Analysis

The above analysis was also applied to the MRSI data. Results were visualised using colour overlays (green—normal, blue—infiltrative, red—high-grade/necrotic) on conventional MRI to

allow assessment of classification results in the absence of histopathological ground truth for all voxels in the MRSI dataset.

Since we have a small number of MRSI datasets for each tumour grade, and the components derived by data reduction will depend on the input dataset, we propose a take-one-out method for the MRSI analysis as an added estimate of the stability of the overall process of tissue segmentation by data reduction and clustering. This is unlike the standard leave-one-out process, and assesses the consistency of data clustering (in the absence of definitive tissue classification for every MRSI voxel) rather than the accuracy of classification. Cluster stability (CS) was defined as the percentage difference between classification labels of all MRSI voxels before and after one patient dataset was removed prior to the DR and clustering procedure. CS was computed separately for each tissue segment $q \in \{1,2,3\}$ such that,

$$7) \quad CS_q = \sum_{\hat{p}} \frac{c(\hat{p}, q)}{\hat{n}} \times 100 \text{ where } c(\hat{p}, q) = \begin{cases} 0 & \text{if } c(p) = q = c(\hat{p}) \\ 1 & \text{if } c(p) = q \neq c(\hat{p}) \end{cases},$$

where $c(p)$ is the voxel classification that is computed from the entire MRSI dataset and $c(\hat{p})$ is the voxel classification computed over the set of MRSI voxels $\hat{p} \in \{1,2,\dots,\hat{n}\}$ after removal of one patient. Clustering performance was evaluated by CS and BER, which was computed by comparing colour overlays with histopathological diagnosis for each patient.

RESULTS

Dimensionality Reduction

Figure 1 shows eigenvector results represented as spectra for DR by PCA+ICA (Figure 1(b)-(d)) and LE (Figure 1(e)-(g)) for SVS data and for MRSI data (Figure 1(i)-(n)). Common spectral features were observed for the eigen-decomposed spectra derived by both DR techniques for both datasets. Eigenvector 1 (Figure 1(b), (e), (i), and (l)) shows characteristic NAA, tCr, tCho and mI peaks in proportions typical of normal brain tissue. Eigenvector 3 (Figure 1(d), (g), (k), and (n)) shows the absence of NAA and includes large lipid/macromolecules signals at 1.3 and 0.9 ppm, which are characteristic of tumour core with necrotic regions in high-grade tumours. Eigenvector 2 shows elevated mI and reduced NAA for all DR results (Figure 1(c), (f), (j), and (m)) and is likely associated with the presence of glial cells and neuronal cell loss found in infiltrative tumour regions. For the PCA+ICA analysis there is also elevated tCho (Figure 1(c) and (j)). Eigenvector 2 shows less consistency across all analyses, but has characteristics that can be associated with Grade-II gliomas and regions of tumour infiltration. Overall there appears to be more consistency between the two spectroscopic datasets for the eigenvectors derived by LE-DR than by PCA+ICA.

SVS Data Cluster Analysis

Figure 2(a)-(d) shows eigenvalue scatter plots coloured according to cluster group (normal brain tissue—green, Grade-II—blue, Grade-IV—red). Cluster group assignment is based on the similarity of their mean metabolite spectra (Figure 2(e)-(p)) to the known spectra of normal brain and the mean spectra of gliomas of different grades as found in other brain tumour studies (33,34) and also shown for the current single voxel data in Figure 3(a) to (c). For the purpose of

labelling, Figure 2(i) is also assigned as Grade-II, although it closely resembles normal brain spectra there is slightly elevated tCho and the presence of a small lactate doublet, both of which are Grade-II characteristics. The scatterplots show that LE-DR increases the separation between tissue classes by unfolding the data into a more linear distribution compared to PCA+ICA. This particularly affects the combination of PCA+ICA with k-means, which does not make such a clear separation of the metabolic characteristics of normal tissue and low-grade tumours as shown by the mean and standard deviation spectra (Figure 2(e) and (i)).

Both clustering methods have significantly lower mean silhouette widths after PCA+ICA than LE-DR (Wilcoxon signed-rank test $P < 0.001$, Table 2) indicating the presence of data outliers. Better clustering after LE-DR is potentially attributable to unfolding of the original high-dimensional data into a more linear low-dimensional data structure.

Smaller classification errors were obtained for k-means clustering by LE than by PCA+ICA, whereas low classification errors were found for AHC after both DR techniques (Table 2). Normal spectra were only misclassified by the combination of k-means clustering and PCA+ICA. The improved classification by k-means after LE-DR is indicated by higher sensitivity and specificity for classification of normal and Grade-II tumours (Table 4).

The mean and standard deviation spectra for the tissue classes according to histopathologically classification of a biopsy are shown in Figure 3(a)-(c) and are similar to the spectra derived from all the DR and cluster analyses except for that by PCA+ICA with k-means (Figure 2(i)). Examples of outlying spectra are identified for normal tissue (Figure 3(d)), Grade-II gemistocytic astrocytoma (Figure 3(e)) and Grade-IV glioblastoma (Figure 3(f)). The normal spectra are generally well clustered in the LE-DR eigenspace (Figure 3(g)), even for a normal spectrum with low-grade characteristics of elevated choline to creatine due to the rising baseline

from poor water suppression (Figure 3(d)). Inconsistency between diagnosis and classification is found in the Grade-II and Grade-IV cases. Here the outlying Grade-II spectrum (Figure 3(e) and (h)) has high-lipid components and is misclassified as Grade-IV. The Grade-IV glioblastoma spectrum (Figure 3(f) and (i)) has no lipid component and is misclassified as Grade-II.

MRSI Data Cluster Analysis

Eigenvalue scatter plots Figure 4(a)-(d) show the LE method to provide a more linear distribution of eigenvalues than the PCA+ICA method in accordance with the SVS data (Figure 2(a)-(d)). For each DR technique both k-means and AHC methods show similar clustering characteristics (as seen by comparing Figure 4(a) and (c) or Figure 4(b) and (d)). Mean spectra of the clusters for all four methods show similarities for normal tissue characteristics (Figure 4(e)-(h)) and high-grade necrotic core (Figure 4(m)-(p)) as found using SVS data (Figure 2(e)-(h) and Figure 2(m)-(p)). Differences in mean spectra are found in the cluster potentially attributable to low-grade tumour characteristics (PCA+ICA Figure 4(i) and (k); LE-DR Figure 4(j) and (l)). Here the PCA+ICA mean spectra have a slight reduction in NAA and elevation in tCho more likely associated with Grade-II tumour infiltrative regions. The LE-DR mean spectra has a greater elevation in tCho, greater reduction in NAA and presence of lipid and macromolecules peaks (1.3 and 0.9ppm) and so would be associated with infiltration that also includes some high-grade tumour regions. For the purposes of labelling we refer to this as the tumour infiltration cluster.

Visualisation of AHC results overlain on conventional MRI allowed the anatomical locations of segmented tissue types to be visually assessed. Six exemplary cases are illustrated in Figure 5 consisting of 2 Grade-II (1 diffuse astrocytoma Figure 5(a) and (g); 1 gemistocytic astrocytoma

Figure 5(b) and (h)), 2 Grade-III (anaplastic astrocytomas) and 2 Grade-IV (glioblastomas). For 7 out of the 30 MRSI datasets, PCA+ICA misclassified normal tissue voxels that were distant from or contralateral to the main lesion as tumour (Table 3) for k-means (data not shown) and for AHC (Figure 5).

To provide an assessment of the accuracy of AHC techniques for correctly identifying patients with low-grade (Grade-II) or high-grade (Grade-III and IV) tumours, each patient MRSI colour map was considered to represent high-grade tumour if any voxel was coloured red (i.e., high-grade/necrotic core) and low-grade if only coloured blue or green (i.e., the infiltration cluster). Good accuracy and low BER (better than for the SVS data) was achieved for all analysis combinations for Grade-II vs. Grade-IV classification (Table 3). However, the inclusion of Grade-III in the high-grade group reduced classification performance (Table 3). Cystic regions in Grade-II and Grade-III cases are graded as high-grade by PCA+ICA (Figure 5(a) and (d)) due to the presence of lactate, which is effectively misinterpreted as a lipid component in the data reduction. This is in contrast to LE-DR, where all cystic regions are clustered and classified as infiltration regions. Hence there is a higher BER for Grade-II vs. IV classification with PCA+ICA than for LE-DR (Table 3). Conversely the grading of cystic regions in some Grade-III as high-grade by PCA+ICA (Figure 5(d)) provides slightly higher classification accuracy compared to LE-DR for Grade-II vs. Grade-III and Grade-IV (Table 3).

The best performance across all measures (Table 2 and Table 3) was achieved by LE-DR with AHC. In general, this provides anatomically reasonable classification of low-grade regions within (Figure 5(h)) or up to the edge of the T_2 -weighted MRI abnormality (Figure 5(g)). For high-grade tumours there is generally a core region (red) surrounded by an infiltrative (blue) margin that extends beyond the post-gadolinium T_1 -weighted margin (Figure 5(k)) or lies within

the T₂-weighted hyper-intense region (Figure 5(l)). Grade-III tumours were mixed in terms of these characteristics (Figure 5(i) and (j)).

For Peer Review

DISCUSSION

In this study we have developed a new metric to qualitatively and quantitatively characterise brain tissue of patients into normal, tumour infiltration, and tumour core segmentations using a nonlinear one-step DR by LE on ¹H MRS data. Overall, our results suggest that an LE-DR+AHC framework provides improved tissue classification compared to conventional linear PCA+ICA with k-means methods. The classification results of the SVS analysis also suggest the visual segmentations provided by MRSI colour overlays for necrotic tumour core and infiltration have validity (Figure 5(g)-(l)).

A limitation of the current MRSI study is that in-vivo multiple-voxel data has no histopathological validation available on a voxel-by-voxel basis. The gold standard of brain tumour diagnosis is histopathological analysis of a biopsy, with the highest grade tumour cells determining the clinical grade and subsequent patient treatment. However, this may not fully represent the average tumour characteristic over large MRS voxels. Consequently, errors in “ground truth” may affect classifier derivation and accuracy evaluation (35). With our unsupervised classification we have found that for both SVS and MRSI data LE-DR provides a natural clustering of spectra that agrees well with the clinical and histopathological diagnosis of Grade-II and Grade-IV gliomas and to the anatomical extent of these tumours by MRI (Figure 5). The apparent Grade-II regions surrounding a Grade-IV core (Figure 5(k) and (l)) have been observed in other MRSI studies (3,11). Partial volume of high-grade tumour core with surrounding infiltrative tumour will lead to a continuum of spectral characteristics; hence with unsupervised clustering of the MRSI data there is some mixing of the Grade-II and high-grade necrotic spectra (Figure 4(j) and (l)) with LE-DR. However, this appears less of a problem in

terms of classification and characterisation (Table 3) than the effects of partial volume mixing of Grade-II and normal tissue when using PCA+ICA data reduction (Figure 4(i) and (k)).

Classification of Grade-III tumours remains a challenge due to heterogeneous tumour characteristics. In this study we assumed the existence of three distinct tissue components. In a preliminary study (data not shown) 4 LE components were tested, but provided no added benefit in tumour tissue classification, with the additional sub-class representing partial volume of grey and white matter regions.

Although we may expect a linear signal path from tissue metabolite concentration to MRS signal intensity and that spectra from mixed tissue types will add in linear proportion, our results suggest benefits for application of nonlinear data analysis techniques. In the presence of large heterogeneity found in tumours and adjacent brain tissue there is no reason to suppose simple linear functions will provide the best relationship between MR spectrum (visual data) and tissue class (world state) (36). The LE method provides very narrow eigenvector spectral peaks for normal brain components (Figure 1(e)) and broader peaks for tumour (Figure 1(f) and (g)) in contrast to PCA+ICA (Figure 1(b)-(d)). This suggests that LE-DR is potentially detecting factors associated with nonlinear effects in metabolite line shape in tumour tissue. Nonlinear effects also occur due to distortions in relative metabolite peak heights for peripheral MRSI voxels due to chemical shift artefacts; however, in the current study these were reduced by excluding outer edge voxels of the MRSI volume. Variations in metabolite line-width and nonlinear variations in peak height may be caused by heterogeneity of tumour tissue. Necrosis, fibrosis, oedema and haemorrhage may all contribute to microscopic magnetic field inhomogeneities at the cellular level and cause larger scale heterogeneity across the MRSI volume and spectrum distortions.

The LE-DR method appears to unfold the MRS data to a more linear data structure than PCA+ICA. A consequence of this is a more linear geometry for data clustering that leads to better cluster stability following LE-DR than compared PCA+ICA. It is apparent that LE-DR unfolds the MRS data into a structure of spectral characteristics from normal brain to high-grade tumour characteristics that provides improved classification accuracies and lower error rates. Our classification accuracies of 93% with SVS and 91% with MRSI for distinguishing Grade-II from Grade-IV tumours compares well to other PR analyses. Classification accuracy of 93% has been obtained by combining long-echo and short-echo spectra (37); however, this would be difficult to routinely achieve due to time constraints. A more recent approach used Non-negative Matrix Factorisation (NMF) to accomplish the DR task for SV tumour spectra (38); however, NMF methods converge to local minima and various initialisations provide different DR and thus require an elaborate initialisation scheme (38). Ortega-Martorell et al. reached 22% BER with long echo SVS and 7% BER with short echo SVS for supervised classification of Grade-II, Grade-IV and normal spectra (38). In contrast, our unsupervised learning framework, which requires minimum parameter manipulation and no extra supervised training procedure, obtained best classification accuracy for LE-DR+AHC with an 11% BER for SVS and 8% BER for MRSI data. Further work with a more supervised approach to defining the clusters from LE-DR should achieve improved classification, but needs validation such as with appropriately localised biopsy classification (39) or from longitudinal studies to confirm the presence of tumour infiltration and grade from tumour growth characteristics and patient outcome (40).

CONCLUSIONS

This work presents a threefold contribution to automated analysis of brain tumour ^1H MRS data. Firstly, to the best of our knowledge this is the first time a nonlinear DR method has been applied to MRSI data. Secondly, we have demonstrated a one-step DR using nonlinear LE in comparison to ICA (which relies on data pre-processing using PCA). LE-DR reduces the risk of loss of relevant data features and leads to better classification accuracy and clustering stability measures. Thirdly, promising brain tissue segmentations have been achieved using LE-DR with unsupervised clustering of the data. Clustering using k-means or hierarchical algorithms and colour visualisation techniques indicate the possible benefits for the nonlinear LE-DR method in determining different MRS characteristics for low- and high-grade gliomas and for identifying infiltration into normal brain tissue. With further validation these techniques could aid targeted biopsies and improve surgical and radiotherapy planning.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1

Number of Subjects Studied					
Dataset	Grade II	Grade III	Grade IV	Normal	In total
Experiment 1: SV MRS	24	0	31	3	58
Experiment 2: MRSI	10	7	13	0	30
Number of MRS Voxels Analysed					
Dataset	From Grade II	From Grade III	From Grade IV	From Normal	In total
Experiment 1: SV MRS	24	0	31	79	134
Experiment 2: MRSI	722	509	734	0	1965

Table 2

Test	k-means Clustering		AHC	
	PCA+ICA	LE	PCA+ICA	LE
Silhouette Analysis (SVS)	0.4±0.3	0.6±0.2	0.5±0.2	0.6±0.2
Tumour vs. Normal Classification (SVS)	67%	100%	100%	100%
Classification Accuracy / BER (SVS)	60% / 30%	93% / 11%	93% / 11%	93% / 11%

Table 3

Test	k-means Clustering		AHC	
	PCA+ICA	LE	PCA+ICA	LE
Cluster Stability (MRSI)	10%±19%	2%±1%	5%±2%	1%±1%
Normal tissue misclassified as GII (MRSI)	Yes (7/30)	No	Yes (7/30)	No
GII vs. GIV Accuracy / BER (MRSI)	91% / 9%	91% / 8%	91% / 9%	91% / 8%
GII vs. (GIII + GIV) Accuracy / BER (MRSI)	77% / 20%	73% / 20%	77% / 20%	73% / 20%

Table 4

k-means Clustering		Predicted Group Membership				Sens. and Spec.		
		Nor	GII	GIV	Total	Sens.	Spec.	
PCA+ICA	Brain Tissue Types	Nor	35	44	0	79	44%	100%
		GII	0	21	3	24	88%	54%
		GIV	0	7	24	31	77%	97%
LE	Brain Tissue Types	Nor	79	0	0	79	100%	100%
		GII	0	24	0	24	100%	91%
		GIV	0	10	21	31	68%	100%
Hierarchical Clustering		Predicted Group Membership				Sens. and Spec.		
		Nor	GII	GIV	Total	Sens.	Spec.	
PCA+ICA	Brain Tissue Types	Nor	79	0	0	79	100%	100%
		GII	0	21	3	24	88%	95%
		GIV	0	6	25	31	81%	97%
LE	Brain Tissue Types	Nor	79	0	0	79	100%	100%
		GII	0	21	3	24	88%	95%
		GIV	0	6	25	31	81%	97%

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ACKNOWLEDGEMENT

This work was supported by Cancer Research UK project, grant number C1459/A13303, and data was obtained during the EU FP7 eTUMOUR project.

For Peer Review

REFERENCES

1. Hall WA. The safety and efficacy of stereotactic biopsy for intracranial lesions. *Cancer* 1998;82:1749–1755.
2. Favre J, Taha JM, Burchiel KJ. An analysis of the respective risks of hematoma formation in 361 consecutive morphological and functional stereotactic procedures. *Neurosurgery* 2002;50:48–56.
3. De Edelenyi FS, Rubin C, Estève F, et al. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images. *Nat Med* 2000;6:1287–1289.
4. Jackson RJ, Fuller GN, Abi-Said D, Lang FF, Gokaslan ZL, Shi WM, Wildrick DM, Sawaya R. Limitations of stereotactic biopsy in the initial management of gliomas. *Neuro. Oncol.* 2001;3:193–200.
5. Howe FA, Opstad KS. ¹H MR spectroscopy of brain tumours and masses. *NMR Biomed.* 2003;16:123–131.
6. Tate AR, Majós C, Moreno A, Howe FA, Griffiths JR, Arús C. Automated classification of short echo time in in vivo ¹H brain tumor spectra: A multicenter study. *Magn. Reson. Med.* 2003;49:29–36.
7. Devos A, Lukas L, Suykens JAK, et al. Classification of brain tumours using short echo time ¹H MR spectra. *J. Magn. Reson.* 2004;170:164–175.
8. Simonetti AW, Melssen WJ, De Edelenyi FS, van Asten JJ A, Heerschap A, Buydens LMC, De Edelenyi FS. Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. *NMR Biomed.* 2005;18:34–43.
9. Georgiadis P, Kostopoulos S, Cavouras D, Glotsos D, Kalatzis I, Sifaki K, Malamas M, Solomou E, Nikiforidis G. Quantitative combination of volumetric MR imaging and MR spectroscopy data for the discrimination of meningiomas from metastatic brain tumors by means of pattern recognition. *Magn. Reson. Imaging* 2011;29:525–535.
10. De Edelenyi FS, Simonetti, F. A, Postma WG, Huo R, Buydens LMC. Application of independent component analysis to ¹H MR spectroscopic imaging exams of brain tumours. *Anal. Chim. Acta* 2005;544:36–46.
11. Wright AJ, Fellows G, Byrnes TJ, Opstad KS, McIntyre DJO, Griffiths JR, Bell BA, Clark CA, Barrick TR, Howe FA. Pattern recognition of MRSI data shows regions of glioma growth that agree with DTI markers of brain tumor infiltration. *Magn. Reson. Med.* 2009;62:1646–1651.

12. Luts J, Laudadio T, Idema AJ, Simonetti AW, Heerschap A, Vandermeulen D, Suykens JAK, Van Huffel S. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. *NMR Biomed.* 2009;22:374–390.

13. Lee JM. *Riemannian Manifolds: An Introduction to Curvature.* Springer, New York; 1997.

14. Bartlett MS, Movellan JR, Sejnowski TJ. Face recognition by independent component analysis. *Neural Networks, IEEE Trans.* 2002;13:1450–1464.

15. Schölkopf B, Smola A, Müller K-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* 1998;10:1299–1319.

16. Tenenbaum JB, De Silva V, Langford JC. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science.* 2000;290:2319–2323.

17. Law MHC, Jain AK. Incremental nonlinear dimensionality reduction by manifold learning. *Pattern Anal. Mach. Intell. IEEE Trans.* 2006;28:377–391.

18. Roweis ST, Saul LK. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science.* 2000;290:2323–2326.

19. Saul LK, Roweis ST. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *J. Mach. Learn. Res.* 2003;4:119–155.

20. Belkin M, Niyogi P. Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural Comput.* 2003;15:1373–1396.

21. Chen L, Tsang IW, Xu D. Laplacian Embedded Regression for Scalable Manifold Regularization. *IEEE Trans. Neural Networks Learn. Syst.* 2012;23:902–915.

22. Wu R, Yu Y, Wang W. ScaLE: Supervised and Cascaded Laplacian Eigenmaps for Visual Object Recognition Based on Nearest Neighbors. 2013 IEEE Conf. Comput. Vis. Pattern Recognit. 2013:867–874.

23. De Silva V, Tenenbaum J. Global versus local methods in nonlinear dimensionality reduction. *Adv. Neural Inf. Process. Syst.* 2003:721–728.

24. Charpiat G, Faugeras O, Keriven R, Maurel P. Distance-based shape statistics. *Acoust. Speech Signal Process.* 2006;5:925–928.

25. Luxburg U. A tutorial on spectral clustering. *Stat. Comput.* 2007;17:395–416.

26. Hyvärinen A, Karhunen J, Oja E. *Independent component analysis.* John Wiley & Sons; 2004 pp. 1–11, 287–289.

27. Hyvärinen A, Oja E. A fast fixed-point algorithm for independent component analysis. *Neural Comput.* 1997;1492:1483–1492.
28. Draper BA., Baek K, Bartlett MS, Beveridge JR. Recognizing faces with PCA and ICA. *Comput. Vis. Image Underst.* 2003;91:115–137.
29. MacKay D. Chapter 20. An Example Inference Task: Clustering. In: *Information Theory, Inference and Learning Algorithms*. Cambridge University Press; 2003. pp. 284–292.
30. Press W, Teukolsky S, Vetterling W, Flannery B. Section 16.4. Hierarchical Clustering by Phylogenetic Trees. In: *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge University Press; 2007. pp. 868–881.
31. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 1987;20:53–65.
32. Chen Y, Lin C. Combining SVMs with various feature selection strategies. In: Guyon I, Nikravesh M, Gunn S, Zadeh LA, editors. *Feature Extraction: Foundations and Applications*. Springer Berlin Heidelberg; 2006. pp. 315–324.
33. Tate AR, Underwood J, Acosta DM, et al. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. *NMR Biomed.* 2006;19:411–434.
34. Opstad KS, Ladroue C, Bell BA, Griffiths JR, Howe FA. Linear discriminant analysis of brain tumour ¹H MR spectra: a comparison of classification using whole spectra versus metabolite quantification. *NMR Biomed.* 2007;20:763–770.
35. Yang G, Raschke F, Barrick TR, Howe FA. Classification of Brain Tumour ¹H MR Spectra : Extracting Features by Metabolite Quantification or Nonlinear Manifold Learning? In: *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*. IEEE. ; 2014. pp. 1039–1042.
36. Prince SJD. *Computer vision: models, learning and inference*. 1st ed. Cambridge University Press; 2012. pp. 145–147.
37. García-Gómez J, Tortajada S, Vidal C. On the use of long TE and short TE SV MR spectroscopy to improve the automatic brain tumor diagnosis. Technical Report, Katholieke University Leuven; 2007; pp. 1–25.
38. Ortega-Martorell S, Lisboa PJG, Vellido A, Julià-Sapé M, Arús C. Non-negative matrix factorisation methods for the spectral decomposition of MRS data from human brain tumours. *BMC Bioinformatics* 2012;13:38.
39. McKnight TR, von dem Bussche MH, Vigneron DB, Lu Y, Berger MS, McDermott MW, Dillon WP, Graves EE, Pirzkall A, Nelson SJ. Histopathological validation of a three-

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

dimensional magnetic resonance spectroscopy index as a predictor of tumor presence. J. Neurosurg. 2002;97:794–802.

40. Wald LL, Nelson SJ, Day MR, et al. Serial proton magnetic resonance spectroscopy imaging of glioblastoma multiforme after brachytherapy. J. Neurosurg. 1997;87:525–534.

For Peer Review

Figure and Table Captions

Figure 1: Summary of pre-processed data and derived DR eigenvectors for SVS (a-g) and MRSI data (h-n). All pre-processed SVS and MRSI spectra are showed overlain in (a) and (h), respectively. For the SVS data, (b-d) show the three eigenvector components obtained after application of PCA+ICA and (e-g) are the three eigenvector components associated with the LE solution. For MRSI data analysis, the three independent components obtained after application of PCA+ICA are illustrated (i-k) with the three eigenvector components associated with the LE solution (l-n). The eigenvectors for each analysis are displayed in order for the tissue class they appear to represent: Eigenvector 1—normal brain; Eigenvector 2—Grade-II or infiltration; Eigenvector 3—Grade-IV or necrotic tissue. All spectra are displayed with a 0 to 4ppm x-axis and the spectra truncated to 0.2 – 4ppm as used in the analysis.

Figure 2: SVS data clustering results (a-d) and mean spectra for the various clusters (e-p). Scatter plots show the results of k-means (a and b) and agglomerative hierarchical clustering (c and d) when applied to the PCA+ICA (a and c) and LE-DR (b and d). Colour coding of the scatter plots is according to the tissue class most likely associated to the mean spectra of each cluster (e-p) with: green—normal brain tissue; blue—Grade-II; red—Grade-IV. Mean (solid red curve) and one standard deviation (dashed green curves) spectra are illustrated below each scatter plot. All spectra are displayed with a 0 to 4ppm x-axis and the spectra truncated to 0.2 – 4ppm as used in the analysis.

Figure 3: Assessment of the LE-DR clustering of the SVS experiment with respect to the clinical and histological based “ground truth”. Mean (solid red curve) and standard deviation spectra (dashed green curves) are shown for: (a) 79 normal; (b) 24 Grade-II; and (c) 31 Grade-IV classes. Panels (d-f) show non-typical cases of: (d) a normal spectrum with a Grade-II feature of elevated Cho to Cr ratio due to baseline distortion; (e) a Grade-II tumour spectrum with prominent Grade-IV features; and (f) a Grade-IV tumour spectrum with no lipid signals and therefore Grade-II characteristics. These atypical spectra are indicated within the LE-DR scatter plot (g), which is colour coded according to ground truth labels: green—normal brain tissue; blue—Grade-II; red—Grade-IV. Image (h) shows a FLAIR image of Grade-II case (e). Image (i) shows the T₂-weighted image of Grade-IV case (f). The planned SVS location is shown by red box in both cases. All spectra are displayed with a 0 to 4ppm x-axis and the spectra truncated to 0.2 – 4ppm as used in the analysis.

Figure 4: The MRSI data clustering results (a-d) and mean spectra for the various clusters (e to p). Scatter plot show the results of k-means (a and b) and agglomerative hierarchical clustering (c and d) when applied to the PCA+ICA (a and c) and LE-DR (b and d). Colour coding of the scatter plots is according to the tissue class most likely associated to the mean spectra of each cluster (e to p) with: green—normal brain tissue; blue—infiltrative tumour; red—high grade tumour/necrotic core. Mean (solid red curve) and one standard deviation (dashed green curves) spectra are illustrated for the clusters underneath each scatter plot. All spectra are displayed with a 0 to 4ppm x-axis and the spectra truncated to 0.2 – 4ppm as used in the analysis.

Figure 5: RGB overlays on conventional images showing the tissue group assigned to the spectrum in each MRSI voxel after agglomerative hierarchical clustering of the complete MRSI dataset. The top row illustrates results using PCA+ICA and the bottom row shows results using LE-DR data. Two Grade-II cases (left two columns), two Grade-III cases (middle two columns), and two Grade-IV cases (right two columns) are shown. Colour coding is: green—normal brain; blue—infiltrative tumour; red—high-grade/necrotic tumour. Non-coloured voxels within the MRSI grid were removed by our quality control criteria.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1: Summary of the number of patients studied and the number of spectroscopic voxels analysed for each tissue class. For the SVS experiment, we applied our methods to 55 SV tumour spectra combined with 79 normal spectra derived from the MRSI data of 3 normal controls. For the MRSI experiment, we applied our method to 1965 2D CSI spectra derived from Grade-II, Grade-III and Grade-IV patients. No extra normal control data were added into the MRSI analysis because the rectangular 2D CSI grid included enough contralateral normal voxels from the tumour patients.

Table 2: Summary evaluation of all the methods for SVS data (55 patients with 55 tumour spectra plus 79 spectra extracted from 3 healthy controls using MRSI). Bold text indicates the best performer or best performers if all equal.

Table 3: Summary evaluation of all the methods for MRSI data (30 patient cases with 1965 voxels). Bold text indicates the best performer or best performers if all equal.

Table 4: Summary of the quantitative validation of clustering techniques on DR results in the SVS experiment with respect to histopathological ground truth. Nor: normal brain tissue; GII: Grade-II tumour tissue; GIV: Grade-IV tumour tissue; Sens.: sensitivity; Spec.: specificity.

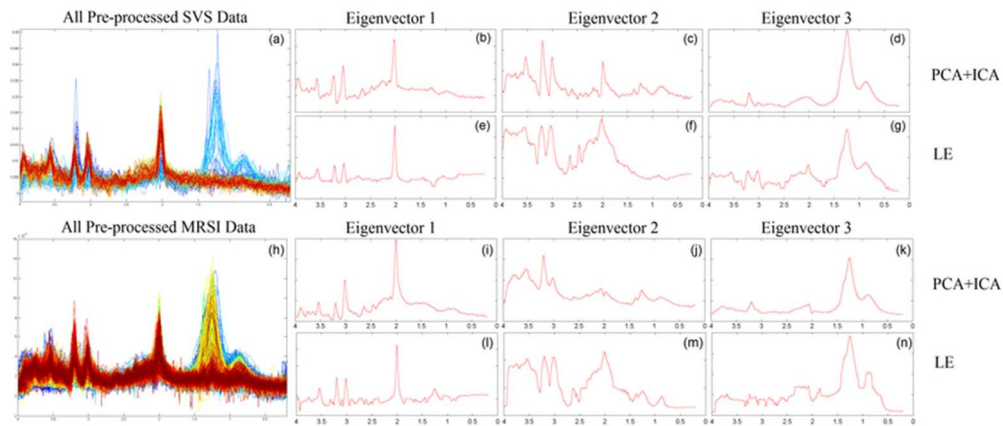


Figure 1: Summary of pre-processed data and derived DR eigenvectors for SVS (a-g) and MRSI data (h-n). All pre-processed SVS and MRSI spectra are showed overlain in (a) and (h), respectively. For the SVS data, (b-d) show the three eigenvector components obtained after application of PCA+ICA and (e-g) are the three eigenvector components associated with the LE solution. For MRSI data analysis, the three independent components obtained after application of PCA+ICA are illustrated (i-k) with the three eigenvector components associated with the LE solution (l-n). The eigenvectors for each analysis are displayed in order for the tissue class they appear to represent: Eigenvector 1—normal brain; Eigenvector 2—Grade-II or infiltration; Eigenvector 3—Grade-IV or necrotic tissue. All spectra are displayed with a 0 to 4ppm x-axis and the spectra truncated to 0.2 – 4ppm as used in the analysis.

34x14mm (600 x 600 DPI)

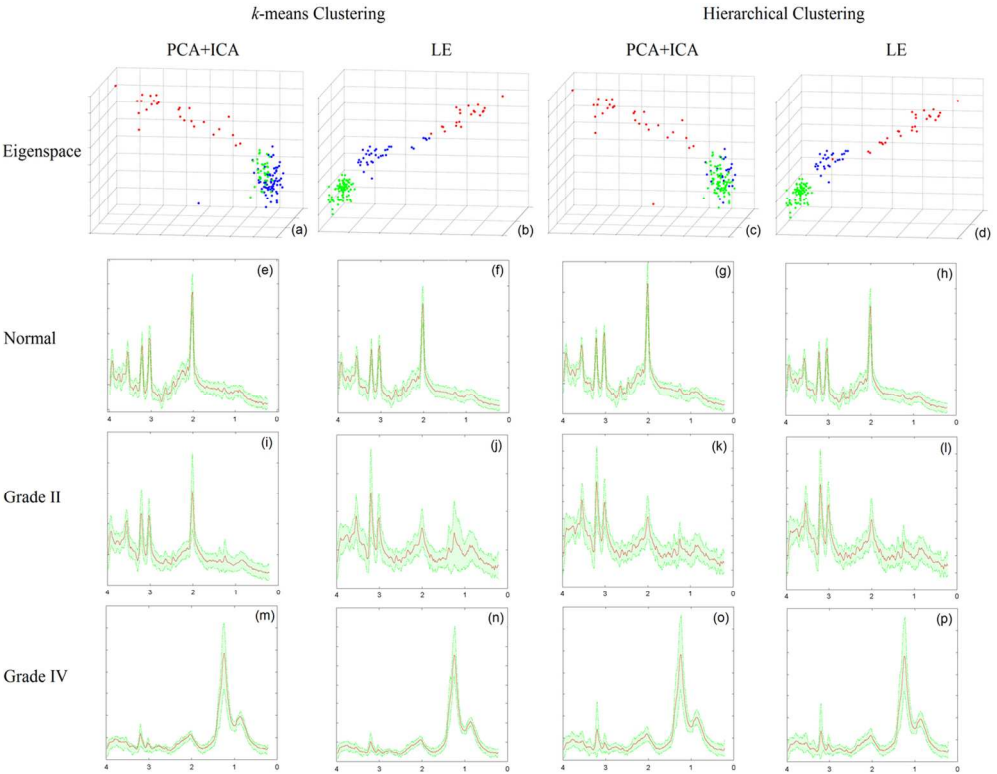


Figure 2: SVS data clustering results (a-d) and mean spectra for the various clusters (e-p). Scatter plots show the results of k-means (a and b) and agglomerative hierarchical clustering (c and d) when applied to the PCA+ICA (a and c) and LE-DR (b and d). Colour coding of the scatter plots is according to the tissue class most likely associated to the mean spectra of each cluster (e-p) with: green—normal brain tissue; blue—Grade-II; red—Grade-IV. Mean (solid red curve) and one standard deviation (dashed green curves) spectra are illustrated below each scatter plot. All spectra are displayed with a 0 to 4ppm x-axis and the spectra truncated to 0.2 – 4ppm as used in the analysis.
60x47mm (600 x 600 DPI)

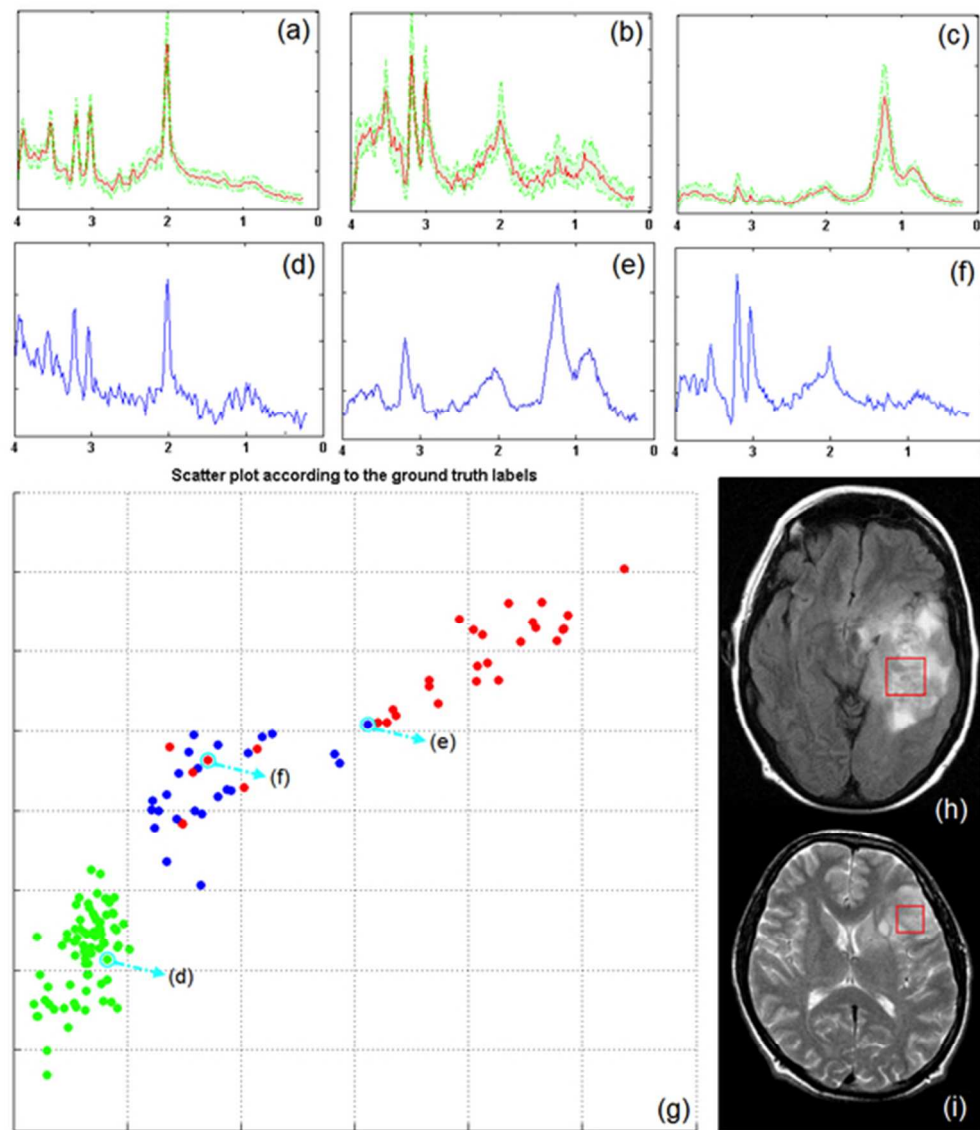


Figure 3: Assessment of the LE-DR clustering of the SVS experiment with respect to the clinical and histological based “ground truth”. Mean (solid red curve) and standard deviation spectra (dashed green curves) are shown for: (a) 79 normal; (b) 24 Grade-II; and (c) 31 Grade-IV classes. Panels (d-f) show non-typical cases of: (d) a normal spectrum with a Grade-II feature of elevated Cho to Cr ratio due to baseline distortion; (e) a Grade-II tumour spectrum with prominent Grade-IV features; and (f) a Grade-IV tumour spectrum with no lipid signals and therefore Grade-II characteristics. These atypical spectra are indicated within the LE-DR scatter plot (g), which is colour coded according to ground truth labels: green—normal brain tissue; blue—Grade-II; red—Grade-IV. Image (h) shows a FLAIR image of Grade-II case (e). Image (i) shows the T2-weighted image of Grade-IV case (f). The planned SVS location is shown by red box in both cases. All spectra are displayed with a 0 to 4ppm x-axis and the spectra truncated to 0.2 – 4ppm as used in the analysis.

24x28mm (600 x 600 DPI)

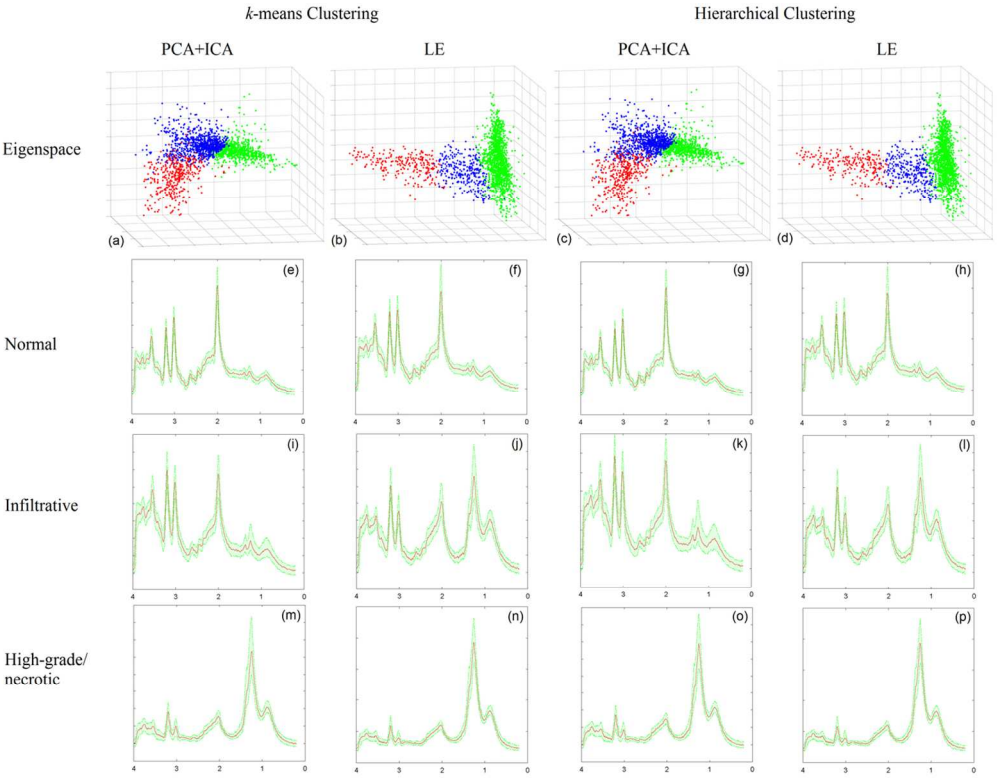


Figure 4: The MRSI data clustering results (a-d) and mean spectra for the various clusters (e to p). Scatter plot show the results of k-means (a and b) and agglomerative hierarchical clustering (c and d) when applied to the PCA+ICA (a and c) and LE-DR (b and d). Colour coding of the scatter plots is according to the tissue class most likely associated to the mean spectra of each cluster (e to p) with: green—normal brain tissue; blue—infiltrative tumour; red—high grade tumour/necrotic core. Mean (solid red curve) and one standard deviation (dashed green curves) spectra are illustrated for the clusters underneath each scatter plot. All spectra are displayed with a 0 to 4ppm x-axis and the spectra truncated to 0.2 – 4ppm as used in the analysis.

60x47mm (600 x 600 DPI)

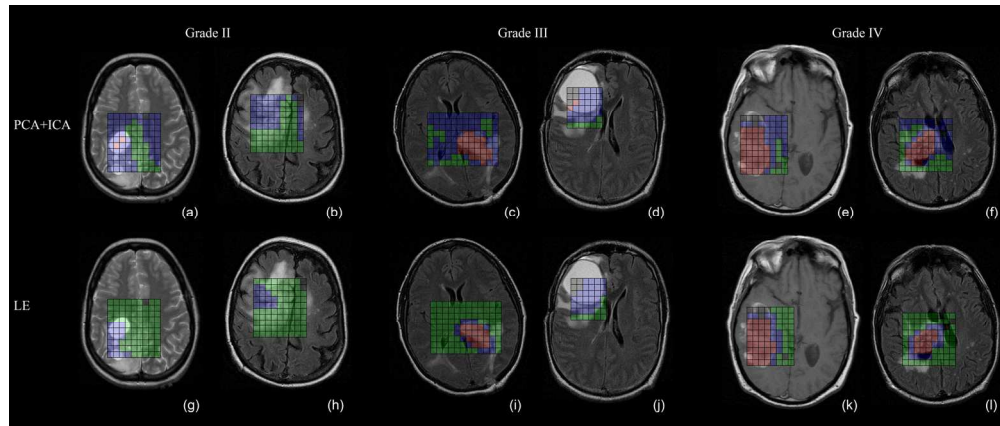


Figure 5: RGB overlays on conventional images showing the tissue group assigned to the spectrum in each MRSI voxel after agglomerative hierarchical clustering of the complete MRSI dataset. The top row illustrates results using PCA+ICA and the bottom row shows results using LE-DR data. Two Grade-II cases (left two columns), two Grade-III cases (middle two columns), and two Grade-IV cases (right two columns) are shown. Colour coding is: green—normal brain; blue—infiltrative tumour; red—high-grade/necrotic tumour. Non-coloured voxels within the MRSI grid were removed by our quality control criteria.

86x36mm (600 x 600 DPI)