Discrete Wavelet Transform-Based Whole-Spectral and Subspectral Analysis for Improved Brain Tumor Clustering Using Single Voxel MR Spectroscopy

Guang Yang*, *Member, IEEE*, Tahir Nawaz, Thomas R. Barrick, Franklyn A. Howe, and Greg Slabaugh, *Senior Member, IEEE*

***Abstract*—Many approaches have been considered for automatic grading of brain tumors by means of pattern recognition with magnetic resonance spectroscopy (MRS). Providing an improved technique which can assist clinicians in accurately identifying brain tumor grades is our main objective. The proposed technique, which is based on the discrete wavelet transform (DWT) of whole-spectral or subspectral information of key metabolites, combined with unsupervised learning, inspects the separability of the extracted wavelet features from the MRS signal to aid the clustering. In total, we included 134 short echo time single voxel MRS spectra (SV MRS) in our study that cover normal controls, low grade and high grade tumors. The combination of DWT-based whole-spectral or subspectral analysis and unsupervised clustering achieved an overall clustering accuracy of 94.8% and a balanced error rate of 7.8%. To the best of our knowledge, it is the first study using DWT combined with unsupervised learning to cluster brain SV MRS. Instead of dimensionality reduction on SV MRS or feature selection using model fitting, our study provides an alternative method of extracting features to obtain promising clustering results.**

***Index Terms*—Brain tumor, clustering, dimension reduction, discrete wavelet transform, glioma grade, magnetic resonance spectroscopy, unsupervised learning.**

I. Introduction

MAGNETIC resonance imaging (MRI) is a widely-used modality that facilitates the diagnosis and prognosis of brain tumors. Standard MRI sequences are routinely used to differentiate among various brain tumor types based on qualitative visual analyses of the represented soft tissue contrast. Indeed, more than 120 classes of brain tumors are known [1], which are categorized into four grades depending on the level of malignancy by the world health organization (WHO) [2]. The grading from low to high (I-IV) represents malignancy levels from biologically least aggressive to most aggressive brain tumors as shown by histological criteria, e.g., invasiveness, vascularity, and tumor growth rate [1]. Gliomas are the most common primary brain tumor and pretreatment assessment of grade is required; however, the sole use of standard MRI sequences may be insufficient for an accurate diagnosis [3]. The current gold standard for diagnosis of a suspicious abnormal mass is the histopathological analysis of a biopsy sample [4]. However, due to tumor heterogeneity a tumor may be undergraded if the area of greatest malignancy is not selected for biopsy.

Alternatively, *in vivo* ^{1}H magnetic resonance spectroscopy (MRS) can be used to noninvasively inspect the biochemical information of the metabolites present in the living tissue, and can improve characterization of human brain tumors compared to using standard MRI alone. There are up to 12 different metabolites in the brain that can be measured using ^{1}H MRS at clinical field strengths of 1.5 T or 3T [5]. In particular, single voxel (SV) MRS extracts metabolic information of a specific region of interest, and it is a unique noninvasive tool to aid classification of human brain tumors with appropriate spectral analysis such as with pattern recognition [6].

Analysis of ^{1}H MRS data for data clustering and tissue classification generally requires some form of data reduction, either to reduce the noise or extract the most salient features. Mainardi *et al.* [7] designed a quantification model for *in vivo* MRS parameters using the discrete wavelet packet decomposition (WPD). Mahmoodabadi *et al.* proposed a modified frequency ordered WPD method combined with fuzzy classification framework to analyze pediatric metabolic brain diseases using wavelet transform based features extracted from both MRS and diffusion-weighted imaging [8]. Tiwari *et al.* investigated combining wavelet (Gabor and Haar filters) features extracted from both T_2-weighted MRI and MRS modalities, applied to supervised prostate cancer detection [9]. The discrete wavelet transform (DWT) has the advantage of providing multiresolution discriminatory information [10] from different acquisition modalities, including but not limited to digital signals and images [9]. However, there is very limited research in the literature on fully-automating an unsupervised brain tumor data clustering using DWT based analysis that does not require labeled data or incur possible overfitting during the training procedure.

Here, we hypothesize that unsupervised learning based clustering on extracted DWT features can improve brain tumor grading compared to dimension reduction techniques, which are based on second and higher-order statistics (e.g., PCA

Manuscript received January 30, 2015; revised June 1, 2015; accepted June 17, 2015. Date of publication June 22, 2015; date of current version November 20, 2015. This work was supported by CRUK C1459/A13303 and a City University London Pump Priming Grant. *Asterisk indicates corresponding author.*

*G. Yang is with the Neuroscience Research Centre, Cardiovascular and Cell Sciences Institute, St. George's, University of London, London SW17 0RE, U.K., and National Heart and Lung Institute, Imperial College London, London SW3 6NP, U.K. (e-mail: gyang@sgul.ac.uk).

T. Nawaz was with the City University London, and is currently with Computational Vision Group, University of Reading.

T. R. Barrick and F. A. Howe are with the St. George's, University of London.

G. Slabaugh is with the City University London.

This paper has supplemental material available online at http://ieeexplore.ieee.org (File size: 1 MB).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2015.2448232

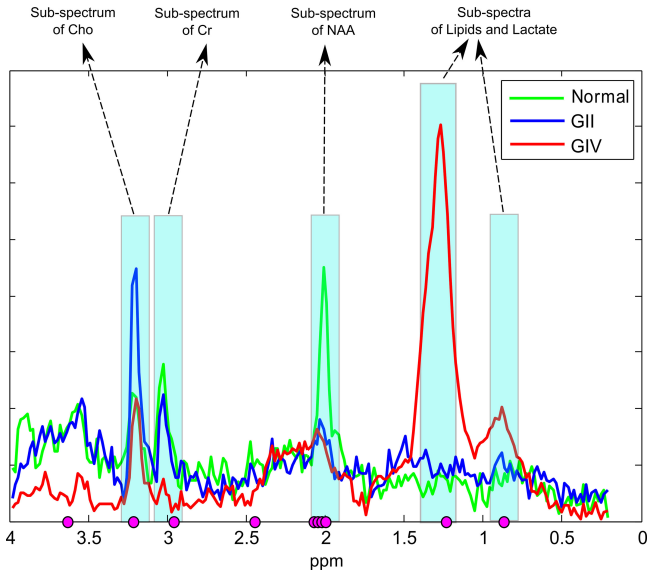

Fig. 1. Decomposition of a whole MR spectrum into a set of subspectra (subregions colored in cyan) corresponding to different metabolites. Representative grade II (GII) tumor, grade IV (GIV) tumor and normal spectra are shown in blue, red and green colors. The selected features using mRMR are shown with magenta circles.

and ICA) or manifold learning based nonlinear methods [e.g., Laplacian Eigenmaps (LE)]. We extract DWT features of the whole-spectra and subspectra (as shown in Fig. 1), and hypothesize that local information from only a few key metabolites of the subspectra is sufficient to distinguish between tumor grades, because those metabolites exhibit discriminative characteristics for specific tumor grades [6].

To accomplish this, we build a feature vector by using the DWT coefficients of the whole-spectra or by encoding the nonparametric statistics of the computed DWT coefficients of the subspectra corresponding to different metabolites. Then in an agglomerative hierarchical clustering framework the MR spectra belonging to different tumor grades are separated. We show the effectiveness of the proposed method on SV MRS data (134 spectra), acquired from normal brain tissue and from low and high grade gliomas. The proposed methods show encouraging performance by achieving an unsupervised clustering accuracy of 94.8% for both whole-spectral and subspectral analysis that outperforms our previous analyses of this data. Instead of extracting eigen-decomposed features using dimensionality reduction techniques, this study explores DWT features using whole-spectral and subspectral analysis, and obtains promising clustering results for separating different brain tissue types. Section II details the materials and main methods of this study. Section III demonstrates our experimental results followed by elaborated discussions (see Section IV) and a conclusion (see Section V).

II. Materials and Methods

A. Data Acquisition and Patient Subjects

MR data were obtained at St. George's University of London using a 1.5-T scanner (GE Healthcare, Milwaukee, WI, USA), which was equipped with 22 mT/m gradients and a quadrature head coil. Written informed consent was obtained from all participants in accordance with local ethics procedures. Either biopsy or resected tumor tissue samples obtained as part of the patients' clinical diagnosis or treatment were used to provide a histological diagnosis of the tumor type and grade as the overall gold standard (ground truth).

TABLE I
Number of Patients Studied and Number of Spectroscopic Voxels Analyzed for Each Tissue Class

	Tissue Class			
	Normal	Grade II	Grade IV	Total
Number of Subjects Studied	3	24	31	58
Number of MRS Voxels Analyzed	79	24	31	134

In total $\mathrm{N} = 134$ SV MRS were obtained including 24 Grade II (GII) tumors (two oligodendroglioma, three oligoastrocytoma, three fibrillary astrocytoma, four gemistocytic astrocytoma and 12 diffuse astrocytoma) and 31 Grade IV (GIV, glioblastoma multiforme). A further 79 MR spectra were obtained from three normal controls using multiple voxel MRS with the same acquisition parameters (i.e., which had compatible TR/TE) as the SV MRS (see Table I).

All SV MRS data were acquired at short echo time (TE) using the GE developed point-resolved spectroscopic sequence (PRESS) protocol [repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, 2048 data points with 2500 Hz bandwidth].

An expert panel (including spectroscopists, pathologists and radiologists) validated the brain tissue types included in this study as part of the eTUMOUR project, with a histopathological diagnosis of the central nervous system tumors according to WHO criteria [2]. Individual voxels were placed to encompass predominantly viable tumor tissue as much as possible and avoid areas of pure necrosis. Apodization in the time domain was performed using a half Hann window followed by a fast Fourier transform and automatic phasing according to [11]. Each spectrum was referenced to both N-acetyl Aspartate (NAA at 2 ppm and a search region $2.20 - 1.80$ ppm) and Choline (Cho at 3.21 ppm and a search region $3.30 - 3.12$ ppm) for chemical shift alignment, and then truncated to the chemical shift range of 4.0 to 0.2 ppm. In addition, the phased real part of the spectra were used for further analysis [12]–[14]. Each whole spectrum consisted of $\mathrm{M} = 198$ data points representing the majority of metabolic information.

B. Feature Extraction

All SV MRS data were stored as a 2-D matrix, $\mathbf{S} \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$, where $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{\mathrm{N}})$ contains each spectrum as a column vector $\mathbf{s}_{\mathrm{n}} \in \mathbb{R}^{\mathrm{M}}$. The column vectors were L_2 normalized

$$\mathbf{s}_n^* = \frac{\mathbf{s}_n}{\sqrt{\sum_m^{\mathrm{M}} \mathbf{s}_{nm}^2}} \quad \forall n \in \{1, 2, \ldots, \mathrm{N}\} \tag{1}$$

TABLE II
PARAMETER SETTINGS OF THE DWT METHOD USING WHOLE-SPECTRAL ANALYSIS AND SUBSPECTRAL ANALYSIS (WBF: WAVELET BASIS FUNCTIONS; LD: LEVEL OF DECOMPOSITION)

Whole-spectral	WBF		LD	Accuracy (Mean ± Standard Deviation)	Best Accuracy (with WBF and LD Settings)
	Daubechies (DbAll)	Db1	[1–12]	83.1% ± 1.6%	94.8% (with Db1 and seven levels of decomposition)
		Db2	[1–12]	82.5% ± 2.1%	
		Db3	[1–12]	78.3% ± 2.5%	
	Coiflets (CoifAll)	Coif1	[1–12]	81.0% ± 2.5%	**94.8**% (with **Coif1** and **six** levels of decomposition)
		Coif2	[1–12]	83.2% ± 1.6%	
		Coif3	[1–12]	79.8% ± 2.2%	
	Symlets (SymAll)	Sym1	[1–12]	83.2% ± 1.6%	92.5% (with Sym1 and seven or nine levels of decomposition)
		Sym2	[1–12]	83.2% ± 1.6%	
		Sym3	[1–12]	83.1% ± 1.6%	

Sub-spectral	WBF		LD	Window Sizes for Each Metabolite	Best Accuracy (with WBF and Window Sizes)
	Daubechies	Db1	1		
		Db2	1		
		Db3	1		
	Coiflets	Coif1	1	ω = [0.02 ppm, 0.04 ppm, . . . , 0.28 ppm, 0.30 ppm]	**94.8**% (with **Coif1** and $\omega^{\mathbf{NAA}}$ = **0.16 ppm,** $\omega^{\mathbf{Cho}}$ = **0.16 ppm,** $\omega^{\mathbf{Cr}}$ = **0.04 ppm,** $\omega^{\mathbf{Lip1}}$ = **0.18 ppm, and** $\omega^{\mathbf{Lip2}}$ = **0.20 ppm**)
		Coif2	1		
		Coif3	1		
	Symlets	Sym1	1		
		Sym2	1		
		Sym3	1		

Bold text indicates the best parameter settings.

to form the matrix $\mathbf{S}^* = \{\mathbf{s}_i^*\}_{i=1}^{\mathrm{N}}$ with $\mathbf{S}^* \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$, which is a set of normalized MR spectra. N is the number of the spectra and M is the number of the data points of each spectrum.

We used the DWT to encode the MR spectral information. For both the whole-spectral and subspectral analysis, we compared widely-used wavelet basis functions including Daubechies (Db1, Db2, Db3), Coiflets (Coif1, Coif2, Coif3) and Symlets (Sym1, Sym2, Sym3).

For the whole-spectral analysis, we tested multiple levels (i.e., one to 12 levels) of decomposition. At each level of resolution or octave, the spectral signal $\mathbf{s}_i^*$ is convolved ($\otimes$) simultaneously with a high-pass filter (ϕ_h) and a low-pass filter (ψ_l) to obtain the corresponding coefficients (h_c and l_c) as $\mathrm{h}_c = (\mathrm{s}_\mathrm{n}^* \otimes \phi_\mathrm{h}) \downarrow 2$ and $\mathrm{l}_c = (\mathrm{s}_\mathrm{n}^* \otimes \psi_\mathrm{l}) \downarrow 2$ with the subsampling or decimation ($\downarrow$) by a factor of 2 after each pass through the paired filters [15]. The feature vectors are computed for all $\mathbf{s}_i^*$ to obtain the feature matrix, of size $134 \times \mathrm{F_L}$, in which $\mathrm{F_L}$ is the size of the approximation and detail coefficients at different levels $\mathrm{L} = 1, 2, \ldots 12$ (see Table II).

For the subspectral analysis by given $\mathbf{s}_i^*$ we built the feature vector, $\mathbf{f}_i$, as follows. We extracted information for the key metabolites by taking the subspectra windowed symmetrically around their respective ppms (see Fig. 1). We performed a mutual information based feature selection (i.e., mRMR, minimum redundancy and maximum relevance method [16]) to identify the key metabolites, which are the most powerful discriminants in terms of separating the three tissue types (Normal, GII and GIV). The only parameter specified in the mRMR method is the number of features, which was set to 10 in order to obtain a moderate sample per feature ratio to avoid overfitting [17]. The mRMR method resulted in ten selected features (see Fig. 1) at 3.62, 3.24, 2.99, 2.49, 2.05, 2.03, 2.01, 1.99, 1.19, and 0.85 ppm. For subspectral analysis we chose dominant high signal to noise ratio biochemicals [18] whose peak areas are mostly strongly

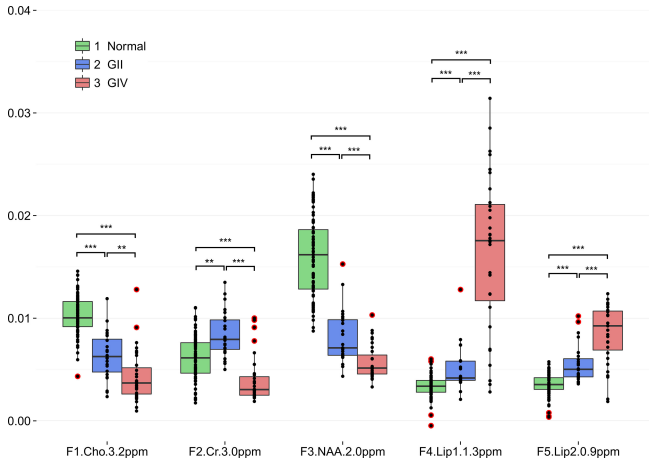

Fig. 2. Boxplot of the five selected features among three tissue types (Normal: green; GII: blue; GIV: red). Red circled dots are the outliers of each group (*** indicates $p < 0.0001$ and ** indicates $p < 0.001$).

associated with these regions: NAA (main singlet at 2.05 ppm), Cho (3.21 ppm), Creatine (Cr) (3.02 ppm) and lipids and macromolecules (main peaks at 1.3 and 0.9 ppm). The excluded points of 3.62 and 2.49 ppm include strongly overlapping multiplet peaks from Myo-inositol with glutamate and glutamine (Glx) and those from Glx with NAA, respectively. Note in high-grade tumors there may also be a contribution from lactate (doublet centered at 1.33 ppm). Boxplots of the five selected features among three different tissue types are shown in Fig. 2, and statistical significances were given by two-sample Wilcoxon rank-sum test between each two tissue types (significance level of $p < 0.01$ subject to the Bonferroni correction [19]). In general, NAA and Cr are decreased and Cho increased in tumors compared to normal brain [20]. Additionally, the lipid and lactate signal observable at short TE increase with tumor grade

TABLE III
PERFORMANCE COMPARISON OF THE PROPOSED DWT METHOD (WHOLE-SPECTRAL ANALYSIS USING COIF1 WITH SIX LEVELS OF DECOMPOSITION OR SUBSPECTRAL ANALYSIS USING COIF1 WITH ONE LEVEL OF DECOMPOSITION) WITH PCA, PCA + ICA AND NONLINEAR LE METHODS

Method	Brain Tissue Types	Predicted Group Membership			Precision (*P*)	Recall (*R*)	F-Score (*F*)	Accuracy (*A*)	Silhouette Statistics ($\mu_{SW} \pm \sigma_{SW}$)	BER
		Normal	GII	GIV						
PCA	Normal	71	8	0	1.00	0.90	0.95	0.873	0.521 ± 0.230	0.140
	GII	0	21	3	0.60	0.88	0.71			
	GIV	0	6	25	0.89	0.81	0.85			
PCA + ICA	Normal	79	0	0	1.00	1.00	1.00	0.933	0.551 ± 0.213	0.106
	GII	0	21	3	0.78	0.88	0.82			
	GIV	0	6	25	0.89	0.81	0.85			
LE	Normal	79	0	0	1.00	1.00	1.00	0.933	0.612 ± 0.201	0.106
	GII	0	21	3	0.78	0.88	0.82			
	GIV	0	6	25	0.89	0.81	0.85			
DWT	Normal	79	0	0	1.00	1.00	1.00	0.948	0.743 ± 0.278	0.078
	GII	0	23	1	0.79	0.96	0.87			
	GIV	0	6	25	0.96	0.81	0.88			

[20]. The subspectra for NAA, Cho, Cr, Lipid (0.9 ppm) and lipid and lactate (1.3 ppm) were denoted as $\mathbf{s}_i^{*\mathrm{NAA}}$, $\mathbf{s}_i^{*\mathrm{Cho}}$, $\mathbf{s}_i^{*\mathrm{Cr}}$, $\mathbf{s}_i^{*\mathrm{Lip1}}$, and $\mathbf{s}_i^{*\mathrm{Lip2}}$, respectively. The window sizes for each s_i^* were fixed and denoted as w^{NAA}, w^{Cho}, w^{Cr}, w^{Lip1}, and w^{Lip2}, respectively. We applied the DWT using the single-level implementation of Mallat's approach [10] on each of the $\mathbf{s}_i^{*\mathrm{NAA}}$, $\mathbf{s}_i^{*\mathrm{Cho}}$, $\mathbf{s}_i^{*\mathrm{Cr}}$, $\mathbf{s}_i^{*\mathrm{Lip1}}$, and $\mathbf{s}_i^{*\mathrm{Lip2}}$ signals to obtain the corresponding set of (approximation and detail) coefficients $\mathbf{D}_i^{\mathrm{NAA}}$, $\mathbf{D}_i^{\mathrm{Cho}}$, $\mathbf{D}_i^{\mathrm{Cr}}$, $\mathbf{D}_i^{\mathrm{Lip1}}$, $\mathbf{D}_i^{\mathrm{Lip2}}$, respectively. The higher levels of decomposition were not needed given the smaller subspectra window sizes compared to the whole-spectral analysis, and single-level decomposition was expected to be sufficient to effectively capture the frequency information.

The choice of the window sizes (w) for different metabolites can influence the performance of the system. In addition, the window size must be kept small to avoid overlap between of adjacent metabolites, and we also need enough data points for further data clustering. We analyzed the effect of the variation that $w = [0.02\ \text{ppm}, 0.04\ \text{ppm}, \ldots, 0.28\ \text{ppm}, 0.30\ \text{ppm}]$, i.e., 15 sets of window sizes [−0.01 ppm, 0.01 ppm], [−0.02 ppm, 0.02 ppm], ..., [−0.14 ppm, 0.14 ppm], [−0.15 ppm, 0.15 ppm around each peak of the metabolites) with respect to clustering performance (see Table III). The window size for each metabolite was varied in turn with the window sizes of the remaining metabolites kept fixed, and inherent to this process is the assumption that the window sizes are independent in terms of optimization. The first key metabolite was Cho (centered at 3.21 ppm) and the window was varied with window size for the other metabolites fixed at the initial smallest value. Subsequently, the window sizes for Cr centered at 3.02 ppm, NAA at 2 ppm, Lip2 at 1.3 ppm, and Lip1 at 0.9 ppm were optimized in turn.

Inspired by [21], we further encapsulated the distribution of the computed DWT coefficients nonparametrically for a metabolite, e.g., NAA, as follows:

$$\mathbf{f}_i^{\mathrm{NAA}} = \left\{\min\left(\mathbf{D}_i^{\mathrm{NAA}}\right),\ \mathrm{Q}_{25}\left(\mathbf{D}_i^{\mathrm{NAA}}\right),\ \mathrm{Q}_{50}\left(\mathbf{D}_i^{\mathrm{NAA}}\right),\ \mathrm{Q}_{75}\left(\mathbf{D}_i^{\mathrm{NAA}}\right),\ \max\left(\mathbf{D}_i^{\mathrm{NAA}}\right)\right\} \quad (2)$$

where $\mathbf{f}_i^{\mathrm{NAA}}$ denotes the feature vector that encodes the nonparametric statistics for the DWT coefficients ($\mathbf{D}_i^{\mathrm{NAA}}$) of $\mathbf{s}_i^{*\mathrm{NAA}}$ using the minimum coefficient value ($\min\left(\mathbf{D}_i^{\mathrm{NAA}}\right)$), 25th percentile ($\mathrm{Q}_{25}\left(\mathbf{D}_i^{\mathrm{NAA}}\right)$), 50th percentile ($\mathrm{Q}_{50}\left(\mathbf{D}_i^{\mathrm{NAA}}\right)$), 75th percentile ($\mathrm{Q}_{75}\left(\mathbf{D}_i^{\mathrm{NAA}}\right)$), and the maximum coefficient value ($\max\left(\mathbf{D}_i^{\mathrm{NAA}}\right)$). $\mathbf{f}_i^{\mathrm{NAA}}$ comprehensively captures the information of $\mathbf{s}_i^{*\mathrm{NAA}}$ by encoding the overall distribution of its coefficients [21]. Similarly to (2), the feature vectors for $\mathbf{s}_i^{*\mathrm{Cho}}$, $\mathbf{s}_i^{*\mathrm{Cr}}$, $\mathbf{s}_i^{*\mathrm{Lip1}}$, and $\mathbf{s}_i^{*\mathrm{Lip2}}$ can be computed and denoted as $\mathbf{f}_i^{\mathrm{Cr}}$, $\mathbf{f}_i^{\mathrm{Cho}}$,$\mathbf{f}_i^{\mathrm{Lip1}}$,$\mathbf{f}_i^{\mathrm{Lip2}}$, respectively. The feature vector, $\mathbf{f}_i$, is therefore defined as follows:

$$\mathbf{f}_i = \left\{\mathbf{f}_i^{\mathbf{NAA}}, \mathbf{f}_i^{\mathbf{Cho}}, \mathbf{f}_i^{\mathbf{Cr}}, \mathbf{f}_i^{\mathbf{Lip1}}, \mathbf{f}_i^{\mathbf{Lip2}}\right\} \quad (3)$$

and $\mathbf{f}_i$ is a 25-dimensional row vector. Using (3), the feature vectors are computed for all $\mathbf{s}_i^*$ to obtain the feature matrix, of size 134×25.

Unsupervised learning based hierarchical clustering is then performed on the feature matrix extracted using whole-spectral or subspectral analysis as described below.

C. Data Clustering

To quantitatively validate and compare the efficacy of our DWT based feature extraction method to previous studies, we applied agglomerative hierarchical clustering algorithms to the feature extraction outputs. Compared to widely used *k*-means clustering, hierarchical clustering requires no initialization settings, and thus can avoid possible local minima that could trap the *k*-means algorithm.

For hierarchical clustering a dissimilarity measure was specified (the Euclidean distance) between disjoint groups of observations according to pairwise dissimilarities between the observations in the two groups. An agglomerative (i.e., bottom-up) paradigm was used which recursively merges pairs of clusters into a single cluster at each level [22]. Pairs were merged based on the smallest intergroup dissimilarity and representation of the recursive binary agglomeration was achieved using dendrograms (i.e., rooted binary trees).

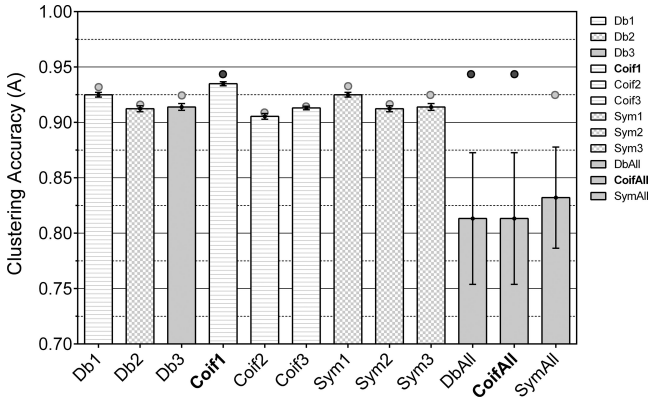

Fig. 3. Comparison between different wavelet basis functions in terms of the obtained clustering accuracy (*A*) for both whole-spectral and subspectral analysis. Error bars represent the standard error of the mean (SEM). For the whole-spectral analysis, SEM was calculated with respect to different decomposition levels. For the subspectral analysis, SEM was computed with respect to various window sizes. Circles above the error bars indicate the maximum accuracy achieved by different wavelet basis functions.

D. Performance Assessment

We evaluated the performance of the proposed method from a clinical point of view using the following measures: precision (P), recall/sensitivity (R), F-score (F) and clustering accuracy (A). $P = \frac{TP}{TP+FP}$ where TP and FP are the number of true positives (correct estimations) and false positives (incorrect estimations). $R = \frac{TP}{TP+FN}$ where FN is the number of false negatives (missed estimations). In addition, $F = 2\frac{P \times R}{P+R}$. Furthermore, A provides the overall clustering accuracy as a ratio of the number of correct clustering ($\mathrm{N_{Correct}}$) and total number of input spectra (N) that is $A = \frac{\mathrm{N_{Correct}}}{\mathrm{N}}$. P, R, F and $A \in [0, 1]$, and the higher P, R, F and A, the better the performance. In addition, we also calculated the balanced error rate [23] (BER), which is the average of the errors on each class, and is suitable for unbalanced datasets.

We also evaluated the discriminative ability of the feature vector, $\mathbf{f}_i$, in terms of quantifying intracluster tightness and intercluster separability. To this end we used the silhouette statistics that is computed for the ith data point as follows [24]: $\mathrm{sw}_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \in [-1, 1]$, where a_i is the average distance of the ith data point to the remaining points within the same cluster and b_i is the minimum average distance of the ith data point to any of the remaining clusters. We used the mean (μ_{SW}) and the standard deviation (σ_{SW}) of sw_i values as performance indicators. Higher μ_{SW} with lower σ_{SW} indicate better partitioning of the clustering results and hence the better discriminative ability of the feature used. $\mu_{\mathrm{SW}} > 0.5$ indicates a proper partitioning whereas $\mu_{\mathrm{SW}} < 0.2$ indicates an improper partitioning [24].

III. Results

Fig. 3 shows the comparison of the widely-used wavelet basis functions based on the clustering accuracy (A) for both whole-spectral and subspectra analysis. Examples of using Daubechies (Db1, Db2, Db3), Coiflets (Coif1, Coif2, Coif3) and Symlets (Sym1, Sym2, Sym3) of the whole spectra are shown as DbAll, CoifAll and SymAll. For the whole spectra, we tested multiple levels of decomposition with standard error of the mean (SEM) shown in the Fig. 3. The minimum accuracy (40.3%) was obtained using Coif1 with three levels of decomposition while the maximum accuracy (94.8%) was achieved by Coif1 using six levels or Db1 using seven levels of decomposition (see Table II). The Coif1 basis function achieved the highest clustering accuracy with a low level of decomposition, hence this represents the best option for the whole-spectral analysis. In addition, for the subspectral analysis, we display the SEM with respect to various window sizes (see Fig. 3). We tested different values for w^{NAA}, w^{Cho}, w^{Cr}, w^{Lip1}, and w^{Lip2} as aforementioned to maximize clustering accuracy for each wavelet basis function. In so doing, we obtained the best clustering accuracy (94.8% in Table II) using the Coif1 basis function with window sizes found to be $w^{\mathrm{NAA}} = 0.16$ ppm, $w^{\mathrm{Cho}} = 0.16$ ppm, $w^{\mathrm{Cr}} = 0.04$ ppm, $w^{\mathrm{Lip1}} = 0.18$ ppm and $w^{\mathrm{Lip2}} = 0.20$ ppm, respectively. The whole-spectral analysis (using Coif1 basis function with six levels of decomposition) and subspectral analysis (using Coif1 with one level of decomposition) performed identically (see Table II). For the whole-spectral analysis, clustering accuracies obtained by the Coif1 basis function with different decomposition levels showed no significant difference compared to the results of using the Db1 basis function, but showed significant differences compared to the results of other basis functions. For the subspectral analysis, clustering accuracies obtained by the Coif1 basis function with various window sizes showed significant differences compared to the results of using other basis functions (nonparametric Kruskal–Wallis test with Dunn's multiple comparisons was performed with a significance level of $p < 0.05$).

Comparison results between DWT and other methods are provided quantitatively (see Table III). From a clinical point of view, the whole (six levels of decomposition) and subspectral (one level of decomposition) DWT method using Coif1 (from this point onwards referred to as the proposed DWT method with the best settings as seen in Table II) showed the best performance for GII, GIV and normal spectra, as reflected by its highest precision (P), recall (R) and F-score (F) values (see Table III). For the case of normal spectra, PCA+ICA, LE and the proposed DWT method obtained $P = R = 1$ meaning that all of the normal spectra were perfectly separated without any false positives ($FP = 0$) and false negatives ($FN = 0$). PCA also obtained a precision of 1 but its recall rate of $R = 0.9$ is lower, as it misclassified eight normal spectra ($FN = 8$). For the case of GIV spectra, PCA, PCA + ICA and LE produce $FP = 3$ each with $P = 0.89$. The proposed DWT method produced $FP = 1$ and thus resulted in a better $P = 0.96$. On the other hand, all methods missed 6 GIV spectra with $R = 0.81$. For the case of GII spectra, the proposed DWT method outperformed the remaining methods both in terms of precision and recall. This method produced the least number of incorrect classifications ($\mathrm{N} - \mathrm{N_{Correct}} = 7$) thus obtaining the best A (see Table III), but incorrectly classified one GII spectrum as a GIV spectrum, and six GIV spectra as GII spectra. In this case, the incorrect clustering of GIV spectra in the GII cluster, show spectral characteristics that make it nontrivial to distinguish them as

GIV spectra. Moreover, the proposed DWT method showed the highest $\mu_{\mathrm{SW}} = 0.743$, demonstrating better separability of clusters of the feature, $\mathbf{f}_i$, than for those extracted using PCA, PCA + ICA and LE methods (see Table III). In terms of σ_{SW} LE shows the smallest variation in the silhouette values (see Table III).

IV. Discussion

Overall, the experimental results suggest that our wavelet based feature extraction (by whole-spectral or subspectral analysis) and clustering provides the maximum accuracy of 94.8% for tissue separation, which demonstrates an improvement compared to the PCA, PCA + ICA, and nonlinear LE methods. In addition, our accuracies for distinguishing GII from GIV tumors using DWT and hierarchical clustering compares well to the results of Garcia-Gomez *et al.* [25], who achieved 92.58% accuracy for distinguishing low from high grade gliomas by combining data from SV long echo and short echo data, which would be difficult to routinely achieve with MRS acquisitions due to time limits for patient scanning. A more recent approach [26] used nonnegative matrix factorization (NMF) to accomplish the feature extraction task for SV tumor spectra. However, NMF methods inevitably converge to local minima and various initializations provide different dimensionality reduction; therefore, NMF requires an elaborate initialization scheme as discussed in previous investigation [26]. In addition, we achieved similar BER (0.078 in Table II) using an unsupervised learning based clustering as Ortega-Martorell *et al.* [26] obtained with a supervised learning based framework. The short echo SV MRS data used in the current study is a subset of the INTERPRET and eTUMOUR project data used in [25] and [26] and are compatible for acquisition parameters and ground truth diagnosis from the histopathological and clinical information.

Interestingly, our results on whole spectra showed large variance of the clustering accuracy with respect to different decomposition levels. The decomposition of the MRS signals using multiple levels of resolution present with much larger variance in accuracy compared to the subspectra analysis while varying the window sizes. Suggesting subspectral DWT analysis may be more stable. In addition, average results of using different basis functions of the subspectral analysis are superior to the average results of using the whole spectra (see Fig. 3). However, the maximum clustering accuracy determined from the whole spectra is 94.8% using the Coif1 basis function and six levels of decomposition, which is still superior to PCA, PCA + ICA and nonlinear LE methods. Nevertheless, the minimum clustering accuracy is 40.3% using Coif1 basis function with three levels of decomposition indicating that DWT analysis on whole spectra is unstable with respect to decomposition levels.

The improvement provided by the proposed DWT method is a better separation between GII and GIV tumors. There is one misclassified GII spectrum as GIV [see Fig. 4(a) and (c)], and six misclassified GIV as GII [one example shown in Fig. 4(b) and (d)]. The one misclassified GII spectrum has unexpectedly high peaks of Lipid [see Fig. 4(c)]. In contrast, all misclassified GIV spectra have very low lipid signal, and so resemble GII spectra. The misclassification of these tumor spectra may be due to the fact that our ground truth is defined by local biopsy or resected tumor samples that do not accurately represent the tissue that has been investigated by MRS. The tumor biopsy samples used for histopathology have typical volumes in mm^3, whereas the SV MRS data are acquired from a volume in cm^3 [14]. In particular, the highest grade tumor cells observed in the biopsy determine the clinical assignation of tumor grade and may only represent a small proportion of the tissue in the MRS voxel. Additionally GIV tumors are frequently large heterogeneous masses that have areas of low-grade appearance by ^{1}H MRS [27], and the most malignant region may not have been sampled by a single voxel placement.

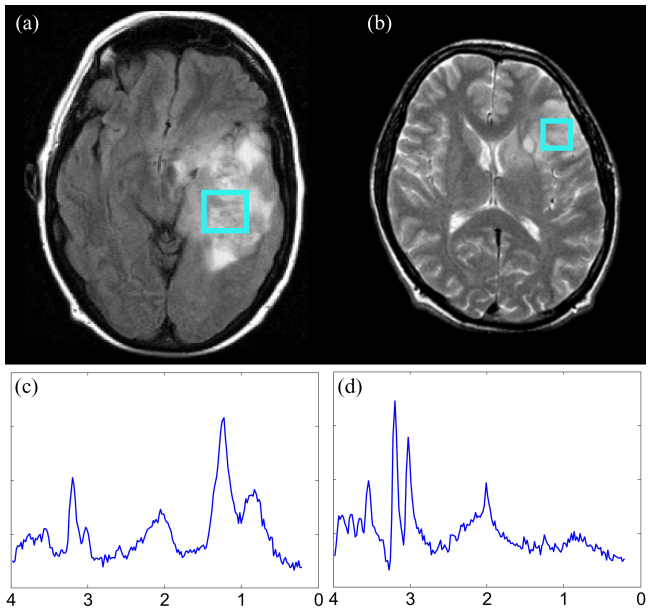

Fig. 4. (a) FLAIR image of the misclassified GII case; (b) T2-weighted image of one misclassified GIV case; (c) Blue curve: SV MRS signal of (a); (d) Blue curve: SV MRS signal of (b); Planned SV MRS acquisitions are shown in cyan boxes overlaid on structural MRI images.

Our analysis may have some potential limitations. As aforementioned, the whole-spectral analysis suffers from large variance of the clustering accuracy when decomposition levels of the DWT are varied. For the subspectral analysis, we rely upon predefinition of the key biochemical peaks, which require elaborate tuning prior to application of DWT. However, once the window sizes are fixed, more SV MRS data (i.e., more normal, GII, and GIV SV MRS data) may be added without retuning. In addition, there may be diagnostically useful contributions to the spectra from biochemicals with lower overall visibility. For example quantified levels of myo-inositol, glutathione, glutamate-glutamine have been used in previous tumor classification studies, metabolites whose major contributions to an MR spectrum are outside of our selected spectral regions [13], [18]. Our selection of five key biochemical used in this study may have certain subjectivity. Nevertheless, these metabolites and lipid peaks are well known and widely used, features to discriminate brain tumor grades. For example, Opstad *et al.* [18] indicated that choline, creatine, lactate and lipid (1.3 ppm) were the most discriminative for GII and GIV tumor, and NAA

and lipid (0.9 ppm) were useful for classifying normal spectra. Moreover, we only included GII and GIV patients for this study due to lack of reliable GIII MRS data. However, we can envisage a straightforward application of the current DWT based feature extraction and unsupervised clustering framework for SV MRS dataset incorporating GIII cases.

V. Conclusion

To the best of our knowledge, this is the first study using DWT and unsupervised clustering to separate SV MRS data from different brain tumor types. We tested both DWT based whole-spectral and subspectral analysis, and we have concluded that a subspectral analysis is sufficient by using windowed key metabolites to distinguish different grades of the brain tumor. The achievement is threefold: 1) we compared different DWT settings including various wavelet basis functions for both whole-spectral and subspectral analysis, different window sizes for the subspectral analysis, and multiple levels of decomposition for the whole-spectral analysis, and we have found that Coif1 wavelet obtained the best clustering results; 2) we compared DWT based subspectral analysis with DWT feature extraction on the whole spectra. Quantitative evidence show that our subspectral analysis is more stable irrespective of the window sizes selected for the key metabolites; 3) we also compared to conventional feature extraction methods such as PCA, PCA + ICA, and newly applied nonlinear LE algorithm, and the comparison demonstrated that both our DWT based whole-spectral and subspectral analysis can further improve the separation between GII and GIV tissue types while maintaining the accuracy of separating tumor spectra from normal brain spectra in controls. In summary, our DWT based feature extraction and hierarchical clustering produces promising brain tumor classification that has potential for analysis of larger multicenter datasets and be applicable to automated analysis of the large datasets obtained in multivoxel ^{1}H MRS using chemical shift imaging.

References

[1] V. P. Collins, "Brain tumours: Classification and genes," *J. Neurol. Neurosurg. Psychiatry*, vol. 75, pp. ii2–ii11, Jun. 2004.

[2] D. N. Louis *et al.*, "The 2007 WHO classification of tumours of the central nervous system," *Acta Neuropathol.*, vol. 114, no. 2, pp. 97–109, Aug. 2007.

[3] J. Faehndrich *et al.*, "Neuroradiological viewpoint on the diagnostics of space-occupying brain lesions," *Clin. Neuroradiol.*, vol. 21, no. 3, pp. 123–39, Sep. 2011.

[4] M. Caulo *et al.*, "Data-driven grading of brain gliomas: A multiparametric MR imaging study," *Radiology*, vol. 272, no. 2, pp. 494–503, 2014.

[5] R. de Graaf, *In Vivo NMR Spectroscopy*. New York, NY, USA: Wiley, 2007.

[6] A. Devos *et al.*, "Classification of brain tumours using short echo time 1H MR spectra," *J. Magn. Reson.*, vol. 170, no. 1, pp. 164–175, Sep. 2004.

[7] L. T. Mainardi *et al.*, "A wavelet packets decomposition algorithm for quantification of in vivo (1)H-MRS parameters," *Med. Eng. Phys.*, vol. 24, no. 3, pp. 201–208, Apr. 2002.

[8] S. Z. Mahmoodabadi *et al.*, "A novel m CAD for pediatric metabolic brain diseases incorporating DW imaging and MR spectroscopy," *Expert Syst.*, vol. 30, no. 1, pp. 21–33, Feb. 2013.

[9] P. Tiwari *et al.*, "Multimodal wavelet embedding representation for data combination (MaWERiC): Integrating magnetic resonance imaging and spectroscopy for prostate cancer detection," *NMR Biomed.*, vol. 25, no. 4, pp. 607–19, Apr. 2012.

[10] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

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

[12] G. Yang *et al.*, "Nonlinear Laplacian Eigenmaps dimension reduction of in-vivo magnetic resonance spectroscopic imaging analysis," in *Proc. Int. Soc. Magnetic Resonance Med. 21st Annu. Meet.*, 2013, p. 1967.

[13] G. Yang *et al.*, "Classification of brain tumour 1H MR spectra: Extracting features by metabolite quantification or nonlinear manifold learning?," in *Proc. Biomed. Imag. 11th Int. Symp.*, 2014, pp. 1039–1042.

[14] G. Yang *et al.*, "Manifold learning in MR spectroscopy using nonlinear dimensionality reduction and unsupervised clustering," *Magn. Reson. Med.*, 2014. DOI: 10.1002/mrm.25447.

[15] M. Weeks, *Digital Signal Processing Using MATLAB and Wavelets*. Boston, MA, USA: Jones & Bartlett, 2010.

[16] H. Peng *et al.*, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[17] R. L. Somorjai *et al.*, "Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: Curses, caveats, cautions," *Bioinformatics*, vol. 19, no. 12, pp. 1484–1491, Aug. 2003.

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

[19] H. Abdi, "The Bonferonni and Šidák corrections for multiple comparisons," in *Encyclopedia of Measurement and Statistics*. Newbury Park, CA, USA: Sage, 2007, pp. 103–107.

[20] F. A. Howe and K. S. Opstad, "1H MR spectroscopy of brain tumours and masses," *NMR Biomed.*, vol. 16, no. 3, pp. 123–131, May 2003.

[21] T. Nawaz *et al.*, "Trajectory clustering for motion pattern extraction in aerial videos," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 1016–1020.

[22] W. Press *et al.*, "Section 16.4. Hierarchical clustering by phylogenetic trees," in *Numerical Recipes: The Art of Scientific Computing*. New York, NY, USA: Cambridge Univ. Press, 2007, pp. 868–881.

[23] Y. Chen and C. Lin, "Combining SVMs with various feature selection strategies," in *Feature Extraction: Foundations and Applications*. Berlin, Germany: Springer, 2006, pp. 315–324.

[24] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

[25] J. García-Gómez *et al.*, "On the use of long TE and short TE SV MR spectroscopy to improve the automatic brain tumor diagnosis," Biomedical Data Mining Research Line, Katholieke Univ. Leuven, Leuven, Belgium, Tech. Rep. 07-55. pp. 1–25, 2007.

[26] S. Ortega-Martorell *et al.*, "Non-negative matrix factorisation methods for the spectral decomposition of MRS data from human brain tumours," *BMC Bioinformat.*, vol. 13, no. 1, pp. 1–20, Jan. 2012.

[27] F. Raschke *et al.*, "1H 2D MRSI tissue type analysis of gliomas," *Magn. Reson. Med.*, vol. 73, no. 4, pp. 1381–1389, Jun. 2014.

Authors' photographs and biographies not available at the time of publication.