

Voxel-based clustered imaging by multiparameter diffusion tensor images for glioma grading



Rika Inano^{a,b}, Naoya Oishi^{b,*}, Takeharu Kunieda^a, Yoshiki Arakawa^a, Yukihiro Yamao^{a,b}, Sumiya Shibata^{a,b}, Takayuki Kikuchi^a, Hidenao Fukuyama^b, Susumu Miyamoto^a

^aDepartment of Neurosurgery, Kyoto University Graduate School of Medicine, Kyoto, Japan

^bHuman Brain Research Center, Kyoto University Graduate School of Medicine, Kyoto, Japan

ARTICLE INFO

Article history:

Received 2 June 2014

Received in revised form 15 July 2014

Accepted 5 August 2014

Available online 7 August 2014

Keywords:

Glioma grading

Diffusion tensor imaging

Voxel-based clustering

Self-organizing map

K-means

Support vector machine

ABSTRACT

Gliomas are the most common intra-axial primary brain tumour; therefore, predicting glioma grade would influence therapeutic strategies. Although several methods based on single or multiple parameters from diagnostic images exist, a definitive method for pre-operatively determining glioma grade remains unknown. We aimed to develop an unsupervised method using multiple parameters from pre-operative diffusion tensor images for obtaining a clustered image that could enable visual grading of gliomas. Fourteen patients with low-grade gliomas and 19 with high-grade gliomas underwent diffusion tensor imaging and three-dimensional T1-weighted magnetic resonance imaging before tumour resection. Seven features including diffusion-weighted imaging, fractional anisotropy, first eigenvalue, second eigenvalue, third eigenvalue, mean diffusivity and raw T2 signal with no diffusion weighting, were extracted as multiple parameters from diffusion tensor imaging. We developed a two-level clustering approach for a self-organizing map followed by the K-means algorithm to enable unsupervised clustering of a large number of input vectors with the seven features for the whole brain. The vectors were grouped by the self-organizing map as protoclusters, which were classified into the smaller number of clusters by K-means to make a voxel-based diffusion tensor-based clustered image. Furthermore, we also determined if the diffusion tensor-based clustered image was really helpful for predicting pre-operative glioma grade in a supervised manner. The ratio of each class in the diffusion tensor-based clustered images was calculated from the regions of interest manually traced on the diffusion tensor imaging space, and the common logarithmic ratio scales were calculated. We then applied support vector machine as a classifier for distinguishing between low- and high-grade gliomas. Consequently, the sensitivity, specificity, accuracy and area under the curve of receiver operating characteristic curves from the 16-class diffusion tensor-based clustered images that showed the best performance for differentiating high- and low-grade gliomas were 0.848, 0.745, 0.804 and 0.912, respectively. Furthermore, the log-ratio value of each class of the 16-class diffusion tensor-based clustered images was compared between low- and high-grade gliomas, and the log-ratio values of classes 14, 15 and 16 in the high-grade gliomas were significantly higher than those in the low-grade gliomas ($p < 0.005$, $p < 0.001$ and $p < 0.001$, respectively). These classes comprised different patterns of the seven diffusion tensor imaging-based parameters. The results suggest that the multiple diffusion tensor imaging-based parameters from the voxel-based diffusion tensor-based clustered images can help differentiate between low- and high-grade gliomas.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Abbreviations: ADC, apparent diffusion coefficient; AUC, area under the curve; BET, FSL's Brain extraction Tool; BLSOM, batch-learning self-organizing map; CI, confidence interval; CNS, central nervous system; DTcl, diffusion tensor-based clustered image; DTI, diffusion tensor imaging; DWI, diffusion-weighted imaging; EPI, echo planar image; FA, fractional anisotropy; FDT, FMRIB's diffusion toolbox; FLAIR, fluid-attenuated inversion-recovery; FSL, FMRIB Software Library; HGG, high-grade glioma; KM, K-means; KM++, K-means++; L1, first eigenvalue; L2, second eigenvalue; L3, third eigenvalue; LGG, low-grade glioma; LOOCV, leave-one-out cross-validation; MD, mean diffusivity; MP-RAGE, magnetization-prepared rapid gradient-echo; MRI, magnetic resonance imaging; PET, positron emission tomography; ROC, receiver operating characteristic; ROI, region of interest; S0, raw T2 signal with no diffusion weighting; SOM, self-organizing map; SVM, support vector machine; T1WI, T1-weighted image; T1Wice, contrast-enhanced T1-weighted image; T2WI, T2-weighted image; WHO, World Health Organization.

* Corresponding author at: Human Brain Research Center, Kyoto University Graduate School of Medicine, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan.

E-mail address: noishi@kuhp.kyoto-u.ac.jp (N. Oishi).

1. Introduction

Gliomas are the most common primary neoplasms of the central nervous system (CNS), and are classified according to a grading system, commonly that of the World Health Organization (WHO), on the basis of their histological appearance. Tumour grading is an important factor that influences the choice of therapy, such as adjuvant radiation and chemotherapy (Louis et al., 2007b).

Patients with low-grade gliomas (LGGs) (WHO grade II) may live for a long time, and the 5-year survival rate is 42%–92% (Sanai and Berger, 2008). In contrast, patients with high-grade gliomas (HGGs) (WHO grades III and IV) have a worst prognosis of brain tumours (Law et al., 2006); particularly, glioblastoma (WHO grade IV) develops rapidly (Ohgaki and Kleihues, 2007), and the 5-year survival rate is only 2% (McLendon and Halperin, 2003). Therefore, patients with HGGs need to be treated as soon as possible and more aggressively with chemotherapy and radiation. Thus, it is important to accurately classify gliomas into low or high grades to provide the best treatment for patients.

Magnetic resonance imaging (MRI) is essential for non-invasively diagnosing the existence, extent and characteristics of brain tumours. Different MRI sequences are used for evaluation and include T1-weighted image (T1WI), contrast-enhanced T1-weighted image (T1WIce), T2-weighted image (T2WI), diffusion-weighted imaging (DWI) and fluid-attenuated inversion-recovery (FLAIR) sequences. The images can provide much information about tumours, such as tumour morphology, the presence of enhancement, intra-tumoural haemorrhage or peri-tumoural oedema and can be helpful to predict tumour grade. The presence of contrast enhancement is often regarded as a sign of malignancy. Watanabe et al. reported that enhancement was present in 11 of 12 HGGs in their study, and histological examination revealed that areas of enhancement were related to neovascularity in tumour tissue or tumour cell infiltration (Watanabe et al., 1992). However, it was also reported that 9% of malignant gliomas lacked enhancement and 48% of LGGs were enhanced (Scott et al., 2002). These studies suggested that T1WIce was less useful than expected for prediction of glioma grade. Furthermore, gadolinium-based contrast agents, which are typically used in MRI, can cause side effects. Acute reactions after injection of gadolinium may cause flushing and nausea as minor reactions and hypotension and bronchospasm as intermediate reactions. In addition to these side effects, severe reactions are all symptoms of minor and intermediate reactions and sometimes cause cardiac arrest (Thomsen, 2003). Thus, T1WIce cannot be used for definitive pre-operative glioma grading because of insufficient information or side effects.

Some previous studies have used other MRI sequences without contrast agents, including diffusion tensor imaging (DTI), for glioma grading. Diffusion is sensitive to water movement, particularly along axonal fibres. DTI provides useful information about diffusion measurements and enables calculation of several parameters from DTI. Because tumour cells of gliomas mainly invade along white matter tracts (Scherer, 1938), we believe that DTI is a potentially useful sequence because of its sensitivity to white matter abnormalities (Filippi et al., 2001). Fractional anisotropy (FA) and apparent diffusion coefficient (ADC) calculated from DTI are more sensitive indicators of the integrity of white matter and tumour infiltration than are T1WI or T2WI (Price et al., 2003). Thus, DTI parameters can have an important role in the assessment of tumours. It has been reported that compared with white matter, HGGs show a mixture of hyper- and iso-intensities in DWI (Tien et al., 1994; Stadnik et al., 2001). One study found that lower ADC values corresponded to increased cellularity and HGGs (Kao et al., 2013). However, another study found no significant difference in ADC values between LGGs and HGGs (Lam et al., 2002). The FA values of LGGs were significantly lower than those of HGGs (Inoue et al., 2005; Kao et al., 2013), whereas another study showed low FA ratios in the tumour centres of both LGGs and HGGs (Goebell et al., 2006). These previous studies suggest that glioma grading with a single parameter of MRI remains controversial.

Recently, a pattern recognition method using multiple parameters has been applied to predict tumour grading. In a study, support vector machine (SVM), which is a widely used supervised machine-learning method because of its remarkable performance of classification, was applied and involved 161 features extracted from manually defined regions of interest (ROIs) on T1WI, T1WIce, T2WI, FLAIR and perfusion MRI using a contrast agent, and a combination of multiple features that differentiated HGGs and LGGs with an accuracy of 87.8%, sensitivity of 84.6% and specificity of 95.5% was reported (Zacharaki et al., 2009). Another study used a self-organizing map (SOM) based on a competitive learning algorithm, which is a type of neural network unsupervised learning, with seven features extracted from wavelet-filtered ADC, ADC, FLAIR and T2WI for each voxel (Vijayakumar et al., 2007). SOM was labelled for seven tissue classes, including low- and high-grade tumours, in a supervised manner using 700 voxel-based training pattern vectors. Although the sample size was small (four patients with low- and six with high-grade tumours), the method differentiated low-grade tumour from other tissues, with a sensitivity of 88% and a specificity of 98%, and high-grade tumour from other tissues, with a sensitivity of 87% and a specificity of 93%. Although pattern recognition methods with multiple parameters and a supervised manner can be useful for prediction of tumour grading, they have some problems in clinical applications. In voxel-based labelling, because it is impossible to examine the pathology of each voxel, supervised voxel-based labelling can be inaccurate and cause rater bias. Furthermore, complicated features make it difficult to determine the most sensitive parameter for characterizing grading. Therefore, a pattern recognition method with multiple uncomplicated parameters without supervised information can be helpful to predict glioma grade. Furthermore, SOM is well-known to its visualization and would help to lead to a novel classification.

This study aimed to develop a new method using multiple DTI-based parameters for voxel-based clustered images in an unsupervised manner that can be used to visually grade gliomas. We also determined if the method is really helpful for pre-operative prediction of glioma grade in a supervised manner.

2. Materials and methods

2.1. Subjects

We retrospectively reviewed 111 patients who were aged 6–87 years and had newly diagnosed and histologically confirmed diffusely infiltrative gliomas, defined according to the WHO classification (Louis et al., 2007a), between March 2010 and June 2013 in Kyoto University Hospital. We classified grade II as LGG ($n = 36$) and grades III and IV gliomas as HGG ($n = 75$) in this study. Patients with LGGs had 22 diffuse astrocytomas, eight oligodendrogliomas, four oligoastrocytomas and two mixed oligoastrocytoma. Patients with HGGs had 17 anaplastic astrocytomas, three anaplastic oligoastrocytomas, two anaplastic oligodendrogliomas and 53 glioblastomas. Among these patients, 51 underwent DTI and magnetization-prepared rapid gradient-echo (MP-RAGE). We excluded 13 patients who had undergone previous tumour resections or exposures to radiotherapy or chemotherapy prior to DTI acquisition. We also excluded three patients whose tumours were located around the temporal basal regions that were severely influenced by distortions of DTI (Mangin et al., 2002) and one patient because of appreciable motion artefacts in MP-RAGE. We excluded one patient <12 years of age because FA values in the frontal white matter are significantly lower in children aged 8–12 years than in adults because of less myelination in children (Klingberg et al., 1999). Consequently, 33 patients (22 men, 11 women) were enrolled in the study. Thirty-two patients had undergone tumour resections, and one had only undergone a biopsy. Twenty-one tumours were located in the frontal region, seven in temporal, two in parietal, one in occipital and two in frontoparietal (Table 1). This study was approved by the Ethics Committee of the Kyoto University Graduate School

Table 1
Summary of patient data.

Histopathology	n	WHO grade	Age (years)	Location
High-grade gliomas	19	III and IV	51.7 ± 18.6	
Glioblastoma	13	IV	54.1 ± 18.0	Frontal, parietal, temporal, frontoparietal
Anaplastic astrocytoma	4	III	33.5 ± 4.8	Frontal, temporal
Anaplastic oligoastrocytoma	2	III	72.0 ± 5.0	Frontal
Low-grade gliomas	14	II	42.7 ± 13.4	
Diffuse astrocytoma	6	II	47.5 ± 15.8	Frontal, parietal, frontoparietal
Oligoastrocytoma	3	II	40.3 ± 15.5	Frontal, temporal
Oligodendroglioma	4	II	38.0 ± 3.7	Frontal, temporal
Mixed oligoastrocytoma	1	II	40.0	Frontoparietal

Age (years) is given as means ± standard deviation.

of Medicine (C 570), and written informed consent was obtained from all patients.

2.2. MRI data acquisition and pre-processing

MRI images were scanned on a 3 Tesla Trio (Siemens, Erlangen, Germany) equipped with an eight-channel phased-array head coil.

DWI in an axial orientation used the following parameters: repetition time = 10,500 ms; echo time = 96 ms; flip angle = 90°; field of view = 192 × 192 mm; slices = 70; and voxel size = 2 × 2 × 2 mm. Diffusion weighting was isotropically distributed along 81 directions by using a b-value of 1500 s/mm² (Jones et al., 1999). Nine volumes with no diffusion weighting (b = 0 s/mm²) were also acquired at points throughout acquisition. A set of diffusion-weighted data was acquired during a scanning time of approximately 18 min.

MP-RAGE using the following parameters was used to acquire three-dimensional T1-weighted anatomical images: repetition time = 2000 ms; echo time = 4.38 ms; flip angle = 8°; field of view = 176 × 192 mm; slices = 160; and voxel size = 1 × 1 × 1 mm. A dual-gradient field map in an axial orientation was also obtained by using the following parameters: repetition time = 511 ms; echo time 1/echo time 2 = 5.19/7.65 ms; flip angle = 60°; field of view = 192 × 192 mm; slices = 46; and voxel size = 3 × 3 × 3 mm.

DTI data were analysed using FSL [FMRIB Software Library v5.0.2.2, <http://www.fmrib.ox.ac.uk/fsl> (Smith et al., 2004)]. The data were corrected for eddy currents and head motion using affine registration to the first b = 0 reference volume. The data were also corrected for geometric distortions occurring in an echo planar image (EPI) (Jezzard and Balaban, 1995) by FUGUE, which is a part of the FSL tool for EPI correction of distortions caused by static magnetic field inhomogeneities (Jenkinson et al., 2012). Field inhomogeneities were measured by using a field-map image, and EPI was unwarped according to field-map data. Seven features, including DWI, FA, first eigenvalue (L1), second eigenvalue (L2), third eigenvalue (L3), mean diffusivity (MD) and raw T2 signal with no diffusion weighting (S0), were extracted from DTIs using the FMRIB's diffusion toolbox (FDT) program (Smith et al., 2004). The diffusivities derived from DTI measurements were separated into components parallel (L1) and perpendicular (L2 and L3) to the white matter tract. These components are referred to as MD, (L1 + L2 + L3)/3. FA, which assigns values between 0 and 1,

represents how strongly water diffuses in the direction of the principal eigenvector (Holodny et al., 2001).

2.3. Feature extraction for clustering

The overview of the processing pipeline in the study is depicted in Fig. 1. The summary is as follows:

- 1 Feature extraction from DTI.
- 2 Clustering using SOM followed by K-means (KM).
- 3 Visualization of whole brain images by diffusion tensor-based clustering images (DTcls).
- 4 Classification using DTcls by SVM.

Features for unsupervised clustering were extracted from voxels on the seven intensity-normalized diffusion tensor images sampled at every 64 (4 × 4 × 4) voxels within the binary mask image obtained by FSL's Brain Extraction Tool (BET). The number of extracted features was 3552 ± 315 (mean ± SD) for each subject. The features of all subjects were then stacked and used for input vectors defined as $T = \{b_m, 1 \leq m \leq n\}$, where b_m is the m th vector and n is the number of vectors (117,232 in the study). The individual input vector (b_m) is defined as $b_m = [x_1, x_2, x_3, x_4, x_5, x_6, x_7]$, where $x_1, \dots, x_7$ are components of the input vector. The components of each input vector (b_m) were extracted from DWI, ADC, MD, S0, L1, L2 and L3 images in the study.

2.4. Unsupervised clustering

The extracted feature vectors were used for calculating voxel-based clustered images. We applied the two-level clustering approach of SOM (Kohonen, 1995) and the KM algorithm (MacQueen, 1967) for unsupervised clustering (Fig. 1). First, a large number of input vectors were clustered into a much larger than the expected number of clusters, defined as 'protoclusters', by SOM. Then, the protoclusters were classified into the expected number of clusters, defined as 'clusters', by KM. The KM algorithm is a popular partition algorithm for clustering. Similar two-level clustering approaches have been reported (Chuang et al., 1999; Vesanto and Alhoniemi, 2000; Thomassey and Happiette, 2007) and applied in the fields of medicine (Wang et al., 2002) and others (Beccali et al., 2004; Lu et al., 2006). The two-level clustering approach has the following two important benefits compared with the KM algorithm,

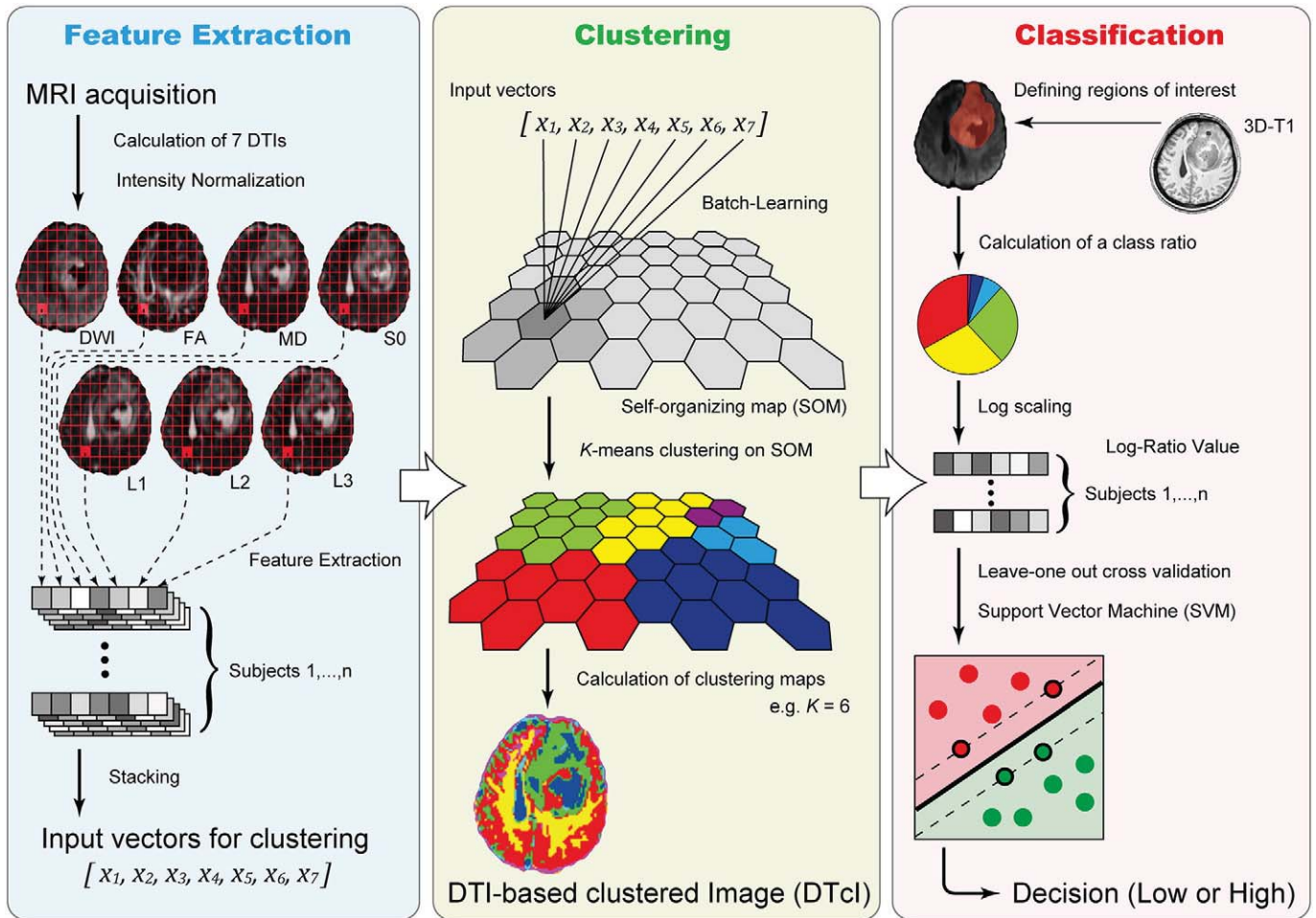


Fig. 1. Simplified graphical overview of the processing pipeline.

which is one of the most famous and effective unsupervised clustering approaches: One is noise reduction. It is known that the KM algorithm is very sensitive to outliers (Velmurugan and Santhanam, 2010). Because protoclusters are local averages of the input vectors and outliers can be filtered out, they are more robust to outliers than are the original vectors. The other is the reduction of the computational cost. Furthermore, the two-dimensional arrangement of the larger protoclusters by SOM can be easily visualized, thus, providing useful information about the features (Jin et al., 2004).

2.4.1. Unsupervised clustering: SOM

SOM is based on a competitive learning algorithm, which is a type of neural network unsupervised learning. It typically comprises hexagonal and two-dimensional grids of map units (also called neurons) defined as $H = \{w_i; 1 \leq i \leq K \times L\}$. Here, K and L are the numbers of columns and rows, respectively, w_i represents the weight vectors assigned to the i th unit of the SOM architecture, defined as $w_i = [v_1, \dots, v_d]$, where d is the dimension of the weight and $v_1, \dots, v_d$ are components of the weight vector. The dimension of the weight vector was seven in this study.

A major problem in the standard sequential SOM algorithm is that the results differ according to the input order. Therefore, the batch-learning algorithm for the SOM (batch-learning self-organizing map; BLSOM) (Kohonen, 1995) was used in the study. The results by BLSOM are consistent because its learning does not depend on the input order. Other advantages of BLSOM compared with the standard sequential SOM are that no learning rate has to be specified by the user of the

algorithm and convergence of the input vectors is faster towards their final values (Brugger et al., 2008). The following algorithm was used for BLSOM:

- 1 Initialize the weight vectors $w_i(0)$ of all map units.
- 2 Calculate the Voronoi sets $V_i(t) = \{b \mid d(b, w_i(t)) < d(b, w_k(t)) \forall_{k \neq i}\}$ and the sums $s_i(t) = \sum_{j=1}^{|V_i(t)|} b_j$.
- 3 Update the weight vectors according to $w_i(t+1) = \sum_j h_{ij}(t) s_j(t) / \sum_j |V_i(t)| h_{ij}(t)$.
- 4 Repeat steps 2 and 3 until a predefined number of steps τ .

Here, $w_i(t)$ is a weight vector in the t th step and $h_{ij}(t)$ is the neighbourhood function defined by $h_{ij}(t) = \exp(-\|r_i - r_j\|^2 / 2\sigma^2(t))$, where r_i and r_j are the i th and j th units of the SOM architecture, respectively, $\|r_i - r_j\|$ denotes the Euclidean distance between r_i and r_j , and $\sigma(t)$ is the neighbourhood radius defined by $\sigma(t) = \sigma_{final} + (\sigma_{initial} - \sigma_{final})(1 - t/\tau)$, where $\sigma_{initial}$ and σ_{final} are the neighbourhood radii in the initial and final steps, respectively. In the study, the initial weight vectors $w_i(0)$ in step 1 were defined on the basis of principal component analysis of the input vectors b_m by $w_i(0) = \bar{b} + 10s_1\{(k-K/2)/K\}a_1 + 10s_2\{(l-L/2)/L\}a_2$, where $\bar{b}$ is the average vector of b_m , a_1 and a_2 are the eigenvectors of the first and second principal components and s_1 and s_2 are the standard deviations of the first and second principal components. k and l are the k th column and l th row in the i th unit of the SOM architecture, respectively. The parameters for BLSOM used in the study were defined as $K = 20$, $L = 20$, $\sigma_{initial} = 3.0$, $\sigma_{final} = 2.0$ and $\tau = 100$ based on the previous studies. Vijayakumar et al. used SOM for segmentation of brain tumour on MRI with pre-defined

parameters of $K = 20$, $L = 20$ (Vijayakumar et al., 2007). Furthermore, Vesanto and Alhoniemi who reported two-level clustering suggested that the number of protocusters N was determined as $N = k_{\max}^2$, where $k_{\max}$ was an estimated maximum number of clusters (Vesanto and Alhoniemi, 2000). Because we predefined $k = 20$ from the viewpoint of a pathological estimation, the number of protocusters was 400 according to the formula. According to the previous report and formula, we think the 20×20 matrix may be acceptable. Chavez-Alvarez et al. reported that the neighbourhood function was Gaussian with an initial radius (σ_{initial}) of 3.0 when $K = 20$ and $L = 20$ in their study (Chavez-Alvarez et al., 2014). To determine a final radius (σ_{final}), we tested BLSOM with a set of parameters (0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 3) according to the previous study (Ehsani and Quiel, 2008). Lower final radius than 2.0 could cause stepping-stone like clusters by the $KM++$ with $k = 20$, which was the maximum number in the study, partly because of overfitting. Therefore, we finally applied $\sigma_{\text{final}} = 2.0$, which did not cause any stepping-stone like clusters. As to τ , the number of total iterations of SOM, some previous neuroimaging studies applied 100 iterations for SOM and suggested that 100 iterations were sufficient to establish convergence of SOM, and, therefore we followed them (Peltier et al., 2003; Liao et al., 2008). We have implemented BLSOM software (in-house developed products) developed in C++.

One of the greatest advantages of SOM is the powerful visualization (Vesanto and Alhoniemi, 2000), which enables SOM to be suited for data understanding or survey. There are two common methods to visualize the results of SOM, U-Matrix and Component Planes. Especially, the Component Planes can show the information of each parameter in each map unit and associations between clusters and variables (Alhoniemi et al., 1999). We used the Component Planes to visualize the relation between variables of each DTI-based parameter clearly. Using the Component Planes, SOM may allow us to evaluate the classification options different from a fixed diagnosis (Wang et al., 2002).

2.6. Unsupervised clustering: KM clustering

The KM algorithm is a classic statistical clustering method (MacQueen, 1967) computed in an off-line mode and does not perform competitive learning as does SOM. Its objective, for K clusters, is to iteratively minimize the within-class inertia by assigning the feature vectors to the nearest cluster centre and update its value. The number of clusters has to be determined prior to calculation.

The algorithm of the standard KM is as follows:

- 1 Start by initializing C that will contain K cluster centres $c_k \in \mathbb{R}^d$ such that $1 \leq K \leq N$. The set $C = \{c_1, c_2, \dots, c_K\}$ will be initialized with the vectors x_j randomly chosen from the data set $X \subset \mathbb{R}^d$.
- 2 Assign each vector x_j from the data set to the nearest centre c_k using the Euclidean distance metric, $d(x_j, c_k) = \|x_j - c_k\|$.
- 3 Update the new cluster centres c_k with the average value of its members by $c_k = (1/C_k) \sum_{x_j \in C_k} x_j$, where C_k represents the number of elements in the respective cluster.
- 4 Repeat steps 2 and 3 if any partition was modified since last iteration.

Although the procedure will always terminate, the standard KM might converge to a local minimum because it uses vectors chosen at random from the data set to initialize the clusters. Therefore, we used a modified version of KM, called the K-means++ ($KM++$) algorithm (Arthur and Vassilvitskii, 2007), that chooses centres at random from the data points but weighs the data points according to their squared distance from the closest centre already chosen. The $KM++$ algorithm shows drastic improvement in both speed and accuracy compared with the classic algorithm. The $KM++$ algorithm is defined as follows:

- 1a Take one centre c_1 , chosen uniformly at random from X .
- 1b Take a new centre c_i , choosing $x \in X$ with probability $D(x)^2 / \sum_{x \in X} D(x)^2$

1c Repeat step 1b until we have taken K centres altogether.

2–4. Proceed as with the standard KM algorithm.

In addition to $KM++$, we repeated $KM++$ trials and selected the best cluster among different clusters from multiple $KM++$ trials by the silhouette index (Rousseeuw, 1987). It calculates the silhouette index for each datum, average silhouette index for each cluster and overall average silhouette index for the total data set. Using the method, each cluster could be represented by so-called silhouette, which is based on the comparison of its tightness and separation. The average silhouette index can be applied for evaluation of clustering validity. The overall average silhouette index (SI) is defined by $SI = (1/K) \sum_{j=1}^K S_j$, where S_j is a silhouette local coefficient defined by $S_j = (1/n_j) \sum_{i=1}^{n_j} s_i$, where s_i is a silhouette index for the i -th object defined by $s_i = (b_i - a_i) / \max(b_i, a_i)$, where a_i is the mean distance between object i and objects of the same class j and b_i is the minimum mean distance between object i and objects in class closest to class j . In the study, 1000 $KM++$ trials were performed in each K . We have little prior knowledge about the number of K , and it can differ according to what users want to know (e.g. tumour detection, tumour grading or outcome prediction) using the clustered images. In the present study, we empirically chose the numbers of $K = 4, 6, 8, 10, 12, 16, 20$ according to the number of estimated segmentations in the abnormal brain with a glioma. We implemented these algorithms in the in-house SOM software, which enabled the use of BLSOM followed by $KM++$.

2.7. Diffusion tensor-based clustered image (DTCl)

After unsupervised clustering by BLSOM followed by $KM++$, 400 protocusters (weighted vectors) with K -class label information were generated. Label information of the nearest protocuster using the Euclidean distance metric was assigned to each voxel on the seven intensity-normalized diffusion tensor images within the binary mask image from FSL's BET. Finally, voxel-based images with K -class label information were obtained. We called them DTcls (Fig. 1). Using the unsupervised clustering method, DTcls can be easily obtained without any initial segmentation.

2.8. Classification using DTCl: definition of ROI

Gliomas generally show unclear and irregular boundaries with discontinuities and variety. ROIs were manually traced by two of the authors in the DTI space according to abnormalities on MP-RAGE, without any knowledge of the clinical or pathological data. It should be noted that ROIs were defined on the basis of abnormal signal intensities in MP-RAGE, which could include tumour as well as oedema, necrosis and cystic parts. For simplicity, we defined ROIs per four axial slices from the top slice of the abnormal regions and counted the total number of voxels in each patient. If the number of voxels was <400, we redefined ROIs per two axial slices to increase the number of total voxels. Finally, the number of voxels in each ROI ranged from 326 to 9316. We only used these ROIs for feature extraction for SVM as mentioned in the next subsection.

2.9. Classification using DTCl: feature extraction for tumour grading

The ratio of each class in DTcls was calculated from ROIs in each subject. Then, the common logarithmic value of the ratio was calculated by $\log_{10}(p + 10^{-2})$, where p is a ratio of each class (%) and used for input features to SVM. The features were defined as log-ratio values according to $U = \{x_i \in \mathbb{R}^K, 1 \leq i \leq N\}$, where N is the number of subjects (Fig. 1).

2.9.1. Classification using DTCl: SVM

SVM (Vapnik, 1998) is a widely used method because of its remarkable classification performance and the simplicity of its theory and

implementation. Accordingly, we chose a linear kernel SVM as a classifier to distinguish between LGGs and HGGs. The hyperparameter (C) of the linear kernel SVM was optimized by using a two-step grid-search technique with five-fold cross-validation according to the recommendation described in a practical guide to SVM classification (Hsu et al., 2003; Ota et al., 2014). First, the best value of C_{coarse} was found by a coarse grid-search on $\log_2 C = -5, -3, \dots, 15$. Then, the best value of C_{fine} was obtained by a fine grid-search on $\log_2 C = C_{coarse} - 2, C_{coarse} - 1.75, \dots, C_{coarse} + 1.75, C_{coarse} + 2$. The best C_{fine} was used to generate the final classifier for each training set.

A LOOCV strategy was used to assess the classification performance because the strategy is widely used in machine learning and allows using most of the data for training (Dosenbach et al., 2010). During LOOCV, each subject is designated as a test subject in turn, while the remaining subjects are used to train the SVM classifier. The decision function derived from the training subjects is then used to classify or calculate a decision value about the test subject. After all LOOCV repeats, the accuracy, sensitivity and specificity for all folds are averaged together to generate the final accuracy, sensitivity and specificity estimate. We also evaluated decision values (Chang and Lin, 2011) for receiver operating characteristic (ROC) curves and area under the curve (AUC). Furthermore, we repeated the LOOCV strategy 100 times to calculate confidence intervals (CIs) of these estimates.

We used C++ and the LIBSVM library (Chang and Lin, 2011; software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>) to implement a linear kernel SVM with a two-step grid-search technique and a LOOCV strategy.

2.10. Statistical analysis

To determine if the classification performances were significantly different according to the number of K in the KM++ method ($K = 4, 6, 8, 10, 12, 16, 20$), we repeated the LOOCV strategy 100 times. AUCs in different K were then analysed by one-way ANOVA followed by Tukey's multiple comparison tests. Differences were considered significant for $p < 0.05$.

Then, to evaluate the behaviour of the classifier in the K -class that revealed the best classification performance, we used the pROC library for R to generate ROC curves with 95% CIs computed with 2000 stratified bootstrap replicates (Robin et al., 2011).

Wilcoxon–Mann–Whitney tests with exact p -values and CIs calculated by a permutation test were used to compare the log-ratio values of each class in the K class that revealed the best classification performance between the LGG and HGG groups (Hothorn et al., 2006). Because of the multiple comparisons in K classes, a Bonferroni correction for multiple comparisons was applied, and differences between the groups were considered to be significant at a level of $p < 0.05/K$.

The ratios of normalized intensities on the seven diffusion tensor images of each class in the K class that revealed the best classification performance were analysed with the bootstrapped 95% CIs. The statistical software package R version 3.0.2 (The R Foundation for Statistical Computing, <http://www.r-project.org/>) was used to perform all statistical analyses.

3. Results

3.1. Unsupervised clustering

Fig. 2 illustrates the Component Planes in seven DTI-based variables by the SOM analysis. Visual inspection of the SOM patterns demonstrated that the Component Planes of DWI and FA were obviously different from the others. Although general patterns of MD, S0, L1, L2 and L3 Component Planes seemed similar, the details among them differed. In the case of $K = 16$ (Fig. 2), for example, FA values in class number 13, DWI values in class number 15 and MD, S0, L1, L2 and L3 values in class number 4 were highest among all classes. SOM also showed that the DWI component of class numbers 1, 2, 9, 12 and 13 had variations from low to high values, and that the FA component of class numbers 9, 10, 12 and 14 had variations from low to high values. Class numbers 12, 13 and 16 in the L1 Component Plane were higher than those of the MD, S0, L2 and L3 Component Planes. Class numbers 2, 6, 7 and 8 in the MD, L1, L2 and L3 Component Planes were higher than

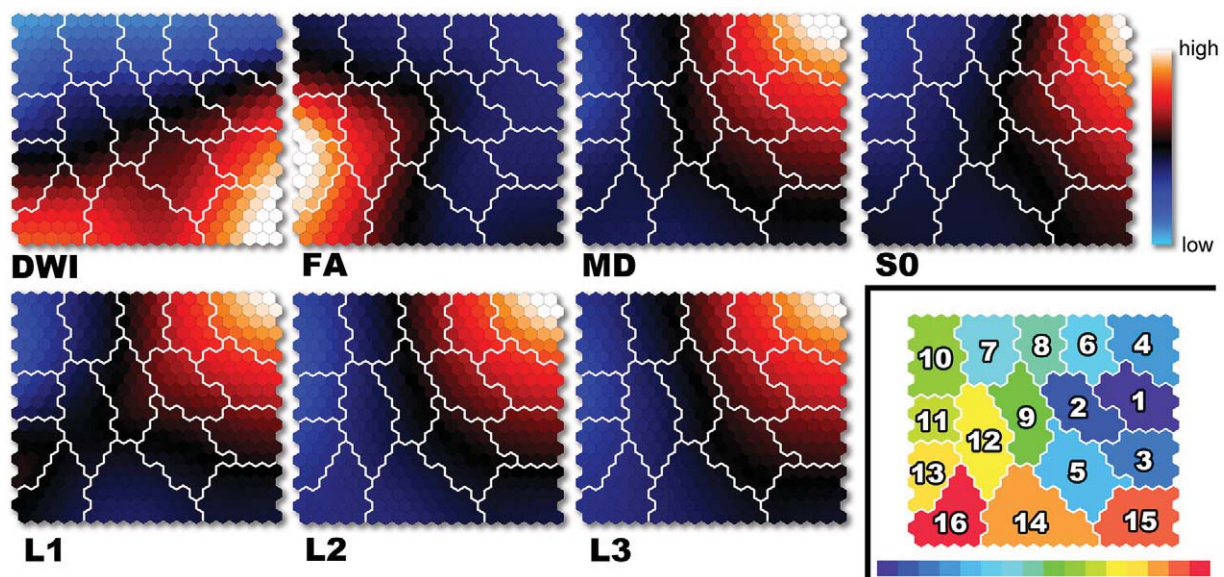


Fig. 2. Visualization of seven DTI-based variables on Component Planes with SOM. Each node (protocluster) is colored from blue to red according to the intensities in each diffusion tensor image. The white lines between nodes denote inter-class borderlines obtained by KM++ with $K = 16$ on SOM. SOM component planes can help to interpret detailed intensity profiles or patterns in each diffusion tensor image (lower right). The 16-class cluster map on the 20×20 SOM. Each class number corresponds to intensity on DTI-based clustered images. DWI = diffusion-weighted imaging; FA = fractional anisotropy; L1 = first eigenvalue; L2 = second eigenvalue; L3 = third eigenvalue; MD = mean diffusivity; S0 = raw T2 signal without diffusion weighting.

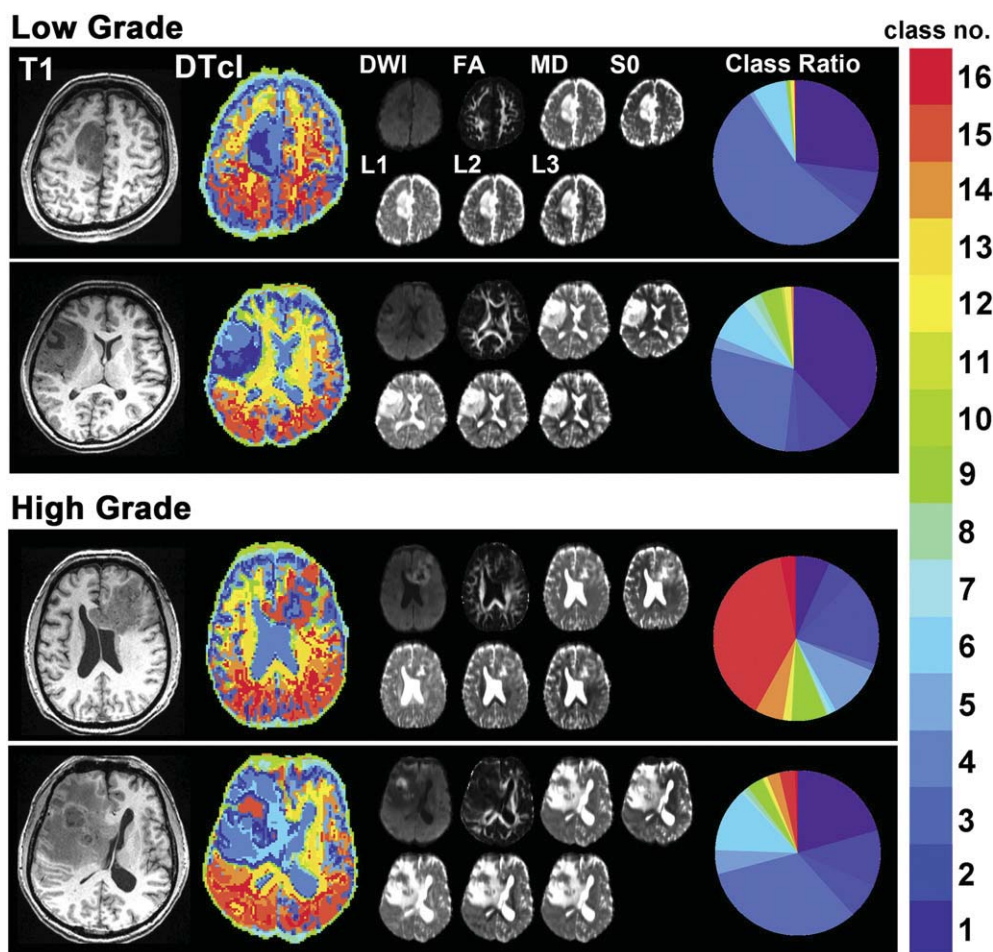


Fig. 3. The representative cases of low- (upper) and high- (lower) grade gliomas, including the 16-class DTcIs that showed the highest classification performance. The T1-weighted images, DTcIs, seven diffusion tensor images and the ratios in each class number are shown for each patient. DWI = diffusion-weighted imaging; FA = fractional anisotropy; L1 = first eigenvalue; L2 = second eigenvalue; L3 = third eigenvalue; MD = mean diffusivity; S0 = raw T2 signal without diffusion weighting. Each colour on DTcIs and circular charts correspond to each class number, shown in the colour bar.

those in the S0 Component Plane. Class number 9 in the MD and L1 Component Planes was higher than that in the S0, L2 and L3 Component Planes.

Representative cases of LGGs and HGGs are shown in Fig. 3. Although the boundaries of LGGs could be clearly recognized, it was much difficult to recognize the boundaries of HGGs. Furthermore, DTcIs revealed few warm coloured classes, such as class numbers 14, 15 and 16 in LGGs, whereas there were more warm coloured classes in HGGs than in LGGs. Thus, the clear differentiations between LGGs and HGGs on DTcIs could be visually recognized.

3.2. SVM classification using DTcI

The performances of LOOCV using DTcI and SVM are shown in Fig. 4. The differences in AUCs were significant among the classes [$F(6, 693) = 246.1, p < 10^{-168}, \eta_p^2 = 0.68$]. Tukey's post-hoc tests showed that AUC was significantly higher for the 16-class DTcIs than for others ($p < 0.001$). The tests also showed that AUCs were significantly higher for the 12- and 20-class DTcIs than for the 4-, 6-, 8- and 10-class DTcIs ($p < 0.001$). There were no significant differences in AUCs between the 12-class and 20-class DTcIs. The tests also showed that AUC was significantly lower for the 4-class DTcIs than for the others ($p < 0.001$). AUC of the 16-class DTcIs was the highest among classes (0.912; 95% CI = 0.903–0.922) (Fig. 5). The sensitivity, specificity and accuracy of the 16-class DTcIs were 0.848 (95% CI = 0.845–0.852), 0.745 (95% CI = 0.733–0.757) and 0.804 (95% CI = 0.800–0.809), respectively. In contrast, AUC of the 4-class DTcIs was the lowest (0.729; 95% CI = 0.718–0.739). There were no significant group differences in AUCs between in the 6-, 8- and 10-class DTcIs (0.856, 0.844 and 0.850, respectively).

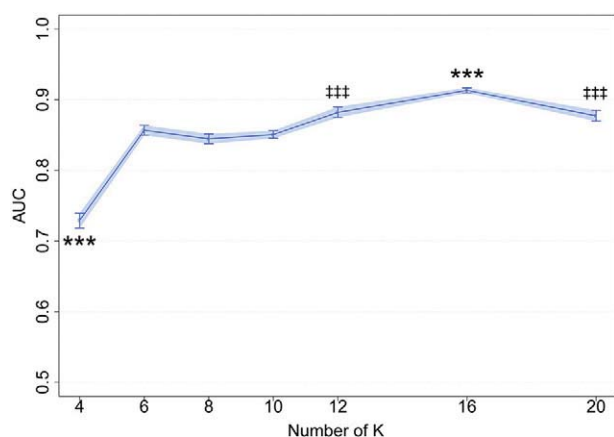


Fig. 4. Plots of AUC versus the number of K in the KM++ method. Values are means and error bars, and light blue shades represent 95% CIs. *** $p < 0.001$ (versus all the rest). ### $p < 0.001$ (versus $K = 4, 6, 8, 10, 16$), one-way ANOVA followed by Tukey's multiple comparison tests. The 16-class diffusion tensor-based clustered images significantly showed the highest AUC (0.912; 95% CIs = 0.903–0.922).

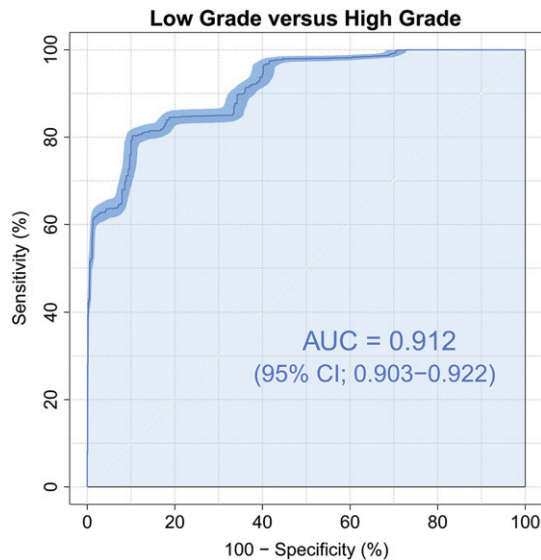


Fig. 5. ROC curves (dark blue line), with AUC and 95% CIs shown in blue shades surrounding the dark blue line, for differentiating high-grade from low-grade gliomas by using the 16-class diffusion tensor-based clustered images.

3.3. Differences in log-ratio values

The log-ratio values of each class of the 16-class DTcls that had the highest classification performance were compared between LGGs and HGGs (Fig. 6). The values of class numbers 14, 15 and 16 were significantly higher in HGGs than in LGGs ($p < 0.005$, $r = 0.52$; $p < 0.001$, $r = 0.60$; $p < 0.001$, $r = 0.73$; respectively). The values of class numbers 12 and 13 also revealed higher trends in HGGs ($p < 0.01$, $r = 0.48$; $p < 0.01$, $r = 0.50$; respectively).

3.4. Ratio of DTI-based parameters

The ratios of normalized intensities of the seven diffusion tensor images for each class number in the 16-class DTcls that revealed the highest classification performance are shown in Fig. 7.

As mentioned above, the ratios of class numbers 14, 15 and 16 were significantly higher in HGGs than in LGGs. The chart patterns of class numbers 14 and 15 seemed similar and comprised high DWI values and low FA values. Class number 15 had the highest DWI values among all. In FA, class number 14 had higher values than class number 15. The variables of class number 16 comprised high FA and DWI values and were different from those of class numbers 14 and 15. All three classes included low values in MD, S0, L1, L2 and L3. Although the variables of class numbers 3 and 5 included high DWI values and low values for

other features, there were no significant differences in the log-ratio values between LGGs and HGGs ($p = 0.28$, $r = -0.19$; $p = 0.84$, $r = 0.03$; respectively).

The indices of class numbers 12 and 13 also revealed higher trends in HGGs. The chart patterns of class numbers 12 and 13 were very different from those of class numbers 14, 15 and 16. The highest FA values were seen in class number 13. The variables of class numbers 12 and 13 included low DWI values. Low values in MD, S0, L1, L2 and L3 were seen in class numbers 12, 13, 14, 15 and 16.

4. Discussion

4.1. Study overview

In this study, we investigated a two-step procedure for predicting glioma grade. In the first step, the unsupervised clustering method with SOM followed by KM++ was used to obtain voxel-based DTcls with multiple DTI-based parameters. DTcls enabled visual grading of gliomas. In the second step, the validity of DTcls for glioma grading was assessed in a supervised manner using SVM. The 16-class DTcls revealed the highest classification performance for predicting the glioma grade. The sensitivity, specificity, accuracy and AUC of the 16-class DTcls for differentiating HGGs and LGGs were 0.848, 0.745, 0.804 and 0.912, respectively. The classifier in the 16-class DTcls showed that the ratios of class numbers 14, 15 and 16 were significantly higher and those of class numbers 12 and 13 showed higher trends in HGGs than in LGGs. Thus, these results indicate that our clustering method of seven parameters can be useful for determining glioma grade visually, despite not using a complicated combination of a high number of features from many modalities.

4.1.1. Clustering method

The two-level clustering approach was used in our study since it has the following two important benefits: noise reduction and computational cost. Because of the character of KM++ mentioned in the Materials and methods section, outliers extracted from DTI parameters can make its clustering accuracy worse. When BLSOM is applied prior to KM++, outliers can be filtered out and the clustering accuracy will be better. The AUC only with the KM++ algorithm without BLSOM was 0.646 with $K = 16$ and remarkably worse than that with the two-level clustering approach. Another important benefit is the reduction of the computational cost. In our study, the KM++ was repeated 1000 times to obtain more stable results. The computational time of the two-level clustering approach for 1000 KM++ trials was 210 s (204 s for BLSOM and 6 s for 1000 KM++ trials) for 117,232 input vectors in the study. On the other hand, the computational time only for the 1 KM++ trial without BLSOM was 305 s and around 85 hours for 1000 KM++ trials.

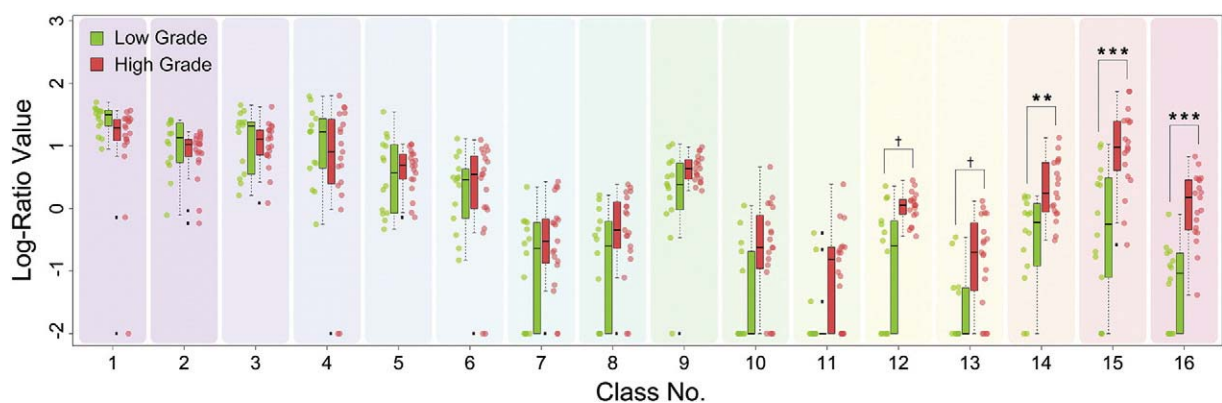


Fig. 6. Strip chart and box plots showing median, interquartile range, inner fence and outliers (○) for log-ratio values of each class by 16-class diffusion tensor-based clustered images in patients with low- (green) and high- (red) grade gliomas. *** $p < 0.001$, ** $p < 0.005$, + $p < 0.01$ by exact Wilcoxon–Mann–Whitney rank sum tests.

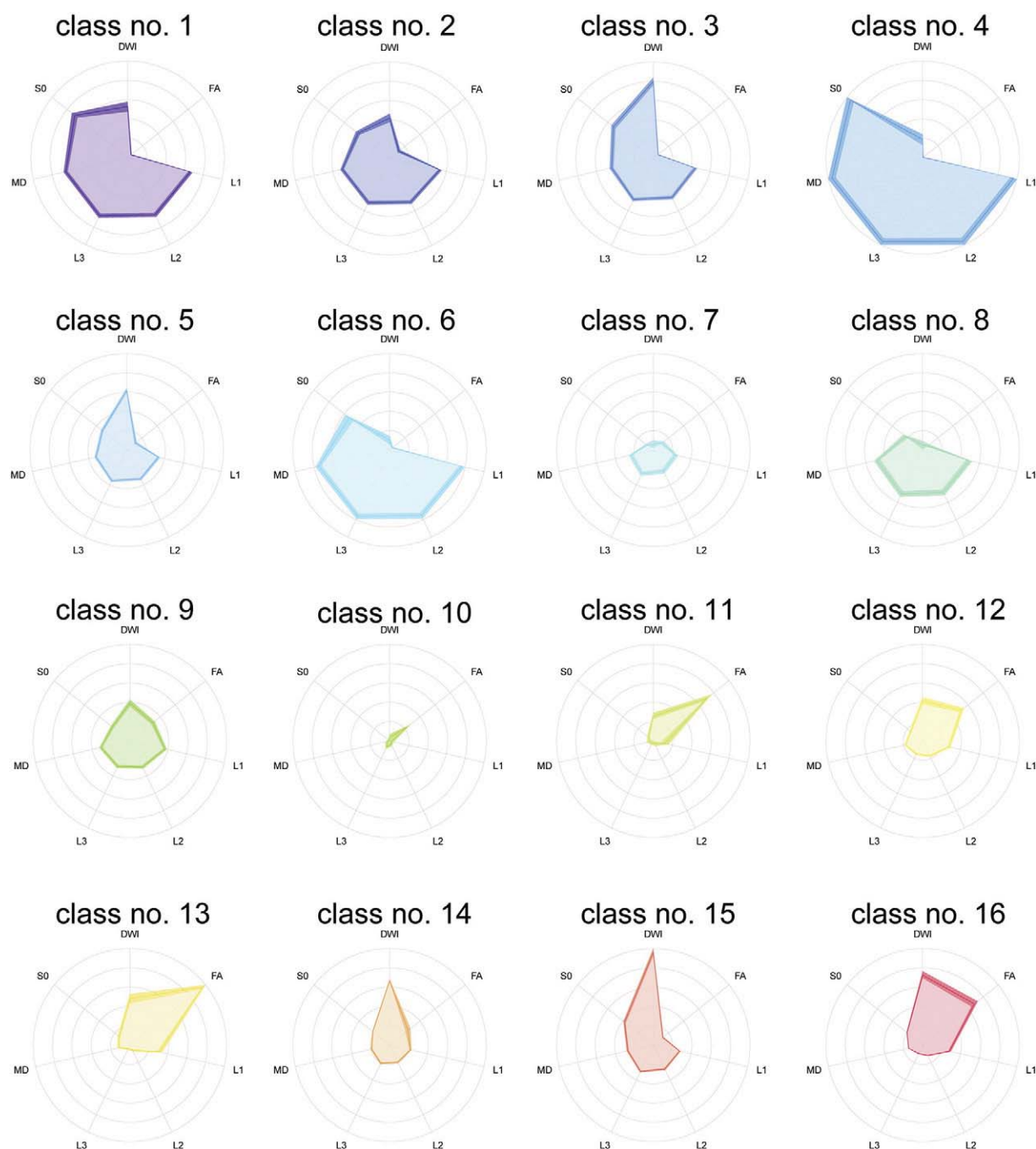


Fig. 7. Radar charts of seven DTI-based variables in each class by 16-class diffusion tensor-based clustered images. Shades surrounding dark-coloured lines represent bootstrapped 95% CIs. DWI = diffusion-weighted imaging; FA = fractional anisotropy; L1 = first eigenvalue; L2 = second eigenvalue; L3 = third eigenvalue; MD = mean diffusivity; S0 = raw T2 signal without diffusion weighting.

For these facts, the two-level approach can be effective especially for clustering of a larger data set.

Deciding optimal parameters for SOM is not easy as previous studies mentioned in their studies. Although we applied most of the parameters according to previous studies (Vesanto and Alhoniemi, 2000; Vijayakumar et al., 2007; Ehsani and Quiel, 2008; Chavez-Alvarez et al., 2014) in the study, it remains unclear, as mentioned in the Materials and methods section, whether these parameters for SOM lead to the best performance or not. The parameters might be verified by undertaking a prospective, randomized controlled study.

Our segmentation method does not need any initial segmentation for defining tumour lesions because features were extracted from the whole brain. Indeed, the DTIs can segment the brain as some parts of

normal and abnormal areas unintentionally, but the method does not need any initial segmentation for defining tumour lesions and it is an important advantage of unsupervised clustering methods. When defining tumour lesions as an initial segmentation, it is needed to draw regions of interest intentionally or decide automatically which voxel is tumour, oedema, necrosis or normal tissue with a supervised clustering method. However, it is impossible to decide the correct boundary between normal and abnormal pathology on MRI. The voxel out of the boundary may include tumour cells considered from the infiltrative nature of glioma, which may influence grading of gliomas. We think that clustering for images of gliomas without an initial segmentation is an indispensable advantage and our method can satisfy this point.

4.2. Number of classes in DTcls

The 16-class DTcls had the best classification performance between HGGs and LGGs in this study. It is assumed that brain tumour images can be segmented at least into four classes (white matter, grey matter, CSF and abnormality) (Rajini and Bhavani, 2012). Within abnormalities, they can be consisted of tumour cells (high/low), gliosis, oedema, necrosis, haemorrhage, and the mixed structure of some of them. Therefore, when we consider the combination of those, several kinds of classes may be reasonable. Furthermore, we found the same cluster in grey matter and in tumours. Class numbers 14–16 that had significantly higher HGG values were seen in grey matter and showed low MD values, which corresponded to increased cellularity (Lam et al., 2002; Kao et al., 2013). This finding may indicate high cellularity within tumour areas. However, it is difficult to say on the basis of our results which class would fit to what tissue. Pathological studies of each class in DTcls by biopsy or resection could clarify the relationship.

4.3. Multiple parameters in DTI

We selected L1, L2 and L3, that are the basis of DTI-related parameters, in order to compare previous pathological studies using those parameters. We also selected DWI, FA, ADC and S0 that are familiar to neurosurgeons in order to compare previous studies that used one of those parameters as a single parameter for glioma grading. In the previous study, Griffith et al. researched the pathologies of MD $[(L1 + L2 + L3)/3]$, L1 and radial diffusivity $[(L2 + L3)/2]$ and compared each parameter. Although MD was calculated from L1, L2 and L3, they have different information at least from the viewpoint of pathology (Griffith et al., 2012). We therefore think that each parameter has an important role clinically and pathologically and selected these parameters in the study. Some previous studies used one parameter, especially ADC, for tumour grading. Although one study suggested that HGGs with lower ADC values corresponded to increased cellularity (Kao et al., 2013), another reported that there was no significant difference in ADC values between LGGs and HGGs (Lam et al., 2002). These studies suggest that differentiating glioma grade on the basis of only one parameter is difficult. In our study, we used seven DTI-based parameters, including DWI, FA, MD, S0, L1, L2 and L3, and the results suggested that the DTI-based multiple parameters were useful for predicting glioma grade. However, it remains unclear which parameters have the most influence on clustering tumours. We identified unique DWI and FA Component Planes that can be predictive of malignancy. A previous study revealed no significant differences in DWI between LGGs and HGGs and indicated that predicting tumour grade only by using DWI had a limitation (Kono et al., 2001). Class numbers 14, 15 and 16 that were significantly higher in HGGs showed similar DWI values but different patterns of values in other DTI-based parameters. In addition to the results, we also calculated the AUC if using the combination of only DWI and FA, only DWI and only FA. However, the AUC of two parameters of DWI and FA was 0.642, 0.645, and 0.637, respectively and they all were much lower than the AUC of seven parameters. The additional analyses also support the idea that combined multiple parameters could be useful for predicting the grade of gliomas. Conversely, although FA values of HGGs were higher than LGGs in a previous study (Kao et al., 2013), class numbers 14 and 15 showed low FA values. Although general patterns of MD, S0, L1, L2 and L3 Component Planes in SOM seemed similar, their subtle differences revealed by multi-variate pattern analysis may be important for predicting grade. In fact, a pathological study using MRI reported that higher astrocytosis correlated with higher ADC and L1 (Griffith et al., 2012). Most cases of reactive astrocytosis can be identified from the infiltrating edge of an astrocytoma (Ironsides, 1994), and it may be difficult to distinguish diffuse astrocytoma from astrocytosis (Camelo-Piragua et al., 2011). Axonal damage and cellular infiltration led to a reduction of L1 (Boretius et al., 2012), and radial diffusivity $[(L2 + L3)/2]$ has been proposed as a marker of demyelination (Klawiter et al., 2011). Furthermore, loss of oligodendrocytes

correlated with L2 and L3 but not L1 (Griffith et al., 2012). Although we extracted seven parameters from DTI, it remains unclear which combination of the parameters is the best or if additional parameters can increase the accuracy for glioma grading. Several features other than the seven parameters, for example, eigenvectors, might provide better results if included. Further studies are needed to assess which combination of parameters is best for glioma grading.

One of the limitations in our study was the influence of distortions on voxel-based clustering. Original DWIs for DTI were acquired by using EPI to reduce acquisition time and artefacts related to physiological motion. Unfortunately, this fast acquisition scheme is highly sensitive to eddy currents induced by the large diffusion gradients and to another distortion induced by susceptibility artefacts (Mangin et al., 2002). A prominent source of artefacts for EPI is the effect of inhomogeneities close to tissue–air and tissue–bone interfaces, such as those around the frontal sinus and petrous bone (Jezzard and Clare, 1999). To reduce artefacts by eddy currents, the data were corrected by using the FDT program (Smith et al., 2004). Furthermore, the FUGUE program was used to reduce susceptibility artefacts for EPI that depend on the subject's head geometry. Only DTI-based parameters were used for glioma grading in our study, although conventional MRI sequences, such as T1WI, T1Wice and T2WI, were scanned. The reason is that these conventional images are spatially different from DTI because of distortions. The differences require more accurate image registration between them for calculating voxel-based images. Multiple parameters of DTI-based parameters as well as conventional MRI parameters combined with an accurate registration algorithm may generate voxel-based clustered images that predict the grade of gliomas more accurately.

4.4. Ratio of DTI-based parameters in LGG and HGG

The ratios of class numbers 14, 15 and 16 were significantly higher in HGGs than in LGGs. Class numbers 14, 15 and 16 seemed similar with regard to high DWI values. There was no significant difference between LGGs and HGGs in a past study (Kono et al., 2001), suggesting that grading tumours only by DWI is controversial. In our study, class numbers 3 and 5, which were not different between LGGs and HGGs, also had high DWI values. These results support the controversy regarding DWI and suggest that glioma grading only by DWI may be difficult.

Class numbers 14 and 15 had low FA values. In contrast, class number 16 had high FA values. Some researchers have reported that changes in FA in gliomas may indicate tumour cell infiltration (Schluter et al., 2005; Kallenberg et al., 2013). A reduction in FA seemed to be the common denominator among structural abnormalities (Wieshmann et al., 1999). Lower FA values were seen in LGGs in one study (Inoue et al., 2005), whereas another study showed no significant differences between LGGs and HGGs (Goebell et al., 2006). In our study, both increases and decreases in FA were seen in tumours. Particularly, class numbers 13 and 16 showed that high FA values were also included in HGGs.

Class numbers 14–16, which had significantly higher values in HGGs, seemed to have low MD values. However, most classes (numbers 2, 3, 5, 7, 8, 9, 10, 11, 12 and 13) with low MD values were not different between HGGs and LGGs. These results suggest that it is difficult to predict glioma grade only by using one parameter, such as DWI, FA or MD as discussed in many previous studies.

4.5. Glioma grading

The current gold standard for determining glioma grade is histopathological assessment. However, the limitations of histopathological assessment because of the heterogeneity of gliomas are well known. First, there is a possibility of sampling error. Because only a few small pieces of tissue are assessed and we tend to examine T1-enhancing lesions, the most malignant tissue may not be obtained. Sampling errors may occur especially for biopsy only (Law et al., 2006). Even if we resect most tumours, a thorough investigation is not always made in the

resected tumour, and this is a histopathological limitation at present. Several studies have been conducted to evaluate more accurate selection of targets by using positron emission tomography (PET) images with ^{18}F -labelled fluorodeoxyglucose and ^{11}C -labelled methionine to reduce sampling errors (Levivier et al., 1995; Pirotte et al., 2004). However, the radiotracer injection in PET is invasive because of radiation exposure. By using DTcls, we can non-invasively predict the grade of gliomas with accuracy and may perform regional grading of glioma, which is useful for targeting biopsy. Gliomas are heterogeneous tumours. If we can preoperatively predict the regional grading of a tumour, we can know which region must be resected, including peritumoural oedematous lesions. Second, some LGGs develop into HGGs, and $\geq 10\%$ of gliomas dedifferentiate into more malignant grades (Law et al., 2006). We cannot know when tumour grade progresses. By using regional grading based on DTcls, we can clarify when LGGs progress into HGGs during follow-up and provide an appropriate adjuvant treatment at the optimum time. The potential benefit of the proposed method in the present study could be emphasized by undertaking a prospective, randomized controlled study.

5. Conclusion

This study applied a two-level clustering approach, which consisted of SOM followed by the KM++ algorithm, for unsupervised clustering of a large number of input vectors with multiple features by DTI. The greatest point of the method is to enable obtaining novel clustering images called DTcls, that can give visual grading of glioma and be helpful in differentiating between LGGs and HGGs without pathological information. Our new approach could lead to more accurate non-invasive grading and more appropriate treatment.

Acknowledgements

We thank Drs. N. Sawamoto and S. Urayama for their technical assistance with the MRI acquisition. This work was partly supported by the following: Grant-in-Aid for Young Scientists (B) 24791296 from the Japan Society for the Promotion of Science (JSPS) (to N.O.), Grant-in-Aid for Young Scientists (B) 25861273 from JSPS (to T. Kikuchi), Grant-in-Aid for Scientific Research (C) 24592159 from JSPS (to T. Kunieda), Grant-in-Aid for Scientific Research (C) 25462254 from JSPS (to Y.A.), Grant-in-Aid for Challenging Exploratory Research 25670623 and Scientific Research (B) 24390343 from JSPS (to S.M.), Grant-in-Aid for Scientific Research on Innovative Areas, 'Glial assembly: a new regulatory machinery of brain function and disorders' (to H.F. and N.O.), and 'Development of BMI Technologies for Clinical Application' under the Strategic Research Program for Brain Sciences by the Ministry of Education, Culture, Sports, Science and Technology of Japan (to N.O.).

References

- Alhoniemi, E., Hollmen, J., Simula, O., Vesanto, J., 1999. Process monitoring and modeling using the self-organizing map. *Integrated Computer-Aided Engineering* 6, 3–14.
- Arthur, D., Vassilvitskii, S., 2007. k-means++: the advantages of careful seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms Society for Industrial and Applied Mathematics*, Philadelphia, PA, USA.
- Beccali, M., Cellura, M., Lo Brano, V., Marvuglia, A., 2004. Forecasting daily urban electric load profiles using artificial neural networks. *Energy Conversion and Management* 45, 2879–2900.
- Boretius, S., Escher, A., Dallenga, T., Wrzoss, C., Tammer, R., Brück, W., et al., 2012. Assessment of lesion pathology in a new animal model of MS by multiparametric MRI and DTI. *Neuroimage* 59, 2678–2688. <http://dx.doi.org/10.1016/j.neuroimage.2011.08.05121914485>.
- Brugger, D., Bogdan, M., Rosenstiel, W., 2008. Automatic cluster detection in Kohonen's SOM. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, 19 442–459 <http://dx.doi.org/10.1109/TNN.2007.90955618334364>.
- Camelo-Piragua, S., Jansen, M., Ganguly, A., Kim, J.C., Cosper, A.K., Dias-Santagata, D., et al., 2011. A sensitive and specific diagnostic panel to distinguish diffuse astrocytoma from astrocytosis: chromosome 7 gain with mutant isocitrate dehydrogenase 1 and p53. *Journal of Neuropathology and Experimental Neurology* 70, 110–115. <http://dx.doi.org/10.1097/NEN.0b013e31820565f921343879>.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 1–27.
- Chavez-Alvarez, R., Chavoya, A., Mendez-Vazquez, A., 2014. Discovery of possible gene relationships through the application of self-organizing maps to DNA microarray databases. *PLoS One* 9, e93233. <http://dx.doi.org/10.1371/journal.pone.009323324699245>.
- Chuang, K.H., Chiu, M.J., Lin, C.C., Chen, J.H., 1999. Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy C-means. *IEEE Transactions on Medical Imaging* 18, 1117–1128. <http://dx.doi.org/10.1109/42.81932210695525>.
- Dosenbach, N.U., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., et al., 2010. Prediction of individual brain maturity using fMRI. *Science* (New York, N.Y.) 329, 1358–1361. <http://dx.doi.org/10.1126/science.119414420829489>.
- Ehsani, A.H., Quiel, F., 2008. Application of self organizing map and SRTM data to characterize yardangs in the Lut desert, Iran. *Remote Sensing of Environment* 112, 3284–3294.
- Filippi, C.G., Ulug, A.M., Ryan, E., Ferrando, S.J., van Gorp, W., 2001. Diffusion tensor imaging of patients with HIV and normal-appearing white matter on MR images of the brain. *AJNR. American Journal of Neuroradiology* 22, 277–283 <http://dx.doi.org/10.1177/000061230102200507>.
- Goebell, E., Paustenbach, S., Vaeterlein, O., Ding, X.Q., Heese, O., Fiehler, J., et al., 2006. Low-grade and anaplastic gliomas: differences in architecture evaluated with diffusion-tensor MR imaging. *Radiology* 239, 217–222. <http://dx.doi.org/10.1148/radiol.238305005916484348>.
- Griffith, J.L., Shimony, J.S., Cousins, S.A., Rees, S.E., McCurnin, D.C., Inder, T.E., et al., 2012. MR imaging correlates of white-matter pathology in a preterm baboon model. *Pediatric Research* 71, 185–191. <http://dx.doi.org/10.1038/pr.2011.3322258130>.
- Holodny, A.L., Ollenschlaeger, M.D., Liu, W.C., Schuder, M., Kalnin, A.J., 2001. Identification of the corticospinal tracts achieved using blood-oxygen-level-dependent and diffusion functional MR imaging in patients with brain tumors. *AJNR. American Journal of Neuroradiology* 22, 83–88 <http://dx.doi.org/10.1177/000061230102200507>.
- Hothorn, T., Hornik, K., van de Wiel, M., Zeileis, A., 2006. A lego system for conditional inference. *American Statistician* 60, 257–263.
- Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2003. A Practical Guide to Support Vector Classification. Department of Computer Science, National Taiwan University.
- Inoue, T., Ogasawara, K., Beppu, T., Ogawa, A., Kabasawa, H., 2005. Diffusion tensor imaging for preoperative evaluation of tumor grade in gliomas. *Clinical Neurology and Neurosurgery* 107, 174–180. <http://dx.doi.org/10.1016/j.clineuro.2004.06.01115823671>.
- Ironsides, J.W., 1994. Update on central nervous system cytopathology. II. Brain smear technique. *Journal of Clinical Pathology* 47, 683–688 <http://dx.doi.org/10.1136/jclinpath.1994.01521979382>.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. *Neuroimage* 62, 782–790. <http://dx.doi.org/10.1016/j.neuroimage.2011.09.01521979382>.
- Jezzard, P., Balaban, R.S., 1995. Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 34, 65–73 <http://dx.doi.org/10.1006/mrm.1995.0005>.
- Jezzard, P., Clare, S., 1999. Sources of distortion in functional MRI data. *Human Brain Mapping* 8, 80–85 <http://dx.doi.org/10.1002/hbm.100524596>.
- Jin, H., Shum, W.H., Leung, K.S., Wong, M.L., 2004. Expanding self-organizing map for data visualization and cluster analysis. *Information Sciences* 163, 157–173.
- Jones, D.K., Horsfield, M.A., Simmons, J., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 42 (3), 515–525 <http://dx.doi.org/10.1006/mrm.1999.0005>.
- Kallenberg, K., Goldmann, T., Menke, J., Strik, H., Bock, H.C., Stockhammer, F., et al., 2013. Glioma infiltration of the corpus callosum: early signs detected by DTI. *Journal of Neuro-Oncology* 112, 217–222. <http://dx.doi.org/10.1007/s11060-013-1049-y23344787>.
- Kao, H.W., Chiang, S.W., Chung, H.W., Tsai, F.Y., Chen, C.Y., 2013. Advanced MR imaging of gliomas: an update. *BioMed Research International* 2013, 1–14. <http://dx.doi.org/10.1155/2013/97058623862163>.
- Klawiter, E.C., Schmidt, R.E., Trinkaus, K., Liang, H.F., Budde, M.D., Naismith, R.T., et al., 2011. Radial diffusivity predicts demyelination in ex vivo multiple sclerosis spinal cords. *Neuroimage* 55, 1454–1460. <http://dx.doi.org/10.1016/j.neuroimage.2011.01.00721238597>.
- Klingberg, T., Vaidya, C.J., Gabrieli, J.D., Moseley, M.E., Hedehus, M., 1999. Myelination and organization of the frontal white matter in children: a diffusion tensor MRI study. *Neuroreport* 10, 2817–2821 <http://dx.doi.org/10.1097/000061230102200507>.
- Kohonen, T., 1995. *Self-Organizing Maps*. Springer, Berlin.
- Kono, K., Inoue, Y., Nakayama, K., Shakudo, M., Morino, M., Ohata, K., et al., 2001. The role of diffusion-weighted imaging in patients with brain tumors. *AJNR. American Journal of Neuroradiology* 22, 1081–1088 <http://dx.doi.org/10.1177/000061230102200507>.
- Lam, W.W., Poon, W.S., Metreweli, C., 2002. Diffusion MR imaging in glioma: does it have any role in the pre-operation determination of grading of glioma? *Clinical Radiology* 57, 219–225. <http://dx.doi.org/10.1053/crad.2001.07411952318>.
- Law, M., Oh, S., Babb, J.S., Wang, E., Ingles, M., Zagzag, D., et al., 2006. Low-grade gliomas: dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging – prediction of patient clinical response. *Radiology* 238, 658–667. <http://dx.doi.org/10.1148/radiol.238204218016396838>.
- Levivier, M., Goldman, S., Pirotte, B., Brucher, J.M., Balériaux, D., Luxen, A., et al., 1995. Diagnostic yield of stereotactic brain biopsy guided by positron emission tomography with [^{18}F]fluorodeoxyglucose. *Journal of Neurosurgery* 82, 445–452. <http://dx.doi.org/10.3171/jns.1995.82.3.04457861223>.
- Liao, W., Chen, H., Yang, Q., Lei, X., 2008. Analysis of fMRI data using improved self-organizing mapping and spatio-temporal metric hierarchical clustering. *IEEE Transactions on Medical Imaging* 27, 1472–1483. <http://dx.doi.org/10.1109/TMI.2008.92398718815099>.

- Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., 2007a. WHO Classification of Tumours of the Central Nervous system fourth edition. WHO, Lyon, <http://dx.doi.org/10.1007/s00401-007-0243-417618441>.
- Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., et al., 2007b. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathologica* 114, 97–109. <http://dx.doi.org/10.1007/s00401-007-0243-417618441>.
- Lu, H.C., Hsieh, J.C., Chang, T.S., 2006. Prediction of maximum daily ozone level using combined neural network and statistical characteristics. *Atmospheric Research* 81, 124–139.
- MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Symposium on Math, Statistics, and Probability*, Berkeley University of California Press, CA.
- Mangin, J.F., Poupon, C., Clark, C., Le Bihan, D., Bloch, I., 2002. Distortion correction and robust tensor estimation for MR diffusion imaging. *Medical Image Analysis* 6, 191–198. <http://dx.doi.org/10.1006/mi.2002.12270226>.
- McLendon, R.E., Halperin, E.C., 2003. Is the long-term survival of patients with intracranial glioblastoma multiforme overstated? *Cancer* 98, 1745–1748. <http://dx.doi.org/10.1002/cncr.1166614534892>.
- Ohgaki, H., Kleihues, P., 2007. Genetic pathways to primary and secondary glioblastoma. *American Journal of Pathology* 170, 1445–1453. <http://dx.doi.org/10.2353/ajpath.2007.07001117456751>.
- Ota, K., Oishi, N., Ito, K., Fukuyama, H., 2014. A comparison of three brain atlases for MCI prediction. *Journal of Neuroscience Methods* 221, 139–150. <http://dx.doi.org/10.1016/j.jns.2014.01.018>.
- Peltier, S.J., Polk, T.A., Noll, D.C., 2003. Detecting low-frequency functional connectivity in fMRI using a self-organizing map (SOM) algorithm. *Human Brain Mapping* 20, 220–226. <http://dx.doi.org/10.1002/hbm.1014414673805>.
- Pirotte, B., Goldman, S., Massager, N., David, P., Wikler, D., Lipszyc, M., et al., 2004. Combined use of 18F-fluorodeoxyglucose and 11C-methionine in 45 positron emission tomography-guided stereotactic brain biopsies. *Journal of Neurosurgery* 101, 476–483. <http://dx.doi.org/10.3171/jns.2004.101.3.047615352606>.
- Price, S.J., Burnet, N.G., Donovan, T., Green, H.A., Peña, A., Antoun, N.M., et al., 2003. Diffusion tensor imaging of brain tumours at 3T: a potential tool for assessing white matter tract invasion? *Clinical Radiology* 58, 455–462. <http://dx.doi.org/10.1054/cr.2003.38314>.
- Rajini, H.N., Bhavani, R., 2012. Automatic MR brain tumor detection using possibilistic C-means and K-means clustering with color segmentation. *International Journal of Computer Applications* 56, 11–17.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., et al., 2011. pROC: an open-source package for R and S+ to analyse and compare ROC curves. *BMC Bioinformatics* 12, 77. <http://dx.doi.org/10.1186/1471-2105-12-7721414208>.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20, 53–65.
- Sanai, N., Berger, M.S., 2008. Glioma extent of resection and its impact on patient outcome. *Neurosurgery* 62, 753–764. <http://dx.doi.org/10.1227/00006123-200806000-00018>.
- Scherer, H.J., 1938. Structural development in gliomas. *American Journal of Cancer* 34, 333–351.
- Schlüter, M., Stieltjes, B., Hahn, H.K., Rexilius, J., Konrad-verse, O., Peitgen, H.O., 2005. Detection of tumour infiltration in axonal fibre bundles using diffusion tensor imaging. *International Journal of Medical Robotics + Computer Assisted Surgery: MRCAS* 1, 80–86. <http://dx.doi.org/10.1002/rcs.3117518394>.
- Scott, J.N., Brasher, P.M., Sevic, R.J., Rewcastle, N.B., Forsyth, P.A., 2002. How often are nonenhancing supratentorial gliomas malignant? A population study. *Neurology* 59, 947–949. <http://dx.doi.org/10.1213/00006123-200209000-00059>.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., et al., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23 (Suppl. 1), S208–S219. <http://dx.doi.org/10.1016/j.neuroimage.2004.07.05115501092>.
- Stadnik, T.W., Chaskis, C., Michotte, A., Shabana, W.M., van Rompaey, K., Luyt, R., et al., 2001. Diffusion-weighted MR imaging of intracerebral masses: comparison with conventional MR imaging and histologic findings. *AJNR. American Journal of Neuroradiology* 22, 969–976. <http://dx.doi.org/10.3171/ajnr.2001.22.5.969>.
- Thomassey, S., Happiette, M., 2007. A neural clustering and classification system for sales forecasting of new apparel items. *Applied Soft Computing* 7, 1177–1187.
- Thomsen, H.S., 2003. Guidelines for contrast media from the European Society of Urogenital Radiology. *AJR. American Journal of Roentgenology* 181, 1463–1471. <http://dx.doi.org/10.2214/ajr.181.6.181146314627556>.
- Tien, R.D., Felsberg, G.J., MacFall, J., Friedman, H., Brown, M., 1994. MR imaging of high-grade cerebral gliomas: value of diffusion-weighted echoplanar pulse sequences. *AJR. American Journal of Roentgenology* 162, 671–677. <http://dx.doi.org/10.2214/ajr.162.3.81095208109520>.
- Vapnik, V.N., 1998. *The Nature of Statistical Learning Theory*. Springer Verlag, New York.
- Velmurugan, T., Santhanam, T., 2010. Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points. *Journal of Computer Science* 6, 363–368.
- Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks* / a Publication of the IEEE Neural Networks Council, 11, 586–600. <http://dx.doi.org/10.1109/72.84673118249787>.
- Vijayakumar, C., Damayanti, G., 2007. Pant R, Sreedhar CM. Segmentation and grading of brain tumors on apparent diffusion coefficient images using self-organizing maps. *Computerised Medical Imaging and Graphics* 31, 473–484. <http://dx.doi.org/10.1016/j.cmi.2007.05.008>.
- Wang, J., Delabie, J., Aasheim, H.C., Smeland, E., Myklebost, O., 2002. Clustering of the SOM easily reveals distinct gene expression patterns: results of a reanalysis of lymphoma study. *BMC Bioinformatics* 3, 1–9.
- Watanabe, M., Tanaka, R., Takeda, N., 1992. Magnetic resonance imaging and histopathology of cerebral gliomas. *Neuroradiology* 34, 463–469. <http://dx.doi.org/10.2214/ajr.150.3.1500436452>.
- Wieshmann, U.C., Clark, C.A., Symms, M.R., Franconi, F., Barker, G.J., Shorvon, S.D., 1999. Reduced anisotropy of water diffusion in structural cerebral abnormalities demonstrated with diffusion tensor imaging. *Magnetic Resonance Imaging* 17, 1269–1274. <http://dx.doi.org/10.1054/mri.1999.076712>.
- Zacharaki, E.I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E.R., et al., 2009. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 62, 1609–1618. <http://dx.doi.org/10.1002/mrm.2214719859947>.