Journal of Advanced Research 16 (2019) 15–23

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Original Article

Gaussian mixture model for texture characterization with application to brain DTI images

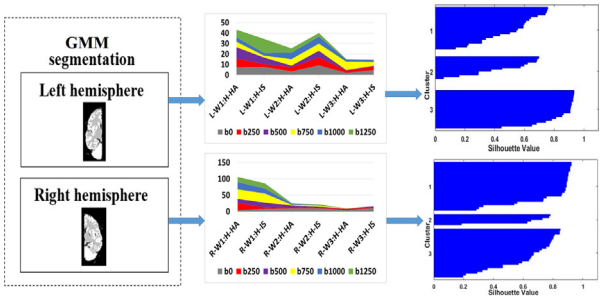
Luminita Moraru [a,*], Simona Moldovanu [a,b], Lucian Traian Dimitrievici [a], Nilanjan Dey [c], Amira S. Ashour [d], Fuqian Shi [e], Simon James Fong [f], Salam Khan [g], Anjan Biswas [g,h,i]

[a] Faculty of Sciences and Environment, Modelling & Simulation Laboratory, Dunarea de Jos University of Galati, 47 Domneasca Str., 800008, Romania
[b] Department of Computer Science and Engineering, Electrical and Electronics Engineering, Faculty of Control Systems, Computers, Dunarea de Jos University of Galati, Romania
[c] Techno India College of Technology, West Bengal 740000, India
[d] Department of Electronics and Electrical Communication Engineering, Faculty of Engineering, Tanta University, 31512, Egypt
[e] College of Information and Engineering, Wenzhou Medical University, Wenzhou 325035, PR China
[f] Department of Computer and Information Science, Data Analytics and Collaborative Computing Laboratory, University of Macau, Taipa, Macau SAR 999078, PR China
[g] Department of Physics, Chemistry and Mathematics, Alabama A&M University, Normal AL-35762, USA
[h] Department of Mathematics, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[i] Department of Mathematics and Statistics, Tshwane University of Technology, Pretoria 0008, South Africa

HIGHLIGHTS

- A Gaussian mixture model to classify the pixel distribution of main brain tissues is introduced.
- A hemisphere approach is proposed.
- Mixing probabilities at the sub-class and class levels are estimated.
- The k-means algorithm optimizes the parameters of the mixture distributions.
- A difference in the mixing probabilities between hemispheres is determined.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 20 October 2018
Revised 31 December 2018
Accepted 1 January 2019
Available online 4 January 2019

Keywords:
Gaussian mixture model
Brain hemispheres
Weight distribution
Weighted Euclidean distance
Clustering
Cluster validity

ABSTRACT

A Gaussian mixture model (GMM)-based classification technique is employed for a quantitative global assessment of brain tissue changes by using pixel intensities and contrast generated by b-values in diffusion tensor imaging (DTI). A hemisphere approach is also proposed. A GMM identifies the variability in the main brain tissues at a macroscopic scale rather than searching for tumours or affected areas. The asymmetries of the mixture distributions between the hemispheres could be used as a sensitive, faster tool for early diagnosis. The k-means algorithm optimizes the parameters of the mixture distributions and ensures that the global maxima of the likelihood functions are determined. This method has been illustrated using 18 sub-classes of DTI data grouped into six levels of diffusion weighting (b = 0; 250; 500; 750; 1000 and 1250 s/mm^2) and three main brain tissues. These tissues belong to three subjects, i.e., healthy, multiple haemorrhage areas in the left temporal lobe and ischaemic stroke. The mixing probabilities or weights at the class level are estimated based on the sub-class-level mixing probability estimation. Furthermore, weighted Euclidean distance and multiple correlation analysis are applied to analyse the dissimilarity of mixing probabilities between hemispheres and subjects. The silhouette data

Peer review under responsibility of Cairo University.
* Corresponding author.
E-mail address: luminita.moraru@ugal.ro (L. Moraru).

https://doi.org/10.1016/j.jare.2019.01.001

evaluate the objective quality of the clustering. By using a GMM in the present study, we establish an important variability in the mixing probability associated with white matter and grey matter between the left and right hemispheres.

Introduction

Accurate and non-invasive methods capable of detecting and correcting localized affected areas of the brain are of substantial interest because there is significant variability in the location and extent of such areas. Moreover, there is substantial variability among individuals. The inter-subject variability of brain structures is apparent for normal subjects as well as for patients with various brain injuries [1,2].

The statistical analysis presented in this paper is based on Gaussian models [3–10]. A Gaussian mixture model (GMM) is a probabilistic model based on a Gaussian distribution for expressing the presence of sub-populations/sub-classes within an overall population/class without requiring the identification of the sub-class of interest (observational data) [11]. That is, a GMM learns to detect injured tissues using healthy patient data. Banfield and Raftery [4] considered both Gaussian and non-Gaussian models to specify the features of the clusters and to estimate which features likely to differ between clusters. These authors applied the proposed method to brain MRI images to identify similarly anatomical structures. Fraley and Raftery [6] systematically reviewed how finite mixture models provide a statistical framework for application in clustering, the effectiveness of a certain clustering method and the influence of outliers on cluster analysis. Paalanen et al. [7] investigated several estimation methods for GMMs, enabling an improvement in the representation and discrimination of patterns. Kim et al. [9] proposed a method to assess gross brain abnormalities using a GMM, with at least two Gaussian components to allocate a specific mixing probability to each subject. Then, the assigned mixture probabilities are tested between the studied groups. The authors stated that a GMM is an effective method in terms of computing resources because it does not incorporate any subject or group-specific parameters. Another popular approach is the multivariate Gaussian method, which is used to identify features and discriminate between different classes in various applications, such as hazardous chemical agents [12], moving parts of electric motors under normal conditions and those with bearing failure [13]. However, the usage of a single feature results in a single fundamental class with variables exhibiting smooth behaviour. Consequently, certain errors in the estimation of the probability density function (pdf) occur, and the discrimination between classes fails. A GMM treats analysed data as a mixture of component distributions, and the main challenge here is to correctly estimate the model parameters such as the weight of the *i*th component, which can be interpreted as the a priori probability, mean or covariance matrix of the normally distributed random variable. A GMM is a simple physical and data-driven model; it permits a flexible characterization of unusual distributions of pixels and provides a quantitative analysis of DTI data [14]. DTI captures vital information and plays a significant role in *in vivo* studies of anatomical structures in brain regions [15]. Generally, brain tissues such as grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF) are considered classes; a few of their pixel features are predominantly measurable and serve as model parameters. A GMM assigns a probability to each pixel if it belongs to only one class. These parameters are learned using the expectation-maximization (EM) algorithm [16,17]. More recently, various associations between prior knowledge on human neuroanatomy and conditional probability distributions that can characterize various brain tissues or anatomy classes were made feasible by considering a GMM associated with various neuroanatomical labels [18,19].

In this study, a GMM-based classification scheme to identify the variability in the main brain tissue in DTI images (rather than searching for tumour areas) is proposed. To consider the signal intensity and contrast effects on image quality, we used multiple b-values. Han et al. [20] reported that brain imaging using a high b-value likely improves both the contrast between tissues and the capability of detecting less prominent lesions. As a probabilistic model, a GMM can characterize arbitrary mixture distributions composed of WM, GM and CSF with unknown parameters. A k-means algorithm is used to optimize the means, variances and mixture probability of the mixture distributions and to ensure that a global maximum of the likelihood function is achieved. The weighted Euclidean distance (wd) is used to validate the capability of a GMM to discriminate between the mixing probabilities across the studied classes. Moreover, a multiple correlation analysis between the left and right hemispheres based on the established mixing probabilities is performed. Finally, the silhouette plot (size and width) is used to evaluate the clustering validity. DTI images of a normal subject without a history of head injury or cerebrovascular disease (denoted H), a patient with multiple haemorrhage areas in the left temporal lobe (HA) and another with ischaemic stroke (IS) were studied.

Methodology

Finite GMM with m components

Brain DTI images contain heterogeneous sub-classes, and the analysis based on the mixture of models is adequate to model the entire distribution containing numerous sub-classes. Different tissues such as WM, GM, and CSF or lesioned tissues aggregate their intensities and contrast, which is essentially decided by the b-values in DTI, under different Gaussian curves with distinct mean and covariance parameters. The highest probability of classifying each pixel as belonging to the WM, GM and CSF or lesioned tissues is the basis of a GMM. The mean values of the weights of each brain tissue and for each b-value are projected onto a vector space with a three-dimensional feature [w1, w2, and w3]. This vector contains the intensities of the pixels for each available b-value. First, healthy subject-specific information is integrated into the algorithm using the GMM by means of a training stage. During this training stage, the weights, mean, and variance for each individual Gaussian density are determined. These factors are prior probabilities for parameters required in the initialization step of the EM algorithm. Then, the GMM parameters for the test data are established using the maximum likelihood and an iterative EM algorithm. The GMM generates a vector space, which contains the probability function of the data computed using the intensities of the pixels and their discriminative weights or mixture probability, for each available b-value and for each considered tissue class. The vector space is passed to the predictive model to capture the discriminative subject-specific information regarding the brain injuries from MRI images.

This approach is followed in an inter-hemisphere analysis, and the mixture weight functions are determined as a posteriori probability to prevent the repetition of an excessively large number of univariate analyses during the characterization of each subject. The brain segmentation into hemispheres involves the following steps: (i) skull stripping based on an irrational mask for filtration and binary morphological operations [21]; and (ii) mid-sagittal axis detection based on the location of the inter-hemispheric fissure and determination of the image centroid, as reported in [22].

In a preliminary step, an image histogram that provides raw information about the pdf of the pixel values is analysed. The number of components or Gaussian sources is established at three, according to the multimodal distribution in the histograms of the pixel distribution. The finite mixture model is based on the assumption that each finite mixture has similar probability distributions for each sub-class; however, inside the sub-class, different multivariate probability density distributions and different parameters are present [3]. For an image, let X denote a vector of pixel intensities $X = \{x_i\},\ i = \overline{1,N}$. X is a feature vector of the observation data for a specific subject and a specific b-value. This vector describes a sub-class, which, in turn, belongs to a class $\Omega_m,\ m = 1,2,3$. The probability function at the observation x_i is expressed as

$$f(x_i,\theta) = \sum_{m=1}^{3} \bar{w}_m f_m(x_i|b_j), \quad \forall j = \overline{1,6} \tag{1}$$

$f_m(x_i|b_j)$ denotes a component of the Gaussian mixture or 'mixture distribution', and $\bar{w}_m$ is the prior distribution of the pixel $x_i, x_i \in \Omega_m$ for each sub-class corresponding to b-values and is called mixing proportions or weights. The weights must satisfy the following conditions to be valid: $\sum_{m=1}^{3} \bar{w}_m = 1$ and $0 < \bar{w}_m \leqslant 1$, for each sub-class. As an a priori distribution, it is obtained by observing a healthy patient. Each density function of the mixture component $f_m(x_i|b_j)$ is characterized by a mean μ_i and a variance Σ and is a univariate normal pdf expressed as

$$f_m(x_i|b_j) = \frac{1}{\sqrt{2\pi\Sigma}} \exp\left(-\frac{1}{2}(X-\mu_i)^T \Sigma^{-1}(X-\mu_i)\right),$$
$$\forall m = 1,2,3 \; and \; \forall j = \overline{1,6} \tag{2}$$

$\theta = \left(\left\{\bar{w}_i\right\}, \mu_i, \Sigma\right)$ denotes a set of the main parameters of the GMM to be estimated by the EM algorithm. The likelihood function of the training vectors based on the probability function (2) is as follows:

$$L(\theta) = \prod_{i=1}^{N} \sum_{m=1}^{3} \bar{w}_m f(x_i|b_j), \quad \forall j = \overline{1,6} \tag{3}$$

where N is the total number of pixels in the image. The function permits one to establish the statistical model parameters. There is incomplete data in θ; therefore, these partially observed parameters in θ are estimated using an EM algorithm [23]. As an iterative algorithm, EM starts by using an initial model Ω, estimates a new model Ω', and in the next iteration, this new model Ω' becomes an initial model to determine a new model, Ω'', etc. This process is repeated until a predefined convergence condition is achieved. In the studied problem, the mixing proportions are estimated and subsequently validated.

(i) In the initialization step, the mean, variance and mixing coefficients of the training data were estimated for $m = 3$ classes. Established during the training stage, these parameters are the prior distribution. This distribution is used to initialize the value of the probability.

(ii) In the expectation step, based on the current estimated parameters θ, the EM algorithm computes the posterior probability using the current parameter values established in the initialization step; furthermore, the algorithm estimates the posterior probability that an observation x_i belongs to a sub-class j over all feasible assignments of data points to Gaussian sources, as

$$w_{ij} = \frac{\bar{w}_m f(x_i|b_j)}{f(x_i)}, \quad i = \overline{1,N}; \quad j = \overline{1,6}; \quad m = 1,2,3 \tag{4}$$

Then, for each sub-class j, the prior probabilities are computed by averaging the posterior probabilities for each class as

$$\langle w_j \rangle = \frac{1}{N} \sum_{i=1}^{N} w_{ij}, \quad j = \overline{1,6} \tag{5}$$

(iii) In the maximization step, the values of the old estimated parameters θ are updated or re-estimated by computing the maximum likelihood estimates of θ with the expected membership values [24,25]. The derivative of function (3) is determined and equated to zero. The determination of the global maxima of the likelihood functions implies the determination of those parameters that maximize the probability of observing the data. The new mean, variance and weight parameters are estimated, and the likelihood function is evaluated.

The iterative process continues through the expectation and maximization steps successively until convergence. Convergence implies that the changes in the parameters become small enough, i.e., it stops when $|\theta_i - \theta_{i-1}| \ll \varepsilon$, where $\varepsilon = 0.00001$ is small enough to assure no significant changes from one iteration to the next exist. If the convergence criterion is not attained, the algorithm returns to the expectation step.

k-means algorithm for clustering

The k-means algorithm is used to assess the data clustering for the selected number of classes ($m = 3$) [26,27]. Each mixture component is associated with a class or cluster based on identical estimated statistical parameters. The data produced by a GMM are clusters with centroids at the means. In the GMM, the EM algorithm converges to the local maxima of the likelihood function. However, the EM algorithm has a drawback, namely, it fails when the covariance matrix associated with a sub-class is singular or the number of observations is reduced. According to the GMM, three initial centroids are defined based on the set of parameters θ established in the maximization step of the EM algorithm. Thirty-five restarts of the k-means algorithm were executed to ensure that a global maximum is determined for each data set. That is, k-means optimizes the mean, variance and weight of the mixture distributions.

Weighted Euclidean distance and multiple correlation

To validate the ability of the GMM mixtures to differentiate between different subjects in a hemisphere approach, we used the wd between two j-dimensional vectors [28]:

$$wd_{H-IS}^{m} = \sqrt{\sum_j \left(\frac{w_{ij}^{H}}{s_j^{H}} - \frac{w_{ij}^{IS}}{s_j^{IS}}\right)^2} \text{ and } wd_{H-HA}^{m} = \sqrt{\sum_j \left(\frac{w_{ij}^{H}}{s_j^{H}} - \frac{w_{ij}^{HA}}{s_j^{HA}}\right)^2} \tag{6}$$

where w_{ij} is given by Eq. (4) for the studied sub-classes, and s_j is the corresponding standard deviation. H denotes the normal subject,

HA a patient with multiple haemorrhage areas in the left temporal lobe and IS another with ischaemic stroke. The wd balances the contributions of the variables in the computation of distance. The weight attached to the *j*th variable in a vector is related to the standard deviation of each distribution s_j^H, s_j^{IS} and s_j^{HA} [29]. A continuous image belongs to a metric space that uses metrics exhibiting the following properties: non-negativity, identity of indiscernibles, symmetry and triangle inequality. The wd is appropriate for measuring the dissimilarity of the two given mixing probabilities because other metrics fail to exhibit a few of these properties, e.g., the Kullback–Leibler distance fails to exhibit the symmetry property, and the Hellinger distance fails to exhibit the triangle inequality. The Minkowski and Mahalanobis distances are general formulations of the Euclidean distance. Hershey and Olsen [30] reported that the Kullback–Leibler divergence metric between two GMMs cannot be analytically applicable and that the algorithm is highly time consuming. Durrieu et al. [31] demonstrated that the Kullback–Leibler divergence metric can be approximated only.

Moreover, an inter-hemisphere multiple correlation analysis between the mixing probabilities was performed to characterize the association of the grey level intensities and contrast for the selected injured subjects and healthy subjects. The multiple correlation coefficients between the independent variables HA and *IS* and the dependent variable H are defined as

$$R^m_{H-(IS,HA)} = \sqrt{\frac{\left(r^m_{IS,H}\right)^2 + \left(r^m_{HA,H}\right)^2 - 2r^m_{IS,H}r^m_{HA,H}r^m_{IS,HA}}{1-\left(r^m_{IS,HA}\right)^2}} \tag{7}$$

where $r^m_{IS,H}$, $r^m_{HA,H}$, $r^m_{IS,HA}$, $m = 1, 2, 3$ are the covariance between the two random variables in each of the pairs IS and *H*, *HA* and *H* and IS and *HA*, respectively [32]. Accordingly, the correlation coefficient between two sub-classes *j* that belong to a class *m* for two random variables is illustrated in the example below:

$$r^m_{H,HA} = \frac{\sum_{i,j} w^H_{ij} w^{HA}_{ij} - K\langle w^H_i\rangle\langle w^{HA}_i\rangle}{\sqrt{\sum_{i,j}\left[\left(w^H_{ij}\right)^2\right] - K\left(\langle w^H_i\rangle\right)^2}\sqrt{\sum_{i,j}\left[\left(w^{HA}_{ij}\right)^2\right] - K\left(\langle w^{HA}_i\rangle\right)^2}} \tag{8}$$

where w_{ij} and $\langle w_i\rangle$ are defined in the expectation step, and $K = 10$ is the number of samples (i.e., images) for each sub-class.

Clustering validation

The analysis is focused on three main brain tissues (i.e., GM, WM and CSF), and an a priori assumption of three-class clustering is considered. The goal is to examine whether these classes reflect the actual clustering structure of the data or whether these data were partitioned into a few artificial groups in the context of the GMM [33]. The quality of the clustering analysis is addressed, and the silhouette index and silhouette plots are used as the validation criteria [34]. If compact and clearly separated clusters are obtained, the targeted tissues were considered well classified. Let a multivariate data w_{ij} be separated into m clusters, A_m, $w_{ij} \in A_m = A_{mH} \cup A_{mHA} \cup A_{mIS}$, $m = 1, 2, 3$. Let us suppose that the GM tissue is described by $w^H_{1i}, i = \overline{1,N}$ and $A_1 = A_{1H} \cup A_{1HA} \cup A_{1IS}$. We define the average dissimilarity of w^H_{1i} with all the other points k of the same cluster having the vector norm $|A_1|$, as follows:

$$\lambda^H_1 = \frac{1}{|A_1|-1}\sum_{k=1}^{N} \| w^H_{1i} - w^H_{1k} \| 1_A(w^H_{1k}), \quad i \neq k \tag{9}$$

where $\| \|$ denotes a 2-norm (L^2). Further, λ^{H-HA}_1 and λ^{H-IS}_1 describe the average dissimilarity of the mixing probability of H with all the points belonging to other clusters HA and IS, respectively:

$$\lambda^{H-HA}_1 = \frac{1}{|A_1|-1}\sum_{k=1}^{N} \| w^H_{1i} - w^{HA}_{1k} \| 1_A(w^{HA}_{1k}) \text{ and } \lambda^{H-IS}_1 = \frac{1}{|A_1|-1}\sum_{k=1}^{N} \| w^H_{1i} - w^{IS}_{1k} \| 1_A(w^{IS}_{1k})$$

The smallest average dissimilarity to another cluster is defined as $\lambda^*_1 = \min\left\{\lambda^{H-HA}, \lambda^{H-IS}\right\}$. The silhouette index is

$$s(i) = \frac{\lambda^*_1 - \lambda^H_1}{\max\left\{\lambda^*_1, \lambda^H_1\right\}} = \begin{cases} 1 - \frac{\lambda^H_1}{\lambda^*_1} & if \quad \lambda^H_1 < \lambda^*_1 \\ 0 & if \quad \lambda^H_1 = \lambda^*_1 \\ \frac{\lambda^*_1}{\lambda^H_1} - 1 & otherwise \end{cases} \tag{10}$$

From Eq. (10), $s(i) \in [-1, 1]$, and if $s(i) \approx -1$, the least effective situation manifests. This method is also used for WM and CSF. Silhouette plots facilitate the interpretation of cluster analysis results because they are independent of the clustering algorithm used and rely only on the actual partition of the 'objects' [34].

Subjects, image acquisition, and processing

The algorithm flow is presented in Fig. 1.

Three subjects (age range 36–60 y; one female and two males) underwent MRI scans. A subject presented multiple haemorrhage areas in the left temporal lobe (male, 48 y), and another presented with IS in the left frontal lobe (female, 60 y, median 8-mo post-stroke). The third subject was a healthy patient (male, 36 y). A series of DTI images were acquired using a pulsed gradient spin-echo sequence in 15 directions and five b-values ($b1 = 250$ s/mm^2; $b2 = 500$ s/mm^2; $b3 = 750$ s/mm^2; $b4 = 1000$ s/mm^2; $b5 = 1250$ s/mm^2). Moreover, images without diffusion gradients ($b0 = 0$ s/mm^2) and with otherwise identical imaging parameters were acquired. A total of 190 images were tested. A b-value encompasses information regarding the strength and timing of the gradients used to generate diffusion-weighted images. Larger b-values provide better contrast among tissues. The selection of b-value continues to be a challenge and strongly depends on the investigated anatomical features or pathology, field strength and average number of signals. In the case of the GMM, the mixing probabilities depend on the experimental conditions, i.e., the diffusion effect or b-value. Multiple b-values permit the use of a small sample size because each data set exhibits characteristics unique to it. The within-subject correlation is avoided by summarizing each mixing probability sequence with a single number. In this case, only a comparison of the statistics between the classes (see data in Table 3) is performed. Averaging repeated measurements is a reasonable choice, especially when the effect of the injury is maintained quite steadily over acquisition time.

For the data acquisition, a 1.5-T MRI scanner was operated (Philips Medical Systems, Best, Netherlands). The diffusion-weighted scans utilized a system with six-channel sensitivity encoding (SENSE) for faster scanning (FS = 1.5). The imaging parameters were as follows: 3D gradient echo with echo time ranging from 83 to 110 ms; repetition time ranging from 6500 to 7800 ms (it varies between subjects); bandwidth = 1070 Hz/pixel; flip angles (2- and 6-); voxel resolution ranging from 2.5 to 3.0 mm; and slice thickness = 4 mm. The acquisition matrix was 128 × 128. The standard Digital Imaging and Communications in Medicine (DICOM) image dataset was used.

Approval for the study was obtained from the Research Ethics Committee of the Dunarea de Jos University of Galati and Saint Andrew Hospital. Voluntary and written informed consent was obtained from each participant. The privacy policy is based on DICOM Confidential [35].

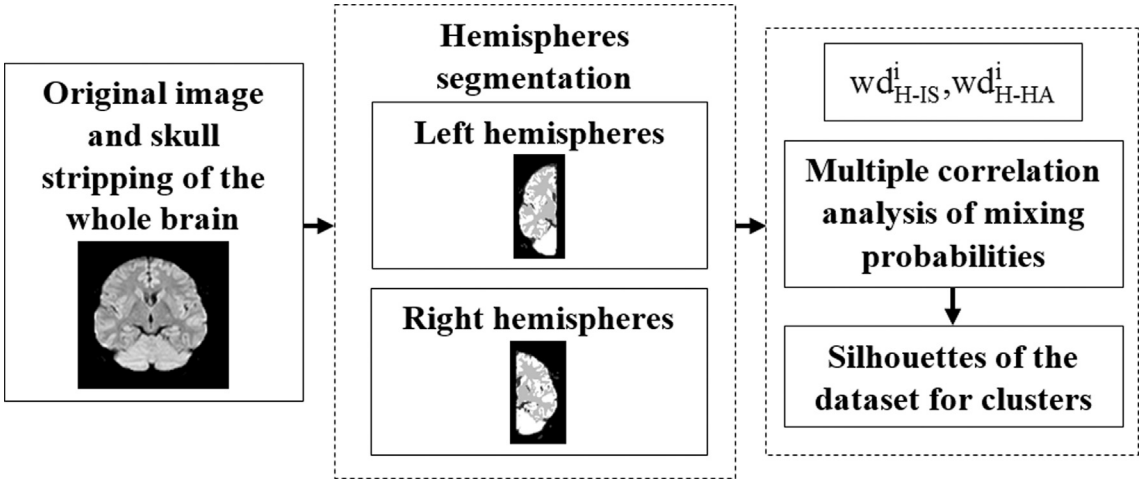


Fig. 1. Algorithm scheme.

Results

The proposed GMM-based classification approach in a brain hemisphere framework is aimed at facilitating the identification of variability in the main brain tissue in DTI images and circumventing the subsequent processing for detecting tumours or lesions. For example, a DTI image (b = 500 s/mm^2) of a healthy subject and the results of the GMM classification and hemisphere segmentation are shown in Fig. 2.

The estimated weights (Eq. (5)) across the entire control group (H) and for each injured group (IS and HA) are presented in Table 1 (for the left hemisphere) and Table 2 (for the right hemisphere). The data in Tables 1 and 2 present details on the difference in the averaged weights or mixing probabilities between the left and right hemispheres for each subject and over the entire range of diffusion gradient values. There are no differences in mixing probabilities for each tissue class between the left and right hemispheres for H class. This result indicates the 'normality' of the healthy subject. For HA and IS, visible differences in the mixing probabilities are presented.

In Fig. 3, the estimated average wds (Eq. (6)) for all diffusion gradients and for each brain hemisphere and subject are presented.

Fig. 3 indicates that the proposed approach exhibits the ability to highlight the differences between brain tissues in the right and left hemispheres for each level of diffusion weighting and subject category. This distance balances the contributions of the variables by considering the standard deviation of each distribution.

The correlation matches images characterized by various intensities and contrasts, albeit with largely similar local intensity variations. The results of the correlation analysis (Eqs. (7) and (8)) are presented in Table 3. First, the correlation between each pair of classes has been investigated. The results indicate that classes HA and IS are not correlated because the correlation coefficient is near zero. This observation leads to the following hypothesis: H is the dependent variable, and HA and IS are not correlated and are the independent variables. Therefore, the multiple correlation coefficient is computed according to Eq. (7).

As the data in Table 3 indicate, for the CSF class (index 3), HA and IS do not correlate with H for neither the left or right hemisphere. The results for the WM class (index 2) illustrate that, for the left hemisphere, HA and IS are marginally correlated with H. The correlation increases by approximately 50% for the right hemisphere. For the GM class (index 1), HA and IS correlate well with H for the right hemisphere and do not correlate with H for the left hemisphere.

The resulting silhouette plots (Eq. (10)) for the whole brain and the left and right hemispheres are displayed in Fig. 4.

The average silhouette width is approximately 0.9, i.e., 90% of the selected clusters are considered the optimal number of clusters (Table 4). The a priori selection of the three main brain tissues or

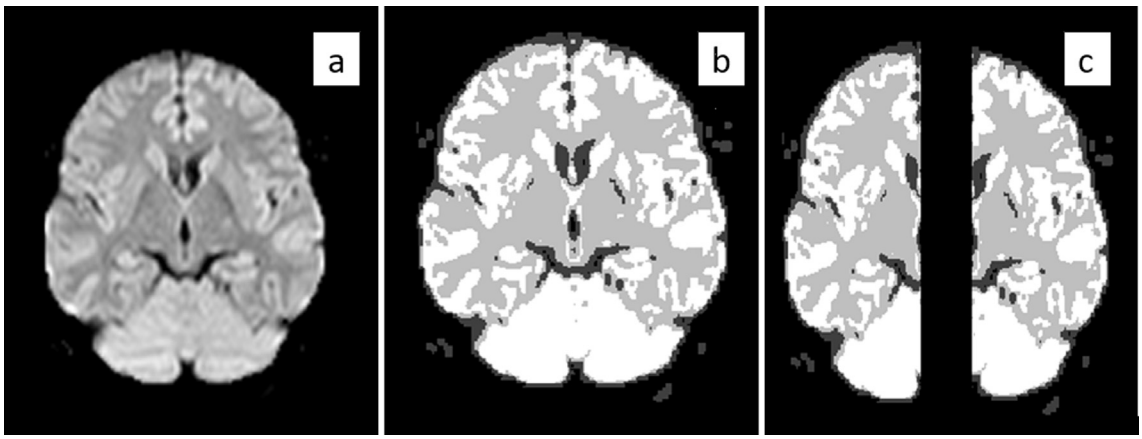


Fig. 2. DTI brain image of a healthy patient for b = 500 s/mm^2. (a) Skull stripping of the whole brain; (b) result of GMM classification; (c) hemisphere segmentation.

Table 1
GMM average mixing probability for the left hemisphere with and without diffusion gradients. The data are summarized for three mixing probabilities (w1 for GM, w2 for WM and w3 for CSF) and for three subjects H, HA and IS.

	$\langle w_1\rangle \pm s_j^H$	$\langle w_1\rangle \pm s_j^{HA}$	$\langle w_1\rangle \pm s_j^{IS}$	$\langle w_2\rangle \pm s_j^H$	$\langle w_2\rangle \pm s_j^{HA}$	$\langle w_2\rangle \pm s_j^{IS}$	$\langle w_3\rangle \pm s_j^H$	$\langle w_3\rangle \pm s_j^{HA}$	$\langle w_3\rangle \pm s_j^{IS}$
b0	0.30 ± 0.012	0.26 ± 0.008	0.22 ± 0.014	0.55 ± 0.049	0.58 ± 0.048	0.55 ± 0.064	0.14 ± 0.045	0.15 ± 0.045	0.17 ± 0.061
b250	0.32 ± 0.021	0.30 ± 0.017	0.29 ± 0.020	0.52 ± 0.046	0.52 ± 0.051	0.54 ± 0.059	0.14 ± 0.051	0.15 ± 0.051	0.15 ± 0.0.047
b500	0.32 ± 0.027	0.30 ± 0.020	0.29 ± 0.021	0.52 ± 0.056	0.54 ± 0.057	0.55 ± 0.048	0.14 ± 0.045	0.15 ± 0.047	0.15 ± 0.051
b750	0.33 ± 0.023	0.29 ± 0.023	0.29 ± 0.021	0.54 ± 0.015	0.55 ± 0.052	0.56 ± 0.053	0.13 ± 0.007	0.15 ± 0.042	0.15 ± 0.041
b1000	0.32 ± 0.021	0.29 ± 0.020	0.29 ± 0.019	0.52 ± 0.048	0.55 ± 0.053	0.55 ± 0.058	0.14 ± 0.050	0.15 ± 0.052	0.15 ± 0.048
b1250	0.33 ± 0.025	0.28 ± 0.018	0.28 ± 0.016	0.55 ± 0.059	0.55 ± 0.055	0.55 ± 0.056	0.14 ± 0.044	0.15 ± 0.050	0.15 ± 0.049

Table 2
GMM average mixing probability for the right hemisphere with and without diffusion gradients. The data are summarized for three mixing probabilities (w1 for GM, w2 for WM and w3 for CSF) and for three subjects H, HA and IS.

	$\langle w_1\rangle \pm s_j^H$	$\langle w_1\rangle \pm s_j^{HA}$	$\langle w_1\rangle \pm s_j^{IS}$	$\langle w_2\rangle \pm s_j^H$	$\langle w_2\rangle \pm s_j^{HA}$	$\langle w_2\rangle \pm s_j^{IS}$	$\langle w_3\rangle \pm s_j^H$	$\langle w_3\rangle \pm s_j^{HA}$	$\langle w_3\rangle \pm s_j^{IS}$
b0	0.32 ± 0.016	0.33 ± 0.018	0.28 ± 0.013	0.54 ± 0.041	0.53 ± 0.050	0.57 ± 0.039	0.12 ± 0.032	0.13 ± 0.041	0.14 ± 0.048
b250	0.34 ± 0.022	0.33 ± 0.041	0.30 ± 0.021	0.51 ± 0.053	0.53 ± 0.051	0.53 ± 0.059	0.14 ± 0.045	0.13 ± 0.044	0.15 ± 0.047
b500	0.34 ± 0.022	0.34 ± 0.029	0.30 ± 0.028	0.50 ± 0.050	0.51 ± 0.059	0.52 ± 0.055	0.14 ± 0.045	0.14 ± 0.044	0.15 ± 0.050
b750	0.35 ± 0.016	0.34 ± 0.029	0.28 ± 0.023	0.52 ± 0.007	0.48 ± 0.042	0.52 ± 0.041	0.13 ± 0.039	0.13 ± 0.039	0.16 ± 0.043
b1000	0.33 ± 0.018	0.33 ± 0.029	0.30 ± 0.029	0.51 ± 0.054	0.51 ± 0.060	0.53 ± 0.055	0.14 ± 0.044	0.14 ± 0.045	0.15 ± 0.047
b1250	0.34 ± 0.019	0.33 ± 0.026	0.30 ± 0.023	0.54 ± 0.053	0.51 ± 0.057	0.51 ± 0.061	0.14 ± 0.044	0.14 ± 0.045	0.15 ± 0.046

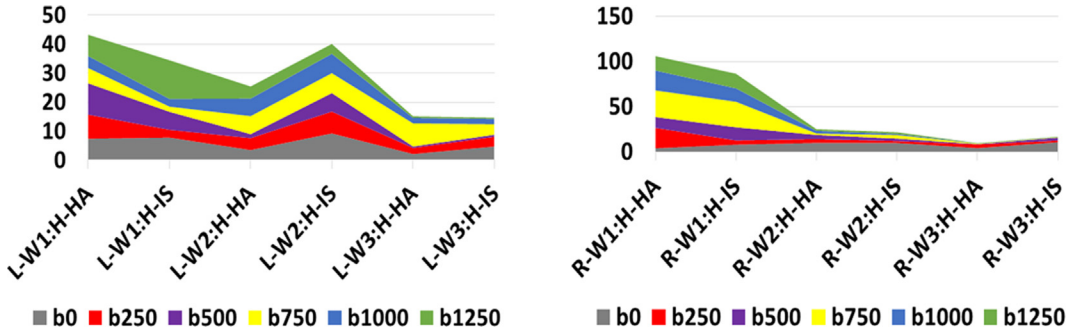


Fig. 3. Average weighted Euclidean distances for pairs of probability density function distributions of mixtures probability of GMM. Estimation is performed for all diffusion gradients and for each brain hemisphere. L denotes the left hemisphere, and R denotes the right hemisphere.

Table 3
Correlation coefficients and multiple correlation coefficients.

	Correlation coefficient									Multiple correlation coefficient		
	$r^1_{HA,H}$	$r^2_{HA,H}$	$r^3_{HA,H}$	$r^1_{IS,H}$	$r^2_{IS,H}$	$r^3_{IS,H}$	$r^1_{IS,HA}$	$r^2_{IS,HA}$	$r^3_{IS,HA}$	$R^1_{H-(IS,HA)}$	$R^2_{H-(IS,HA)}$	$R^3_{H-(IS,HA)}$
Left hemisphere	0.658	−0.421	0.214	0.654	0.515	−0.612	0.214	−0.031	0.295	0.528	0.429	0.545
Right hemisphere	0.751	−0.773	0.654	0.443	0.339	0.622	0.336	−0.214	0.345	0.714	0.699	0.564

'natural determination' is validated and performs best with respect to the hemisphere approach. The width of cluster 2 (HA subject) is not significantly high for CSF and GM in the left hemisphere. This narrow silhouette is interpreted as a spread of the point inside the cluster and as a slightly inadequate separation of the cluster.

Discussion

A different classification scheme based on the GMM for identifying the variability in the main brain tissue through a hemisphere approach (rather than by searching for tumours or lesion areas) was presented. A whole-brain imaging analysis is labour intensive and tends to be biased towards structural anatomical boundaries. Brain asymmetry analysis is a tool for analysing the neuroanatomical basis of disorders with an assumed developmental aetiology, such as dyslexia, autism and schizophrenia, in men and women [36–39]. Most studies have focused on exploring the asymmetry of the WM structure. Furthermore, these studies are mostly based on a region of interest in an image data set that is specified by users. Our results follow these observations and enlarge the applicability of the research of hemispheric specialization to the apparent difference in statistical features to reveal abnormal asymmetries of the statistical distribution of the main brain tissues. By using multiple b-values, we constructed a tool to evaluate Gaussian diffusion based on the decreased degree of diffusion-related signal attenuation with the increased b-value.

Mixture distribution models such as a GMM expresses the presence of the sub-class in a class without requiring that the sub-class of interest (observational data) be identified [11]. That is, a GMM

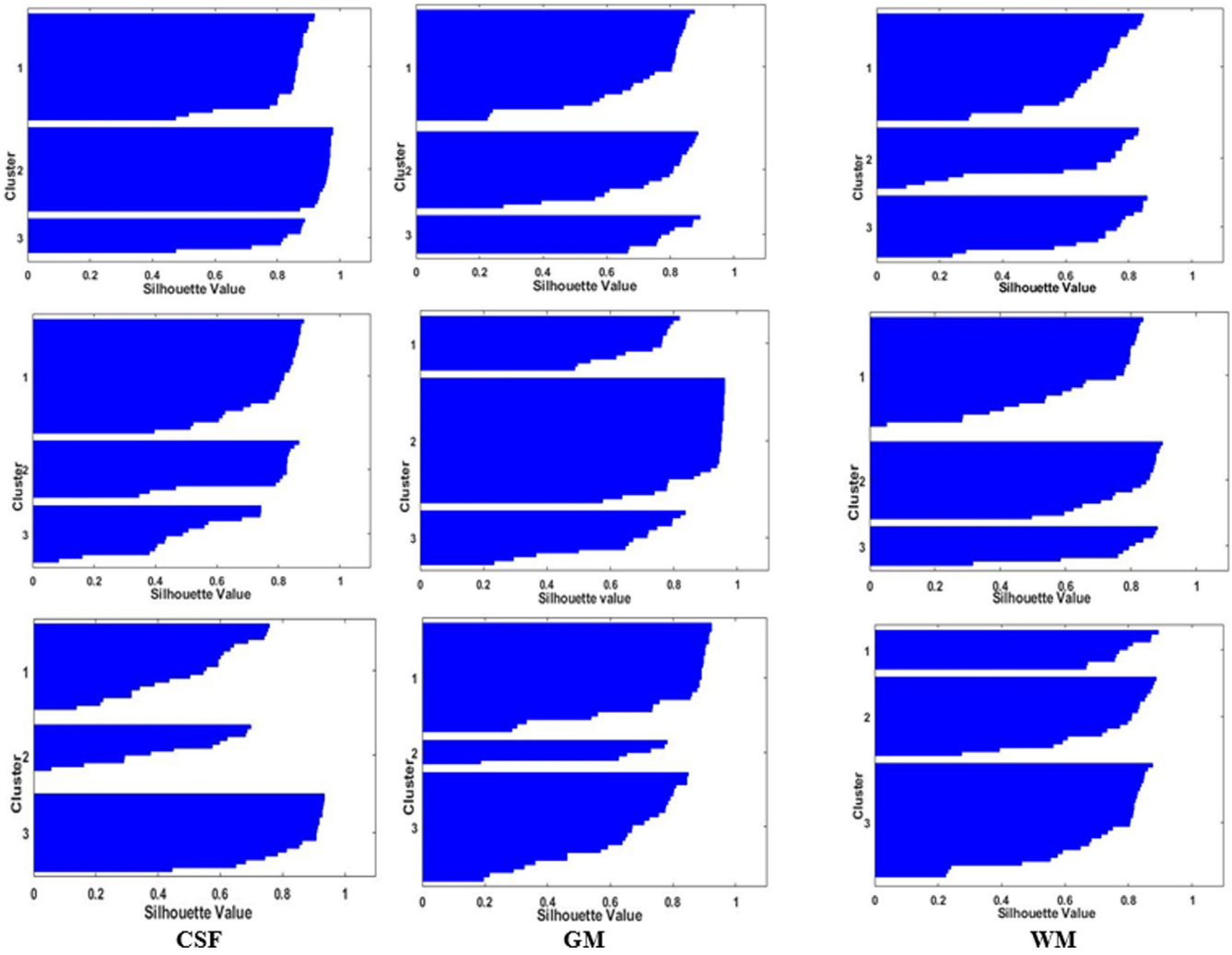


Fig. 4. Silhouettes of a data set for three clusters (line 1 on the silhouette plot corresponds to healthy subjects, line 2 for HA and line 3 for IS). Row 1: whole brain; Row 2: right hemisphere; Row 3: left hemisphere.

Table 4
Average silhouette width for evaluating clustering validity.

Class	Whole brain	Left hemisphere	Right hemisphere
H	0.9176	0.935	0.9284
HA	0.9829	0.9774	0.9326
IS	0.9989	0.8578	0.9296

expresses the probability distribution of the observational data in a class. We are focused on the three main brain tissues; thus, a GMM extracts a class's characteristic from a sub-class. As a mixture distribution model, a GMM does not seek the sub-class's information identification; since a GMM can simultaneously provide the observational data about the class, it also provides a statistical inference about the characteristic of the sub-class. Generally, a GMM requires the number of components to be specified in advance for analysing the data, i.e., inputting the number of components m (Eq. (1)) present in the mixture is necessary [40]. For ten classes of univariate distributions (including Gaussian distributions), Khalafal–Hussaini and Ahmad [41] established that all the finite mixtures generated by the family of parameters are identifiable. Chen et al. [42] reported that a finite mixture model with k components ($k = 2$ and $k \geq 3$) appears to provide consistent results in both cases ($k = 2$ and $k = 3$). Moreover, these authors claimed the absence of evidence that indicates $k \geq 4$.

The images contain multiple regions with different intensity distribution characteristics. Pixels with similar characteristics will cluster together. However, pixel classification as either CSF, GM, or WM can have a < 100% probability of belonging to a certain brain tissue. In this case, a low mixing probability can be interpreted as a possibility that a pixel has lower percentages of content of the various tissues, as data in Tables 1 and 2 showed for GM and CSF.

Specifying that the pdfs were estimated and that the mixing probabilities were computed based on the individual pixel distribution is necessary. Therefore, the highlighted differences in the probability density function originate from the particular feature of each mixing probability. The GMM analysis through the hemisphere approach evidently indicates the 'normality' of the healthy subject. There is no difference in the mixing probabilities between the left and right hemispheres for any of the classes. In contrast, a GMM with mixture probabilities tested between the left and right hemispheres for the injured subjects (both HA and IS) indicated differences and permitted the estimation of the effect of disease on the pixel distribution. A higher difference is captured for the w_2 mixing probability characteristic of WM, according to the

restricted diffusion mechanisms. The difference in the diffusion coefficient between normal WM water diffusion and diseased tissue indicates the loss of WM microstructural elements. w_1 exhibits smaller albeit visible differences. The proposed approach transcends limitations identified by Schmithorst et al. [43], which have determined developmental differences between males and females in the brain structure or various pathologies such as alcoholism [44] and schizophrenia [45].

To validate these observations, we computed the wd based on the mixing probability between pairs of subjects. One of each pair's members is a healthy subject (H), whereas the other is a patient with one of the studied diseases (HA and IS). Fig. 3 shows the results of significant variation in the wd values. An evident difference between the brain hemispheres is apparent when a paired comparison is performed between the parameters of the healthy subject and the patients with HA and IS diseases. The differentiation of mixing probabilities for the left and right hemispheres by wd provides a simple tool for assessing the variability in the main brain tissue in DTI images. To cross-verify this hypothesis, we performed a multiple correlation analysis. This analysis highlighted that HA and IS do not correlate and are the independent variables and that H is the dependent variable. The multiple correlation analysis reinforced the conclusion that for an uninjured right hemisphere, the mixing weights for HA and IS correlated well with those for H, and for the left hemisphere, the correlation weakened and indicated brain injuries. A weak correlation between the injured subjects and the control for the left hemisphere validates the variability of mixing probabilities inside the class. Furthermore, because we have a priori grouping of our data into three clusters, the silhouette plots graphically validate that the analysed 'objects' are grouped into three natural clusters. The explanation for the less wide silhouette of cluster 2 (HA subject) for the CSF and GM in the left hemisphere (Fig. 4) lies in the spread of the points inside the cluster. The data reported in Table 4 validate the initially assumed hypothesis for the three classes used in the GMM.

To our knowledge, no studies have been reported on a GMM based on pixel intensities and contrast and following a hemisphere approach to assist with brain injury diagnosis. A GMM approach with fMRI data has been proposed to explore hemispheric lateralization for language production or the human visual system in genome-wide association [46–49]. The hemispheric functional lateralization index probability density function was modelled separately for both the hemispheres using a mixture of n Gaussian components with fMRI data [47]. Furthermore, recent developments used an alternative thresholding approach based on model fit as part of mixture distribution to demonstrate that mixture modelling provides satisfactory results for the human visual system [48]. A study carried out by Kherif and Muller [49] on subjects with aphasia caused by IS demonstrated that GMMs are capable of dissociating between the sub-groups of the subject based on the main sources of variability in fMRI (i.e., handedness, sex, and age). Moreover, the authors reported that the GMM in combination with fMRI and automated lesion detection techniques is a reliable method for analysing how a normal language function is sustained notwithstanding brain injuries in the critical area. A recent study performed by Pepe et al. [50] investigated the local statistical shape analysis of gross cerebral hemispheric surface asymmetries through the brain's morphological features (i.e., surface vertices) to establish the correspondence between the hemispheric surfaces. The proposed statistical method was tested on a small sample of healthy patients and first-episode neuroleptic-naïve patients with schizophrenia.

A few limitations of the proposed approach are the following: (i) an important topic for further studies is the monotone change of relative pixel numbers with the age of the patients. A decrease of the relative number of pixels from the brain tissues as the age of the patient increases exists and age-related changes were found in the mean and variance. GMM can be affected by this finding and further examinations with more extensive age classes are required and (ii) in the current study, the optimal number of Gaussian components (m) for GMM was determined based on histogram distribution. Other methods like the Akaike Information Criterion or Bayesian Information Criterion can be used to determine this number.

Employing a GMM provides flexibility in terms of pixel spatial distributions that can be associated with a specific pathology. This method may be used to automatically detect brain microstructural differences, which exhibit statistical characteristics different from those for the same hemisphere in a normal subject, when mixing weights are considered. A major advantage is that the statistical approach over hemispheres accurately identifies the structural variations in the brain tissues by using a small number of data samples to estimate the GMM parameters. Another advantage of using the GMM for this application is that it is based on unsupervised learning; in addition, it can be rapid and to a certain extent, capable of circumventing the subsequent processing for detecting tumours or lesions. Moreover, the proposed approach is unbiased, not operator dependent and circumvents the region of interest. The main drawbacks of a DTI acquisition system (such as noise, vibration and movement artefacts) can be overcome by this multimodal approach.

Conclusions

This study, based on the asymmetries of mixture distributions between the left and right hemispheres in the human brain, can improve and more effectively assist in early diagnosis. This study is a collection of cross-sectional data samples of different subjects. The main advantage of this approach is that it is very simple, fast and can summarize the existing differences between subjects. An important source of variability in the probability density function distribution for the w_2 (associated with the WM) and w_1 (associated with the GM) mixing probabilities between the left and right hemispheres was established. The differences between the subjects in terms of mixing probabilities were also reflected by the variation in the wds. The GMM approach, mixing probabilities and wd measure represent practical and convenient tools for large-scale meta-analysis of DTI data without searching for delimitation of tumour/affected areas. Specifically, two advantages were identified. The statistical approach over the hemispheres accurately identifies structural variations in brain tissues by using a small number of data samples to estimate the GMM parameters; moreover, it is unbiased, operator independent and circumvents the region of interest.

Conflict of interest

The authors have declared no conflict of interest.

Compliance with Ethics requirements

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5). Informed consent was obtained from all patients for being included in the study.

References

[1] Thiebaut de Schotten M, Ffytche DH, Bizzi A, Dell'Acqua F, Allin MPG, Walshe M, et al. Atlasing location, asymmetry and inter-subject variability of white

matter tracts in the human brain with MR diffusion tractography. NeuroImage 2011;54(1):49–59.
[2] Kim N, Branch CA, Kim M, Lipton ML. Whole brain approaches for identification of microstructural abnormalities in individual patients: comparison of techniques applied to mild traumatic brain injury. PLoS One 2013;8(3):59382.
[3] McLachlan G, Peel D. Finite mixture models: Wiley series in probability and mathematical statistics. Wiley (NY) 2000.
[4] Banfield J, Raftery AE. Model-based Gaussian and non-Gaussian clustering. Biometrics 1993;49:803–21.
[5] Duda RO, Hart PE, Stork DG. Pattern classification. 2nd ed. NY: John Wiley & Sons; 2002.
[6] Fraley C, Raftery AE. Model-based clustering, discriminant analysis, and density estimation. J Am Stat Assoc 2002;97:611–31.
[7] Paalanen P, Kamarainen JK, Ilonen J, Kalviainen H. Feature representation and discrimination based on Gaussian mixture model probability densities-practices and algorithms. Pattern Recognit 2006;39(7):1346–58.
[8] Xia Y, Ji Z, Zhang Y. Brain MRI image segmentation based on learning local variational Gaussian mixture models. Neurocomputing 2016;204:189–97.
[9] Kim N, Heo M, Fleysher R, Branch CA, Lipton ML. A Gaussian mixture model approach for estimating and comparing the shapes of distributions of neuroimaging data: diffusion measured aging effects in brain white matter. Front Public Health 2014;2:32.
[10] Chandra S. On the mixtures of probability distributions. Scand J Stat 1977;4 (3):105–12.
[11] McLachlan G, Peel D. Finite mixture models. USA: Wiley-Interscience; 2000.
[12] Ilonen J, Kamarainen JK, Kälviäinen H, Anttalainen O. Automatic detection and recognition of hazardous chemical agents. In: Proceeding of the 14th international conference on digital signal processing. p. 1345–8.
[13] Lindh T, Ahola J, Kamarainen JK, Kyrki V, Partanen J. Bearing damage detection based on statistical discrimination of stator current. In: Proceeding of the 4th IEEE international symposium on Diagnostics for electric machines, power electronics and drives. p. 177–81.
[14] Minati L, Węglarz WP. Physical foundations, models, and methods of diffusion magnetic resonance imaging of the brain: a review. Concepts Magn Reson 2007;30A:278–307.
[15] Guo Z, Wang Y, Lei T, Fan Y. Zhang X. DTI image registration under probabilistic fiber bundles tractography learning. Biomed Res Int 2016:4674658.
[16] Balafar MA. Gaussian mixture model based segmentation methods for brain MRI images. Artif Intell Rev 2014;41:429–39.
[17] Wang J. Discriminative Gaussian mixtures for interactive image segmentation. In: Proceeding of the IEEE international conference on acoustics, speech and signal processing (ICASSP). p. 386–96.
[18] Ghosh N, Sun Y, Turenius C, Behanu B, Obenaus A, Ashwal A, et al. Computational analysis: a bridge to translational stroke treatment. In: Lapchak PA, Zhang JH, editors. Translational stroke research, from target selection to clinical trials. Switzerland: Springer; 2012. p. 884–6.
[19] Pounti O, van Leemput K. Simultaneous whole brain segmentation and white matter lesion detection using contrast-adaptive probabilistic models. In: Crimi A, Menze B, Maier O, editors. Brain lesion: glioma, multiple sclerosis, stoke and traumatic brain Injuries. Germany: Springer; 2016. p. 9–20.
[20] Han C, Zhao L, Zhong S, Wu X, Guo J, Zhuang X, et al. A comparison of high b-value vs standard b-value diffusion-weighted magnetic resonance imaging at 3.0 T for medulloblastomas. Br J Radiol 2015;88(1054):20150220.
[21] Moldovanu S, Moraru L, Biswas A. Robust skull stripping segmentation based on irrational mask for magnetic resonance brain images. J Digit Imaging 2015;28(6):738–47.
[22] Moldovanu S, Moraru L, Biswas A. Edge-based structural similarity analysis in brain MR images. JMIHI 2016;6:539–46.
[23] Bishop CM. Pattern recognition and machine learning. USA (NY): Springer; 2006.
[24] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser 1977;39(1):1–38.
[25] Redner R, Homer W. Mixture densities, maximum likelihood and the EM algorithm. SIAM 1984;26(2):195–239.
[26] Gnanadesikan R. Methods for statistical data analysis of multivariate observations. NY: Wiley Series in Probability and Statistics; 2011.
[27] Moldovanu S, Obreja C, Moraru L. Threshold selection for classification of MR brain images by clustering method. In: AIP conference proceedings 2015;1694:040005-1-7.
[28] Greenacre M. Correspondence analysis in practice. 2nd ed. USA: Chapman&Hall/CRC, Taylor and Francis Group; 2007.
[29] Greenacre M. Ordination with any dissimilarity measure: a weighted Euclidean solution. Ecology 2017;98(9):2293–300.
[30] Hershey JR, Olsen PA, Approximating the Kullback Leibler divergence between Gaussian mixture models. In: IEEE International conference on acoustics, speech and signal processing 2007;4:317–20.
[31] Durrieu JL, Thiran J, Kelly F. Lower and upper bounds for approximation of the Kullback-Leibler divergence between Gaussian mixture models. In: IEEE international conference on acoustics, speech and signal processing. p. 4833–6.
[32] Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. USA: Lawrence Erlbaun Associates; 1988.
[33] Sugar CA, James GM. Finding the number of clusters in a data set: an information theoretic approach. J Am Stat Assoc 2003;98:750–63.
[34] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.
[35] Rodríguez González D, Carpenter T, van Hemert JI, Wardlaw J. An open source toolkit for medical imaging de-identification. Eur Radiol 2010;20:1896–904.
[36] Watkins KE, Paus T, Lerch JP, Zijdenbos A, Collins DL, Neelin P, et al. Structural asymmetries in the human brain: a voxel-based statistical analysis of 142 MRI scans. Cereb Cortex 2001;11(9):868–77.
[37] Gong G, Jiang T, Zhu C, Zang Y, Wang F, Xie S, et al. Asymmetry analysis of cingulum based on scale-invariant parameterization by diffusion tensor imaging. Hum Brain Mapp 2005;24(2):92–8.
[38] Park HJ, Kubicki M, Shenton M, Guimond A, McCarley R, Maier S, et al. White matter hemisphere asymmetries in healthy subjects and in schizophrenia: a diffusion tensor MRI study. NeuroImage 2004;23(1):213–23.
[39] Büchel C, Raedler T, Sommer M, Sach M, Weiller C, Koch MA. White matter asymmetry in the human brain: a diffusion tensor MRI study. Cereb Cortex 2004;14(9):945–51.
[40] Nord CL, Valton V, Wood J, Roiser JP. Power-up: a Reanalysis of 'Power Failure' in neuroscience using mixture modeling. J Neurosci 2017;37(34):8051–61.
[41] Khalafal-Hussaini E, Ahmad KD. On the identifiability of finite mixtures of distributions. IEEE Trans Inf Theory 1981;IT-21(5):664–8.
[42] Chen H, Chen J, Kalbfleisch JD. Testing for a finite mixture model with two components. J R Stat Soc Ser B 2004;66(1):95–115.
[43] Schmithorst VJ, Holland SK, Dardzinski BJ. Developmental differences in white matter architecture between boys and girls. Hum Brain Mapp 2008;29 (6):696–710.
[44] Pfefferbaum A, Rosenbloom M, Deshmukh A, Sullivan E. Sex differences in the effects of alcohol on brain structure. Am J Psychiatry 2001;158:188–97.
[45] Highley JR, DeLisi LE, Roberts N, Webb JA, Relja M, Razi K, et al. Sex-dependent effects of schizophrenia: an MRI study of gyral folding, and cortical and white matter volume. Psychiatry Res 2003;124:11–23.
[46] Thompson WK, Wang Y, Schork AJ, Witoelar A, Zuber V, Xu S, et al. An empirical Bayes mixture model for effect size distributions in genome-wide association studies. PLoS Genet 2015;11(12):e1005717.
[47] Mazoyer B, Zago L, Jobard G, Crivello F, Joliot M, Perchey G, et al. Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness. PLoS One 2014;9(6):e101165.
[48] Bielczyk NZ, Walocha F, Ebel PW, Haak KV, Llera A, Buitelaar JK, et al. Thresholding functional connectomes by means of mixture modeling. NeuroImage 2018;171(1):402–14.
[49] Kherif F, Muller S. Early prognosis models in Aphasia. In: Toga AW, editor. Brain mapping. An encyclopedic reference. Elsevier Inc Editors; 2015. p. 807–21.
[50] Pepe A, Zhao L, Koikkalainen J, Hietala J, Ruotsalainen U, Tohka J. Automatic statistical shape analysis of cerebral asymmetry in 3D T1-weighted magnetic resonance images at vertex-level: application to neuroleptic-naïve schizophrenia. Magn Reson Imaging 2013;31:676–87.