frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 29 March 2016
doi: 10.3389/fncom.2016.00025

Positive Unanimous Voting Algorithm for Focal Cortical Dysplasia Detection on Magnetic Resonance Image

Xiaoxia Qu[1,2], Jian Yang[1], Shaodong Ma[1], Tingzhu Bai[1] and Wilfried Philips[2]*

[1] Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Electronics, Beijing Institute of Technology, Beijing, China, [2] Department of Telecommunications and Information Processing (IPI-TELIN-iMinds), Ghent University, Ghent, Belgium

Focal cortical dysplasia (FCD) is the main cause of epilepsy and can be automatically detected via magnetic resonance (MR) images. However, visual detection of lesions is time consuming and highly dependent on the doctor's personal knowledge and experience. In this paper, we propose a new framework for positive unanimous voting (PUV) to detect FCD lesions. Maps of gray matter thickness, gradient, relative intensity, and gray/white matter width are computed in the proposed framework to enhance the differences between lesional and non-lesional regions. Feature maps are further compared with the feature distributions of healthy controls to obtain feature difference maps. PUV driven by feature and feature difference maps is then applied to classify image voxels into lesion and non-lesion. The connected region analysis then refines the classification results by removing the tiny fragment regions consisting of falsely classified positive voxels. The proposed method correctly identified 8/10 patients with FCD lesions and 30/31 healthy people. Experimental results on the small FCD samples demonstrated that the proposed method can effectively reduce the number of false positives and guarantee correct detection of lesion regions compared with four single classifiers and two recent methods.

Keywords: epilepsy, focal cortical dysplasia, magnetic resonance images, lesion detection

OPEN ACCESS

Edited by:
Anke Meyer-Baese, Florida State University, USA

Reviewed by:
Dara M. Cannon, National University of Ireland, Ireland
Nia Goulden, Bangor University, UK

***Correspondence:**
Jian Yang
jyang@bit.edu.cn

Received: *16 December 2015*
Accepted: *07 March 2016*
Published: *29 March 2016*

Citation:
Qu X, Yang J, Ma S, Bai T and Philips W (2016) Positive Unanimous Voting Algorithm for Focal Cortical Dysplasia Detection on Magnetic Resonance Image. Front. Comput. Neurosci. 10:25. doi: 10.3389/fncom.2016.00025

INTRODUCTION

Focal cortical dysplasia (FCD) is the main cause of epilepsy, which is a chronic illness of human brain that affects 50–65 million people worldwide (Bernasconi and Bernasconi, 2011). FCD, as a brain malformation of neocortical development, can be eliminated by respective surgery (Despotovic et al., 2011). Neurologists use magnetic resonance (MR) imaging as a non-invasive clinical tool during surgical planning to determine the location of the FCD lesion (Antel et al., 2003). The FCD volumes can be ranged from tiny size of 734 mm^3 to large size of 80,726 mm^3 (Colliot et al., 2006). The location of the FCD lesion is completely random in its cortical boundary distribution. Different patients have lesions at different locations within the cortex which has complex gyral structure. The lesion in the MR image exhibits three features, namely, cortical thickening, blurring of the gray/white matter (GM/WM) junction, and hyper-intensity signal within lesional region compared with other cortical regions (**Figure 1**; Bernasconi et al., 2001).

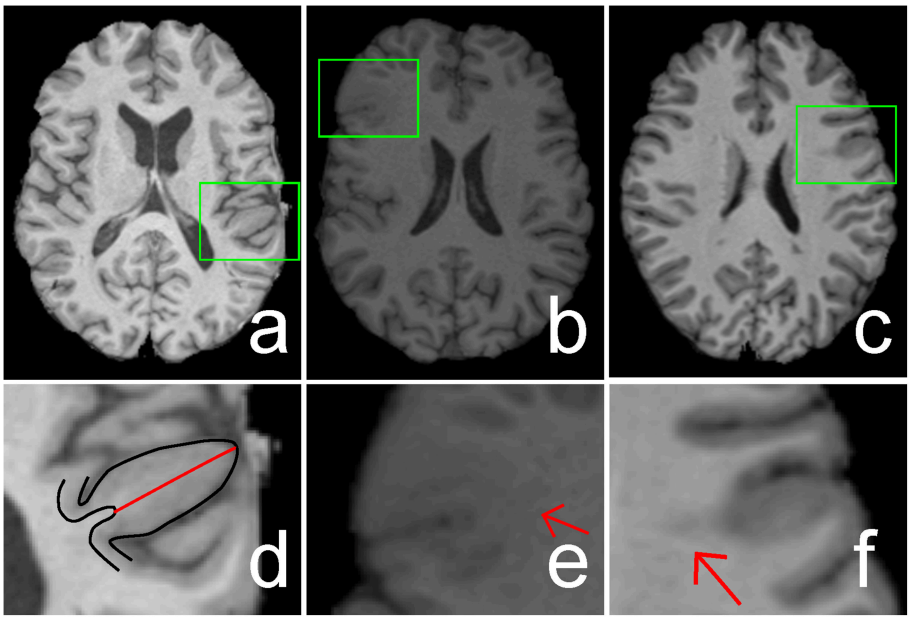

FIGURE 1 | Illustrations of cortical thickening (A), blurring of gray matter (GM)/white matter (WM) junction (B), and hyper-intensity signal (C) within focal cortical dysplasia (FCD) regions (indicated by green rectangles). The FCD regions in **(A–C)** are enlarged and shown in **(D–F)**, respectively. The black lines in image **(D)** are the boundaries of GM or cortex, whereas the length of the red line shows the increased cortical thickness. Image **(E)** shows that the GM and WM have a blurry boundary (indicated by a red arrow). Hyper-intensity signal within the WM region is indicated by a red arrow in image **(F)**.

However, an MR image consists of many slices; for example, 175 sagittal slices exist for each MR image of one patient. Thus, visual detection of lesions is time consuming. Diagnosis results are prone to present large variances when the neurologist is tired because of heavy workload from observing MR images. The image-based diagnostic method is very much subjective, because it heavily depends on the neurologist's personal experience and knowledge.

Numerous studies have attempted to develop image processing methods to facilitate the detection and localization of FCD lesions to enhance the efficiency and accuracy for diagnosis of FCD lesions on MR images. Many feature maps have been proposed to enhance the contrasts of lesional and non-lesional regions. The following maps are used to detect distinct features: GM thickness map for cortical thickening (Antel et al., 2002); gradient map for blurring of GM/WM junction (Bernasconi et al., 2001) relative intensity map for hyper-intensity signal (Bernasconi et al., 2001), mean GM density map (Kassubek et al., 2002), junction image for blurring of GM/WM junction (Huppertz et al., 2005), and complex diffusion map for blurring of GM/WM junction and cortical thickening (Rajan et al., 2009). A new feature named GM/WM boundary (GWB) width map was developed to improve the detection of blurry GM/WM junction within FCD lesional region (Qu et al., 2014). All these feature maps have increased contrasts between lesional and non-lesional regions and can be utilized to assist lesion detection.

Several classification methods, such as threshold method (Kassubek et al., 2002; Huppertz et al., 2005), Bayesian classifier (Antel et al., 2003), support vector machine (Loyek et al., 2008), and neural network classifier (Besson et al., 2008a,b), have been applied to classify image voxels or regions into positives (lesional) and negatives (non-lesional) to automatically detect the lesion. Threshold methods consider the intensity of T1 MR data as a feature, compare the intensities of patients with those of healthy controls, and classify the voxels of images into lesional and non-lesional. The Bayesian classifier, support vector machine, and neural network classifiers do not only analyze the intensity feature but also consider advanced feature maps as classification criterion, such as GM thickness and gradient maps. These classification methods can correctly classify most of the FCD regions as positive, but numerous false positive (FP) regions are also produced (Besson et al., 2008a,b). Moreover, the shapes, locations, and sizes of the FCD lesions remarkably vary in patients. Thus, automated FCD detection remains a challenge.

In this study, we developed a new framework for positive unanimous voting (PUV) to reduce the FP regions for automatic FCD detection. Unanimous voting combines multiple classifiers, which can be used to improve the performance (such as classification errors) of single classifiers (Wozniak et al., 2014). The proposed framework mainly consists of five parts, namely, image normalization, feature determination, unanimous voting for feature classification, region connection analysis, and evaluation.

Image normalization improves the image quality for subsequent study. In feature determination, maps for GM thickness, gradient, relative intensity, and GWB were considered features for distinguishing lesional and non-lesional voxels. In addition, feature distributions of healthy controls were integrated into the computation of feature differences between test images and healthy controls.

Unanimous voting involves the classification of image voxels into lesional or non-lesional. Multiple classifiers were previously combined to improve the performance of single classifiers (Wozniak et al., 2014). These combinations mainly include majority voting, voting of one against all, and unanimous voting. We selected unanimous voting with respect to positive results to combine multiple classifiers to reduce the FP. Feature value distributions of lesional and non-lesional regions were overlapped. Thus, in this study, the decision function of classification was constructed utilizing the naïve Bayesian (NB), linear discriminant analysis (LDA), Mahalanobis discriminant analysis (MDA), and quadratic discriminant analysis (QDA) classifiers (Duda et al., 2000), which were chosen as the basic classifiers for optimization.

Considering that the high similarity in features of lesional and non-lesional regions, the basic classifiers are comprised of the classifiers that calculate both the mean and the variance in each class. In addition, these classifiers are as stable in convergence as the support vector machine and neural network based methods are. A large false positive rate can cause classifiers based on support vector machine or neural network methods not be able to converge. Moreover, for this particular application the parametric estimation of the support vector machine and neural network methods can be overly complicated. Therefore, the NB, LDA, MDA, and QDA classifiers were chosen as basic classifiers for detection.

Connected region analysis was designed to remove tiny fragment regions and allow the classification to extend into the subject level. Each classified image contained voxels falsely classified as positive. These voxels formed irregular small areas called fragment regions. Connected region analysis measured the size values of these tiny fragment regions. Regions smaller than a pre-set value were relabeled as negative. In subject level classification, each subject was recognized as patient when the refined image of the subject contained voxels classified as positive (lesional). Otherwise, the subject was considered healthy.

The true positive (TP) rate (TPR), FP rate (FPR), precision, recall, and F score (F_s) were chosen as evaluation metrics to evaluate the performance of the proposed methods. The evaluation values were computed by comparing the classified images with the ground truth. Classification results of the different feature groups were compared to validate the effectiveness of the features. To demonstrate the performance of the proposed method, it was compared with single classifiers, such as NB, LDA, MDA, and QDA, in terms of the classified results. Moreover, the two-stage Bayesian classifier (TSBC; Antel et al., 2003) and surface-based LDA (SLDA; Hong et al., 2014) methods were also evaluated using the same metrics as the proposed method.

The main contributions of this work could be summarized as follows. (1) A detection framework using positive results based on unanimous voting of multiple classifiers was proposed to classify images into lesional and non-lesional. Thus, the detection results provided less FP voxels than single classifier-based methods. Experimental results clearly demonstrated that the proposed work was more effective than the single classifiers and the two state-of-the-art methods for FCD detection. (2) Mean representations of the healthy model were integrated into feature determination, which could differentiate lesional and non-lesional regions more efficiently. For example, the averaged F score resulted from different classifiers using FG6 is bigger than which using FG3 (0.071 vs. 0.052). Here, the FG6 integrates the mean representations of the healthy model, while FG3 does not. (3) Connected region analysis removed tiny FP fragments and extended the evaluation of the voxel level to the subject level.

METHODS

The framework for detecting FCD lesion on T1-weighted MR image is illustrated in **Figure 2**. The framework involved five major processing steps. (1) MR images were normalized. (2) Features were determined to compute and group features, which enhanced the differences between lesional and non-lesional regions. The feature groups were then evaluated and selected to establish the best group for further analysis. (3) Unanimous voting driven by the selected best feature group (FG_{best}) for feature classification was performed to classify the images into lesional and non-lesional regions. (4) Connected region analysis was utilized to reclassify the falsely classified tiny fragment regions into negative to refine the classified results. (5) Refined images were evaluated as the final results.

Image Normalization

Image standardization (Nyul and Udupa, 1999) was performed to refine the original raw image to a predefined scale ranging from 0 to 255. The original images were first oriented into the right-posterior-inferior direction. The brain extraction tool proposed by Smith et al. (Smith, 2002) was applied to extract the brain region on the oriented images because FCD lesions only occur in brain regions. Intensity non-uniformity (also called bias field) was corrected using the expectation–maximization algorithm (Zhang et al., 2001) on all images to obtain consistent intensities of the extracted brain region.

Each histogram of the 3D images has two peaks which correspond to gray matter and white matter regions, respectively. To insure all the images are normalized properly, we took one standardized image of healthy people as reference. When the areas under the two peaks on the histograms of the query and the reference images were almost overlapped (90% overlapped), the intensities of the query image were taken as normalized adequately.

After the intensity normalization, the space related normalization was performance. We tri-linearly interpolated images into the same resolution to standardize voxel size. We used rigid registration to roughly align the images and unify the space of the images, followed by affine registration to further standardize the images. The reference image for registration was the MNI152 brain T1-weighted MR image from the Montreal Neurological Institute (Montreal Neurological Institute). We could only align general brain regions, not match details such as gyral GM regions, because brain structures of different subjects exhibit distinct topological structures. Thus, we could scale the brain regions into similar sizes while maintaining the topological

structure and details. Consequently, differences in total brain volume between various subjects were adjusted. Brain tissues that were not anatomically related to FCD lesions (e.g., brain cerebellum, brain stem, striatum, and thalamus) were removed from the registered images. The brain atlas of images from MNI (Mazziotta et al., 2001; Diedrichsen et al., 2009) was considered the template for elimination.

Feature Determination

Feature determination was used to compute feature maps, which could enhance contrasts between lesional and non-lesional regions. Feature determination involved four steps. First, features were computed from MR images. Second, average feature representations of healthy controls were calculated. Third, feature differences between the test images and healthy controls were generated. Fourth, feature groups were selected.

The brain tissues were first segmented to acquire tissue space information before the feature maps were computed. We segmented the brain data into partial volume maps of GM, WM, and cerebrospinal fluid using a hidden Markov random field model and expectation-maximization algorithm to consider the effect of partial volume (Zhang et al., 2001).

We computed the GM thickness map (denoted by F_1), gradient map (F_2), relative intensity map (F_3; Bernasconi et al., 2001; Antel et al., 2002), and GWB width map (F_4; Qu et al., 2014) for each individual subject on the pre-processed images to capture the features of FCD regions in the MR images. These features formed a feature vector $\boldsymbol{F} = \{F_1, F_2, F_3, F_4\}$, and they are shown in **Figure 3**. Lesional regions had larger values than non-lesional regions on the GM thickness map, GWB width map, and relative intensity map. The gradient map and lesional regions with blurring of GWB had lower values than the non-lesional regions because lower values indicate blurry regions.

Average feature representations of healthy controls referred to the mean values in sliding local windows of images. Images of different subjects could be compared because the images of the subjects were registered into the same MNI space using image normalization (Section Image Normalization). The mean model of healthy controls for the i-th feature map F_i is as follows:

$$F_{\mu,i}(\boldsymbol{v}) = \frac{1}{K \cdot N} \sum_{k=1}^{k} \left(\sum_{n=1}^{N} F_i^{(k)}(\boldsymbol{v}_n) \right) \quad (1)$$

where $F_{\mu,i}(\boldsymbol{v})$ is the average value for the i-th feature map of all healthy controls at voxel $\boldsymbol{v}$, $\boldsymbol{v}_n$ is the coordinate of the n-th neighboring voxel within a 3D local window centered on voxel $\mathbf{v}$, N is the total number of neighboring voxels, $F_i^{(k)}(\boldsymbol{v}_n)$ is the feature value of the k-th healthy control at voxel $\boldsymbol{v}_n$, and K is the total number of healthy controls. The example of mean features of healthy controls is shown in **Figure 4**. The feature value distributions of healthy controls were not uniform, that is, they varied in different locations in the human brain.

Feature difference measured variations between feature $\boldsymbol{F}$ and average feature of healthy controls $F_{\mu,i}$. The procedure is illustrated in **Figure 5**. F_{d1} eliminated the influence of

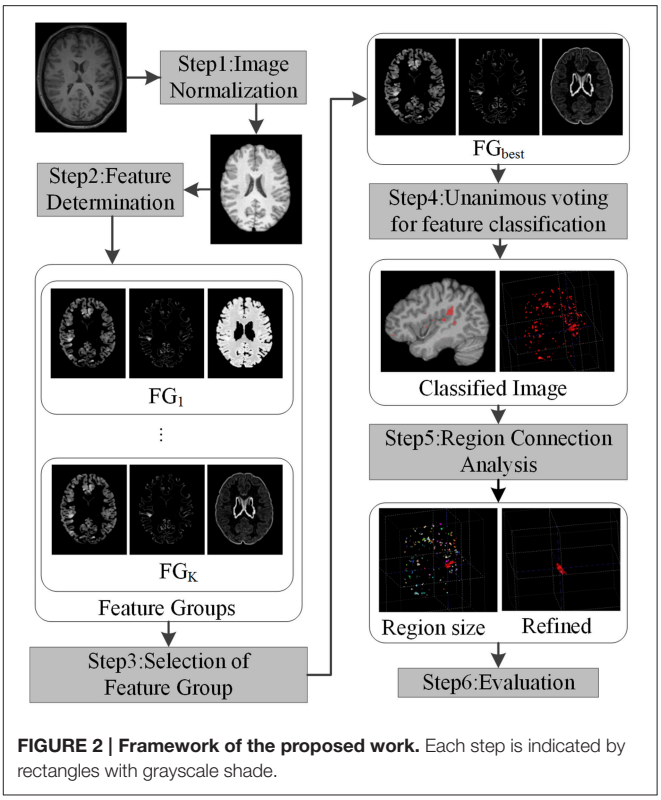

FIGURE 2 | Framework of the proposed work. Each step is indicated by rectangles with grayscale shade.

distributions on healthy controls. Thus, the feature difference of GM thickness map denoted by F_{d1} showed the increased cortical thickness better than the GM thickness F_1. The feature difference was denoted by vector $\boldsymbol{F}_d = \{F_{d1}, F_{d2}, F_{d3}, F_{d4}\}$, and it can be computed as follows:

$$F_{di}(\boldsymbol{v}) = \mathrm{F}_i(\boldsymbol{v}) - \mathrm{F}_{\mu,i}(\boldsymbol{v}) \quad (2)$$

Examples of feature differences between the test images and healthy controls were generated and are illustrated in **Figure 6**. The feature difference $\boldsymbol{F}_d$ images enhanced contrasts between lesional and non-lesional regions compared with feature $\boldsymbol{F}$.

Features and feature differences were then grouped according to their actual meanings, such as detecting cortical thickening or measuring the relative cortical thickening in contrast to healthy controls, to determine a good combination of features for classification (**Table 1**). Feature group 1 (FG_1) included three basic FCD features, namely, maps of GM thickness, gradient, and relative intensity, which had been proven to be effective for FCD detection (Antel et al., 2003). We assembled the features into FG_2 and FG_3 to evaluate the performance of the GM/WM width map with respect to FCD detection. FG_{1-3} were features in which the feature distributions of healthy models were not considered, whereas FG_{4-6} were feature differences that included the influence of healthy controls.

Several classifiers, denoted by $C_1, C_2, \ldots, C_M$, were used to classify image voxels driven by different feature groups $\mathrm{FG_k}$ to select $\mathrm{FG_{best}}$. The results of each classifier were evaluated by F_s, and the resulting E_m (m = 1, 2 …, M) is shown

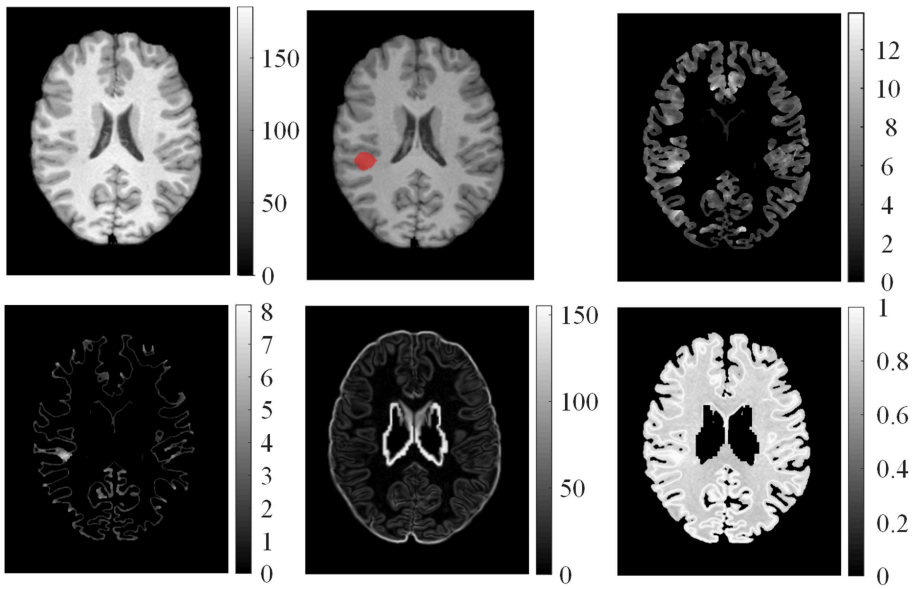

FIGURE 3 | Examples of features computed from T1 MR images. From left to right on the top row: normalized image, ground truth where the FCD lesion is labeled in red, and GM thickness map. On the bottom row: GM/WM boundary (GWB) width map, gradient map, and relative intensity map.

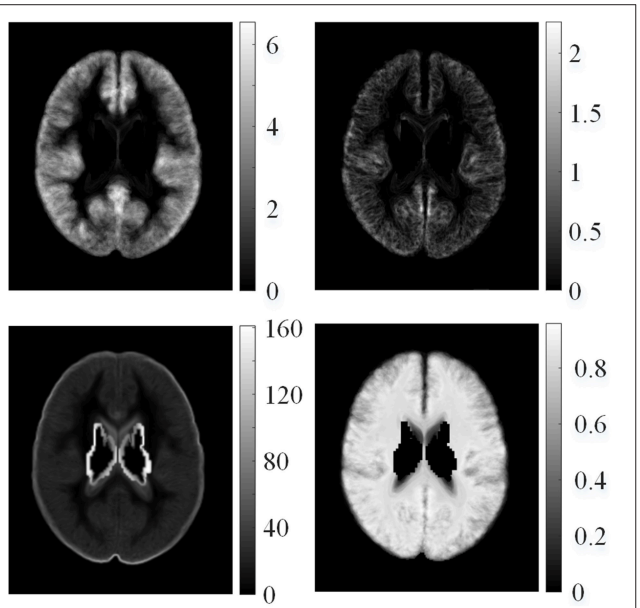

FIGURE 4 | Examples of mean features of healthy controls. From left top row, clockwise: mean GM thickness map, mean GWB width map, mean gradient map, and mean relative intensity map of healthy controls.

in **Figure 7**. Evaluation values of different classifiers for each feature group were averaged, with the resulting average value denoted by $A_{EK}\left(k = 1, 2, \ldots, K\right)$, where k is the index of the feature group. $\mathrm{FG_{best}}$ was the feature group with the highest value of $\mathrm{A_{Ek}}$. Classifiers $C_1, C_2, \ldots, C_M$ were the NB, LDA, MDA, and QDA classifiers, which were also applied for classification using unanimous voting for feature classification

TABLE 1 | Feature Groups.

	F_1	F_2	F_3	F_4		F_{d1}	F_{d2}	F_{d3}	F_{d4}
$\mathrm{FG_1}$	+	+	+	−	$\mathrm{FG_4}$	+	+	+	−
$\mathrm{FG_2}$	+	+	+	+	$\mathrm{FG_5}$	+	+	+	+
$\mathrm{FG_3}$	+	−	+	+	$\mathrm{FG_6}$	+	−	+	+

"+" means that the feature is present in the feature group, "−" indicates absence of feature. F_1, F_2, F_3, and F_4 are the maps of gray matter (GM) thickness, gradient, relative intensity, and gray/white matter boundary width map. F_{d1}, F_{d2}, F_{d3}, and F_{d4} are the corresponding feature differences of F_1, F_2, F_3 and F_4.

and will be introduced in Section Unanimous Voting for Feature Classification.

Unanimous Voting for Feature Classification

Unanimous voting for feature classification was performed to classify image voxels into lesional and non-lesional using the proposed PUV method. The basic principle of this step is explained in **Figure 8**. The features in $\mathrm{FG_{best}}$ were the inputs for this step. Image voxels were first classified by basic classifiers $C_1, C_2, \ldots, C_M$. Then, unanimous voting of the classified images was performed to reclassify voxels falsely identified as positive into negative. We attempted to obtain an object function to achieve unanimous voting for feature classification instead of classifying images on multiple stages to simplify the procedure.

Each voxel $\boldsymbol{v}$ of an image was classified into two classes, namely, positive (lesional) and negative (non-lesional), denoted by w_1 and w_2, respectively. The feature vector of each voxel is denoted by $\boldsymbol{x} = \left\{x_1, x_2, \ldots, x_f\right\}$, where $\boldsymbol{x}$ is composed of feature

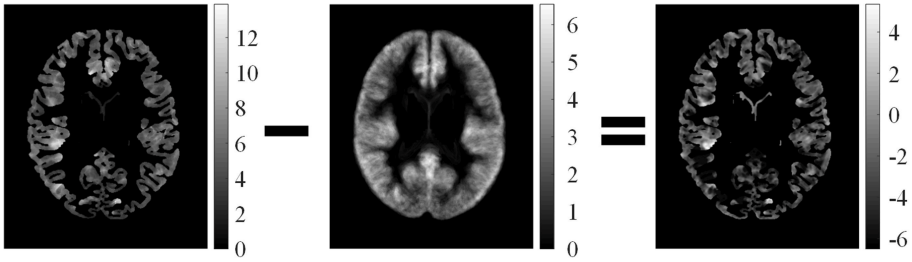

FIGURE 5 | Illustrations of computed feature differences between test images and healthy controls. From left to right: GM thickness map of one patient on axial view, mean values of GM thickness map of healthy controls, and feature difference in GM thickness map.

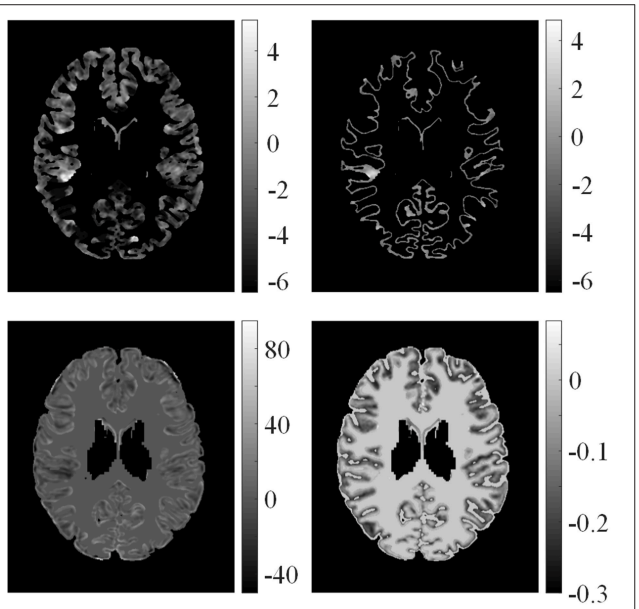

FIGURE 6 | Examples of feature differences between test images and healthy controls. From left top row, clockwise: feature differences of GM thickness map, GWB width map, gradient map, and relative intensity map from one patient.

$\boldsymbol{F}$ and feature difference $\boldsymbol{F}_d$. Each voxel was classified based on its feature vector.

The NB classifier model is expressed as follows:

$$p_{NB}\left(w_i|\,\boldsymbol{x}\right) = p_{NB}\left(\boldsymbol{x}|w_i\right) p_{NB}\left(w_i\right) / \left(\sum p_{NB}\left(\boldsymbol{x}|w_i\right) p_{NB}\left(w_i\right)\right), \tag{3}$$

where $p_{NB}\left(\boldsymbol{x}|w_i\right)$ is the likelihood of w_i with respect to $\boldsymbol{x}$, $p_{NB}\left(w_i\right)$ is the prior probability, and $p_{NB}\left(w_i|\mathrm{x}\right)$ is the posterior probability. Likelihood was computed as follows:

$$p_{NB}\left(\boldsymbol{x}|w_i\right) = \prod_{f=1}^{d}\left(2\pi\sigma_{f,w_i}\right)^{-0.5}\exp\left(-\left(\boldsymbol{x}_f - \mu_{f,w_i}\right) / \left(2\left(\sigma_{f,w_i}\right)^2\right)\right), \tag{4}$$

where μ_{f,w_i} and σ_{f,w_i} are the mean and standard deviation values of the f-th feature with respect to class w_i, respectively. Posterior probability of the *LDA* is as follows:

$$p_{LDA}\left(w_i|\boldsymbol{x}\right) = \mu_{w_i}\Sigma^{-1}\boldsymbol{x}^T - \frac{1}{2}\,\mu_{w_i}\,\Sigma^{-1}\,\mu_{w_i}^T + ln\left(p_{LDA}\left(w_i\right)\right), \tag{5}$$

whereμ_{w_i} is the mean of class w_i, and Σ is the pooled covariance matrix. The posterior probability of the *QDA* classification is as follows:

$$p_{QDA}\left(w_i|\boldsymbol{x}\right) = \mu_{w_i}\,\Sigma_{w_i}^{-1}\,\boldsymbol{x}^T - \frac{1}{2}\,\mu_{w_i}\,\Sigma_{w_i}^{-1}\,\mu_{w_i}^T + ln\left(p_{QDA}\left(w_i\right)\right), \tag{6}$$

where Σ_{w_i} is the variance matrix of class w_i. The Mahalanobis distance between sample x and class w_i in the *MDA* classifier is as follows:

$$D_{MDA}\left(w_i, \boldsymbol{x}\right) = \sqrt{\left(\boldsymbol{x} - \mu_{w_i}\right)^T\Sigma^{-1}\left(\boldsymbol{x} - \mu_{w_i}\right)}. \tag{7}$$

We reclassified the voxels identified as positive by the basic classifiers as negative or positive, because the results of basic classifiers contain a large number of FPs. The class labels $L_j\left(v\right)$were defined as 1, if $v \in w_1$. Otherwise, $L_j\left(v\right) = 0$. v is the location of the 3D voxel in the MR image, and j is the index of the classifier. L_1, L_2, L_3, and L_4 are labels of classifiers NB, LDA, QDA, and MDA, respectively. The combination of multiple classifiers using unanimous voting can be formulated as follows:

$$L_{PUV}\left(\mathbf{v}\right) = \prod_{j=1}^{N} L_j\left(\mathbf{v}\right) \tag{8}$$

where N is the total number of classifiers. When L_1, L_2,L_3 and L_4 were all labeled as 1, $L_{PUV}\left(v\right)$ was set to 1. Otherwise, $L_{PUV}\left(v\right)$ was set to 0. The final labeled $L_{PUV}\left(v\right)$ is defined as follows:

$$\begin{aligned} L_{PUV}\left(\mathbf{v}\right) = {} & sign\left(p_{NB}\left(w_1|\,\boldsymbol{x}\right)/p_{NB}\left(w_2|\boldsymbol{x}\right)\right) \\ & \bullet sign\left(p_{LDA}\left(w_1|\boldsymbol{x}\right) - p_{LDA}\left(w_2|\boldsymbol{x}\right)\right) \\ & \bullet sign\left(p_{QDA}\left(w_1|\boldsymbol{x}\right) - p_{QDA}\left(w_2|\boldsymbol{x}\right)\right) \\ & \bullet sign\left(p_{MDA}\left(w_1|\boldsymbol{x}\right) - p_{MDA}\left(w_2|\boldsymbol{x}\right)\right) \end{aligned} \tag{9}$$

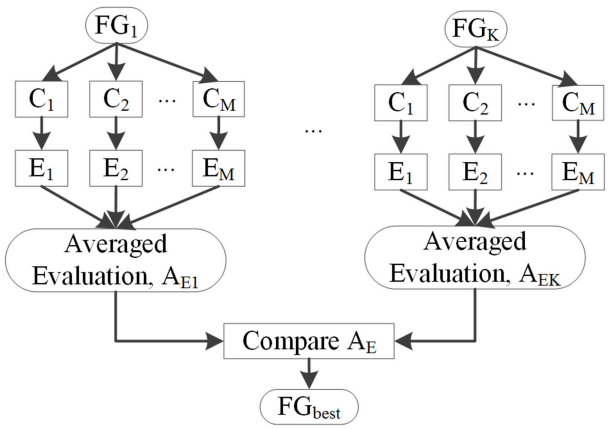

FIGURE 7 | Selection of feature groups.

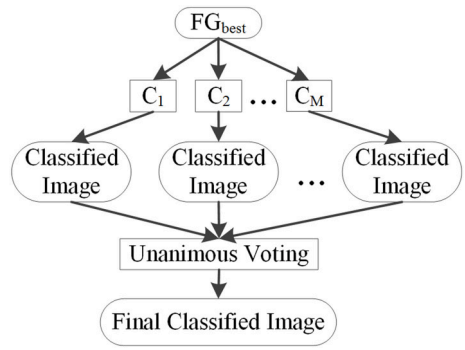

FIGURE 8 | Basic principle of unanimous voting for feature classification.

which can be calculated as follows:

$$L_{PUV}(\mathbf{v}) = sign\left(\prod_{f=1}^{d}\frac{\sigma_{f,w_2}}{\sqrt{\sigma_{f,w_1}}}\exp\left(\frac{x_f-\mu_{f,w_2}}{2\sigma_{f,w_2}^2}-\frac{x_f-\mu_{f,w_1}}{2\sigma_{f,w_1}^2}\right)\right)$$
$$\bullet sign\left(\left(\mu_{w_1}-\mu_{w_2}\right)\sum^{-1}\mathbf{x}^{\mathrm{T}}-\tfrac{1}{2}\left(\mu_{w_1}\Sigma^{-1}\mu_{w_1}^{\mathrm{T}}+\mu_{w_2}\Sigma^{-1}\mu_{w_2}^{\mathrm{T}}\right)\right)$$
$$\bullet sign\left(\mu_{w_1}\Sigma_{w_1}^{-1}\left(\boldsymbol{x}^T-\tfrac{1}{2}\mu_{w_1}^T\right)-\mu_{w_2}\Sigma_{w_2}^{-1}\left(\boldsymbol{x}^T-\tfrac{1}{2}\mu_{w_2}^T\right)\right)$$
$$\bullet sign\left(\frac{\sqrt{\left(\boldsymbol{x}-\mu_{w_2}\right)^T\Sigma^{-1}\left(\boldsymbol{x}-\mu_{w_2}\right)}/}{\sqrt{\left(\boldsymbol{x}-\mu_{w_1}\right)^T\Sigma^{-1}\left(\boldsymbol{x}-\mu_{w_1}\right)}}\right) \quad (10)$$

Equation (10) is the final classification decision of the proposed *PUV* method. Similar to Equation (8), when $L_{PUV}(v)$ was equal to 1, v was labeled as w_1. Otherwise, $L_{PUV}(v)$ was equal to 0, and v was labeled as w_2.

Region Connection

Region connection analysis was performed to refine the voxel-based classification results by removing the tiny fragmented regions. This process extended the classification to the subject level, in which each subject was recognized as either a patient or healthy person. This analysis was performed as follows. First, lesional regions in the classified images were labeled as 1, whereas healthy regions were labeled as 0. Second, the morphological opening with local window of $L_w \times L_w \times L_w$ was then used to process regions labeled as 1, such that small noise voxels were removed, and weak connected regions were separated. Third, we labeled every connected region uniquely and measured the size of each connected region by counting the number of voxels. Fourth, lesional regions smaller than a threshold T_s were relabeled as non-lesional. Otherwise, the regional label was retained. Thus, tiny fragmented regions were removed from the lesional results.

Each subject was classified as either a healthy control or a patient after the tiny fragments were removed. If a subject's image contained voxels classified as lesional, the subject was considered a patient. Otherwise, the subject was considered a healthy control. Thus, the classification was extended from the voxel level to the subject level.

Evaluation

We compared the classified images against the ground truth using the evaluation metrics, namely, TPR, FPR, and F_s, to evaluate the performance of the classification methods. Evaluations were composed of voxel- and subject-based evaluation. Voxel evaluation computes the number of correctly and incorrectly classified voxels, whereas subject evaluation calculates the number of correctly and incorrectly classified subjects. We analyzed voxel-based evaluation as an example, and the results are discussed below.

First, a TP is a correctly identified lesional voxel, whereas an FP is a non-lesional voxel incorrectly identified as a lesional voxel. A false negative (FN) is a lesional voxel incorrectly identified as a non-lesional voxel, whereas a true negative (TN) is correctly identified as a non-lesional voxel.

Second, the TPR defines how many correct positive results occur among all positive samples, that is, $\mathrm{TPR} = \#\mathrm{TP}/(\#\mathrm{TP} + \#\mathrm{TN})$ (hereafter, # means the number of the parameter cited). FPR defines the number of incorrect positive results that occur among all negative samples, that is, $\mathrm{FPR} = \#\mathrm{FP}/(\#\mathrm{FP} + \#\mathrm{TN})$.

Third, we used F_s to evaluate the trade-off between precision and recall. Precision is the probability that the positive results are TP and defined as $\mathrm{precision} = \#\mathrm{TP}/(\#\mathrm{TP} + \#\mathrm{FP})$. Recall (similar to TPR) indicates the percentage of positive regions in ground truth classified as positive. F_s can be interpreted as the weighted average of precision and recall and defined as $F_s = 2 \times \mathrm{precision} \times \mathrm{recall}/(\mathrm{precision} + \mathrm{recall})$. F_s reaches its best value at 1 and worst score at 0. A larger F_s indicates better trade-off between precision and recall.

EXPERIMENTAL RESULTS AND DISCUSSION

Experimental Data

We studied the T1-weighted MR images of 10 patients with FCD lesions and 31 healthy controls (one image per subject). Images were acquired at Ghent University Hospital on a Siemens 3T MR scanner. Each image consisted of $256 \times 256 \times 176$ voxel matrices with a resolution of $0.8594 \times 0.8594 \times 0.9$ mm. A doctor manually

delineated the FCD lesions in the images of FCD patients prior to this study. All subjects were processed to be anonymous before this study to protect their privacy. The data used in this study were extracted from a retrospective study that was approved by the local Ethics Committee of the Ghent University Hospital. The 10 patients and 31 healthy people involved in our study have provided written consent. All patients suffered from epilepsy due to FCD have been confirmed by clinical examinations.

Experimental Design

The experiments were designed to validate features and classifications of the proposed framework. The classifications driven by different feature groups were compared using F_s to assess the effectiveness of the features, and the results are shown in **Figure 9**. The proposed method was compared with the NB, LDA, MDA, and QDA classifiers in terms of F_s (**Figure 9**) and the vivid classified images (**Figure 10**) to demonstrate that the proposed method could improve the performances of single classifiers. Moreover, the TSBC (Antel et al., 2003) and SLDA (Hong et al., 2014) methods previously developed for FCD detection were also compared with the proposed method using the evaluation metrics (**Table 2**) and classified images (**Figure 11**). Classifications were constructed using the leave-one-out-cross-validation because of the limited MR images. Each test image was classified based on the classifier trained on the rest of the images in the study.

The parameters were as follows. The total number of neighboring voxels N was set to 27 when the average features of healthy controls were calculated. This value indicated that three voxels were set in each dimension of the 3D images. The local window size for morphological opening L_w was set to 3 in the connected region analysis step. The threshold of the region size was $T_s = 900$, which was equivalent to the physical size of $112.5\,\text{mm}^3 = 900 \times (0.5\,\text{mm})^3$.

Comparison of PUV and Single Classifiers

Figure 9 presents F_s of the classified images generated from different classifications. Among all feature groups, FG_6 showed the highest mean F_s, where A_{E6} of FG_6 was equal to 0.0713, whereas A_{E1-5} of FG_{1-5} ranged from 0.0406 to 0.6893. Therefore, FG_6 was FG_{best} in this study, and subsequent classifications were based on FG_{best}.

NB and PUV presented better F_s-values than other methods for FG_1, FG_2, and FG_3. This result indicated that NB and PUV were more suitable for FCD detection than LDA, QDA, and MDA using features that were not compared with the healthy controls. The F_s-values of FG_5 and FG_6 using NB and PUV were obviously larger than the other classified results. This characteristic indicated that the feature groups that contained feature difference of GWB width map (F_{d4}) could improve the performance of FCD detection with a suitable classifier, because FG_5 and FG_6 included F_{d4}, contrary to other feature groups.

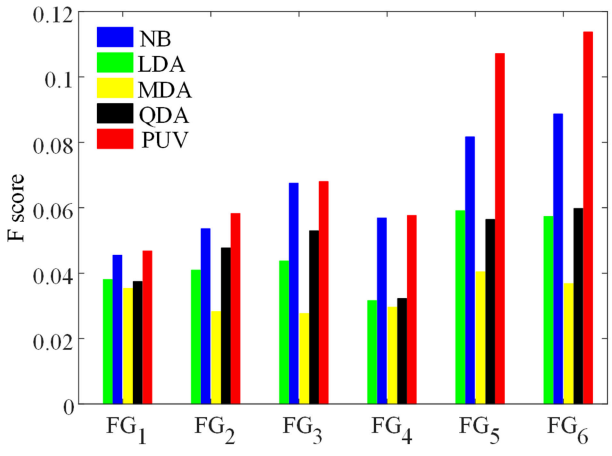

FIGURE 9 | Comparison of F-scores of the classification results from different feature groups and classification methods. The feature groups FG_{1-6} are defined in **Table 1**. FG_1 includes maps of GM thickness, gradient, and relative intensity. FG_2 includes maps in FG_1 and the GM/WM width map. FG_3 includes maps of GM thickness, relative intensity and GM/WM width. The values of FG_{4-6} are feature differences that included the influence of healthy controls, and corresponded to the FG_{1-3}, respectively.

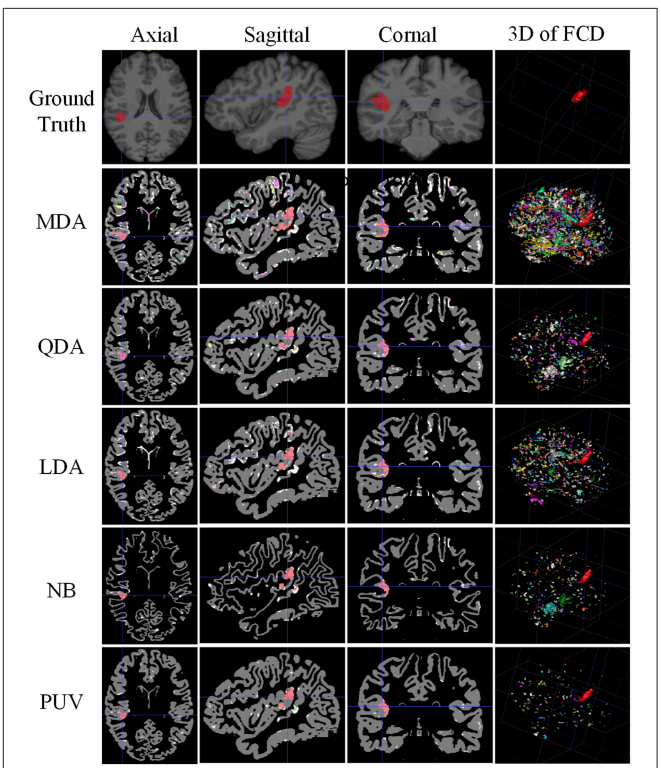

FIGURE 10 | Comparison of examples of classified images from positive unanimous voting (PUV) and single classifiers used in this study. From top to bottom, images of each row are ground truth and classified images from Mahalanobis discriminant analysis, quadratic discriminant analysis, linear discriminant analysis, naïve Bayesian, and PUV. From left to right, columns show axial, sagittal, coronal, and 3D views. For illustration purposes, the different connected regions of positive results are described by different colors on the 3D views. FCD regions are colored red.

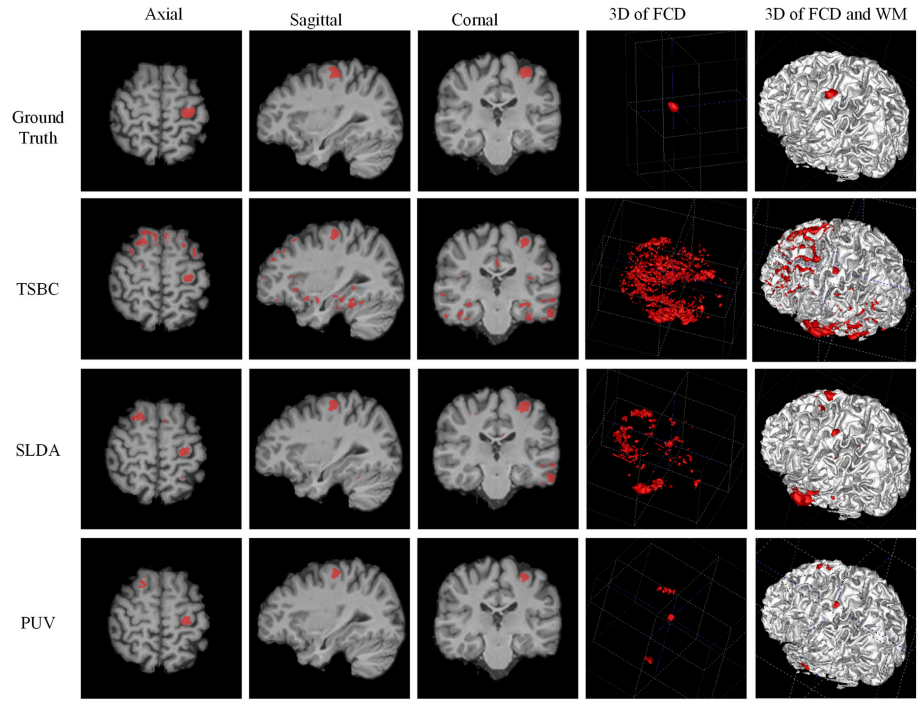

FIGURE 11 | Examples of classification results of the 3D image from patient 8 with FCD lesions. From left to right rows: axial, sagittal, coronal, and two 3D views. In the ground truth image (the first row), the FCD regions are in red. From the results of two-stage Bayesian classifiers, surface-based LDA, and PUV methods (from second row to the bottom row), the voxels and vertex identified as positive (FCD) are also in red.

Thus, PUV using the FG_6 provided the best F_s, indicating that the PUV with FG_6 exhibited better trade-off between precision and recall than other assemblies of feature groups and classifiers in this study.

Figure 10 illustrates the experimental results from MDA, QDA, LDA, NB, and PUV driven by FG_{best}. All classification methods correctly identified the true FCD as positive on the 2D images (axial, sagittal, and coronal views) compared with the ground truth, but different numbers of FP were also obtained. In addition, PUV could reduce the FP regions as shown by the fourth column in **Figure 10**, because it had the lowest number of color regions in 3D views among all classifications in this study. This phenomenon indicated that PUV facilitated the identification of the true FCD regions among positive results.

Comparison of PUV and Two Existing Methods of FCD Detection

Figure 11 demonstrates a comparison of PUV, TSBC, and SLDA. All the three methods successfully classified the FCD region as positive from the axial, sagittal, and coronal views. SLDA and TSBC produced significantly more FP results than PUV using the 3D images. This result indicated that PUV could reduce the FP results, such that the TP regions were easy to delineate from all voxels identified as positive by a classifier.

Table 2 describes the evaluation results from the F_s, TPR, and FPR-values of PUV1, PUV2, TSBC, and SLDA. PUV1 and PUV2 had lower mean TPRs than TSBC and SLDA, because the combination of multiple classifiers relabeled parts of the true lesional region as negative when reducing the FP results.

TABLE 2 | Evaluation results of positive unanimous voting (PUV), two-stage Bayesian classifier [3], and surface-based linear discriminant analysis using F-score, true positive rate, and false positive rate.

	F-score		TPR		FPR	
	Mean	***SD***	**Mean**	***SD***	**Mean**	***SD***
TSBC	0.0708	0.1131	0.3933	0.2754	0.0536	0.0270
SLDA	0.2060	0.1454	0.5374	0.2979	0.0212	0.0241
PUV1	0.1251	0.1324	0.3008	0.1657	0.0160	0.0091
PUV2	0.3039	0.1865	0.2377	0.1801	0.0211	0.0022

PUV1 and PUV2 represent the results before and after the connected region size analysis step, respectively.

PUV2 showed a larger mean F_s than TSBC and SLDA, whereas PUV1 had a larger mean F_s than TSBC but lower mean F_s than SLDA. This phenomenon suggested that the connected region analysis step could help improve the trade-off between precision and recall of FCD detection. The *SD*-values of F_s were large for all methods relative to the mean values, because the size of the lesions had larger ranges from 499 to 23,667 mm^3 ($\mu \pm \sigma$ was 4179 $\pm$ 6598) in this study. For example, the lesional size of patient number 10 was 23,667 mm^3, which was large. The

Fs-values of this patient were 0.38 for TSBC, 0.45 for SLDA, and 0.69 for PUV2. By contrast, the lesional size of patient number 3 was 1754 mm^3, which was small. The *Fs*-values of this patient were 0.04 for TSBC, 0.06 for SLDA, and 0.28 for PUV2.

PUV2 provided lower mean FPR and considerably lower *SD*-values of the FPR than TSBC. This result suggested that the proposed framework had lower and more stable probability of falsely classifying non-lesional regions as positives than TSBC. PUV2 provided similar mean FPR but smaller *SD*-values of the FPR compared with SLDA. Thus, PUV2 presented comparable ability of correctly classifying non-lesional regions as SLDA, but the ability of PUV2 was remarkably more stable than that of SLDA.

Results of Connected Region Analysis and Evaluation on Subject Level

Connected region analysis extended the classification results from the voxel level to the subject level (**Figure 12**). Each subject, even the healthy controls (black triangles), had voxels falsely classified as lesional. However, most true lesional regions were larger than the falsely classified regions in healthy controls (red dots vs. black triangles), confirming that relabeling the small positive regions as negative was reasonable. After relabeling all small/tiny positive regions smaller than T_s (size of 900 voxels or 112.5 mm^3) as negative, eight out of 10 patients were correctly identified as patients, whereas 30 of 31 healthy controls were correctly classified as healthy. Thus, TPR was 80% and FPR was 3.3% on the subject level, indicating that the proposed framework produced promising results for FCD detection.

Limitations

This study was limited by four factors. First, region size analysis possibly split the TP regions. For example, two red dots are shown in **Figure 12** when the subject index was 6, which indicated that the true lesional region of patient number 6 was split into two clusters. In future studies, methods of splitting and merging of regions might address this issue. Second, this study used a limited number of single classifiers. Including more effective single classifiers might also further improve automatic FCD detection. Third, the proposed framework was affected by registration accuracy, because the feature difference step in the proposed framework compared images from different subjects and the volumes of images remarkably varied in different subjects. PUV and SLDA both depended on registration accuracy, contrary to TSBC. However, PUV and SLDA produced better performance for FCD detection. Thus, comparing registered images from different subjects was necessary, although the comparative results were dependent on registration. Fourth, the detection performance is affected by the intensity normalization, since the feature computation are related to the intensity values. When the images are obtained from different scanners, the intensity scale should be carefully normalized and the intensity non-uniformity has to be carefully adjusted. Using good evaluation methods to guarantee the intensities are properly corrected, might further improve the detection performance.

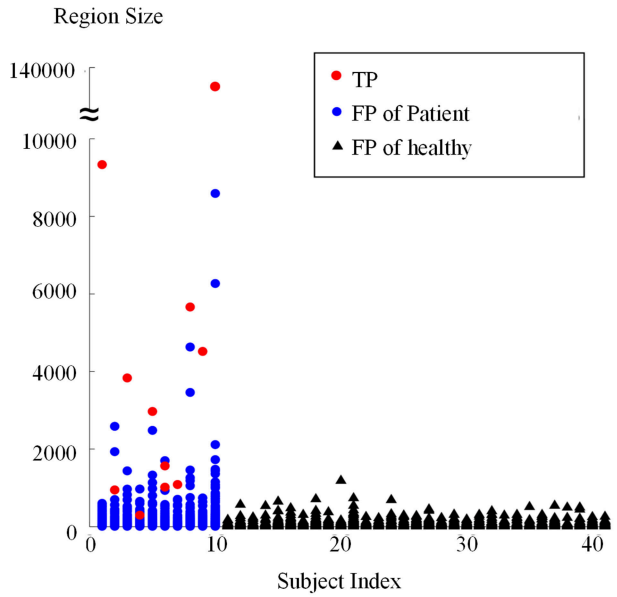

FIGURE 12 | Sizes of regions classified as positive from patients and healthy controls using connected region analysis. The red dots are the region size values of the true FCD lesions, blue dots are the region sizes of FP regions in patients, and black triangles are the region sizes of FP regions of healthy controls.

CONCLUSION

In this study, we proposed a new framework based on PUV to achieve automated detection of FCD lesions. Feature determination of the proposed method enhanced the contrasts between lesional and non-lesional regions. Selection of feature group found FG_{best} in this study. Unanimous voting for feature classification categorized image voxels into positives or negatives. Finally, the connected region refined the classified images through relabeling the small fragment region as negative.

Experiments on the small number of the FCD samples reveal that the proposed framework produced higher F_s-values and lesser number of FP regions compared with single classifier techniques, such as NB, LDA, QDA, and MDA, facilitating the easy identification of true FCD lesions. The proposed framework had lower and more stable probability of falsely classifying non-lesional regions as positives compared with TSBC. The proposed framework presented comparable ability of correctly classifying the non-lesional regions compared with SLDA. However, the ability of the proposed work was much more stable. After tiny fragments were relabeled as negative using connected region analysis, 8/10 patients and 30/31 healthy controls were correctly recognized.

The intensity scale standardization and intensity non-uniformity correction might affect the generalizability of the proposed method when the images are over scanners with variable intensity bias patterns. In the future, to apply the proposed method on more data sets that from different

scanners, intensity scale standardization and intensity non-uniformity correction should be cautiously operated before tissue segmentation and feature computation. The reason is that two limitations of the proposed work are non-ignorable: (1) the method is tested on the small number of FCD samples; (2) the images are from same scanner. The intensity scale standardization and intensity non-uniformity correction on the data sets from different canners, might change the intensities of different brain tissue. The tissue segmentation and feature computation highly depend on the intensities of tissues, and directly affect the detection performance. Therefore, the proposed method might not provide detection results as good as this study when more data sets from different canners are considered, due to the preprocessing in terms of intensity scale standardization and intensity non-uniformity correction may change the information of tissues.

Based on the experiments in this study, the proposed work was simple, easy to reproduce, and may become a useful tool to assist doctors in detecting and diagnosing FCD lesions.

AUTHOR CONTRIBUTIONS

XQ performed the experiments, analyzed the data sets and wrote the manuscript. JY designed the work and finalized this manuscript. SM modified the manuscript and checked all the figures and texts. WP and TB designed the topic of the work and provided the framework of the topic.

ACKNOWLEDGMENTS

This work was financially supported by the National Basic Research Program of China (2013CB328806), the Key Projects in the National Science and Technology Pillar Program (2013BAI01B01), National Hi-Tech Research and Development Program (2015AA043203), the China Scholarship Council (CSC), and the co-funding of the CSC from Ghent University. We thank doctor Karel Deblaere from Ghent University Hospital scanned the patients and healthy people using the MRI machine, and delineated the lesional regions as a ground truth for evaluating the experimental results.

REFERENCES

Antel, S. B., Bernasconi, A., Bernasconi, N., Collins, D. L., Kearney, R. E., Shinghal, R., et al. (2002). Computational models of MRI characteristics of focal cortical dysplasia improve lesion detection. *Neuroimage* 17, 1755–1760. doi: 10.1006/nimg.2002.1312

Antel, S. B., Collins, D. L., Bernasconi, N., Andermann, F., Shinghal, R., Kearney, R. E., et al. (2003). Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis. *Neuroimage* 19, 1748–1759. doi: 10.1016/S1053-8119(03)00226-X

Bernasconi, A., Antel, S. B., Collins, D. L., Bernasconi, N., Olivier, A., Dubeau, F., et al. (2001). Texture analysis and morphological processing of magnetic resonance imaging assist detection of focal cortical dysplasia in extra-temporal partial epilepsy. *Ann. Neurol.* 49, 770–775. doi: 10.1002/ana.1013

Bernasconi, A., and Bernasconi, N. (2011). Advances in MRI for 'cryptogenic' epilepsies. *Nat. Rev. Neurol.* 7, 99–108. doi: 10.1038/nrneurol.2010.199

Besson, P., Bernasconi, N., Colliot, O., Evans, A., and Bernasconi, A. (2008a). Surface-based texture and morphological analysis detects subtle cortical dysplasia. *Med. Image Comput. Comput. Assist. Interv.* 11, 645–652. doi: 10.1007/978-3-540-85988-8_77

Besson, P., Colliot, O., Evans, A., and Bernasconi, A. (2008b). "Automatic detection of subtle focal cortical dysplasia using surface-based features on MRI," in *The 5th IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro* (Paris).

Colliot, O., Antel, S. B., Naessens, V. B., Bernasconi, N., and Bernasconi, A. (2006). *In vivo* profiling of focal cortical dysplasia on high-resolution MRI with computational models. *Epilepsia* 47, 134–142. doi: 10.1111/j.1528-1167.2006.00379.x

Despotovic, I., Segers, I., Platisa, L., Vansteenkiste, E., Pizurica, A., Deblaere, K., et al. (2011). Automatic 3D graph cuts for brain cortex segmentation in patients with focal cortical dysplasia. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2011, 7981–7984. doi: 10.1109/IEMBS.2011.6091968

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., and Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *Neuroimage* 46, 39–46. doi: 10.1016/j.neuroimage.2009.01.045

Duda, R. O., Hart, P. E., and Stork, D. G. (2000). *Pattern Classification, 2nd Edn*. New York, NY: John Wiley & sons, Inc.

Hong, S. J., Kim, H., Schrader, D., Bernasconi, N., Bernhardt, B. C., and Bernasconi, A. (2014). Automated detection of cortical dysplasia type II in MRI-negative epilepsy. *Neurology* 83, 48–55. doi: 10.1212/WNL.0000000000000543

Huppertz, H., Grimm, C., Fauser, S., Kassubek, J., Mader, I., Hochmuth, A., et al. (2005). Enhanced visualization of bluffed gray-white matter junctions in focal cortical dysplasia by voxel-based 3D MRI analysis. *Epilepsy Res.* 67, 35–50. doi: 10.1016/j.eplepsyres.2005.07.009

Kassubek, J., Huppertz, H., Spreer, J., and Schulze-Bonhage, A. (2002). Detection and localization of focal cortical dysplasia by voxel-based 3D MRI analysis. *Epilepsia* 43, 596–602. doi: 10.1046/j.1528-1157.2002.41401.x

Loyek, C., Woermann, F. G., and Nattkemper, T. W. (2008). *Detection of Focal Cortical Dysplasia Lesions in MRI Using Textural Features*. Berlin; Heidelberg: Springer.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., et al. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1293–1322. doi: 10.1098/rstb.2001.0915

Nyul, L. G., and Udupa, J. K. (1999). On standardizing the MR image intensity scale. *Magn. Reson. Med.* 42, 1072–1081.

Qu, X., Platisa, L., Despotovic, I., Kumcu, A., Bai, T., Deblaere, K., et al. (2014). Estimating blur at the brain gray-white matter boundary for FCD detection in MRI. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2014, 3321–3324. doi: 10.1109/EMBC.2014.6944333

Rajan, J., Kannan, K., Kesavadas, C., and Thomas, B. (2009). Focal Cortical Dysplasia (FCD) lesion analysis with complex diffusion approach. *Comput. Med. Imaging Graph.* 33, 553–558. doi: 10.1016/j.compmedimag.2009.05.004

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062

Wozniak, M., Grana, M., and Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Inform. Fusion* 16, 3–17. doi: 10.1016/j.inffus.2013.04.006

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20, 45–57. doi: 10.1109/42.906424

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Qu, Yang, Ma, Bai and Philips. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.