Cortical feature analysis and machine learning improves detection of “MRI-negative” focal cortical dysplasia

Bilal Ahmed, Carla E. Brodley, Karen E. Blackmon, Ruben Kuzniecky, Gilad Barash, Chad Carlson, Brian T. Quinn, Werner Doyle, Jacqueline French, Orrin Devinsky, Thomas Thesen

1. Introduction

Despite advances in pharmacotherapy for the treatment of epilepsy, approximately one-third of patients remain to have seizures refractory to medications [1]. For patients with treatment-resistant epilepsy (TRE), the best option for achieving seizure freedom is often surgical resection [2]. Advances in MRI have revolutionized the diagnosis of FCD, resulting in higher success rates for resective epilepsy surgery. However, many patients with histologically confirmed FCD have normal presurgical MRI studies (‘MRI-negative’), making presurgical diagnosis difficult. The purpose of this study was to test whether a novel MRI post-processing method successfully detects histopathologically verified FCD in a sample of patients without visually appreciable lesions. We applied an automated quantitative morphometry approach which computed five surface-based MRI features and combined them in a machine learning model to classify lesional and nonlesional vertices. Accuracy was defined by classifying contiguous vertices as “lesional” when they fell within the surgical resection region. Our multivariate method correctly detected the lesion in 6 of 7 MRI-positive patients, which is comparable with the detection rates that have been reported in univariate vertex-based morphometry studies. More significantly, in patients that were MRI-negative, machine learning correctly identified 14 of 24 FCD lesions (58%). This was achieved after separating abnormal thickness and thinness into distinct classifiers, as well as separating sulcal and gyral regions. Results demonstrate that MRI-negative images contain sufficient information to aid in the in vivo detection of visually elusive FCD lesions.

Focal cortical dysplasia (FCD) is the most common cause of pediatric epilepsy and the third most common lesion in adults with treatment-resistant epilepsy. Advances in MRI have revolutionized the diagnosis of FCD, resulting in higher success rates for resective epilepsy surgery. However, many patients with histologically confirmed FCD have normal presurgical MRI studies (‘MRI-negative’), making presurgical diagnosis difficult. The purpose of this study was to test whether a novel MRI post-processing method successfully detects histopathologically verified FCD in a sample of patients without visually appreciable lesions. We applied an automated quantitative morphometry approach which computed five surface-based MRI features and combined them in a machine learning model to classify lesional and nonlesional vertices. Accuracy was defined by classifying contiguous vertices as “lesional” when they fell within the surgical resection region. Our multivariate method correctly detected the lesion in 6 of 7 MRI-positive patients, which is comparable with the detection rates that have been reported in univariate vertex-based morphometry studies. More significantly, in patients that were MRI-negative, machine learning correctly identified 14 of 24 FCD lesions (58%). This was achieved after separating abnormal thickness and thinness into distinct classifiers, as well as separating sulcal and gyral regions. Results demonstrate that MRI-negative images contain sufficient information to aid in the in vivo detection of visually elusive FCD lesions.
algorithms to detect FCD lesions in patients classified as MRI-negative following conventional radiological analysis of scans acquired through a standard epilepsy protocol. The novelty of our approach is that it utilizes specific strategies to model the biological features of FCD lesions. For example, we train separate classifiers on abnormally thick versus abnormally thin lesional regions to model these features separately, which can vary by FCD lesion subtype [7]. Additionally, we train separate classifiers for the gyral wall, sulcus, and crown to optimize detection of bottom-of-the-sulcus lesions [17].

2. Materials and methods

2.1. Participants

Participants were selected from a large registry of patients with epilepsy treated at the New York University School of Medicine Comprehensive Epilepsy Center who signed consent for a research MRI scanning protocol. Criteria for inclusion in this study included the following: (1) completion of a high resolution T1-weighted MRI scan, (2) surgical resection to treat focal epilepsy, and (3) diagnosis of FCD on neuropathological examination of the resected tissue. These selection criteria resulted in a sample of 31 patients with FCD. Demographic- and seizure-related information of both MRI-positive and MRI-negative patients is presented in Table 1. In addition, MRI scans using identical imaging parameters from a total of 62 neurotypical controls were acquired (31 females; ages 17–65; mean age = 33; SD = 12.5). Exclusion criteria for the control group included any history of psychiatric or neurological disorders.

2.2. Image acquisition

2.2.1. Imaging for research

Imaging for the research protocol was performed at the New York University Center for Brain Imaging on a Siemens Allegra 3T scanner. Clinical and research sequences included: (i) fluid-attenuated inversion recovery (FLAIR) images (2 mm slice thickness), (ii) magnetization-prepared rapid gradient echo (MPRAGE) images (2–6 mm slice thickness). The research T1-weighted MPRAGE images used in our analyses were included in the set of images reviewed by the clinical radiology team. Conventional visual analysis was followed by a registration process that involved morphing the reconstructed, it was further refined by classifying all white matter vertices in the MRI volume to create the gray/white matter boundary. The gray/white matter junction was delineated up to submillimeter accuracy by further refining the gray/white matter surface. After refining the gray/white matter junction, the pial surface was located by deforming the surface outward. Each segmentation and reconstruction underwent manual inspection and editing, when necessary. However, the high image quality and gray–white contrast in the initial images resulted in minimal editing requirements for both patient and control scans. Surface reconstruction was followed by a registration process that involved morphing the reconstructed surface to an average spherical representation that accurately matched sulcal and gyral features across individual subjects while minimizing metric distortion [19].

2.4. Morphometric feature extraction

Five cortical features were computed at each vertex. These included (i) cortical thickness, (ii) gray/white matter contrast, (iii) sulcal depth, (iv) mean curvature, and (v) Jacobian distortion.

(ii) tessellation of the gray/white matter boundary, (iii) inflation of the folded surface, and (iv) correction of topological defects. Once the surface was reconstructed, it was further refined by classifying all white matter vertices in the MRI volume to create the gray/white matter boundary. The gray/white matter junction was delineated up to submillimeter accuracy by further refining the gray/white matter surface. After refining the gray/white matter junction, the pial surface was located by deforming the surface outward. Each segmentation and reconstruction underwent manual inspection and editing, when necessary. However, the high image quality and gray–white contrast in the initial images resulted in minimal editing requirements for both patient and control scans. Surface reconstruction was followed by a registration process that involved morphing the reconstructed surface to an average spherical representation that accurately matched sulcal and gyral features across individual subjects while minimizing metric distortion [19].

2.2.2. Clinical imaging

Clinical imaging sequences for radiological review were acquired at the NYU Department of Radiology on a 3-Tesla Siemens scanner. Clinical sequences were variable across patients but commonly included high-resolution T1-weighted MPRAGE (magnetization-prepared rapid gradient echo) images, T2-weighted images (axial and coronal, varying slice thickness from 1 to 3 mm), and fluid-attenuated inversion recovery (FLAIR) images (2–6 mm slice thickness). The research T1-weighted MPRAGE images used in our analyses were included in the set of images reviewed by the clinical radiology team. Conventional visual analysis of the clinical scans resulted in an MRI diagnosis of FCD in 7 patients (MRI-positive) and a “normal” report in 24 patients (MRI-negative). The higher number of MRI-negative patients in this sample may be due to a tendency for patients with more complex, MRI-negative epilepsy to be referred to our level 4 epilepsy treatment center.

2.3. Surface reconstruction

The research MRI sequences were processed using the FreeSurfer software package (http://surfer.nmr.mgh.harvard.edu/), which performs automated tissue segmentation to recreate 3D representations of the cortical surfaces from structural MRI scans [18]. Briefly, after skull stripping, the method [18] involves (i) segmentation of the white matter,
2.5. Normalization of parameters

In preparing the data for the machine learning classifier, the cortical features from each patient are z-score normalized using the mean and standard deviation calculated from the control population, on a vertex-by-vertex basis.

2.6. Lesion and resection tracing

For MRI-positive patients, an expert on epileptogenic malformations on cortical development and who is board-certified in neurology and neurophysiology (RK), reviewed the clinical MRI report and manually traced the outer regions of the visible lesions on the morphometric T1-weighted 3D volume scan based on the lesional areas identified in the initial clinical report or during surgical conference. When available, the visual detection was aided by T2-weighted FLAIR images from the standard clinical epilepsy MRI protocol. For MRI-negative patients, the postoperative T1-weighted image (with the resection area removed) was rigid-body coregistered to the (intact) preoperative T1-weighted image using FLIRT[25]. The brain resection area was manually traced on the postsurgical MRI scan by a trained technician blinded to patient self. For example, labeling “lesional” vertices in MRI-positive cases involves subjective tracing of the FCD lesion. Moreover, in the absence of an MRI-visible lesion, lesional vertices are delineated by the extent of the tissue removed in surgery, which may include a gradation from abnormal to normal tissue. From a supervised machine learning perspective, treating all the resected vertices in the case of MRI-negative patients as being lesion introduces substantial false positives into the training data, which can have adverse effects on classifier accuracy[30].

The fact that the resection zones in MRI-negative patients include both lesional tissue and nonlesional tissue is problematic for training classifiers. Hong et al.[16] address this problem by utilizing a preprocessing step that not only includes the generation of texture maps[31,32] but also requires human expertise and intervention to visually identify and trace lesions. In our approach, we used cortical thickness to reduce the impact of false-positive label noise. Cortical thickness is the most prominent feature on T1-weighted imaging in FCD[7,26,33]. We, therefore, trained the classifier on the vertices inside the resection zone that showed the highest degree of thickness abnormality both in terms of thickening and thinning.

Normal tissue classification was performed on data from control subjects in order to control false negatives in the labeled data that can arise because the lifetime seizure burden of a given patient can lead to cortical abnormalities outside the seizure onset zone[34–37] or the possibility of additional nonepileptogenic dysplastic lesions[38]. In addition, patients who are suffering from epilepsy due to developmental factors may have additional lesions that either are not epileptogenic or have latent epileptogenicity. Based on these considerations, we chose not to include nonlesional vertices from the subjects as negative instances in our training data.

2.8. Machine learning classification

Machine learning algorithms are ideally suited for dysplasia detection in that they can incorporate multiple quantitative MRI measures, making maximum use of all relevant data available. The goal of the machine learning classification model was to accurately differentiate contiguous clusters of lesional vertices from nonlesional vertices in a single patient. Accuracy was defined by classifying contiguous vertices as “lesional” when they fell within the manually traced lesion or resection region for MRI-positive and MRI-negative patients, respectively, and “nonlesional” when they fell outside of these regions.

Designing an appropriate classification scheme for detecting FCD under these constraints has three important challenges. First, class label noise arises from subjectivity in delineating the lesion zone (either a manually traced MRI-visible zone or resection-defined zone). Second, the anatomic complexity and heterogeneity in folded cortical tissue reduce the ability to discern lesional tissue from the normal cortex, which is one of the reasons why a large number of lesions remain elusive to human perception in routine radiological evaluation[28]. Third, class imbalance[29] results from a ratio of substantially fewer lesional to nonlesional vertices for a particular patient. The class imbalance problem is further compounded by the higher availability of healthy control data compared to patient data. We address each of these challenges in turn.

2.8.1. Addressing class label noise

Optimizing classifiers for detecting FCD lesions relies on accurately labeling vertices as “lesional” or “nonlesional” in the training data. Class label noise can arise from errors in human decision-making and subjectivity, in addition to the anatomical complexity of the brain itself. For example, labeling “lesional” vertices in MRI-positive cases involves subjective tracing of the FCD lesion. Moreover, in the absence of an MRI-visible lesion, lesional vertices are delineated by the extent of the tissue removed in surgery, which may include a gradation from abnormal to normal tissue. From a supervised machine learning perspective, treating all the resected vertices in the case of MRI-negative patients as being lesion introduces substantial false positives into the training data, which can have adverse effects on classifier accuracy[30].
2.8.2. Reducing cortical complexity

Anatomical complexity of the cortical convolution may account for why many lesions remain undetected in radiological MRI evaluations. The folding of the cortex varies across individuals, and it can hinder the visibility of subtle FCD lesions that may be hidden deep within the folds. Recent studies have shown that subtle FCD lesions occur with higher frequency at the bottom of the sulcus. Given these observations, we designed a stratified classification scheme composed of different classifiers that were trained separately for sulcal, wall, and gyral regions. We separated the data into three nonoverlapping levels where (i) sulcal depth in the range [−2, −1] represents vertices that are part of the sulcus, (ii) [1, 2] represents vertices residing on the gyrus, and (iii) the vertices in between (i.e., with a sulcal depth of [−1, 1]) were labeled as wall vertices. Partitioning the vertices into these three groups meant that we needed to calculate the two thresholds for mitigating label noise per sulcal level, which resulted in a total of six distinct thresholds (i.e., 1. thin/sulcus, 2. thick/sulcus, 3. thin/gyrus, 4. thick/gyrus, 5. thin/wall, and 6. thick/wall).

In other words, for each sulcus level X, we trained two separate classifiers, which differed in how the training data for lesional vertices were collected. This problem arises out of having substantially fewer vertices labeled as “lesional” than vertices labeled as “nonlesional.” Such an imbalance in training data can result in classifiers that are biased towards the majority class [29]. To address this issue, we used a “bagging” approach [39]. We construct a set of “base-level” classifiers, each trained using logistic regression, using an iterative-reweighted least squares (IRLS) algorithm [40]. Each base-level classifier is trained on all the minority class instances (lesional vertices) and an equal-sized random sample of majority class instances (nonlesional instances). A “bag” of ten “base-level” classifiers was trained for each of the resulting six subsets of vertices. To classify a vertex as lesional or nonlesional, we first used its sulcal depth to choose the two correct bags of classifiers (e.g., if the sulcal depth was “sulcus”, we use the “thin/sulcus” and “thick/sulcus” classifiers). Next, each of the ten base-level classifiers was applied for each bag. The final classification was obtained by a majority vote of their predictions. The overall training and testing phases of the proposed classification scheme are shown in Figs. 1A and B, respectively.

2.9. Experimental method

A leave-one-out cross-validation (LOOCV) strategy was used to test the performance of the classifier on unseen data. In each run, we left out a single subject from the data and trained a classifier on vertices belonging to all the remaining subjects and the controls. The output of each logistic regression classifier within the bag is the probability that the given input vertex belongs to the positive (lesional) class. To convert this probability into a class label, we defined a threshold \( \rho = 0.95 \) for the output values such that the vertices that have a predicted probability above \( \rho \) were deemed lesional and those that fall below \( \rho \) were considered normal. After classifying each vertex of the test subject, the results were postprocessed [26] to remove spurious detections by defining the detected cluster as a set of contiguous lesional vertices having a surface area greater than or equal to 50 mm\(^2\), an approach similar to [41] where the threshold was determined as the area of the largest cluster detected by the classifier in the control population.

To determine detection values, patients were regarded as true positives if any of the remaining clusters partially or completely overlapped with the lesion/resection area. Outside clusters were considered false positives. It should be kept in mind that the resulting detection outside the lesion/resection zone may actually represent other malformations in the cortex that either have escaped visual inspection or were not part of the seizure onset zone. Thus, the statistics provided here represent a lower bound on actual classifier performance.

2.9.1. Performance evaluation metrics

We use three metrics to quantify and contrast the performance of our classification scheme with the baseline univariate approach. These include the true positive rate (TPR), the false positive rate (FPR), and the Dice coefficient (DC) [42]. The DC is a set similarity metric that is a special case of the kappa statistic [43]. It is commonly used to measure the accuracy of segmentation in medical images when ground truth is available [44,45]. We use DC to measure the overlap between the final detected clusters (after postprocessing) and the resection or expert-traced lesion for a test patient.

Let, the resection/lesion zone be represented by a binary vector \( M_{\text{label}} \in \{0,1\} \), and let \( M_{\text{pred}} \in \{0,1\} \) be the binary vector representing the detection results. The metrics are then defined as follows:

\[
\text{TPR}(M_{\text{label}}, M_{\text{pred}}) = \frac{|M_{\text{pred}} \cap M_{\text{label}}|}{|M_{\text{label}}|} \\
\text{FPR}(M_{\text{pred}}, M_{\text{label}}) = \frac{|M_{\text{pred}} \cap M_{\text{label}}|}{|M_{\text{label}}|}
\]

Fig. 1. Different steps involved in the (A) training and (B) test phases of the proposed classification scheme. Note that the lesion reduction step is applied only to the training patients. For a test subject, we calculate two labels per vertex: one from each thick/thin classifier. The final label of the vertex is calculated as the maximum of both predicted labels.
\[ \text{DC}(M_{\text{pred}}, M_{\text{label}}) = \frac{2|M_{\text{pred}} \cap M_{\text{label}}|}{|M_{\text{pred}}| + |M_{\text{label}}|} \]

where \( |M| \) represents the first norm of the binary vector and, in our case, translates to the number of vertices marked as lesional and \( M \) represents an inverted mask, such that the original 0 values are replaced with 1 and vice versa.

### 3. Results

In this section, we review the performance of our proposed classification scheme and contrast it with the baseline z-score-based method. We also provide empirical evidence to support our design decisions, i.e., classifier stratification, mask reduction, and bagging.

#### 3.1. Overview of the detection results

For MRI-positive patients, both the z-score and machine learning approaches were found to perform identically and accurately detected lesions in 6 out of 7 patients, yielding an 86% detection rate (Table 2). Machine learning correctly identified a significantly larger proportion (t(7) = 3.3, p < 0.05) of the lesional area (mean = 20.14%) compared to the z-score approach (mean = 16.03%) as quantified by the TPR. However, the differences in the DC values were found to be not statistically significant. The false-positive rate was significantly lower for the z-score approach (mean = 1.4%) compared to the machine learning approach (mean = 2.4%; t(7) = 5.1, p < 0.01) (see Table 2). Fig. 2A shows an example of a detected lesion in an MRI-negative patient. Detailed results for both approaches are listed in Table 2.

For patients with MRI-negative lesions, the machine learning approach significantly outperformed the z-score-based method. The z-score-based method correctly detected lesions in 9 out of the 24 patients (37%), whereas the machine learning approach correctly detected clusters inside the resection zone for 14 patients (58%) (see Table 2). Fig. 2B shows an example of a detected lesion in an MRI-negative subject. The overall true-positive rate was significantly higher (t(23) = 3.04, p < 0.01) in the machine learning approach (mean = 2.5%) compared to the z-score approach (mean = 1.1%). The DC values for the machine learning approach (mean = 3.68%) were also significantly superior (t(23) = 3.04, p < 0.01) to the baseline (mean = 1.87%). However, the false-positive rate was also significantly higher (t(23) = 5.65, p < 0.001) in the machine learning approach (mean = 1.0%) compared to the z-score approach (0.6%). Detailed results are shown in Table 2.

#### 3.2. Sensitivity analysis of design decisions

In order to determine whether correcting for cortical complexity by stratifying classifiers by sulcal depth results in improved detection rates, we reran the training phase in the leave-one-out cross-validation without this correction (note that we retain bagging and mask reduction). As depicted in Tables 3 and 4 (compare the ML column to column "A"), the true-positive rate dropped from 20.1% to 12.9% in the MRI-positive group and lesion detection dropped from 58% to 33% in the MRI-negative group. This suggests that different feature combinations might be more prevalent in specific regions (e.g., sulcus, gyrus, and wall), which is consistent with the observation of region-specific dysplasia subtypes (e.g., bottom-of-the-sulcus dysplasia).

In order to correct for the class label noise problem, we employed a strategy to reduce vertices labeled as "lesional" to those that were significantly thicker or thinner than "nonlesional" vertices. We tested the improvement in detection rates when utilizing this strategy by rerunning our analysis without mask reduction (note that we retained stratification and bagging for this experiment). The results are again depicted in Tables 3 and 4 (compare the ML column to column "B") and show a drop in detection rates for both the MRI-positive group (from 6/7 to 3/7) and the MRI-negative group (from 14/24 to 3/24 detections). This indicates that class label noise is a significant issue for both groups that can be corrected by utilizing a mask reduction strategy with a separate threshold for cortical thickening and cortical thinning.

Our last experiments examined the impact of bagging on the results (we eliminated bagging and retained stratification and mask reduction). We see from Tables 3 and 4 that eliminating bagging resulted in the most substantial drop in performance; the TPR of the MRI-positive group dropped from 20.1% to 2.1%, and for the MRI-negative group, the detection rate dropped from 14/24 to 0/24. In other words, failing to correct for the class imbalance problem resulted in zero detection of MRI-negative FCD lesions. This strongly supports the use of such bagging and stratified classifiers in future machine learning models for FCD detection. Fig. 3 summarizes our results and contrasts the rate of detection for MRI-negative patients under different variations in the design of the machine learning approach.

### 4. Discussion

Our results demonstrate that surface-based morphometry, coupled with a multivariate classification scheme that is adapted for FCD lesion data, can successfully detect epileptogenic FCD lesions on MRIs that were previously interpreted as normal by neuroradiologists. This approach correctly identified epileptogenic regions in 58% of MRI-negative patients compared to 37% when using univariate statistics. A separate analysis showed that while the best detectors of FCD lesions...
were cortical thickness and GWC, features commonly used in the visual diagnosis of FCD, measures of cortical complexity, such as curvature and Jacobian distortion, also contributed strongly to lesion detection. This finding suggests that MRI images contain ample information about focal epileptogenic lesions but do so to a degree and in a complexity that may not be appreciable by visual inspection alone.

Malformations of cortical development are the third most frequent disease entity associated with TRE, and FCD is the underlying pathology in 75% of these cases [46]. Resection of FCD tissue is critical to seizure control; therefore, it is an important target for MRI evaluation during presurgical assessment. The presurgical detection of a lesion informs intraoperative planning and provides a valuable target that, when surgically resected, can lead to a substantial improvement in postsurgical outcome [47,48]. Indeed, surgical success in patients with neocortical epilepsy and a concordant MRI lesion is drastically improved (66%) compared to cases without lesions (29%) [49].

The application of machine learning algorithms to the detection of FCD lesions resulted in a unique set of challenges that are specific to this clinical population, requiring innovative solutions. The existence of these challenges and the improvement in classification when solutions were implemented offer a unique perspective on the biological complexity of focal cortical dysplasia. One such challenge was the presence of abnormal surfaces outside of the histopathologically confirmed focal dysplastic region. Although these are considered to be statistical false positives, alternative explanations must also be considered, such as the following: (1) the presence of a dysplastic cortex outside of the seizure onset zone may or may not have latent epileptogenic potential [38] and/or (2) the burden of intractable seizures on brain structure could result in subtle abnormalities (e.g., atrophy and gliosis) that may be difficult to distinguish from developmental aberrations [34–37,50]. Both of these possibilities could impact postsurgical outcomes and are thus worth further exploration. For example, magnetoencephalography or intracranial electroencephalography could be used to determine whether there is abnormal electrophysiology in these “false-positive” regions. Tracking postsurgical outcomes could determine whether a greater extent of “extralesional” abnormalities is associated with suboptimal postsurgical seizure control or functional outcomes.

An additional challenge for machine learning algorithms is the heterogeneity of pathological and MRI features in FCD. For example, FCD lesions might contain small diameter cells that may result in an abnormally thin cortex on MRI [51] or large dysmorphic cells that may result in an abnormally thick cortex on MRI [7]. We observed improved classification rates when we stratified labeling of “lesion” vertices in the lesion zone/resection zone based on separate thresholds for cortical thickening and thinning. Additionally, specific regions of the cortical architecture may be more vulnerable to dysplastic pathology. Focal cortical dysplasia lesions occur with higher frequency at the bottom of the sulcus, potentially reflecting different “micromechanical” tensions that enhance pathophysiological vulnerability in the sulcal bottom [22]. We observed improvement when we stratified classifiers trained separately for sulcal, wall, and gyral regions. The improvement in our model after implementing such solutions suggests that similar stratification strategies should be employed in future FCD lesion detection efforts.

The resection zones of MRI-negative patients include both lesional tissue and nonlesional tissue; therefore, the resection zone cannot be treated as a gold standard for training classifiers. Hong et al. [16] utilize a mask reduction step in which texture maps [31,32] are used to manually trace the lesion for MRI-negative patients that have type II FCD. This preprocessing not only entails the generation of texture maps but also requires specific human expertise to identify lesions. In the proposed approach, we use cortical measures to reduce the resection mask, such that the resected regions that are not significantly different from the regions outside the resection zone are not used for training the classifier. We hypothesize that this approach will accurately classify FCD lesions in a sample of patients with verified MRI-negative FCD lesions.

The methodological approach used in the current study to improve lesion detection of MRI-negative images has a number of significant advantages. First, it works with most existing scanners and sequences and does not require advanced imaging technologies. Second, as we learn more about FCD, stratification of larger data sets into distinct FCD subtypes [52] can be incorporated into future training sets to help the system learn specific subtype features and potentially classify FCD by subtype. Third, such a method can be fully automated and thus, with minimal effort, can augment visual inspection by yielding targets for closer evaluation by neuroradiologists. This latter point is important, given the fact that the visual detection of FCD on MRI varies

![Fig. 2. Detection results for the ML-based approach for (A) an MRI-positive and (B) an MRI-negative patient. The inflated lateral and medial cortical surfaces show the original expert-traced lesion (A) or the resection zone (B) as the regions outlined by the white solid curve. The significant lesional clusters discovered by the ML-based approach are shown in yellow. The MRI slice on the right shows the abnormal area corresponding to the clusters discovered inside the lesion/resection on the actual brain volume. (For interpretation of the references in this figure legend, the reader is referred to the web version of this article.)](image108x596 to 478x741)

### Table 3

A comparison of detection results using the z-score-based method and the ML method only for MRI-positive subjects with different variations in the design of the ML approach.

<table>
<thead>
<tr>
<th>Patient</th>
<th>z-Score</th>
<th>ML</th>
<th>(A)</th>
<th>(B)</th>
<th>(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TPR</td>
<td>FPR</td>
<td>TPR</td>
<td>FPR</td>
<td>TPR</td>
</tr>
<tr>
<td>NY49</td>
<td>11.8</td>
<td>1.0</td>
<td>24.8</td>
<td>2.3</td>
<td>8.7</td>
</tr>
<tr>
<td>NY53</td>
<td>20.3</td>
<td>2.6</td>
<td>27.7</td>
<td>4.5</td>
<td>23.5</td>
</tr>
<tr>
<td>NY123</td>
<td>29.8</td>
<td>3.7</td>
<td>31.3</td>
<td>4.5</td>
<td>31.4</td>
</tr>
<tr>
<td>NY143</td>
<td>16.4</td>
<td>0.6</td>
<td>20.0</td>
<td>2.0</td>
<td>26.4</td>
</tr>
<tr>
<td>NY156</td>
<td>26.1</td>
<td>1.2</td>
<td>25.7</td>
<td>2.1</td>
<td>20.1</td>
</tr>
<tr>
<td>NY178</td>
<td>–</td>
<td>0.5</td>
<td>0.9</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>NY194</td>
<td>7.8</td>
<td>0.1</td>
<td>11.5</td>
<td>0.6</td>
<td>–</td>
</tr>
<tr>
<td>Mean (%)</td>
<td>16.0</td>
<td>1.4</td>
<td>20.1</td>
<td>2.4</td>
<td>12.9</td>
</tr>
</tbody>
</table>

(A) No stratification along the sulcal values, (B) stratifies the data based on the sulcal depth values but does not reduce the lesion mask, and (C) uses stratification and lesion reduction, but it does not use bagging. The TPR and FPR are measured as a percentage.
widely among raters, and is highly dependent on the experience of the evaluator.

4.1. Limitations

In the current study, training data in MRI-negative cases were derived from resection areas that were defined by intracranial electrophysiology. Focal cortical dysplasia pathology was present in the resection area in all patients; however, nonlesional tissue may have also been resected. We reduced this problem by applying a mask reduction step, and this increased performance. In future research studies, this step can be improved by accurately coregistering the pathological sample with the MRI, allowing the matching of pathological and MRI slices.

In addition, our sample of MRI-negative patients was disproportionately higher than MRI-positive patients, which may not reflect the proportions seen at other neurological clinics. This likely represents a bias in patient referrals to our level 4 epilepsy center, which offers intensive neurodiagnostic monitoring for patients with treatment-resistant epilepsy that is difficult to localize. Our results offer a potential advancement of neurodiagnostic tools for this more challenging population. However, the case–control methods that we utilize in our approach require a large healthy control MRI data set with identical scanning parameters as those of the patient and thus cannot be readily applied in any clinical center. Further investigations with combined data sets from different scanners and institutions are needed to create methods for making these analyses feasible with different scanning sequences across centers. Finally, automated detection and classification of lesions should not replace careful visual analysis with a high lesional probability. These results should always be interpreted in the context of all available patient information collected during presurgical evaluation.

5. Conclusion

In summary, we have demonstrated that a quantitative morphometric method using surface-based brain modeling, combined with machine learning algorithms and novel strategies to deal with the complexity of cortical malformations, results in improved detection of FCD. Improved detection of neocortical structural lesions is likely to increase the number of patient referrals to specialized tertiary epilepsy centers for surgical consideration and, in many cases, may decrease the delay between initial diagnosis and surgery. This has significant implications for improved seizure and cognitive outcomes in patients with FCD and concomitant epilepsy.
Acknowledgments

We would like to thank John Brumm, Omar Khan, and Christine Chao for help with data processing and Rachel Jurd for comments. This research is supported by FACES—Finding a Cure for Epilepsy and Seizures (TF) and the Epilepsy Foundation (KEB, BA, and CC).

Conflict of interest

The authors declare no competing financial interests.

References