

# Automated focal cortical dysplasia detection on brain MR images using image analysis

N. Alsahanova, V. Yarkin, M. Sharaev, O. Bronov, V. Bychenko, A. Marinets, E. Syrkashev, O. Karpov, E. Burnaev, A. Bernstein, E. Spodarev

**Abstract**—Focal cortical dysplasias (FCDs) represent one of the most frequent causes of pharmaco-resistant focal epilepsy. Despite improved clinical imaging methods over the past years, FCD detection remains challenging, as FCDs vary in location, size, and shape and can even be unrecognized. In this work we propose novel FCD detection algorithms, based on distributions of grey and white matter intensities. Based on these algorithms, we created three-dimensional feature maps where the brain areas with intensity and morphological abnormalities are highlighted.

We collected 3-dimensional T1-weighted, T2-weighted and fluid-attenuated inversion recovery (FLAIR) magnetic resonance images (MRI) of 90 patients with clinically confirmed FCD. All images were preprocessed and segmented for further analysis. Gray and white matter distributions are analyzed with a set of algorithms and compared to a normal database, resulting in the detection of clusters. To evaluate the detection performance of our method, we calculated metrics considered common in similar works.

Feature maps based on T1-weighted images junction maps reached lesion detection in 71% of subjects (64 of 90), where the lesion hits top 10 detection map clusters. On the other hand, curvature detection maps revealed poor performance with detection on 11% of subjects.

Additionally, we integrated the information from all feature maps in classification models to achieve automated detection tool for FCD lesions. Best model sensitivity reached 78%.

Presented algorithms demonstrated interpretable and computational efficient automated detection of the FCD. Created feature maps and information combining classification models carry viable information in the diagnostic routine of drug-resistant focal epilepsy.

**Index Terms**—Focal cortical dysplasia, voxel-based analysis, blurring, brain segmentation, statistical image analysis.

This work was funded by Russian Science Foundation [grant number: 21-71-10136] to M.S. (machine learning models for FCD detection); E. Spodarev gratefully acknowledges the partial financial support for his sabbatical leave in the fall 2021 by DAAD grant "Ostpartnerschaften" as well as by Skoltech, Moscow.

N. Alsahanova, V. Yarkin, M. Sharaev, E. Burnaev, A. Bernstein are with the Skolkovo Institute of Science and Technology, Moscow, 121205, Russia (e-mails: N.Alsahanova@skoltech.ru, V.Yarkin@skoltech.ru, M.Sharaev@skoltech.ru, E.Burnaev@skoltech.ru, A.Bernstein@skoltech.ru).

O. Bronov, A. Marinets, O. Karpov are with the Pirogov National Medical and Surgical Center, Moscow, 105203, Russia.

V. Bychenko, E. Syrkashev are with the Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology, Moscow, 117997, Russia.

E. Spodarev is with the Institute of Stochastics, Ulm University, Ulm, D-89069, Germany (e-mail: evgeny.spodarev@uni-ulm.de).

## I. INTRODUCTION

### A. Clinical task

Epilepsy is one of the most common neurological disorders characterized by chronic and recurrent epileptic seizures. MRI is nowadays a substantial diagnostic tool aiming at detecting potentially epileptogenic lesions for neurosurgery planning. Over the past decades, the quality of MR images has substantially improved not only by equipment but also by further development of epilepsy-specific protocols and particular sequences (e.g. high resolution FLAIR in focal cortical dysplasia, i.e., a cortical malformation resulting from abnormal proliferation of neurons and abnormal cortical organization during pregnancy).

Despite the efforts to improve MRI acquisition protocols, about 20-30% of patients with focal epilepsy still remain non-lesional after visual inspection of their MR images. Detection of pathological changes by "manual" analysis of MRI requires the radiologist to visually inspect a large number of images obtained in various scanning modes using a significant number of routine operations, and therefore the overall process can take several hours. For example, to detect and localize an epileptogenic focus, routine diagnostic techniques include analyzing images obtained in different pulse sequences with different resolution and signal-to-noise ratio, evaluation of the intensity of the white and gray matter tissues, as well as the presence or absence of focal changes. Therefore, to detect and localize an epileptogenic focus, an experienced radiologist may need to analyze from a few hundred to several thousand two-dimensional MRI slices obtained by various scanning protocols. Additionally, the low contrast between healthy tissue and pathology can make it difficult to reliably detect small hidden lesions, requiring the expertise of highly experienced radiologists. Computerized MRI analyses thus have been introduced to enhance visualization of especially subtle lesions that might be overlooked by conventional visual inference with the aim of reducing the proportion of non-lesional patients.

Typical MRI features of FCD include abnormal gyral contours, thickening of the cortex, abnormal differentiation of the grey-white matter boundary, and sometimes signal hyperintensity in T2-weighted images. In subtle cases, however, diagnosis is time-consuming and difficult, and conventional MRI can be unrevealing in a high percentage of patients. Therefore, attempts have been made to facilitate lesion de-

tection by modern image post-processing strategies.

### B. Current state of the art and difficulties

1) *Image features and statistical approaches*: Previously a voxel-based method has been developed for post-processing of MRIs based on algorithms of the freely available software for statistical parametric mapping (SPM) and on additional simple calculations and filters; this approach has been used in a multitude of international studies, see for example [15]. Here, from a high-resolution T1-weighted 3D MRI dataset three new feature maps are derived which characterize in comparison with a normal database three different potential features of FCD: the abnormal extension of grey matter into white matter, blurring of the grey-white matter junction, and abnormal thickness of the cortical ribbon. This morphometric MRI analysis has proven to facilitate the detection of subtle FCD (and potentially other epileptogenic lesions) and to increase the diagnostic yield of MRI in that respect, although, in the original approach, the recognition of FCD in the new feature maps was still based on visual examination, and a clear cut-off differentiating between truly dysplastic brain areas and false-positive findings highlighted in these maps was missing. In contrast to the conventional visual analysis of MR images, however, computer-based quantitative post-processing techniques such as this morphometric MRI analysis already offer several advantages, e.g. an inherent comparison with a normal database, a comprehensive and gapless 3D high-resolution analysis of MRI data with reduced risk for misinterpretation due to partial volume effects, and the enhanced visualization of different potential and complementing imaging features of cortical malformations which are not always easily assessable by pure visual examination.

A large number of works discuss the difficulties of detecting FCD in visual manual analysis of MRI. As mentioned above, FCD patterns on MRI are connected with local thickening of the cortex and / or violation of the border between white and gray matter and / or focal atrophy and / or altered MR signal at different pulse sequences [3].

2) *Machine learning and computer vision approaches*: There is a large number of papers in which the methods of machine learning (ML) and computer vision are used for segmentation/detection of medical images (MRI, compute tomography, ultrasound, electroencephalography and electrocardiography, etc.) to detect/isolate tumors, lesions and other pathological changes. One of the most adequate approaches, and, at the same time, one of the most difficult in the field of computer vision, is semantic segmentation; the relevance of this approach over the past 30 years is constantly growing. Convolutional neural networks (CNN) that are part of Deep Learning (DL) technologies show the best results in many areas, including medical data diagnostics.

However, the vast majority of papers [33], [8], [12] on the application of this kind of methods do not take the features of the data into account, inherent in most cases that are studied in clinical practice, namely:

- MRI obtained using different pulse sequences have varying intensities of the MR signal,

- for some pulse sequences, the radiologist does not have a complete data array, but only a set of 2D images taken from slices examined at irregular intervals; in turn, these 2D slice images have a higher resolution, and non-isotropic pixel/voxel sizes are used,

- only a part of MRI will have full segmentation masks; for other MR images, either only bounding boxes will be indicated, or just the binary answer 0/1 - the presence/absence of epilepsy, etc.

For this reason, the straight use of existing deep learning methods for the detection of MRI described above cannot be carried out, and it is challenging to try new data preprocessing and feature extraction techniques before applying DL models.

In the last decade, more elaborated technical advances have been established and were recently combined with ML techniques in first studies [16], [20], [19] to enable a fully automated detection of FCDs with high sensitivity. Beside voxel-based methods, surface-based analyses have been applied by other research groups also providing promising results especially in combination with multivariate ML techniques [1], as demonstrated in MRI samples of limited size 24.

Further studies mainly by the Dr. Bernasconi research group could repeatedly confirm the feasibility of multivariate ML techniques in the detection of previously MRI-negative FCDs in a fully automated manner [14], [13]. A study by Jin et al. [16] applied surface-based morphometry combined with an automated classifier to a large cohort of patients with FCD type II whose MRIs were obtained in different epilepsy centers. Despite the heterogeneity of MRI scanners and protocols, they found a robust detection performance of their method. Currently, the data-driven analysis of disease factors provides a novel appraisal of the continuum of interindividual variability, which is likely determined by multiple interacting pathological processes so that in recent publications, incorporating interindividual variability is likely to improve clinical prognostics [20].

Although unbiased approaches are well established, automated detection methods fail in 20% to 40% of patients, particularly with subtle FCD, and suffer from high false-positive rates. In the meanwhile, current approaches successfully use unsupervised learning in FCD detection [19]; although first studies even used deep CNN classifiers in multicenter data (148 patients' MRI [12]), the optimization of the techniques still remains a challenge given the complexity of the MRI phenotypes.

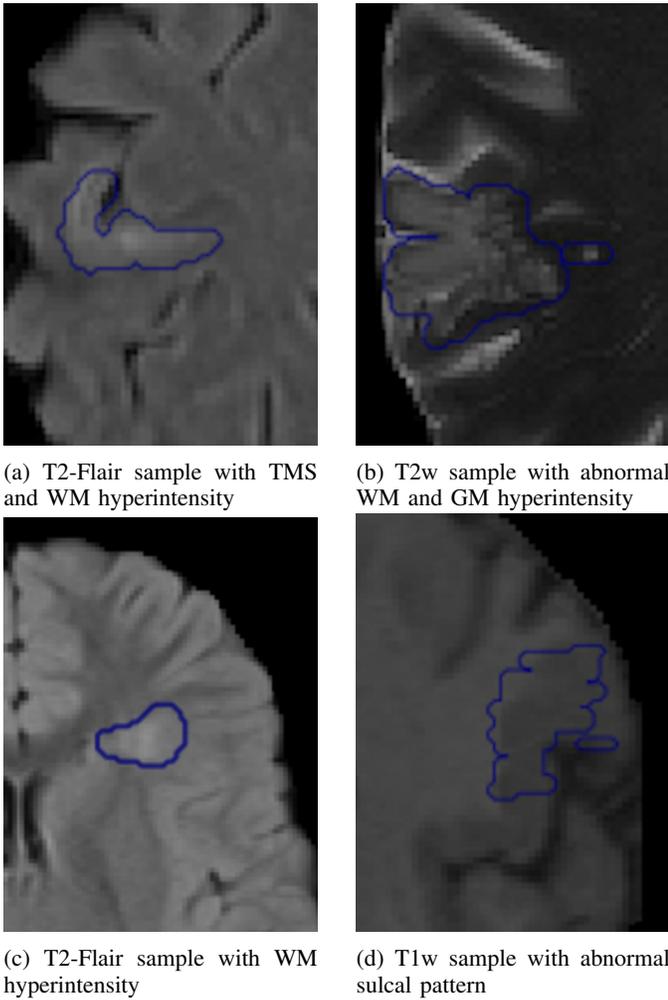
### C. Local FCD image properties and novel statistical features

In this work we studied distinctive image features that represent visual signs of FCD lesions.

1) *Local FCD image properties*: The following five image features of cortical dysplasia were analysed:

Transmantle sign (TMS) - signal change extending from the subcortical white matter in direction to the ventricle (Fig. 1a).

White matter (WM) hyperintensity - abnormal signal change in the subcortical white matter according to the normal cortex (Fig. 1a, 1c).



**Fig. 1:** Examples of FCD image properties with corresponding manual segmentation

Cortical gray matter (GM) signal intensity - signal change involving the entire thickness of T1w, T2w, Flair cortex (Fig. 1b).

Abnormal sulcal pattern (Fig. 1d) - abnormal anatomy of the gyrus and sulcus, different from the standard brain model, without changing signal characteristics of the brain structures. It could represent both extended zones and local change in the course of gyrus and sulcus [31].

Segmental hypoplasia/atrophy - underdevelopment of gray and white matter of the brain due to conditions during embryonic and perinatal development or due to impaired trophic/blood supply to the brain.

**2) Features:** In this paper, we propose three image features and show that they perform well detecting different kinds of brain lesions in our MRI database:

- the *concentration rate*  $CR_g$  of  $g$  gray scales close to white is designed to locate lesions characterized by gray or white matter hyperintensity,
- the *local  $\alpha$ -Rényi entropy*  $R_\alpha$  of gray scales with  $\alpha = 2$  performs best finding gray or white matter hyperintensity areas and transmatle sign,
- the *anisotropic sample variance* approach helps detecting

blurred gray-white matter junction.

Although these features are not new, their use in FCD detection context is certainly novel.

#### D. Paper structure and aims

After an initial data and image preprocessing description, we introduce and explore novel statistical features targeting lesion-specific properties on MR images. Special focus here is to assess and quantify local (near the lesion - non-lesion border) feature properties.

Next, on in-house original dataset we calculate baseline quality lesion segmentation metrics for well-established conventional features like gray matter thickness, curvature, gray-white matter junction maps (blurring) and sulcal depth.

Similar to [7], we trained a feed-forward network, multi-layer perceptron (MLP), to make a prediction of FCD location in a simple, interpretable manner. We trained this model on all features and calculated their importances for DL model.

## II. METHODS

First, we give a brief mathematical description of image features that we use in the segmental MRI analysis. Then, we give a detailed account of all image processing steps needed for final FCD lesion detection, including normalization, post-processing as well as neural network training.

Let  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$  be the set of all natural numbers and zero. For  $d = 2, 3$ , consider a lattice

$$L_d = \{1, \dots, M\}^d$$

of  $M^d > 0$  pixels or voxels  $v$  making up a (for simplicity, quadratic or cubic)  $d$ -dimensional gray scale MR image

$$X : L_d \rightarrow \mathbb{N}_0 \cap [0, 255],$$

where  $X(v)$  is the brain MR gray scale value at a pixel/voxel  $v$ . Here, the gray levels 0 and 255 correspond to black and white colors, respectively. Let  $W_m(v) = v + [-m, m]^d \cap L_d$  be a cubic mesh of side length  $(2m + 1) < M$  with its central vertex at  $v \in L_d$ . We will use it as a scanning window moving around our MR image  $X$ .

#### A. Novel MRI features and their properties

In what follows, we consider pixel/voxel-based features

$$F_X : L_d \rightarrow \mathbb{R}$$

which are computed from a local knowledge of MRI  $X$ , i.e.,  $F_X(v)$  depend of the values of  $X$  within the scanning window  $W_m(v)$ . Such a feature  $F_X$  is used to detect an anomalous area  $A_X$  of brain MRI  $X$  taking the level set

$$A_X = \{v \in L_d : F_X(v) \notin [u_{\gamma/2}(v), u_{1-\gamma/2}(v)]\}$$

for some small error of the first kind  $\gamma > 0$  where  $u_\beta(v)$  are empirical quantiles of  $F_X(v)$  at level  $\beta \in (0, 1)$  computed from the training set of all MRI of a healthy hemisphere of brains from our dataset.

Now we briefly describe three image features  $F_X$  which help finding specific kinds of lesions in brain MRI. Borrowed from mathematics and statistics, they have been never used in brain lesion detection before.

1) *Concentration rate*: Following the well-known concentration measures from econometrics like the *Lorenz curve* and *Gini coefficient* (cf. e.g. [11], [32]), the *concentration rate* (CR) of gray scales of  $X$  can be introduced as

$$CR_{g,k}(v) = \sum_{i=(2m+1)^d-g-k+1}^{(2m+1)^d-k} X_{(i)}, \quad (1)$$

where  $g \in \mathbb{N}$  is the number of highest gray scales of  $X$  truncated on  $k$  taken into account, and  $X_{(i)}$  is the  $i$ -th highest order statistic of gray scales of  $X$  in our scanning window  $W_m(v)$ .

If  $g$  and error level  $\gamma > 0$  are small enough, the statistic  $CR_{g,k}$  finds regions within the brain MR image subject to high concentration of white scales. In our 3D MRI database,  $CR_{g,k}$  is sensitive to lesions of gray or white matter hyperintensity.

2) *Local entropy of gray scales*: For calculating the entropy feature, we used the commonly extracted gray-level co-occurrence matrix (GLCM) [29], [6]. This matrix shows the frequency that two voxels of specific gray-level intensities are at a certain distance away from each other in a specified image direction. We computed GLCM for distances from 1 to 4 within 13 main directions. Then, we summed up matrices for all distances to get one matrix. Let  $f_{i,j}$  be an element of this matrix of size  $256^2$ . Renyi entropy of order  $\alpha > 0$ ,  $\alpha \neq 1$  given by

$$H_\alpha(v) = (1 - \alpha)^{-1} \log \left( \sum_{i,j=0}^{255} f_{i,j}^\alpha \right)$$

is a measure of "disorder" of gray scales [26], [27].

For our 3D brain MR images, we locally computed entropy estimates  $\hat{H}_\alpha$  for the values  $\alpha = 2, 3, \dots, 19, 20$ . The feature  $\hat{H}_2$  performed best to discover lesions characterized by white or gray matter hyperintensities or transmantle sign.

3) *Anisotropic sample variance*: Second-order methods using covariances and variograms are common in geostatistics [5], [30] to detect changes in dependence structure of a spatially located feature. With a brain MR image at hand, this means the difference in thickness of white and gray/black layers of the image which may indicate the presence of a lesion. Using the gray scales  $X(t)$ ,  $t \in W_m(v)$  of a  $d$ -dimensional MRI  $X$  ( $d = 2, 3$ ), we define the *local empirical variance* at a pixel/voxel  $v \in L_d$  as

$$\hat{V}(v) = \frac{1}{(2m+1)^d - 1} \sum_{t \in W_m(v)} [X(t) - \bar{X}(v)]^2,$$

where

$$\bar{X}(v) = \frac{1}{(2m+1)^d} \sum_{t \in W_m(v)} X(t)$$

is the local sample mean of the gray scales. Differences in the values of  $\hat{V}(v)$  in various directions along or across the boundary of gray-white matter for healthy and deceased brain areas help finding specific kinds of lesions such as blurred gray-white matter junction or transmantle sign.

## B. Conventional features

Here some classical features used in lesion detection of MRI are described.

1) *Cortical Thickness*: Cortical Thickness is measured as a distance between white and cortical surfaces [3] for each vertex of cortical reconstruction. White and Cortical surfaces are obtained by inflating a mesh model to the boundary between gray matter and white as well as gray matter and cerebrospinal fluid, respectively. FreeSurfer routinely has a hard limit of 5 mm on the cortical thickness, so we ran the command "mris-thickness" expert option parameter that allowed a maximum value of 10 mm.

2) *Sulcal depth*: The dot product of the movement vector of the cortical surface during inflation is used to calculate the sulcal depth. Shallow gyral areas of the brain move inwards during inflation and have a negative value whereas deep sulcal areas move outwards and have a positive value [9].

3) *Curvature*: It was calculated as mean curvature (1/R mm) of the white matter surface with a Gaussian smoothing kernel applied [25].

4) *Blurring*: To detect and localize areas of increased gray-white matter indicative of FCD, the so-called "junction images" were calculated [15], [17], considered to brighten abnormal blurring of the gray-white matter junction. Junction images were obtained with several steps. First individual lower and upper intensity thresholds were calculated for T1w images according functions:

$$\begin{aligned} T_{lower\ threshold} &= Mean_{GM} + \delta \cdot Std_{GM}, \\ T_{upper\ threshold} &= Mean_{WM} - \delta \cdot Std_{WM}, \end{aligned} \quad (2)$$

where mean (*Mean*) and standard deviation (*Std*) of masked T1w are computed by probability maps of gray matter (GM) and white matter (WM) respectively,  $\delta$  is a parameter chosen experimentally so that

$$T_{lower\ threshold} < T_{upper\ threshold}.$$

We also explored "junction images" for T2w and Flair images. Thresholds for them were calculated according functions:

$$\begin{aligned} T_{lower\ threshold} &= Mean_{WM} + \delta \cdot Std_{WM}, \\ T_{upper\ threshold} &= Mean_{GM} - \delta \cdot Std_{GM}. \end{aligned} \quad (3)$$

Each voxel with a gray value between these thresholds was set to 1, while the remaining voxels were set to zero resulting in a binary image. To brighten clustered regions with voxels equal to 1, the binary image was filtered with convolutional kernel of ones with shape  $5^3$ , as was done in [15]. After that, a normalization was performed. The method of normalization will be described in Section II-D.

## C. Features with poor performance

In contrast to the above features showing good success in lesion detection of some kinds, the following more conventional features from image analysis did not perform well and are therefore discarded in the sequel. Those are, for instance, Frangi [10], Meijering [21] and Hessian [22] filters that were designed to extract thin image objects such as blood vessels, neurites, wrinkles and cracks. Modified energy approach based

on GLCM and normally used for texture analysis shows no lesion detection as well.

#### D. Normalization

The last part of calculation of a junction image (cf. Section II-B.4) is a normalization on healthy patients. It is an important step because it compensates the variability of junction between gray and white matters in different brain regions. In previous works [15], the normalization was performed by subtracting the mean of images obtained for healthy patients.

In our case, we had only six brain images of control subjects, therefore we extracted hemisphere of brain images from FCD dataset where no FCD lesions were found. In the result we had 95 left and 94 right healthy hemispheres of brains. We used quantiles to normalize junction images and features described in Section II-A. The use of quantiles is widespread in image analysis because a quantile is a more robust statistic than the sample mean. We calculated different sets of quantiles  $q_l$  and  $q_s$ , where  $q_l$  is the larger quantile and  $q_s$  is the smaller one. The normalized features were calculated according function:

$$X(v)_n = \begin{cases} X(v) - q_l, & \text{if } X(v) > q_l; \\ 0, & \text{if } q_s \leq X(v) \leq q_l; \\ q_s - X(v), & \text{if } X(v) < q_s, \end{cases}$$

where  $X(v)$  is the MRI value in a voxel  $v$  before normalization,  $X(v)_n$  is the normalized voxel value. The confidence levels for quantiles ( $q_l, q_s$ ) were taken from the set  $\{(0.98, 0.02); (0.975, 0.025); (0.95, 0.05); (0.925, 0.075); (0.9, 0.1)\}$ . The pair of quantiles that after normalization gives the highest precision was chosen for different features.

#### E. Postprocessing

The postprocessing procedure was done similarly as in [12]. The feature map was binarized by equating to 1 everything that is not equal to 0. Then it underwent a postprocessing routine entailing morphologic erosion, dilation, and extraction of connected components ( $> 75$  voxels as in [12]) to remove noise. This procedure resulted in non-overlapping clusters. The resulting binary map was multiplied by the initial feature map to save information about the brightness of clusters after normalization.

Features as thickness, curv and sulc did not undergo normalization in the way described in Section II-D because quantile normalization has lowered the quality metrics on these features. So for them postprocessing took place in a different way. Maps of these features were binarized at thresholds equal to 10% and 90% quantiles to form separate clusters. Further steps are the same as for other features.

#### F. Model design and data sampling

1) *ANN classification*: Automatic lesion detection was performed by using an artificial neural network (ANN). We did not use convolutional neural networks (CNN) as features were already extracted. To utilize computational models derived from MRI characteristics, single voxel values from 15 maps

were used as input values for ANN. Thus, ANN is trained to classify each voxel of input feature maps on dysplastic and non-dysplastic classes. We tried one-layer and two-layers perceptrons with 5 to 15 nodes per layer. We used ReLU activation layers between linear layers and a sigmoid layer in the end.

2) *Training and inference workflow*: Before the training procedure, we use histogram standardization [23] and z-normalization for all the features. The usage of these normalization techniques allows us to increase the quality of the results.

Training of ANN on all voxels from a brain is time consuming, so it was decided not to choose all voxels. We chose only dysplastic voxels and random  $N$  non-dysplastic voxels. We tuned the number of non-dysplastic voxels per subject,  $N$ , included in the training procedure using cross-validation. This number was chosen from the set:  $\{3, 5, 8, 10, 12\} \times 10^5$ . For validation, we took all voxels in validation brains.

As a loss function for the network training, we use a variant of the Dice loss  $D$ , based on the Sørensen–Dice coefficient  $DSC$ , see Eq. (4), (5). This coefficient is widely used as a measure of the segmentation quality [2]:

$$DSC = \frac{2 P_b \cdot Q_b}{|P_b| + |Q_b|} = \frac{2 TP}{2 TP + FP + FN}, \quad (4)$$

$$D = 1 - DSC, \quad (5)$$

where in Eq. (4), the binarized volumetric network output  $Q_b$  indicates predicted probabilities of the FCD voxels, which is compared with the binarized ground-truth layout  $P_b$ . It can also be rewritten in terms of  $TP$ ,  $FP$  and  $FN$ , indicating the total number of True Positive, False Positive and False Negative voxels, respectively.

We trained the network to minimize the loss function, using backpropagation within the PyTorch framework [24]. For the optimization, we used Adam method [18], applying it for 60 epochs with learning rate  $\lambda = 0.001$  and  $\beta_0 = 0.9$ ,  $\beta_1 = 0.99$ . Learning rate was decreased in 0.2 times each epoch. Batch size was 1.

The simplicity of the workflow was chosen on purpose, since the main goal of this model is to estimate the validity of the proposed features for FCD detection and not the model per se.

#### G. Computing Resources and Data Availability

All computations and experiments were performed on a desktop with Intel(R) CPU, i-7, 256 GM RAM, GPU x4 GTX1080Ti 12 GB, HDD x4 11 TB.

The source code is available by the link [https://github.com/NadezhdaAlsahanova/FCD\\_features](https://github.com/NadezhdaAlsahanova/FCD_features) (version 1.0.0 at GitHub).

### III. RESULTS

#### A. Demographics

The overall cohort comprised 90 patients with FCD (45 female patients, mean  $\pm$  SD age  $26 \pm 13.9$  years) with MRI positive of FCD Type II. Visual inspection reported

localization of lesions in front parietal lobe (n=52), temporal lobe (n=35) and occipital lobe (n=6), multilobar cases (n=2). Multilobar cases comprised two lesions in left frontal lobe and two lesions in left frontal lobe and temporal lobe. Table with lesion localization and visual features is in Supplementary materials that is available in the supporting documents. White matter hyperintensity was mostly presented (n=66) as a visual feature of FCD Type II on T2w images.

## B. Metrics

1) *Voxel-Level Performance*: To evaluate the effectiveness of the proposed features and models, we calculated several metrics: precision, DSC, sensitivity and specificity. Precision, sensitivity and specificity are defined as

$$Precision = \frac{TP}{TP + FP},$$

$$Sensitivity = \frac{TP}{TP + FN},$$

$$Specificity = \frac{TN}{TN + FP},$$

where  $TP$  and  $FP$  are defined above. We considered precision to be the target metric, since doctors are more interested not in the absolute accuracy of finding the lesion, but in detecting several lesion regions from which they will be able to determine the exact lesion. Other metrics were considered as they are often used in similar works.

We decided to measure metrics by threshold to evaluate segmentation performance. So we chose a threshold between 0 and maximum brightness of a feature. After that we calculated numbers of true positive ( $TP$ ), false positive ( $FP$ ), true negative ( $TN$ ) and false negative ( $FN$ ) voxels for the areas of the features above a threshold. It was done to determine whether the maximum values of the features are in the lesion area.

Then the metrics were calculated as follows:

$$metric = metric(TP, TN, FP, FN) \cdot I\{(TP+TN) > 800\},$$

where  $TP$ ,  $TN$ ,  $FP$ ,  $FN$  depend on a threshold,  $I\{(TP + TN) > 800\}$  is an indicator function to check that the volume of found areas is not too small. 800 voxels is the first quartile of the size of lesions in our dataset.

2) *Patient-Level Performance*: After the post-processing stage described in Section II-E, all features maps have clusters of different brightness. Since the doctors are able to see the brightest clusters better, we ordered clusters by their average brightness. Moreover, the greater the deviation from quantiles of a healthy brain, the higher the brightness of clusters. Then, for each patient, the brightness order of clusters that intersect with the lesion was found. The percentage of patients for whom this order does not exceed 10 is a metric of sensitivity.

## C. Explore novel features and their local properties

To evaluate feature performance on the border between lesion and healthy tissue, we chose only parts of the brains that contain local areas around lesions. These areas include lesions

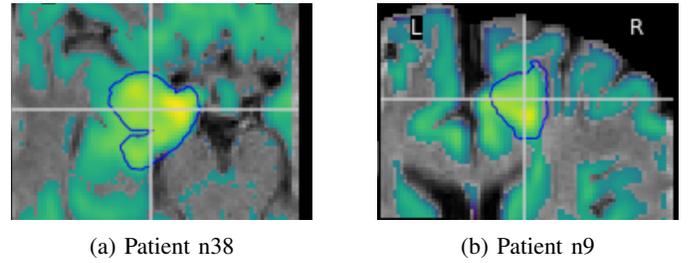


Fig. 2: Examples of lesion detection with the concentration rate feature. Lesion areas are outlined with a blue line.

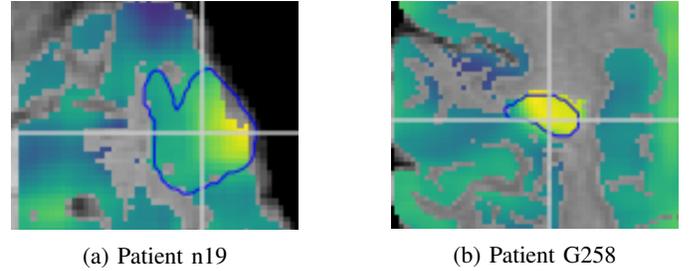


Fig. 3: Examples of lesion detection with entropy feature. Lesion areas are outlined with a blue line.

and several gyri around them to see the difference between the values of features on healthy and diseased voxels. In this subsection we describe the results of calculating our features in these local areas.

1) *Concentration rate  $CR_g$* : The feature described by formula (1) depends on three parameters:  $g$ ,  $k$ ,  $m$ . The side length of the scanning cube  $2m + 1$  was chosen from numbers  $\{3, 5, 7, 9, 11\}$ . The parameter  $g$  was chosen between 15 and the number of voxels in a cube with step equal to 15. The parameter  $k$  was chosen from 1 to 15, it should be small to discard only outliers. The highest difference between the values in healthy and diseased voxels appeared to be with  $m$ ,  $g$  and  $k$  equal to 4, 105 and 15, respectively. It can be seen in Fig. 2 how  $CR_g$  finds lesion areas.

The mean values of precision and sensitivity for these parameters are equal to 0.35 and 0.55.

2) *Local  $\alpha$ -Rényi entropy  $H_\alpha$* : Two main parameters for entropy feature are  $m$  and  $\alpha$ . The parameter  $m$  for side length of the scanning cube  $2m + 1$  was chosen from a subset  $\{2, 4, 6, 8, 10\}$ .

The parameter  $\alpha$  was chosen between 2 and 20. The best parameters, which gave the maximum difference between diseased and healthy voxels, are  $m = 10$  and  $\alpha = 2$ , see Fig. 3. The average of precision and sensitivity for these parameters are equal to 0.15 and 0.70, respectively.

3) *Anisotropic sample variance approach*: The parameter to be specified for this feature is  $m \in \{2, 4, 6, 8, 10\}$ .

The best value of  $m$  is equal to 8. Precision and sensitivity for this parameter are equal to 0.07 and 0.80, respectively. The performance of this feature is shown in Fig. 4. It can be seen that values inside the lesion area are smaller than outside, and after normalization this difference is still kept noticeable.

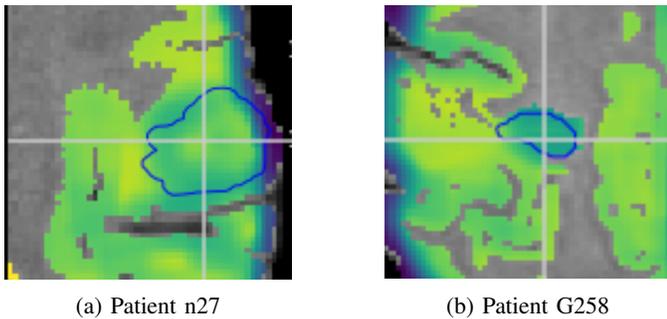


Fig. 4: Example of lesion detection with sample variance. Lesion areas are outlined with a blue line.

#### D. Metrics on features

1) *Cortical thickness, sulcal depth, curvature*: Patient-averaged voxel-level metrics for morphological features are given in Table I. All metrics were calculated for features without post-processing. It can be seen that the cortical thickness gave the highest result. However, even this result is small compared with proposed features. A patient-level metric was calculated for features that undergo post-processing procedure described in Section II-E. According to this metric, the lesion is in the top 10 bright clusters for 34%, 13% and 9% of patients for cortical thickness, curvature and sulcal depth respectively.

The average computational time needed to calculate cortical thickness, curvature and sulcation within MRIs is 2 hours as a part of freesurfer recon-all pipeline.

2) *Blurring feature*: It was calculated using formulas (2) and (3). The parameter  $\delta$  for T1w was chosen from 0.1 to 0.5 with step 0.1 and for T2w and Flair from 0.01 to 0.05 with step 0.01. The best coefficients according to the results of the experiments were 0.4, 0.03 and 0.04 for T1w, T2w and Flair, respectively. Further, normalization with different quantiles was performed for the obtained maps. The precision of blurring features, depending on the selected quantile, is presented in Table II. It can be seen that the quantile normalization gives better results than the standard mean normalization.

All metrics were calculated for features after normalization and post-processing. Voxel-level metrics (Table I) show good results. But it seems that the blurring calculated on T1w is superior. However, there are patients whose T2w-Blurring or T2Flair-Blurring give better results (see an example in Fig. 6).

According to patient-level metrics, the lesion is in the top 10 bright clusters for 71%, 43% and 23% of patients for blurring on T1w, T2w and Flair, respectively. Moreover, it can be seen for which patients a lesion got into the top 10 clusters in Fig. 5.

The average CPU time is needed to process the blurring feature with normalization is about 15 seconds.

3) *Concentration Rate*: Voxel-level metrics for CR feature are shown in Table I. This feature on Flair images gave the highest average precision compared to all tested features. Although T2Flair-CR gave better voxel-level metrics, there are several patients whose T2w-CR has better quality than T2Flair-CR feature, see examples in Fig. 7.

For 31% and 54% of patients, a lesion is in top 10 clusters for CR on T2 and Flair, respectively.

TABLE I: Patient-averaged voxel-level metrics

| Feature        | Precision | DSC   | Sensitivity | Specificity |
|----------------|-----------|-------|-------------|-------------|
| Blurring T1    | 0.158     | 0.100 | 0.149       | 0.998       |
| Blurring T2    | 0.062     | 0.070 | 0.189       | 0.998       |
| Blurring Flair | 0.019     | 0.029 | 0.208       | 0.994       |
| CR T2          | 0.053     | 0.031 | 0.068       | 0.675       |
| CR Flair       | 0.179     | 0.107 | 0.119       | 0.677       |
| Entropy        | 0.132     | 0.127 | 0.207       | 0.931       |
| Variance       | 0.060     | 0.033 | 0.032       | 0.767       |
| Thickness      | 0.013     | 0.007 | 0.658       | 0.869       |
| Sulc           | 0.003     | 0.006 | 0.363       | 0.962       |
| Curv           | 0.002     | 0.004 | 0.265       | 0.965       |

TABLE II: Dependence of precision on quantiles after normalization

| Quantiles     | Blurring |       |       |
|---------------|----------|-------|-------|
|               | T1       | T2    | Flair |
| mean          | 0.109    | 0.005 | 0.011 |
| (0.9,0.1)     | 0.125    | 0.032 | 0.016 |
| (0.925,0.075) | 0.135    | 0.036 | 0.017 |
| (0.95,0.05)   | 0.151    | 0.031 | 0.018 |
| (0.975,0.025) | 0.160    | 0.044 | 0.019 |
| (0.98,0.02)   | 0.158    | 0.062 | 0.019 |

On average, it takes 61 seconds to calculate this feature with subsequent normalization.

4) *Entropy*: It can be seen from Table I, that the entropy feature yields best results in terms of Dice metric. Moreover, the lesion was in the top 10 bright clusters for 39% of patients. The MRI with high value of entropy feature is seen in Fig. 8.

The calculation of this feature takes as much as 6 thousand seconds on average. This is due to the computationally expensive calculation of the GLC matrix.

5) *Variance*: Despite the low voxel-level metrics (Table I) compared to other proposed features, the variance feature showed results on those patients where other features did not work (Fig. 5). An example is patient n43 whose variance feature is in Fig. 8. According to patient-level metric, the lesion was in the top 10 bright clusters for 37% of patients.

The calculation of this feature takes about 92 seconds on average.

#### E. Results for ANN model and XGBoost model

In this section, we are addressing two questions: (1) whether different models, operating on multimodal MRI, have significant segmentation diagnostic value and (2) how established feature maps impact the classification performance of the models.

1) *ANN performance*: To evaluate the diagnostic ability of the deep-learning model, we assess stratified  $k$ -fold cross-validation metrics with the number of folds  $k = 9$  within FCD lesion localization. Our model is trained on the subset  $\mathcal{F}$  of features considered in the paper plus the original T1, T2 and Flair images and WM and GM segmentation maps. Respective metrics are reported in Table III.

A model trained on raw images only performs poorly. Voxel level DSC, precision metrics and patient-wise sensitivity are low. This can be explained by the simplicity of the models.

A model trained on all features demonstrated detection of confirmed FCDs in 7 of 10 subjects on average, therefore, the



Fig. 5: Indication that a lesion is found in top 10 brightest clusters for a patient. Points are the marks that top 10 brightest clusters intersect with lesion for a patient.

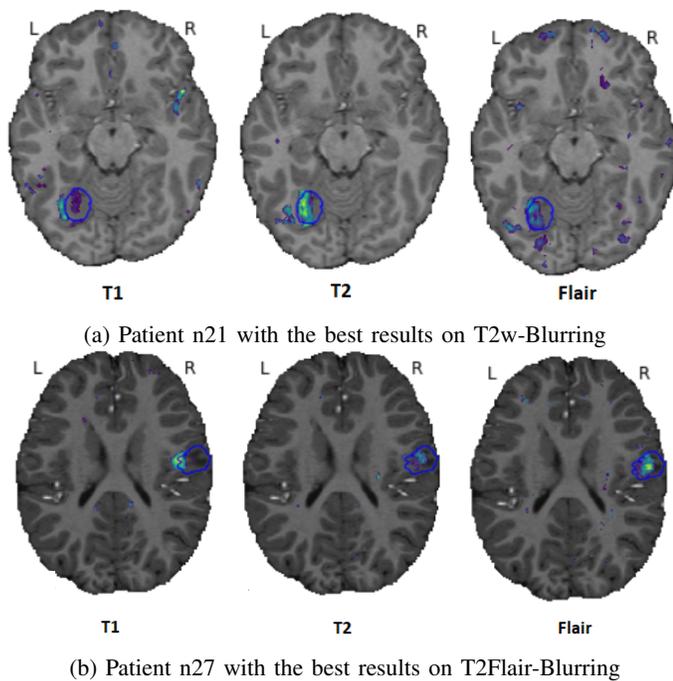


Fig. 6: Lesion detection with T2w-Blurring and T2Flair-Blurring. Lesion areas are outlined with a blue line.

patient-wise sensitivity was around 70%. Voxel-level metrics of the model obtained as mean results of the 9-fold cross validation (Precision = 0.231, DSC = 0.209, Sensitivity = 0.691) indicate relatively high sensitivity and DSC metrics.

In Fig. 9 we demonstrate representative examples of FCD detection by the ANN model in the validation subset. Blue contours mark lesions detected by a model.

2) *XGBoost performance*: To train XGBoost model [4], we assess  $k$ -fold cross validation metrics with the number of folds  $k = 9$ . Our model is trained on the subset  $\mathcal{F}$  of features considered in the paper plus the original T1, T2 and Flair images and WM and GM segmentation maps. Early stopping was applied to interrupt the training process whilst monitoring the balanced accuracy value. We used half of training points as validation set for early stopping. Voxel-wise metrics are

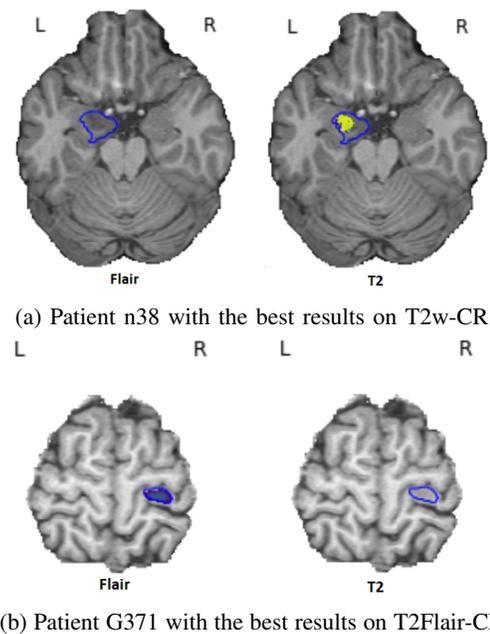


Fig. 7: Lesion detection with T2w-CR and T2Flair-CR. Lesion areas are outlined with a blue line.

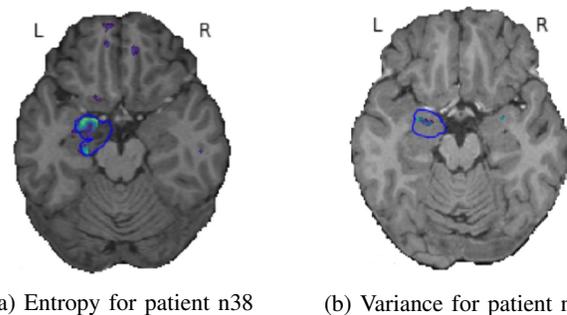


Fig. 8: Good examples of lesion detection with Entropy and Sample Variance. Lesion areas are outlined with a blue line.

reported in Table IV.

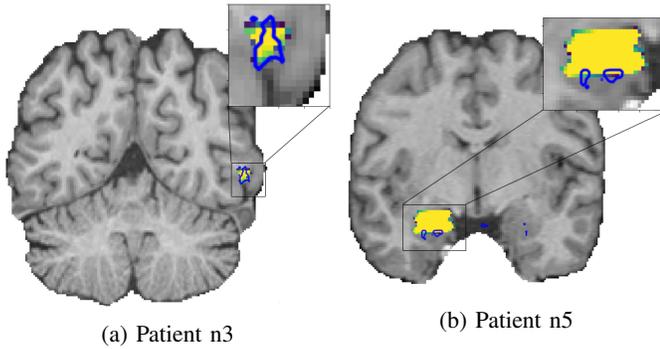
We can see that patient-wise sensitivity is higher than for MLP results, while voxel-wise metrics are lower. It is due to

| Model input | Precision | DSC   | Sensitivity | Patient sensitivity | FP        |
|-------------|-----------|-------|-------------|---------------------|-----------|
| T1 only     | 0.076     | 0.081 | 0.870       | 28%                 | $8 \pm 4$ |
| T1 + all    | 0.231     | 0.209 | 0.691       | 71%                 | $2 \pm 2$ |

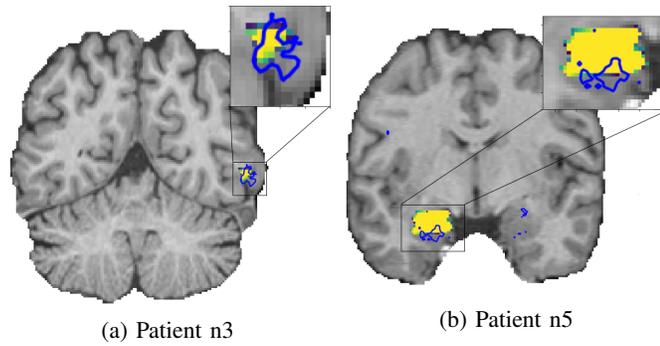
**TABLE III:** Cross validation with 9 folds metrics for MLP trained with different inputs: with T1, T2, Flair only and with all available features, including T1, T2 and Flair. FP is a number of false positive clusters.

| Precision | DSC   | Sensitivity | Patient sensitivity | FP        |
|-----------|-------|-------------|---------------------|-----------|
| 0.265     | 0.221 | 0.663       | 78%                 | $3 \pm 3$ |

**TABLE IV:** Cross validation with 9 folds metrics for XGBoost model trained with all available features, including T1, T2 and Flair. FP is a number of false positive clusters.



**Fig. 9:** Slices of the segmentation examples from validation subset. Yellow is true label, blue contour is a prediction of MLP model.

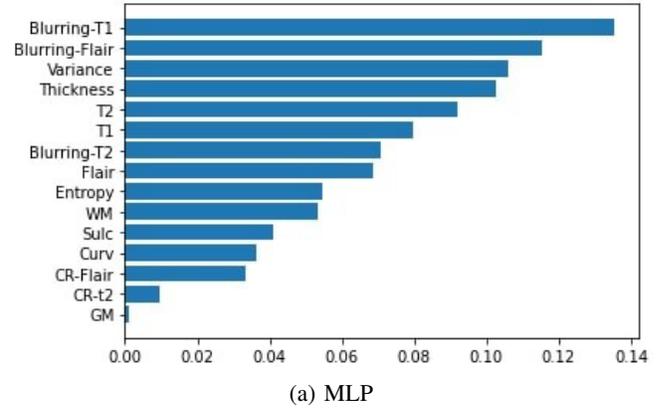


**Fig. 10:** Slices of the segmentation examples from validation subset. Yellow is true label, blue contour is a prediction of XGBoost classifier.

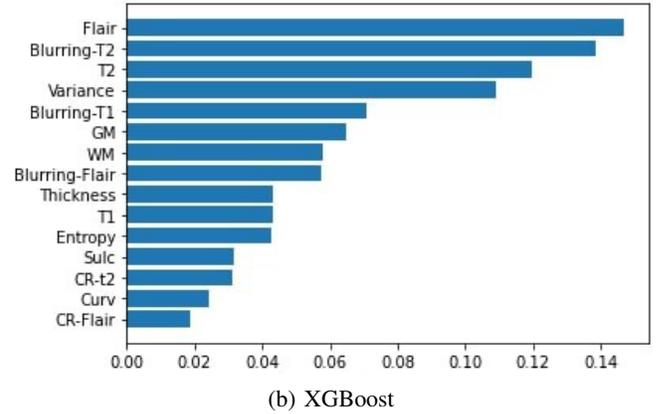
the fact that the mean number of false positive clusters for XGBoost predictions is higher than for MLP predictions. The examples of XGBoost predictions can be seen in Fig. 10.

**3) Feature importance:** Locally Interpretable Model Agnostic Explanations (LIME [28]) is a model-agnostic explanation method that approximates any black box model with a local, interpretable model to explain each individual prediction. We used LIME to find features importance for both models. Because LIME explains a model locally, we calculated importances of all features as the mean absolute value of the LIME values for validation points.

MLP and XGBoost feature importances are given in Fig. 11.



(a) MLP



(b) XGBoost

**Fig. 11:** Feature importance.

It can be seen that the most important features for both models include Blurring and Variance features. Also, all of proposed features are valuable for models.

#### IV. DISCUSSION

Although magnetic resonance imaging is currently the most widely used diagnostic tool, the detection of FCD lesions using MRI is still unsatisfactory. As a result, there is a need for more accurate and efficient diagnostic methods. In our study we evaluated methods for automatic FCD detection based on features analysis and machine learning (ML) techniques. To create MR images feature maps, we collected high-quality MR images sequences (3D T1, Flair and T2 images) that suffice ILAE Neuroimaging Task Force recommendations on Harness-Mri protocol. To train ML models, we collected 3D volumetric FCD lesion segmentations. Extracted features demonstrated significant detection performance of FCD lesions, however, we noticed the improvement of patient-wise sensitivity of XGBoost classifier (78%) compared to best performing Blurring detection map (71%) computed on T1w image. Patient-wise feature maps specificity is high (0.99) and demonstrates FCD detection in mostly top 10 brightest clusters, this is comparable to the results of XGboost with a tendency to result  $3 \pm 3$  false positives. However, using feature maps requires normalization and postprocessing with heuristic arguments that could interfere the clinical workflow integration.

The results indicate that the introduced features can be easily interpreted and have substantial diagnostic relevance for doctors. Furthermore, the focus on contrast features in this work is a key difference from recent studies related to FCD that focus on deep learning models. Also, we showed that classification models like MLP and XGBoost classifier overall sensitivity was 71% and 78% respectively.

In the future work, we plan to validate proposed statistical features on collected two-site dataset (n=203) with manually voxel-wise segmented FCD lesions by two experienced radiologists. This validation will help to further improve the efficiency and accuracy for automatic FCD detection in patients with focal epilepsy.

In conclusion, the results of our study suggest that the use of machine learning algorithms as for automatic FCD detection based on contrast feature maps is a promising approach. With further validation and refinement, these methods could become valuable tools for improving the diagnosis and treatment of patients with focal epilepsy. The development of accurate and efficient diagnostic methods for FCD has the potential to greatly improve patient outcomes and quality of life.

## REFERENCES

- [1] B. Ahmed, C. E. Brodley, C. E. Blackmon, R. Kuzniecky, G. Barash, C. Carlson, B. T. Quinn, W. Doyle, J. A. French, O. Devinsky, and T. Thesen. Cortical feature analysis and machine learning improves detection of "MRI-negative" focal cortical dysplasia. *Epilepsy & Behavior*, 48:21–28, 2015.
- [2] A. Carass, S. Roy, A. Gherman, J. S. Reinhold, A. Jesson, T. Arbel, O. Maier, H. Handels, M. Ghafoorian, B. Platel, A. Birenbaum, H. Greenspan, D. L. Pham, C. M. Crainiceanu, P. A. Calabresi, J. L. Prince, W. L. G. Roncal, R. T. Shinohara, and I. Oguz. Evaluating white matter lesion segmentations with Refined Sørensen-Dice analysis. *Scientific Reports*, 10(1), May 2020.
- [3] F. Cendes, W. H. Theodore, B. H. Brinkmann, V. Sulc, and G. D. Cascino. Neuroimaging of epilepsy. *Handbook of clinical neurology*, 136:985–1014, 2016.
- [4] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.
- [5] J.-P. Chilès and P. Delfiner. *Geostatistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2012. Modeling spatial uncertainty.
- [6] R. D. Chitalia and D. Kontos. Role of texture analysis in breast MRI as a cancer biomarker: A review. *Journal of Magnetic Resonance Imaging*, 49(4):927–938, 2019.
- [7] B. David, J. Kröll-Seger, F. Schuch, J. Wagner, J. Wellmer, F. Woermann, B. Oehl, W. Paesschen, T. Breyer, A. Becker, H. Vatter, E. Hattungen, H. Urbach, B. Weber, R. Surges, C. Elger, H.-J. Huppertz, and T. Rüber. External validation of automated focal cortical dysplasia detection using morphometric analysis. *Epilepsia*, 62, 02 2021.
- [8] N. M. Dipu, S. A. Shohan, and K. Salam. Deep learning based brain tumor detection and classification. pages 1–6, 2021.
- [9] B. Fischl, M.I. Sereno, and A.M. Dale. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2):195 – 207, 1999.
- [10] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale vessel enhancement filtering. In *International conference on medical image computing and computer-assisted intervention*, pages 130–137. Springer, 1998.
- [11] J. L. Gastwirth. The estimation of the Lorenz curve and Gini index. *Rev. Econom. Statist.*, 54:306–316, 1972.
- [12] R. S. Gill, H.-M. Lee, B. Caldirou, S.-J. Hong, C. Barba, F. Deleo, L. D'Incerti, V. C. M. Coelho, M. Lenge, M. Semmelroch, D. V. Schrader, F. Bartolomei, M. Guye, A. Schulze-Bonhage, H. Urbach, K. C. Cho, F. Cendes, R. Guerrini, G. Jackson, R. E. Hogan, N. Bernasconi, and A. Bernasconi. Multicenter validation of a deep learning detection algorithm for focal cortical dysplasia. *Neurology*, 97(16):e1571–e1582, September 2021.
- [13] S. Hong, B. C. Bernhardt, D. Schrader, N. Bernasconi, and A. Bernasconi. Whole-brain MRI phenotyping in dysplasia-related frontal lobe epilepsy. *Neurology*, 86:643 – 650, 2016.
- [14] S. Hong, H. Kim, D. Schrader, N. Bernasconi, B. C. Bernhardt, and A. Bernasconi. Automated detection of cortical dysplasia type II in MRI-negative epilepsy. *Neurology*, 83:48 – 55, 2014.
- [15] H.-J. Huppertz, C. Grimm, S. Fauser, J. Kassubek, I. Mader, A. Hochmuth, J. Spreer, and A. Schulze-Bonhage. Enhanced visualization of blurred gray–white matter junctions in focal cortical dysplasia by voxel-based 3D MRI analysis. *Epilepsy research*, 67(1-2):35–50, 2005.
- [16] B. Jin, B. Krishnan, S. Adler, K. Wagstyl, W. Hu, S. E. Jones, I. M. Najm, A. V. Alexopoulos, K. Zhang, J. Zhang, M. Ding, Sh. Wang, and Z. I. Wang. Automated detection of focal cortical dysplasia type II with surface-based magnetic resonance imaging postprocessing and machine learning. *Epilepsia*, 59:982 – 992, 2018.
- [17] J. Kassubek, F. Juengling, J. Jüngling, B. Hellwig, J. Spreer, and C. H. Lücking. Thalamic gray matter junctions in unilateral parkinsonian resting tremor: a voxel-based morphometric analysis of 3-dimensional magnetic resonance imaging. *Neuroscience Letters*, 323:29–32, 2002.
- [18] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.
- [19] D. A. Lee, B. J. Kim, H. Lee, S. E. Kim, and K. M. Park. Network characteristics of genetic generalized epilepsies: Are the syndromes distinct? *Seizure*, 82:91–98, 2020.
- [20] D. A. Lee, H. Lee, H. C. Kim, and K. M. Park. Alterations of structural connectivity and structural co-variance network in focal cortical dysplasia. *BMC Neurology*, 21, 2021.
- [21] E. Meijering, M. Jacob, J.-C. F. Sarría, P. Steiner, H. Hirling, and M. Unser. Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry Part A*, 58A(2):167–176, 2004.
- [22] C.-C. Ng, M. H. Yap, N. Costen, and B. Li. Automatic wrinkle detection using hybrid Hessian filter. In *Lecture Notes in Computer Science*, page 609–622. Springer, 2015.
- [23] L.G. Nyul, J.K. Udupa, and Xuan Zhang. New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, 19(2):143–150, 2000.
- [24] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.
- [25] R. Pienaar, B. Fischl, V. Caviness, N. Makris, and P. E. Grant. A methodology for analyzing curvature in the developing brain from preterm to adult. *Imaging Syst. Technol.*, 18:42 – 68, 2008.
- [26] A. Rényi. A few fundamental problems of information theory. *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.*, 10:251–282, 1960.
- [27] A. Rényi. On measures of entropy and information. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 547–561. Univ. California Press, Berkeley, Calif., 1961.
- [28] M. T. Ribeiro, S. Singh, and C. Guestrin. "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1135–1144, 2016.
- [29] A. Suresh and K. Shunmuganathan. Image texture classification using gray level co-occurrence matrix based statistical features. *European Journal of Scientific Research*, 75:591–597, 04 2012.
- [30] H. Wackernagel. *Multivariate Geostatistics: An Introduction with Applications*. Springer Berlin Heidelberg, 2013.
- [31] Z.I. Wang, A.V. Alexopoulos, Z. Jones, S.E. Jaisani, I.M. Najm, and R.A. Prayson. The pathology of magnetic-resonance-imaging-negative epilepsy. *Modern Pathology*, 26:1051–1058, 2013.
- [32] S. Yitzhaki and E. Schechtman. *The Gini methodology*. Springer Series in Statistics. Springer, New York, 2013. A primer on a statistical methodology.
- [33] R. Zhang and A. Chung. Medq: Lossless ultra-low-bit neural network quantization for medical image segmentation. *Medical Image Analysis*, 73:102200, 2021.