Advances in Artificial Intelligence and Machine Learning; Research 2 (3) 422-440 Received 13-06-22; Accepted 29-06-22; Published 07-07-22

Lesion Segmentation in Paediatric Epilepsy Utilizing Deep Learning Approaches

Azad Aminpour azad.aminpour@ontariotechu.ca
Faculty of Science, Ontario Tech University,
Oshawa, Ontario,
Canada

Mehran Ebrahimi mehran.ebrahimi@ontariotechu.ca
Faculty of Science, Ontario Tech University,
Oshawa, Ontario,
Canada

Elysa Widjaja elysa.widjaja@sickkids.ca
The Hospital for Sick Children (SickKids),
Toronto, Ontario,
Canada

Corresponding Author: Mehran Ebrahimi

Copyright © 2022 Aminpour A, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Focal cortical dysplasia (FCD) is one of the most common lesions responsible for drug-resistant epilepsy, and is frequently missed by visual inspection. FCD may be amenable to surgical resection to achieve seizure freedom. By improving lesion detection the surgical outcome of these patients can be improved. Image processing techniques are a potential tool to improve the detection of FCD prior to epilepsy surgery. In this research, we propose and compare the performance of two type of models, Fully Convolutional Network (FCN) and a multi-sequence FCN to classify and segment FCD in children with drug-resistant epilepsy. This experiment utilized the volumetric T1-weighted, T2 weighted and FLAIR sequences. The whole slice FCN models were applied to each sequence separately while the multi-sequence model leverages combined information of all three sequences simultaneously. A leave-one-subject-out technique was utilized to train and evaluate the models. We evaluated subjectwise sensitivity and specificity, which corresponds to the ability of the model to classify those with or without a lesion. We also evaluated lesional sensitivity and specificity, which expresses the ability of the model to segment the lesion and the dice coefficient to evaluate lesion coverage. Our data consisted of 80 FCD subjects (56 MR-positive and 24 MR-negative) and 15 healthy controls. Performance of whole slice FCN was best on T1-weighted, followed by T2-weighted and lowest with FLAIR sequences. Multi-sequence model performed better than the T1 whole slice FCN, and detected 98% vs. 93% respectively MR-positive cases, and 92% vs. 88% respectively MR-negative cases, as well as achieved lesion coverage of 74% vs. 67% respectively for MR-positive cases and 68% vs. 64% for MR

Citation: Aminpour A, et al. Lesion Segmentation in Paediatric Epilepsy Utilizing Deep Learning Approaches. Advances in Artificial Intelligence and Machine Learning. 2022;2(3):28.

negative cases. The dice coefficient for the multi-sequence model was 57% and for whole slice FCN was 56% for MR-positive cases. In the test cohort of six new cases, the multi-sequence model detected 4 out of 6 cases where the predicted lesion had 56% overlap with the actual lesion. This work showed that deep learning methods in particular fully convolutional networks are a promising tool for classification and segmentation of FCD. Additional work is required to further improve lesion classification and segmentation, particularly for small lesions, as well as to train and test optimal algorithms on a larger multi-center dataset.

Keywords: Epilepsy, Focal Cortical Dysplasia (FCD), Convolutional Neural Network (CNN), Segmentation, Multi-sequence MRI

1. INTRODUCTION

Epilepsy is one of the most common neurological disorders in children with an incidence rate of 4 to 9 in 1000 per year [1,2]. Drug-resistant epilepsy is a serious neurological disorder, which can lead to detrimental effects on children's cognition and psychosocial development [3,4], and 5 to 9 times higher mortality rate than the general population [5-7]. Focal cortical dysplasia (FCD), a brain malformation, is one of the most common lesions responsible for drug-resistant epilepsy in children. The MRI features of FCD are frequently subtle and may be overlooked or may not be detected by visual assessment. A lesion is seen on MRI in 30% to 85% of patients with drug resistant epilepsy [8]. In other words, 15% to 70% of patients do not have an abnormality reported on MRI. The patients who do not have an abnormality reported on MRI could have a subtle lesion that is not detected by the radiologist. Patients in whom a lesion is not detected on pre-surgical evaluation with MRI are considered to have MR-negative focal epilepsy (or non-lesional epilepsy). Up to 72% of patients with MR-negative epilepsy have FCD reported on histopathology [9-13].

Children with focal drug-resistant epilepsy may be treated with epilepsy surgery to remove the lesion responsible for epilepsy. Patients who have MR-negative epilepsy have poorer seizure-free epilepsy surgery outcome [14], and may have increased use of invasive electroencephalography (EEG) monitoring for surgical planning. Therefore, it is critical to improve our ability to detect FCD, as discovering a previously undetected lesion can increase the success of surgery in curing focal drug-resistant epilepsy, and change pre-surgical planning of epilepsy surgery. Visual assessment of MRI is subjective and highly dependent on the expertise of the observer. Thus there is a need for a more advanced and objective tool for analyzing the MRI data to improve detection of FCD. Machine learning algorithms offer the potential to detect subtle structural changes, which may not be identifiable on visual inspection of MRI.

Recently, artificial intelligence (AI) techniques based on Deep Neural Networks (DNN) have emerged and been applied to many fields, such as computer vision and natural language processing [15,16]. Deep neural network methods also have been applied in various fields of medicine, including identifying lesions on medical imaging [17]. In general, DNN works on vector shaped inputs. However, a particular type of DNN has been proposed where it can take images as input and learn kernels or filters to extract features out of the image,

then performs classification or regression to get the desired output. These deep networks are called “Convolutional Neural Networks (CNN)” [18].

In this study, we applied two types of deep learning-based methods, Fully Convolutional Network (FCN) and a multi-sequence FCN for the task of FCD detection and localization. FCN is a special type of Convolutional Neural Network (CNN) that takes an input image and produces the same size output. It is used in the image-to-image translation problems such as segmentation. The whole slice FCN models were applied to the 2D MRI slices of T1, T2 and FLAIR sequences separately while the multi-sequence model combined information from MRI slices of T1, T2 and FLAIR weighted sequences simultaneously.

2. LITERATURE REVIEW

Previous studies have utilized a variety of image processing algorithms, including voxel-based morphometric analysis [25,26] or surface-based algorithms to detect FCD [23,24,27-29]. These algorithms extracts various features of FCD such as cortical thickness, gray-white junction, sulcal depth, and cortical fold, and subsequently integrating these features within a classical machine learning framework for the classification of FCD. These approaches require domain experts to extract the required features. It is time-intensive to post-process the data, and errors in processing are propagated throughout the algorithm. We have chosen state-of-the-art deep learning approach using CNN to overcome some of the limitations of classical computer aided tools to identify FCD, which does not require feature extraction and could learn optimal features automatically without human intervention.

The performance of recent classical works is summarized in TABLE 1. The first column is the work and the year. The second column shows their extracted features out of the MRIs along with the classification algorithm they have used for FCD detection. The last column illustrates their results in terms of subjectwise specificity and sensitivity.

More recently, a few studies have utilized deep learning approaches to detect FCD [30-34]. Gill et al. [34], trained two networks on patches extracted from MRI in patients with FCD. The algorithm was trained on volumetric T1 and FLAIR MRI data from a single site and evaluated on independent data from seven sites. Two networks were developed, consisting of three stacks of convolution and max-pooling layers with 48, 96 and 2 feature channels and $3 \times 3 \times 3$ kernels. The first network is trained to recognize lesional voxels and the second network is trained to reduce the number of misclassified voxels. The authors found that the classifier showed excellent sensitivity (91%, 61/67 lesions detected) and specificity (95%, no findings in 36/38 healthy controls). David et al. [33], have combined conditional generative adversarial networks (cGAN) with 3D CNN and showed that the addition of cGAN to CNN can improve the sensitivity of subject-wise classification from 81% to 93%, and the specificity from 71% to 96%. Wang et al. [31], have used patch-wise CNN with five convolutional layers, a max pooling layer and two fully connected layers to evaluate FCD, and found that this technique successfully classified 9 out of 10 FCD cases. These studies have used CNN to classify subjects as lesional or non-lesional but have not evaluated the performance of CNN for localizing the lesion. Feng et al. [30], have used a six-layer CNN, consisting of two convolutional layers, two pooling layers and two fully connected layers,

Table 1: Summary of some the classical approaches to problem of FCD detection and localization. The first column is the work and the year. The second column shows their features along with the classification algorithm. The last column illustrates their results in terms of subjectwise specificity and sensitivity.

Author and Year	Method	Results
Antel et al. (2003) [19]	Cortical thickening, blurring of GW junction, gray level hyperintensity and textural features. Bayesian classifier and Fishers discriminant ratio.	Specificity of 83% Sensitivity 100%
Yang et al. (2011) [20]	Statistical features on cortical thickness. Naïve Bayes classifier.	Specificity of 62% Sensitivity 81%
Strumia et al. (2013) [21]	Textural features such as cortical thickness, and spatial tissue maps. Naïve Bayes classifier.	Specificity of 95% Sensitivity 100%
Ahmed et al. (2014) [22]	Patches extracted from coarse to fine resolutions of the image. Random forest.	MR Pos Specificity of 90% MR Neg Specificity 80%
Adler et al. (2017) [23]	Morphometric and textural features. Neural network.	Specificity of 73%
Kulaseharan et al (2019) [24]	Morphometric and textural features. Two-step Bayesian classifier.	Specificity of 94% Sensitivity 100%

in conjunction with activation maximization and convolutional localization to classify and localize FCD on FLAIR sequence in 12 MRI-negative epilepsy patients. The authors reported subject-wise recall or detection rate of 83% to 100%, and dice coefficient of 53% to 71% for localizing FCD.

Bijay et al. [32], have also applied CNN to the FLAIR sequence in 43 subjects with FCD and found subject-wise recall of 83%, and dice coefficient of 52% for localizing FCD. Thomas et al. [35], proposed a multi-resolution attention based model which is based on U-net architecture [36]. They have used ResPaths [37], and Attention Gates [38], to overcome the segmentation gap between the decoder and encoder parts of U-net. They used 26 FCD subjects where they yielded 92% subject-wise sensitivity with lesion coverage of 60%. Previous publications based on CNN to detect FCD have been applied to adults or mixed adults and pediatric population.

The results for the recent deep learning-based works are presented in TABLE 2. These numbers are on different datasets used by each team, and to the best of our knowledge, they used MR-positive cases only except for Feng et al. [30], where they have considered both MR-positive and MR-negative.

Most of the proposed deep learning-based methods for FCD detection are using one sequence (such as FLAIR or T1) at a time. In this paper, we propose integrating information from different sequences simultaneously. The use of several sequences has been explored previously by Gill et al. [34]. However, they used each sequence consecutively, where a network is trained on one sequence and then it is trained on another sequence and their outputs are

Table 2: FCD detection and localization performance

MR-positive Subjects					
	Subjectwise Sens.	Subjectwise Spec.	Lesional Sens.	Lesional Spec.	Dice
David et al.[33]	93	96	-	-	-
Gill et al. [34]	91	95	-	-	-
Wang et al. [31]	90	85	-	-	78
Feng et al. [30]	100	-	59	99	71
Bijay et al. [32]	83	-	40	-	52
Thomas et al. [35]	92	-	60	-	62
MR-negative Subjects					
	Subjectwise Sens.	Subjectwise Spec.	Lesional Sens.	Lesional Spec.	Dice
Feng et al. [30]	83	-	51	99	53

combined at the end. Here we plan to incorporate information from various sequences at the training phase simultaneously, using a multi-sequence model. We have also explored the use of different sequences independently through a two dimensional FCNs.

3. MATERIALS AND METHODS

3.1 Data

Our dataset consisted of T1-weighted, T2 weighted and FLAIR sequences of paediatric subjects. The MRI data was acquired from the Hospital for Sick Children. This study has the approval of Research Ethics Board. the dataset included 80 subjects with suspected FCD on MRI (56 MR-positive and 24 MR-negative), and 15 healthy controls. The mean age of the MR-positive FCD patients was 11.15 years with standard deviation (SD) 3.22 years; there are 36 males and 20 females. For MR-negative cases, we had 14 male and 10 female subjects with mean age of 11.71 (SD=30.78) years. For healthy controls, the mean age was 13.87 (SD=2.94) years, including 6 males and 9 females. The baseline characteristics of the FCD subject's MRI data are presented in TABLE 3. Patients and controls underwent MRI on Philips 3 T scanner (Philips Medical System, Best, Netherlands) using 8-channel head coil with the same imaging parameters. Patients underwent high-resolution epilepsy protocol, which included volumetric T1-weighted (TR/TE = 4.9/2.3 msec, ST = 0.9 mm, FOV = 22 cm, matrix = 220 × 220), axial and coronal FLAIR (TR/TE = 10,000/140 msec, ST = 3 mm, FOV = 22 cm, matrix = 316 × 290) and axial and coronal T2-weighted (TR/TE = 4200/80 msec, ST = 3 mm, FOV = 22 cm, matrix = 400 × 272). All patients underwent epilepsy surgery resection and had post operative Computed Tomography (CT) or volumetric T1-weighted MRI on the same scanner. All controls underwent volumetric T1- weighted, axial T2- weighted, and axial FLAIR imaging, using the same parameters as for the patients.

Table 3: Characteristics of MRI data, lesion or surgery location and surgical outcome, SD is the standard deviation.

	MR-positive	MR-negative
Mean age (SD) in years	11.15 (3.22)	11.71 (3.78)
Sex # (%)		
Male	36 (64.29 %)	14 (58.33 %)
Female	20 (35.71 %)	10 (41.67 %)
Mean age at seizure onset (SD) in years	5.57 (3.95)	6.02 (4.03)
Seizure frequency, # (%)		
Daily	27 (48.21 %)	14 (58.33 %)
Weekly	22 (39.29 %)	10 (41.67 %)
Monthly	6 (10.71 %)	0 (0 %)
3-Monthly	1 (1.79 %)	0 (0 %)
Mean number of anti-seizure medications (SD)	2.04 (0.85)	2.33 (0.64)
Location of lesion or surgery,# (%)		
Frontal	23 (41.07 %)	10 (41.33 %)
Temporal	19 (33.93 %)	7 (29.17 %)
Parietal	10 (17.86 %)	5 (20.83 %)
Others	4 (7.14 %)	2 (8.33 %)
ILAE surgical outcome, # (%)		
I	41 (73.21 %)	17 (70.83 %)
II	3 (5.36 %)	4 (16.67 %)
III	4 (7.14 %)	1 (4.17 %)
IV	6 (10.71 %)	1 (4.17 %)
V	2 (3.57 %)	1 (4.17 %)
Histology, # (%)		
FCD I	4 (7.14 %)	7 (29.17 %)
FCD II	32 (57.14 %)	3 (12.5 %)
Oligodendrogliosis	10 (17.86 %)	2 (8.33 %)
Subpial gliosis	9 (16.07 %)	9 (37.5 %)
No abnormality/ others	1 (1.79 %)	3 (12.5 %)

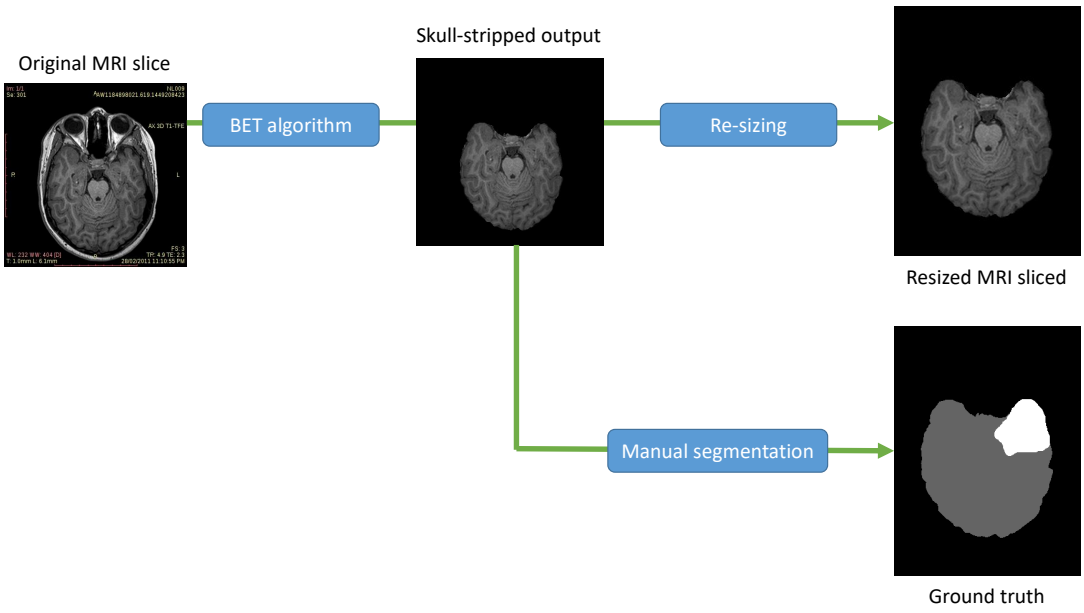


Figure 1: Data pre-processing flow chart.

3.2 Pre-processing

We performed brain extraction on each subject to remove the non-brain tissue using BET (Brain Extraction Tool) algorithm proposed by Smith et. al. in 2002 [39], which is embedded in FSL [1] software package.

Next, each MRI voxel was labelled with three classes, including the background class label. For healthy controls, we created a binary mask where all the voxels with intensity are labelled as 1 and the rest as zero. In FCD subjects, the lesion was manually segmented by a paediatric neuroradiologist to provide the ground truth segmentation masks. To do the manual segmentation on the MRI volume, we utilized "segmentation editor" plug-in within ImageJ software package [2] [40]. The acquired segmentation masks from the neuroradiologist were used to create a 3-dimensional array the same size as each volume where we assigned label 2 to the voxels within the segmented part and 1 to the rest of valued voxels and zero to the background area.

We then resized all the slices to a standard size since deep learning frameworks are sensitive to the input size and are not able to process slices of different sizes during training. We could not stretch or compress the actual brain images due to the sensitivity of the data. Therefore, based on the largest brain size in all slices of all subjects, we resized the slices by adding or removing background voxels. The slice size we chose was 320×400, as it was the largest foreground (brain tissue) segment across all the slices in all MRI sequences. The pre-processing steps are illustrated in FIGURE 1.

[1] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL
[2] https://imagej.nih.gov/ij/

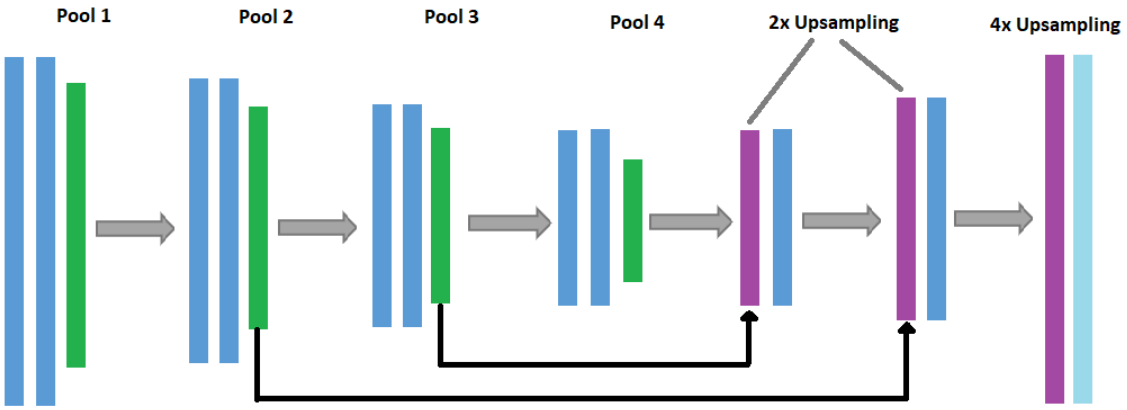


Figure 2: Two-dimensional whole-slice FCN architecture, blue rectangles are 3×3 convolutional layers, and green ones are pooling layers, purple rectangles are up-sampling layers. The black arrows represents the skip connections from pooling layers to the up-sampling layers.

3.3 Network Architecture

3.3.1 Whole-slice FCN

We built upon previous work based on whole-slice CNN for the current work using whole-slice FCN [41]. The feature extraction part of our proposed network is the same as the whole-slice network architecture where we have four blocks of two convolutions and pooling layers. Afterwards, the fully connected layers are replaced by three up-sampling blocks which have one convolutional layer and one up-sampling layer. Up-sampling layers [15], are used in FCNs to increase the spatial dimension of the activation maps. An up-sampling layer is used to produce dense pixel outputs from coarse inputs. Similar to a pooling layer, up-sampling layers do not have learnable parameters and use some interpolation techniques (e.g. nearest-neighbor or bilinear interpolation) to account for missing data. The first two up-sampling layers are "2x $Up - sampling$", where they increase inputs spatial size to twice their initial size. Both up-sampling layers are followed by one convolution layer with kernel size of 3×3. The last up-sampling layer is a "4x $Up - sampling$" layer Followed by 1×1 convolution layer. The network's architecture scheme is presented in FIGURE 2. ReLu activation function is following each layer in the architecture. We have also utilized skip layers [15], to get a more refined up-sampled output by adding the respective down-sampled feature map from the encoder part. Skip connection will improve the up-sampled output at each step which results in a more refined and detailed lesion segmentation. In addition, batch normalization and dropout were used in order to improve training and avoid over fitting.

3.3.2 Multi-sequence FCN

The proposed model consisted of two networks working together to extract features from all sequences and combine them and predict a segmentation map. The architecture is illustrated in FIGURE 3. The feature extraction network was trained on different sequences with subject

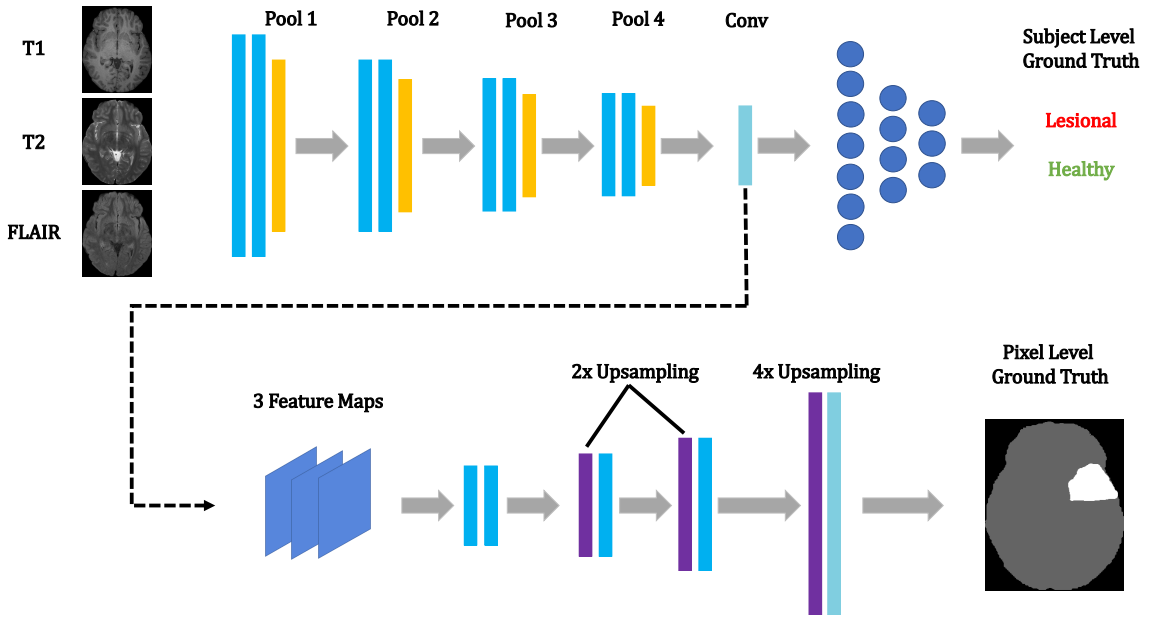


Figure 3: multi-sequence FCN architecture, blue rectangles are 3×3 convolutional layers, light blue rectangles are 1×1 convolutional layers, orange rectangles are 2×2 convolutional layers moving with stride of 2, and purple rectangles are representing upsampling layers.

level labels, then last convolutional layer's output was used as an input for the second network. The second network was an FCN that takes three feature maps of T1, T2 and FLAIR sequences, concatenates them and upsamples the input to the original size, in order to get the segmentation map. The feature extraction CNN's architecture is same as the whole-slice CNN [41], including four blocks of three convolution layers. In each block. The first two convolution were 3×3 filters with stride 1 and zero-padding and the last convolution layer had 2×2 filter sizes with stride 2 for downsampling. The blocks were followed by a 1×1 convolutional layer to decrease the feature maps depth to one. Then three fully connected layers to classify the extracted features into healthy or lesional classes. All layers except the last fully connected layer were using ReLu activation function and the last fully connected layer activation was the Sigmoid function. Dropout was utilized in the fully connected layers to prevent over-fitting, and batch normalization were used in convolutional layers. We used binary cross-entropy function as loss and optimized the model using SGD optimizer.

The upsampling FCN consisted of two convolutional layers followed by three blocks of upsampling. The convelutional layers had 128 filters of size 3×3 with stride of 1 and zero-padding. The first two blocks of upsampling were one 2x upsampling layer and one 3×3 convolutional layer. The last block consisted of one 4x upsampling layer followed by a 1×1 convolutional layer. ReLU was used as activation for all layers except the last convolutional layer which was using Sigmoid function instead. We applied batch normalization after the 3×3 convolutional layers.

Table 4: Confusion Matrix

		Actual	
		1-Positive	2-Negative
Prediction	1-Positive	TP	FP
	2-Negative	FN	TN

3.4 Methods of Measuring Performance

To evaluate the performance of our model, we computed the number of subjects who were correctly identified by the model as well as the number of mis-labelled subjects (TABLE 4) and then evaluated subjectwise sensitivity and subjectwise specificity. Then, we performed the same processes at pixel level to evaluate the segmentation output and subsequently assessed the lesional sensitivity and lesional specificity. Subjectwise sensitivity is the number of all FCD subjects correctly classified (true positive) divided by the total number of FCD subjects. Subjectwise specificity is the count of all control subjects in which no lesion is identified (true negative) divided by the total number of control subjects. Here are the definitions of sensitivity and specificity,

$$Sensitivity = \frac{TP}{TP + FN},$$
$$Specificity = \frac{TN}{TN + FP}.$$

Lesional sensitivity is measured as the sum of all abnormal tissue pixels labelled lesional by the model divided by the total number pixels labelled as lesional by a neuroradiologist in the ground truth masks. Lesional specificity is the sum of healthy tissue pixels in FCD subjects classified as normal divided by the total number of non-lesional pixels for FCD subjects. Lesional sensitivity and lesional specificity will only be calculated for subjects which the model labelled as FCD individually and the final results were averaged over all subjects.

We will also report the dice coefficient for the FCD subjects. Dice coefficient, is a statistical tool which measures the similarity between two sets of data. This index has arguably been the most commonly used method for evaluating segmentation algorithms. The index is calculated using the following equation,

$$Dice = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} = \frac{2 \times TP}{2 \times TP + FP + FN}.$$

4. RESULTS

As the number of subjects was not sufficient for a train-and-test set split, we applied a leave-one-subject-out technique for training and evaluation of both models. In this case, all of the

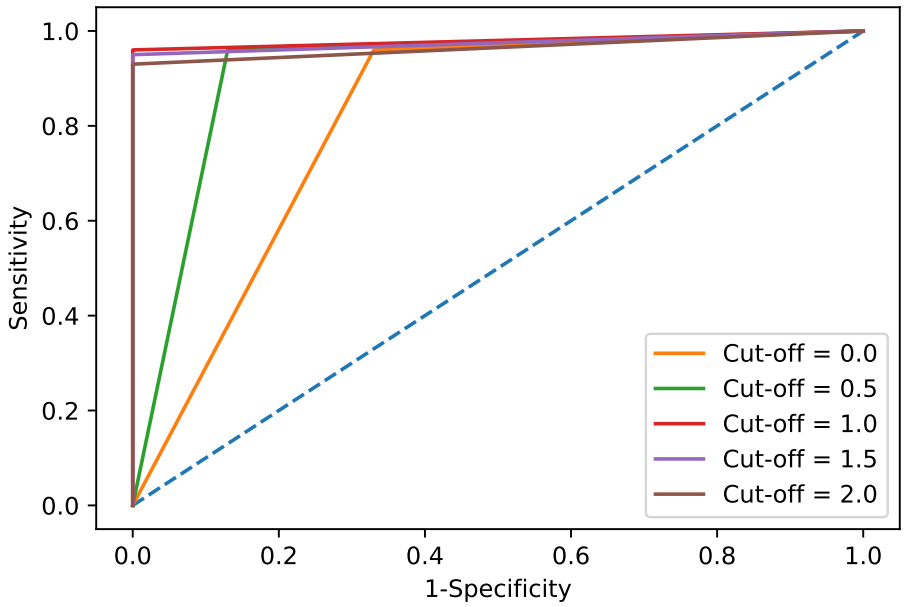


Figure 4: Subjectwise class label ratio cut-off ROC curves, horizontal axis shows (1-subjectwise specificity) and the vertical axis indicates the subjectwise sensitivity. The green curve has the largest AUC which indicates best ratio cut-off value.

subjects except for one, were used as training set and the remaining subject was used as the test subject to evaluate the model.

After obtaining the predictions, to determine the subject's label as either healthy control or FCD, we divided the number of lesional predicted pixels over the healthy predicted pixels and based on this ratio, labelled the subject as FCD or control. FIGURE 4 shows the Receiver Operating Characteristic (ROC) curve for various ratio cut-off values. The ratio cut-off value represents the threshold for defining the subjects as FCD or healthy controls. Subjects with values above the ratio cut-off value were labeled as FCD and thoes with values below the ratio cut-off values were labeled as controls. The threshold value of 1 percent has the largest area under the curve (AUC). Hence we chose 1% as the cut-off value. Then, for each FCD predicted subject we calculated the lesional sensitivity and lesional specificity.

TABLE 5 shows results for two dimensional FCN. We grouped the subjects based on MR-positive and MR-negative cases in TABLE 5 along with all FCD subjects. The model was trained and tested on T1, T2 and FLAIR separately. In case of T1-weighted, two dimensional FCN classified all healthy control cases as healthy. For T1-weighted sequence, the FCN detected 52 out of 56 MR-positive cases with lesion coverage of 67 percent (SD of 4) and dice coefficient of 56 (SD of 14) while detecting 21 out of 24 MR-negative subjects with 64 percent (SD of 3) coverage and dice coefficient of 55 (SD of 10). For T2-weighted sequence, two dimensional FCN model classified all healthy control cases as healthy, while for FLAIR sequences, the model predicted 14 out of 15 as healthy. For T2 sequence, the model detected

Table 5: Two dimensional FCN results for MR-positive and MR-negative subjects in percent (%). SD is the standard deviation.

T1-weighted			
	MR-positive	MR-negative	All Subjects
Subject-wise Sensitivity	93	88	91
Subject-wise Specificity	100	100	100
Lesional Sensitivity (SD)	67 (4)	64 (3)	66 (4)
Lesional Specificity (SD)	98 (2)	97 (1)	98 (2)
Lesional Precision (SD)	51 (19)	49 (14)	50 (18)
Dice (SD)	56 (14)	55 (10)	56 (13)
T2-weighted			
	MR-positive	MR-negative	All Subjects
Subject-wise Sensitivity	88	75	84
Subject-wise Specificity	100	100	100
Lesional Sensitivity (SD)	65 (6)	57 (5)	63 (6)
Lesional Specificity (SD)	97 (2)	96 (1)	97 (2)
Lesional Precision (SD)	36 (14)	33 (9)	35 (13)
Dice (SD)	44 (13)	40 (9)	43 (13)
FLAIR			
	MR-positive	MR-negative	All Subjects
Subject-wise Sensitivity	84	67	79
Subject-wise Specificity	93	93	93
Lesional Sensitivity (SD)	61 (6)	55 (3)	60 (6)
Lesional Specificity (SD)	97 (1)	96 (1)	97 (1)
Lesional Precision (SD)	38 (15)	26 (13)	35 (15)
Dice (SD)	45 (13)	34 (12)	42 (14)

Table 6: multi-sequence Segmentation FCN results for MR-positive and MR-negative subjects in percent (%). SD is the standard deviation.

	MR-positive	MR-negative	All Subjects
Subject-wise Sensitivity	98	92	96
Subject-wise Specificity	100	100	100
Lesional Sensitivity (SD)	74 (5)	68 (9)	72 (7)
Lesional Specificity (SD)	98 (1)	97 (2)	98 (1)
Lesional Precision (SD)	48 (15)	51 (12)	49 (14)
Dice (SD)	57 (12)	56 (8)	57 (11)

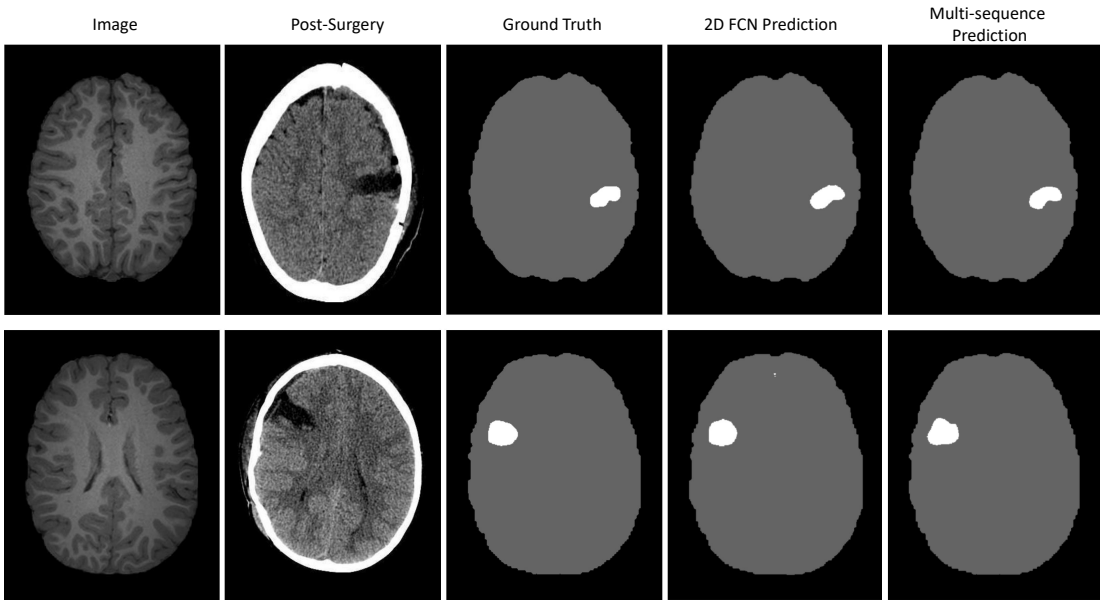
49 out of 56 MR-positive cases with lesion coverage of 65 percent (SD of 6) and 18 out of 24 MR-negative subjects with lesion coverage of 57 percent (SD of 5). For FLAIR sequence, the model detected 47 out of 56 MR-positive cases with lesion coverage of 61 percent (SD of 6) and 16 out of 24 MR-negative subjects with lesion coverage of 55 percent (SD of 3).

In TABLE 6, we grouped the subjects based on MR-positive and MR-negative cases along with all subjects results. The multi-sequence FCN model correctly classified all healthy control cases, and detected 55 out of 56 MR-positive cases and 22 out of 24 MR-negative subjects. For MR-positive cases, we obtained 74 percent (SD of 5) lesion coverage and dice coefficient of 57 (SD of 12) while for MR-negative subjects, the lesion coverage was 68 percent (SD of 9) and dice coefficient of 56 (SD of 8).

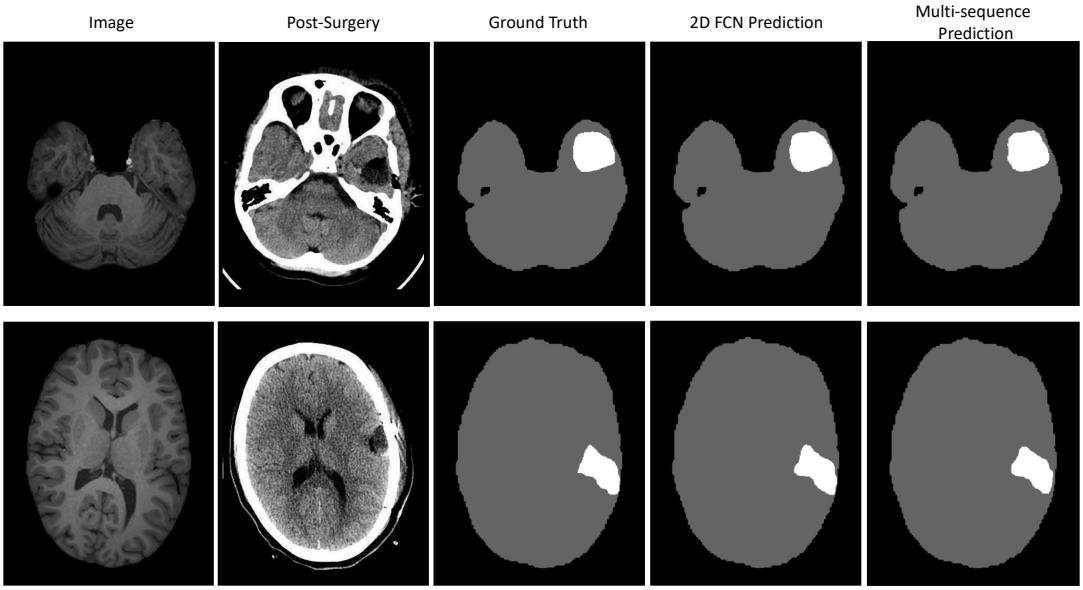
FIGURE 5 illustrates both models predictions, we have 4 subjects along with the pixel level ground truth and the network's output. There are subtle differences between multi-sequence and two dimensional FCN however the multi-sequence model is covering the lesion better. The case of multi-sequence prediction is an over-segmentation case where the FCD predicted pixels outnumbers the ground truth FCD pixels by a small margin. The advantage of multi-sequence model over the two-dimensional FCN is that we achieved a higher subject-wise sensitivity which is the result of the combined information from all sequences simultaneously. Both models segmented the lesion as best as possible while avoiding the healthy tissue.

TABLE 7 shows information related to the relative volume of the lesion for MR-positive and MR-negative cases. Relative volume is computed by dividing number of lesional voxels over healthy voxels which here is presented in percentage. The average relative lesion volume for MR-positive cases was 2.08 percent with a standard deviation of 1.9 percent and for MR-negative cases, the average relative lesion size was 1.8 percent (SD of 0.83). Both FCN models had difficulty detecting smaller and subtle lesions, for example, the average relative lesion size for missed cases by the multi-sequence model was 1.01 percent (3 cases) and the average relative lesion size for the seven missed cases by the 2D segmentation FCN was 1.02 percent.

We tested the multi-sequence model on six new cases to ensure the model did not overfit. Once we obtained the predictions for the location of the lesion, the subjects labels was assigned using the cut-off ratio. Subsequently, we calculated the lesional performance of the model on the new cases. TABLE 8 shows the characteristic of the new patients.



(a) MR-positive subjects



(b) MR-negative subjects

Figure 5: Models output visualization, each row's of images belong to one subject. Top are MR-positive cases and the bottom are MR-negative. Each row, from left to right, columns represent the original slice, post-surgery volumes, pixel-wise ground-truth, 2D FCN's prediction, multi-sequence network's prediction.

Table 7: Lesions relative volumes (mean and Standard deviation(SD)).

	MR-positive	MR-negative
Mean	2.08%	1.80%
Standard deviation	1.91%	0.83%
Minimum	0.44%	1.06%
Maximum	7.80%	5.16%

Table 8: Baseline characteristics of the MRI data for the new patients.

Subject	Age	Sex	Age at seizure onset	Seizure freq.	Number of anti-seizure medications	Location of the lesion	Histology
107	8.94	Male	7	Weekly	2	R posterior temporal	Laser therapy (N/A)
108	8.05	Male	4	Daily	3	R parietal	FCD IIB
109	15.10	Male	2	Monthly	1	L mesial parietal	Laser therapy (N/A)
110	14.52	Male	5	Daily	3	L inferior frontal	FCD IIB
112	16.62	Female	9	Weekly	2	R basal frontal	Laser therapy (N/A)
113	11.45	Male	3	Daily	3	R inferior frontal	FCD IIB

TABLE 9, shows the performance of the multi-sequence and the 2D segmentation FCN on the new cases. The multi-sequence model detected four out of six cases where the predicted lesion had 56 percent overlap with the actual lesion.

Similar to the previous experiments the lesion in the missed cases were very subtle or smaller in size and the average relative lesion size for the two missed cases were 0.72 percent (SD of 0.02) and the average relative lesion size for the correctly classified cases were 2.17 percent (SD of 1.6). Relative lesion size is computed by dividing number of lesional voxels over healthy voxels which here is presented in percentage.

5. DISCUSSION AND CONCLUSION

In this study, we have applied whole-slice and a multi-sequence FCN to volumetric T1, T2, FLAIR imaging to classify and localize FCD in children with drug-resistant epilepsy. We have found that the sensitivity of FCN to classify a lesion was dependent on the model used, with multi-sequence FCN demonstrating higher subject-wise sensitivity than the whole-slice FCN to classify a lesion. In the whole-slice FCN model, a whole slice approach was used as input for the model, which may have reduced the subject-wise sensitivity compared to multi-sequence model. The lesional sensitivity was also dependent on the model used, with multi-sequence FCN demonstrating higher lesional sensitivity than the whole slice FCN to segment a lesion, with dice coefficient of 57% as compared to 56%, 43% and 42%. The multi-

Table 9: multi-sequence FCN results for new cases in percent (%). SD is the standard deviation.

	Subjectwise Sens.	Lesional Sens. (SD)	Lesional Spec. (SD)	Dice (SD)
Muti-sequence FCN	67	56 (9)	99 (0)	58 (6)

sequence model combined information from three distinct sequences which could account for its higher lesional sensitivity. Our models, irrespective of multi-sequence or whole-slice FCN, achieved higher subject-wise sensitivity and dice coefficient compared to the study by Feng et al. [30], and Bijay et al. [32].

This study was conducted using MRI dataset derived from a single institution and from a single scanner, and therefore the performance of the network may differ if there were heterogeneity in MRI datasets from different institutions. Future study that focus on training using heterogeneous datasets from multiple scanners, and then subsequently validating the models to multi-institutional datasets, will improve the generalizability of the models for detecting subtle lesion such as FCD in children. Due to the relatively small dataset, we have not split the data into training and testing datasets, but instead utilize a leave-one-out method of training and testing. Other study using CNN has also utilized a leave-one-out technique for training and testing dataset [31].

We have shown that multi-sequence and whole-slice FCN could identify and localize a lesion in children with drug-resistant epilepsy. The subject-wise sensitivity and lesional sensitivity were higher with multi-sequence FCN than whole-slice FCN. This work showed that deep learning utilizing FCN is promising for identifying and localizing FCD in children with drug-resistant epilepsy [42]. Additional work is needed to further improve lesion identification and localization, particularly for small lesion, as well as to train and test the optimal algorithm on a larger multi-center dataset.

As part of future work, larger datasets are crucial in data-driven models, and also moving to a patch-wise multi-sequence model, which is a more robust way of building representations compared to the current model. Another bottleneck in the multi-sequence model is the labeling of the T2 and FLAIR sequences as they are sharing the same label as T1 even though T2 and FLAIR sequences are 2D MRI images where T1 is a 3D MRI sequence. Improving the labeling and data pre-processing will also improve the representation and could result in a more accurate detection and localization.

6. ACKNOWLEDGEMENTS AND DECLARATIONS

This study is supported by EpLink, The Epilepsy Research Program of the Ontario Brain Institute (OBI). The OBI is an independent non-profit corporation, funded partially by the Ontario government. The opinions, results and conclusions are those of the authors and no endorsement by the Ontario Brain Institute is intended or should be inferred. This research was also supported in part by Natural Sciences and Engineering Research Council of Canada (NSERC). AA would like to acknowledge OntarioTech university for the doctoral graduate international tuition scholarship (GITS). The authors gratefully acknowledge the support of NVIDIA corporation for their donation of Titan V GPU used in this research through its Academic Grant Program.

7. CONFLICT OF INTEREST

The authors declare no conflict of interest.

References

[1] Chin RF, Neville BG, Peckham C, Bedford H, Wade A, et al. Incidence, Cause, and Short-Term Outcome of Convulsive Status Epilepticus in Childhood: Prospective Population-Based Study. The Lancet. 2006;368, 9531:222–229.

[2] Coeytaux A, Jallon P, Galobardes B, Morabia A. Incidence of Status Epilepticus in French-Speaking Switzerland (Epistar). Neurology. 2000;55,5: 693–697.

[3] Hrabok M, Sherman EM, Bello-Espinosa L, Hader W. Memory and Health-Related Quality of Life in Severe Pediatric Epilepsy. Pediatrics. 2013;131:e525–e532.

[4] Rodenburg R, Stams G J, Meijer AM, Aldenkamp AP, Deković M. Psychopathology in Children With Epilepsy: A Meta-Analysis. Journal of pediatric psychology 30, 6 (2005), 453–468.

[5] Shorvon SD, Goodridge DM. Longitudinal Cohort Studies of the Prognosis of Epilepsy: Contribution of the National General Practice Study of Epilepsy and Other Studies. Brain. 2013;136:3497–3510.

[6] Berg AT, Shinnar S, Testa FM, Levy SR, Smith SN, et al. Mortality in Childhood-Onset Epilepsy. Archives of pediatrics adolescent medicine. 2004;158:1147–1152.

[7] Peters A, Callenbach P. Death in Children With Epilepsy: A Population-Based Study. The Journal of Pediatrics. 2003;142:87-87.

[8] Colombo N, Tassi L, Galli C, Citterio A, Russo GL, et al. Focal Cortical Dysplasias: Mr Imaging, Histopathologic, and Clinical Correlations in Surgically Treated Patients With Epilepsy. American Journal of Neuroradiology. 2003;24:724-733.

[9] Krsek P, Hajek M, Dezortova M, Jiru F, Skoch A, et al. 1H MR Spectroscopic Imaging in Patients With MRI-Negative Extratemporal Epilepsy: Correlation With Ictal Onset Zone and Histopathology. European radiology. 2007; 17:2126-2135.

[10] McGonigal A, Bartolomei F, Régis J, Guye M, Gavaret M, et al. Stereoelectroencephalography in presurgical assessment of MRI-negative epilepsy. Brain. 2007;130:3169-3183.

[11] Jeha LE, Najm I, Bingaman W, Dinner D, Widdess-Walsh P, et al. Surgical Outcome and Prognostic Factors of Frontal Lobe Epilepsy Surgery. Brain. 2007;130:574–584.

[12] Lee SK, Lee SY, Kim KK, Hong KS, Lee DS, et al. Surgical Outcome and Prognostic Factors of Cryptogenic Neocortical Epilepsy. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society. 2005;58:525-532.

[13] Nobili L, Francione S, Mai R, Cardinale F, Castana L, et al. Surgical Treatment of Drug-Resistant Nocturnal Frontal Lobe Epilepsy. Brain. 2006;130:561–573.

[14] Téllez-Zenteno JF, Ronquillo LH, Moien-Afshari F, Wiebe S. Surgical Outcomes in Lesional and Non-lesional Epilepsy: A Systematic Review and Meta-Analysis. Epilepsy Research. 2010;89:310-318.

[15] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[16] Krizhevsky A, Sutskever I., Hinton GE. Imagenet Classification With Deep Convolutional Neural Networks. In Advances in neural information processing systems. 2012: 1097–1105.

[17] Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. Journal of digital imaging. 2017;30 : 449–459.

[18] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning, vol. 1. MIT press Cambridge, 2016.

[19] Antel SB, Collins DL, Bernasconi N, Andermann F, Shinghal R, et al. Automated Detection of Focal Cortical Dysplasia Lesions Using Computational Models of Their MRI Characteristics and Texture Analysis. Neuroimage. 2003:19:1748-1759.

[20] Yang CA, Kaveh M, Erickson BJ. Automated Detection of Focal Cortical Dysplasia Lesions on T1-Weighted MRI Using Volume-Based Distributional Features. In 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (2011), IEEE. 865–870.

[21] Strumia M, Ramantani G, Mader I, Henning J, Bai L, et al. Analysis of Structural MRI Data for the Localisation of Focal Cortical Dysplasia in Epilepsy. In Workshop on Clinical Image-Based Procedures. Springer. 2012:25-32.

[22] Ahmed B, Thesen T, Blackmon K, Zhao Y, Devinsky O, et al. Hierarchical Conditional Random Fields for Outlier Detection: An Application to Detecting Epileptogenic Cortical Malformations. In International Conference on Machine Learning. 2014: 1080–1088.

[23] Adler S, Wagstyl K, Gunny R, Ronan L, Carmichael D, et al. Novel Surface Features for Automated Detection of Focal Cortical Dysplasias in Paediatric Epilepsy. NeuroImage: Clinical. 2017;14:18–27.

[24] Kulaseharan S, Aminpour A., Ebrahimi M, Widjaja E. Identifying Lesions in Paediatric Epilepsy Using Morphometric and Textural Analysis of Magnetic Resonance Images. NeuroImage: Clinical. 2019:101663.

[25] Wang I, Oh S, Blümcke I, Coras R, Krishnan B, et al. Value of 7T MRI and Post-processing in Patients With Nonlesional 3T MRI Undergoing Epilepsy Presurgical Evaluation. Epilepsia. 2020.

[26] Demerath T, Rubensdörfer L, Schwarzwald R, Schulze-Bonhage A, Altenmüller D- M, et al. Morphometric MRI analysis: Improved detection of focal cortical dysplasia using the mp2rage sequence. American Journal of Neuroradiology. 2020;41;6:1009-1014.

[27] Maiworm M, Nöth U, Hattingen E, Steinmetz H, Knake S, et al. Improved Visualization of Focal Cortical Dysplasia With Surface-Based Multiparametric Quantitative MRI. Frontiers in Neuroscience. 2020;14:622.

[28] Jin B, Krishnan B, Adler S, Wagstyl K, Hu W, et al. Automated Detection of Focal Cortical Dysplasia Type II With Surface-Based Magnetic Resonance Imaging Postprocessing and Machine Learning. Epilepsia. 2018;59:982–992.

[29] Roca P, Mellerio C, Chassoux F, Rivière D, Cachia A, et al. Sulcus-Based MR Analysis of Focal Cortical Dysplasia Located in the Central Region. PloS one. 2015;10:e0122252.

[30] Feng C, Zhao H, Li Y, Wen J. Automatic Localization and Segmentation of Focal Cortical Dysplasia in Flair-Negative Patients Using a Convolutional Neural Network. Journal of Applied Clinical Medical Physics. 2020;21:215–226.

[31] Wang H, Ahmed SN, Mandal M. Automated Detection of Focal Cortical Dysplasia Using a Deep Convolutional Neural Network. Computerized Medical Imaging and Graphics. 2020; 79: 101662.

[32] Dev KB, Jogi PS, Niyas S, Vinayagamani S, Kesavadas C, et al. Automatic Detection and Localization of Focal Cortical Dysplasia Lesions in MRI Using Fully Convolutional Neural Network. Biomedical Signal Processing and Control. 2019;52:218–225.

[33] Bastian D, Sailesh C, Fabiane S, Bernd W, Elke H, et al. Conditional Generative Adversarial Networks Support the Detection of Focal Cortical Dysplasia. Organization for Human Brain Mapping,Abstract Submission. 2019.

[34] Gill RS, Hong SJ, Fadaie F, Caldairou B, Bernhardt BC, et al. Deep Convolutional Networks for Automated Detection of Epileptogenic Brain Malformations. In International Conference on Medical Image Computing and Computer-Assisted Intervention .Springer.2018;490-497.

[35] Thomas E, Pawan S, Kumar S, Horo A, Niyas S, et al, J. Multi-res-attention unet: A CNN Model for the Segmentation of Focal Cortical Dysplasia Lesions From Magnetic Resonance Images. IEEE Journal of Biomedical and Health Informatics. 2020.

[36] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical image computing and computer-assisted intervention. Springer. 2015:234–241.

[37] Ibtehaz N, Rahman MS. Multiresunet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. Neural Networks.2020;121:74–87.

[38] Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, et al. Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images. Medical image analysis. 2019;53:197–207.

[39] Smith SM. Fast Robust Automated Brain Extraction. Human brain mapping. 2002;17:143–155.

[40] Collins T J. Imagej for Microscopy. Biotechniques. 2007;43: S25–S30.

[41] Aminpour A, Ebrahimi M, Widjaja E. Lesion Localization in Paediatric Epilepsy Using Patch-Based Convolutional Neural Network. In International Conference on Image Analysis and Recognition, Springer. 2020:216–227.

[42] Aminpour A. A Deep Learning Approach to Focal Cortical Dysplasia Segmentation in Children with Medically Intractable Epilepsy. University of Ontario Institute of Technology (Ontario Tech University). 2021.