DETECTION OF MRI-NEGATIVE FOCAL CORTICAL DYSPLASIA USING UNCERTAINTY-INFORMED BAYESIAN DEEP LEARNING: A MULTICENTRE VALIDATION STUDY

Gill RS [1,2*], Hong SJ [1,2], Fadaie F [1,2], Caldairou B [1,2], Lee HM [1,2], Hall J [2], Dudley R [2], Nguyen D [3], Barba C [4], Brandt A [5], Coelho V [6], d'Incerti L [7], Lenge M [4], Semmelroch M [8], Schrader D [9], Deleo F [7], Hogan RE [10], Bartolomei F [11], Guye M [12], Schulze-Bonhage A [5], Cho KH [13], Cendes F [6], Guerrini R [4], Jackson G [8], Bernasconi N [1,2], Bernasconi A [1,2]

[1] Neuroimaging of Epilepsy Laboratory, and [2] Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montréal, Canada; [3] Department of Neurology, Université de Montreal, Canada; [4] Pediatric Neurology Unit and Laboratories, Children's Hospital A. Meyer - University of Florence, Italy; [5] Freiburg Epilepsy Center, Universitätsklinikum Freiburg, Germany; [6] Department of Neurology, University of Campinas, Brazil; [7] Istituto Neurologico Carlo Besta, Milano, Italy; [8] Florey Institute of Neuroscience and Mental Health and University of Melbourne, Victoria, Australia; [9] Department of Pediatrics, Children's Hospital, University of British Columbia, Vancouver, Canada; [10] Department of Neurology, Washington University School of Medicine, St Louis, Missouri, USA; [11] Aix Marseille Univ, INSERM, Institut de Neurosciences des Systèmes, Marseille, France; [12] Aix Marseille Univ, CNRS, CRMBM, Marseille, France; [13] Department of Neurology, Yonsei University College of Medicine, Seoul, Korea

PURPOSE

Focal cortical dysplasia (FCD) is a prevalent surgically-amenable developmental epileptogenic lesion. On MRI, FCD typically presents with cortical thickening, hyperintensity, and blurring of the gray-white matter interface [1]. These changes may be visible to the naked eye, or subtle and overlooked. Despite advances in MRI analytics, current machine learning algorithms [2-5] fail to detect FCD in up to 50% of cases [6]. Also, their deterministic nature prevents risk assessments of predictions, a crucial step in clinical decision-making.

Here, we propose a deep learning algorithm formulated on Bayesian convolutional neural networks (CNN) [7] that provides prediction uncertainty, while leveraging this information to optimize performance. The algorithm was trained and validated on multimodal MRI data using a leave-one-site-out strategy across 9 epilepsy surgery centers across the world, for a total of 249 cases.

METHODS

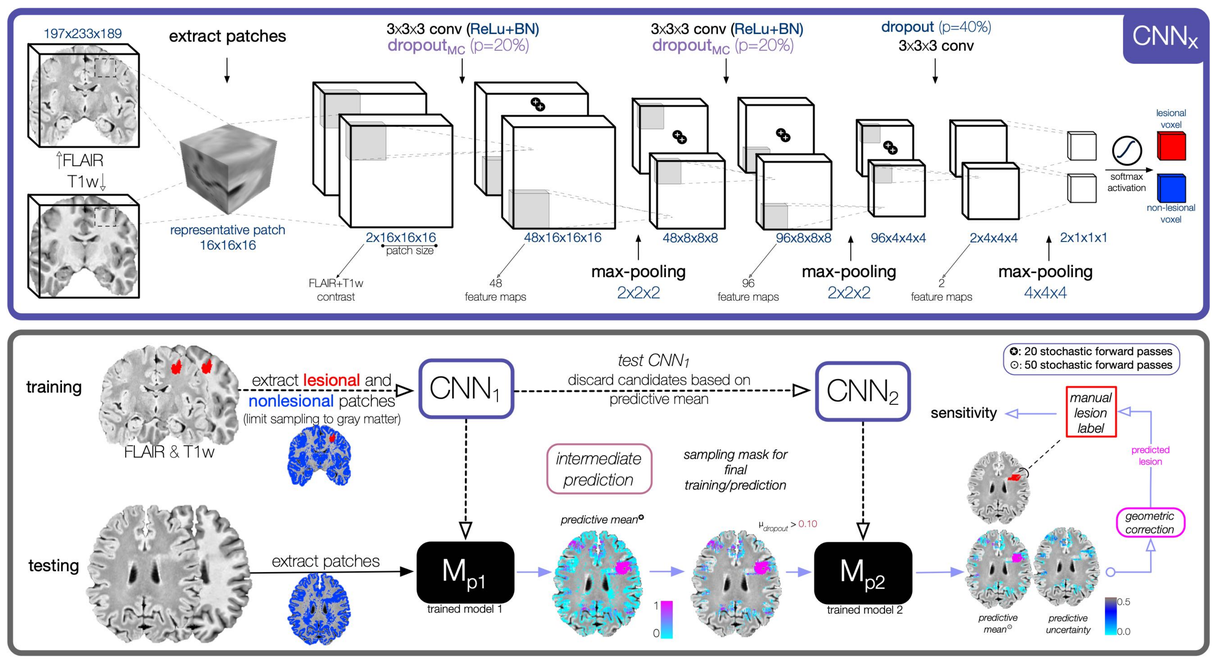

Figure 1. Upper panel: Convolutional network architecture (CNN$_x$) for two-label (lesional vs. non-lesional) classification with three consecutive convolutions (kernel size: 3x3x3, filters: 48,96,2) and max-pooling units, followed by a voxel-wise softmax classification using multimodal (FLAIR+T1w) patches. Each convolution is followed by rectified linear units (ReLu), which introduce non-linearity. Batch normalization (BN) and dropout (p=40%) prevent network overfitting. Dropout$_{MC}$ (p=20%) operation after first and second convolution layers are essential to quantify the epistemic uncertainly using dropout Monte Carlo. Adadelta serves as the gradient descent optimizer to minimize the binary crossentropy loss. **Lower panel**: Training and testing schema using two-stage CNN cascade (CNN$_1$/CNN$_2$) that incorporates uncertainty information. Multimodal patches (centered around the voxel to classify) are sampled from a gray matter mask to reduce training time. This mask is generated using intra-subject z-score of FLAIR, discarding hypointense voxels (z<0.1). Performance (sensitivity) is measured relative to the expert manual labels.

Training. 3D T1-weighted and 3D FLAIR MRI were collected in 148 patients with histologically-verified FCD, across 9 sites. Routine MRI was initially reported as unremarkable in 68 (MRI-negative; 46%) patients. Images underwent intensity inhomogeneity correction and standardization. T1-weighted images were registered linearly to the MNI152 template. FLAIR images were linearly mapped to T1w in MNI space.

Classifier design (Figure 1). Volumetric datasets served as inputs to a two-stage cascaded CNN: the first CNN was designed to maximize sensitivity (*i.e.*, detecting a maximum number of lesional voxels), while the second optimized specificity (*i.e.*, reducing false positives). Stochastic forward passes allow computing the voxel-wise uncertainty in predictions.

Validation. A leave-one-site-out cross-validation tested sensitivity (prediction co-localizing with manual FCD labels) across the 9 sites. Specificity was assessed by testing the model on 38 healthy controls (age: 30±7) and 63 temporal lobe epilepsy (TLE) patients as disease controls (age: 31±8), matched for age and sex to the S1 cohort.

*ravnoor.gill@mail.mcgill.ca
https://noel.bic.mni.mcgill.ca/

neuroimaging of epilepsy laboratory noel

RESULTS

Per-site sensitivity and false positive rates are summarized in the **Table**.

Overall sensitivity was 87% (129/148 FCD lesions detected; 6±5 extra-lesional clusters), including 76% of MRI- cases with lesions initially overlooked on routine exam (**Figure 2**).

Specificity was 89% in healthy controls (4/38; 2±1) and 89% in TLE (7/63; 1±0).

Site	Scanner	Age	Gender (% ♀)	MRI- (%)	Sensitivity MRI+ & MRI-	Sensitivity MRI-	False positive clusters
S1-I	Siemens 3T TrioTim	27±9	49%	32/45 (71%)	38/45 (84%)	25/32 (78%)	7±4
S1-II	Siemens 3T Prisma	18±9	65%	14/17 (82%)	13/17 (76%)	11/14 (79%)	7±4
S2	GE 3T Discovery	11±6	25%	1/8 (13%)	8/8 (100%)	1/1 (100%)	6±5
S3	Philips 3T Achieva	22±17	80%	3/5 (60%)	3/5 (60%)	1/3 (33%)	1±1
S4	Philips 3T Achieva / GE 3T Signa	8±7	36%	0/11	11/11 (100%)	NA	8±6
S5-I	Siemens 3T Prisma	23±14	30%	2/10 (20%)	9/10 (90%)	1/2 (50%)	10±6
S5-II	Siemens 3T TrioTim	13±12	41%	0/12	11/12 (92%)	NA	6±7
S6	Siemens 3T Verio	31±15	63%	2/11 (18%)	11/11 (100%)	2/2 (100%)	3±3
S7	Siemens 3T Skyra / Siemens 3T TrioTim	33±13	33%	7/9 (78%)	8/9 (89%)	6/7 (86%)	8±6
S8	Philips 3T Achieva / Siemens 1.5T Avanto	24±13	43%	1/7 (14%)	6/7 (86%)	0/1 (0%)	6±5
S9	Philips 3T Achieva	26±8	38%	6/13 (46%)	11/13 (85%)	5/6 (83%)	1±2
				68/148 (46%)	**129/148 (87%)**	**52/68 (76%)**	**6±5**

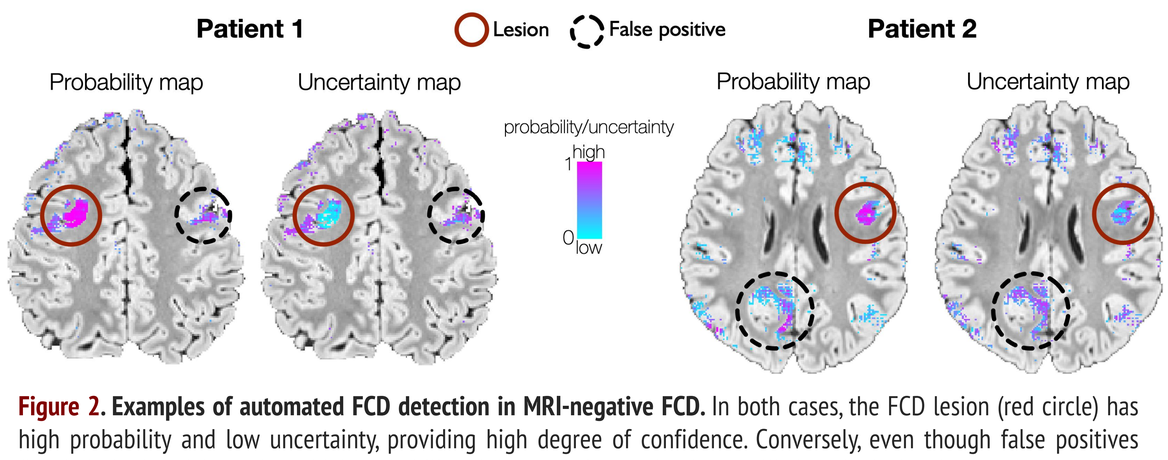

Figure 2. Examples of automated FCD detection in MRI-negative FCD. In both cases, the FCD lesion (red circle) has high probability and low uncertainty, providing high degree of confidence. Conversely, even though false positives (black dashed circle) may have high probability, high uncertainty make them unlikely to be lesional.

CONCLUSION

We present the first multicenter automated FCD detection algorithm based on deep learning with sites contributing data for both training and validation. Our method is unique as it:

1) operates on multi-contrast MRI in voxel-space
2) demonstrates generalizability (robust performance across independent cohorts with varying age and scanner hardware).
3) provides similar sensitivity in MRI- and MRI+ FCD.
4) quantifies a degree of confidence via uncertainty maps, thereby allowing to evaluate non-invasively the FCD lesion relative to putative false positives
5) sets the basis for distributed machine learning through sharing of site-specific training models, rather than patient data

Easy implementation, minimal pre-processing, and performance gains make this classifier the ideal platform for large-scale clinical use, particularly in MRI-negative FCD.

REFERENCES

[**1**] Bernasconi A, et al. Nat Rev Neurol 2011; [**2**] Adler S., et al. NeuroImage Clin 2017; [**3**] Hong SJ, et al. Neurology 2014; [**4**] Gill RS, et al. MICCAI Proceedings 2017; [**5**] Tan YL, et al. Neuroimage 2017; [**6**] Kini LG, et al. Neuroimage Clin 2016; [**7**] Gal Y & Ghahramani Z, arXiv:statML 2015.