Quantitative Imaging of Epileptogenic Lesions in MRI-negative Epilepsy

Ravnoor Singh Gill

Department of Neurology and Neurosurgery
McGill University
Montreal, Quebec, Canada

April 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

To
Nonna, Gurnaaz, Bhopi, and Ambreen

Acknowledgements

The patients and control subjects who took part in this research deserve my utmost gratitude for their interest and willingness to undergo prolonged MRI scanning sessions. Without their invaluable contributions, research experiments in this thesis would not be possible.

I am deeply grateful to my thesis advisors, Drs Andrea Bernasconi and Neda Bernasconi, who have been instrumental in completion of this thesis. I gratefully acknowledge their patience, warmth, kind and insightful advice, and generous supports in all aspects of life, and their exemplary willingness to support their students. They have striven to impart skills critical to succeeding as an independent researcher, from writing to analytical thinking and research methods. I thoroughly enjoyed fruitful discussions that helped situate my research in a clinical context. Their tolerance and openness to untested research ideas is greatly appreciated.

As a senior PhD student, Dr Seok-Jun Hong, provided a soft landing for a naïve first year PhD student, and a conducive environment to ask questions and bounce ideas off, for which I'll be forever grateful. I am thankful to Dr Boris Bernhardt, one of my committee members, for providing generous advice to improve the quality of this work and facilitating insightful discussions during lab and advisory committee meetings.

Many thanks to my colleagues and friends from the lab: Benoit Caldairou, Fatemeh Fadaie, Hyo-Min Lee and Min Liu. Especially, Benoit, who has been a both an active and a passive partner in crime in piloting various research projects. I would like to extend my gratitude to Haran Balendra, Manoja Chandralingam and Maria Athanassiou to help translating the technical and non-technical aspects of my thesis abstract into Française. I would also like to express my deep admiration to all the amazing people at the Montreal Neurological Institute (MNI) for providing a fantastic work environment: Louis Collins, Vladimir Fonov, Claude Lepage, Julien Doyon, Jean-Francois Malouin, Sylvain Milot, Ivan Castanon Niconoff, and Anna Stabile.

I greatly appreciate the examiners for my PhD candidacy exam and defence, Drs Richard Hoge, Ali Khan and Sridar Narayanan, and my IPN mentor, Dr Pierre Lachapelle for dedicating their time to provide wise and generous advice integral to improving the quality of this work. I am also

thankful to the staff at the Integrated Program in Neuroscience (IPN): Vivian Omune, Katherine Vanka, Ben Browning, Alexander DeGuise, Ed Ruthazer and Joseph Rochford.

I have been extremely fortunate to be the recipient of several awards during my doctoral journey. Thus, I would like to gratefully acknowledge the Fonds de recherche du Québec - Santé (FRQS), MNI, the American Epilepsy Society (AES), Organization for Human Brain Mapping (OHBM), the IPN, and others for generous financial support. Specifically, PhD fellowship (FRQS), Young Investigator Awards (AES), Merit Abstract Award (OHBM), Kenelm M. Winslow and Tom Gevas Travel Awards, (MNI), IPN Star Award, and MICCAI/NIH Travel Award.

Finally, I am deeply indebted to my family for their constant support, love, space, time, patience, and inspiration: Nonna, Gurnaaz, Bhopi, and Ambreen, for this thesis would not be feasible without them.

Thank you.

Contributions of Authors

All manuscripts in this thesis are the product of a close collaboration between my supervisors (**Andrea Bernasconi, MD** and **Neda Bernasconi, MD PhD**) and I. We jointly designed the experiments, planned the statistical analysis, interpreted the results, and wrote the manuscripts. I performed all stages of software design and development, image processing, quality control, statistical analysis, and data visualizations. I also drafted the first version of all manuscripts and made all subsequent revisions based on the comments of my supervisors.
The following list summarizes the contributions of the co-authors at the Montreal Neurological Institute and elsewhere.

1. **Benoit Caldairou, PhD**. Assisted in the cross-validation, quality control and software development in Project 4; and image processing in Projects 2–4.
2. **Seok-Jun Hong, PhD**. Assisted in study design and image processing of Project 2 and provided guidance on aspects of thesis writing.
3. **Boris Bernhardt, PhD**. *Director, MICA lab at the BIC*. Advised on study design and statistical analyses for Projects 1–2.
4. **Tommaso Mansi, PhD**, *Director, AI/ML and Digital Health, Janssen R&D Data Science*. Provided advice on the development of deep learning classifier and aided in the interpretation of results in Project 3.

Publications

PEER-REVIEWED ARTICLES

1. B. Caldairou, N. A. Foit, C. Mutti, F. Fadaie, **R. Gill**, H. M. Lee, T. Demerath, H. Urbach, A. Schulze-Bonhage, A. Bernasconi, et al. "MRI-Based Machine Learning Prediction Framework to Lateralize Hippocampal Sclerosis in Patients With Temporal Lobe Epilepsy". *Neurology* 97:16, 2021, e1583–e1593.
2. F. Fadaie, H. M. Lee, B. Caldairou, **R. S. Gill**, V. Sziklas, J. Crane, B. C. Bernhardt, S.-J. Hong, A. Bernasconi, and N. Bernasconi. "Atypical functional connectome hierarchy impacts cognition in temporal lobe epilepsy". *Epilepsia* 62:11, 2021, pp. 2589–2603.
3. **R. S. Gill**, H.-M. Lee, B. Caldairou, S.-J. Hong, C. Barba, F. Deleo, L. D'Incerti, V. C. Mendes Coelho, M. Lenge, M. Semmelroch, D. V. Schrader, F. Bartolomei, M. Guye, A. Schulze-Bonhage, H. Urbach, K. H. 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, 2021, e1571–e1582.
4. H. M. Lee, F. Fadaie, **R. Gill**, B. Caldairou, V. Sziklas, J. Crane, S.-J. Hong, B. C. Bernhardt, A. Bernasconi, and N. Bernasconi. "Decomposing MRI phenotypic heterogeneity in epilepsy: a step towards personalized classification". *Brain*, 2021.
5. H. M. Lee, **R. Gill**, F. Fadaie, K. H. Cho, M. C. Guiot, S.-J. Hong, N. Bernasconi, and A. Bernasconi. "Unsupervised machine learning reveals lesional variability in focal cortical dysplasia at mesoscopic scale". *NeuroImage: Clinical*, 2020, p. 102438.
6. A. Bernasconi, F. Cendes, W. H. Theodore, **R. S. Gill**, M. J. Koepp, R. E. Hogan, G. D. Jackson, P. Federico, A. Labate, A. E. Vaudano, I. Blümcke, P. Ryvlin, and N. Bernasconi. "Recommendations for the use of structural magnetic resonance imaging in the care of patients with epilepsy: A consensus report from the International League Against Epilepsy Neuroimaging Task Force". *Epilepsia* 60:6, 2019, pp. 1054–1068.
7. **R. S. Gill**, B. Caldairou, N. Bernasconi, and A. Bernasconi. "Uncertainty-Informed Detection of Epileptogenic Brain Malformations Using Bayesian Neural Networks". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham. 2019, pp. 225–233.

8. S.-J. Hong, H.-M. Lee, **R. Gill**, J. Crane, V. Sziklas, B. C. Bernhardt, N. Bernasconi, and A. Bernasconi. "A connectome-based mechanistic model of focal cortical dysplasia". *Brain* 142:3, 2019, pp. 688–699.
9. **R. S. Gill**, S.-J. Hong, F. Fadaie, B. Caldairou, B. C. Bernhardt, C. Barba, A. Brandt, V. C. Coelho, L. d'Incerti, M. Lenge, et al. "Deep Convolutional Networks for Automated Detection of Epileptogenic Brain Malformations". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham. 2018, pp. 490–497.
10. **R. S. Gill**, S.-J. Hong, F. Fadaie, B. Caldairou, B. Bernhardt, N. Bernasconi, and A. Bernasconi. "Automated Detection of Epileptogenic Cortical Malformations Using Multimodal MRI". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, MICCAI*. Springer, Cham, 2017, pp. 349–356.
11. S.-J. Hong, B. Bernhardt, **R. S. Gill**, N. Bernasconi, and A. Bernasconi. "Connectome-based pattern learning predicts histology and surgical outcome of epileptogenic malformations of cortical development". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham. 2017, pp. 390–397.
12. S.-J. Hong, B. C. Bernhardt, **R. S. Gill**, N. Bernasconi, and A. Bernasconi. "The spectrum of structural and functional network alterations in malformations of cortical development". *Brain* 140:8, 2017, pp. 2133–2143.

MANUSCRIPTS IN PROGRESS

1. **R. S. Gill**, B. Caldairou, A. Bernasconi, and N. Bernasconi. "Combined Automated Hippocampal Segmentation and Focus Lateralization in Temporal Lobe Epilepsy".
2. **R. S. Gill**, F. Deleo, S.-J. Hong, F. Fadaie, B. Caldairou, B. Bernhardt, N. Bernasconi, and A. Bernasconi. "MRI-negative Epilepsy: A Systematic Review & Meta-analysis".

MEETING PRESENTATIONS

1. F. Fadaie, H. M. Lee, B. Caldairou, **R. S. Gill**, V. Sziklas, J. Crane, S.-J. Hong, A. Bernasconi, and N. Bernasconi. "Cognitive deficits in temporal lobe epilepsy reflect atypical functional connectome hierarchy". In: *International Epilepsy Congress of the International League Against Epilepsy*. Virtual, 2021.

2. H. M. Lee, F. Fadaie, **R. S. Gill**, B. Caldairou, V. Sziklas, J. Crane, S.-J. Hong, B. C. Bernhardt, A. Bernasconi, and N. Bernasconi. "Decomposing MRI phenotypic heterogeneity in epilepsy has clinical relevance". In: *The 75th Annual Meeting of the American Epilepsy Society*. Chicago, IL, USA, 2021.

3. **R. S. Gill**, B. Caldairou, N. Bernasconi, and A. Bernasconi. "Automated hippocampal subfields segmentation using deep learning". In: *The 74th Annual Meeting of the American Epilepsy Society*. Virtual, 2020.

4. H. M. Lee, F. Fadaie, **R. S. Gill**, B. Caldairou, V. Sziklas, J. Crane, S.-J. Hong, B. C. Bernhardt, A. Bernasconi, and N. Bernasconi. "Decomposing the complex imaging signature of temporal lobe epilepsy using latent phenotypes". In: *International Epilepsy Congress of the International League Against Epilepsy*. Virtual, 2020.

5. H. M. Lee, F. Fadaie, **R. S. Gill**, N. Foits, B. Caldairou, S.-J. Hong, A. Bernasconi, and N. Bernasconi. "Latent whole-brain disease factors in temporal lobe epilepsy". In: *The 74th Annual Meeting of the American Epilepsy Society*. Virtual, 2020.

6. **R. S. Gill**, B. Caldairou, N. Bernasconi, and A. Bernasconi. "Uncertainty-Informed Detection of Epileptogenic Brain Malformations Using Bayesian Neural Networks". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Shenzhen, China, 2019.

7. **R. S. Gill**, S.-J. Hong, F. Fadaie, B. Caldairou, H. M. Lee, J. Hall, R. Dudley, D. Nguyen, C. Barba, A. Brandt, V. C. Coelho, L. d'Incerti, M. Lenge, et al. "Detection of MRI-Negative Focal Cortical Dysplasia Using Uncertainty-Informed Bayesian Deep Learning: A Multicentre Validation Study". In: *The 73rd Annual Meeting of the American Epilepsy Society*. Baltimore, MD, USA, 2019.

8. H. M. Lee, F. Fadaie, **R. S. Gill**, B. Caldairou, S.-J. Hong, A. Bernasconi, and N. Bernasconi. "MRI-based Individualized Prediction of Medication Response and Surgical Outcome: A Whole-brain Approach Based on Factor Modeling". In: *The 73rd Annual Meeting of the American Epilepsy Society*. Baltimore, MD, USA, 2019.

9. H. M. Lee, **R. S. Gill**, F. Fadaie, S.-J. Hong, N. Bernasconi, and A. Bernasconi. "MRI Structural Heterogeneity Within and Across Focal Cortical Dysplasia: A Data-Driven Approach Based on Consensus Clustering". In: *The 73rd Annual Meeting of the American Epilepsy Society*. Baltimore, MD, USA, 2019.

10. **R. S. Gill**, S.-J. Hong, F. Fadaie, B. Caldairou, B. Bernhardt, C. Barba, V. Coelho, M. Lenge, M. Semmelroch, F. Bartolomei, M. Guye, F. Cendes, R. Guerrini, G. Jackson, N. Bernasconi, and A. Bernasconi. "Deep convolutional neural networks for detection of cor-

tical dysplasia: a multicenter validation". In: *Organization for Human Brain Mapping Annual Meeting*. Singapore, 2018.

11. **R. S. Gill**, S.-J. Hong, F. Fadaie, B. Caldairou, B. Bernhardt, C. Barba, V. Coelho, M. Lenge, M. Semmelroch, F. Bartolomei, M. Guye, F. Cendes, R. Guerrini, G. Jackson, N. Bernasconi, and A. Bernasconi. "Automatic Detection of Focal Cortical Dysplasia Using Deep Learning: A Multicentre Validation Study". In: *The 72nd Annual Meeting of the American Epilepsy Society*. New Orleans, LA, USA, 2018.

12. S.-J. Hong, H. Min-Lee, **R. S. Gill**, N. Bernasconi, and A. Bernasconi. "Connectome-Based Biotyping of Focal Cortical Dysplasia Type II". In: *The 72nd Annual Meeting of the American Epilepsy Society*. New Orleans, LA, USA, 2018.

13. S.-J. Hong, H. Min-Lee, **R. S. Gill**, N. Bernasconi, and A. Bernasconi. "Relation Between Lesion Topography and Sulcal Pits in Focal Cortical Dysplasia Type II". In: *The 72nd Annual Meeting of the American Epilepsy Society*. New Orleans, LA, USA, 2018.

14. S.-J. Hong, H. Min-Lee, **R. S. Gill**, B. Bernhardt, N. Bernasconi, and A. Bernasconi. "Focal cortical dysplasia: relation between lesion topography and sulcal pits". In: *Organization for Human Brain Mapping Annual Meeting*. Singapore, 2018.

15. **R. S. Gill**, S.-J. Hong, F. Fadaie, B. Caldairou, B. Bernhardt, N. Bernasconi, and A. Bernasconi. "Automated detection of subtle focal cortical dysplasia using multi-contrast MRI". In: *The 71st Annual Meeting of the American Epilepsy Society*. Washington, DC, USA, 2017.

16. S.-J. Hong, B. Bernhardt, **R. Gill**, N. Bernasconi, and A. Bernasconi. "Connectome-based pattern learning predicts histology and surgical outcome of epileptogenic malformations of cortical development". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Quebec City, QC, Canada, 2017.

17. S.-J. Hong, B. Bernhardt, **R. S. Gill**, N. Bernasconi, and A. Bernasconi. "Focal cortical dysplasia: abnormal functional connectome embedding predicts histopathology and surgical outcome". In: *The 71st Annual Meeting of the American Epilepsy Society*. Washington, DC, USA, 2017.

18. S.-J. Hong, B. C. Bernhardt, **R. S. Gill**, N. Bernasconi, and A. Bernasconi. "Multimodal connectome organization across the spectrum of cortical malformations". In: *The 70th Annual Meeting of the American Epilepsy Society*. Houston, TX, USA, 2016.

Abstract

Background. More than a third of patients with epilepsy suffer from seizures that are resistant to antiepileptic drugs. Drug-resistant epilepsy is a serious condition associated with a structural brain lesion. Temporal lobe epilepsy (TLE) secondary to mesiotemporal sclerosis, and extratemporal lobe neocortical epilepsy secondary to focal cortical dysplasia (FCD), a malformation due to abnormal neuronal proliferation and cortical organization, are the two most common drug-resistant epilepsies amenable to surgery. Surgical removal of the lesion is the only effective treatment to control seizures, limit their adverse effects on cognition and reduce risks of injury and death. Despite advances in MRI analytics, current algorithms are not optimized to accurately detect subtle lesions, a scenario in ~50% of referrals for pre-surgical evaluation. Since MRI criteria to localize the surgical target are missing, these "MRI-negative" patients undergo hospitalizations for invasive intracranial EEG monitoring (SEEG). Notably, there is a lack of objective criteria to ascribe the status of MRI-negative, perpetuating biases in the literature with a significant number of patients being misclassified. Indeed, patients considered as MRI-negative based on visual evaluation before surgery are not necessarily non-lesional, as retrospective quantitative analysis detects lesions concordant with histopathology. Consequently, misdiagnosis or delayed diagnosis results in lower chances for post-surgical seizure freedom compared to patients with positive MRI.

Objective. To objectively define MRI-negative and develop and validate novel approaches to improve the yield of MRI to resolve hard to detect epileptogenic lesions.

Methods. We first performed a systematic review and meta-analyses to assess the consistency of the criteria used to ascribe MRI-negative status in focal epilepsy (*Project 1*). Subsequently, we employed a bipartite approach in developing algorithms to detect FCD, which rely on the integration of multiple imaging modalities through i) surface-based sampling that respects cortical topology and provides accurate inter-subject correspondence (*Project 2*), and ii) minimally preprocessed volumetric approach that facilitates high generalization performance combining deep learning with uncertainty estimation for risk stratification (*Project 3*). Finally, we developed an algorithm for hippocampal subfield segmentation using deep learning and assessed its lateralization performance in TLE (*Project 4*).

Results. In *Project 1*, a systematic review of 196 studies demonstrated variability in ascribing MRI-negative status. Narrative synthesis summarized the clinical, demographic, and presurgical diagnostics profile of the included studies demonstrating that MRI-negative patients more often undergo SEEG (76% *vs.* 54%), are less frequently operated (74% *vs.* 86%) and have a less favor-

able seizure outcome (62% *vs.* 74%) relative to MRI-positive patients. Unsupervised clustering of the diagnostic modalities revealed three distinct groups (*MRI-dominant*, *nuclear-imaging*, and *limited-MRI-information*) with significant associations observed across several outcomes (MRI reporting and post-processing, and SEEG). The metanalyses revealed favorable post-surgical seizure outcome in 75% of MRI-positive cohorts relative to 59% in MRI-negative, and that MRI post-processing is associated with two-fold gain in diagnostic yield over qualitative review of MRI alone. In *Project 2*, we developed a classifier that leveraged surface-based descriptors of morphology and intensity derived from multicontrast MRI to accurately identify subtle FCD lesions initially overlooked on routine radiological assessment. The algorithm demonstrated excellent sensitivity (83%, 34/41 lesions detected) and specificity (92%, no findings in 35/38 controls). In *Project 3*, we propose a novel volume-based detection algorithm that combines deep learning with uncertainty estimation for risk stratification. The learner yielded the highest sensitivity (93%; 137/148 FCD detected) to date in histologically verified MRI-negative FCD. Minimal pre-processing and generalizability across age and MRI hardware in cohorts sampled from nine tertiary epilepsy centers worldwide makes this classifier ideal for clinical diagnostics. Finally, in *Project 4*, *DeepPatch*, a volume-based subfield segmentation method that combines patch-based analysis with deep learning, demonstrated Dice of > 88% across hippocampal subfields in healthy controls and TLE patients. Importantly, we showed clinical utility by accurately lateralizing the seizure focus in 95% of cases (91% in MRI-negative).

Significance. Our findings advocate for a central role of MRI post-processing in pre-surgical epilepsy diagnostics. Our integrated approach combining the analysis of multiple contrasts with advanced statistical learning techniques across diverse multisite datasets is designed to create open-source generalizable algorithms with the potential for broad clinical translation with low technical debt.

Résumé

Contexte. Plus d'un tiers des patients épileptiques souffrent de crises résistantes aux médicaments antiépileptiques. L'épilepsie résistante aux médicaments est une maladie grave associée à une lésion cérébrale structurelle. L'épilepsie du lobe temporal (ELT) secondaire à une sclérose mésiotemporale et l'épilepsie néocorticale du lobe extratemporel secondaire à une dysplasie corticale focale (DCF), une malformation due à une prolifération neuronale et une organisation corticale anormales, sont les deux épilepsies pharmaco-résistantes les plus courantes pouvant être opérées. L'ablation chirurgicale de la lésion est le seul traitement efficace pour contrôler les crises, limiter leurs effets indésirables sur la cognition et réduire les risques de blessures et de décès. Malgré les progrès des analyses imagerie par résonance magnétique (IRM), les algorithmes actuels ne sont pas optimisés pour détecter avec précision les lésions subtiles, ce qui est le cas dans ~50% des cas référés pour une évaluation pré-chirurgicale. Comme les critères d'IRM permettant de localiser la cible chirurgicale sont manquants, ces patients « IRM-négatifs » sont hospitalisés pour une surveillance invasive par EEG intracrânienne (SEEG). Ce manque de critères objectifs pour attribuer le statut d'IRM-négatif perpétue les biais dans la littérature en augmentant le nombre de patients mal classés. En effet, les patients considérés comme négatifs à l'IRM sur la base d'une évaluation visuelle avant la chirurgie ne sont pas nécessairement non-lésionnels, car une analyse quantitative rétrospective détecte des lésions concordantes avec l'histopathologie. Par conséquent, un mauvais diagnostic ou un diagnostic tardif entraîne un plus haut taux de de crises post-chirurgicales chez ces patients par rapport aux patients dont l'IRM est positif.

Objectif. Définir objectivement le statut IRM-négatif et développer et valider de nouvelles approches pour améliorer le rendement de l'IRM afin de résoudre les lésions épileptogènes difficiles à détecter.

Méthodes. Nous avons d'abord effectué une revue systématique et méta-analyses pour évaluer la cohérence des critères utilisés pour attribuer le statut IRM-négatif dans l'épilepsie focale (*Projet 1*). Par la suite, nous avons utilisé une approche bipartite pour développer des algorithmes de détection de l'épilepsie focale, qui reposent sur l'intégration de plusieurs modalités d'imagerie par le biais i) d'un échantillonnage basé sur la surface qui respecte la topologie corticale et fournit une correspondance inter-sujet précise (*Projet 2*), et ii) d'une approche volumétrique minimalement prétraitée qui facilite une généralisation de haute performance combinant « deep learning » avec l'estimation de l'incertitude pour la stratification du risque (*Projet 3*). Enfin, nous avons développé

un algorithme pour la segmentation des sous-divisions hippocampiques à l'aide de « deep learning » et évalué sa performance de latéralisation dans le ELT (*Projet 4*).

Résultats. Dans le *Projet 1*, une revue systématique de 196 études a démontré une variabilité dans l'attribution du statut négatif à l'IRM. Une synthèse narrative a résumé le profil clinique, démographique et diagnostic pré-chirurgical des études incluses, démontrant que les patients négatifs à l'IRM subissent plus souvent un SEEG (76% contre 54%), sont moins souvent opérés (74% contre 86%) et ont une évolution moins favorable des crises (62% contre 74%) par rapport aux patients positifs à l'IRM. Le regroupement non supervisé des modalités de diagnostic a révélé trois groupes distincts (*IRM-dominante*, *imagerie nucléaire* et *information limitée par IRM*) avec des associations significatives observées pour plusieurs résultats (rapport et post-traitement par IRM et SEEG). Les méta-analyses ont révélé une évolution favorable des crises post-chirurgicales dans 75% des cohortes positives à l'IRM contre 59% des cohortes négatives à l'IRM, et que le post-traitement de l'IRM est associée à un rendement diagnostique deux fois plus élevé que l'examen qualitatif exclusif de l'IRM. Dans le cadre du *Projet 2*, nous avons mis au point un classificateur qui exploite les descripteurs de surface de la morphologie et de l'intensité dérivés de l'IRM multicontraste pour identifier avec précision les lésions subtiles de la DCF initialement négligées lors de l'évaluation radiologique de routine. L'algorithme a démontré une excellente sensibilité (83%, 34/41 lésions détectées) et spécificité (92%, aucune découverte chez 35/38 témoins). Dans le *Projet 3*, nous proposons un nouvel algorithme de détection basé sur le volume qui combine « deep learning » et l'estimation de l'incertitude pour la stratification du risque. L'algorithme a démontré la sensibilité la plus élevée (93% ; 137/148 DCF détectés) à ce jour dans les DCF négatifs à l'IRM vérifiés histologiquement. Un prétraitement minimal et une bonne généralisation à travers les groupes d'âges et le matériel d'IRM utilisé dans des cohortes échantillonnées dans neuf centres d'épilepsie tertiaires du monde entier rendent ce classificateur idéal pour le diagnostic clinique. Enfin, dans le *Projet 4*, *DeepPatch*, une méthode de segmentation des sous-champs basée sur le volume qui combine l'analyse basée sur les patchs et « deep learning », a démontré un « Dice » de > 88% dans les sous-champs de l'hippocampe chez les témoins sains et les patients atteints de ELT. Il est important de noter que nous avons démontré une utilité clinique de latérisation avec précision le foyer des crises dans 95% des cas (91% dans les cas négatifs à l'IRM).

Signification. Nos résultats favorisent un rôle central du post-traitement IRM dans le diagnostic pré-chirurgical de l'épilepsie. Notre approche intégrée, combinant l'analyse de contrastes multiples avec des techniques d'apprentissage statistique avancées sur divers ensembles de données multicentriques, est conçue pour créer des algorithmes généralisables « open-source » avec le potentiel d'une large traduction clinique avec une faible dette technique.

Original Contributions

Project 1. *Defining MRI-negative Epilepsy*
We conducted a comprehensive systematic review of 196 studies and synthesized evidence to assess the ambiguity in criteria used to ascribe MRI-negative status. Our findings revealed that MRI-negative categorization lacks objectivity – across MRI reporting of contrasts, parameters, rater expertise, post-processing, and invasive EEG. These findings suggest that invasive diagnostics underutilize MRI both in acquisition protocols and post-processing. Subsequently, we presented meta-analytic evidence based on 70 studies demonstrating favorable post-surgical seizure outcome in MRI-positive relative to the MRI-negative cohorts. Finally, we present the first meta-analytic evidence based on 40 studies that supports the role of MRI post-processing in facilitating two-fold improvement in diagnostic yield over standard radiological evaluation.

Project 2. *Surface-based Automated Detection of MRI-negative Focal Cortical Dysplasia*
We developed a novel *in vivo* surface-based automated focal cortical dysplasia (FCD) detection algorithm exploiting the complementary diagnostic power of T1-weighted and T2-weighted FLAIR contrasts, together with a synthetic FLAIR/T1 map designed to increase the sensitivity for co-occuring FLAIR hyperintensity and T1w hypointensity at the grey and white matter interface. Our histologically validated method assessed intra- and sub-cortical FCD features on multi-contrast MRI and provided high detection performance (83%; 34/41 lesions detected), coupled with high specificity (92%; no findings in 35/38 healthy controls). Operating on routine MRI sequences, this approach optimizes the detection of subtle FCD lesions overlooked by standard radiological evaluation.

Project 3. *Automated Detection of Focal Cortical Dysplasia using Deep Learning*
We proposed the first multicenter-validated deep learning detection algorithm formulated on Bayesian convolutional neural networks that provide prediction uncertainty, while leveraging this information to optimize performance. The algorithm was trained and validated on minimally pre-processed multimodal MRI data in histologically verified MRI-negative FCD across 9 epilepsy surgery centers worldwide. Sensitivity was 93% (137/148 lesions detected: 85% in MRI-negative FCD) using a *leave-one-site-out* cross-validation strategy, and specificity was 89% in healthy and disease controls, providing the highest performance to date. By combining predictions with a confidence-based risk stratification, this classifier provides means to triage putative lesional candidates. Notably, this study provides Class III level of case-control diagnostic evidence that deep

learning on multimodal MRI accurately identifies FCD in patients with epilepsy initially diagnosed as MRI-negative.

Project 4. *Automated Hippocampal Subfield Segmentation using Deep Learning*
We propose *DeepPatch*, a volume-based subfield segmentation method that leverages both patch-based analysis, which optimizes label fusion and image matching by compactly representing anatomy, shape, texture and intensity, and fully deep convolutional neural networks that offer hierarchical feature learning ability. *DeepPatch*, operating on widely available T1-weighted MRI, yields remarkable performance, both in healthy controls and TLE patients, with Dice > 87% across hippocampal subfields. Segmentations obtained through the automated algorithm showed high performance to lateralize the seizure focus in 95% of patients (91% in MRI-negative), suggesting clinical utility. The combination of the patch-based framework with hierarchical feature learning capacity of deep neural networks captures efficiently the complex shape deformations and displacements, which are particularly prevalent in disease.

Contents

List of Figures

List of Tables

Abbreviations

ASHS	Automatic Segmentation of Hippocampal Subfields
AUC	Area Under the Curve
BN	Batch-normalization
BOLD	Blood Oxygen Level Dependent
CE	Cross Entropy
CLASP	Constrained Laplacian Anatomic Segmentation using Proximity
CNN	Convolutional Neural Network
CP	Cortical Plate
CSF	Cerebrospinal Fluid
DG	Dentate Gyrus
DICOM	Digital Imaging and Communications in Medicine
DNA	De-oxyribose Nucleic Acid
DNET	Dysembryoplastic Neuroepithelial Tumor
EEG	Electroencephalography
ENIGMA	Enhancing NeuroImaging Genetics through Meta-Analysis
FCD	Focal Cortical Dysplasia
FDG	Fluorodeoxyglucose
FL	Focal Loss
FLAIR	Fluid-Attenuated Inversion Recovery
fMRI	Functional Magnetic Resonance Imaging (resting-state or task-based)
FOV	Field of View
FP	False Positive
GABA	Gamma-Aminobutyric Acid
GM	Gray Matter
GRADE	Grades of Recommendations, Assessment, Development and Evaluation
GW	Gestational Week
HARNESS	Harmonized Neuroimaging of Epilepsy Structural Sequences
HDF5	Hierarchical Data Format

HIPS	Hippocampus Subfield Segmentation
HS	Hippocampal Sclerosis
HV	Hippocampal Volumetry
ILAE	International League against Epilepsy
INSECT	Intensity Normalized Stereotaxic Environment for the Classification of Tissue
IQR	Inter-Quartile Range
KL	Kullback-Leibler
LDA	Linear Discriminant Analysis
MAP	Morphometric Analysis Program
MCD	Malformations of Cortical Development
MEG	Magnetoencepahology
MPRAGE	Magnetization Prepared Rapid Gradient Echo
MRI	Magnetic Resonance Imaging
mTOR	Mammalian Target of Rapamycin
NIfTI	Neuroimaging Informatics Technology Initiative
OR	Odds Ratio
PET	Positron Emission Tomography
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analyses
QC	Quality Control
ReLU	Rectified Linear Activation
RI	Relative Intensity
ROC	Receiver Operating Characteristics
ROI	Region of Interest
RUSBoost	Random Undersampling with Boosting
SBM	Surface-Based Morphometry
scSE	Concurrent Spatial and Channel Squeeze-and-Excite
SD	Standard Deviation
SE	Squeeze and Excite
SEEG	Stereoencephalography
SPECT	Single-Photon Emission Computed Tomography
SPHARM-PDM	Spherical Harmonics - Point Distribution Models
SUB	Subiculum
SVZ	Subventricular Zone
TE	Time Echo
TI	Inversion Time

TLE	Temporal Lobe Epilepsy
TP	True Positive
TR	Repetition Time
TSC	Tuberous Sclerosis Complex
VBM	Voxel-Based Morphometry
VI	Stochastic Variational Inference
WM	White Matter

Part I

Introduction

I Overview

Epilepsy is a brain disorder characterized by an enduring predisposition to generate epileptic seizures that affects over 50 million people worldwide. Underlying aetiologies include structural, metabolic, and genetic abnormalities [1]. A third of these patients suffer from seizures that are resistant to antiepileptic drugs. Consequences of recurrent seizures are severe with increased risk of injury, dire socioeconomic status, and even sudden death [2, 3]. In a large number of patients with drug-resistant focal onset epilepsy, a structural abnormality is identified; its surgical resection offers patients the chance at a curative treatment [4]. Temporal lobe epilepsy (TLE) secondary to mesiotemporal sclerosis [5], and extratemporal lobe neocortical epilepsy secondary to focal cortical dysplasia (FCD), a malformation due to abnormal neuronal proliferation and cortical organization [6], are the two most common drug-resistant epilepsies amenable to surgery [7]. In both syndromes, early identification of these anomalies allows timely resective surgery, limits the long-term effects of recurrent seizures and medication, and has been shown to have positive consequences on cognitive outcome and brain development [8–11], offering patients the chance at a curative treatment [4].

Recent advances in MRI have revolutionized the clinical management of drug-resistant epilepsies by allowing accurate *in vivo* identification of the primary lesions in increasing number of patients [12, 13]. Both neocortical dysplasias and hippocampal sclerosis constitute a spectrum of histopathology, clinical, and radiological manifestations [6, 7, 14–18]. Although the lesion generally presents with distinct radiological features, the degree and patterns of these phenotypes vary significantly across patients, possibly due to different underlying histopathological anomalies [19]. Notably, lesions with mild anomalies are often unremarkable on standard clinical MRI (referred to as MRI-negative), substantially affecting clinical management because of the difficulty in defining the surgical target [20].

Notably, epilepsies initially considered MRI-negative are not necessarily nonlesional since retrospective histopathological analysis in approximately 30–50% of MRI-negative patients reveals the presence of epileptogenic lesions – mainly FCD [21–23], mild hippocampal sclerosis [24–26], or subtle isolated neocortical or hippocampal gliosis [27, 28]. While mild cyto- and myeloarchitectural anomalies may not be directly visible on conventional imaging, they manifest quantifiably on high-resolution MRI, especially in conjunction with advanced computational modeling [29, 30].

Similarly, retrospective assessment of preoperative MRI scans, guided by quantitative structural image analysis, can frequently identify a lesion [20, 31]. These findings reinforce the importance of obtaining high-quality multimodal datasets interpreted by MRI experts to evaluate and treat patients with MRI-negative epilepsy.

Critically, there's no consensus on the definition of diagnostic workup that leads to a patient objectively being ascribed the MRI-negative status. In fact, there's significant variability in clinical workflows, expertise of personnel [32] assessing the radiological images, image acquisition protocols across site; in addition to myriad of biases that stem from human involvement in diagnostic radiology [33]. This is further complicated by several ambiguous definitions in the literature used to attribute MRI-negative status.

The overall goal of this thesis is to develop objective methods to improve the yield of MRI to resolve hard to detect epileptogenic lesions. Specifically, we first performed a comprehensive assessment of the literature to demonstrate biases associated with MRI-negative terminology in focal epilepsy. Subsequently, we employed a bipartite approach to developing algorithms to detect FCD, which rely on the integration of multiple imaging modalities through i) surface-based sampling that respects cortical topology and provides accurate inter-subject correspondence, and ii) minimally preprocessed volumetric approach that facilitates high generalization performance combining deep learning with uncertainty estimation for risk stratification.

Finally, we developed an algorithm for hippocampal subfield segmentation using minimally preprocessed MRI and assessed its lateralization performance in TLE.

The thesis is organized as follows:

1. Review of the background literature (Chapter 2).
2. Systematic review and meta-analyses of the literature to demonstrate current biases and variability in the MRI-negative literature (Chapter 3).
3. Automated surface-based machine-learning algorithm to detect FCD lesions initially overlooked by standard radiological diagnosis. (Chapter 4).
4. Deep learning method to detect FCD using minimally preprocessed multimodal volumetric data combined with uncertainty estimation and multi-site validation (Chapters 5, 6 and 7).
5. Deep learning classifier with label fusion for hippocampal subfield segmentation combined with seizure focus lateralization in TLE (Chapter 8).
6. Key findings and significance (Chapter 9).

2 Background

Temporal lobe epilepsy secondary to mesiotemporal sclerosis, and extratemporal lobe neocortical epilepsy secondary to FCD are the two most common drug-resistant epilepsies amenable to surgery [11, 34, 35] . Early identification of these anomalies allows timely resective surgery, limiting the long-term effects of recurrent seizures and medication, and has been shown to have positive consequences on cognitive outcome. Indeed, uncontrolled epilepsy is detrimental to the brain [36, 37] with long-term socio-economic consequences [38, 39]. The definition of epilepsies with unrevealing MRI is a moving target that constantly evolves with advances in diagnostic technology. While a consensus exists that the primary histopathological substrates are subtle dysplasias and isolated hippocampal gliosis, the MRI signature of these entities is not yet fully defined, mainly because of the lack of a unified methodology to evaluate their underlying structural changes. In this chapter, we discuss the aetiology of these two focal epilepsy syndromes, review the literature on localizing structural lesions on MRI, and advances in machine learning to optimize and automate lesion detection.

2.1 Malformations of cortical development

Any perturbation during in utero corticogenesis can trigger a cascade of molecular and biomolecular changes potentially resulting in malformations of cortical development (MCD) [40, 41]. These malformative processes may include alterations or dysregulation: of the proliferation rate, symmetric or asymmetric division patterns, mitosis, apical or basal attachment of progenitors and progenitor apoptosis [42, 43]. The most common manifestations are epilepsy, developmental delay and/or motor abnormalities of tone, movement and posture [41, 44]. The diagnostic pathway for MCDs is complex owing to wide variations in presentation and etiology, thereby hampering timely and adequate management [45]. Given that a timely and precise clinical diagnosis is critical for optimizing therapeutic approaches, developing effective biomarkers to accurately identify and characterize malformative lesions is a clinical priority.

2.2 Normal neurodevelopment

Human brain development is an intricate series of precisely orchestrated molecular cascades involving a complex interplay of diverse array of genes. The whole process is divided into three distinct phases: neuronal proliferation, migration, and post-migratory cortical organization [46–48]. An overall snapshot of these stages is depicted in Figure 2.1.

2.2.1 Neuronal proliferation: the formation of the ventricular and subventricular zones

Upon culmination of embryogenesis (4th gestational week, GW), the human forebrain forms a smooth sheet entirely occupied by neuroepithelial cells. During the early period of this phase, these cells divide symmetrically at the margin of the ventricle, gradually increasing the number of progenitor cells [47, 49], resulting in an increase in surface area and thickness of the ventricular zone (VZ) [50]. Constituent neurons are generated in the proliferative transient embryonic zones (such as the VZ and subventricular zone, SVZ), which are situated near the surface of the cerebral lateral ventricles. Starting at approximately the 5th GW, progenitor cells (also called radial glia cells) in the VZ begin to switch from symmetric to asymmetric cell division [51], the former resulting in a daughter and a radial glial cell, while the latter develops into either a post-mitotic neuron or an intermediate progenitor cell [52]. Notably, an accumulation of intermediate progenitor cells creates a new distinct compartment above the VZ, namely the SVZ [53, 54]. While the neuronal proliferation in the VZ begins to taper off beyond the 18th GW [55], the SVZ maintains its original proliferative role, actively producing more pyramidal neurons throughout the whole developmental period [56] (Figure 2.1A).

2.2.2 Neuronal migration: radial and tangential mode

A cardinal feature of the developing brain is that newborn neurons must translocate from their site of origin to their target regions at varying distances. Following the 7th GW, within the cortex, these neurons leave the ventral SVZ to reach their appropriate location within the developing cortical plate (CP) [57]. The CP, a primitive structure of neocortical gray matter (GM) in the mature brain [47], begins to develop as newborn neurons initiate the translocation of cell bodies from the proliferative zone to their target layers in the CP – transforming into the future six-layered cortex [57]. This elaborate cell movement, dubbed neuronal migration [58, 59] (Figure 2.1B), has two different modes: i) the radial migration, in which post-mitotic cells migrate vertically upwards from the VZ along a radial glial scaffold, and ii) the tangential migration, in which cells migrate parallel to the pial surface while being directed by surrounding molecular cues [60]. In general, neurons in the VZ take the radial mode of migration and accumulate in

the CP in a phenomenon known as the inside-out gradient of neurogenesis, where each generation of migrating post-mitotic neuron bypasses the previous one [46, 61, 62]. In other words, early born neurons form a layer at the most basal level of the CP and subsequently, the younger neurons travel through layers of older neurons to accumulate above them, gradually expanding the layers in outward direction. Eighty percent of these radially migrating neurons develop into excitatory glutamatergic neurons in the CP [62]. Conversely, neurons that develop into GABAergic inhibitory interneurons migrate tangentially by traveling several hundred micrometers parallel to the pial surface; these neurons were largely produced in the subpallium – a ventral forebrain area that contains the lateral and medial ganglionic eminences [60, 62–64]. These two structures eventually give rise to the basal ganglia and amygdala in the adult brain [63].

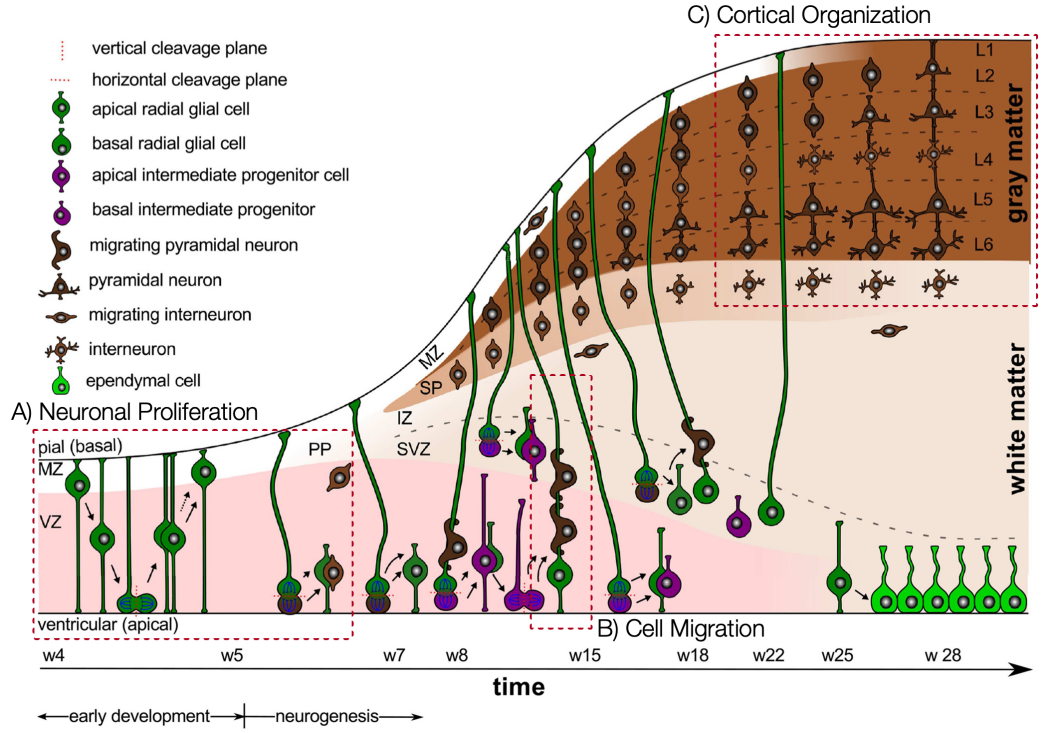


Figure 2.1: **An overall snapshot of corticogenesis**. **A**) Neuronal proliferation, **B**) Cell migration, and **C**) Cortical organization. Modified from Budday *et al.* [48] (licensed under CC-BY-4.0).

2.2.3 Post-migratory cortical organization

The period following the 22nd GW marks the most significant time for post-migratory cortical differentiation. The process occurs along two orthogonal directions: horizontal (areal) and vertical (laminar). After areal differentiation, the rostral regions of neocortex become specialized in

executive and motor functions, while caudal regions are engaged in somatosensory, auditory, and visual inputs [65]. Two distinct hypotheses have been proposed for this functional specialization: i) the proto-cortex and ii) the proto-map hypothesis [65, 66]. The proto-cortex hypothesis postulates that early formed uniform cortical areas selectively receive diverse somatosensory and high-order neuronal information from the thalamus [67]. This constant input from the thalamus signals the premature cortical areas to progressively identify and consolidate their final functions. In contrast, the proto-map hypothesis [68] suggests that during a period of neuronal proliferation, differential gene expression in newborn neurons in VZ guides them to attract appropriate thalamic inputs. Cortical differentiation also occurs along the vertical direction (Figure 2.1C). Indeed, laminar development that commences during the migration phase continues until late corticogenesis. While several theoretical and empirical studies have tried to decipher the mechanisms aiding this process, evidence is currently converging on the emergent effects of substantive interactions between layer-specific genes such as *Cux1-2* or *Foxp2* and a set of proteins diffused in extracellular matrix such as Reelin [69–71], which collectively regulate the layer positioning of migrating neurons. The cortex begins to manifest areas fully developed into six layers around 18th GW [46, 61]. Simultaneously, it also undergoes cellular differentiation, development of cell body [72], selective cell death and extensive axonal and dendritic expansion [48]. At 28th GW, the arborization of apical dendrites and tangential axons from early-generated neurons pervades layer I, and radial glial cells in subcortical layers either disappear or become astrocytes. During 24–34th GW, axons undergo myelination, gradually transforming the original intermediate zone into mature white matter (WM) tissue [40, 46, 73].

2.3 Cortical malformations subtypes

The updated taxonomy proposes a classification system based on the putative onset timing of malformative process [6] secondary to i) early abnormal neuronal and glial proliferation or apoptosis; ii) abnormal neuronal migration; iii) abnormal late post-migrational organization (see Table 2.1 for clinical characteristics of each group). Group 1 includes FCD (an isolated lesion associated with dysmorphic neurons and balloon cells). Group 2 includes periventricular or subcortical heterotopia (abnormally arrested neuronal clusters along the ventricular wall or between cortex and lateral ventricles) and classic lissencephaly (smooth brain). Group 3 includes polymicrogyria (excessive number of small gyri and shallow sulci) and mild cortical dysplasia (subtle cortical dyslamination, ectopic WM neurons). While this classification is generally useful and readily applicable in practical clinical diagnosis, actual genetic aetiology and behavioral outcomes across patients are often far more heterogeneous than what the taxonomy represents [74]. In subsequent paragraphs,

Table 2.1: Classification and clinicogenetic features of malformations of cortical development

Developmental stage	**Cortical malformation**	**Genetic etiology**†	**Clinical features**	**Incidence**†
Neuroglial proliferation	Microcephaly	ASPM, CDKRAP5, MCPH1	Mental retardation, not generally associated with epilepsy	< 1%
	Megalencephaly	PI3K-AKT	Mental retardation, early onset seizures	2%
	Focal cortical dysplasia Tuberous sclerosis	DEPDC5, NPRL3, MTOR, TSC1, TSC2	Normal-to-severe cognitive dysfunction	20–40%
Neuronal migration	Periventricular heterotopia	ARGGEF2, FLNA, LIS1	Neurodevelopmental delay, adolescent onset seizures	2–20%
	Subcortical band heterotopia	DCX, LIS1	Mental retardation, epilepsy	9%
	Lissencephaly	DCX, LIS1	Severe language deficit and social interaction, epilepsy	< 1%
Post-migratory development	Polymicrogyria	GPR56	Intellectual disability, movement disorder, seizures	5–16%
	A mild cortical dysplasia (without dysmorphic neurons)	Unknown	Cognitive decline, early onset epilepsy	13%

† Genetic etiology and incidence information based on [6, 44, 74–80]

we will review literature focusing on FCD and its variants – the primary research interest of the current thesis.

2.3.1 Focal cortical dysplasia

Focal cortical dysplasia (FCD) is one of the most frequently observed pathologies in drug-resistant extra-temporal lobe focal epilepsy (up to 50%) [7]. This malformation encompasses a broad spectrum of histopathological abnormalities including cortical disorganization as a cardinal characteristic. Associated features include cytopathology (large dysmorphic neurons and balloon cells) and gliosis caused by proliferation and hypertrophy of astrocytes [19]. Based on these anomalies, a new classification proposed a three-tiered classification system Table 2.2 [16]: i) FCD Type I is characterized by an isolated malformation with abnormal cortical layering, either showing persistence of vertical developmental microcolumns (IA) or loss of the horizontal hexalaminar structure (IB), or both (IC); ii) Type II presents with completely disorganized cortical layering and specific

cytopathology including dysmorphic neurons, either isolated (IIA) or together with balloon cells (IIB); iii) Type III comprises architectural abnormalities associated with either hippocampal sclerosis (IIIA), tumors (IIIB), vascular malformations (IIIC) or other lesions acquired during early life (IIID).

Compromised microcolumns in Type IA dysplasia is characterized by greater than 8 neurons (representing 2 standard deviations in healthy controls) aligned in a vertical direction along the cortex, predominantly in layers 3–4 [81]. The tissue harboring such microcolumns presents with a reduced cell size and increased neuronal densities, as well as a tendency of decreased cortical thickness, compared to healthy cortices [81, 82]. In stark contrast, dysmorphic neurons – the defining feature of Type II lesions – express either a pyramidal or an interneuronal phenotype and are characterized by significantly enlarged soma and nucleus relative to normal cortex [82]. While they share some commonalities with dysmorphic neurons (including a gigantic cell body and accumulated intermediate filaments [83]), balloon cells present with multiple displaced nuclei, and while being electrically silent [84, 85] have been implicated in mediating heightened inflammatory and immune responses [86]. So far only Type-II lesion has been systematically associated with specific genetic etiology (*e.g.,* DEPDC5, NPRL3, mTOR, TSC1, TSC2) [44, 77]. Abnormal expression of these genes disrupts normal signaling of mammalian target of rapamycin (mTOR), a core developmental pathway which governs initial cell growth and proliferation, resulting in immature balloon cells and dysmorphic neurons [85]. Both somatic and germline mutations have been identified in genes encoding mTOR cascade regulatory proteins in association with Type-IIA and Type IIB FCDs. The majority of these mutations have been somatic mutations occurring with highly variable allelic frequency, detected by whole-exome or whole-genome sequencing of DNA extracted from resected specimens compared with blood DNA from the same individuals [77, 78]. Interestingly, several recent studies have identified both germline and somatic mutations in other mTOR pathway regulatory genes in Type IIA and Type IIB FCDs. For example, loss-of-function mutations in DEPDC5 have been identified in Type IIA and Type IIB FCD specimens. Germline frameshift and splice-site DEPDC5 mutations were identified in Type IIB FCD specimens [87], while whole-exome sequencing identified a nonsense variant of DEPDC5 in Type IIA [88]. Thus, evidence seems to suggest that Type II FCDs are indeed mTORopathies at both molecular and cellular level [77]. This has led to pharmacological mTOR inhibitors being suggested and investigated as a potential therapeutic target for clinical trials [77, 89], and might represent a viable avenue for precision medicine in epilepsy based on the known disease mechanisms, used in isolation or in conjunction with surgery [90].

Table 2.2: Three-tiered classification system of focal cortical dysplasia [16]

Cortical dysplasia Type-I (isolated)	Cortical dysplasia with abnormal radial lamination (Type IA)	Cortical dysplasia with abnormal tangential cortical lamination (Type IB)	Cortical dysplasia with abnormal radial and tangential cortical lamination (Type IC)	
Type-II (isolated)	Cortical dysplasia with dysmorphic neurons (Type IIA)		Cortical dysplasia with dysmorphic neurons and balloon cells (Type IIB)	
Type- III (associated with a principal lesion)	Cortical lamination abnormalities in the temporal lobe with hippocampal sclerosis (Type IIIA)	Cortical lamination abnormalities adjacent to a glial or glioneuronal tumor (Type IIIB)	Cortical lamination abnormalities adjacent to vascular malformation (Type IIIC)	Cortical lamination abnormalities adjacent to any other lesion acquired during early life (Type IIID)

2.4 The role of MRI in characterizing FCD

Magnetic resonance imaging (MRI) has revolutionized the management of drug-resistant epilepsy because of its unparalleled ability to visualize epileptogenic lesions *in vivo* [12]. Recent advances in MRI acquisition technology, specifically high-field (3T) and ultra-high field (7T) imaging combined with multiple phased-array head coils, have allowed for an increasingly precise characterization of the primary lesion, facilitating improved description and classification of malformations of cortical development [74, 91–93]. Wang *et al.* [94] probed the diagnostic value of 7T over 3T, while also evaluating the yield facilitated by MRI post-processing, demonstrating its diagnostic utility over conventional visual review. More recently, a systematic review [91] demonstrated the superiority of 7T in detecting lesions missed on 3T or lower. However, beside the unavailability of 7T scanners in most countries, its high cost of operation and maintenance, because of inconsistencies in the definition of MRI-negative, so far it has been infrequent to see dysplasias at 7T that are completely invisible at 3T [95–97]. Moreover, the additional role of MRI post-processing in augmenting lesion localization at 3T or lower fields was not considered, signifying lost opportunity to fully leverage the MRI.

2.4.1 Imaging characteristics of FCDs

Cardinal features of FCD on structural MRI include abnormally thick GM (50–92% of cases) and blurring of the GM-WM interface (60–80% of cases) [7, 98]. Analysis of T2-weighted images reveals GM hyperintensity in 46–92% of lesions and sensitivity of FLAIR images is even higher (71–100%). The typical transmantle sign, a neurodevelopmental remnant of disrupted cell mi-

gration along the radial glial processes, presents as a funnel-shaped hyperintensity extending from the lateral ventricle to the lesion, manifesting in majority of Type II cases [17, 98, 99].

The *in vivo* visibility of dysplastic changes on MRI generally is analogous to the degree of histopathological abnormalities [7]. Even in patients with a Type II dysplasia, as the radiological spectrum on MRI encompasses variable degrees and patterns of GM and WM changes, visual identification can be challenging, particularly when inspecting the convoluted neocortex in two-dimensional slices (see Figure 2.2). Indeed, recent surgical series indicate that up to 33% of Type II and 87% of Type I dysplasia [17, 18, 98] present with "unremarkable" routine MRI. This clinical difficulty has motivated the development of computer-aided methods aimed at analyzing brain morphology and signal intensities. Such procedures provide distinct information through quantitative assessment without the cost of additional scanning time.

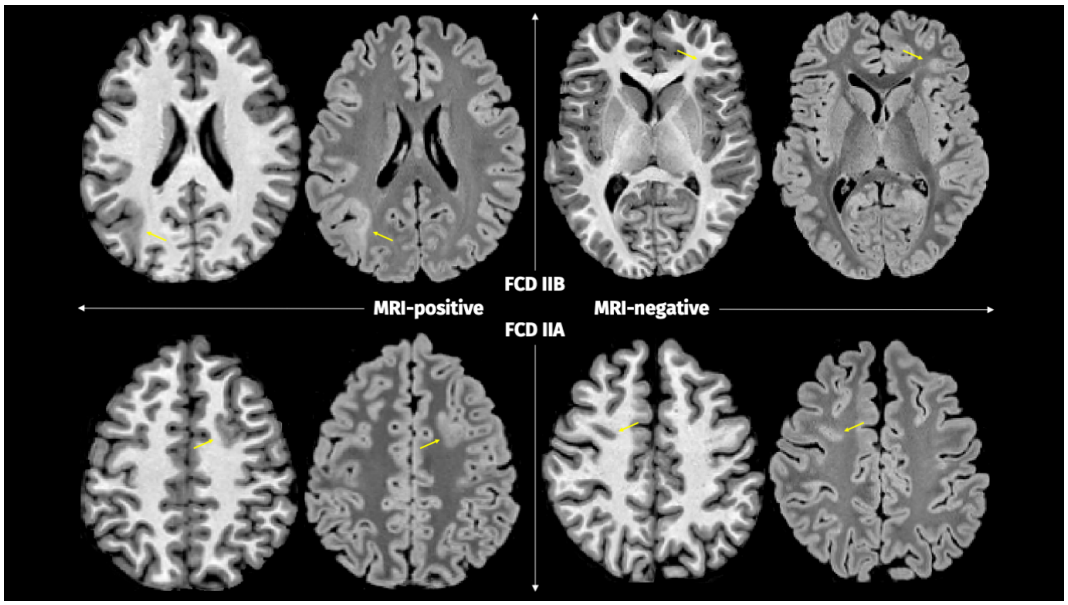


Figure 2.2: **Multimodal MRI of FCD Type-II**. Two representative cases each (T1-weighted on the *left* and T2-weighted FLAIR on the *right*) of FCD Type-IIB (*top panel*) and FCD Type-IIA (*bottom panel*) are illustrated. The *top-right* and *bottom-right* quadrants are MRI-negative, while the *top-left* and *bottom-left* are MRI-positive cases. Lesions are indicated by the yellow arrows.

2.4.2 Voxel-based lesion detection frameworks

Voxel-based morphometry (VBM) is a widely applied imaging post-processing technique, originally developed to probe whole-brain tissue morphology *in vivo* at the population- or group-level by spatial normalization of all subjects to a common stereotaxic or template space [100]. It relies on automated tissue segmentation [*i.e.,* GM, WM and cerebrospinal fluid (CSF)] followed by

gaussian image smoothing to generate voxel-wise brain tissue density maps, on which statistical analyses can be performed. VBM has been co-opted to detect structural abnormalities related to MRI-visible dysplasia in single patients by several clinical and research groups [20]. This automated image processing identifies differences in tissue density at a voxel level, detects increases in GM concentration co-localizing with the lesion in 63–86% of cases. Histopathological confirmation of lesions that eluded visual inspection (despite their relatively large size) [101] suggests that VBM may be applied to investigate patients with MRI-negative epilepsy. Importantly, however, a threshold of 2×SD (or 95th percentile) above the mean GM concentration in healthy controls does not guarantee specificity of findings – false positives may still occur in control subjects at this threshold level. VBM has also been used to analyze intensities derived from quantitative MRI contrasts such as T2 relaxometry, double inversion recovery, and magnetization transfer imaging. These approaches have demonstrated high sensitivity (87–100%) in detecting conspicuous or obvious malformations of cortical development [102–105]. Nevertheless, these techniques may fail to identify areas concordant with clinical and EEG findings in more than two-thirds of MRI-negative cases [102, 103].

The relatively unspecific nature of VBM with respect to pathological characteristics of cortical dysplasia has motivated the search for computational models of morphological imaging features distinctive of the lesion. In the same vein, Bernasconi *et al.* introduced a novel approach to integrate voxel-wise textures and morphological modeling of three main features in the lesion (*i.e.,* cortical thickening, blurred GM-WM junction and relative intensity alterations) into a unified composite map [106] (Figure 2.3A). This semi-automated algorithm demonstrated the value of computer-aided detection of FCD (yielding 88% sensitivity and 95% specificity), substantially improving the detection rate over conventional visual identification. Moreover, integrating these models with the quantification of higher-order image texture features invisible to the human eye, detected ~80% of dysplastic anomalies [107, 108]. Other VBM methods exist, such as the Morphometric Analysis Program (MAP) [109]; this technique relies on a "junction" image – a three-dimensional map which identifies voxels that are unequivocally neither GM nor WM to quantify the likelihood of tissue blurring. Beyond lesion detection, our group developed an automated lesion segmentation algorithm based on level-set-based deformable models was proposed [110] to better delineate the boundaries of the lesion (Figure 2.3B).

Although voxel-based approaches offer an automated and exploratory analytic framework of whole brain structural changes, their utility so far has been limited since they are unable to account for the complex sulco-gyral topology of the human brain while being prone to volumetric (Euclidean geometry) averaging of adjacent and non-adjacent cortical regions across sulci and gyri, potentially increasing the incidence of false positives. Moreover, high anatomical variability in gyrification and sulcation across individuals may reduce the specificity and sensitivity to de-

tect subtle effects. These limitations restricted the clinical utility of voxel-based approaches and catalysed the development and adoption of alternative frameworks.

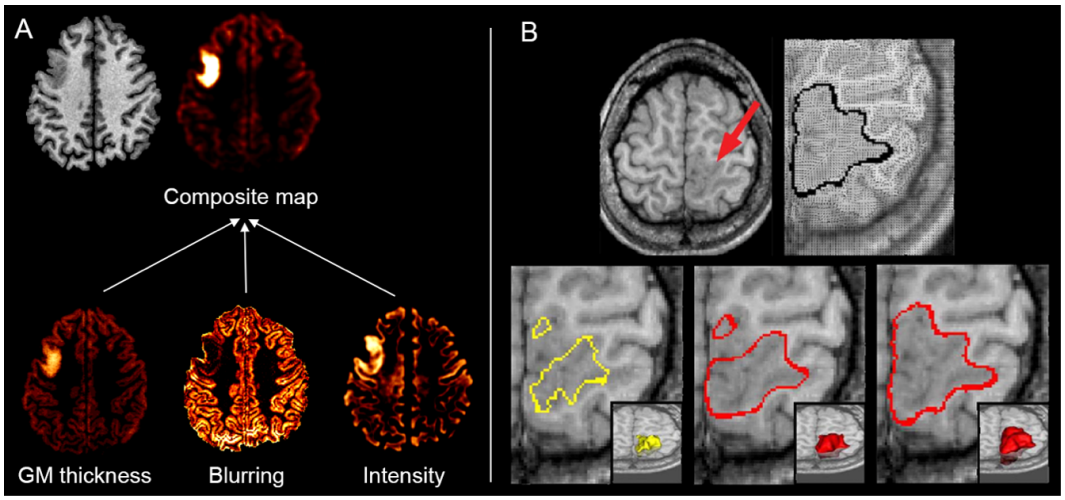


Figure 2.3: **Voxel-wise post-processing for lesion detection and segmentation**. A) *Upper panel*: A cortical dysplasia located at the left frontal lobe is highlighted in a composite map. *Bottom*: Individual features present with increased GM thickness, altered intensity and reduced gradient of the lesion. The composite map is computed as follows [(GM thickness × relative intensity) / intensity gradient]. B) Lesion segmentation. *Upper*: A red arrow indicates the lesion. The final result shown together with gradient vector flow. *Bottom*: After initializing with the lesion detection result of the classifier (yellow), a deformable model gradually expands the lesional boundary (red) following gradient vector flow [Modified with permissions from 110, 111].

2.4.3 Anatomically plausible surface-based analytic framework

In contrast to the early VBM approaches, surface-based morphometric (SBM) measurements offer a more direct and anatomically plausible way to quantify cortical structural integrity as these methods accurately model the sulco-gyral complexity of the convoluted cortical surface. Several image processing methods have been developed to enable the measurement of cortical surfaces along the corresponding points between the GM and WM interface across the entire cortical mantle [112, 113]. The CIVET pipeline developed at the Montreal Neurological Institute, one of the widely used image processing pipelines, leverages the Constrained Laplacian Anatomical Segmentation using Proximity (CLASP) algorithm for cortical surface extraction [114].

Surface-based lesion detection (Chapter 4) in this thesis leveraged the CIVET/CLASP pipeline (Figure 2.4) for surface reconstruction. Briefly, the native-space T1-weighted MRI images serve as input in this pipeline, which undergo a non-uniformity intensity correction [115] and linear registration to standardized stereotaxic template based on the Talairach atlas [116]. Images registered in the stereotaxic space are classified into GM, WM, and CSF using an automated classifier

that takes into account intensity information as well as spatial anatomical priors that were derived from a large training set [117]. Recently, in order to address challenges during segmentation such as regional variation in intensity due to local radiofrequency artifacts or disparities in tissue composition, a novel anatomy-driven algorithm has been proposed [118]. This approach carries out tissue segmentation within a small local parcel which conforms to the cortical anatomy, therefore, maximizing the regional tissue contrast and significantly improving the GM-WM border definition compared to conventional methods. A subsequent partial volume classification step is invoked that improves the detection of deeper buried sulci as well as the discrimination between insular cortex and subcortical GM structures [119]. To generate the model of the cortical surface, the CLASP algorithm iteratively warps a surface mesh to fit the boundary between WM and GM in the classified image. It then expands the WM/GM boundary along a Laplacian map to generate a second outer surface that runs along the GM/CSF boundary [114]. To improve anatomical correspondence of vertices in all subjects, surfaces are then non-linearly aligned to an iteratively generated surface template [120] using a two-dimensional registration procedure that minimizes differences in cortical folding [121]. By achieving a better alignment of sulco-gyral patterns across subjects, this procedure demonstrated higher sensitivity to detect group-level findings compared to conventional voxel-based methods [122]. Subsequent quantification using a battery of morphometric features including cortical thickness, folding complexity and sulcal depth, and modelling of MRI intensities and gradients at varying depths of the cortical ribbon provide an integrative description of whole-brain cortical integrity.

Compared to voxel-based techniques, a surface-based approach preserves cortical topology and quantifies sulco-gyral anomalies, at times the only sign of dysgenesis [123]. Over the last decade, a number of automated FCD detection algorithms have been developed [31]. Recent FCD detection methods rely on surface-based approaches [124–130] allow effective morphology-aware sulco-gyral modelling. The multi-centre epilepsy lesion detection (MELD) project [124, 125] – a multi-site initiative – has endeavoured to mitigate the significant technical burden; the performance, however, has been suboptimal, partly due to their exclusive reliance on surface-based preprocessing. Moreover, the focus is on accessibility and open source rather than developing generalizable algorithms. Notably, while the laborious and arduous pre-processing, and quality control (correction of segmentation and surface extraction errors) has contributed to the high-fidelity of the features facilitated by specialized expertise – this significant technical burden has precluded their broader integration into clinical workflows.

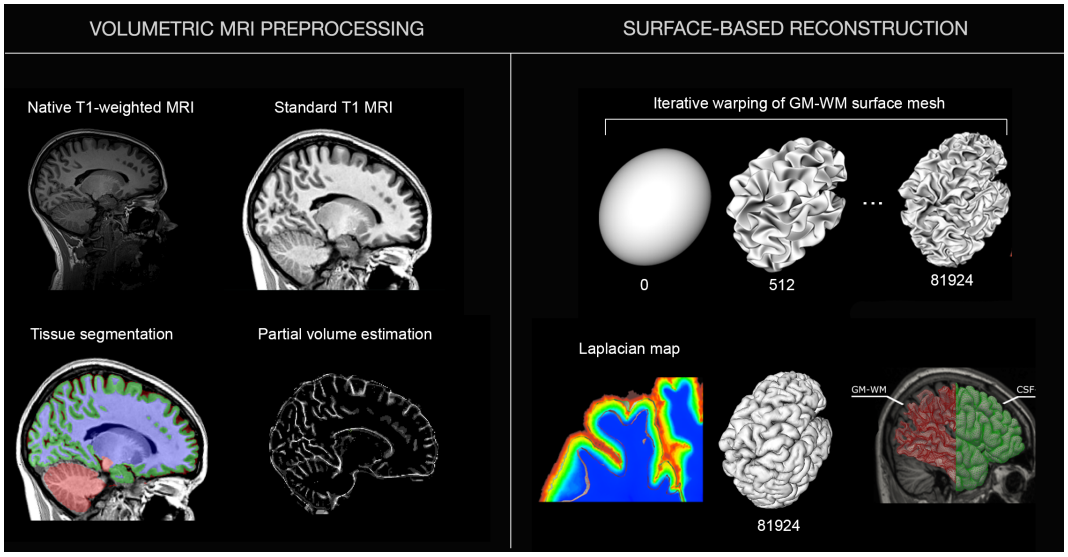


Figure 2.4: **Cortical surface extraction**. Please refer to the subsection 2.4.3 for brief description of the methodology [Modified with permission from 111].

2.5 The role of MRI in characterizing HS

Hippocampal sclerosis (HS) is the most common histological feature of medically refractory temporal lobe epilepsy (TLE) in many patients, characterized on MRI by atrophy and loss of internal structure together with increased T2-weighted signal intensity. The degree of atrophy has been shown to correlate with the severity of neuronal loss in the *cornu ammonis* 1 (CA1) [131]. Notably, the *in vivo* signature of HS is modulated by the severity of cell loss and gliosis [132–134] with subtle forms typified by isolated gliosis [135] often evading detection [136, 137].

MRI volumetry has been the most commonly employed quantitative technique to assess mesiotemporal lobe pathology as it is more sensitive than qualitative visual evaluation – lateralizing the seizure focus in 70% of patients [138]. Manual delineation of the hippocampus is considered to be the gold standard, as it is accurate, reproducible, and sensitive [139–141]. However, it is time-prohibitive and prone to rater-bias (high intra- and inter-rater variability). Consequently, manual volumetry of mesiotemporal lobe structures, remains largely underutilized in the clinic. Furthermore, T2-relaxometry provides a quantitative estimate of T2-weighted signal relative to the qualitative analysis of T2-weighted MRI, demonstrating an increase in sensitivity in detecting mesiotemporal gliosis [142, 143]. Bernasconi *et al.* [143] demonstrated the ability of T2-relaxometry to accurately lateralize the seizure focus in 82% of patients with normal hippocampal volume. Since volumetry provides a global estimate of atrophy, its sensitivity to detect subtle diffuse or focal anomalies is limited. This may explain its failure in 30–40% of patients with ambiguous electroclinical features of drug-resistant TLE, despite histopathology revealing subtle

sclerosis. In these cases, three-dimensional shape analysis has the potential to refine the subfield-specific MRI correlates of hippocampal pathology [134, 144–146]. A recent surface-based method relying on spherical harmonic shape descriptors localized submillimetric variations of volume to model the intrinsic geometric properties in a given structure relative to a template while guaranteeing anatomical correspondence across subjects [147], a prerequisite for reliable group-level inferential statistics. This technique has demonstrated effectiveness in detecting subtle atrophy in patients with normal hippocampal volume.

A fully automated approach is key to capitalizing the predictive potential of hippocampal volume and shape abnormalities for clinical prognostics. Beyond epilepsy, the increased demand to study large cohorts of healthy and diseased populations has motivated the development of automated segmentation procedures. Catalyzed by the advances in MR hardware and sequence technology enabling submillimetric image acquisition with improved signal-to-noise ratio, it is nowadays possible to resolve hippocampal substructure. These subfields or subregions include the dentate gyrus, subiculum, parasubiculum, entorhinal cortex, and the cornu ammonis (CA1-4) regions. Although several delineation protocols have been proposed, consensus is scarce [148, 149]. Consequently, reproducibility becomes problematic since results cannot be adequately compared across studies deploying different protocols. Winterburn *et al.* [150] proposed an *in vivo* high-resolution atlas to segment the hippocampus into five distinct structures in five healthy controls: CA1, CA2-3, CA4-DG, Stratum and Subiculum. Soon after, Kulaga-Yoskovitz *et al.* [151] proposed a segmentation protocol consisting of three structures: CA1-3, CA4-DG and Subiculum. Both protocols were delineated on high-resolution T1- and T2-weighted images and have aided validation efforts in contemporary segmentation algorithms.

Several methods have been developed for MRI-based subfield segmentation [152–154]. ASHS [155] uses a multi-atlas approach coupled with a similarity-weighted voting and a boosting-based error correction. *FreeSurfer* (*v.* 6 [154]) includes a probabilistic atlas-based procedure leveraging high-resolution post-mortem samples. *HIPS* [156] based on multiatlas label fusion method (OPAL [157]) obtained state-of-the-art results with fast inference times. Caldairou *et al.* [158] proposed *SurfPatch*, a surface patch-based segmentation method combining patch-based template library and feature matching. Although these methods yield promising results (average Dice ~88%), most remain compute-intensive, result in subpar segmentation quality relative to the ground truth labels, or fail to demonstrate effectiveness when presented with subtle pathology [159].

2.6 The concept of MRI-negative in focal epilepsies

The definition of epilepsies with a negative MRI reading is a moving target that has constantly evolved in synchrony with advances in diagnostic technology. Several variations of the nomenclature exist in the literature, which include nonlesional, normal, unremarkable, cryptogenic, and MRI-negative, among others. In the past few years, however, it has become increasingly clear that epilepsies initially considered to be cryptogenic are not necessarily nonlesional. Indeed, in 30–50% of those patients who undergo surgery, histological examination of the resected specimens reveals the presence of epileptogenic lesions [17, 160], rendering the term "nonlesional" and "cryptogenic" unsuitable. Finally, the term "MRI-negative" epilepsy better conveys the meaning where presurgical MRI is devoid of an apparent structural abnormality as the most likely cause of the epilepsy.

While a consensus exists that the primary histopathological substrates of MRI-negative epilepsy are subtle hippocampal sclerosis (mainly isolated gliosis) and dysplasias, the MRI signature of these entities remains to be fully defined, primarily due to the lack of a unified methodology in evaluating their structural substrates. This limitation can be partially attributed to sulco-gyral complex layout [20]. The advent of higher-field magnets at ≥3T, combined with phased array coils replacing the conventional quadrature coil, has resulted in quantifiable improvements in image signal-to-noise and contrast-to-noise ratios, yet consensus around their definition on MRI is still lacking. Consequently, many patients with unrevealing MRI undergo intracranial EEG, a procedure that carries risks similar to resective surgery [161, 162] and incurs high costs [163]. Besides, retrospective assessment of preoperative MRI, often guided by quantitative structural image analysis, can unveil these occult lesions [164–166]. These findings reinforce the importance of obtaining high-quality images, interpreted by MRI experts. For example, in most reported cases, images obtained at 3T allow more detailed and complete delineation and characterization of structural changes related to FCD than those obtained at 1.5T, revealing lesions previously unseen.

Recently, HARNESS-MRI protocol proposed by the 2013–2017 Neuroimaging Task Force aimed to standardize best-practice neuroimaging of epilepsy in outpatient clinics and specialized surgery centers alike on conventional scanners (1.5T and 3T), without significantly increasing scan times [167]. The advantages for higher field acquisitions are inevitable. Nevertheless, they are capital-intensive investments and not broadly accessible.

The important role of MRI education in epileptology was recognized in the recent competency-based educational curriculum proposed by the ILAE education taskforce [168], which promotes advanced courses in MRI interpretation and post-processing for identifying lesions on MRI, culminating in a professional certification. The primary goal is to reduce ineffi-

ciencies in clinical decision-making, taking cognisance of the time and cost demands for various diagnostic tests and their potential impact on delays of initial therapy [169]. This updated curriculum signifies an important step in upgrading the neuroradiological expertise in clinical epileptology [32], and thus, potential reduction in lesions overlooked on MRI.

Advanced MRI post-processing remains underutilized as a potent analytic and diagnostic tool. Indeed, MRI post-processing enables objective analysis that is both replicable and rater independent, with each step in the process (such as intensity non-uniformities correction, classification into tissue types, and image registration) being transparent and accessible to quality control. In addition, post-processing techniques can be readily performed on clinical MRI data to successfully identify MRI-negative lesions that evade detection by conventional radiology [105, 106, 166, 170, 171]. In a recent cohort of 188 FCD patients, PET was only marginally better at detecting lesions than conventional MRI review (78.2% *vs.* 75.8%) [172]. This gap could be potentially bridged by adopting MRI post-processing in the routine clinical workflow. Thus, despite the added technical burden, advanced post-processing offer a favorable cost-benefit ratio.

2.7 Leveraging deep learning to improve diagnostics

The seminal work behind today's cutting edge machine learning (and by extension, artificial intelligence) technology is rooted in the neuron-inspired Connectionism school of thought dating back to 1940s [173]. However, it was not until the beginning of the last decade that advances in computer hardware made it computationally tractable and thus, practical for big data applications. Deep learning, in particular, has enabled increasingly substantial performance leaps in computer vision, natural language processing, object detection and recognition [174–177], as well as medical imaging [see 178, 179, for review]. Briefly, a deep learning neural network consists of unimodal or multimodal digitized inputs (images, timeseries, graphs, etc.), which are processed through multiple layers of interconnected units (or 'neurons') that detect and learn progressively high-level features, and ultimately provides a task-specific predictive output [177, 180].

A special class of deep learning algorithms called convolutional neural networks (CNNs) that operate on minimally preprocessed volumetric data are promising candidates for their unprecedented generalization capabilities across the range of clinical specialities, including but not limited to radiology, pathology, dermatology, ophthalmology, cardiology, and brain disorders [179, 181], attaining or exceeding physician-level diagnostic performance [182]. This steadfast adoption has been fueled partly by the increasing realization that current statistical learning methods are inadequate to derive clinically relevant and actionable insights from labelled big data [183, 184]. These deep neural networks can aid interpretation of complex and heterogeneous data including medical scans, pathology slides, skin lesions, retinal images, electrocardiograms, endoscopy, continuous

vital signs monitoring, medical notes entered by physicians, and genomic data to help make medically relevant predictions [179, 183, 185]. Most notably, CNNs can hierarchically extract insights from the data without exploiting domain knowledge to engineer meaningful features.

Although still in their infancy in epilepsy, deep learning algorithms have seen broad spectrum application, ranging from seizure detection based on EEG [186, 187], video and facial analysis [188], MRI-based detection of epileptogenic lesions [189, 190] and lateralizing hippocampal sclerosis in TLE [191], distinguishing epileptic from nonepileptic seizures based on wearable sensors, and outcome prediction for medical and surgical management of epilepsy [see 192, for review]. However, these implementations are generally perceived as cumbersome and difficult to implement and reliably reproduce outside of specialized epilepsy units [193]; and have yet to deliver clinically adequate generalizability and detection capabilities for subtle epileptogenic lesions [20]. Furthermore, most studies rely on small sample sizes with limited external validation, which is essential for broader adoption of models in clinical practice [192].

While lesion detection algorithms in FCD have relied on SBM methods [31], they are being progressively superseded in performance by deep learning techniques using volumetric data [189, 190, 194–196], which require only minimal pre-processing, manual intervention and quality control, thus setting the bases for widespread use in clinical settings. Still, deep learning requires large corpus of expertly labelled annotations (ground truth) to train and optimize the network, both cost- and time-prohibitive endeavors, resulting in suboptimal cohort sizes. Therefore, we propose a patch-based augmentation (extracting several hundred overlapping patches from a single subject) that enables scaling up the data without the requirement of an impractically large cohort size, while leveraging the state-of-the-art advances, in detecting MRI-negative FCD-II lesions (Chapter 5). Compared to previous deep learning methods for FCD detection [190, 195, 196] in which clinical description was scarce to absent, and information on the FCD expert labels and histological validation of lesions was not provided, our study relied on best-practice multimodal MRI, histologically-validated lesions, and a large dataset. Notably, we leveraged uncertainty estimation to derive diagnostic confidence from the classifier predictions, a property insofar unexplored by contemporary implementations for lesion detection. There are two main forms of uncertainty in modelling. Aleatoric uncertainty captures measurement noise, for example, noise due to the instrument miscalibration and faulty sensors or motion blur in MR images. In contrast, epistemic or model uncertainty accounts for uncertainty in the model parameters, capturing the incertitude about which model or underlying distribution generated the collected data [197]. Model uncertainty can be further reduced as more data is observed, while aleatoric uncertainty is irreducible despite adding data [198, 199]. Model uncertainty is therefore important for safety-critical applications to understand examples which diverge from the training data, and small datasets where the training data is sparse; estimating uncertainty is also a desirable property in clinical decision

support. Due to their over-parametrized nature (*i.e.,* large number of parameters relative to the available data), these neural networks can represent several models or solutions (owing to the presence of local optima) that fit the data [197], which can be leveraged to derive uncertainty from the model predictions. In contrast, traditional machine learning approaches are unable to estimate and incorporate uncertainty in predictions [198]. In clinical domains, uncertainty information has insofar been used to evaluate the robustness of predictions in multiple sclerosis [200] and diabetic retinopathy [201], quality control of whole-brain segmentation [202], data enhancement [203], expert label disagreements [204], while also being actively investigated in facilitating explainability and interpretability of black-box AI algorithms [197, 205].

In relation to hippocampal segmentation, recently, due to the widespread adoption of deep learning in medical imaging, there has been a resurgence in volumetric methods aimed at improving accuracy (Dice > 90%). More specifically, CNNs have demonstrated unprecedented results in subfield segmentation, outperforming prior traditional approaches. *DeepHIPS* [206] demonstrated < 1% improvement in Dice overlap indices (averaged across all subfields) over the ResDUnet implementation, while ResDUnet reiterated the same degree of improvement over *HIPS* [156] and 3D-UNet [207]. While most implementations are tested on either the Winterburn [150] and/or Kulaga-Yoskovitz [151] datasets, the exact training and validation data splits are not reported, potentially explaining the ~1% variation in performance across implementations. This might underscore why Dice in isolation is not a good comparative performance metric, especially in the clinical context where 1% difference is unlikely to make a meaningful impact on the diagnosis. Moreover, while several other methods have been proposed, the closed source nature of most implementations hinders fair comparisons. In Chapter 8, we propose a multi-scale patch-based paradigm to segment subfields using a UNet variant in TLE patients. To mitigate the issue of unfair comparisons, rather than rely solely on reported metrics, we propose testing the efficacy of our algorithm on seizure focus lateralization on an independent validation cohort of patients that were never used to train the model.

✦✦✦

Part II

Projects

3 MRI-negative Epilepsy: A Systematic Review & Meta-analysis

Preface

Despite its prevalence in the medical parlance, the term "MRI-negative" and its variations (normal MRI, nonlesional, and cryptogenic) remain subjectively defined as the absence of any structural lesion on MRI. Moreover, this nomenclature is prone to bias and is highly correlated with the field strength, the MRI protocol [167], post-processing [20] and expertise of the reviewer [32], particularly when dealing with subtle lesions. Critically, identifying the epileptogenic substrate on MRI and its surgical resection offers drug-resistant epilepsy patients the chance at a curative treatment [4].

This first study, thus, aimed to synthesize evidence assessing the ambiguity in criteria used to ascribe MRI-negative status. Subsequently, we evaluated the diagnostic yield of MRI post-processing compared to standard visual evaluation and assessed the effects of MRI diagnostic status (MRI-negative *vs.* MRI-positive) on post-surgical outcome in two separate metanalyses.

Manuscript I

In preparation

MRI-negative Epilepsy: A Systematic Review & Meta-analysis

Ravnoor S. Gill [1], Francesco Deleo [3], Seok-Jun Hong [1], Fatemeh Fadaie [1], Benoit Caldairou [1], PhD; Hyo M. Lee [1], Boris Bernhardt [2], Neda Bernasconi [1], MD PhD; Andrea Bernasconi, MD [1]

[1] Neuroimaging of Epilepsy Laboratory and Department of Neurology and Neurosurgery, [2] Multimodal Imaging and Connectome Analysis Laboratory, Montreal Neurological Institute, McGill University, Montreal, Canada; [3] Istituto Neurologico Carlo Besta, Milano, Italy

Abstract

Objective. To synthesize evidence assessing ambiguity in criteria used to define MRI-negative in the context of a systematic review and meta-analysis.

Methods. In a systematic review of Medline, Embase and Cochrane databases – from Jan 1, 1990, to Mar 31, 2021, we identified English-language observational cohorts (N > 5) with MRI-negative drug-resistant focal epilepsy. Studies were first assessed semi-quantitatively using unsupervised machine learning, allowing stratification based on the co-occurrence of various neuroimaging modalities. Within the identified clusters, we assessed the rigour of diagnostic reporting in terms of MRI contrasts and parameters, MRI post-processing, rater expertise and incidence of SEEG. A subset of studies served to analyse effects of MRI-negative status on post-surgical seizure outcome, and MRI post-processing on diagnostic yield using random-effects meta-analytic pooling.

Results. A total of 196 studies were included in the systematic review. Overall, 91 (46%) studies provided data exclusively on MRI-negative patients, while 105 (54%) provided data on mixed cohorts with MRI-negative and MRI-positive patients, for a total of 7,436 MRI-negative (median age 30, IQR 27–36) and 4,585 MRI-positive (median age 31, IQR 27–36). Unsupervised clustering based on diagnostic modalities revealed three groups: *MRI-dominant* , *nuclear-imaging* and *limited-MRI-information*. Rigour of MRI parameter reporting ($ORs > 5.3$, $ps < 10^{-5}$) and post-processing was exemplified in the *MRI-dominant* group as opposed to the *limited-MRI-information* group ($\chi^2 = 37.5$, $p_{Bonferroni} < 10^{-8}$). The *limited-MRI-information* group was less likely to report the rater expertise ($ORs > 15.6$, $p_{Bonferroni} < 10^{-8}$). Finally, SEEG was the hallmark of the *nuclear-imaging* (Odds Ratio, $ORs > 3.7$, $ps < 0.02$) group.

Meta-analysis of 61 studies reporting post-surgical outcome yielded a significant pooled proportion ($\chi^2 = 13.9$, $p < 0.01$) of favorable outcome in 75% of MRI-positive (95% CI 67–84%; $I^2 = 81\%$) and 59% of MRI-negative patients (95% CI 50–59%; $I^2 = 56\%$). In addition, 50 studies reporting diagnostic yield resulted in a significant pooled OR ($\chi^2 = 4.2$, $p = 0.04$) of 11.4 for MRI post-processing (95% CI 7.3–17.8; $I^2 = 43\%$) and 5.9 for qualitative MRI review (95% CI 3.5–9.8; $I^2 = 36\%$), a two-fold (11.4/5.9) gain in diagnostic yield over qualitative review.

Conclusions. Our systematic review and meta-analyses support the need for an objective definition for ascribing MRI-negative status. Beside the value of detailing field strength, sequences and parameters and the expertise of the reader, our findings suggest that MRI post-processing should be systematically performed when assessing patients with medically intractable seizures, particularly those with suspected MRI-negative epilepsy in search of prevalent epileptogenic lesions, specifically hippocampal sclerosis or focal cortical dysplasia.

3.1 Introduction

In clinical practice, epilepsy patients, particularly those with drug-resistant seizures who are candidates for surgery, are commonly dichotomized based on lesion visibility on structural MRI. Indeed, even though the characterization of the seizure focus relies on the convergence of evidence stemming from seizure semiology, EEG, neuroimaging and neuropsychology [193, 208], a positive MRI is the strongest prognostic factor for postoperative seizure freedom [10, 209]. Conversely, patients with non-diagnostic MRI currently represent an utmost challenge [210]. Indeed, notwithstanding long and costly hospitalizations for EEG monitoring with intracerebral electrodes [162, 211], surgery is less likely to be performed [22, 212]. A previous meta-analysis conducted a decade ago stipulated that the odds of seizure freedom are lower in the absence of a lesion on MRI [4, 213].

Patient categorization based on MRI underlines the inherent importance of imaging. Yet, a binary distinction has also created biases in the literature, with many patients with histologically-verified lesions being initially misclassified as MRI-negative [4, 7, 214]. While there are indications that several criteria, such as appropriate MRI protocol [167], the expertise of the reviewer [32, 215] and post-processing [20] may optimize the detection of previously unrecognized lesions, the degree to which these criteria are applied across studies remains unknown. In other words, in many studies the criteria used to label a patient as MRI-negative are undefined.

MRI-negative epilepsy has profound implications on treatment strategies and results. Our purpose was to perform a systematic review to synthesize evidence assessing the consistency of criteria used to define MRI-negative. In addition, we conducted the first meta-analysis evaluating the diagnostic yield of post-processing compared to standard visual evaluation. Another meta-analysis assessed effects of MRI diagnostic status on post-surgical outcome.

3.2 Methods

3.2.1 Search strategy and selection criteria

We conducted a systematic review and meta-analysis in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines [216]. We searched Medline, Embase, and the Cochrane Central Register of Controlled Trials databases for studies published in English language between January 1990 and March 2021, querying the following terms: "epilepsy `OR` seizure" `AND` "cryptogenic `OR` nonlesional `OR` MRI-negative `OR` normal MRI". We included studies conducted in human adults or children with drug-resistant focal epilepsy and screened the reference lists of all included studies for additional relevant publications. We excluded reviews, case series with fewer than five patients, conference abstracts, drug trials, epidemiological

studies, as well as those on new-onset and drug-responsive epilepsies, genetic generalized epilepsies, and encephalopathies.

3.2.2 Study screening and selection

Database query and initial screening was performed by one rater (RSG) for the entire list. After removal of duplicates across databases, screening based on abstracts was done and full-text articles were retrieved. Each full-text article was assessed independently by the two raters (RSG and FD); disagreements were resolved by consensus. Based on 2,530 articles, we obtained an excellent inter-rater agreement (Cohen's *Kappa*, $\kappa = 0.95$).

3.2.3 Data extraction and quality evaluation

For the included studies, we extracted the following information using a standardized form: (i) *MRI definition*: MRI-negative, non-lesional, normal MRI, cryptogenic; (ii) *Demographics and clinical characteristics*: sample size, age, gender syndromic definition or lobar seizure focus localization, surgical intervention histopathology, post-surgical outcome; (iii) *MRI acquisition parameters*: field strength, contrasts, voxel resolution (when not explicitly mentioned, it was derived from in-plane resolution and slice thickness, or field-of-view and matrix size); (iv) *MRI evaluation*: visual only or combined with image postprocessing (voxel- or surface-based morphometry, texture or morphometric analysis, volumetry, relaxometry); (v) *Qualification of the reader*: general radiologist, radiologist specialized in epilepsy, neurologist; (vi) *Non-MRI diagnostics*: MEG, PET, SPECT, intracranial stereo EEG (SEEG), EEG-fMRI, and spectroscopy.

Owing to the sparsity of evidence resulting from considerable variability in reporting of methods, narrative synthesis was used to semi-quantitatively summarize evidence on the above characteristics (i-vi) in all studies. A subset contributed to the meta-analysis, which quantitatively addressed differences in the post-surgical seizure outcome (Engel-I, > 1 year follow up), as well as diagnostic yield of MRI post-processing compared to standard visual assessment.

3.2.4 Data synthesis and statistical analysis

Narrative synthesis

To reduce the inherent variability of qualitative evidence, we leveraged data-driven unsupervised clustering to identify co-occurrence of diagnostic modalities (full methodology detailed in Appendix A). In each modality-driven cluster, we counted the number of studies reporting: (i) MRI acquisition parameters; (ii) Qualification of the reader; (iii) Type of MRI post-processing. Chi-squared (χ^2) statistics were used to test the relationships between modality-clusters, followed by

post-hoc Fisher's exact test reported as Odds Ratio (*OR*). Two-tailed statistics were thresholded at $p < 0.05$ and corrected for multiple comparisons using the Bonferroni-Holms procedure.

Meta-analysis

To evaluate effects of MRI diagnostic status on post-surgical seizure outcome, we included studies in which this information was independently reported for MRI-negative and MRI-positive cohorts (excluding those in which this distinction was absent). To compute the diagnostic yield of MRI quantitation, for each study the number of patients reported as MRI-negative before and following the diagnostic procedure (qualitative MRI review *vs.* MRI post-processing) were counted.

We used *Meta* [217] and *Metafor* [218] toolboxes implemented in *R* (*version* 4.1.2; R-project.org), employing the inverse variance method for single proportions (post-surgical seizure outcome) and Mantel-Haenszel method for binary outcomes (reversal of diagnosis). Outcomes were expressed as transformed proportions and OR with 95% confidence intervals (CI), with $p < 0.05$ as significant. Outlier case diagnostics removed studies in which the CI did not overlap with that of the estimated pooled effect [219], *i.e.,* any individual outlier study differing significantly from the overall "population" effect. Random effects models accounted for potential between-study heterogeneity (I^2), which quantifies the percentage of variation across studies due to the observed effects (or effect sizes) as compared to random error (chance) alone. I^2 was categorized into low (≥25%), moderate (≥50%) and high (≥75%). As studies with high effect size are more likely to be published (since small sample sizes require larger effects to reach statistical significance), we evaluated publication bias [220] using the *Eggers'* test (significance threshold set at $p < 0.10$).

We assessed the quality of evidence according to the Grades of Recommendations, Assessment, Development and Evaluation (GRADE) scale [221, 222], which segregates the quality of evidence across the different outcomes.

3.2.5 Standard protocol approvals, registrations, and patient consents

Individual studies were approved by local ethics committees. No institutional approval or consent was sought since the study collected and synthesized nonidentifiable data from published literature.

3.2.6 Data availability

Data are available to qualified investigators on a reasonable request to the corresponding author.

3.3 Results

3.3.1 Study selection – Screening and eligibility

Figure 3.1 summarizes the PRISMA flowchart. We screened 2,530 articles identified by the tripartite database search based on titles and abstracts. Of these, we excluded 2,149 articles as follows: reviews/opinion papers and case reports (n = 1147), studies on genetic generalized epilepsies (n = 473), conference abstracts (n = 458), and studies on encephalopathies (n = 71). Full text of the remaining 381 with an additional 46 studies identified from the reference lists were assessed for eligibility. Following the exclusion of 231 articles [new-onset and drug-responsive epilepsies (n = 89), those including only MRI-positive patients (n = 64), sample size less than 5 (n = 36), drug trials (n = 25), epidemiological studies (n = 17)], the final eligible dataset consisted of 196 articles, totaling 12,021 patients.

3.3.2 Systematic review – Narrative synthesis

The systematic review included 196 studies conducted between January 1990 and March 2021; it comprised 91 studies assessing MRI-negative patients only and 105 analysing both MRI-negative and MRI-positive. Overall, MRI as the sole diagnostic modality was used in 30% (59/196) of studies and in combination with non-MRI modalities (PET and/or SPECT) in 46% (90/196). In 12% (23/196) of studies, only non-MRI modalities were used (5 PET, 10 SPECT, 5 PET+SPECT, 2 MEG, 1 spectroscopy); in 24% (47/196) no imaging modalities were reported. In relation to the MRI field strength, 40% (79/196) of studies used a 1.5T scanner, 20% (39/196) a 3T system; combinations of 1.5T and 3T were used in 13% (26/196), 1.5T, 3T and 7T in 1.5% (3/196), and 3T and 7T in 0.5% (1/196). Notably, 48 studies (24%) did not report the field strength.

To gather the clinical and demographic information (Figure 3.2), we merged cohorts across the 196 studies for a total of 7,436 MRI-negative (median age 30, interquartile range [IQR] 27–36) and 4,585 MRI-positive patients (median age 31, IQR 27–36). Compared to MRI-positive, MRI-negative patients underwent more often SEEG (76% *vs.* 54%; $p < 0.05$ uncorrected), were less frequently operated (74% *vs.* 86%) and had a less favourable seizure outcome (62% *vs.* 74%; $p_{Bonferroni} < 0.05$). More lesions were reported on histology in MRI-positive than MRI-negative (87% *vs.* 66%, $p_{Bonferroni} < 0.05$). Stratifying post-surgical seizure outcome based on histology, FCD type II was found in almost equal number of patients in both cohorts (422 *vs.* 374); however, a favorable seizure outcome was less frequent in cases reported as MRI-negative (67% *vs.* 75%, $p_{Bonferroni} < 0.1$). Conversely, hippocampal sclerosis was almost twice as frequent in MRI-positive than MRI-negative (542 *vs.* 277), with higher proportion of seizure freedom (90% *vs.* 75%, $p_{Bonferroni} < 0.05$). Gliosis was two times more frequent in MRI-negative than MRI-positive (274

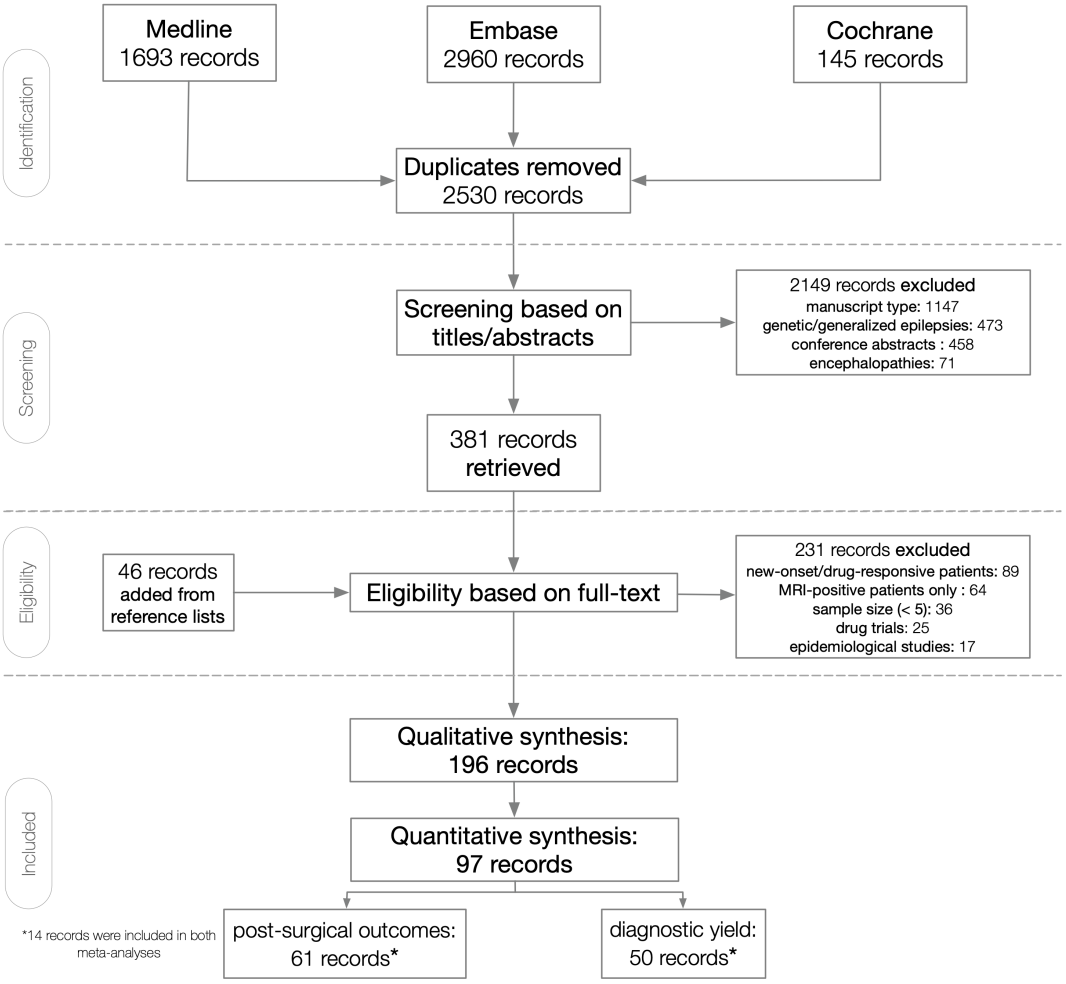


Figure 3.1: PRISMA flow chart of the study screening and inclusion process.

vs. 131) with equally unfavorable seizure outcome; notably, the majority of MRI-positive with gliosis were TLE patients.

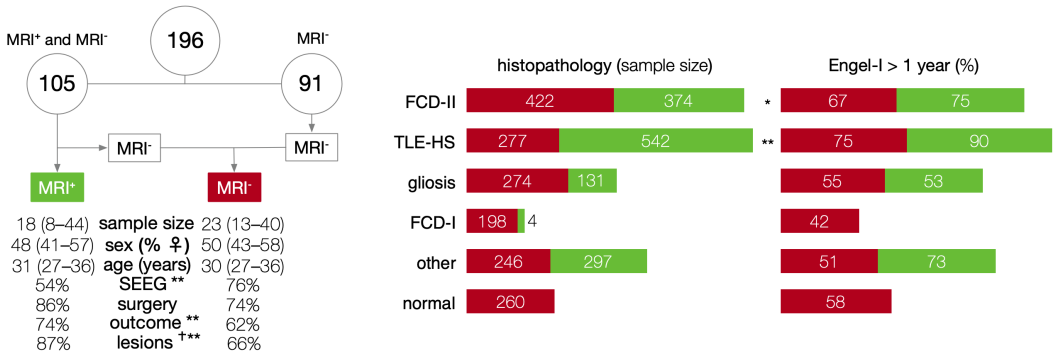


Figure 3.2: Clinical and demographic information across 196 records. Results are reported separately for MRI-positive (MRI^+; green) and MRI-negative (MRI^-; red) patients. Sample size, gender and age are reported as median (inter-quartile range), prevalence of intracranial EEG (SEEG), surgery, and outcome are expressed as mean. Plots in the right panel report seizure outcome stratified by histopathology. Other refers to benign tumors (DNET and hamartoma), and cortical malformations other than FCD I/II. **/*: significant difference ($p < 0.05$) / trend ($p < 0.1$). †: Lesions are represented as a percentage of patients that underwent surgery.

3.3.3 Systematic review – Cluster analysis

Based on the co-occurrence of MRI and non-MRI modalities, unsupervised clustering identified three groups of studies with distinct diagnostic profiles (Figure A.1A): i) *MRI-dominant*, including 61 articles using conventional MRI contrasts; ii) *nuclear-imaging*, comprising 85 articles using nuclear imaging (in combination with MRI); *limited-MRI-information*, including 50 studies lacking information on imaging modalities. Figure A.1B illustrates the cluster composition of MRI and non-MRI modalities.

Group differences between clusters are reported in Figure 3.3. There was a significant association between groups and reporting of MRI parameters ($\chi^2 = 42.67$, $p_{Bonferroni} < 6{\times}10^{-10}$). Of the 77 studies that reported details pertinent to T1-weighted MRI (in-plane voxel resolution and/or slice thickness), 65% were in the *MRI-dominant* compared to 35% for the *nuclear-imaging* group. Specifically, *MRI-dominant* studies were 10 times more likely to report T1-weighted parameters than *nuclear-imaging* ($OR = 9.76$, $p_{Bonferroni} < 5{\times}10^{-9}$), whereas such information was completely missing in the *limited-MRI-information* group. In addition, of the 43 studies reporting T2-weighted parameters, 70% were in the *MRI-dominant* compared to only 30% in the *nuclear-imaging* group ($OR = 5.36$, $p_{Bonferroni} < 10^{-5}$) and none in the *limited-MRI-information* group.

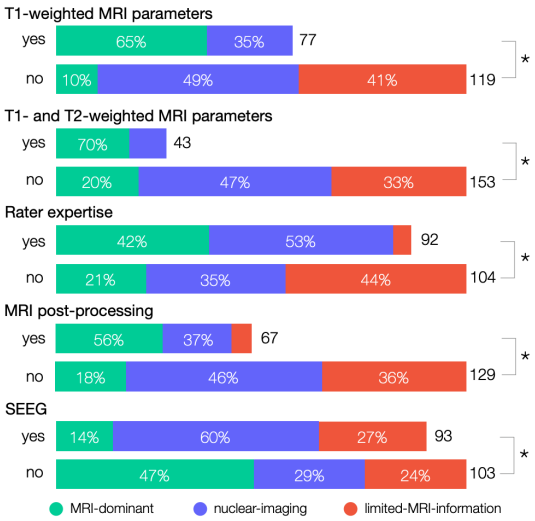


Figure 3.3: Incidence of reporting MRI parameters, rater expertise, MRI post-processing and SEEG among the 196 studies stratified by unsupervised clustering into three groups: *MRI-dominant* (dark cyan), *nuclear-imaging* (cyan), and *limited-MRI-information* (light cyan). The total number of studies across clusters are beside each horizontal bar. *: significant group differences ($p < 10^{-6}$).

There was a significant association between groups and reporting of rater expertise ($\chi^2 = 41.42$, $p_{Bonferroni} < 10^{-9}$) with the *MRI-dominant* and *nuclear-imaging* more likely (20 and 16 times, respectively) to report it than the *limited-MRI-information* group ($ORs > 15.65$, $p_{Bonferroni} < 10^{-8}$). In addition, there was a significant association between groups and incidence of MRI post-processing ($\chi^2 = 37.52$, $p_{Bonferroni} < 7{\times}10^{-9}$), with *MRI-dominant* group 4 times more likely to perform such analysis than the *nuclear-imaging* group ($OR = 3.96$, $p_{Bonferroni} < 0.003$).

Compared to the *limited-MRI-information*, *MRI-dominant* and *nuclear-imaging* groups were more likely to perform MRI post-processing (19 and 5 times, respectively; $ORs > 4.8$, $p_{Bonferroni} < 0.03$). Conversely, the *nuclear-imaging* and *limited-MRI-information* groups were more likely to perform SEEG than *MRI-dominant* group (7 and 4 times, respectively; $ORs > 3.7$, $p_{Bonferroni} < 0.02$).

Finally, evaluating the relative incidence of the various definitions of MRI-negative status, we observed a significant association between groups and definitions (*i.e.,* cryptogenic, non-lesional, MRI-negative, or normal MRI; $\chi^2 = 16.74$, $p_{Bonferroni} < 0.01$); specifically, compared to the *limited-MRI-information* group, the *MRI-dominant* group was 8 times more likely to use the term "MRI-negative" than "non-lesional" ($OR = 7.50$, $p < 0.05$; Figure A.2).

3.3.4 Meta-analysis

The effect of MRI diagnostic status on post-surgical outcome (Figure 3.4) was assessed in 61 studies (13 contributed both MRI-negative and MRI-positive cohorts) for a total of 4,335 patients; outlier case diagnostics led to exclusion of 4 studies, resulting in a final tally of 3,008 patients. Engel-I outcome was less frequently reported in MRI-negative than MRI-positive patients (57 studies, 59% [95% CI 55–63%] *vs.* 13 studies, 75% [95% CI 67–84%]; $\chi^2 = 13.94$, $p < 0.01$). Statistical between-study heterogeneity was moderate for MRI-negative ($\chi^2 = 56\%$; $p < 0.01$) and substantial for MRI-positive ($\chi^2 = 81\%$; $p < 0.01$). Conversely, visual inspection of funnel plots for random-effects models (Figure A.3, *left*) revealed an asymmetry in distribution, confirmed by the *Eggers'* test ($t = -3.96$, $p < 0.05$), indicative of a publication bias [223]; such a bias suggests that statistically significant results are more likely to be published, while studies with large effect size and small samples are not adequately represented.

The effect of the MRI analyses on diagnostic yield (Figure 3.5) was assessed in 50 studies including 2,123 patients; outlier case diagnostics led to exclusion of 10 studies, bringing the final tally to 1,242 patients. Relative to qualitative MRI review, MRI post-processing was associated with improved diagnostic yield (16 studies, 563 patients, *OR* 5.81 [95% CI 3.50–9.83] *vs.* 24 studies, 679 patients; *OR* 11.41 [95% CI 7.30–17.81]; $\chi^2 = 4.21$, $p = 0.04$). Statistical heterogeneity was low for both MRI post-processing and qualitative MRI studies ($\chi^2 = 43\%$, $p = 0.01$ *vs.* $\chi^2 = 36\%$; $p = 0.07$) indicating consistency among studies, *i.e.,* the individual effect size in either subgroup was relatively close to their subgroup mean effect size. There was no evidence of publication bias, which was confirmed by the symmetric appearance of the Funnel plot for random-effects model (Figure A.3, *right*), and corroborated by the *Eggers'* test ($t = 1.26$, $p > 0.05$).

The GRADE recommendation yielded an overall score of 1.0 and 2.5 for evidence pertaining to post-surgical outcome and diagnostic yield of MRI post-processing, reflecting low to moderate and moderate quality of evidence, respectively (Tables A.1 and A.2). GRADE profiles (Tables A.3 and A.4) include an explicit judgment of the evidence for each of the outcomes (MRI-negative and MRI-positive; qualitative MRI and MRI post-processing).

3.4 Discussion

The role of MRI in the diagnosis and treatment of epilepsy is undisputed, particularly when surgery is being considered. However, despite repeated recommendations and guidelines, adoption of best practice MRI is still variable, with many advances not fully transferred into clinical care [167]. The most dramatic implication of this translational gap relates to lesion identification, with the risk that MRI-positive patients may be wrongly labelled as MRI-negative. Paradoxically, while appending MRI-negative status has profound implications in terms of treatment strategies

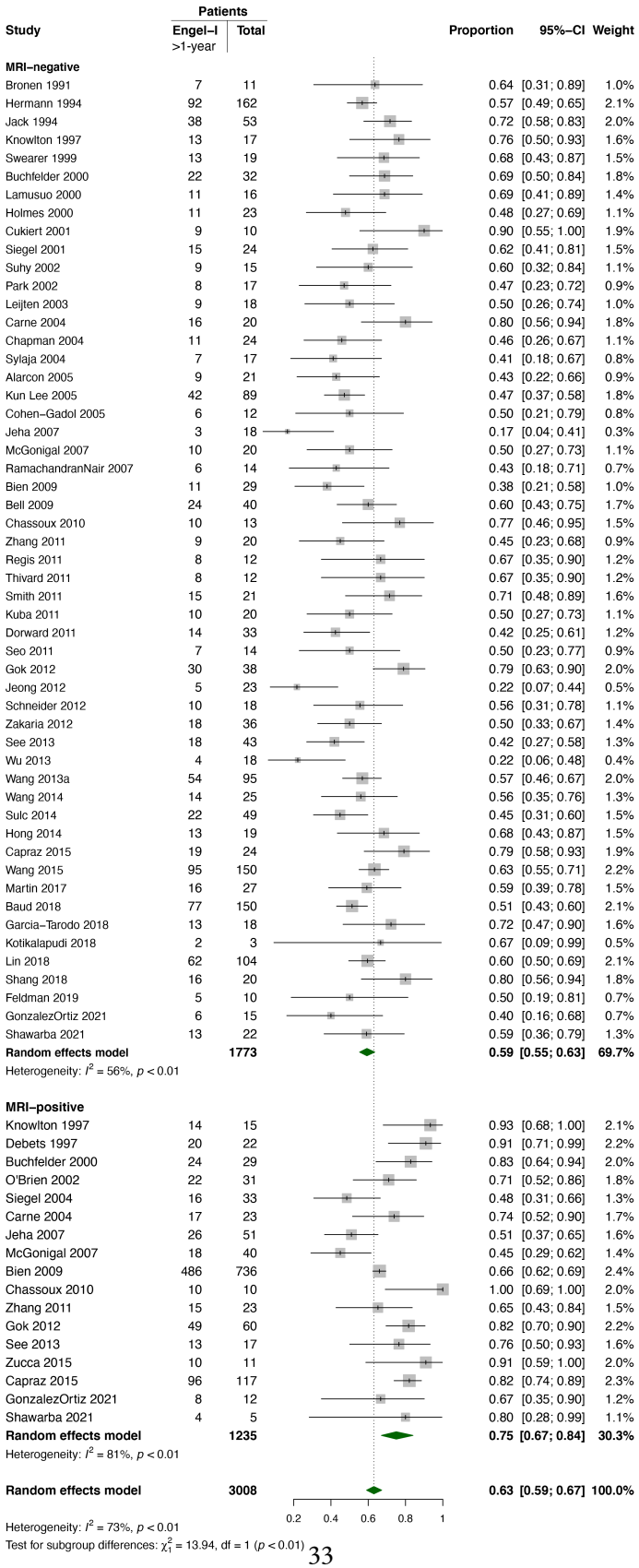


Figure 3.4: **Meta-analysis on post-surgical outcomes**. Forest plots depict the effect estimates for the association between post-surgical seizure freedom [Engel-I outcome (> 1-year follow-up)] and MRI-diagnostic status [MRI-neg (n = 1773) or MRI-pos (n = 1235)].

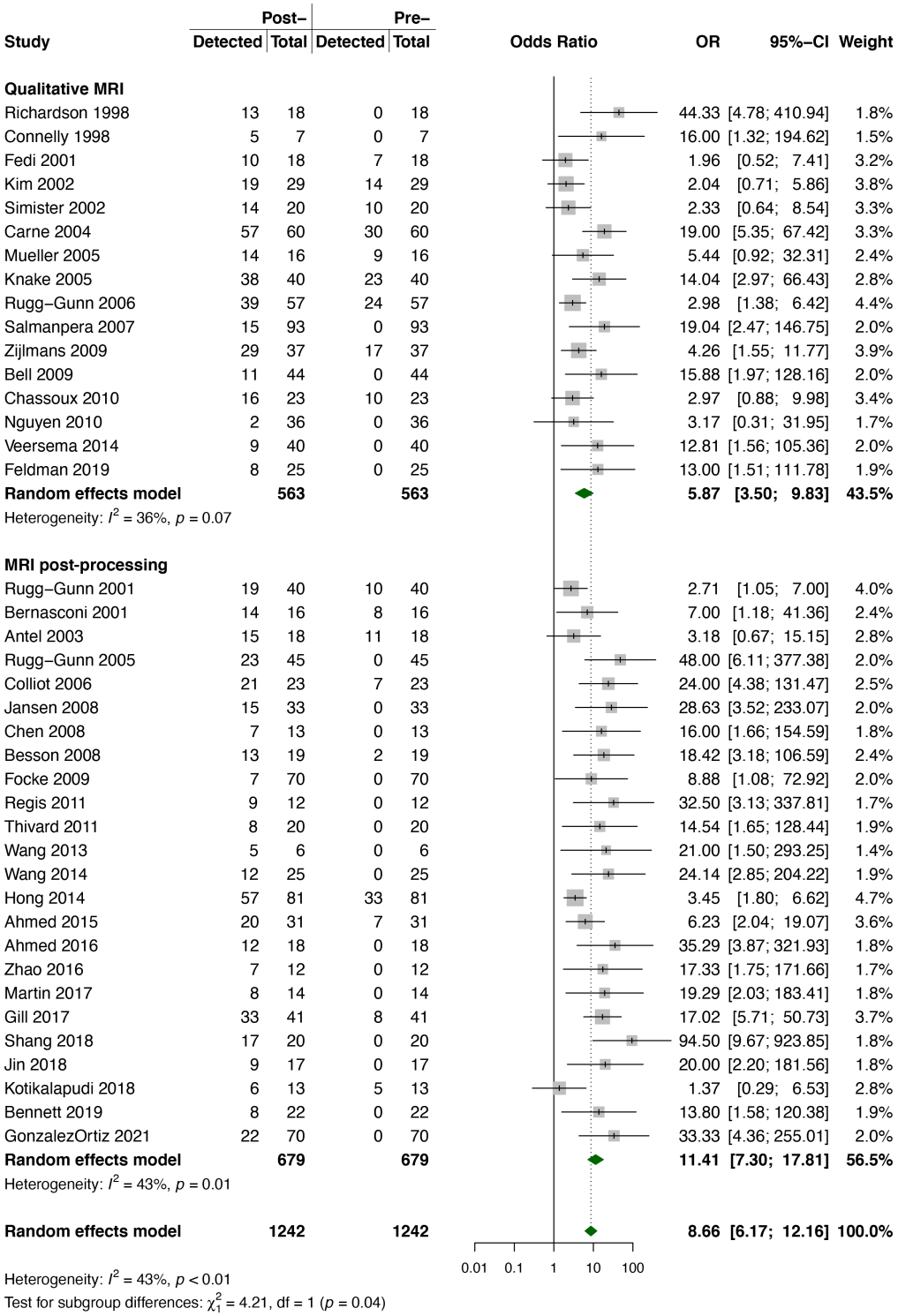


Figure 3.5: **Meta-analysis on diagnostic yield**. Forest plots depict the summary effect estimates for the association between diagnostic yield and MRI analysis procedures [post-processing (n = 679) *vs.* qualitative (n = 563)].

and seizure outcome, our systematic review showed that, in many studies published over the last three decades, this categorization has consistently lacked objectivity. Specifically, our quantitative summary estimates of almost 200 studies identified substantial variability in practices to designate the MRI status. Notably, more than a third did not report any information on the MRI procedure itself or relied on non-MRI modalities, such as PET and SPECT, to assign the MRI-negative status. As a likely consequence, MRI-negative patients underwent more often SEEG and had lower number of surgeries, with less favourable seizure outcome compared to MRI-positive patients. The results of the narrative synthesis based on 12,021 patients are in agreement with similar observations on smaller cohorts [22, 224–227]. Importantly, however, the number of lesions ultimately found on histology was similar among MRI-positive and MRI-negative patients, validating quantitatively the clinical impression that many patients labelled as MRI-negative present indeed with a structural pathology that remains overlooked on imaging. The second implication is that for the same histological substrate, when surgery is not guided by MRI findings, outcome is poorer, likely due to incomplete resection of the epileptogenic lesion.

To gain further insights into factors underpinning the heterogeneity in the definition of the MRI-negative status, we leveraged unsupervised machine learning and stratified studies based on diagnostic modalities. The algorithm identified three distinct classes comprising similar number of studies. The *MRI-dominant* class exemplified methodological rigour, detailing information on contrasts and acquisition parameters, the expertise of the reviewer, and the use of post-processing more often relative to the *nuclear-imaging* class. Conversely, the *limited-MRI-information* class lacked any criteria to ascribe MRI-diagnostic status. Notably, SEEG was more often performed in studies in *nuclear-imaging* and *limited-MRI-information*. Thus, this data-driven stratification further demonstrated the methodological variability and weaknesses in the criteria used for MRI diagnostics.

The meta-analysis on post-surgical outcome confirmed the results of the narrative synthesis, namely that seizure freedom after surgery is more often achieved in MRI-positive compared to MRI-negative patients with proportions of 75% and 59%, respectively. Admittedly, while there was no overlap of confidence intervals among the two groups, focussing our analysis primarily on MRI-negative manifested with a wider interval due to the smaller MRI-positive cohort. Moreover, while studies within the MRI-positive subgroup displayed high inconsistency (due to between-study heterogeneity) and publication bias (evident from the funnel plots), there were no concerns regarding imprecision (due to the inclusion of relatively few patients) or indirectness (outcomes measured using proxies). Evidently, a recent systematic review and meta-analysis focusing solely on MRI-positive FCD patients [228] revealed favourable post-surgical outcome in 70% of patients, a proportion close to the one we found. Similarly, our data on post-surgical outcomes recapitulate the results of a past metanalysis reporting 70% of MRI-positive and 46%

of MRI-negative patients achieving seizure freedom [4]. Hence, our review provides best current evidence for an association between MRI-negative status and unfavorable post-surgical outcome. Notably, the virtually stable proportion of MRI-negative with unfavorable outcome despite constant improvements in MRI technology and analytical tools over a period of more than 10 years demonstrates a translational gap and should be taken as an incentive to increase educational efforts and a cultural change in the epilepsy community. This argument is further supported by the meta-analysis on diagnostic yield showing that MRI post-processing is twice as likely to reverse the MRI-negative status relative to the visual analysis. Specifically, we observed a four-fold improvement in diagnostic yield (from 13% to 52%) using MRI post-processing, relative to a two-fold improvement (from 26% to 53%) when relying on a qualitative image review. While the individual studies across subgroups had minor weaknesses regarding publication bias, namely the absence of small studies with small effect sizes, there were no concerns regarding inconsistency, imprecision, or indirectness. Hence, to the best for our knowledge, this review provides the first meta-analytic evidence for an association between MRI post-processing and gain in diagnostic yield, largely driven by the ability of MRI post-processing to unveil subtle or subthreshold anomalies inconspicuous to the unaided human eye.

In summary, our systematic review and meta-analyses support the need for an objective definition for ascribing MRI-negative status. Beside the value of detailing field strength, sequences and parameters and the expertise of the reader, our findings suggest that MRI post-processing should be systematically performed when assessing patients with medically intractable seizures, particularly those with suspected MRI-negative epilepsy in search of prevalent epileptogenic lesions, specifically hippocampal sclerosis or focal cortical dysplasia [23].

4 Multimodal Surface-based Automated Detection of Focal Cortical Dysplasia

Preface

While previous voxel-based computational approaches improved the visibility of the lesion, they have primarily focused on large- or medium-size visible FCD. When obvious features of gray matter thickening and blurring are absent, abnormal sulco-gyral patterns may be an alternative marker of cortical dysgenesis. Indeed, the resection of subtle cortical gyral abnormalities in patients with MRI reported as normal may result in favorable surgical outcome [22]. Sulco-gyral anomalies are, however, are difficult to discriminate visually when images are inspected on orthogonal planes in the volumetric domain. The intrinsic limitations of voxel-based approaches (*e.g.,* suboptimal image registration and spatial smoothing that disregards the true cortical geometry) has motivated the shift to surface-based morphometric paradigms that permit anatomically plausible modeling of subtle cortical dysplasia while respecting the highly convoluted geometry of the neocortex. In addition, FCD characterization is best achieved when using multiple MRI contrasts that probe different facets of pathology. However, previous studies have exploited single contrasts alone, or analyzed multiple contrasts in isolation within univariate frameworks.

The purpose of this study was to implement a multivariate surface-based detection framework to detect subtle FCD using T1-weighted gradient echo, T2-weighted FLAIR and FLAIR/T1w ratio images using statistical machine-learning techniques, and to evaluate its diagnostic yield relative to conventional visual inspection.

Manuscript 2

Published as R. S. Gill, S.-J. Hong, F. Fadaie, B. Caldairou, B. Bernhardt, N. Bernasconi, and A. Bernasconi. "Automated Detection of Epileptogenic Cortical Malformations Using Multimodal MRI". in: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Ed. by M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. S. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, and Z. Lu. Springer International Publishing, Cham, 2017, pp. 349–356. ISBN: 978-3-319-67558-9. DOI: 10.1007/978-3-319-67558-9_40

Detection of Epileptogenic Cortical Malformations using Multimodal MRI

Ravnoor S. Gill[1], Seok-Jun Hong[1], Fatemeh Fadaie[1], Benoit Caldairou[1], Boris C. Bernhardt[1], Neda Bernasconi[1], Andrea Bernasconi[1]

[1] Neuroimaging of Epilepsy Laboratory, McConnell Brain Imaging Center, Montreal Neurological Institute, Montreal, QC, Canada

Abstract

Focal cortical dysplasia (FCD), a malformation of cortical development, is a frequent cause of drug-resistant epilepsy. This surgically-amenable lesion is histologically characterized by cortical dyslamination, dysmorphic neurons, and balloon cells, which may extend into the immediate subcortical white matter. On MRI, FCD is typically associated with cortical thickening, blurring of the cortical boundary, and intensity anomalies. Notably, even histologically-verified FCD may not be clearly visible on preoperative MRI. We propose a novel FCD detection algorithm, which aggregates surface-based descriptors of morphology and intensity derived from T1-weighted (T1w) MRI, T2-weighted fluid attenuation inversion recovery (FLAIR) MRI, and FLAIR/T1w ratio images. Features were systematically sampled at multiple intracortical/subcortical levels and fed into a two-stage classifier for automated lesion detection based on ensemble learning. Using 5-fold cross-validation, we evaluated the approach in 41 patients with histologically-verified FCD and 38 age-and sex-matched healthy controls. Our approach showed excellent sensitivity (83%, 34/41 lesions detected) and specificity (92%, no findings in 35/38 controls), suggesting benefits for presurgical diagnostics.

4.1 Introduction

Focal cortical dysplasia (FCD), a malformation of cortical development, is a prevalent cause of drug-resistant epilepsy. Its surgical removal is currently the only treatment option to arrest seizures. Cardinal histopathological features of FCD include cortical dyslamination associated with various intra-cortical cytological anomalies, namely dysmorphic neurons (FCD Type-IIA) and balloon cells (FCD Type-IIB) [19].

On MRI, FCD is typically associated with varying degrees of cortical thickening, blurring of the interface between the grey and white matter and anomalous intensity profiles. Notably, even histologically-verified FCD may not be clearly visible on MRI [20]. Previous studies using voxel- [101, 229] or surface-based methods [130, 230, 231] modeled limited numbers of features derived from T1 or T2-weighted MRI, except for one [124] combining them in a small pediatric cohort. All these methods, however, were mainly applied to large- to medium-sized lesions visible on MRI and provided limited sensitivity [31]. Moreover, histological validation was present in 30% of cases only.

The current work proposes a novel in vivo surface-based automated detection algorithm modeling FCD at various depths within the cortex and the subcortical white matter. Our method exploits the diagnostic power of T1-weighted gradient echo and T2-weighted FLAIR contrasts together with a synthetic FLAIR/T1w ratio map; the latter was specifically designed to increase the sensitivity for co-occuring FLAIR hyperintensity and T1w hypointensity present at the interface between the grey and white matter. Our framework was validated in a cohort of 41 patients with histopathologically proven FCD.

4.2 Methods

4.2.1 MRI acquisition

Multimodal MRI was acquired on a 3T Siemens TimTrio using a 32-channel head coil, including a 3D T1-weighted MPRAGE (T1w; TR = 2300 ms, TE = 2.98 ms, flip angle = 9°, FOV = 256 mm^2, voxel size = 1×1×1 mm^3) and fluid-attenuated inversion recovery (FLAIR; TR = 5000 ms, TE = 389 ms, TI = 1800 ms, flip angle = 120°, FOV = 230 mm^2, voxel size = 0.9×0.9×0.9 mm^3).

T1w MRI underwent intensity inhomogeneity correction [115] followed by intensity standardization, linear registration to MNI152 space, and classification into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) [118]. GM-WM and GM-CSF surface models were reconstructed using CLASP, an algorithm relying on intensity and geometric constraints [114]. Surface-based registration that aligned individual subjects based on cortical folding was used to

increase across-subjects correspondence in measurement locations [120]. T1w images were linearly co-registered to FLAIR. After similar pre-processing, FLAIR images were divided by T1w images to generate a FLAIR/T1w ratio map; this ratio allows for additional correction of B1 intensity non-uniformity after N3 correction. Hyperintensities exceeding 1 SD from the mean ratio within the brain mask were excluded, generating the ratio image.

4.2.2 Multi-surface generation

To examine intracortical GM, we positioned 3 surfaces between the inner and outer cortical surfaces at 25%, 50%, and 75% cortical thickness, guided by a straight line providing vertex-correspondence across surfaces [232]. Although these surfaces do not necessarily reflect cortical laminae, they capture relative differences along the axis perpendicular to the cortical mantle. To assess the WM immediately beneath the cortex, we generated 3 equidistant surfaces guided by a Laplacian field running between the GM-WM interface and the ventricles, with between-surface intervals adapted to the resolution of each modality.

4.2.3 Feature extraction

The following features were extracted in the native space of a given contrast to minimize interpolation.

a) *Intensity-based features*. For each modality, *i.e.,* T1w, FLAIR, and FLAIR/T1w, we divided voxel-wise intensities by the mean GM-WM boundary intensity; this value was normalized with respect to the mode of the respective intensity histogram [233]. Normalized intensities were linearly interpolated at each surface-point (or vertex) of intra- and subcortical surface models. We did not sample intensity on the GM-CSF surface to avoid CSF contamination; values for all other surfaces were analytically corrected for partial volume effects [234]. We also computed intensity gradients in perpendicular and tangential direction relative to cortical surfaces, to model radial and tangential dyslamination [130].

b) *Morphology*. Cortical thickness was calculated as the Euclidean distance between vertices on the GM-WM and GM-CSF surfaces [114]. Small FCD lesions often occur at the bottom of a sulcus and display curvature changes [230]. We thus computed sulcal depth for each vertex as the shortest geodesic distance from a gyral crown and measured absolute mean curvature along the 50% intracortical surface [235].

c) *Feature profiling*. We assigned at each vertex a unique vector of intra- and sub-cortical intensity and morphological features smoothed using a surface-based 5 mm full-width-at-half-maximum Gaussian kernel and z-normalized with respect to the distribution in healthy controls. For each intensity feature, we calculated an average across the 4 intracortical (25%, 50%, 75%, GM-WM in-

terface) and 3 subcortical surfaces. For each individual, we thus obtained 3 morphological maps (cortical thickness, sulcal depth, curvature), 6 intensity maps (T1w, FLAIR, FLAIR/T1w at intracortical and subcortical levels), 6 corresponding gradient maps (3 tangential, 3 perpendicular) together with their asymmetries yielding a total of 30 features.

4.3 Experiments

4.3.1 Subjects

Our patient cohort consisted of 41 patients (20 males; mean±SD age = 27 ± 9 years) admitted to the MNI for the investigation of drug-resistant focal epilepsy. The presurgical workup included neurologic examination, assessment of seizure history, MRI, and video-EEG telemetry. In 33 (80%) patients, lesions were initially not seen on conventional radiological inspection of pre-operative MRI. Since the MRI was initially reported as unremarkable, the location of the seizure focus was established using intracerebral EEG; retrospective inspection revealed a subtle FCD in the seizure onset region in all. All patients underwent surgery and the diagnosis FCD was histopathologically verified.

The control group consisted of 38 age- and sex- matched healthy individuals (19 males; mean±SD: age = 30 ± 7 years).

4.3.2 Manual lesion segmentation

Two experts, blinded to clinical information, independently segmented FCD lesions on co-registered T1w and FLAIR MRI. Inter-rater Dice agreement index (D = 2|M1∩M2| / [|M1|+|M2|]; M1, M2: 1st, 2nd label; M1∩M2: intersection of M1 and M2) was 0.91 ± 0.11. Their consensus volume label was intersected with the surface models, thereby generating a surface-based lesion label.

4.3.3 Classification paradigm

We built a two-stage classifier (Figure 4.1). A first vertex-wise classification was designed to maximize sensitivity (*i.e.,* detecting a maximum number of lesional clusters), whereas a subsequent cluster-wise classification aimed at improving specificity (*i.e.,* removing false positives while maintaining optimal sensitivity).

We used an ensemble of RUSBoosted decision trees to systematically test detection performance across both classification stages. RUSBoost [236] is a hybrid sampling/boosting algorithm, which can learn while mitigating the class imbalance problem that occurs from the presence of

imbalance between high number of non-lesional vertices in a given subject compared to lesional vertices.

Vertex-level classification. RUSBoost randomly undersampled from the pool of non-lesional vertices until a balanced distribution was achieved. AdaBoost [237], a common boosting algorithm, then iteratively built an ensemble of base learners (decision trees). During each iteration, higher penalty weights were assigned to misclassified vertices to improve classification accuracy for the next round. After the final iteration, all trained base learner models participate in a weighted vote to classify test vertices as lesional or non-lesional. On the resulting predictions, a cluster was defined as a collection of vertices that form 6-connected neighbors on the discrete cortical surface mesh. Lesional (true positives) and non-lesional (false positives) clusters were fed into subsequent cluster-level classification.

Cluster-level classification. For each cluster, we assessed the overall load of anomalies by computing the Mahalanobis distance (a multivariate z-transform between each patient's feature vector and the corresponding distribution in controls). For the set of vertices displaying the highest distance, we used the 15 original features (excluding their asymmetries) to compute the following higher-order features: statistical moments (mean, standard deviation, skewness, kurtosis, moment, and their asymmetries) representing the shape of the distribution of each feature, spatial location (as determined by anatomical parcellation and 3D coordinates, and lesion size. This process generated a total of 95 features per subject. Classification then proceeded using these cluster features in the same manner as vertex-wise classification.

Feature selection. In both classification stages, we used an ensemble of extremely randomized trees in conjunction with RUSBoost to select features. This procedure introduced more randomness during feature selection and training, which has been shown to improve the bias/variance tradeoff and performance compared to conventional tree-based classifiers [238].

Partitioning of training and test datasets. At each stage, classifiers were trained using 5-fold cross validation with 10 iterations: 4 folds of data were used for model training (200 RUSBoosted decision trees), and the remaining for testing (*i.e.,* lesion detection). For feature selection, nested 5-fold cross-validation was implemented. Optimal features were determined across 10 iterations and finally averaged. Selected features were used for classifier training and testing. This procedure permits a conservative assessment of performance and generalizability for previously unseen FCD cases.

Evaluation of classification accuracy. Performance was assessed relative to manual labels. Sensitivity was the proportion of patients in whom a detected cluster co-localized with the lesion label. Specificity was determined with respect to controls (*i.e.,* proportion of controls in whom no FCD

lesion cluster was falsely identified). We also report the number of clusters detected in patients remote from the lesion label (*i.e.,* false positives).

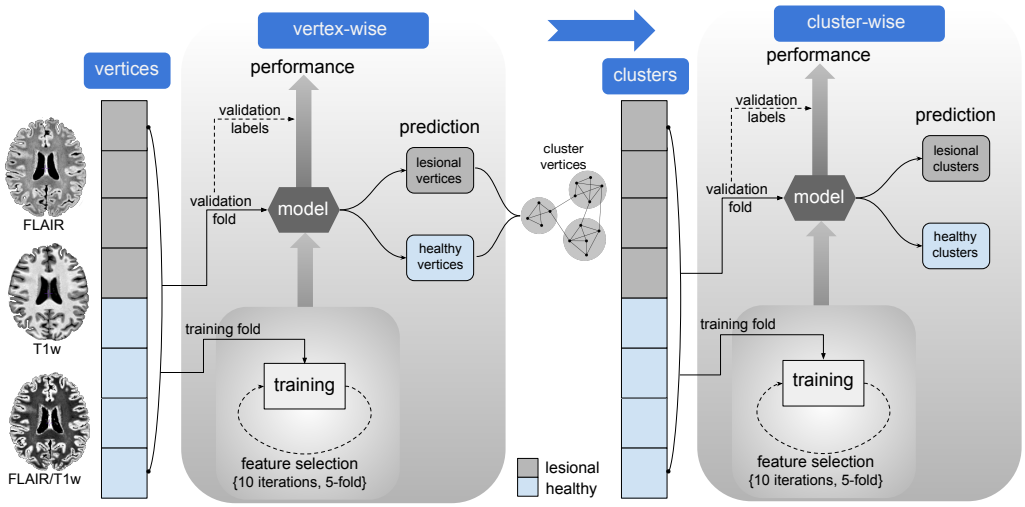


Figure 4.1: Vertex-wise and cluster-wise classification schema. See text for details.

4.4 Results

At the first vertex-wise stage, the classifier detected all but 4 lesions (37/41 = 90% sensitivity). However, it also detected false positives (mean ± SD clusters: 25 ± 23 in patients; 7 ± 5 in controls). Subsequent cluster-wise classification guaranteed a sensitivity of 83% (34/41 clusters co-localized with the label), while it dramatically reduced the number of false positives (mean±SD clusters = 4 ± 5). We also obtained a high specificity of 92% with only a single cluster in 3 healthy subjects. The results for both stages are summarized in the Table 4.1. An example is shown in Figure 4.2.

The highest normalized weights (*i.e.,* more than 10% of total weighting) were as follows: Vertex-wise classification was driven by perpendicular gradient in 69% of cases (derived from: FLAIR in 25%, T1w in 24%, FLAIR/T1w in 20%), followed by subcortical normalized intensity (derived from T1w in 31%; FLAIR in 16%, FLAIR/T1w in 15%). Cluster-wise classification was largely driven by cortical thickness (34%), followed by FLAIR subcortical normalized intensity (33%), and lesion size (33%).

4.5 Discussion

The current study presents a novel machine intelligence system for the automated detection of FCD lesions. Our approach was developed and evaluated in a consecutive cohort of patients with

Table 4.1: Summary of classification results

Classifier	Sensitivity	FPs in patients (mean±SD)	FPs in controls (mean±SD)	Specificity (w.r.t Controls)
Vertex-wise	90% (37/41)	25 ± 23	7 ± 5	N/A
Cluster-wise	83% (34/41)	4 ± 5	0.08 ± 0.27	92% (35/38)

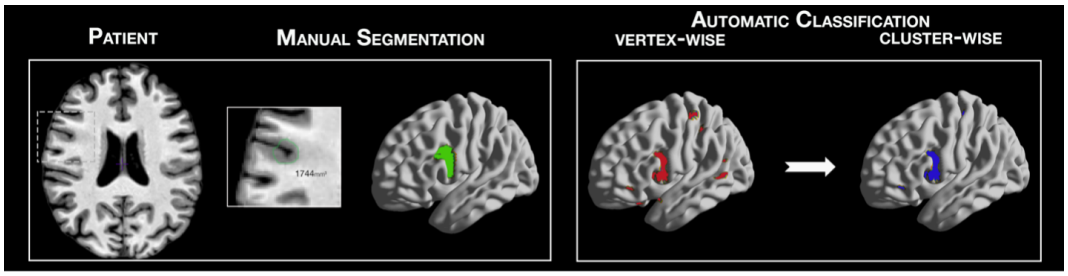


Figure 4.2: *Left*: Axial T1w MRI showing the cortex harboring the FCD (dashed square), a close-up of the outline of the lesion label (solid green line), and its surface projection. *Right*: vertex-wise and subsequent cluster-wise automated classification results.

histopathologically-validated lesions, with the majority of patients having lesions that were initially overlooked on standard radiological exam. Core to our approach was a surface-based integration of morphological markers as well as intensity and textural features derived from co-registered T1w and FLAIR data as well as their ratio. The latter was specifically chosen to enhance co-occuring FLAIR hyperintensity and T1w hypointensity occurring at the junction between the grey and white matter, an important FCD feature, so far not addressed [239].

To classify FCD lesions in a high-dimensional dataset, we chose a non-parametric boosted decision tree ensemble that can capture complex decision boundaries while avoiding overfitting. Circumventing class imbalance is another critical issue in lesion detection. Therefore, to reduce bias against minority class (*i.e.,* lesional vertices), we implemented random-undersampling with adaptive boosting [236, 237], with robust 5-fold cross-validation; this, together with a comprehensive multi contrast modelling of FCD features, likely contributed to the highest performance to date compared to previous studies [101, 124, 130, 229–231].

Operating on two sequential levels, our approach resulted in both high sensitivity and specificity. Notably, the number of false positive findings in controls was rather modest (only 1 cluster per subject in 3/38 healthy controls) as was the number of extra-lesional clusters in patients. Notably, as our patients were seizure free after surgery, we are confident that these extra-lesional clusters were indeed false positives. Our automated algorithm provided a 4 times higher detection rate than conventional radiological visual inspection, which allowed identifying only 20% of the lesions.

In conclusion, we designed our protocol to attain high specificity, which is critical in the management of patients who undergo presurgical evaluation for medically intractable seizures. This histologically-validated new FCD detection method, providing the highest performance to date compared to the literature, has the potential to become a useful clinical tool to assist in the diagnosis of subtle lesions that are frequently overlooked by conventional means of analysis.

5 Automated Detection of Focal Cortical Dysplasia using Deep Learning

Unified Preface

Early voxel-based computational approaches aimed at improving the lesion visibility primarily focused on large- or medium-sized FCD. Surface-based morphometry methods (SBM) were devised to better visualize sulco-gyral anomalies, which are difficult to visually discriminate on orthogonal planes. Despite their high fidelity in localizing lesions, current benchmark automated detection fails in 20–40% of patients, particularly those with subtle FCD. Moreover, they suffer from high false positive rates, and incur a significantly higher technical debt in terms of feature engineering and manual quality control. Nonetheless, the intrinsic limitations of voxel-based approaches that motivated the shift to SBM paradigms can be circumvented by deep learning (DL) methods, which have progressively superseded contemporary SBM methods in performance using minimally preprocessed volumetric data. Notably, DL methods require minimal manual intervention and quality control, and obviate feature engineering via their inherent hierarchical feature learning capabilities, enabling unprecedented discovery of knowledge from large heterogenous datasets, thus setting the bases for widespread clinical use.

Effective clinical translation of computer-aided diagnostic algorithms necessitates reliable performance against variable environments. Lack of harmonization in image acquisition protocols across specialized centers often leads to over-fitting on a single dataset and low reproducibility resulting from poor generalization performance. Thus, to make novel imaging biomarkers effective and generalizable, performance on clinical data should be evaluated independently across multiple sites with diverse cohorts. In the subsequent chapters, we address generalizability and risk-stratification in a large multi-center cohort using DL applied to volumetric multimodal MRI.

This chapter is structured as a collection of three manuscripts as follows. First, we designed a CNN that leverages multimodal MRI to harness the added diagnostic value of FLAIR contrast

to detect FCD lesions in a pre-dominantly MRI-negative cohort (Chapter 4). Second, we used approximate Bayesian variational inference to sample the model's epistemic uncertainty (Chapter 5) and leveraged it to assign a lesional ranking to triage multiple putative lesional candidates (Chapter 6), while also comparing the diagnostic yield of this Bayesian DL classifier to one that does not incorporate uncertainty in predictions. Finally, we developed and validated an automated framework to detect subtle FCD lesions based on Bayesian DL (developed in Chapters 5 and 6) and evaluated the generalization performance across nine tertiary epilepsy centers worldwide with diverse cohorts, scanners, and field-strengths (Chapter 7). The framework provided a measure of diagnostic confidence that enables risk stratification by incorporating Bayesian uncertainty estimation. Importantly, this study provided Class III evidence that DL on multimodal MRI accurately identifies FCD in epilepsy patients initially diagnosed as MRI-negative.

Manuscript 3

Published as R. S. Gill, S.-J. Hong, F. Fadaie, B. Caldairou, B. C. Bernhardt, C. Barba, A. Brandt, V. C. Coelho, L. d'Incerti, M. Lenge, 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, pp. 490–497. DOI: 10.1007/978-3-030-00931-1_56

Deep convolutional networks for automated detection of epileptogenic brain malformations

Ravnoor S. Gill [1], Seok-Jun Hong [1], Fatemeh Fadaie [1], Benoit Caldairou [1], Boris C. Bernhardt [1,2], Carmen Barba [3], Armin Brandt [4], Vanessa C. Coelho [5], Ludovico d'Incerti [6], Matteo Lenge [3], Mira Semmelroch [7], Fabrice Bartolomei [8], Fernando Cendes [5], Francesco Deleo [6], Renzo Guerrini [3], Maxime Guye [9], Graeme Jackson [7], Andreas Schulze-Bonhage [4], Tommaso Mansi [10], Neda Bernasconi [1], Andrea Bernasconi [1]

1) Neuroimaging of Epilepsy Laboratory, McConnell Brain Imaging Center, Montreal Neurological Institute, Montreal, QC, Canada; 2) Multimodal Imaging and Connectome Laboratory, McConnell Brain Imaging Center, Montreal Neurological Institute, Montreal, QC, Canada; 3) Children's Hospital A. Meyer-University of Florence, Florence, Italy; 4) Freiburg Epilepsy Center, University Medical Center, Freiburg, Germany; 5) University of Campinas, Brazil; 6) Istituto Neurologico Carlo Besta, Milano, Italy; 7) The Florey Institute of Neuroscience and Mental Health and The University of Melbourne, Victoria, Australia; 8) Aix-Marseille Univ, INS, Marseille, France; 9) Aix-Marseille Univ, CNRS, Marseille, France; 10) Siemens Medical Solutions, Medical Imaging Technologies, Princeton, NJ

Abstract

Focal cortical dysplasia (FCD) is a prevalent surgically-amenable epileptogenic malformation of cortical development. On MRI, FCD typically presents with cortical thickening, hyperintensity, and blurring of the gray-white matter interface. These changes may be visible to the naked eye, or subtle and be easily overlooked. Despite advances in MRI analytics, current surface-based algorithms fail to detect FCD in 50% of cases. Moreover, arduous data pre-processing and specialized expertise preclude widespread use. Here we propose a novel algorithm that harnesses feature-learning capability of convolutional neural networks (CNNs) with minimal data pre-processing. Our classifier, trained on a patch-based augmented dataset derived from patients with histologically-validated FCD operates directly on MRI voxels to distinguish the lesion from healthy tissue. The algorithm was trained and cross-validated on multimodal MRI data from a single site (S1) and evaluated on independent data from S1 and six other sites worldwide (S2–S7; 3 scanner manufacturers and 2 field strengths) for a total of 107 subjects. The classifier showed excellent sensitivity (S1: 87%, 35/40 lesions detected; S2–S7: 91%, 61/67 lesions detected) and specificity (S1: 95%, no findings in 36/38 healthy controls; 90%, no findings in 57/63 disease controls). Easy implementation, minimal pre-processing, high performance, and generalizability make this classifier an ideal platform for large-scale clinical use, particularly in "MRI-negative" FCD.

5.1 Introduction

Focal cortical dysplasia (FCD), a malformation of cortical development, is a frequent cause of drug-resistant epilepsy. This surgically-amenable lesion is characterized on histology by altered cortical laminar structure and cytological anomalies together with gliosis and demyelination, which may extend into the underlying white matter [16]. On MRI, FCD typically presents with cortical thickening, hyperintensity, and blurring of the gray-white matter interface. These changes may be visible to the naked eye on T1- and T2-weighted MRI, or subtle and easily overlooked [20].

Over the last decade, a number of automated algorithms have been developed [31]. Contemporary FCD detection methods rely on surface-based approaches [124, 129, 130, 240], which allow to effectively model sulco-gyral morphology. While they have shown effectiveness, they have been mainly used as a proof of principle and applied to lesions previously seen on MRI, but rarely validated histologically. Despite advances in MRI analytics, current algorithms fail to detect subtle FCD [20]. Importantly, since training and validation have been performed on data from the same center and scanner, generalizability to independent cohorts remains unclear. Finally, arduous preprocessing and specialized expertise preclude their broader integration into clinical workflows.

Conventional machine-learning systems require careful engineering and considerable domain knowledge to design features from which the classifier can learn patterns. Conversely, convolutional neural networks (CNNs), a class of deep neural networks, have the capacity to extract a hierarchy of increasingly complex features from the data [177]. In biomedical imaging, CNNs have gained popularity in brain tissue classification, and segmentation of brain tumors and multiple sclerosis plaques (see Litjens *et al.* [178] for review). To the best of our knowledge, no study has deployed CNNs to detect cortical brain malformations.

Exploiting the complementary diagnostic power of T1- and T2-weighted contrasts, we propose a novel algorithm with minimal data pre-processing and which harnesses feature-learning proficiency of CNNs to distinguish FCD from healthy tissue directly on MRI voxels. Our algorithm was trained and tested on data from a single site (S1) and tested on independent data from S1 and six sites worldwide (S2–S7), for a total of 107 individuals. Furthermore, it was tested against a benchmark surface-based algorithm, making this study the first deep-learning approach for FCD detection with multicentric validation.

5.2 Methods

5.2.1 MRI acquisition

At S1, multimodal MRI was acquired on a 3T Siemens TimTrio using a 32-channel head coil, including: 3D T1-weighted MPRAGE (T1w; TR = 2300 ms, TE = 2.98 ms, flip angle = 9°, FOV

= 256 mm^2, voxel size = 1×1×1 mm^3), and T2-weighted 3D fluid-attenuated inversion recovery (FLAIR; TR = 5000 ms, TE = 389 ms, TI = 1800 ms, flip angle = 120°, FOV = 230 mm^2, voxel size = 0.9×0.9×0.9 mm^3).

5.2.2 Image pre-processing

For all datasets, T1w and FLAIR images underwent intensity non-uniformity correction [115] and normalization. T1w images were then linearly registered (affine, 9 degrees of freedom) to the age-appropriate MNI152 symmetric template (1×1×1 mm^3) stratified across seven age-groups [0–4.5, 4.5–8.5, 7–11, 7.5–13.5, 10–14, 13–18.5, 18.5–43 years old] [241]. Age-appropriate templates minimize the interpolation effects of linear registration, thereby limiting blurring effects that may mimic lesional tissue and manifest as false positives. FLAIR images were linearly mapped to T1w images in MNI space. Skull-stripping was performed to exclude non-brain tissue.

5.2.3 Patch-based input sampling

Balanced inputs based on 3D volumetric images. Data imbalance is a challenging issue in FCD lesion detection where the number of healthy voxels significantly outweighs pathological voxels (< 1% of total voxels). To prevent biasing the classifier towards healthy voxels, we constructed a patch-based dataset by randomly under-sampling the healthy voxels such that the feature set was composed of equal number of examples from both classes. To this end, we sub-sampled multi-contrast 3D patches from the co-registered 3D T1w and FLAIR images, with each input image modality representing a channel. The data was normalized within each input modality with zero mean and unit variance. For each normalized training image, we computed 3D patches (16×16×16) centered on the voxel of interest. The set of all computed patches were aggregated as P = n×2×16×16×16, where n and 2 denote the number of training patches and input MRI modalities, respectively.

Sampling heuristics. On a per-subject level (1.7 million patches × 32 kB/patch = 26.3GB), the training is quite memory-intensive to complete within a reasonable timeframe. To circumvent this issue, we sampled only hyperintense voxels based on the FLAIR contrast by thresholding the subject-level z-normalized images and discarding the bottom 10 percentile intensities. This thresholding yielded a crude gray matter mask, which covered the hyperintense white matter as well. This approach is also biologically meaningful as FCD lesions are primarily located in the gray matter [19]; moreover, both their gray matter and white matter components are consistently hyperintense on FLAIR [242].

5.2.4 Network architecture and design

A typical convolutional neural network (CNN) consists of three stages: convolutions, nonlinearity, and pooling. Here, we designed two identical CNNs whose weights are optimized independently. This two-phase cascaded training procedure has been shown to allow efficient training in both CNNs [243, 244] and conventional machine learning [129, 130] paradigms when the distribution of labels is unbalanced. CNN1 was trained to maximize putative lesional voxels, while CNN2 reduced the number of misclassified voxels (*i.e.,* removing false positives while maintaining optimal sensitivity). Each fully convolutional network was composed of three stacks of convolution and max-pooling layers with 48, 96 and 2 filters, respectively. The rectified linear activation (ReLU) non-linearity function was applied to the first two of the three convolutional layers. Softmax non-linearity was used after the final convolution to normalize the result of the kernel convolutions into a binominal distribution over the healthy and lesional labels. See Figure 5.1 for network parameters.

5.2.5 Classification paradigm

Training algorithm. We used a validation set (75/25 training data split) to optimize the CNN weights. The training set is used to adjust the weights of the neural network, while the validation set measures the performance of trained CNN after each epoch and continues until the validation error plateaus. The model is randomly initialized, and network parameters learns iteratively via the adaptive learning rate method (*AdaDelta*) by minimizing the binary cross-entropy loss. Binary cross-entropy loss is mathematically defined as:

$$crossentropy(p, q) = -(p \cdot \log q + (1 - p) \cdot \log(1 - q)) \tag{5.1}$$

where: p is the true/label distribution, and q is the model/predicted distribution.

Regularization contingencies, including batch-normalization (BN) and Dropout were implemented to prevent overfitting to the training data. At each iteration, BN regularization was implemented after the first two of the three convolutional layers and Dropout (p = 0.4) before the last layer, thereby randomly deactivating 40% of the units (or network connections).

Inference/Testing algorithm. The proposed pipeline was trained on the S1 cohort of 40 consecutive patients with histologically-confirmed FCD lesions. This trained model cascade then served probabilistic predictions on unseen datasets acquired at S1–S7 sites. For each test subject, input images were first partitioned into patches with voxel sampling limited to the FLAIR mask (intra-subject z-score > 0.1). The balanced patch dataset was evaluated using CNN_1, which effectively discards improbable lesion candidates. The remaining voxels (threshold > 10%) were re-evaluated by CNN_2 to obtain the final probabilistic lesion mask. Since, the cost of misclassifying the le-

sion as healthy tissue is severe, we applied a conservative threshold (> 10%) on the probabilistic prediction masks. A simple post-processing routine involving successive morphological erosion, dilation, and extraction of connected components (> 75 voxels), was executed to remove flat blobs and noise. The final segmentation masks were compared to manual expert annotations of the lesions.

5.3 Experiments

5.3.1 Subjects

We studied retrospective cohorts with FCD lesions histologically-confirmed after surgery from seven tertiary epilepsy centers worldwide (n = 107). The presurgical workup included neurologic examination, assessment of seizure history, neuroimaging, and video-EEG telemetry. Since the routine MRI was initially reported as unremarkable in 56 patients (52%), the location of the seizure focus was established using intracranially-implanted electrodes; in all patients, retrospective inspection revealed a subtle FCD in the seizure onset region.

Training cohort. The primary site (S1) comprised 40 patients (20 males, 35 adults; mean±SD age = 27 ± 9 years).

Independent testing cohorts. Independent test cohorts comprised 67 histologically-confirmed FCD (37 adults and 30 children; mean±SD age = 33 ± 11 years, 9 ± 6 years, respectively) from six sites with different scanners, and field strengths (1.5T, 3T). The control group consisted of 38 healthy individuals (age = 30 ± 7 years) and 63 disease controls with temporal lobe epilepsy (TLE) and histologically-verified hippocampal sclerosis (age = 31 ± 8), matched for age and sex to S1 cohort.

5.3.2 Performance evaluation

Evaluation of classification for S1. Two experts segmented independently 40 lesions on co-registered T1w and FLAIR images. Inter-rater dice agreement index [D = 2|M1∩M2|/(|M1|+|M2|) (M1: 1st label, M2: 2nd label; M1∩M2: intersection of M1 and M2] was 0.91 ± 0.11. The union of the two ground truth labels served to train the classifier. The classifier was trained using 5-fold cross validation repeated 20 times. Sensitivity was the proportion of patients in whom a detected cluster co-localized with the lesion label. Specificity was determined with respect to controls (*i.e.,* proportion of controls in whom no FCD lesion

cluster was falsely identified), and disease controls with TLE. We also report the number of clusters detected in patients remote from the lesion label (*i.e.,* false positives).

Evaluation of classifier generalizability. We tested the sensitivity of the classifier trained on S1 was tested on a held-out dataset of eight FCD patients from S1 and 59 independent FCD datasets from S2–S7. For the cross-site unbiased reporting of results blinded to clinical information, the prediction maps (in stereotaxic space) were sent back to respective sites to confirm or dispute the detection of the lesion.

Comparison with a benchmark surface-based classifier. We analysed the S1 dataset using a previously published method [129] based on an ensemble of RUSBoosted decision trees across two classification stages, which uses a total of 30 intensity and morphology features calculated on multimodal T1-weighted MPRAGE and FLAIR images. The classifiers were trained using 5-fold cross validation averaged across 10 iterations.

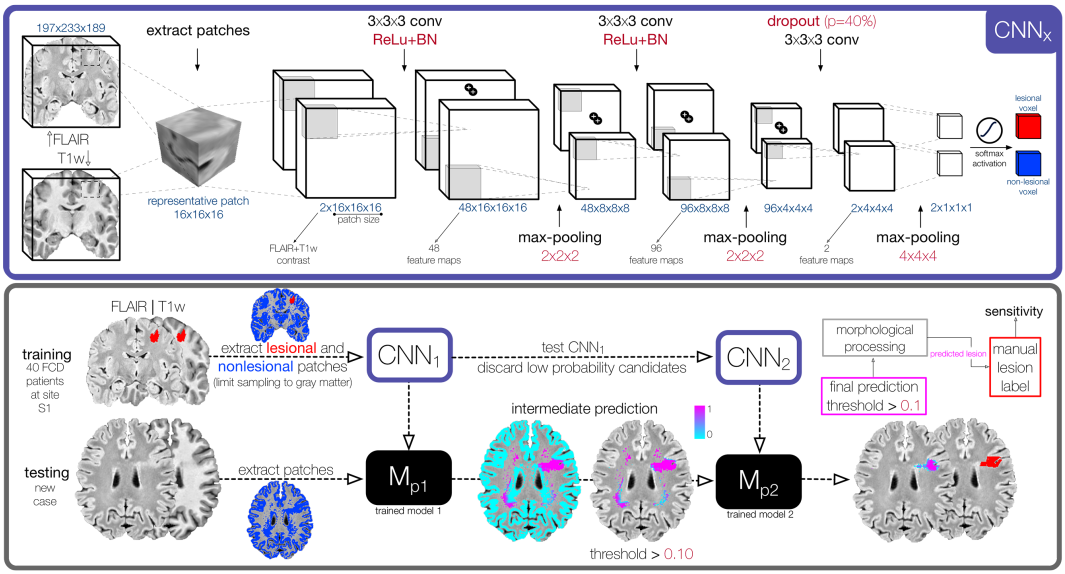


Figure 5.1: *Top panel*: Convolutional network architecture (CNN_x) for two-label (lesional *vs.* non-lesional) classification. *Bottom panel*: Training and testing schema using two-stage CNN_x cascade (CNN_1/CNN_2).

5.4 Results

The 5-fold cross-validation of the CNNs resulted in a sensitivity of 87 ± 4%, with an average of 35/40 lesions detected. In these cases, 2 ± 1 extra-lesional clusters were also detected. Specificity was 95% in healthy controls (3 ± 1 clusters in 2/38) and 90% in TLE (1 ± 0 cluster in 7/63).

For cross-dataset classification at seven sites, overall sensitivity was 91% (61/67 lesions detected) with 3 ± 2 extra-lesional clusters observed in 47/67 cases. Per-site sensitivity for S1–S7 was 100% (8/8 lesions detected, 2 ± 2 extra-lesional clusters), 86% (17/19, 4 ± 2), 89% (8/9, 2 ± 1), 75% (6/8, 2 ± 1), 100% (5/5, 5 ± 2), 91% (10/11, 2 ± 3), and 100% (7/7, 2 ± 2), respectively. Stratifying patients based on age, sensitivity in children (2-18.5 years old) was 90% (27/30 FCD detected, 4 ± 3 extra-lesional clusters) while in adults (> 19 years old) it was 92% (34/37, 3 ± 2). Figure 5.2 shows test case examples.

Training and testing a surface-based classifier based on S1 dataset yielded a lower performance with a sensitivity of 83 ± 2% (33/40 lesions detected), with 4 ± 5 extra-lesional clusters. Specificity was 92% in healthy controls (1 ± 0 cluster in 3/38).

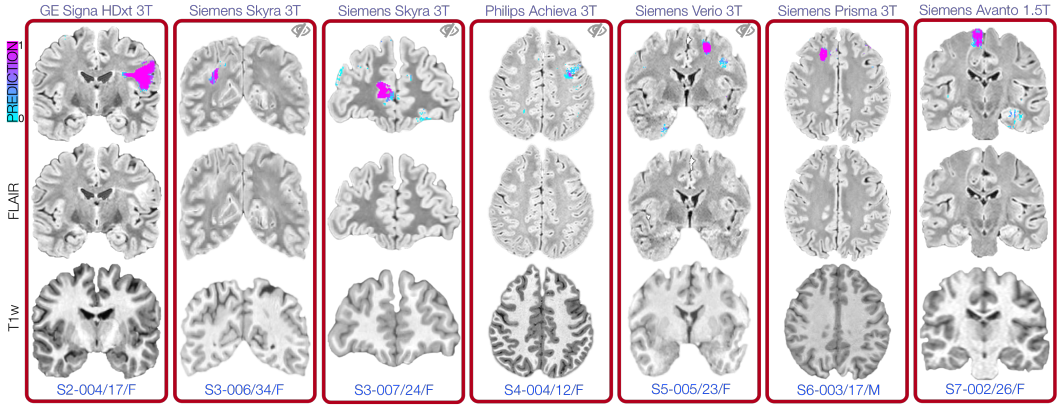


Figure 5.2: Classification results using the cascaded CNN_x trained on 40 FCD patients at site S1 (Siemens TrioTim 3T) to demonstrate generalizability for lesion detection along three axes of heterogeneity: scanner type, field strength (top labels), and age (bottom labels). The seven cases obtained using different scanners at six sites (excluding S1) are shown. The top row indicates the strength of prediction overlaid on the FLAIR, while the second/third rows show the corresponding FLAIR and T1w, respectively. The bottom labels are read as site-patient-ID/age/gender. MRI-negative cases are identified with 👁.

5.5 Discussion

We present the first deep learning method to segment FCD, with multicentric validation. Operating on routine multi-contrast MRI in voxel-space, our algorithm provides the highest performance to date. Furthermore, we demonstrated generalizability of a model trained on a single-site dataset by showing robust performance across independent cohorts from various centres worldwide with different age, scanner hardware and sequence parameters. Notably, > 50% of lesions were missed by conventional radiological inspection.

Operating at two consecutive levels, our classifier resulted in both high sensitivity and specificity. The number of false positive findings in healthy and disease controls were rather modest. Even though our algorithm was trained on an adult dataset, its performance was equally good in children. With respect to the latter, the use of age-appropriate templates taking into account the developmental trajectories, *i.e.,* age-varying tissue contrast, white matter myelination and cortical maturation, is likely to have contributed to the excellent performance by limiting the interpolation effects that would have occurred during registration using an adult template. Moreover, the overall high performance across cohorts strongly suggests that the network learns and optimizes parameters specific to FCD pathology, a fact validated by histological confirmation in all cases.

Compared to a state-of-the-art surface-based classifier, both sensitivity and specificity were higher using the current algorithm. Applying a surface-based approach to S2–S7 would have been challenging due to the large variability in image quality, which would require site-specific fine-tuning of algorithm parameters. A comprehensive comparison is part of future work. In addition, owing to the considerable time investment to manually correct brain tissue segmentation and surface extraction errors, which may have negative downstream effects on the fidelity of features extracted, the current approach is both time-effective and superior.

In conclusion, easy implementation, minimal pre-processing, significant performance gains and inference time of < 6 minutes/case make this classifier an ideal platform for large-scale clinical use, particularly in "MRI-negative" FCD.

6 Uncertainty-informed detection of epileptogenic brain malformations using Bayesian neural networks

Manuscript 4

Published as R. S. Gill, B. Caldairou, N. Bernasconi, and A. Bernasconi. "Uncertainty-informed detection of epileptogenic brain malformations using bayesian neural networks". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 225–233. DOI: 10.1007/978-3-030-32251-9_25

Uncertainty-informed detection of epileptogenic brain malformations using Bayesian neural networks

Ravnoor S. Gill, Benoit Caldairou, Neda Bernasconi, Andrea Bernasconi

Neuroimaging of Epilepsy Laboratory, McConnell Brain Imaging Center, Montreal Neurological Institute (MNI), Montreal, QC, Canada

Abstract

Focal cortical dysplasia (FCD) is a prevalent surgically-amenable epileptogenic malformation of cortical development. On MRI, FCD typically presents with cortical thickening, hyperintensity, and blurring of the gray-white matter interface. These changes may be visible to the naked eye, or subtle and be easily overlooked. Despite advances in MRI analytics, current machine learning algorithms fail to detect FCD in up to 50% of cases. Moreover, the deterministic nature of current algorithms does not allow conducting risk assessments of such predictions, an essential step in clinical decision-making. Here, we propose an algorithm formulated on Bayesian convolutional neural networks (CNN) providing information on prediction uncertainty, while leveraging this information to improve classification performance. Our classifier was trained on a patch-based augmented dataset derived from 56 patients with histologically-validated FCD to distinguish the lesion from healthy tissue. The algorithm was trained and cross-validated on multimodal 3 Tesla MRI data. Compared to a non-Bayesian learner with the same network architecture and complexity, the uncertainty-informed Bayesian CNN classifiers showed significant improvement in sensitivity (89% *vs.* 82%; $p < 0.05$) while specificity was high for both classifiers. We demonstrate empirically the effectiveness of our uncertainty-informed CNN algorithm, making it ideal for large-scale clinical diagnostics of FCD.

6.1 Introduction

Focal cortical dysplasia (FCD), a malformation of cortical development, is a frequent cause of drug-resistant epilepsy. This surgically-amenable lesion is characterized on histology by altered cortical laminar structure and cytological anomalies together with gliosis and demyelination, which may extend into the underlying white matter [16]. Complete lesion resection is associated with good post-surgical outcome [245]. On MRI, FCD typically presents with cortical thickening, hyperintensity and blurring of the gray-white matter interface. These changes may be visible to the naked eye on T1- and T2-weighted MRI, or subtle and easily overlooked (often referred to as "MRI-negative" FCD) [20].

Over the last decade, a number of automated FCD detection algorithms have been developed [31]. Recent methods rely on surface-based approaches [124, 129, 130], which allow to effectively model sulco-gyral morphology. Nevertheless, current algorithms fail to detect subtle FCD [20]. Also, arduous pre-processing and specialized expertise preclude their broader integration into clinical workflows. Importantly, they have not quantified the degree of uncertainty in predictions, a desirable information particularly when training datasets are small.

Convolutional neural networks (CNN) extract hierarchically increasing complex features from the data [177] without the need for user-defined feature engineering; they have achieved state-of-the-art performances in medical imaging (see Litjens *et al.* [178] for review). A standard approach to assess the reliability of CNN predictions is to rely on the probabilities obtained from the Softmax layer. However, such raw confidence scores may be miscalibrated [246]. Also, CNN predictions are typically deterministic. In contrast, traditional Bayesian machine learning assigns the degree of uncertainty (or confidence) to predictions through a probability density function. In clinical domains, uncertainty information has insofar been used to evaluate the robustness of predictions in multiple sclerosis [247] and diabetic retinopathy [201].

Dropout variational inference approximates Bayesian inference in models with large number of learnable parameters, for which exact Bayesian inference is computationally intractable [248]. This training strategy includes dropout after every convolutional layer; subsequently, a Monte Carlo dropout procedure, applied during testing, samples the posterior distribution to provide predictions [198]. The model's epistemic uncertainty is then derived from the mean and variance of the distribution of predictions. Here, we exploited the complementary diagnostic power of T1- and T2-weighted contrasts paired with an uncertainty-informed Bayesian CNN. Compared to a non-Bayesian CNN classifier with the same network architecture and complexity, the uncertainty-informed Bayesian classifier showed significant improvement in sensitivity, while maintaining high specificity.

6.2 Methods

6.2.1 MRI acquisition

Multimodal MRI was acquired on a 3T Siemens TimTrio using a 32-channel head coil, including: 3D T1-weighted MPRAGE (T1w; TR = 2300 ms, TE = 2.98 ms, flip angle = 9°, FOV = 256 mm^2, voxel size = 1×1×1 mm^3), and T2-weighted 3D fluid-attenuated inversion recovery (FLAIR; TR = 5000 ms, TE = 389 ms, TI = 1800 ms, flip angle = 120°, FOV = 230 mm^2, voxel size = 0.9×0.9×0.9 mm^3).

6.2.2 Image pre-processing

T1w and FLAIR images underwent intensity non-uniformity correction [115] and normalization. T1w images were then linearly registered (affine, 9 degrees of freedom) to the MNI152 symmetric template (1×1×1 mm^3) [249]. FLAIR images were linearly mapped to T1w images in MNI space. Skull-stripping excluded non-brain tissue.

6.2.3 Patch-based data augmentation

MRI sampling procedure. To prevent biasing the classifier towards healthy voxels, we constructed a patch-based dataset by randomly under-sampling healthy voxels such that the feature set was composed of equal number of examples from both classes. To this end, we sub-sampled multi-contrast 3D patches from the co-registered 3D T1w and FLAIR images, with each input image modality representing a channel. The data was normalized within each input modality with zero mean and unit variance. For each normalized training image, we computed 3D patches (16×16×16) centered on the voxel of interest. The set of all multimodal patches served as training dataset.

Sampling heuristics. We sampled hyperintense voxels based on FLAIR contrast by thresholding the subject-level z-normalized images and discarding the bottom 10 percentile intensities. This thresholding yielded a crude gray matter mask partially extending into the hyperintense white matter. This approach is biologically plausible since FCD lesions are primarily located in the gray matter [19]; moreover, their gray matter and white matter components are consistently hyperintense on FLAIR [242].

6.2.4 Network architecture and design

A typical convolutional neural network (CNN) consists of three stages: convolutions, nonlinearity, and pooling. We used a two-phase cascaded CNN training architecture [189] in which

weights of two identical CNNs are optimized independently, a procedure yielding efficient training when the distribution of labels is unbalanced [129, 130]. CNN1 was trained to maximize putative lesional voxels, while CNN2 reduced the number of misclassified voxels (*i.e.,* removing false positives while maintaining optimal sensitivity). Each fully convolutional network was composed of three stacks of convolution (filter size: 3×3×3) and max-pooling layers with 48, 96 and 2 filters, respectively. The rectified linear activation (ReLU) non-linearity function was applied to the first two of the three convolutional layers. Softmax non-linearity was used after the final convolution to normalize the result of the kernel convolutions into a binominal distribution over healthy and lesional labels. See Figure B.1 (Appendix B) for detailed network parameters.

6.2.5 Classification paradigm

Training algorithm. We used a validation set (75/25 training data split) to optimize the CNN weights. The training set is used to adjust the weights of the neural network, while the validation set measures the performance of the trained CNN after each epoch and continues until the validation error plateaus. The model is randomly initialized, and network parameters are learned iteratively via the adaptive learning rate method (AdaDelta) by minimizing the binary cross-entropy loss. Batch-normalization (BN) and Dropout were implemented to prevent overfitting.

Uncertainty estimation. We computed the posterior distribution $p(w|X, y)$, where X, y is the training dataset and w is the learned weights of the CNN. In practice, while the solution of this posterior is analytically intractable, variational inference (VI) methods approximate it with a parameterized distribution $q(w)$, while θ summarizes network parameters over a space of functions, and $x*$ represents a new input point (see Equation 6.1). The first term in Equation 6.2 maximizes the likelihood of the training data X, y, whereas the second term approximates the true distribution $p(w)$ by $q(w)$. Gal Ghahramani [248] empirically associates Equation 6.2 with dropout training to approximate the intractable integral with Monte Carlo sampling. This results in the conventional Softmax loss for dropout networks, for which units are dropped by drawing from a Bernoulli prior with probability drop for setting a unit to zero. The Kullback-Leibler (KL) term in Equation 6.2 was shown to correspond to a $L2$-regularization term in dropout networks.

$$p\big(\theta \mid X, y, x^{*}\big) \approx \int p(\theta \mid x^{*}, w) p\big(w \mid X, y\big) dw \approx \int p(\theta \mid x^{*}, w) q(w) dw \tag{6.1}$$

$$\mathcal{L} := \int p\big(y \mid X, w\big) q(w) dw - KL\big(q(w) \parallel p(w)\big) \tag{6.2}$$

Inference algorithm. The proposed pipeline (Figure 6.1) was trained and cross-validated using a 5-fold scheme on a cohort of 56 consecutive patients with histologically-confirmed FCD lesions. This trained model cascade served probabilistic predictions on held-out fold data. For

each test subject, input images were first partitioned into patches with voxel sampling limited to the FLAIR mask (intra-subject Z-score > 0.1). The balanced patch dataset served as input to CNN1. To discard improbable lesion candidates, we applied the following thresholding criteria. For the non-Bayesian classifier, we used a single forward pass at > 0.1. For the Bayesian classifier, we used separately the mean ($\mu_{bayesian}$) and variance ($\sigma_{bayesian}$; *i.e.,* uncertainty) of 20 stochastic forward passes thresholded at > 0.1 and > 0.05, respectively. For each of the three thresholding schemes, the resulting candidate voxels served as the input mask to sample patches for CNN2. For the non-Bayesian classifier, the remaining voxels (threshold > 0.1) provided the final probabilistic lesion mask. For each Bayesian experiment, we computed the mean and uncertainty of the predictions resulting from 50 forward passes and thresholded the mean at > 0.7. These thresholds are empirically determined by limiting the average cluster-level false positive rate to below 5 per patient. Finally, a simple post-processing routine involving successive morphological erosion, dilation, and extraction of connected components (> 75 voxels) removed flat blobs and noise. The final segmentation masks were compared to manual expert annotations of the lesions.

6.3 Experiments

6.3.1 Subjects

We studied a cohort of FCD lesions histologically-confirmed after surgery at a tertiary epilepsy center (n = 56). The pre-surgical workup included assessment of seizure history, video-EEG telemetry, and clinical neuroimaging. Since routine MRI was initially reported as unremarkable in 45 patients (80%), the location of the seizure focus was established using intracranially-implanted electrodes; in all, retrospective inspection revealed a subtle FCD in the seizure onset region.

Training and cross-validation cohort. The dataset comprised 56 patients (28 females, 45 adults; mean±SD age = 26 ± 10 years).

Independent testing cohorts. The control group consisted of 38 healthy individuals (age = 30 ± 7 years) and 63 disease controls with temporal lobe epilepsy (TLE) and histologically-verified hippocampal sclerosis (age = 31 ± 8), matched for age and sex to training cohort.

6.3.2 Performance evaluation

Evaluation of classification. Two experts manually segmented independently 56 lesions on co-registered T1w and FLAIR images. Inter-rater Dice agreement index was 0.93 ± 0.10. The union of the two ground truth labels served to train the classifier. The classifier was trained using 5-fold cross validation repeated 5 times. Sensitivity was the proportion of patients in whom a detected cluster co-localized with the lesion label. Specificity was determined with respect to controls (*i.e.,*

proportion of controls in whom no FCD lesion cluster was falsely identified), and disease controls with TLE. We also report the number of clusters detected in patients remote from the lesion label (*i.e.,* false positives; FP).

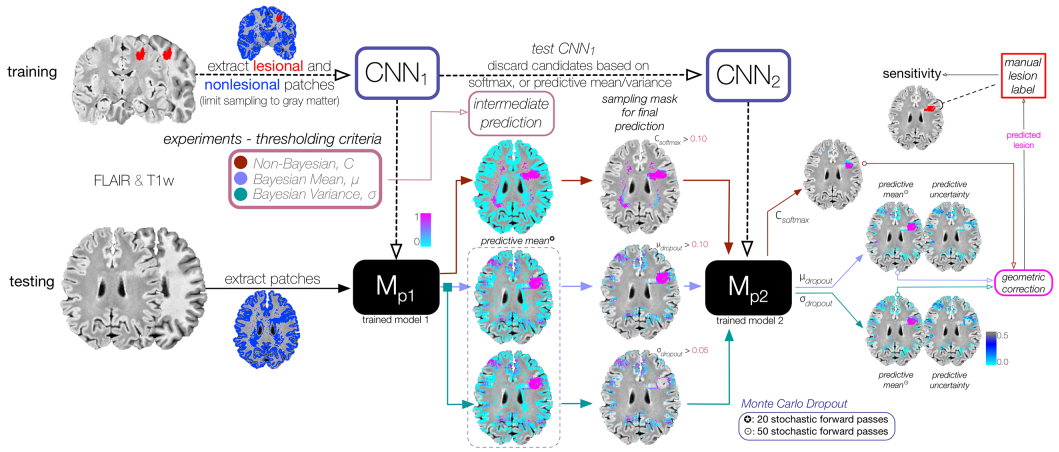


Figure 6.1: Training and testing scheme using the two-stage CNN cascade (CNN_1 and CNN_2) that incorporates uncertainty information using dropout Monte Carlo.

6.4 Results

The 5-fold cross-validation of the Bayesian CNN classifiers resulted in a sensitivity of 89%, with an average of 50/56 lesions detected, compared to 82% using the non-Bayesian CNN, at an identical cluster-wise FP rate. Non-parametric permutation tests (one-tailed, 10,000 iterations) assessing the pair-wise predictive accuracy based on area under the curves (AUCs) showed that sensitivity of the Bayesian CNNs was significantly higher than the non-Bayesian CNN (see Table 6.1).

Voxel-wise receiver operating characteristics (ROC) curves are shown in Figure 6.2A. Higher AUC scores signify better classification performance. Uncertainty values positively correlated with predictive probabilities at the individual level for both the mean-based thresholding (healthy controls: Pearson's $r = 0.81 \pm 0.03$, $p < 0.05$; TLE disease controls: Pearson's r: 0.77 ± 0.04, $p < 0.05$) and uncertainty-based thresholding (healthy controls: 0.78 ± 0.04, $p < 0.05$; TLE: 0.81 ± 0.03, $p < 0.05$). Specificity was 84% in healthy controls (no findings in 32/38; 1 ± 0 FPs) using Bayesian CNNs and the non-Bayesian CNN. Specificity was 87% (no findings in 55/63; 1 ± 0 FPs) in TLE controls using Bayesian CNNs, and slightly higher at 92% (no findings in 58/63; 1 ± 0 FPs) using the non-Bayesian CNN.

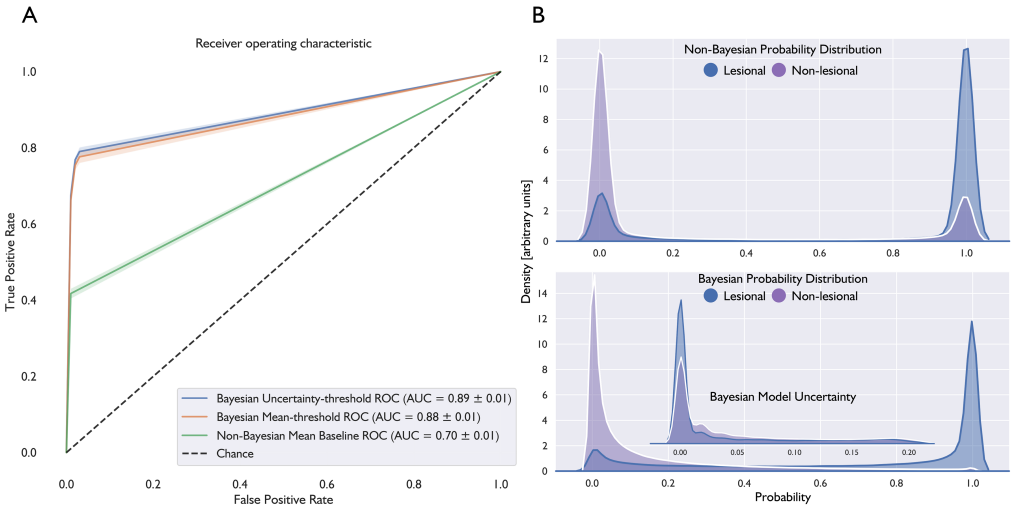


Figure 6.2: **A**. Receiver operating characteristic (ROC) curves of the three CNN classifiers. The opaque error line represents the ±1 standard deviation of the area under the curve (AUC) around the mean AUC (solid colored line). The dotted line represents the AUC for a random classifier. **B**. The posterior predictive distributional profiles for FCD lesions and non-lesional tissue of the non-Bayesian CNN (top panel) and Bayesian CNN (bottom panel–only mean based thresholding depicted). The Bayesian model uncertainty is shown (inset) in the bottom panel.

Table 6.1: Performance metrics for the three CNN classifiers. Sensitivity is derived after averaging across 5 trials and thresholding to aggregate voxel as clusters. The rate of false positives (FP) clusters is averaged across patients with FCD. The Dice index represents FCD lesion coverage compared to manual labeling.

CNN Classifier	**Sensitivity**	**FP**	**Dice**	**AUC Permutation Tests**
Non-Bayesian (C1)	82% (46/56)	4 ± 5	0.49	—
Bayesian (C2; mean-based threshold)	89% (50/56)	5 ± 4	0.47	C2 > C1 ($p < 0.05$)
Bayesian (C3; uncertainty-based threshold)	89% (50/56)	5 ± 5	0.47	C3 > C1 ($p < 0.05$)

6.5 Discussion

We present the first deep learning method for automated FCD detection trained and validated on histologically verified data from multiple centers worldwide. The classifier uses T1- and T2-weighted FLAIR, contrasts available on most recent MR scanners [167], operates on 3D voxel space without laborious pre-processing and feature extraction, and pairs predictions with confidence. It yields the highest performance to date with an overall sensitivity of 93% and 89% specificity, both in healthy and disease controls. Importantly, deep learning detected MRI-negative FCD with 85% sensitivity, thus offering a considerable gain over standard radiological assessment. Results were generalizable across cohorts with variable age, hardware, and sequence parameters. Taken together, such characteristics and performance promise unprecedented potential for broad clinical translation. We present the first deep learning-based method to segment FCD that leverages uncertainty for clinical decision-making with the highest sensitivity to date. Notably, epistemic uncertainty is important for safety-critical applications and instances with small datasets [198].

Our framework exploits uncertainty both during the intermediate testing and the final prediction. The calibration of posterior probabilities during the intermediate step is apparent in Figure 6.2B (group evaluation) and Figure B.2 (individual evaluation in Appendix B) showing that Bayesian CNN is more effective in separating tissue classes than the non-Bayesian CNN, a result attributable to fitting multiple hyperplanes in the former rather than just one. This also explains why the Bayesian classifier detected lesions that were missed by the non-Bayesian learner (as exemplified by P2 in Figure B.2).

While being superior to the non-Bayesian classifier, both Bayesians CNNs yielded equal performance at the patient level. Notably, the number of FPs in healthy and disease controls were minimal, and highly correlated with their degree of uncertainty. Moreover, the overall high performance across cohorts strongly suggests that the CNN learns and optimizes parameters specific to FCD pathology, a fact validated by histological confirmation in all cases.

FCD lesions manifest on a biological spectrum ranging from subtle to severe. Subtle FCD resembles the healthy cortex and may thus present with high uncertainty, while the predictive mean may be high or low. Within the automated framework, the choice of parameters is based on the whole dataset since it is not possible to anticipate where a new, test FCD may lie along the spectrum. While we have shown that the automated approach is effective, a human-in-the-loop is more appropriate for the final clinical decision. Therefore, on account of the variability stemming from overlap in distributions of lesional and non-lesional tissue (as is evident from Figure 6.2B and Figure B.2), the uncertainty map would be best suited for an individualized analysis where the clinician rules out false positives in conjunction with converging evidence from other

independent exams. Thus, uncertainty estimates can be used to refer uncertain predictions to experts for further evaluation This is especially important when considering that 80% of the FCD lesions detected by the CNN were missed by conventional radiological inspection. Finally, these estimates have the added benefit of being readily computed without the need to re-train the existing models or increasing model or time complexity.

In conclusion, easy implementation, minimal pre-processing, significant performance gains coupled with uncertainty information about predictions make our CNN classifier an ideal platform for large-scale clinical use, particularly in "MRI-negative" FCD.

7 Multicenter validated detection of focal cortical dysplasia using deep learning

Manuscript 5

Published as R. S. Gill, H.-M. Lee, B. Caldairou, S.-J. Hong, C. Barba, F. Deleo, L. D'Incerti, V. C. Mendes Coelho, M. Lenge, M. Semmelroch, D. V. Schrader, F. Bartolomei, M. Guye, A. Schulze-Bonhage, H. Urbach, K. H. 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, 2021, e1571–e1582. DOI: 10.1212/WNL.0000000000012698

Multicenter validated detection of focal cortical dysplasia using deep learning

Ravnoor S. Gill, PhDc [1]; Hyo M. Lee, PhDc [1]; Benoit Caldairou, PhD [1]; Seok-Jun Hong, PhD [1]; Carmen Barba, MD [2]; Francesco Deleo, MD [3]; Ludovico D'Incerti, MD [4]; Vanessa C. Mendes Coelho, MD [5]; Matteo Lenge, PhD [2]; Mira Semmelroch, PhD [6]; Dewi Schrader, MD [7]; Fabrice Bartolomei, MD [8]; Maxime Guye, MD PhD [9]; Andreas Schulze-Bonhage, MD [10]; Horst Urbach, MD [10]; Kyoo Ho Cho, MD [11]; Fernando Cendes, MD PhD [5]; Renzo Guerrini, MD [2]; Graeme Jackson, MD [6]; R. Edward Hogan, MD [12]; Neda Bernasconi, MD PhD [1]; Andrea Bernasconi, MD [1]

[1] Neuroimaging of Epilepsy Laboratory, Montreal Neurological Institute, McGill University, Montreal, Canada; [2] Children's Hospital A. Meyer-University of Florence, Italy; [3] Epilepsy Unit, Fondazione IRCCS Istituto Neurologico C. Besta, Milano, Italy; [4] Fondazione IRCCS Istituto Neurologico C. Besta, Milano, Italy; [5] University of Campinas, Brazil; [6] The Florey Institute of Neuroscience and Mental Health and The University of Melbourne, Victoria, Australia; [7] British Columbia Children's Hospital, Vancouver, Canada; [8] Aix Marseille University, INSERM UMR 1106, INS, Marseille, France; [9] Aix Marseille University, Marseille, France; [10] Freiburg Epilepsy Center, Universitätsklinikum Freiburg, Germany; [11] Yonsei University College of Medicine, Seoul, Korea; [12] Washington University School of Medicine, St Louis, Missouri, USA

Abstract

Objective. To test the hypothesis that a multicenter-validated computer deep learning algorithm detects MRI-negative focal cortical dysplasia (FCD).

Methods. We used clinically-acquired 3D T1-weighted and 3D FLAIR MRI of 148 patients (median age, 23 years [range, 2–55]; 47% female) with histologically-verified FCD at nine centers to train a deep convolutional neural network (CNN) classifier. Images were initially deemed as MRI-negative in 51% of cases, in whom intracranial EEG determined the focus. For risk stratification, the CNN incorporated Bayesian uncertainty estimation as a measure of confidence. To evaluate performance, detection maps were compared to expert FCD manual labels. Sensitivity was tested in an independent cohort of 23 FCD cases (13 ± 10 years). Applying the algorithm to 42 healthy and 89 temporal lobe epilepsy disease controls tested specificity.

Results. Overall sensitivity was 93% (137/148 FCD detected) using a *leave-one-site-out* cross-validation, with an average of six false positives per patient. Sensitivity in MRI-negative FCD was 85%. In 73% of patients, the FCD was among the clusters with the highest confidence; in half it ranked the highest. Sensitivity in the independent cohort was 83% (19/23; average of five false positives per patient). Specificity was 89% in healthy and disease controls.

Conclusions. This first multicenter-validated deep learning detection algorithm yields the highest sensitivity to date in MRI-negative FCD. By pairing predictions with risk stratification this classifier may assist clinicians to adjust hypotheses relative to other tests, increasing diagnostic confidence. Moreover, generalizability across age and MRI hardware makes this approach ideal for pre-surgical evaluation of MRI-negative epilepsy.

Classification of evidence. This study provides Class III evidence that deep learning on multimodal MRI accurately identifies FCD in epilepsy patients initially diagnosed as MRI-negative.

7.1 Introduction

Focal cortical dysplasia (FCD), a surgically-amenable developmental epileptogenic brain malformation, presents with cortical thickening on T1-weighted MRI, as well as hyperintensity and blurring of the gray-white matter interface on FLAIR images. While these features are often visible to the naked eye, FCD may be overlooked and only found at surgery [20]. MRI-negative patients represents a major diagnostic challenge [210].

Currently, benchmark automated detection methods fail in 20–40% of patients [128–130, 251], particularly those with subtle FCD, and suffer from high false positive rates [31]. Conversely, deep neural networks outperform state-of-the-art methods at disease detection [see 178, 179, for review]. Specifically, convolutional neural networks (CNNs) learn abstract concepts from high-dimensional data alleviating the challenging task of hand-crafting features [177]. The integration of convolutional operators that implicitly encode spatial covariance of neighboring voxels (rather than treating each voxel independently) with nonlinearity capturing complex patterns and variability is expected to optimize the detection of the full FCD spectrum. Notably, with regards to diagnostic performance, the deterministic nature of conventional algorithms does not permit risk assessment of the automated decisions, a requirement to be integrated into clinical diagnostic systems. Alternatively, Bayesian CNNs provide a distribution of predictions from which the mean and variance can be computed, the latter being interpreted as a measure of uncertainty [201].

Here, we tested the hypothesis that a multicenter-validated computer deep learning algorithm operating directly on T1-weighted and FLAIR MRI voxel detects MRI-negative focal cortical dysplasia (FCD).

7.2 Methods

We present an automated algorithm trained and validated on a multicenter dataset of patients with histologically confirmed FCD. We ruled out sources of spectrum bias [252] by evaluating specificity against healthy individuals as well as a disease control cohort of patients with temporal lobe epilepsy (TLE) and histologically confirmed hippocampal sclerosis (HS). To minimize incorporation bias [252], the classifier was iteratively trained and tested using a *leave-one-site-out* scheme; *i.e.,* the classifier was trained iteratively on all sites minus the one held-out for testing; this guaranteed the out-of-fold validation in which tested cohorts were never part of the training. Moreover, the classifier trained on the full dataset was tested on an independent cohort of patients that were never part of training. According to the Classification of evidence schemes of the American Academy of Neurology (`https://www.neurology.org/sites/default/files/ifa/loe.pdf`) [253], this study satisfies the rating for Class III evidence for diagnostic accuracy, demonstrating that

deep learning operating on multimodal MRI has significant diagnostic value, including in MRI-negative patients, with 85% sensitivity.

7.2.1 Subjects

We studied consecutive retrospective cohorts from nine tertiary epilepsy centers worldwide with histologically validated FCD lesions collected from October 2012 to January 2018 and in whom both 3D T1-weighted MRI and 3D FLAIR were acquired as part of the clinical presurgical investigation [167]. The TLE cohort included both patients with MRI-visible HS (n = 49; comparable to MRI-positive FCD) and those in whom the MRI was unremarkable, but the histological examination of the surgical specimen revealed the presence of subtle HS (n = 40; comparable to MRI-negative, histology-positive FCD). Patients had been investigated for drug-resistant epilepsy with a standard presurgical workup including neurological examination, assessment of seizure history, neuroimaging, and video-EEG recordings.

On histological examination of the surgical specimen [16], FCD Type-II was defined as disrupted cortical lamination with dysmorphic neurons in isolation (IIA, n = 70) or together with balloon cells (IIB, n = 78). At a mean±SD postoperative follow-up of 31.2 ± 14.4 months (range: 12-78 months), 103 patients (70%) became seizure-free (Engel-I), 33 (22%) had rare disabling seizures (Engel-II), nine (6%) had worthwhile improvement (Engel-III) and three (2%) had no improvement (Engel-IV); in patients with Engel-III and IV, the resection was incomplete as the FCD encroached eloquent areas in primary cortices (7 in sensorimotor, 2 in primary visual and 3 in language areas); the residual lesion and extent of resection were evaluated on post-operative MRI.

7.2.2 Standard protocol approvals, registrations, and patient consent

The Ethics Committees and institutional review boards at all participating sites (S1–S9) approved the study, and written informed consent was obtained from all participants.

7.2.3 MRI acquisition and image processing

High-resolution 3D T1-weighted and 3D FLAIR MRI images were acquired in all individuals [167]. All images were obtained on 3T scanners; one site provided additional cases with 1.5T MRI. Imaging parameters are listed on Table C.1 (available from Dryad: `doi.org/10.5061/dryad.h70rxwdgm`). MRI data was de-identified; files were converted from DICOM to NIfTI with header anonymization. T1-weighted images were linearly registered to the MNI152 symmetric template. FLAIR images were linearly mapped to T1-weighted MRI in MNI space. T1-weighted and FLAIR underwent intensity non-uniformity correction [115] followed by intensity

standardization with scaling of values between 0 and 100. Finally, images were skull-stripped using an in-house deep learning method (*v.* 1.0.0; https://github.com/NOEL-MNI/deepMask; doi.org/10.5281/zenodo.4521716) trained on manually corrected brain masks from patients with cortical malformations. Two experts manually segmented lesions on co-registered T1-weighted and FLAIR images; inter-rater Dice agreement was 0.92 ± 0.10 [calculated as $2 \cdot |M_1 \cap M_2| / (|M_1| + |M_2|)$, where M_1 = label 1, M_2 = label 2, $M_1 \cap M_2$ = intersection of M_1 and M_2].

7.2.4 Classifier design

Figures 7.1 and C.1 (Appendix C; also available from Dryad: doi.org/10.5061/dryad.h70rxwdgm) illustrate the design. The full methodology is described in Appendix C (also available from Dryad: doi.org/10.5061/dryad.h70rxwdgm).

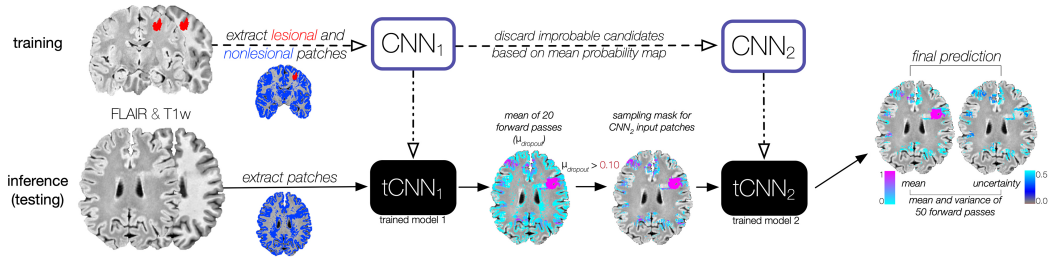


Figure 7.1: **Classifier design**. Training and inference (or testing) workflow. In the cascaded system the output of CNN-1 serves as an input for CNN-2. CNN-1 maximizes the detection of lesional voxels; CNN-2 reduces the number of misclassified voxels, removing false positives (FPs) while maintaining optimal sensitivity. The training procedure (indicated by dashed arrows) operating on T1-weighted and FLAIR MRI, extracts 3D patches from lesional and non-lesional tissue to yield tCNN-1 (trained model 1) and tCNN-2 (trained model 2) models with optimized weights (indicated by vertical dashed-dotted arrows). These models are then used for subject-level inference. For each unseen subject, the inference pipeline (solid arrows) uses tCNN-1 and generates a mean ($\mu_{dropout}$) of 20 predictions (forward passes); the mean map is then thresholded voxel-wise to discard improbable lesion candidates ($\mu_{dropout} > 0.1$). The resulting binary mask serves to sample the input patches for the tCNN-2. A mean probability and uncertainty maps are obtained by collating 50 predictions; uncertainty is transformed into confidence. The sampling strategy (identical for training and inference) is only illustrated for testing.

Data sampling and network architecture. In each individual, we thresholded FLAIR images by z-normalizing intensities and discarding the bottom 10 percentile intensities; this internal thresholding resulted in a mask containing voxels within the grey matter (GM) and its interface with the white matter (WM). This mask was then used to extract 3D patches (*i.e.,* regions of interest centered around a given voxel) from co-registered 3D T1-weighted and FLAIR images, which served as input to the network. Notably, 3D patches seamlessly sampled the FCD across orthogo-

Table 7.1: Demographics and dataset stratification for cross-site validation. *Abbreviations*. S = site; N: sample size; yrs = years; I and II refer to different MRI scanners for the same site

TESTING DATASET				**TRAINING DATASETS**			
Site	**N**	**Age (mean±SD yrs)**	**% Female**	**Sites**	**N**	**Age (mean±SD yrs)**	**% Female**
S1-I	45	27 ± 9	49%	S1-II, S2–S9	103	20 ± 13	46%
S1-II	17	18 ± 9	65%	S1-I, S2–S9	131	23 ± 13	44%
S2	08	11 ± 6	25%	S1, S3–S9	140	23 ± 12	48%
S3	05	22 ± 17	80%	S1–S2, S4–S9	143	23 ± 12	45%
S4	11	8 ± 7	36%	S1–S3, S5–S9	137	24 ± 12	44%
S5-I	10	23 ± 14	30%	S1–S4, S5-II, S6–S9	138	23 ± 13	48%
S5-II	12	13 ± 12	42%	S1–S4, S5-I, S6–S9	136	22 ± 12	47%
S6	11	31 ± 15	64%	S1–S5, S7–S9	137	22 ± 12	45%
S7	09	33 ± 13	33%	S1–S6, S8–S9	139	22 ± 13	47%
S8	07	24 ± 13	43%	S1–S7, S9	141	22 ± 13	47%
S9	13	26 ± 8	38%	S1–S8	135	22 ± 13	47%

nal planes and tissue types. We designed a cascaded system in which the output of the first CNN (CNN-1) served as input to the second (CNN-2). CNN-1 aimed at maximizing the detection of lesional voxels; CNN-2 reduced the number of misclassified voxels, removing false positives (FPs) while maintaining optimal sensitivity. In brief, for each test subject, 3D T1-weighted and FLAIR patches were fed to CNN-1. To discard improbable lesional candidates, the mean of 20 forward passes (or predictions) was thresholded at > 0.1 (equivalent to rejecting bottom 10 percentile probabilities); voxels surviving this threshold served as the input to sample patches for CNN-2.

Estimation of prediction uncertainty. Bayesian inference in deep CNNs with large number of parameters is computationally intensive [248]. By probabilistically excluding neurons (or units) after every convolutional layer during training, the Monte Carlo dropout method [254] simulates an ensemble of neural networks with diverse architectures, thus preventing overfitting without compromising on accuracy. This procedure provides a distribution of posterior probabilities at each voxel resulting from multiple stochastic forward passes through the classifier; their variance provides a measure of uncertainty. Here, we used the mean and variance of 50 voxel-wise forward passes to generate probability and uncertainty maps. The mean probability map was binarized by thresholding at > 0.7 (empirically determined by setting the cluster-level FP rate to < 6) and underwent a post-processing routine entailing morphological erosion, dilation and extraction of connected components (> 75 voxels) to remove flat blobs and noise, a procedure that resulted

in non-overlapping clusters. To evaluate performance, this detection map was compared to the manual expert annotation.

Transforming uncertainty into confidence and ranking. For each cluster of the detection map, we estimated confidence by computing the median uncertainty across its voxels; we then aggregated uncertainties across all clusters and normalized values between 0 and 1 to obtain a measure of confidence. All clusters were then ranked based on their confidence estimates with the highest confidence cluster as rank 1, second highest confidence cluster rank 2, and so on until all clusters surviving the threshold (probability > 0.7 and spatial extent > 75 voxels) had been ranked. Confidence maps were evaluated together with a diagram plotting lesion probability against lesion ranking, with rank 1 signifying highest confidence to be lesional, regardless of cluster size.

Performance evaluation. To assess performance, we employed a *leave-one-site-out* cross-validation by which the classifier trained on eight sites was tested iteratively on the held-out site until all sites had served as testing set. A minimum of one voxel co-localizing with the manually segmented FCD (ground truth label) was deemed as TP, any detection not co-localizing as FP. Consistent with previous FCD detection literature [124, 126, 130, 231], we deemed partial overlap to be sufficient without requiring the detection to be completely within the expert label. Demographics and dataset stratification are shown in Table 7.1. In addition, we evaluated the algorithm trained on the complete dataset of the 148 FCD patients on an independent cohort of 23 FCD cases (11 females; 13 ± 10 years; 70% MRI-negative) from S1 and S2.

Patient-level (*i.e.,* lesion-level) evaluation metrics included sensitivity $(P, L) = |P_1 \cap L_1|/|L_1|$ and specificity $(P, L) = |P_0 \cap L_0|/|L_0|$, where P is the model prediction and L the ground truth label; L_1 and L_0 signify voxels predicted as positive (lesional) and negative (not lesional), while P_1 and P_0 represent the same for model predictions. We evaluated specificity as the absence of any findings by applying the algorithm trained on the complete dataset of FCD patients to healthy controls and TLE disease controls; in other words, specificity was calculated as the proportion of healthy or disease controls in whom no FCD lesion cluster was falsely identified. Site-wise area under the receiver operating characteristic curve (AUC) evaluated voxel-wise classification performance (*i.e.,* the true positive (TP) *vs.* FP rate) stratified by sites.

We evaluated the spatial relation between lesional clusters and FPs in patients as well as healthy and disease controls. To this end, we generated a lesional probability map by overlaying all manually segmented FCD labels; the Dice coefficient quantified the overlap between the FCD probability map and both the group-wise probability and uncertainty maps of FPs.

Pearson's correlation quantified associations between probability and uncertainty, and between age and the number of FPs. Biserial correlation evaluated association between MRI-negative status and the number of FPs. *Spearman's* correlation quantified association between lesion rank

and probability. Nonparametric permutations (10,000 iterations with replacement) tested group differences at $p < .05$ (two-tailed), with *Bonferroni* correction for multiple comparisons.

7.2.5 Data availability statement

These datasets are not publicly available as they contain information that could compromise the privacy of research participants. The source code and pre-trained model weights are available for download at `https://github.com/NOEL-MNI/deepFCD` (*v.* 1.0.0; `https://doi.org/10.5281/zenodo.4521706`). In addition, a derivative dataset composed of lesional and non-lesional patches from 148 FCD patients is available as a Hierarchical Data Format (HDF5) dataset (available from Zenodo: `https://doi.org/10.5281/zenodo.3239446`).

7.3 Results

Demographics. The primary site (S1) comprised 62 FCD patients (33 females; mean±SD age = 25 ± 10 years) and control groups consisting of age- and sex- matched healthy individuals (n = 42; 22 females; 30 ± 7 years), and patients with TLE and histologically verified HS (n = 89; 47 females; age: 31 ± 8). Across the remaining eight sites (S2–S9), the cohort comprised 86 FCD patients (36 females; age: 20 ± 14). In 75 patients (51%) in whom routine MRI evaluation was initially reported as unremarkable in the initial readings of the neuroradiologists at each participating center, the location of the seizure focus was established using intracranial EEG.

Patient-level performance. The classifier's overall sensitivity based on *leave-one-site-out* cross-validation was 93% (137/148 FCD lesions detected), with 6 ± 5 FP clusters per patient. Stratifying children and adults, sensitivity was 98% for the former (52/53; 7 ± 5 FP clusters) and 89% (85/95; 5 ± 5 FP) for the latter. Notably, 85% of MRI-negative and 100% of MRI-positive lesions were detected. When testing the classifier on the independent cohort (using the model trained on the complete dataset of the 148 FCD patients), overall sensitivity was 83% (19/23 FCD lesions detected; 5 ± 3 FP clusters per patient) with 100% of MRI-positive and 75% of MRI-negative lesions detected. Specificity was 90% in healthy (4/42 with 2 ± 1 FP clusters) and 89% in TLE disease controls (10/89, 1 ± 0 FP cluster). With respect to the latter, specificity was similar between MRI-positive HS (92%; 5/49, 1 ± 0 FP cluster) and MRI-negative HS (88%; 5/40, 1 ± 0 FP cluster). Per-site sensitivity and FP rates are shown in Table 7.2.

Voxel-wise performance. The median AUC was 0.83 (range, 0.72–0.87) indicative of high sensitivity (high TP rates) and specificity (low FP rates), with comparable performance across sites.

Analysis of confidence. In 73% of patients, the FCD lesion was among the five clusters with the highest confidence; in half of them, it ranked the highest, with a mean probability of 72% (95%

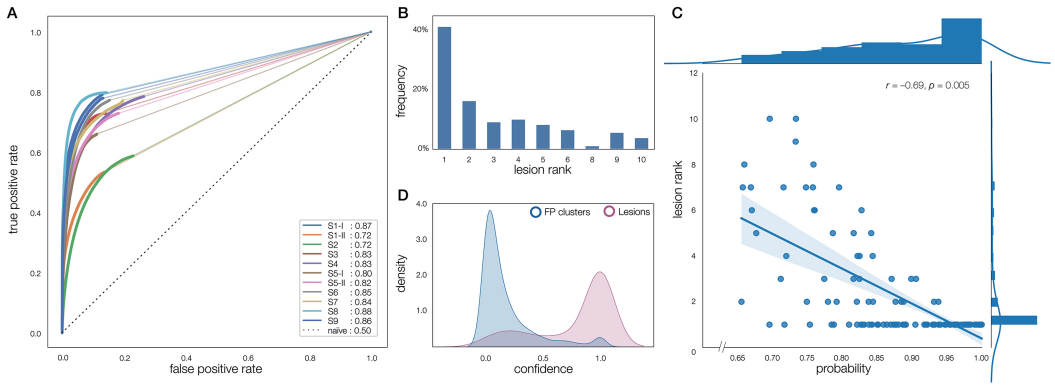


Figure 7.2: **Performance evaluation**. **A**. Site-wise area under the receiver operating characteristic curve (AUC) using the *leave-one-site-out* cross-validation (solid colored lines with values; black dotted line represents a naïve classifier). **B**. Frequency of lesions according to their rank. Rank 1 signifies highest confidence to be lesional. 73% of lesions were distributed across ranks 1 to 5. **C**. Lesion rank plotted against probability of being lesional shows a significant correlation with FCD voxels having low rank values (high confidence) and high probability. **D**. Distribution (kernel density estimation) of confidence for lesional and false positive (FP) clusters; lesions exhibit high confidence values, while FP clusters show low confidence.

Table 7.2: Site-specific demographics and performance metrics

Site	**N**	**Age (mean±SD yrs)**	**% Female**	**MRI+/ MRI−**	**Sensitivity**		**FPs**
					All patients	**MRI−**	
S1-I	45	27 ± 9	49%	13/32	39/45 (87%)	26/32 (81%)	7 ± 4
S1-II	17	18 ± 9	65%	2/15	15/17 (88%)	13/15 (87%)	7 ± 4
S2	08	11 ± 6	25%	5/3	8/8 (100%)	3/3 (100%)	6 ± 5
S3	05	22 ± 17	80%	2/3	5/5 (100%)	3/3 (100%)	1 ± 1
S4	11	8 ± 7	36%	11/0	11/11 (100%)	n/a	8 ± 6
S-I	10	23 ± 14	30%	8/2	9/10 (90%)	1/2 (50%)	10 ± 6
S5-II	12	13 ± 12	42%	11/1	12/12 (100%)	1/1 (100%)	6 ± 7
S6	11	31 ± 15	64%	6/5	11/11 (100%)	5/5 (100%)	3 ± 3
S7	09	33 ± 13	33%	2/7	8 /9 (89%)	6/7 (86%)	8 ± 6
S8	07	24 ± 13	43%	6/1	6/7 (86%)	0/1 (0%)	6 ± 5
S9	13	26 ± 8	38%	7/6	13/13 (100%)	6/6 (100%)	1 ± 2
Total	148	23 ± 13	47%	49/51%	137/148 (93%)	64/75 (85%)	6 ± 5
Indep	23	13 ± 10	48%	30/70%	19/23 (83%)	12/16 (75%)	5 ± 3

confidence interval, 69%–76%; Figure 7.2B). Lesion rank negatively correlated with probability, *i.e.,* the lower the rank, the higher the probability of being lesional ($r = -0.69$, $p = 0.005$; Figure 7.2C). Moreover, confidence for a cluster to be lesional centered around 1 (*i.e.,* 100% confidence), while for FPs it centered around zero (Figure 7.2D). Representative MRI-negative cases are shown in Figures 7.3 and 7.4.

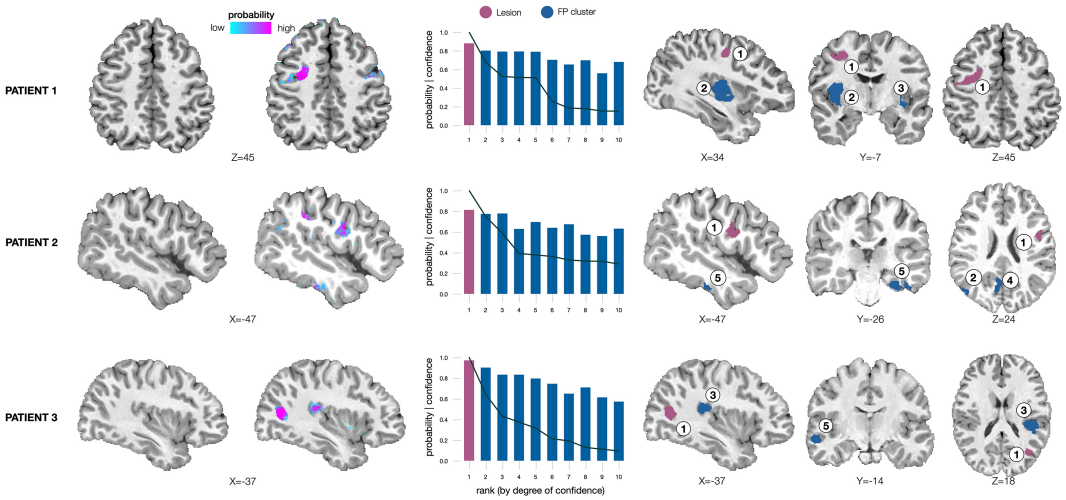


Figure 7.3: **Automated detection of MRI-negative FCD**. The left panels show the T1-weighted MRI and the prediction probability maps with the lesion circled. The middle plots show the probability of the lesion and false positive (FP) clusters sorted by their rank; the superimposed line indicates the degree of confidence for each cluster. The right panels illustrate the location of the FCD lesion (rank 1, highest confidence; purple) and FP clusters (ranks 2–5; blue). In these cases, the lesion has both highest confidence (rank 1) and high probability (> 0.8).

Spatial distribution of FCD and FPs. The majority of FCD lesions were located within the frontal lobes (Figure 7.5A). Overall, FPs in patients, healthy and disease controls (Figure 7.5B) were found in the insula and the parahippocampus (Dice overlap with FCD: 21%, 22% and 34%, respectively). Notably, FPs in healthy and disease controls overlapped to a greater extent (Dice: 52%) and exhibited low confidence to be lesional (*i.e.,* high uncertainty); conversely, FPs in FCD patients tended to display high confidence to be lesional ($p = 0.013$). Coordinates for the lesion and FPs are listed on Tables C.2 and C.3, respectively (also available from Dryad: `doi.org/10.5061/dryad.h70rxwdgm`). The incidence of FP clusters was negatively correlated with age ($r = -0.23$, $p = 0.004$), namely the younger the patients the higher the number of FPs. Number of FPs was not significantly different between MRI-positive and MRI-negative patients.

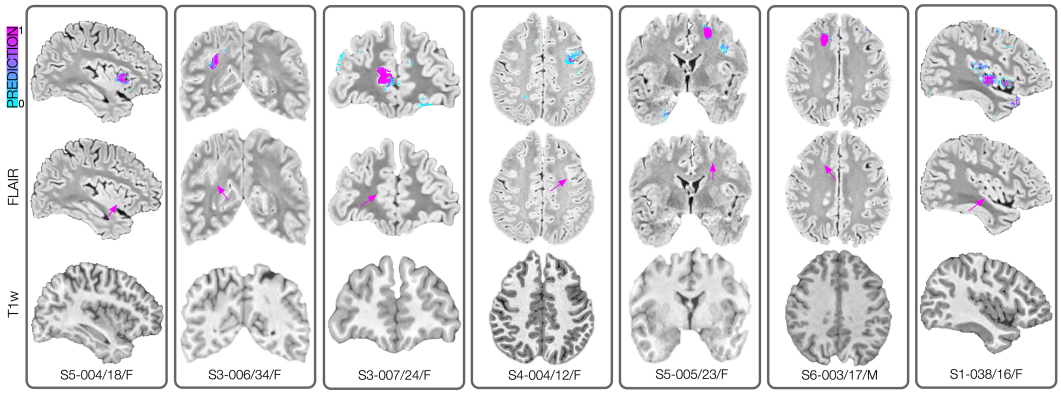


Figure 7.4: **Representative FCD detection examples**. Seven representative MRI-negative FCD lesions across sites are shown (*top row*: prediction overlaid on the FLAIR; the flame scale indicates the probability strength). The bottom labels are interpreted as site-patient-ID/age/gender. The arrows indicate the ground-truth lesion location.

7.4 Discussion

MRI-negative FCD represents a major diagnostic challenge. To define the epileptogenic area patients undergo long and costly hospitalizations for EEG monitoring with intracerebral electrodes, a procedure that carries risks similar to surgery itself [162, 211]. Moreover, patients without MRI evidence for FCD are less likely to undergo surgery and consistently show worse seizure control compared to those with visible lesions [4, 213]. Here, we present the first deep learning method for automated FCD detection trained and validated on histologically verified data from multiple centers worldwide. The classifier uses T1- and T2-weighted FLAIR, contrasts available on most recent MR scanners [167], operates in 3D voxel space without laborious pre-processing and feature extraction, and pairs predictions with confidence. It yields the highest performance to date with a sensitivity of 93% using a *leave-one-site-out* cross-validation and 83% when tested on an independent cohort, while maintaining a high specificity of 89% both in healthy and disease controls. Importantly, deep learning detected MRI-negative FCD with 85% sensitivity, thus offering a considerable gain over standard radiological assessment. Results were generalizable across cohorts with variable age, hardware and sequence parameters. Taken together, such characteristics and performance promise potential for broad clinical translation. Notwithstanding these advantages, good quality scans are essential to guarantee valid results, as motion can mimic lesions [167]; we thus advise against analysing low-quality motion-corrupted scans.

Deep learning: moving beyond conventional automated FCD detection

Over the last decade, several automated FCD detection algorithms have been developed, the most recent relying on surface-based representations [126, 129, 130, 231]. While the majority operate on

T1-weighted MRI, recent methods have combined T1-weighted and T2-weighted MRI for improved performance [124, 129]. A few have used shallow (single layer) artificial neural networks [124, 128]. Notably, all require arduous pre-processing, including manual corrections of tissue segmentation and surface extraction, thus precluding integration into clinical workflow. Importantly, they rely on domain knowledge to engineer features. These procedures generally fail to detect subtle lesions [31]. In comparison, our approach offers several advantages. Firstly, to optimize lesion detection across the FCD spectrum, we leveraged the power of CNNs that recursively learn complex properties from the data itself. Secondly, contrary to previous medical imaging applications relying on 2D orthogonal sampling, we extracted 3D patches to model the spatial extent of FCD across multiple slices and tissue types. Operating in true volumetric domain allowed assessing the spatial neighborhood of the lesion, whereas prior surface-based methods have considered each vertex location independently. Thirdly, restricting training to the GM reduced nearly infinite dataset to a manageable finite set. Finally, by relying on subject-wise feature normalization, rather than group-wise, our implementation obviates the need for a matched normative dataset, an expensive and time-intensive undertaking. Compared to previous deep learning methods [190, 195, 196] in which clinical description was scarce to absent, and information on the FCD expert labels and histological validation of lesions was not provided, our study relied on best-practice multimodal MRI, histologically-validated lesions, and a large dataset. Moreover, in previous work FLAIR images in presumably MRI-positive patients were acquired with inter-slice gap ranging from 0.5 to 1.0 mm [195, 196], and the acquisition parameters for the 3D T1-weighted images were different from those in healthy and disease controls [190].

Notwithstanding practical advantages of our method, a general limitation of deep learning is the reduced transparency of the process leading to the predictions, a consequence of the high dimensionality of learned features. The trade-off is a richer encoding and learning of complex spatial covariances of intensity and morphology that is beyond the ability of human eye. To maximize transparency and validity, we trained our algorithm on manual expert labels of histologically-validated FCD lesions. In addition to a rigorous cross-validation design, including applying the classifier to a totally independent cohort of FCD patients, our predictions were stratified according to confidence to be lesional. Notwithstanding these precautions, as for many diagnostic tests, the convergence of findings with independent tests is essential to increase confidence even further.

Estimation of generalizability is key to any diagnostic method. To guarantee unbiased evaluation, training and testing datasets should remain distinct. We thus devised a strategy in which the model was iteratively trained on patient data from all sites, except the one held-out. This guaranteed out-of-distribution validation in which tested cohorts were never part of the training. This *leave-one-site-out* cross-validation simulated a real-world scenario with optimal bias-variance trade-off compared to conventional train-test split of k-folds; it also exploited the full richness of

data during training and the out-of-distribution samples from a single site during testing. Moreover, the classifier trained on the full dataset was tested on a totally independent cohort of patients that were never part of training. Consistent high performance across cohorts, as well as modest FPs in healthy and disease controls, demonstrate that our cascaded CNN classifier learns and optimizes parameters specific to FCD, a fact validated by histological confirmation.

Human-in-the-loop machine learning: key to clinical translation

In machine learning, human-in-the-loop refers to the need for human interaction with the learner to improve human performance, machine performance, or both. Human involvement expedites the efficient labeling of difficult or novel cases that the machine has previously not encountered, reducing the potential for errors, a requirement of utmost importance in healthcare. In FCD, the outcome of surgery depends heavily on the identification of the lesion; it is thus crucial to decide which putative lesional clusters are significant. In this context, thresholding the final probabilistic mean map is essential to evaluate the balance between true positive and false positives. Notably, to guarantee an objective assessment of sensitivity and specificity across cohorts, in this study we defined an empirical threshold. However, in clinical practice, a judicious approach would imply adaptive thresholding of the maps at single-patient level, taking into account independent tests. Indeed, in 5/11 of undetected MRI-negative cases, the lesion could be resolved when modulating the threshold in light of seizure semiology and electrophysiology. Besides thresholding, confidence is pivotal in any diagnostic assessment, an aspect so far neglected. To fill this gap, we incorporated a Bayesian uncertainty estimation that enables risk stratification. In practical terms, we ranked putative lesional clusters in a given patient based on confidence, thus assisting the examiner to gauge the significance of all findings. In 73% of cases the FCD was among the top five clusters with the highest confidence to be lesional; in half of them it ranked the highest. In the remaining 27%, lesions manifested with low confidence; in a real-world scenario, when location is unknown (*i.e.,* no FCD label is available), a concerted evaluation including electro-clinical and other imaging tests is likely to increase diagnostic certainty [255]. While the good performance of our classifier is also attributable to the richness of the training set including a large spectrum of anatomical locations, eleven MRI-negative FCD remained unresolved, with six located in the orbitofrontal cortex, an area for which limited data was available for training. The prospective use of our classifier trained on the entire cohort would likely reclaim these lesions.

The analysis of the spatial distribution of FPs was moderately comparable across FCD patients, healthy and disease controls, mainly involving the insula and parahippocampal region bilaterally. A possible explanation may lie in the similarity of the cytoarchitectonic signature of these cortices with FCD histopathological traits. Notably, the three-layered cortex of the hippocampal formation, the transitional mesocortex of the parahippocampus and the mesocortex-like insula present

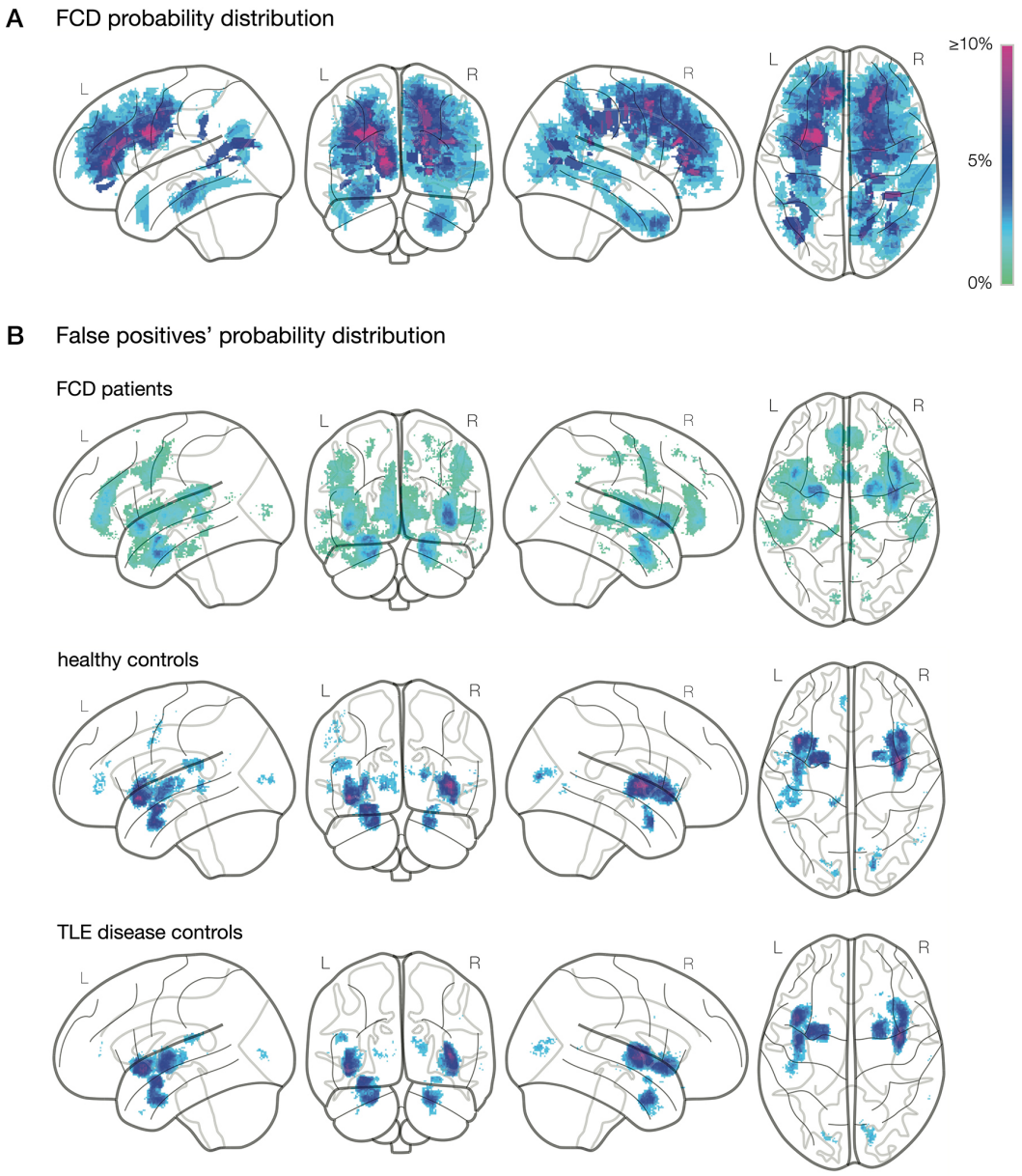


Figure 7.5: **Probability distributions of FCD and false positives**. **A**. Lesional probability maps of manually labelled FCD lesions superimposed on glass brains. **B**. Probability maps of confidence of FP clusters across cohorts. Colors indicate proportions (in %) of lesional (A) and FPs (B) voxels.

with indistinct boundaries between laminas compared to the typical six-layered neocortex [256, 257]; these cortices may thus mimic dyslamination and blurring. Notably, however, our algorithm detected 3/3 FCD lesions in the insula with high degree of confidence. Since these lesions were provided by different sites, the *leave-one-site-out* strategy guaranteed that each training set had at least one lesion. Nonetheless, adding more lesions to the training set would increase the classifier's ability to learn better discriminative features in the insular region. Alternatively, an impact of developmental trajectory [258] on FPs is suggested by high prevalence in younger patients, possibly in relation to age-varying tissue contrast, cortical myelination and maturation, which may also manifest as lesion-like on MRI. Conversely, registration errors are less likely in our voxel-based method as compared to surface-based algorithms. For the latter, to align a subject's brain into a standardized stereotaxic space registration strongly depends on the accuracy of GM/WM segmentation, while our method does not require tissue segmentation. Notably, some FPs were only seen in FCD cases, particularly in fronto-central regions and tended to gather around the lesion, suggesting subthreshold peri-lesional anomalies not included in the manually-segmented label [130, 242]. Given the favorable surgical outcome, a biological explanation for FPs in our FCD cohort may thus include a combination of normal cytoarchitectural nuances and non-epileptogenic peri-lesional developmental anomalies. In a previous study [130], we found FPs to manifest as abnormal sulcal depth, while the FCD lesions had higher cortical thickness relative to controls. Sulcal abnormalities in cortical malformations have been described in the proximity and at a distance of MRI-visible lesions and are thought to result from disruptions of neuronal connectivity and WM organization [130, 230]. Finally, it is also plausible that some FP clusters my represent dysplastic tissue, an entity so far reported only in five cases [259].

While our algorithm was trained on histologically verified FCD-II lesions and is mainly aimed at identifying MRI-negative FCD, it is possible that it could identify difficult-to-detect low grade tumors that may resemble dysplastic lesions, a rare instance occurrence since these tumors are generally easy to see on routine MRI. Regardless, the dilemma of differentiation of FCD from low grade tumors uniquely based on MRI features may arise; the differential diagnosis is then evaluated using additional tests, including MR spectroscopy. On the other hand, our algorithm may be useful in identifying associated often-occult dysplastic lesions in the peritumoral area [260].

Federated machine learning: a path to the future

Traditional machine learning adopts a centralized approach that requires training datasets to be aggregated in a single center. A significant obstacle to clinical adoption of such strategy is privacy and ethical concerns. Federated learning [261], on the other hand, is a distributed approach that enables multi-institutional collaboration without sharing patient data. Our proposed approach of patch-based data augmentation is privacy-preserving since only a portion of each patient's data

is collated and randomized before exposure to the neural network, an implementation that can be flexibly re-configured to support federated learning. As the data corpus diversifies and expands to include more edge cases, performance and confidence of future classifiers will inevitably improve.

8 Automated Hippocampal Subfield Segmentation using Deep Learning

Preface

Hippocampal pathology, a hallmark of numerous neurological conditions, manifests as atrophy localized to a specific subfield or spanning multiple subfields, emphasizing the need for accurate segmentation. The rising demand for large-scale data analysis has motivated the development of automated segmentation algorithms [262, 263]. Accurate segmentation of the hippocampal formation is particularly relevant for effective individualized diagnostics in temporal lobe epilepsy (TLE), as this structure is the hallmark site of pathology. The current implementations force users to undertake a significant technical debt, and are cumbersome to implement, replicate and validate on independent datasets.

The purpose of this study was to develop and validate an automated framework to segment hippocampal subfields using deep learning and evaluate the generalization performance in pathology. We propose *DeepPatch*, a volume-based subfield segmentation method that combines patch-based analysis (similar to the methodology proposed in Chapter 5), which optimizes label fusion and image matching by compactly representing anatomy, shape, texture and intensity, and fully convolutional deep neural networks (CNN) that offers hierarchical feature learning ability.

Manuscript 6

In preparation

Automated Hippocampal Subfield Segmentation using Deep Learning

Ravnoor S. Gill[1], Benoit Caldairou[1], Andrea Bernasconi[1], Neda Bernasconi[1]

[1] Neuroimaging of Epilepsy Laboratory, McConnell Brain Imaging Center, Montreal Neurological Institute, Montreal, QC, Canada

Abstract

Hippocampal pathology, a hallmark of numerous neurological conditions, manifests generally as atrophy localized to a specific or multiple subfields, emphasizing the need for accurate segmentation, both for understanding disease mechanisms and individualized diagnostics. We propose an automated subfield segmentation procedure which combines deep learning with patch-based template library for feature matching. Our algorithm was trained on a dataset of manually segmented labels obtained on submillimetric T1-weighted MRI. Validation experiments showed equally high accuracy in healthy individuals and patients with Dice ranging from 87% to 92% across hippocampal subfields. Segmentations obtained through the automated algorithm showed high performance to lateralize the seizure focus in 95% of cases (91% in MRI-negative), suggesting clinical utility.

8.1 Introduction

Temporal lobe epilepsy (TLE), the most common syndrome in adults, is pathologically defined by varying degrees of neuronal loss and gliosis in the hippocampus and adjacent structures [136]. On MRI, marked hippocampal sclerosis (HS) appears as atrophy and signal hyperintensity, generally more severe ipsilateral to the seizure focus. Accurate identification of hippocampal atrophy as a marker of HS is crucial for deciding the side of surgery. While volumetry has been one of the first computational analyses applied to TLE [25, 131, 139, 264, 265], the need for accurate localization of pathology has motivated a move from whole-structure volumetry to surface-based approaches allowing a precise mapping of anomalies along the hippocampal axis. Surface modeling based on spherical harmonics [147] has been particularly performant [266], as it guarantees precise inter-subject correspondence.

Manual hippocampal volumetry is time-prohibitive and prone to rater bias. These challenges, together with rising demand to study larger patient cohorts, have motivated the shift towards automated segmentation, setting the basis for large-scale clinical use. Initial methods for whole hippocampal segmentation used a single template or deformable models constrained by shape priors obtained from neurotypical individuals [267–270]. More recent approaches rely on multiple templates and label fusion; by selecting a subset of atlases from a template library which best fit the structure to segment, thereby accounting for inter-subject variability, these approaches have provided increased performance over single template approaches [271–274].

Advances in MR acquisition hardware and sequence technology, which enable submillimetric resolution and improved signal-to-noise ratio, have facilitated accurate identification of hippocampal subfields or subregions, including the dentate gyrus, subiculum and the cornu ammonis (CA1-4) regions [151]. Several methods have been developed for MRI-based subfield segmentation [144, 152, 153, 158, 275], providing state-of-the-art results (average Dice ~88%) with fast inference times.

Recently, the widespread adoption of deep learning [180] in medical imaging has promoted a resurgence in volumetric segmentation methods. Unlike contemporary algorithms, deep learning does not require the data to be extensively preprocessed, thus eliminating the need to build template libraries. More specifically, the ability of convolutional neural networks to learn important features from multimodal data during the training process rather than using hand-crafted features has enabled them to outperform traditional approaches, with Dice overlap indices exceeding 90% in both healthy [206, 276] and atrophic [277] hippocampi. Their applications to the task of whole hippocampal and subfield segmentation are rather recent [156, 206, 207, 276–278]. Yet so far studies have been limited in terms of sample size, and absence of validation in independent datasets in health and disease has precluded assessment of generalizability.

We propose *DeepPatch*, a method for hippocampal subfield segmentation that combines label fusion and patch matching with fully convolutional networks [279], leveraging the hierarchical feature learning capabilities of deep neural networks and the efficiency of patch-similarity label fusion [157]. Validation was performed using a publicly available 3T dataset of manual segmentations of CA1-3, CA4-DG and subiculum (SUB) on T1-weighted MRI of healthy individuals [151], and a cohort of TLE patients with histologically verified HS. In addition, we tested the performance of *DeepPatch* on a seizure focus lateralization task.

8.2 Methods

8.2.1 Subjects

The MNI-HISUB25 dataset [151] (https://www.nitrc.org/projects/mni-hisub25) composed of MR images and manually drawn labels (CA1-3, CA4-DG, and SUB) from 25 healthy subjects (31 ± 7 years, 13 females) served as the training and validation dataset. A clinical dataset of 76 TLE patients (39/76 MRI-negative; 35 ± 10 years, 47 females; manually-segmented subfield labels available in 32/76) with identical imaging parameters were used for validation of hippocampal subfield segmentation, and TLE seizure focus lateralization (Table 8.1)

8.2.2 MRI acquisition and image preparation

MRI data were acquired on a 3 Tesla Siemens TimTrio scanner using a 32-channel head coil. We obtained a submillimetric T1-weighted 3D magnetization-prepared rapid-acquisition gradient echo (MPRAGE) [repetition time (TR): 3000 ms; echo time (TE): 4.32 ms; inversion time (TI): 1500 ms; flip angle: 7°; matrix size: 336×384; FOV: 210×229 mm^2; 0.6×0.6×0.6 mm^3 isotropic resolution; to increase the signal-to-noise ratio, two identical scans were acquired, motion corrected, and averaged into a single volume]. T2-weighted images were obtained using a 2D turbo spin-echo sequence (TR: 10 810 ms; TE: 81 ms; flip angle: 119°; matrix size: 512×512; FOV: 203×203 mm^2, 60 coronal slices angled perpendicular to the hippocampal long axis, slice thickness: 2 mm, resulting in 0.4×0.4×2.0 mm^3 anisotropic resolution). Images underwent automated correction for intensity non-uniformity [115] and intensity standardization. T1-weighted scans were resampled to 0.4×0.4×0.4 mm^3 resolution in the MNI152 stereotaxic space. To augment the sample size, right hippocampi were flipped to the left.

8.2.3 Classifier design

DeepPatch matches patches from an atlas- to a training-set based on a similarity function [280]. An atlas patch is the combination of the intensity patch and its corresponding label patch. All

Table 8.1: Validation cohort characteristics

	Age	Female (%)	SEEG (%)	TLE (L/R)	HS/G	Engel Class I (%)
MRI-positive (n = 37)	37 ± 11	51%	14%	19/18	36/0	94%
MRI-negative (n = 39)	34± 10	72%	67%	23/16	6/21	63%

*SEEG: percentage of patients that underwent surgery. *Abbreviations*: HS/G = hippocampal sclerosis/isolated gliosis; SEEG = stereoencephalography; TLE = temporal lobe epilepsy. L (left)/R (right) refers to EEG lateralization.

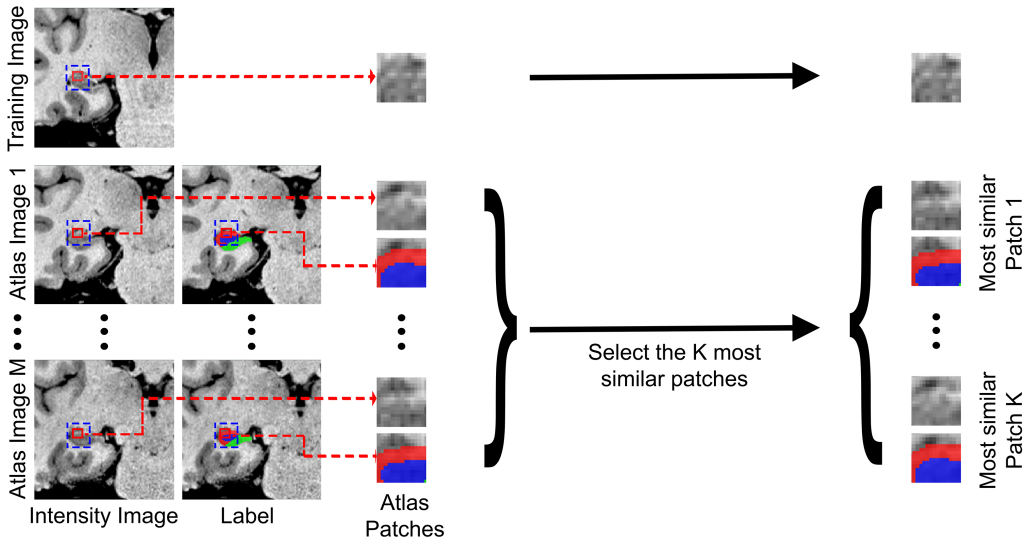


Figure 8.1: **Atlas patch selection**. For each training subfield patch, we search for the corresponding best similar patch in each atlas within a search neighbourhood (blue dashed lines). These atlas patches are ranked, and the most similar K (based on Equation 8.1) are fed to the neural network along with the training patch. The number of atlas patches is treated as a hyperparameter.

atlas patches are concatenated to their corresponding training patches that serve as inputs to the neural network.

Atlas patch selection

Because of the variability in label sizes, we implemented a strategy to balance the number of patches extracted across all labels. Specifically, we ensure that an adequate number of patches is extracted from the subfield boundaries, since these regions provide the most discriminative shape information. To this end, a canny edge filter was applied to the multi-class label map to detect boundary voxels. Patches were then randomly sampled, with equal numbers for each subfield label. In addition, we randomly sample voxels from the inner part of each label. The ratio of patches between the border and inner voxels was treated as a hyperparameter. For each training patch, P_T^j centered at voxel j, extracted from the training image I_T, we can find the most similar patch from each atlas image according to intensity similarity within a cubic search window N^j centered around voxel j. For each atlas A, the most similar atlas patch is defined as:

$$\widehat{P_A^j} = \left\{ P_A^n \mid min_{n \in N^j} \left\| P_T^j - P_A^n \right\|_2^2 \right\} \tag{8.1}$$

Finally, all selected atlas patches are ranked so that we can choose the K patches most similar to the training patch. Subsequently, the training patch and its most similar patches are jointly fed to the neural network for training (see Figure 8.1).

Network architecture

The proposed fully convolutional neural network (Figure 8.2) comprises three streams based on an encoder-decoder architecture: i) atlas-unique; ii) target-patch; and iii) atlas-fusion. Specifically, each candidate atlas patch (intensity and label) is concatenated to the training patch (intensity only) and propagated through the atlas-unique stream. The target-patch stream propagates the input training patch. There can be as many atlas-unique streams as the selected atlases (Figure 8.3 illustrates only two atlas-unique streams). The target-patch and atlas-unique streams are based on a UNet [281] architecture adapted for end-to-end (*i.e.,* the input and the output dimensions match) image segmentation. The downstream encoding and upstream decoding paths consist of repeating blocks separated by down-sampling and up-sampling, respectively. Finally, the atlas-fusion stream includes a series of connections linking different spatial scales of information that propagates from the atlas-unique to the target-patch streams; it concatenates feature maps generated at specific levels of the network, followed by a convolution layer with a 1×1×1 kernel to fuse them. ReLu [282] introduces nonlinearity; batch normalization provides regularization to

prevent overfitting while also stabilizing the training by averting vanishing gradients and reducing the internal covariate shift [283].

Concurrent spatial and channel squeeze-and-excite (scSE) module. Convolution layer serves as the basic building computational block for all CNN architectures, *i.e.,* hierarchically learning filters that capture local spatial pattern along all the input modalities (or channels) and generate feature maps jointly encoding the spatial and channel information. While this joint encoding of spatial and channel information has garnered more research attention, encoding of the spatial and channel-wise patterns independently has been only marginally explored. An architectural component called squeeze excitation (SE) block [284] attempted to address this issue by explicitly modeling the interdependencies between the channels of feature maps. Its nomenclature is motivated by the fact that the SE block 'squeezes' along the spatial dimension and 'excites' or reweights along the channel (or modality) dimension. *scSE* [285] combines the outputs following recalibration of the feature maps independently along channel and space, explicitly optimizing feature maps to be more informative both spatially and channel-wise. *scSE* blocks were inserted after every encoder and decoder.

Localizer and subfield network

The localizer network is an atlas-free fully convolutional UNet corresponding to the target-patch stream as shown in Figure 8.2. The network weights are trained independently of the subsequent subfield segmentation. The network uses the whole image as input rather than patches for training and testing. The output of the localizer network provides global hippocampal segmentation, which serves as the basis for sampling multi-label subfield patches for the downstream subfield network. The latter comprises three networks as described in section 8.2.3.

Loss function and network optimization

To mitigate the class imbalance issue (significant proportion of the image voxels do not represent the structure of interest), previous work [207, 281] applied a differential weighting to the categorical or multi-class cross-entropy loss function. The large class imbalance, however, overwhelms the cross-entropy loss [286]. To account for imbalance due to size differences across hippocampal subfields, we applied similar weighting (inversely proportion to the subfield size) – smaller subfields are assigned a higher weight. Therefore, we use the focal loss (FL) [286] objective func-

tion that in addition to size-based weighting, also down-weights easy examples and thus focusing training on hard negatives to boost segmentation accuracy of the smallest structures.

$$CE(p) = -\log(p(l)) \tag{8.2}$$

$$FL(p) = (1 - p(l))^{\gamma} \cdot CE \tag{8.3}$$

where CE is the cross-entropy loss, and FL is expressed as a function of CE is the number of labels to obtain, l is the label being considered (0 is background, 1 is CA1-3, 2 is SUB, and 3 is CA4-DG) and $p(l)$ is the probability that a voxel is labeled as l. The focusing parameter γ attributes a stronger penalty to labeling errors compared to the cross-entropy loss function. CE loss is, however, used for segmenting the whole hippocampus using the localizer network. Kaiming Uniform function [287] was used to initialize network weights, subsequently optimized using Adam (learning rate = 10^{-4}) [288].

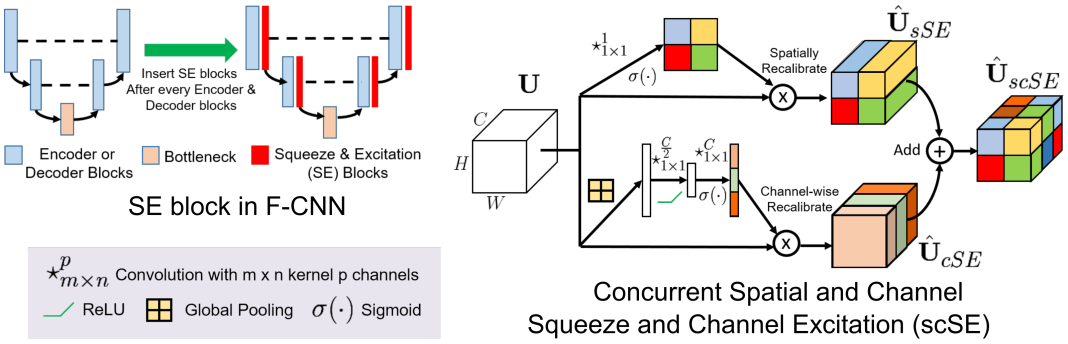


Figure 8.2: **Illustration of network architecture with squeeze & excitation (SE) blocks**. (a) The integration of SE blocks within F-CNN. (b) The architectural design of the integrated *scSE* blocks, for recalibrating the feature map U [Adapted with permission from 285].

.

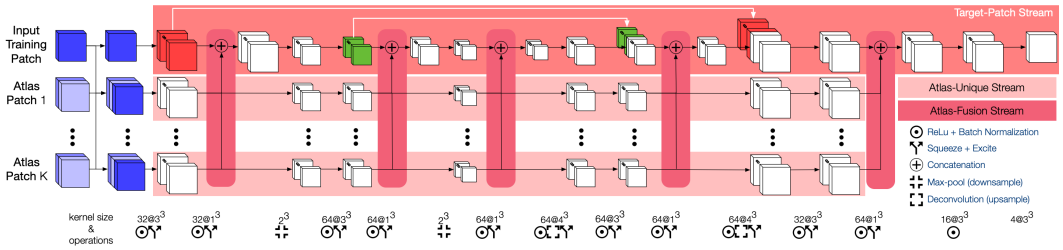


Figure 8.3: **Network construction**. Network architectures and flow of information across different network streams. Each convolutional and deconvolutional layer is followed by a ReLu and a batch normalization unit. ⊕ represents a concatenation of features maps from the target-patch and the atlas-unique streams, followed by a 1×1×1 convolution to fuse these features.

Segmentation inference and refinement

Each testing image is first fed to the localizer network, which provides a coarse global hippocampal segmentation. Then, we extract one patch per voxel and select its K most similar atlas patches as described in the Atlas patch selection. Next, these patches are combined and serve as inputs to the target-patch stream. The results from all testing patches are fused through majority voting to produce the multi-label segmentation. As fusion may introduce small and isolated segmentation errors, we extracted the largest connected component from the binarized network output and intersected it with the original output to obtain the final subfield segmentation.

Network and hyperparameter optimization

Network weight optimization was performed on submillimetric images. Focusing parameter was empirically set to $\gamma = 4$. To accelerate the segmentation during testing, instead of dense sampling with significant overlap between patches, we sampled patches sparsely, at every 100 voxels. This resulted in patches with partial overlap without significant loss of information.

To determine the optimal network configuration, we tested the following domain-informed hyperparameters: i) patch size (32×32×32, and 40×40×40), ii) atlas paths (3, 4, and 5), iii) border-to-inner patch sampling ratio (for differential boundary sampling) with minimum number of inner patches held constant at 20 patches per label per hippocampus (4:1).

Performance evaluation

Performance was evaluated using Dice Similarity Coefficient (DSC) between two labels S_1 and S_2, defined as: $2 \cdot |S_1 \cap S_2| / |S_1 \cup S_2|$, and Bland-Altman plots. Cross-validation was conducted in 25 healthy controls using a *four*-fold scheme; specifically, 25% of cases were used for testing, 37.5% for training and the remaining 37.5% as atlas. We also validated the algorithm on 76 TLE patients with the same MRI parameters – the algorithm was trained on 25 healthy controls (80% training, 20% atlas).

TLE lateralization

We assessed the clinical utility of the current approach using a lateralization task in TLE patients that assessed the accuracy of a linear discriminant analysis (LDA) classifier to lateralize the seizure focus using [158, 289] in individual patients. Briefly, volumetric subfield labels generated using *DeepPatch* are converted to surface meshes and parameterized using spherical harmonics and a point distribution model (SPHARM-PDM) [147] that guarantees correspondence of surface points or vertices across subjects. A medial surface sheet running along the central path of each subfield allowed for vertex-wise sampling of columnar volume and normalized T2 and

FLAIR/T1w signal intensity. These features individually and in combination across each subfield are used as inputs to an LDA classifier. Cross-validation was performed using a *five*-fold scheme, repeated 100 times. Lateralization performance was compared against subfield segmentations generated using *SurfPatch*.

8.2.4 Standard protocol approvals, registrations, and patient consent

The Ethics Committee of the MNI approved the study and written informed consent was obtained from all participants.

8.3 Results

Subfield segmentation

In healthy controls, the average overlap between manual and automated labels was 91.89% ± 1.22 for CA1-3, 87.88% ± 2.19 for CA4-DG and 88.87% ± 2.01 for SUB. Similar high performance was obtained in TLE patients, with an overlap of 90.66% ± 2.32 (CA1-3), 86.63% ± 4.56 (CA4-DG), and 87.82% ± 2.53 (SUB) (see Table 8.2). High correlations and small differences between automated and manual volumes as shown in the Bland-Altman plots in both groups also supported the robustness of the algorithm (Figure 8.4). Representative subfield segmentation examples are shown in Figure 8.5.

Seizure focus lateralization in TLE

The subfield segmentations generated using *DeepPatch* were used to predict to lateralize the side of the seizure focus in TLE. The overall lateralization accuracy based on individual features was significantly better for T2 (90.83 ± 2.22%; $p < 0.05$) than FLAIR/T1w intensity (85.67 ± 2.12%), which were superior to that of volumetry (69.72 ± 3.25%; $p < 0.05$). The combination of T2 and FLAIR/T1w yielded the best overall performance of 95.40 ± 1.53% ($p < 0.05$), with an accuracy of 91.46 ± 2.73% in MRI-negative and 99.54 ± 1.09% in MRI-positive TLE. The lateralization performance was comparable to subfield segmentations generated using *SurfPatch*. Refer Table 8.3 for details.

8.4 Discussion

We propose *DeepPatch* , a volume-based subfield segmentation framework that combines patch-based atlas matching with fully convolutional neural networks. The algorithm operates on T1-weighted images alone, the most frequently used anatomical contrast in major population-level

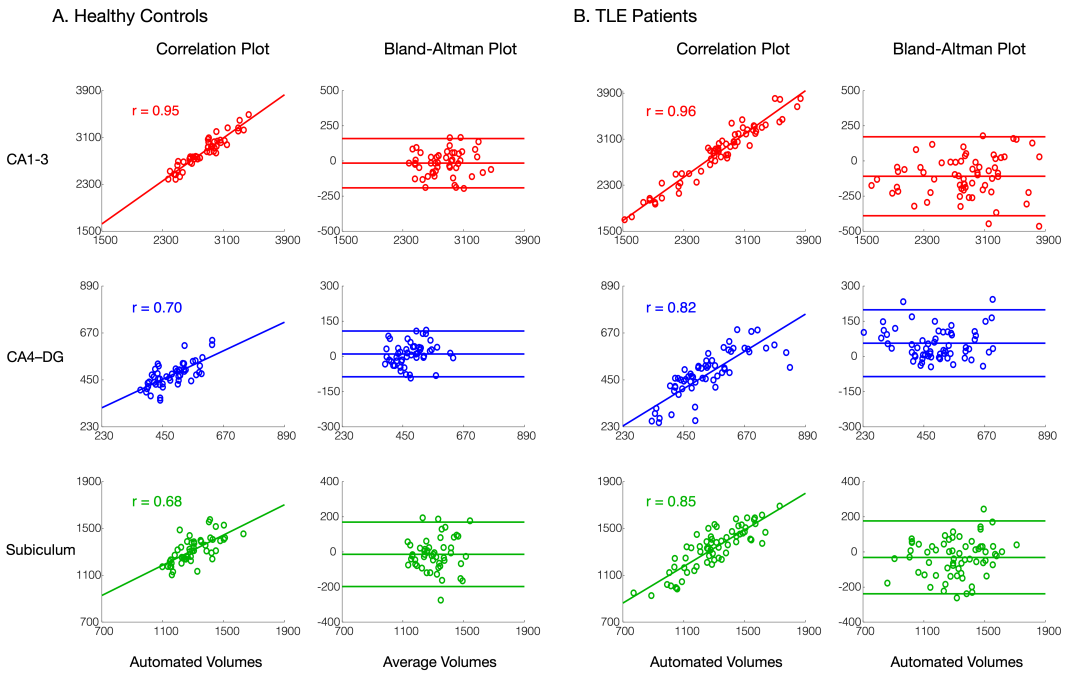


Figure 8.4: High correlation and small differences between automated and manual volumes (in mm^3) as shown by the Bland-Altman plots in both controls (A) and patients (B) supports the robustness of the algorithm.

Table 8.2: Performance comparison between *DeepPatch* and surface- and volume-based segmentation methods validated with the same controls dataset [151] for each subfield and their average.

	% Dice mean (standard deviation)			
	CA1-3	**CA4-DG**	**SUB**	**Average**
Method				
DeepPatch (C)	**91.89 (1.22)**	**87.88 (2.19)**	**88.87 (2.01)**	**89.55 (2.51)**
DeepPatch (TLE)	90.66 (2.32)	86.63 (4.56)	87.82 (2.53)	88.37 (3.67)
DeepHIPS (C) [206]	**92.45 (1.06)**	**88.87 (2.37)**	**89.80 (1.55)**	**90.37 (1.29)**
HIPS (C) (T1w+T2w) [156]	91.15 (1.44)	86.16 (3.39)	87.46 (2.36)	88.26 (2.59)
Manjón *et al.* (C) [278]	90.01 (1.09)	84.04 (1.57)	86.78 (2.96)	86.95 (N/A)
SurfPatch (C) [158]	87.43 (2.47)	82.71 (2.85)	84.95 (2.45)	N/A
Δ_{mean} (Δ_{std})	**−0.56(+0.16)**	**−0.99(−0.18)**	**−0.93(+0.46)**	**−0.82(+1.22)**

C: controls; N/A: not available; T1w/T2w: T1-/T2-weighted contrast; Δ: difference in mean and standard deviation (std) between *DeepPatch* and *DeepHIPS* [206] in controls.

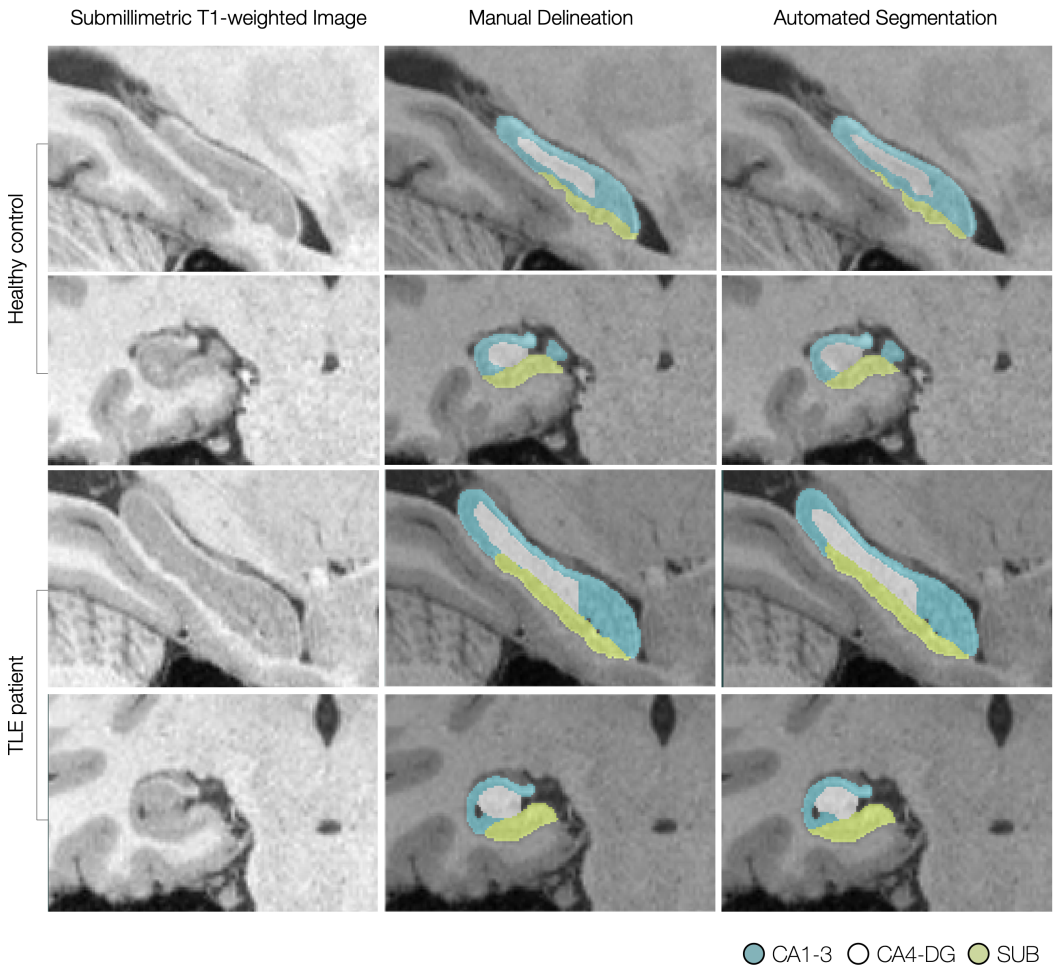


Figure 8.5: Manual delineation and *DeepPatch* automated segmentations relying on submillimetric T1-weighted MRI for a representative healthy control (CA1-3: 91.3%, CA4-DG: 85.2%, SUB: 88.4%) and a TLE patient (93.6%, 90.7%, 90.5%). Note that in the patient, despite the presence of hippocampal malrotation, subfields are adequately segmented.

Table 8.3: TLE lateralization accuracy using *DeepPatch* and *SurfPatch*

	Volume	**T2 Signal**	**FLAIR/T1w**	**T2 + FLAIR/T1w**
DeepPatch				
All patients (n = 76)	69.72 ± 3.25	90.83 ± 2.22 *	85.67 ± 2.12 *	**95.40 ± 1.53 ***
MRI-positive (w/HS; n = 37)	94.89 ± 2.82	99.70 ± 0.85	95.38 ± 1.81	99.54 ± 1.09 *
MRI-negative (n = 39)	45.85 ± 5.65	82.41 ± 4.14 *	76.46 ± 3.97 *	91.46 ± 2.73 *
SurfPatch				
All patients (n = 76)	76.02 ± 2.96	89.08 ± 2.00 *	90.75 ± 2.29	**95.78 ± 1.45 ***
MRI-positive (w/HS; n = 37)	97.32 ± 2.89 *	99.14 ± 1.48	95.32 ± 1.27	98.76 ± 1.35 *
MRI-negative (n = 39)	55.82 ± 4.54	79.54 ± 3.57 *	86.41 ± 4.26 *	92.97 ± 2.35 *

* Increased/best lateralization AUC with respect to at least 1/any model (*Friedman's* $p < 0.05$ corrected for multiple comparisons). HS: hippocampal sclerosis.

MRI initiatives, such as the UK BioBank and Human Connectome Project. Our algorithm outperformed state-of-the-art implementations [156, 158, 206, 278], with remarkable performance across submillimetric images both in health and disease. Importantly, the use of the same publicly available dataset [151] provides confidence in comparing performance across studies.

In healthy controls, segmentation performance was excellent, with Dice overlap indices ranging from 87% to 92% across subfields. With respect to *HIPS*, a platform operating on T1-weighted MRI alone or combined with T2-weighted MRI, our algorithm yielded higher Dice indices with consistently lower variance across all subfields. Notably, T2-weighted contrast is prone to motion artifacts [290], which may be prevalent in patients., which may explain why it underperformed *DeepPatch* despite using an additional contrast. However, *DeepPatch* was unable to match the performance of *DeepHIPS*. Although, the difference in performance was less than 1%, it can be partially explained due to the differential composition of controls in the splits across the training and testing sets, and partially due to the denoising preprocessing implemented. More importantly, the observed variance in performance across implementations tested is generally ~1%. While the rationale for this phenomenon is unclear, we speculate that there could be a ceiling in terms of Dice, attributable to the label noise (errors in ground truth labels). This limitation probably makes Dice or any other overlap metric unsuitable to compare performance, at least in the clinical context where ~1% difference is unlikely to meaningfully impact diagnosis.

The high performance of *DeepPatch* may be attributed to patch-based matching that offers intrinsic modeling of multi-scale intensity features, coupled with sampling heuristics (border-to-inner ratio) to implicitly learn the transitions along the label boundaries. In addition, implementation of weighted focal loss addressed data imbalance, further improving segmentation quality. Due to the subjective trial and error nature of hyperparameter optimization only a limited num-

ber of configurations could be tested; a more exhaustive search may improve performance even further. Notably, *DeepPatch* performance was high despite minimal postprocessing. The combination of the patch-based framework with hierarchical feature learning capacity of deep neural nets offers flexibility in capturing complex shape deformations and displacements, which are particularly prevalent in disease. Indeed, we demonstrate robust performance not only on normative data but also in TLE, underscoring the potential utility of *DeepPatch* in clinical decision making.

In a recent work, we combined automatically-generated hippocampal surface labels using *SurfPatch* with a surface-based classifier for seizure focus lateralization [289]. In this work, as a first step towards a more simplified approach, we replaced *SurfPatch* with *DeepPatch* , a volume-based segmentation. *DeepPatch* algorithm matched the clinical accuracy of *SurfPatch* generated segmentations, but without the additional complexity involved in generating those segmentations. A planned future implementation will integrate the lateralization and segmentation within a single multi-task deep neural network framework [291, 292]. Such an implementation will generate both the subfield segmentations and lateralization (*left* or *right* TLE) labels for a patient, rather than rely on an independent classification paradigm.

Part III

Conclusions

9 Key Findings and Significance

This dissertation describes a series of studies combining a review (systematic review and meta-analysis) with advanced multimodal imaging leveraging MRI-based morphometric and volumetric features with recent advances in machine-learning. The review aimed at demonstrating biases and variability in the definition for MRI-negative epilepsies and emphasizing the value of MRI post-processing in the presurgical evaluation of epilepsy patients. In subsequent studies, we employed bipartite approaches to detecting predominantly MRI-negative epileptogenic lesions using surface-based morphometry and Bayesian deep learning based on minimal preprocessed volumetric MRI to leverage uncertainty in predictions, while also evaluating their comparative yield and establishing clinical utility.

Project 1 (Chapter 3). This project aimed at synthesizing evidence assessing ambiguity in criteria used to define MRI-negative in the context of a systematic review and meta-analysis. Indeed, we show that ascribing the MRI-negative categorization lacks objectivity. This subjectivity manifested as significant associations between modality-derived clusters and several outcomes (MRI reporting of contrasts, parameters, rater expertise, quantitation). Importantly, invasive diagnostics (*i.e.,* intracranial EEG) underutilize MRI both in terms of acquisition protocols and quantitative post-processing. Furthermore, meta-analytic evidence showed seizure freedom in 75% of MRI-positive cohorts relative to 59% in MRI-negative. Finally, we provide first meta-analytic evidence that MRI postprocessing is associated with two-fold gain in diagnostic yield over expert review of MRI alone.

Project 2 (Chapter 4). This project implemented a multivariate surface-based detection framework to detect subtle FCD using T1-weighted, FLAIR and FLAIR/T1w ratio images using statistical machine-learning techniques, and to evaluate its diagnostic yield relative to conventional visual inspection. Features were systematically sampled at multiple cortical levels from multiple contrasts and fed into a two-stage classifier for automated lesion detection based on ensemble learning with intrinsic mitigation for class imbalance. By implementing such an automated classifier relying on FCD morphology and intensity features, we could accurately identify subtle lesions initially overlooked on routine radiological inspection. Our approach showed excellent sensitivity (83%, 34/41 lesions detected) and specificity (92%, no findings in 35/38 controls) to discriminate

true malformative lesion from healthy cortices, which is expected to benefit presurgical diagnostics.

Project 3 (Chapters 5, 6 and 7). The intrinsic limitations of voxel-based approaches that motivated the shift to surface-based morphometry paradigms can be circumvented by deep learning methods using minimally preprocessed volumetric data. In the same vein, first, we designed a neural network that leverages multimodal MRI to harness the diagnostic value of T1-weighted MPRAGE and T2-weighted FLAIR contrasts to detect FCD lesions in predominantly MRI-negative cohorts. Second, we used approximate Bayesian variational inference to sample the model's epistemic uncertainty and leveraged it to assign a ranking to triage multiple putative lesional candidates. Finally, we developed and validated an automated framework to detect the subtle FCD lesions based on Bayesian deep learning and evaluated its generalization performance across nine tertiary epilepsy centers worldwide with diverse cohorts, scanners, and field-strengths. Consequently, we provide Class III evidence that deep learning on multimodal MRI accurately identifies FCD in epilepsy patients initially diagnosed as MRI-negative.

Project 4 (Chapter 8). Establishing the utility of patch-based sampling with neural networks in Project 3, we leveraged a similar strategy coupled with label fusion to develop a hippocampal subfield segmentation algorithm (*DeepPatch*). We implemented patch-based matching to intrinsically model multi-scale intensity features, coupled with sampling heuristics to implicitly learn the transitions along the label boundaries, and mitigate class imbalance issues using focal loss. *DeepPatch* was trained on open-source data of healthy controls with expert manual labels, and subsequently validated on a TLE cohort. We show that the performance is comparable to state-of-the-art in the field, while also demonstrating the clinical utility of the segmentations in a seizure focus lateralization task in predominantly MRI-negative TLE patients.

Significance. The role of MRI in the diagnosis of epilepsy and presurgical evaluation is undisputed. However, despite repeated recommendations and guidelines, adoption of best practice MRI is still variable, with many advances not fully transferred into clinical care [167]. The most dramatic implication of this translational gap relates to lesion identification, with the risk that MRI-positive patients may be mislabeled as MRI-negative. This limitation has resulted in the disproportionate use of invasive EEG studies, with the assumption that electro-clinical hypotheses alone may be sufficient to identify the epileptogenic zone [293]. Paradoxically, while appending MRI-negative status has profound implications in terms of treatment strategies and seizure outcome, our systematic review showed that this categorization has lacked objectivity. Furthermore,

data-driven unsupervised machine learning and stratified studies based on diagnostic modalities identified three distinct classes demonstrating the overall methodological variability and weaknesses of MRI diagnostic criteria. While the consequences of inadequate data reporting are unclear, the current findings make a strong case to standardize reporting of not just MRI relevant information, but all modalities used in the clinical workflow to ascribe the MRI-negative status. The meta-analysis on post-surgical outcome confirmed the results of the narrative synthesis, namely that seizure freedom after surgery is more often achieved in MRI-positive compared to MRI-negative patients. Notably, the virtually stable proportion of MRI-negative with unfavorable outcome despite constant improvements in MRI technology and analytical tools over the last decade demonstrates a translational gap and should be taken as an incentive to increase educational efforts promoting neuroimaging skills, the availability of standardized MRI acquisition protocols and guidelines and a cultural change in the epilepsy community. In addition, large-scale efforts within the open-access and neuroscience community are required to facilitate and democratize access to image analyses techniques. Upskilling and strengthening the core competences of epileptologists will undoubtedly transform traditional clinical decision-making into a modern, systematic, multidisciplinary approach, ultimately circumventing the use of invasive diagnostic methods. This argument is supported by the meta-analysis on diagnostic yield showing that MRI post-processing is twice as likely to reverse the MRI-negative status relative to the expert visual analysis. This gain in diagnostic yield is largely driven by the ability of MRI post-processing to unveil subtle or subthreshold anomalies inconspicuous to the unaided human eye.

While appending MRI-negative status has profound implications in terms of treatment strategies and seizure outcome, our systematic review showed that this categorization has lacked objectivity. Consequently, we directed our efforts to develop algorithms to identify lesions in MRI-negative focal cortical dysplasia. Core to our approach was a surface-based integration of morphological markers as well as intensity and textural features derived from co-registered T1w and FLAIR data as well as their ratio. We implemented a non-parametric boosted decision tree ensemble that integrates random-undersampling (to circumvent class imbalance) with adaptive boosting [236, 237] that can capture complex decision boundaries while avoiding overfitting to classify FCD lesions. This allowed our automated algorithm to provide a 4 times higher detection rate than conventional radiological visual inspection. Regrettably, performance comparisons with competing implementations referenced in the study were impeded due to either the source code or the data not publicly available. In addition, owing to the considerable time investment to manually correct

brain tissue segmentation and surface extraction errors, an alternate approach that is both time-effective and superior in performance was proposed in the subsequent project.

To circumvent the limitation of surface-based methods, we developed the first deep learning-based method to segment FCD that incorporates uncertainty for clinical decision-making, and multicentric validation. The classifier uses T1-weighted MPRAGE and T2-weighted FLAIR, contrasts available on most recent MR scanners [167] and widely available, operates in 3D voxel space without laborious pre-processing and feature extraction. We used a novel leave-one-site-out cross-validation to simulate a real-world scenario, while also testing the algorithm on an independent cohort. Furthermore, by relying on subject-wise data normalization, rather than group-wise, our implementation obviates the need for a matched normative dataset, an expensive and time-intensive undertaking. Regarding specificity, the number of false positive findings in healthy and disease controls were rather modest, while attaining performance parity in adults and children. With respect to the latter, the use of age-appropriate templates considering the developmental trajectories is likely to have contributed to the excellent performance. In addition, our method relies on linear registration of voxel-based data to align a subject's brain into a standardized stereotaxic space to correct for location, orientation, and overall size of the brain without requiring GM/WM segmentation. However, even this requirement can be rendered obsolete altogether in future iterations of the algorithm, by using the MRI in their native stereotaxic space, since intra-subject registration is more critical than the inter-subject to detect colocalizing anomalies in multimodal contrasts. Hence, further limiting the amount of preprocessing required.

To address the technical and logistic challenges required for feature engineering and processing steps involved in the analysis of surface-based data, we have taken steps to ensure wide accessibility by minimizing the technical burden on the clinicians and other end-users by simplifying the workflow from installation to retrieving predictions. This entails integrating all the steps into a streamlined preprocessing routine (including skull stripping, inhomogeneity correction, intensity normalization, and modality co-registration). Notably, the algorithm and all its dependencies, built using open-source software, are containerized to be installed and executed using a single command. This minimizes software installation and prevents licensing issues. Essentially, users are required to only provide the native 3D T1-weighted MPRAGE and T2-weighted FLAIR images to the algorithm. Future iterations of the project will provide a web interface (`https://github.com/NOEL-MNI/noelTexturesPy`) to upload the image pair for each patient, further reducing the friction to obtaining predictions from our algorithm. Taken together, such characteristics and performance promise potential for broad clinical translation.

Compared to previous deep learning methods [190, 195, 196] in which clinical description was scarce to absent, and information on the FCD expert labels and histological validation of lesions was not provided, our study relied on best-practice multimodal MRI, histologically-validated le-

sions, and a large heterogenous dataset. Critically, none of the implementations are open source, thus preventing any performance comparisons.

Federated learning [261], a distributed approach that decentralizes data for multi-institutional collaboration, coupled with human-in-the-loop approach can expedite the labeling of difficult or novel cases that the algorithm has previously not encountered. This iterative procedure of adding more lesions would increase the classifier's ability to learn better discriminative features for lesions under-represented or missing in the training dataset. As the data corpus diversifies and expands to include more edge cases, performance and confidence of future classifiers will inevitably improve. Notably, this hybrid approach lays the groundwork necessary to facilitate a multi-institutional prospective validation of this framework.

We developed *DeepPatch*, a volume-based subfield segmentation framework that combines patch-based atlas matching with fully convolutional neural networks. The algorithm operates on T1-weighted images alone, the most frequently used anatomical contrast in major population-level MRI initiatives, such as the UK BioBank and Human Connectome Project. Our algorithm demonstrated robust performance not only on normative data (publicly available dataset [151]) but also in TLE, underscoring the potential utility of *DeepPatch* in clinical decision making. *DeepPatch*, however, was unable to match the performance of *DeepHIPS*. Although, the difference in performance was less than 1%, it can be partially explained due to the differential composition of controls in the splits across the training and testing sets, and partially due to the denoising preprocessing implemented. However, this can be circumvented by explicitly specifying the split composition, as prescribed by the MIT-BIH arrythmia dataset [294] used widely in electrocardiology. We also observed that the variance in performance across implementations was ~1%. The underlying reason is unclear, but we speculate that there cfould be a ceiling effect in terms of Dice – attributable to the label noise (errors in ground truth labels). This limitation probably makes Dice (or other overlap metrics) unsuitable to compare performance, at least in the clinical context where ~1% difference is unlikely to meaningfully impact diagnosis. Such overlap metrics merely probe the global segmentation rather than inform areas predictive of pathology, with large penalties ascribed to errors in smaller and compact structures like the hippocampal subfields. Moreover, the richness of data in terms of locoregional textures and shape complexity of the subfields and their immediate vicinity is not adequately captured. In the clinical context, where making a diagnosis is prime, testing performance on a downstream performance on clinically validated task is thus, more appealing and pragmatic.

In a recent work, we combined automatically-generated hippocampal surface labels using *SurfPatch* with a surface-based classifier for seizure focus lateralization [289]. In this work, as a first step towards a more simplified approach, we replaced SurfPatch with *DeepPatch*, a volume-based segmentation. *DeepPatch* algorithm matched the clinical accuracy of *SurfPatch* generated seg-

mentations, but without the additional complexity involved in generating those segmentations. A planned future implementation will streamline the workflow by integrating the lateralization and segmentation within a single multi-task deep neural network framework [291, 292]. Such an implementation will generate both the subfield segmentations and lateralization (left or right TLE) labels for a patient, rather than rely on an independent classification paradigm. Moreover, we will offload the trivial task of global hippocampal segmentation to existing frameworks (*e.g.,* nnU-Net [295], HippMapp3r [277], ANTsPyNet [296], etc.) to make the source code more maintainable and accessible, while also significantly reducing duplication of efforts. This also ensures that more rigorous baseline (*e.g.,* ablation-type experiments) comparisons can be easily incorporated into the study design. Hence, this substantial technical debt reduction would pave way for a broader clinical adoption.

✦

In summary, our quantitative evaluation emphasizes the power of MRI post-processing to extract critical information from high dimensional imaging datasets. The proposed methods together with our findings provide avenues to clinically improve lesion detection and better characterization of MRI-negative epilepsy to facilitate improved clinical prognostic and diagnostic description of this challenging patient cohort.

✦✦✦

Part IV

Appendices

A Supplementary Data for Chapter 3

Unsupervised clustering

The purpose of clustering was to identify the co-occurrence of diagnostic modalities, namely studies with similar multimodal diagnostic profile. This procedure was designed to identify groups (or clusters) with higher *intra*-group similarity than *inter*-group similarity, without any prior knowledge about their composition. For each study, we binarized MRI (1.5T, 3T, 7T, T1w, T2w/FLAIR, IR, DWI) and non-MRI (MEG, PET, SPECT, SEEG, EEG-fMRI, and spectroscopy) diagnostic modalities: 1 if certain evaluation was done, 0 otherwise. This data, assembled in a binary matrix, served as input to the clustering algorithm. Since intracranial stereo EEG (SEEG) was an outcome of interest, it was dropped from further analysis to prevent confounding. This resulted in a binary data matrix (A) and computed a dot product to obtain a transformed modality matrix ($X : A_T \times A$, where A_T is a transpose of A, Equation A.1). Each element in the matrix X represents the number of studies that were common to any two modalities.

$$X = AA_T = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1C} \\ x_{21} & x_{22} & \cdots & x_{2C} \\ \vdots & \vdots & \vdots & \vdots \\ x_{R1} & x_{R2} & \cdots & x_{RC} \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{R1} \\ x_{21} & x_{22} & \cdots & x_{R2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1C} & x_{2C} & \cdots & x_{RC} \end{bmatrix} \tag{A.1}$$

We then leveraged spectral clustering [297] using the binary matrix X as the input. The final clustering solution was obtained by taking the majority vote of the class assignment across 1000 randomly initialized 2 bootstrap subsamples, resulting in each of 196 studies being assigned to given clusters. The optimal number of clusters (k=3) was determined using silhouette analysis [298]. *Adjusted Rand index* was used to quantify the consistency of cluster assignment across tandem bootstrap samples.

Table A.1: GRADE evidence for association of MRI-negative status with Engel-I outcome

	Study design (N) = Observational (70 studies)	**Starting score of 2.0**
Risk of bias	Low variability in patient inclusion criteria since drug-resistant patients who were candidates for presurgical evaluation were included, towards an objective outcome (surgical freedom: Engel-I)	No change
Inconsistency	While studies are relatively consistent in direction with the magnitude of point estimates, there was substantial heterogeneity in one of the subgroups (MRI-positive)	−0.5
Indirectness	Consistent outcome timeframe (surgical freedom: Engel-I, > 1-year follow-up) used across both subgroups	No change
Imprecision	Confidence intervals around the pooled estimate are relatively narrow [0.63, (0.59–0.67)]	No change
Effect size	Moderate effect size estimate (Proportion, 0.63)	No change
Publication bias	Not serious, as indicated by funnel plots. Small studies with small effect size seem to be missing. Large to moderate number of studies across both subgroups are present	−0.5
Overall score:	**Moderate to Low**	**1.0**

Table A.2: GRADE evidence for association of type of MRI analysis procedure on diagnostic yield

	Study design (N) = Observational (40 studies)	**Starting score of 2.0**
Risk of bias	Reporting bias controlled for since only studies with repeated measures design (pre- and post- intervention estimates in the same studies) were included.	No change
Inconsistency	Low to moderate statistical heterogeneity was present.	No change
Indirectness	Repeated measures inference to decipher the direct effect of intervention	No change
Imprecision	Confidence intervals around the pooled estimate are relatively narrow [8.66, (6.17–12.16)]	No change
Effect size	Moderate to large (Odds Ratio, 8.66)	+0.5
Publication bias	Undetected, as indicated by funnel plots	No change
Overall score:	**Moderate**	**2.5**

Table A.3: GRADE evidence profile: MRI-diagnostic status with post-surgical seizure freedom

	Number of patients			
No of studies (Design)	**Engel-I**	**Total**	**Proportion (95% CI)**	**Quality**
MRI-negative				
57 (Obs)	995	1773	59% (55–63%)	+++ moderate
MRI-positive				
13 (Obs)	848	1235	75% (67–84%)	++ low

Table A.4: GRADE evidence profile: MRI intervention on diagnostic yield

	Converting MRI-negative to MRI-positive			
No of studies (Design)	**Post-**	**Pre-**	**Odds Ratio (95% CI)**	**Quality**
MRI post-processing				
24 (Obs)	367/679	91/679	11.41 (7.30–17.81)	+++ moderate
Qualitative MRI				
16 (Obs)	299/563	144/563	5.87 (3.50–9.83)	+++ moderate

Abbreviations – **GRADE**: Grading of Recommendations Assessment, Development, and Evaluation; **Obs**: observational; **CI**: confidence interval; **OR**: Odds Ratio

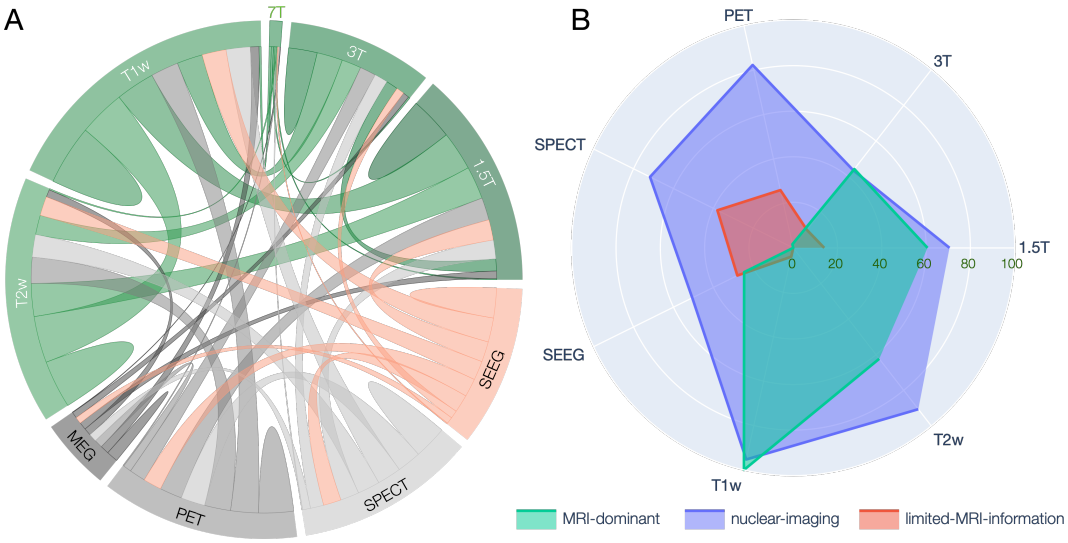


Figure A.1: **A**. The chord diagram depicts the co-occurrence of any two MRI (shades of green) and non-MRI modalities (shades of gray, and orange) across the 196 included studies. The chord links the source and target modalities specified by the color while its width signifies the number of studies. **B**. The radar plot depicts the percent composition of imaging modalities per cluster. Only modalities with composition > 20% are indicated.

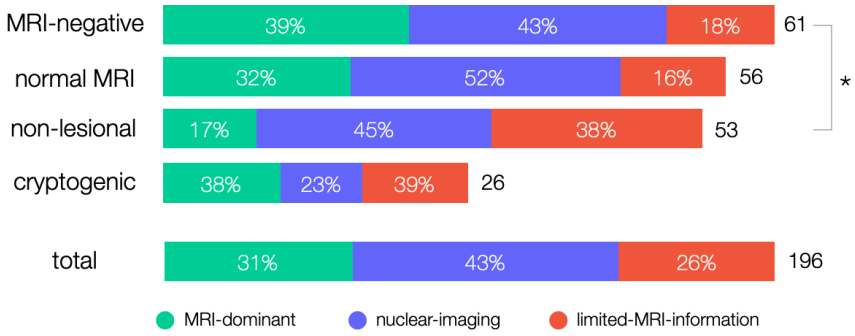


Figure A.2: Terminology used to describe MRI diagnostic status across groups. Note that MRI-dominant was 5 times more likely to use the term "MRI-negative" than "non-lesional" (*: χ^2=7.50, uncorrected $p < 0.05$), while the opposite was true for the limited-MRI-information group.

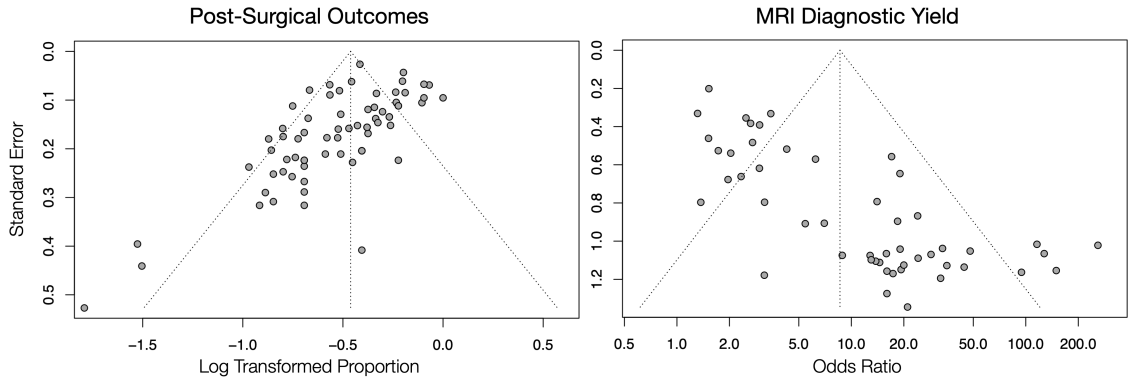


Figure A.3: Funnel plots relating the study-wise effect size to the standard error for the meta-analysis on post-surgical outcomes (*left, log transformed proportion of Engel-I outcome*) and MRI diagnostic yield (*right, Odds Ratio of the diagnostic yield*). Gray solid dots represent a study. The vertical dotted line along the middle of the funnel indicates the average effect size. The outer dotted lines delineate the boundaries within which 95% of studies are expected to lie in the absence of biases and heterogeneity. The funnel plot for the post-surgical outcome is asymmetric (*i.e.,* all studies are not distributed symmetrically around the effect size), indicative of publication bias due to missing small samples/large effect size studies in the right bottom corner. Conversely, the funnel plot for MRI diagnostic yield is symmetric signifying absence of publication bias.

B Supplementary Data for Chapter 6

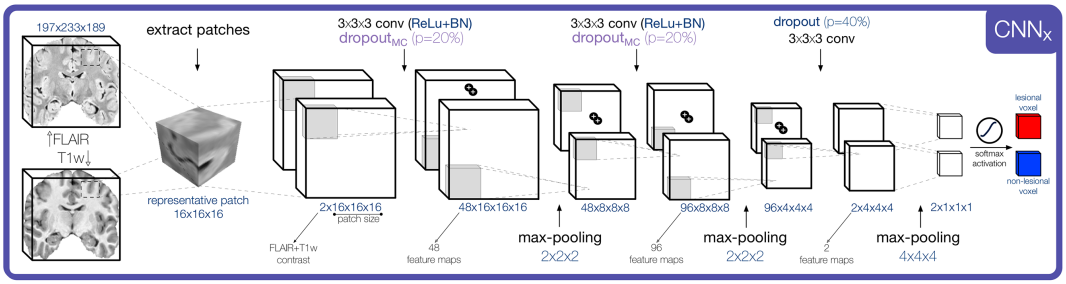
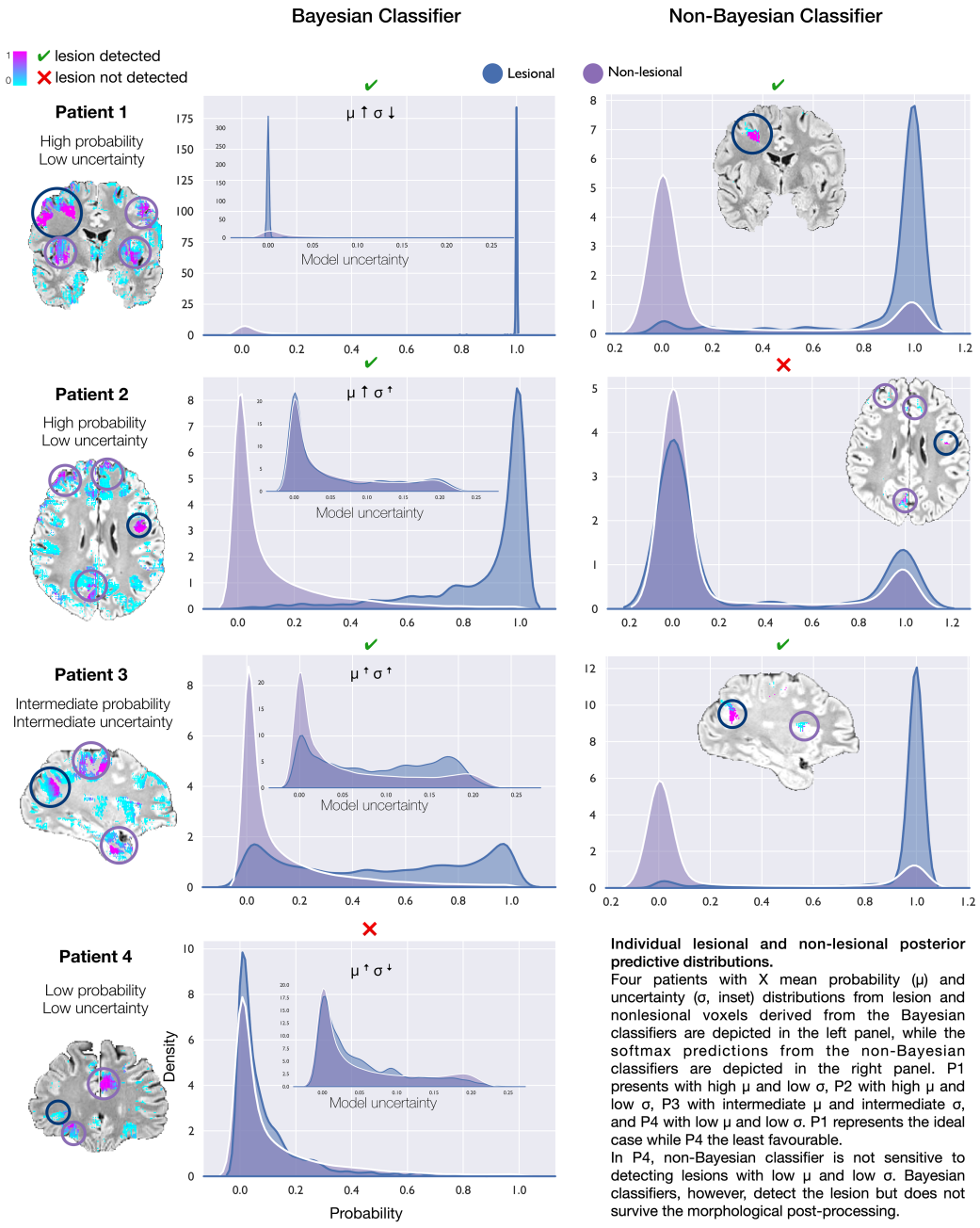


Figure B.1: Convolutional network architecture (CNN_x) for two-label (lesional *vs.* non-lesional) classification with three consecutive convolutions (kernel size: 3×3×3, filters: 48,96,2) and max-pooling units, followed by a voxel-wise softmax classification using multimodal (FLAIR+T1w) patches. Each convolution is followed by rectified linear units (ReLu), which introduce non-linearity. Batch normalization (BN) and *dropout* (p=40%) serve as regularizers to 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, and exclusive to Bayesian classifiers detailed in Classification paradigm of the text. Adadelta serves as the gradient descent optimizer to minimize the binary crossentropy loss.



Individual lesional and non-lesional posterior predictive distributions.
Four patients with X mean probability (μ) and uncertainty (σ, inset) distributions from lesion and nonlesional voxels derived from the Bayesian classifiers are depicted in the left panel, while the softmax predictions from the non-Bayesian classifiers are depicted in the right panel. P1 presents with high μ and low σ, P2 with high μ and low σ, P3 with intermediate μ and intermediate σ, and P4 with low μ and low σ. P1 represents the ideal case while P4 the least favourable.
In P4, non-Bayesian classifier is not sensitive to detecting lesions with low μ and low σ. Bayesian classifiers, however, detect the lesion but does not survive the morphological post-processing.

Figure B.2: Individual lesional and non-lesional posterior predictive distributions

C Supplementary Data for Chapter 7

Classifier design

(a) *Data sampling and patch-based augmentation*. Data augmentation encompasses a suite of techniques that enhance the size and quality of training datasets; in addition, they make training computationally tractable without information loss, prevent overfitting, and increase generalizability [299]. To this end, we sampled hyperintense voxels by thresholding FLAIR images z-normalized with respect to each subject discarding the bottom 10 percentile intensities; this thresholding resulted in a mask containing voxels within the grey matter and its interface with the white matter. The mask was then used to extract 3D patches (dimension: 16×16×16 voxels) from co-registered T1-weighted and FLAIR, thereby modelling the spatial extent of the FCD across orthogonal planes and tissue types. Data was further normalized within each modality with zero mean and unit variance. The set of all computed 3D patches were aggregated as P=n×2×16×16×16 (where n and 2 denote the number of training patches and input MRI modalities, respectively) and randomized. As the number of healthy voxels significantly outweighs lesional ones (< 1% of total voxels), we undersampled patches derived from healthy voxels to obtain equal numbers from both healthy and lesional tissue. This multicontrast patch dataset then served as input to the convolutional neural network (CNN).

(b) *Network architecture*. A typical CNN consists of three stages: convolutions, nonlinearity, and pooling. Recursive and hierarchical abstraction of data at varying depths (or layers) are at the core of the architecture. Specifically, a convolutional operator allows implicit encoding of spatial covariance amongst neighboring voxels (rather than treating each voxel independently), while nonlinearity optimizes performance on fine-grained tasks. The current CNN (see Figure C.1) is composed of three stacks of convolution and max-pooling layers with 48, 96 and 2 filters (size: 3×3×3 for convolution and 2×2×2 or 4×4×4 for max-pooling), respectively, followed by a voxel-wise Softmax activation using multimodal (FLAIR+T1w) patches. The rectified linear activation (ReLU) [282] non-linearity function is applied to the first two of the three convolutional layers. Batch normalization (BN) and dropout [254] ($p_{dropout}$=0.4, namely 40% of the neurons are randomly dropped) serve as regularizers to

prevent overfitting. Dropout Monte Carlo operation ($Dropout_{MC}$; p=0.2, namely 20% of neurons are randomly dropped) after the first and second convolution layers are essential to quantify epistemic model uncertainty [198, 248]. Softmax non-linearity is used after the final convolution to normalize the result of the kernel convolutions into a binominal distribution over healthy and lesional labels. The adaptive learning rate method Adadelta [300] served as the gradient descent optimizer to minimize binary cross-entropy loss, starting with a randomly initialized model. The training set is used to adjust the weights of the CNN, while the validation set (25% of the training set) measures the performance of trained CNN after each epoch and continues until the validation error plateaus to achieve an optimal bias-variance tradeoff, thereby mitigating overfitting.

(c) *Estimating prediction uncertainty*. Assessing uncertainty is an important step towards ensuring the safety and reliability of machine learning systems. Deterministic classifiers provide prediction probabilities without estimating the algorithm's uncertainty or confidence in prediction. This is achieved by sampling a distribution of predictions at test time. Each test prediction varies slightly –the lower the spread of distribution of predictions, the higher the confidence, and vice versa. While probabilities can be normalized between 0 and 1, they are uncalibrated, overconfident estimates for out-of-distribution data, *i.e.,* they do not match the training data distribution [301]. The fundamental reason for miscalibration is failure to account for uncertainty. A neural network can represent many models that are consistent with our data, while classical procedures discard this uncertainty and rely on a single model [197]. On the contrary, Bayesian probabilities are better calibrated since uncertainty is preserved by averaging over several models; in addition, this variance over the predictive distributions provides an estimate of uncertainty that is clinically relevant.
We computed the posterior distribution $p(w|X, y)$, where X, y is the training dataset and w the learned weights of the CNN. In practice, while the solution of this posterior is analytically intractable, variational inference (VI) methods approximate it with a parameterized distribution $q(w)$ (Equation C.1 below).

$$p\big(\theta \mid X, y, x^{\star}\big) \approx \int p(\theta \mid x^{\star}, w) p\big(w \mid X, y\big) dw \approx \int p(\theta \mid x^{\star}, w) q(w) dw \tag{C.1}$$

$$\mathcal{L} := \int p\big(y \mid X, w\big) q(w) dw - KL\big(q(w) \parallel p(w)\big) \tag{C.2}$$

The first term in Equation C.2 above maximizes the likelihood of the training data X, y, whereas the second approximates the true distribution $p(w)$ by $q(w)$. Gal & Ghahramani [248] empirically associated Equation C.2 with dropout training to approximate the intractable integral with Monte Carlo sampling. This results in the conventional Softmax loss for dropout networks, for which units are dropped by drawing from a Bernoulli prior

with probability drop for setting a unit to zero. The Kullback-Leibler (KL) term (see Equation C.2) was shown to correspond to a $L2$-regularization term in dropout networks.

Source code and data availability

The source code and the pre-trained model weights for the FCD detection algorithm are available at `https://github.com/NOEL-MNI/deepFCD` (*v.* 1.0.0; `doi.org/10.5281/zenodo.4521706`), and for the brain extraction at `https://github.com/NOEL-MNI/deepMask` (*v.* 1.0.0; `doi.org/10.5281/zenodo.4521716`). To minimize software installation and obviate licensing (MATLAB or equivalent) issues, the open-source Python-based algorithm and all its dependencies are containerized into a Docker application. The Docker images for `deepFCD` and `deepMask` can be accessed through their respective GitHub repositories under the Packages section.

Pre-trained model weights for CNN-1 and CNN-2 are provided as two separate Hierarchical Data Format (HDF5) files: `noel_deepFCD_dropoutMC_model_1.h5` and `noel_deepFCD_dropoutMC_model_2.h5`, respectively, and located in the `app/weights` folder in the GitHub repository (`https://github.com/NOEL-MNI/deepFCD`).

Anonymized lesional and non-lesional patches derived from 148 FCD patients is available as a HDF5 dataset (available from Dryad: `doi.org/10.5061/dryad.h70rxwdgm`). To create this dataset, for each of the 148 FCD patients, we sampled at most 1,000 cortical patches (or # voxels in the lesion, whichever is lower) of size 16×16×16 within the lesion. The same number of cortical patches were sampled randomly outside the lesion. The resulting lesional and non-lesional patches were concatenated, shuffled (to add another layer of randomization), and saved along with their binary labels (as a compressed HDF5 dataset).

Table C.1: **MRI acquisition parameters across sites**.

Site	MRI scanner	Coil	TR (ms)/TE (ms)/ Flip angle (degrees)		Voxel resolution [x, y, z (mm)]	
			T1-weighted	**FLAIR**	**T1-weighted**	**FLAIR**
S1	Siemens TimTrio 3T	32ch	2300/2.98/9	5000/389/120	1×1×1	0.9×0.9×0.9
	Siemens Prisma 3T	32ch	2300/2.9/9	5000/392/120	1×1×1	0.9×0.9×0.9
S2	GE 3T Discovery	32ch	8/3/12	6283/118/90	0.89×0.89×0.78	0.8×0.55×0.62
S3	Philips Achieva 3T	8ch	7.51/3.55/8	4800/278/90	0.45×0.45×0.45	0.55×0.55×0.56
S4	Philips Achieva 3T	8ch	8.18/3.75/8	6628/315/90	1.00×0.82×0.82	0.69×0.59×0.59
	GE Signa Hdxt 3T	8ch	10.68/4.87/13	11000/129/90	1.00×0.82×0.82	1.00×0.42×0.42
S5	Siemens Prisma 3T	32ch	2000/2.26/12	5000/388/120	1×1×1	1×1×1
	Siemens TrioTrim 3T	32ch	1567/2.15/15	5000/388/120	1×0.6×0.6	1×0.7×0.7
S6	Siemens Verio 3T	12ch	2155/2.7/9	5000/295/120	1×1×1	0.8×0.5×0.7
S7	Siemens Skyra 3T	32ch	1900/2.3/9	5444/392/120	0.9×0.9×0.9	0.92×0.85×0.85
S8	Philips Achieva 3T	8ch	9.9/4.6/8	4800/307/90	1×1×1	0.8×1.04×0.8
	Siemens Avanto 1.5T	8ch	1640/2/12	6667/340/110	0.8×0.8×1.0	1.02×1.02×1
S9	Philips Achieva 3T	8ch	9.6/4.4/8	6446/301/90	0.78×0.94×0.91	0.51×0.56×0.63

Abbreviations. S = site; TR = repetition time; TE = echo time; ms = milliseconds

Table C.2: **Peak location of FCD lesions in MNI space** (see Figure 7.5A).

ID	Peak (x,y,z)	Cluster size	Anatomical distribution
1	-22, 1, 29	320043	8% Front_Mid_R; 7% Front_Sup_R; 6% Front_Sup_L; 5% Front_Mid_L
2	-35, -73, 19	41437	28% Fus_L; 18% Temp_Inf_L; 10% Temp_Mid_L; 8% Occip_Mid_L; 6% Parahipp_L
3	16, -71, 0	8032	64% Ling_R; 22% Fus_R; 7% Calc_R
4	-45, 8, -35	5851	33% Temp_Pole_Sup_L; 30% Temp_Pole_Mid_L; 17% Temp_Inf_L; 12% Temp_Mid_L
5	-26, -37, 33	2727	18% Par_Inf_L; 12% Postc_L; 5% Par_Sup_L
6	-17, 41, -14	522	47% OFmed_L; 16% OFant_L; 12% Rec_L
7	-30, 30, -19	447	46% OFpost_L; 16% OFmed_L; 15% OFant_L
8	-39, -15, -3	386	49% Ins_L; 9% Temp_Sup_L
9	-50, -10, -34	262	100% Temp_Inf_L
10	-7, -42, 31	205	59% Cing_Post_L; 35% Cing_Mid_L
11	54, 31, 18	76	100% Front_Inf_Tri_R
12	-55, -12, 9	59	73% Heschl_L; 27% Temp_Sup_L
13	-19, -45, 72	51	67% Par_Sup_L; 17% Precun_L; 16% Postc_L
14	-16, -64, -4	42	100% Ling_L
15	-19, 31, -14	39	38% OFmed_L; 13% OFant_L; 10% OFpost_L
16	-31, -39, 62	37	100% Postc_L

Lesions (ID) are ranked by size (number of voxels). Anatomical distribution is defined based on percent overlap with the AAL atlas [302].

Abbreviations. Cing: Cingulate; Front: frontal; Fus: fusiform; Ling: Lingual; Occip: Occipital; OF: Orbitofrontal; Parahip: Parahippocampal; Par: Parietal; Precun: Precuneus; Temp: temporal; Rec: Rectus.

Ant/inf/med/mid/post/sup: Anterior, inferior, medial, middle, posterior, superior

Table C.3: **Peak location of false positive clusters in MNI space** (see Figure 7.5B).

ID	**Peak** (x,y,z)	**Cluster size**	**Anatomical distribution**	**Group**
1	-19, -2, -26	69990	13% Ins_L; 8% Cing_Ant_L; 7% Cing_Ant_R; 6% Temp_Mid_L; 5% Parahipp_L; 5% Temp_Sup_L	FCD
2	36, -6, 3	17308	53% Ins_R; 12% Rol_Oper_R	FCD
3	-43, 3, 36	10726	63% Prec_L; 10% Front_Inf_Oper_L; 9% Postc_L; 8% Front_Sup_L	FCD
4	53, -30, -1	8344	52% Temp_Mid_R; 27% Temp_Sup_R; 18% Temp_Inf_R	FCD
5	47, 5, 31	6437	65% Prec_R; 21% Front_Inf_Oper_R; 9% Front_Mid_R	FCD
6	38, 0, 2	5287	56% Ins_R	HC
7	38, -3, 4	5069	73% Ins_R	TLE
8	-41, 8, -6	4378	76% Ins_L; 10% Temp_Sup_L	TLE
9	-40, 14, -8	3896	72% Ins_L	HC
10	5, -20, 50	3556	76% Cing_Mid_R; 10% SMA_R; 9% Precun_R	FCD
11	31, 34, 48	3267	59% Front_Mid_R; 40% Front_Sup_R	FCD
12	-30, -2, -16	2268	47% Parahipp_L; 20% Amy_L	TLE
13	22, -12, 57	2267	49% Front_Sup_R; 26% Prec_R	FCD
14	-58, -28, -23	2132	79% Temp_Inf_L	FCD
15	17, -86, 8	1766	88% Calc_R	FCD
16	-6, -90, 8	1504	77% Calc_L; 15% Ling_L	FCD
17	-27, -1, -16	1454	42% Parahipp_L; 34% Amy_L	HC
18	21, 3, -30	1360	77% Parahipp_R	TLE
19	15, -34, 0	690	16% Precun_R; 13% Hipp_R; 10% Ling_R	FCD
20	18, -56, 18	595	51% Precun_R; 39% Calc_R; 9% Cun_R	FCD
21	22, 4, -25	529	71% Parahipp_R; 21% Amy_R	HC
22	49, -63, 2	437	54% Temp_Mid_R; 46% Temp_Inf_R	FCD
23	40, -10, -33	337	80% Fus_R; 20% Temp_Inf_R	FCD
24	12, 18, 63	317	91% SMA_R; 5% Front_Sup_R	FCD

25	-46, -34, 16	229	73% Temp_Sup_L; 26% Rol_Oper_L	HC
26	53, 5, -10	224	91% Temp_Pole_Sup_R; 9% Temp_Sup_R	FCD
27	-44, -26, 15	223	74% Rol_Oper_L; 26% Temp_Sup_L	TLE
28	17, -83, 8	204	93% Calc_R; 7% Cun_R	HC
29	-8, 14, 67	194	100% SMA_L	FCD
30	-42, -25, 18	147	94% Rol_Oper_L; 6% Supramarg_L	HC
31	-46, -66, 18	121	79% Temp_Mid_L; 15% Angular_L	FCD
32	15, -82, 10	117	98% Calc_R	TLE
33	-66, -17, -20	108	94% Temp_Mid_L; 6% Temp_Inf_L	FCD
34	30, 0, -16	107	97% Amy_R	TLE
35	20, -35, -14	101	71% Fus_R; 25% Parahipp_R	FCD
36	-15, -34, -1	85	49% Hipp_L	HC
37	5, -8, 63	84	100% SMA_R	FCD
38	-51, -51, -12	78	100% Temp_Inf_L	FCD
39	50, -71, 24	73	58% Temp_Mid_R; 42% Occip_Mid_R	FCD
40	-49, -73, 30	60	93% Angular_L	FCD
41	10, 38, 53	57	86% Front_Sup_Medial_R; 14% Front_Sup_R	FCD
42	-10, -90, 5	53	66.% Calc_L; 34 Occip_Sup_L	HC
43	49, -50, 15	50	98% Temp_Mid_R	FCD
44	-37, 45, 31	47	100% Front_Mid_L	FCD
45	-43, -31, 42	47	55% Postc_L; 45% Par_Inf_L	FCD
46	53, -7, -31	47	89% Temp_Inf_R; 11% Temp_Mid_R	FCD
47	12, 45, 49	46	61% Front_Sup_Medial_R; 39% Front_Sup_R	FCD
48	18, -18, 69	46	89% Prec_R; 9% Front_Sup_R	FCD
49	-54, -37, -2	45	100% Temp_Mid_L	HC
50	62, 3, -7	44	68% Temp_Sup_R; 32% Temp_Pole_Sup_R	FCD
51	65, -13, 20	42	100% Postc_R	FCD
52	14, -20, 67	38	50% Prec_R; 39% SMA_R	FCD
53	55, -63, 7	38	100% Temp_Mid_R	FCD
54	-15, -10, 67	37	43% SMA_L; 32% Parac_L; 24% Prec_L	FCD
55	-44, -69, -6	35	54% Temp_Inf_L; 46% Occip_Inf_L	FCD
56	38, 40, 37	34	100% Front_Mid_R	FCD
57	-13, -57, 10	31	87% Calc_L; 13% Precun_L	FCD
58	-55, -44, -14	30	77% Temp_Inf_L; 23% Temp_Mid_L	FCD

False positive clusters (ID) for FCD, healthy controls (HC) and TLE patients ranked by size (number of voxels); anatomical distribution is defined by percent overlap with the AAL atlas [302].

Abbreviations. Amy: Amygdala; Cing: Cingulate; Cun: Cuneus; Front: frontal; Fus: fusiform; Hip: Hippocampus; Ling: Lingual; Occip: Occipital; OF:Orbitofrontal; Parac: Paracentral lobule; Parahip: Parahippocampal; Par: Parietal; Prec: Precentral; Precun: Precuneus; Rol: Rolandic; Supramarg: Supramarginal; Temp: temporal; Rec: Rectus; SMA: Supplementary motor area.

Ant/inf/med/mid/post/sup: Anterior, inferior, medial, middle, posterior, superior

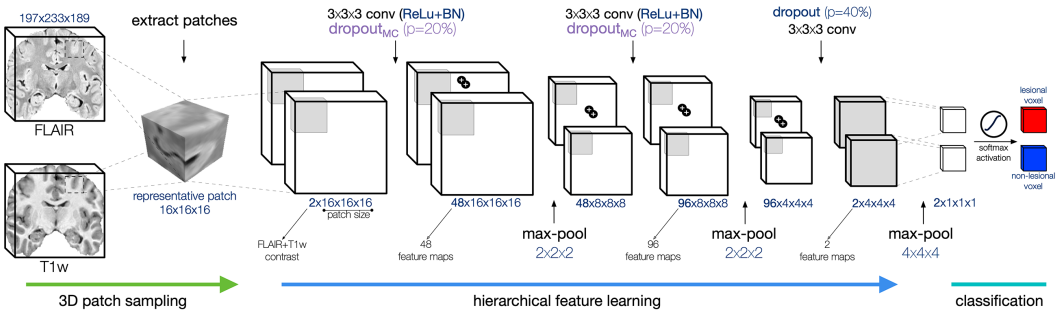


Figure C.1: **Hierarchical patch-based feature learning using CNN**. The CNN workflow entails patch sampling, hierarchical feature learning and classification (colored solid arrows and line) for a single patch. For each test subject, T1-weighted and FLAIR images were first partitioned into 3D patches, which served as input to the CNN. The two-label classification (lesional *vs.* non-lesional) is achieved by performing a series of operations [Convolution (conv), Batch Normalization (BN), ReLu, $Dropout/Dropout_{MC}$, Max Pooling (max-pool), and Softmax activation] at different depths (or layers) to learn (or extract) features from the multimodal patch dataset. The feature dimensions (larger cubes with black outline) at each step are in the format N, where N represents the number of features and P the edge size of the patch. The convolutional operation dilates and contracts the first dimension (2 to 48, 48 to 96, and 96 to 2 feature maps) to learn cross-modal feature dependencies, while the max pooling operation downsamples the data (16×16×16 to 8×8×8, 8×8×8 to 4×4×4, and 4×4×4 to 1×1×1) to learn features at different spatial scales. The gray cubes (within larger cube frames) represent the kernel size for conv (3×3×3) or max-pool (2×2×2 or 4×4×4). These kernels slide over all the voxels in the multimodal patch to generate feature maps. Finally, the softmax activation layer normalizes the network output to probabilities between 0 and 1. The binary cross-entropy loss is computed between the ground truth label and the predicted label, and errors back-propagated through the network iteratively until the performance as tested on the validation set no longer improves. See Classifier design for details. All elements in the illustration are not to scale.

Part V

Bibliography

1. R. S. Fisher, J. H. Cross, C. D'Souza, J. a. French, S. R. Haut, N. Higurashi, E. Hirsch, F. E. Jansen, L. Lagae, S. L. Moshé, J. Peltola, E. Roulet Perez, I. E. Scheffer, A. Schulze-Bonhage, E. Somerville, M. Sperling, E. M. Yacubian, and S. M. Zuberi. "Instruction manual for the ILAE 2017 operational classification of seizure types." *Epilepsia* 58:4, 2017, pp. 531–542.
2. M. R. Keezer, S. M. Sisodiya, and J. W. Sander. "Comorbidities of epilepsy: current concepts and future perspectives." *The Lancet. Neurology* 15:1, 2016, pp. 106–115.
3. O. Devinsky, D. C. Hesdorffer, D. J. Thurman, S. Lhatoo, and G. Richerson. "Sudden unexpected death in epilepsy: epidemiology, mechanisms, and prevention." *The Lancet. Neurology* 15:10, 2016, pp. 1075–1088.
4. J. F. Téllez-Zenteno, L. Hernandez-Ronquillo, F. Moien-Afshari, and S. Wiebe. "Surgical outcomes in lesional and non-lesional epilepsy: a systematic review and meta-analysis." *Epilepsy Research* 89:2-3, 2010, pp. 310–318.
5. J. Engel and N. Salamon. "Temporal Lobe Epilepsy". In: *Brain Mapping*. Ed. by A. W. Toga. Academic Press, Waltham, 2015, pp. 853–860.
6. A. J. Barkovich, R. Guerrini, R. I. Kuzniecky, G. D. Jackson, and W. B. Dobyns. "A developmental and genetic classification for malformations of cortical development: update 2012." *Brain: a journal of neurology* 135:Pt 5, 2012, pp. 1348–1369.
7. J. T. Lerner, N. Salamon, J. S. Hauptman, T. R. Velasco, M. Hemb, J. Y. Wu, R. Sankar, W. Donald Shields, J. Engel, I. Fried, C. Cepeda, V. M. André, M. S. Levine, H. Miyata, W. H. Yong, H. V. Vinters, and G. W. Mathern. "Assessment and surgical outcomes for mild type I and severe type II cortical dysplasia: a critical review and the UCLA experience." *Epilepsia* 50:6, 2009, pp. 1310–1335.
8. S. Wiebe, W. T. Blume, J. P. Girvin, M. Eliasziw, and Effectiveness and Efficiency of Surgery for Temporal Lobe Epilepsy Study Group. "A randomized, controlled trial of surgery for temporal-lobe epilepsy." *The New England journal of medicine* 345:5, 2001, pp. 311–318.
9. S. Fauser, A. Schulze-Bonhage, J. Honegger, H. Carmona, H.-J. Huppertz, G. Pantazis, S. Rona, T. Bast, K. Strobl, B. J. Steinhoff, R. Korinthenberg, D. Rating, B. Volk, and J. Zentner. "Focal cortical dysplasias: surgical outcome in 67 patients in relation to histological subtypes and dual pathology." *Brain: a journal of neurology* 127:Pt 11, 2004, pp. 2406–2418.
10. B. C. Jobst and G. D. Cascino. "Resective epilepsy surgery for drug-resistant focal epilepsy: a review." *JAMA: The Journal of the American Medical Association* 313:3, 2015, pp. 285–293.

11. G. D. Cascino. “Surgical treatment for epilepsy.” *Epilepsy Research* 60:2-3, 2004, pp. 179–186.

12. N. Colombo, N. Salamon, C. Raybaud, C. Özkara, and A. J. Barkovich. “Imaging of malformations of cortical development.” *Epileptic disorders : international epilepsy journal with videotape* 11:3, 2009, pp. 194–205.

13. S. Adler, S. Lorio, T. S. Jacques, B. Benova, R. Gunny, J. H. Cross, T. Baldeweg, and D. W. Carmichael. “Towards in vivo focal cortical dysplasia phenotyping using quantitative MRI.” *NeuroImage: Clinical* 15, 2017, pp. 95–105.

14. N. C. De Lanerolle, J. H. Kim, A. Williamson, S. S. Spencer, H. P. Zaveri, T. Eid, and D. D. Spencer. “A Retrospective Analysis of Hippocampal Pathology in Human Temporal Lobe Epilepsy: Evidence for Distinctive Patient Subcategories”. *Epilepsia* 44:5, 2003, pp. 677–687.

15. J. S. Hauptman and G. W. Mathern. “Surgical treatment of epilepsy associated with cortical dysplasia: 2012 update.” *Epilepsia* 53 Suppl 4:Suppl. 3, 2012, pp. 98–104.

16. I. Blümcke et al. “The clinicopathologic spectrum of focal cortical dysplasias: a consensus classification proposed by an ad hoc Task Force of the ILAE Diagnostic Methods Commission.” *Epilepsia* 52:1, 2011, pp. 158–174.

17. P. Krsek, B. Maton, B. Korman, E. Pacheco-Jacome, P. Jayakar, C. Dunoyer, G. Rey, G. Morrison, J. Ragheb, H. V. Vinters, T. Resnick, and M. Duchowny. “Different features of histopathological subtypes of pediatric focal cortical dysplasia”. *Annals of Neurology* 63:6, 2008, pp. 758–769.

18. P. Krsek, B. Maton, P. Jayakar, P. Dean, B. Korman, G. Rey, C. Dunoyer, E. Pacheco-Jacome, G. Morrison, J. Ragheb, H. V. Vinters, T. Resnick, and M. Duchowny. “Incomplete resection of focal cortical dysplasia is the main predictor of poor postsurgical outcome.” *Neurology* 72:3, 2009, pp. 217–223.

19. S. M. Sisodiya, S. Fauser, J. H. Cross, and M. Thom. “Focal cortical dysplasia type II: biological features and clinical perspectives.” *The Lancet. Neurology* 8:9, 2009, pp. 830–843.

20. A. Bernasconi, N. Bernasconi, B. C. Bernhardt, and D. Schrader. “Advances in MRI for ’cryptogenic’ epilepsies”. *Nature reviews. Neurology* 7:2, 2011, pp. 99–108.

21. A. McGonigal, F. Bartolomei, J. Régis, M. Guye, M. Gavaret, A. Trébuchon-Da Fonseca, H. Dufour, D. Figarella-Branger, N. Girard, J.-C. Péragut, and P. Chauvel. “Stereoelectroencephalography in presurgical assessment of MRI-negative epilepsy.” *Brain: a journal of neurology* 130:Pt 12, 2007, pp. 3169–3183.

22. C. G. Bien, M. Szinay, J. Wagner, H. Clusmann, A. J. Becker, and H. Urbach. "Characteristics and surgical outcomes of patients with refractory magnetic resonance imaging-negative epilepsies." *Archives of neurology* 66:12, 2009, pp. 1491–1499.

23. I. Blümcke et al. "Histopathological Findings in Brain Tissue Obtained during Epilepsy Surgery." *The New England journal of medicine* 377:17, 2017, pp. 1648–1656.

24. D. King, S. S. Spencer, A. Bouthillier, J. Kim, N. de Lanerolle, R. A. Bronen, G. McCarthy, M. Luby, and D. D. Spencer. "Medial temporal lobe epilepsy without hippocampal atrophy". *Journal of Epilepsy* 9:4, 1996, pp. 291–297.

25. N. Bernasconi, A. Bernasconi, Z. Caramanos, F. Dubeau, J. Richardson, F. Andermann, and D. L. Arnold. "Entorhinal cortex atrophy in epilepsy patients exhibiting normal hippocampal volumes." *Neurology* 56:10, 2001, pp. 1335–1339.

26. A. A. Cohen-Gadol, C. C. Bradley, A. Williamson, J. H. Kim, M. Westerveld, R. B. Duckrow, and D. D. Spencer. "Normal magnetic resonance imaging and medial temporal lobe epilepsy: the clinical syndrome of paradoxical temporal lobe epilepsy." *Journal of neurosurgery* 102:5, 2005, pp. 902–909.

27. I. Blümcke, E. Pauli, H. Clusmann, J. Schramm, A. Becker, C. Elger, M. Merschhemke, H.-J. Meencke, T. Lehmann, A. von Deimling, C. Scheiwe, J. Zentner, B. Volk, J. Romstöck, H. Stefan, and M. Hildebrandt. "A new clinico-pathological classification system for mesial temporal sclerosis." *Acta Neuropathologica* 113:3, 2007, pp. 235–244.

28. M. Thom, J. Zhou, L. Martinian, and S. Sisodiya. "Quantitative post-mortem study of the hippocampus in chronic epilepsy: seizures do not inevitably cause neuronal loss." *Brain: a journal of neurology* 128:Pt 6, 2005, pp. 1344–1357.

29. D. Barazany and Y. Assaf. "Visualization of cortical lamination patterns with magnetic resonance imaging." *Cerebral cortex* 22:9, 2012, pp. 2016–2023.

30. F. De Martino, M. Moerel, J. Xu, P.-F. Van de Moortele, K. Ugurbil, R. Goebel, E. Yacoub, and E. Formisano. "High-Resolution Mapping of Myeloarchitecture In Vivo: Localization of Auditory Areas in the Human Brain." *Cerebral cortex* 25:10, 2015, pp. 3394–3405.

31. L. G. Kini, J. C. Gee, and B. Litt. "Computational analysis in epilepsy neuroimaging: A survey of features and methods". *NeuroImage: Clinical* 11, 2016, pp. 515–529.

32. J. von Oertzen, H. Urbach, S. Jungbluth, M. Kurthen, M. Reuber, G. Fernández, and C. E. Elger. "Standard magnetic resonance imaging is inadequate for patients with refractory focal epilepsy." *Journal of neurology, neurosurgery, and psychiatry* 73:6, 2002, pp. 643–647.

33. C. S. Lee, P. G. Nagy, S. J. Weaver, and D. E. Newman-Toker. "Cognitive and system factors contributing to diagnostic errors in radiology." *AJR. American journal of roentgenology* 201:3, 2013, pp. 611–617.

34. R. Kuzniecky and G. D. Jackson. *Magnetic resonance in epilepsy: neuroimaging techniques*. 2005.

35. A. L. Jones and G. D. Cascino. "Evidence on Use of Neuroimaging for Surgical Treatment of Temporal Lobe Epilepsy: A Systematic Review." *JAMA Neurology* 73:4, 2016, pp. 464–470.

36. T. P. Sutula, J. Hagen, and A. Pitkanen. "Do epileptic seizures damage the brain?" *Current opinion in neurology* 16:2, 2003, pp. 189–195.

37. C. D. Whelan et al. "Structural brain abnormalities in the common epilepsies assessed in a worldwide ENIGMA study." *Brain: a journal of neurology* 141:2, 2018, pp. 391–408.

38. S. Wiebe. "Burden of intractable epilepsy." *Advances in neurology* 97, 2006, pp. 1–4.

39. C. E. Begley, E. Beghi, R. G. Beran, D. Heaney, J. T. Langfitt, C. Pachlatko, H. Silfvenius, M. R. Sperling, and S. Wiebe. "ILAE Commission on the Burden of Epilepsy, Subcommission on the Economic Burden of Epilepsy: Final report 1998-2001." *Epilepsia* 43:6, 2002, pp. 668–673.

40. B. D. Semple, K. Blomgren, K. Gimlin, D. M. Ferriero, and L. J. Noble-Haeusslein. "Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species." *Progress in neurobiology* 106-107, 2013, pp. 1–16.

41. L. Subramanian, M. E. Calcagnotto, and M. F. Paredes. "Cortical Malformations: Lessons in Human Brain Development." *Frontiers in cellular neuroscience* 13, 2019, p. 576.

42. A. Pinson, T. Namba, and W. B. Huttner. "Malformations of Human Neocortex in Development - Their Progenitor Cell Basis and Experimental Model Systems." *Frontiers in cellular neuroscience* 13, 2019, p. 305.

43. F. C. Guarnieri, A. de Chevigny, A. Falace, and C. Cardoso. "Disorders of neurogenesis and cortical development." *Dialogues in clinical neuroscience* 20:4, 2018, pp. 255–266.

44. R. Guerrini and W. B. Dobyns. "Malformations of cortical development: clinical features and genetic causes". *The Lancet. Neurology* 13:7, 2014, pp. 710–726.

45. R. Oegema, T. S. Barakat, M. Wilke, K. Stouffs, D. Amrom, E. Aronica, N. Bahi-Buisson, V. Conti, A. E. Fry, T. Geis, D. G. Andres, E. Parrini, I. Pogledic, E. Said, D. Soler, L. M. Valor, M. S. Zaki, G. Mirzaa, W. B. Dobyns, O. Reiner, R. Guerrini, D. T. Pilz, U. Hehr, R. J. Leventer, A. C. Jansen, G. M. S. Mancini, and N. Di Donato. "International consensus recommendations on the diagnostic work-up for malformations of cortical development." *Nature reviews. Neurology* 13:Suppl. 1, 2020, p. 710.

46. P. Rakic. "Evolution of the neocortex: a perspective from developmental biology." *Nature reviews. Neuroscience* 10:10, 2009, pp. 724–735.

47. I. Bystron, C. Blakemore, and P. Rakic. "Development of the human cerebral cortex: Boulder Committee revisited." *Nature reviews. Neuroscience* 9:2, 2008, pp. 110–122.

48. S. Budday, P. Steinmann, and E. Kuhl. "Physical biology of human brain development." *Frontiers in cellular neuroscience* 9, 2015, p. 257.

49. L. Subramanian, M. Bershteyn, M. F. Paredes, and A. R. Kriegstein. "Dynamic behaviour of human neuroepithelial cells in the developing forebrain." *Nature communications* 8:1, 2017, pp. 14167–12.

50. J. H. Lui, D. V. Hansen, and A. R. Kriegstein. "Development and evolution of the human neocortex." *Cell* 146:1, 2011, pp. 18–36.

51. P. Iacopetti, M. Michelini, I. Stuckmann, B. Oback, E. Aaku-Saraste, and W. B. Huttner. "Expression of the antiproliferative gene TIS21 at the onset of neurogenesis identifies single neuroepithelial cells that switch from proliferative to neuron-generating division". *Proceedings of the National Academy of Sciences of the United States of America* 96:8, 1999, pp. 4639–4644.

52. A. Pontious, T. Kowalczyk, C. Englund, and R. F. Hevner. "Role of Intermediate Progenitor Cells in Cerebral Cortex Development". *Developmental Neuroscience* 30:1-3, 2007, pp. 24–32.

53. S. C. Noctor, V. Martínez-Cerdeño, L. Ivic, and A. R. Kriegstein. "Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases". *Nature neuroscience* 7:2, 2004, pp. 136–144.

54. W. Haubensak, A. Attardo, W. Denk, and W. B. Huttner. "Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: A major site of neurogenesis". *Proceedings of the National Academy of Sciences of the United States of America* 101:9, 2004, pp. 3196–3201.

55. R. L. Sidman and P. Rakic. "Neuronal migration, with special reference to developing human brain: a review". *Brain research* 62:1, 1973, pp. 1–35.

56. N. Zecevic, Y. Chen, and R. Filipovic. “Contributions of cortical subventricular zone to the development of the human cerebral cortex”. *The Journal of comparative neurology* 491:2, 2005, pp. 109–122.

57. I. Y. Buchsbaum and S. Cappello. “Neuronal migration in the CNS during development and disease: insights from in vivo and in vitro models”. *Development* 146:1, 2019.

58. J. A. Cooper. “Cell biology in neuroscience: mechanisms of cell migration in the nervous system.” *The Journal of cell biology* 202:5, 2013, pp. 725–734.

59. M. F. Paredes, D. James, S. Gil-Perotin, H. Kim, J. A. Cotter, C. Ng, K. Sandoval, D. H. Rowitch, D. Xu, P. S. McQuillen, J. M. Garcia-Verdugo, E. J. Huang, and A. Alvarez-Buylla. “Extensive migration of young neurons into the infant human frontal lobe”. *Science* 354:6308, 2016, aaf7073–aaf7073.

60. O. Marín and J. L. R. Rubenstein. “Cell migration in the forebrain.” *Annual review of neuroscience* 26:1, 2003, pp. 441–483.

61. M. Hadders-Algra. “Early human motor development: From variation to the ability to vary and adapt.” *Neuroscience and biobehavioral reviews* 90, 2018, pp. 411–427.

62. B. Nadarajah and J. G. Parnavelas. “Modes of neuronal migration in the developing cerebral cortex.” *Nature reviews. Neuroscience* 3:6, 2002, pp. 423–432.

63. D. V. Hansen, J. H. Lui, P. Flandin, K. Yoshikawa, J. L. Rubenstein, A. Alvarez-Buylla, and A. R. Kriegstein. “Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences”. *Nature neuroscience* 16:11, 2013, pp. 1576–1587.

64. O. Marín, M. Valiente, X. Ge, and L.-H. Tsai. “Guiding neuronal cell migrations.” *Cold Spring Harbor perspectives in biology* 2:2, 2010, a001834.

65. M. Sur and J. L. R. Rubenstein. “Patterning and plasticity of the cerebral cortex.” *Science* 310:5749, 2005, pp. 805–810.

66. D. D. M. O’Leary, S.-J. Chou, and S. Sahara. “Area Patterning of the Mammalian Cortex”. *Neuron* 56:2, 2007, pp. 252–269.

67. D. D. O’Leary. “Do cortical areas emerge from a protocortex?” *Trends in neurosciences* 12:10, 1989, pp. 400–406.

68. P. Rakic. “Specification of cerebral cortical areas”. *Science* 241:4862, 1988, pp. 170–176.

69. B. J. Molyneaux, P. Arlotta, J. R. L. Menezes, and J. D. Macklis. “Neuronal subtype specification in the cerebral cortex.” *Nature reviews. Neuroscience* 8:6, 2007, pp. 427–437.

70. E. Förster, Y. Jossin, S. Zhao, X. Chai, M. Frotscher, and A. M. Goffinet. "Recent progress in understanding the role of Reelin in radial neuronal migration, with specific emphasis on the dentate gyrus." *The European journal of neuroscience* 23:4, 2006, pp. 901–909.

71. Y. Hirota and K. Nakajima. "Control of Neuronal Migration and Aggregation by Reelin Signaling in the Developing Cerebral Cortex." *Frontiers in cell and developmental biology* 5, 2017, p. 40.

72. F. E. Poulain and A. Sobel. "The microtubule network and neuronal morphogenesis: Dynamic and coordinated orchestration through multiple players." *Molecular and cellular neurosciences* 43:1, 2010, pp. 15–32.

73. M. A. Holland, K. E. Miller, and E. Kuhl. "Emerging Brain Morphologies from Axonal Elongation." *Annals of biomedical engineering* 43:7, 2015, pp. 1640–1653.

74. R. Kuzniecky. "Epilepsy and malformations of cortical development: new developments." *Current opinion in neurology* 28:2, 2015, pp. 151–157.

75. I. Blümcke, H. V. Vinters, D. Armstrong, E. Aronica, M. Thom, and R. Spreafico. "Malformations of cortical development and epilepsies: neuropathological findings with emphasis on focal cortical dysplasia." *Epileptic disorders : international epilepsy journal with videotape* 11:3, 2009, pp. 181–193.

76. L. A. Jansen, G. M. Mirzaa, G. E. Ishak, B. J. O'Roak, J. B. Hiatt, W. H. Roden, S. A. Gunter, S. L. Christian, S. Collins, C. Adams, J.-B. Rivière, J. St-Onge, J. G. Ojemann, J. Shendure, R. F. Hevner, and W. B. Dobyns. "PI3K/AKT pathway mutations cause a spectrum of brain malformations from megalencephaly to focal cortical dysplasia." *Brain: a journal of neurology* 138:Pt 6, 2015, pp. 1613–1628.

77. P. H. Iffland and P. B. Crino. "Focal Cortical Dysplasia: Gene Mutations, Cell Signaling, and Therapeutic Implications." *Annual review of pathology* 12:1, 2017, pp. 547–571.

78. P. H. Iffland and P. B. Crino. "The role of somatic mutational events in the pathogenesis of epilepsy." *Current opinion in neurology*, 2019.

79. P. Perucca, M. Bahlo, and S. F. Berkovic. "The Genetics of Epilepsy." *Annual review of genomics and human genetics* 21:1, 2020, pp. 205–230.

80. K. A. Myers, D. Johnstone, and D. A. Dyment. "Epilepsy Genetics: current knowledge, applications and future directions." *Clinical genetics*, 2018.

81. M. Hildebrandt, T. Pieper, P. Winkler, D. Kolodziejczyk, H. Holthausen, and I. Blümcke. "Neuropathological spectrum of cortical dysplasia in children with severe focal epilepsies." *Acta Neuropathologica* 110:1, 2005, pp. 1–11.

82. A. Mühlebner, R. Coras, K. Kobow, M. Feucht, T. Czech, H. Stefan, D. Weigel, M. Buchfelder, H. Holthausen, T. Pieper, M. Kudernatsch, and I. Blümcke. "Neuropathologic measurements in focal cortical dysplasias: validation of the ILAE 2011 classification system and diagnostic implications for MRI." *Acta Neuropathologica* 123:2, 2012, pp. 259–272.

83. H. Urbach, B. Scheffler, T. Heinrichsmeier, J. von Oertzen, T. Kral, J. Wellmer, J. Schramm, O. D. Wiestler, and I. Blümcke. "Focal cortical dysplasia of Taylor's balloon cell type: a clinicopathological entity with characteristic neuroimaging and histopathological features, and favorable postsurgical outcome." *Epilepsia* 43:1, 2002, pp. 33–40.

84. C. Cepeda, R. S. Hurst, J. Flores-Hernández, E. Hernández-Echeagaray, G. J. Klapstein, M. K. Boylan, C. R. Calvert, E. L. Jocoy, O. K. Nguyen, V. M. André, H. V. Vinters, M. A. Ariano, M. S. Levine, and G. W. Mathern. "Morphological and electrophysiological characterization of abnormal cell types in pediatric cortical dysplasia." *Journal of neuroscience research* 72:4, 2003, pp. 472–486.

85. I. Marin-Valencia, R. Guerrini, and J. G. Gleeson. "Pathogenetic mechanisms of focal cortical dysplasia." *Epilepsia* 55:7, 2014, pp. 970–978.

86. T. S. Zimmer, D. W. M. Broekaart, M. Luinenburg, C. Mijnsbergen, J. J. Anink, N. S. Sim, I. Michailidou, F. E. Jansen, P. C. van Rijen, J. H. Lee, L. François, J. van Eyll, S. Dedeurwaerdere, E. A. van Vliet, A. Mühlebner, J. D. Mills, and E. Aronica. "Balloon cells promote immune system activation in focal cortical dysplasia type 2B." *Neuropathology and applied neurobiology*, 2021.

87. A. M. D'Gama, Y. Geng, J. A. Couto, B. Martin, E. A. Boyle, C. M. LaCoursiere, A. Hossain, N. E. Hatem, B. J. Barry, D. J. Kwiatkowski, H. V. Vinters, A. J. Barkovich, J. Shendure, G. W. Mathern, C. A. Walsh, and A. Poduri. "Mammalian target of rapamycin pathway mutations cause hemimegalencephaly and focal cortical dysplasia." *Annals of Neurology* 77:4, 2015, pp. 720–725.

88. T. Scerri, J. R. Riseley, G. Gillies, K. Pope, R. Burgess, S. A. Mandelstam, L. Dibbens, C. W. Chow, W. Maixner, A. S. Harvey, G. D. Jackson, D. J. Amor, M. B. Delatycki, P. B. Crino, S. F. Berkovic, I. E. Scheffer, M. Bahlo, P. J. Lockhart, and R. J. Leventer. "Familial cortical dysplasia type IIA caused by a germline mutation in DEPDC5." *Annals of clinical and translational neurology* 2:5, 2015, pp. 575–580.

89. A. Jeong and M. Wong. "Targeting the Mammalian Target of Rapamycin for Epileptic Encephalopathies and Malformations of Cortical Development." *Journal of child neurology* 33:1, 2018, pp. 55–63.

90. R. Guerrini and C. Barba. "Focal cortical dysplasia: an update on diagnosis and treatment." *Expert review of neurotherapeutics*, 2021, pp. 1–12.

91. R. H. G. J. van Lanen, A. J. Colon, C. J. Wiggins, M. C. Hoeberigs, G. Hoogland, A. Roebroeck, D. Ivanov, B. A. Poser, R. P. W. Rouhl, P. A. M. Hofman, J. F. A. Jansen, W. Backes, K. Rijkers, and O. E. M. G. Schijns. "Ultra-high field magnetic resonance imaging in human epilepsy: A systematic review." *NeuroImage: Clinical* 30, 2021, p. 102602.

92. R. S. Desikan and A. J. Barkovich. "Malformations of cortical development". *Annals of Neurology*, 2016.

93. A. J. Barkovich, W. B. Dobyns, and R. Guerrini. "Malformations of cortical development and epilepsy." *Cold Spring Harbor perspectives in medicine* 5:5, 2015, a022392–a022392.

94. I. Wang, S. Oh, I. Blümcke, R. Coras, B. Krishnan, S. Kim, A. McBride, O. Grinenko, Y. Lin, M. Overmyer, T. T. Aung, M. Lowe, M. Larvie, A. V. Alexopoulos, W. Bingaman, J. A. Gonzalez-Martinez, I. Najm, and S. E. Jones. "Value of 7T MRI and post-processing in patients with nonlesional 3T MRI undergoing epilepsy presurgical evaluation." *Epilepsia* 61:11, 2020, pp. 2509–2520.

95. T. J. Veersema, C. H. Ferrier, P. van Eijsden, P. H. Gosselaar, E. Aronica, F. Visser, J. M. Zwanenburg, G. A. P. de Kort, J. Hendrikse, P. R. Luijten, and K. P. J. Braun. "Seven tesla MRI improves detection of focal cortical dysplasia in patients with refractory focal epilepsy." *Epilepsia open* 2:2, 2017, pp. 162–171.

96. A. De Ciantis, C. Barba, L. Tassi, M. Cosottini, M. Tosetti, M. Costagli, M. Bramerio, E. Bartolini, L. Biagi, M. Cossu, V. Pelliccia, M. R. Symms, and R. Guerrini. "7T MRI in focal epilepsy with unrevealing conventional field strength imaging". *Epilepsia* 57:3, 2016, pp. 445–454.

97. R. E. Feldman, B. N. Delman, P. S. Pawha, H. Dyvorne, J. W. Rutland, J. Yoo, M. C. Fields, L. V. Marcuse, and P. Balchandani. "7T MRI in epilepsy patients with previously normal clinical MRI exams compared against healthy controls." *PLoS ONE* 14:3, 2019, e0213642.

98. A. A. Cohen-Gadol, K. Ozduman, R. A. Bronen, J. H. Kim, and D. D. Spencer. "Long-term outcome after epilepsy surgery for focal cortical dysplasia." *Journal of neurosurgery* 101:1, 2004, pp. 55–65.

99. N. Colombo, L. Tassi, F. Deleo, A. Citterio, M. Bramerio, R. Mai, I. Sartori, F. Cardinale, G. Lo Russo, and R. Spreafico. "Focal cortical dysplasia type IIa and IIb: MRI aspects in 118 cases proven by histopathology." *Neuroradiology* 54:10, 2012, pp. 1065–1077.

100. J. Ashburner and K. J. Friston. "Voxel-Based Morphometry—The Methods". *NeuroImage* 11:6, 2000, pp. 805–821.

101. O. Colliot, N. Bernasconi, N. Khalili, S. B. Antel, V. Naessens, and A. Bernasconi. "Individual voxel-based analysis of gray matter in focal cortical dysplasia." *NeuroImage* 29:1, 2006, pp. 162–171.

102. T. M. Salmenpera, M. R. Symms, F. J. Rugg-Gunn, P. A. Boulby, S. L. Free, G. J. Barker, T. A. Yousry, and J. S. Duncan. "Evaluation of Quantitative Magnetic Resonance Imaging Contrasts in MRI-Negative Refractory Focal Epilepsy". *Epilepsia* 48:2, 2007, pp. 229–237.

103. F. J. Rugg-Gunn, S. H. Eriksson, P. A. Boulby, M. R. Symms, G. J. Barker, and J. S. Duncan. "Magnetization transfer imaging in focal epilepsy." *Neurology* 60:10, 2003, pp. 1638–1645.

104. F. J. Rugg-Gunn, P. A. Boulby, M. R. Symms, G. J. Barker, and J. S. Duncan. "Whole-brain T2 mapping demonstrates occult abnormalities in focal epilepsy." *Neurology* 64:2, 2005, pp. 318–325.

105. N. K. Focke, M. R. Symms, J. L. Burdett, and J. S. Duncan. "Voxel-based analysis of whole brain FLAIR at 3T detects focal cortical dysplasia." *Epilepsia* 49:5, 2008, pp. 786–793.

106. A. Bernasconi, S. B. Antel, D. L. Collins, N. Bernasconi, A. Olivier, F. Dubeau, G. B. Pike, F. Andermann, and D. L. Arnold. "Texture analysis and morphological processing of magnetic resonance imaging assist detection of focal cortical dysplasia in extra-temporal partial epilepsy." *Annals of Neurology* 49:6, 2001, pp. 770–775.

107. S. B. Antel, D. L. Collins, N. Bernasconi, F. Andermann, R. Shinghal, R. E. Kearney, D. L. Arnold, and A. Bernasconi. "Automated detection of focal cortical dysplasia lesions using computational models of their MRI characteristics and texture analysis". *NeuroImage* 19:4, 2003, pp. 1748–1759.

108. S. B. Antel, A. Bernasconi, N. Bernasconi, D. L. Collins, R. E. Kearney, R. Shinghal, and D. L. Arnold. "Computational models of MRI characteristics of focal cortical dysplasia improve lesion detection." *NeuroImage* 17:4, 2002, pp. 1755–1760.

109. 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, 2005, pp. 35–50.

110. O. Colliot, T. Mansi, N. Bernasconi, V. Naessens, D. Klironomos, and A. Bernasconi. "Segmentation of focal cortical dysplasia lesions on MRI using level set evolution". *NeuroImage* 32:4, 2006, pp. 1621–1630.

111. S.-J. Hong. "Advanced multimodal imaging in epileptogenic malformations of cortical development". PhD thesis. McGill University: McGill University, 2016.

112. A. M. Dale, B. Fischl, and M. I. Sereno. "Cortical surface-based analysis. I. Segmentation and surface reconstruction." *NeuroImage* 9:2, 1999, pp. 179–194.

113. D. MacDonald, N. Kabani, D. Avis, and A. C. Evans. "Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI." *NeuroImage* 12:3, 2000, pp. 340–356.

114. J. S. Kim, V. Singh, J. K. Lee, J. Lerch, Y. Ad-Dab'bagh, D. MacDonald, J. M. Lee, S. I. Kim, and A. C. Evans. "Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification." *NeuroImage* 27:1, 2005, pp. 210–221.

115. J. G. Sled, A. P. Zijdenbos, and A. C. Evans. "A nonparametric method for automatic correction of intensity nonuniformity in MRI data." *IEEE transactions on medical imaging* 17:1, 1998, pp. 87–97.

116. D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans. "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space." *Journal of computer assisted tomography* 18:2, 1994, pp. 192–205.

117. A. Zijdenbos, R. Forghani, and A. Evans. "Automatic quantification of MS lesions in 3D MRI brain data sets: validation of INSECT". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 1998, pp. 439–448.

118. H. Kim, B. Caldairou, J.-W. Hwang, T. Mansi, S.-J. Hong, N. Bernasconi, and A. Bernasconi. "Accurate cortical tissue classification on MRI by modeling cortical folding patterns." *Human Brain Mapping* 36:9, 2015, pp. 3563–3574.

119. J. Tohka, A. Zijdenbos, and A. Evans. "Fast and robust parameter estimation for statistical partial volume models in brain MRI." *NeuroImage* 23:1, 2004, pp. 84–97.

120. O. Lyttelton, M. Boucher, S. Robbins, and A. Evans. "An unbiased iterative group registration template for cortical surface analysis." *NeuroImage* 34:4, 2007, pp. 1535–1544.

121. S. Robbins, A. C. Evans, D. L. Collins, and S. Whitesides. "Tuning and comparing spatial normalization methods." *Medical Image Analysis* 8:3, 2004, pp. 311–323.

122. S. S. Ghosh, S. Kakunoori, J. Augustinack, A. Nieto-Castanon, I. Kovelman, N. Gaab, J. A. Christodoulou, C. Triantafyllou, J. D. E. Gabrieli, and B. Fischl. "Evaluating the validity of volume-based and surface-based brain image registration for developmental cognitive neuroscience studies in children 4 to 11 years of age." *NeuroImage* 53:1, 2010, pp. 85–93.

123. A. A. Raymond, D. R. Fish, S. M. Sisodiya, N. Alsanjari, J. M. Stevens, and S. D. Shorvon. "Abnormalities of gyration, heterotopias, tuberous sclerosis, focal cortical dysplasia, microdysgenesis, dysembryoplastic neuroepithelial tumour and dysgenesis of the archicortex in epilepsy. Clinical, EEG and neuroimaging features in 100 adult patients." *Brain* 118 (Pt 3):3, 1995, pp. 629–660.

124. S. Adler, K. Wagstyl, R. Gunny, L. Ronan, D. Carmichael, J. H. Cross, P. C. Fletcher, and T. Baldeweg. "Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy". *NeuroImage: Clinical* 14, 2017, pp. 18–27.

125. B. Jin, B. Krishnan, S. Adler, K. Wagstyl, W. Hu, S. Jones, I. Najm, A. Alexopoulos, K. Zhang, J. Zhang, M. Ding, S. Wang, Pediatric Imaging, Neurocognition, and Genetics Study, and Z. I. Wang. "Automated detection of focal cortical dysplasia type II with surface-based magnetic resonance imaging postprocessing and machine learning." *Epilepsia* 59:5, 2018, pp. 982–992.

126. B. Ahmed, C. E. Brodley, K. E. Blackmon, R. Kuzniecky, G. Barash, C. Carlson, B. T. Quinn, W. Doyle, J. French, O. Devinsky, and T. Thesen. "Cortical feature analysis and machine learning improves detection of "MRI-negative" focal cortical dysplasia." *Epilepsy & behavior: E&B* 48, 2015, pp. 21–28.

127. B. Ahmed, T. Thesen, K. E. Blackmon, R. Kuzniekcy, O. Devinsky, and C. E. Brodley. "Decrypting cryptogenic epilepsy: semi-supervised hierarchical conditional random fields for detecting cortical lesions in MRI-negative patients". *The Journal of Machine Learning Research* 17:1, 2016, pp. 3885–3914.

128. Y. Zhao, B. Ahmed, T. Thesen, K. E. Blackmon, J. G. Dy, C. E. Brodley, R. Kuzniekcy, and O. Devinsky. "A Non-parametric Approach to Detect Epileptogenic Lesions using Restricted Boltzmann Machines". In: *the 22nd ACM SIGKDD International Conference*. ACM Press, New York, New York, USA, 2016, pp. 373–382.

129. R. S. Gill, S.-J. Hong, F. Fadaie, B. Caldairou, B. Bernhardt, N. Bernasconi, and A. Bernasconi. "Automated Detection of Epileptogenic Cortical Malformations Using Multimodal MRI". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Ed. by M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. S. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, and Z. Lu. Springer International Publishing, Cham, 2017, pp. 349–356. ISBN: 978-3-319-67558-9. DOI: 10.1007/978-3-319-67558-9_40.

130. S.-J. 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* 83:1, 2014, pp. 48–55.

131. G. D. Cascino, C. R. Jack, J. E. Parisi, F. W. Sharbrough, K. A. Hirschorn, F. B. Meyer, W. R. Marsh, and P. C. O'Brien. "Magnetic resonance imaging-based volume studies in temporal lobe epilepsy: pathological correlations." *Annals of Neurology* 30:1, 1991, pp. 31–36.

132. J. D. Rudie, J. B. Colby, and N. Salamon. "Machine learning classification of mesial temporal sclerosis in epilepsy patients." *Epilepsy Research* 117, 2015, pp. 63–69.

133. M. Goubran, B. C. Bernhardt, D. Cantor-Rivera, J. C. Lau, C. Blinston, R. R. Hammond, S. de Ribaupierre, J. G. Burneo, S. M. Mirsattari, D. A. Steven, A. G. Parrent, A. Bernasconi, N. Bernasconi, T. M. Peters, and A. R. Khan. "In vivo MRI signatures of hippocampal subfield pathology in intractable epilepsy." *Human Brain Mapping* 37:3, 2016, pp. 1103–1119.

134. B. C. Bernhardt, S.-J. Hong, A. Bernasconi, and N. Bernasconi. "Magnetic resonance imaging pattern learning in temporal lobe epilepsy: classification and prognostics." *Annals of Neurology* 77:3, 2015, pp. 436–446.

135. I. Beheshti, D. Sone, N. Maikusa, Y. Kimura, Y. Shigemoto, N. Sato, and H. Matsuda. "FLAIR-Wise Machine-Learning Classification and Lateralization of MRI-Negative 18F-FDG PET-Positive Temporal Lobe Epilepsy." *Frontiers in neurology* 11, 2020, p. 580713.

136. I. Blümcke, M. Thom, E. Aronica, D. D. Armstrong, F. Bartolomei, A. Bernasconi, N. Bernasconi, C. G. Bien, F. Cendes, R. Coras, J. H. Cross, T. S. Jacques, P. Kahane, G. W. Mathern, H. Miyata, S. L. Moshé, B. Oz, C. Özkara, E. Perucca, S. Sisodiya, S. Wiebe, and R. Spreafico. "International consensus classification of hippocampal sclerosis in temporal lobe epilepsy: a Task Force report from the ILAE Commission on Diagnostic Methods." *Epilepsia* 54:7, 2013, pp. 1315–1329.

137. E. Hattingen, S. J. Enkirch, A. Jurcoane, M. Kruse, D. Delev, A. Grote, and A. Becker. "Hippocampal "gliosis only" on MR imaging represents a distinct entity in epilepsy patients." *Neuroradiology* 60:2, 2018, pp. 161–168.

138. A. Bernasconi and N. Bernasconi. "Unveiling epileptogenic lesions: the contribution of image processing." *Epilepsia* 52 Suppl 4:s4, 2011, pp. 20–24.

139. N. Bernasconi, A. Bernasconi, Z. Caramanos, S. B. Antel, F. Andermann, and D. L. Arnold. "Mesial temporal damage in temporal lobe epilepsy: a volumetric MRI study of the hippocampus, amygdala and parahippocampal region." *Brain* 126:Pt 2, 2003, pp. 462–469.

140. R. Kuzniecky, E. Bilir, F. Gilliam, E. Faught, R. Martin, and J. Hugg. "Quantitative MRI in temporal lobe epilepsy: evidence for fornix atrophy." *Neurology* 53:3, 1999, pp. 496–501.

141. G. D. Jackson, A. Connelly, J. S. Duncan, R. A. Grünewald, and D. G. Gadian. "Detection of hippocampal pathology in intractable partial epilepsy: increased sensitivity with quantitative magnetic resonance T2 relaxometry." *Neurology* 43:9, 1993, pp. 1793–1799.

142. A. C. Coan, B. Kubota, F. P. G. Bergo, B. M. Campos, and F. Cendes. "3T MRI quantification of hippocampal volume and signal in mesial temporal lobe epilepsy improves detection of hippocampal sclerosis." *AJNR. American journal of neuroradiology* 35:1, 2014, pp. 77–83.

143. A. Bernasconi, N. Bernasconi, Z. Caramanos, D. C. Reutens, F. Andermann, F. Dubeau, D. Tampieri, B. G. Pike, and D. L. Arnold. "T2 relaxometry can lateralize mesial temporal lobe epilepsy in patients with normal MRI." *NeuroImage* 12:6, 2000, pp. 739–746.

144. H. Kim, T. Mansi, N. Bernasconi, and A. Bernasconi. "Surface-based multi-template automated hippocampal segmentation: application to temporal lobe epilepsy." *Medical Image Analysis* 16:7, 2012, pp. 1445–1455.

145. B. C. Bernhardt, A. Bernasconi, M. Liu, S.-J. Hong, B. Caldairou, M. Goubran, M. C. Guiot, J. Hall, and N. Bernasconi. "The spectrum of structural and functional imaging abnormalities in temporal lobe epilepsy." *Annals of Neurology* 80:1, 2016, pp. 142–153.

146. L. Maccotta, E. D. Moseley, T. L. Benzinger, and R. E. Hogan. "Beyond the CA1 subfield: Local hippocampal shape changes in MRI-negative temporal lobe epilepsy." *Epilepsia* 56:5, 2015, pp. 780–788.

147. M. Styner, I. Oguz, S. Xu, C. Brechbühler, D. Pantazis, J. J. Levitt, M. E. Shenton, and G. Gerig. "Framework for the Statistical Shape Analysis of Brain Structures using SPHARM-PDM." *The insight journal* 1071, 2006, pp. 242–250.

148. P. A. Yushkevich et al. "Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: towards a harmonized segmentation protocol." *NeuroImage* 111, 2015, pp. 526–541.

149. L. E. M. Wisse et al. "A harmonized segmentation protocol for hippocampal and parahippocampal subregions: Why do we need one and what are the key goals?" *Hippocampus* 27:1, 2017, pp. 3–11.

150. J. L. Winterburn, J. C. Pruessner, S. Chavez, M. M. Schira, N. J. Lobaugh, A. N. Voineskos, and M. M. Chakravarty. "A novel in vivo atlas of human hippocampal subfields using high-resolution 3 T magnetic resonance imaging." *NeuroImage* 74, 2013, pp. 254–265.

151. J. Kulaga-Yoskovitz, B. C. Bernhardt, S.-J. Hong, T. Mansi, K. E. Liang, A. J. W. van der Kouwe, J. Smallwood, A. Bernasconi, and N. Bernasconi. "Multi-contrast submillimetric 3 Tesla hippocampal subfield segmentation protocol and dataset." *Scientific data* 2:1, 2015, p. 150059.

152. K. Van Leemput, A. Bakkour, T. Benner, G. Wiggins, L. L. Wald, J. Augustinack, B. C. Dickerson, P. Golland, and B. Fischl. "Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI." *Hippocampus* 19:6, 2009, pp. 549–557.

153. J. Pipitone, M. T. M. Park, J. Winterburn, T. A. Lett, J. P. Lerch, J. C. Pruessner, M. Lepage, A. N. Voineskos, M. M. Chakravarty, and Alzheimer's Disease Neuroimaging Initiative. "Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates." *NeuroImage* 101, 2014, pp. 494–512.

154. J. E. Iglesias, J. C. Augustinack, K. Nguyen, C. M. Player, A. Player, M. Wright, N. Roy, M. P. Frosch, A. C. McKee, L. L. Wald, B. Fischl, K. Van Leemput, and Alzheimer's Disease Neuroimaging Initiative. "A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI." *NeuroImage* 115, 2015, pp. 117–137.

155. P. A. Yushkevich, J. B. Pluta, H. Wang, L. Xie, S.-L. Ding, E. C. Gertje, L. Mancuso, D. Kliot, S. R. Das, and D. A. Wolk. "Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment." *Human Brain Mapping* 36:1, 2015, pp. 258–287.

156. J. E. Romero, P. Coupé, and J. V. Manjón. "HIPS: A new hippocampus subfield segmentation method." *NeuroImage* 163, 2017, pp. 286–295.

157. R. Giraud, V.-T. Ta, N. Papadakis, J. V. Manjón, D. L. Collins, P. Coupé, and Alzheimer's Disease Neuroimaging Initiative. "An Optimized PatchMatch for multi-scale and multi-feature label fusion." *NeuroImage* 124:Pt A, 2016, pp. 770–782.

158. B. Caldairou, B. C. Bernhardt, J. Kulaga-Yoskovitz, H. Kim, N. Bernasconi, and A. Bernasconi. "A surface patch-based segmentation method for hippocampal subfields". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 379–387.

159. J. E. Peixoto-Santos, L. Kandratavicius, T. R. Velasco, J. A. Assirati, C. G. Carlotti, R. C. Scandiuzzi, C. E. G. Salmon, A. C. D. Santos, and J. P. Leite. "Individual hippocampal subfield assessment indicates that matrix macromolecules and gliosis are key elements for the increased T2 relaxation time seen in temporal lobe epilepsy." *Epilepsia* 58:1, 2017, pp. 149–159.

160. L. Tassi. "Focal cortical dysplasia: neuropathological subtypes, EEG, neuroimaging and surgical outcome". *Brain* 125:8, 2002, pp. 1719–1732.

161. P. Widdess-Walsh, L. Jeha, D. Nair, P. Kotagal, W. Bingaman, and I. Najm. "Subdural electrode analysis in focal cortical dysplasia: predictors of surgical outcome." *Neurology* 69:7, 2007, pp. 660–667.

162. E. Hedegärd, J. Bjellvi, A. Edelvik, B. Rydenhag, R. Flink, and K. Malmgren. "Complications to invasive epilepsy surgery workup with subdural and depth electrodes: a prospective population-based observational study." *Journal of neurology, neurosurgery, and psychiatry* 85:7, 2014, pp. 716–720.

163. L. E. Jeha, I. Najm, W. Bingaman, D. Dinner, P. Widdess-Walsh, and H. Lüders. "Surgical outcome and prognostic factors of frontal lobe epilepsy surgery." *Brain: a journal of neurology* 130:Pt 2, 2007, pp. 574–584.

164. D. Sone. "Making the Invisible Visible: Advanced Neuroimaging Techniques in Focal Epilepsy." *Frontiers in Neuroscience* 15, 2021, p. 699176.

165. Z. I. Wang, P. Suwanpakdee, S. E. Jones, Z. Jaisani, A. N. V. Moosa, I. M. Najm, F. von Podewils, R. C. Burgess, B. Krishnan, R. A. Prayson, J. A. Gonzalez-Martinez, W. Bingaman, and A. V. Alexopoulos. "Re-review of MRI with post-processing in nonlesional patients in whom epilepsy surgery has failed." *Journal of neurology* 263:9, 2016, pp. 1736–1745.

166. I. Wang and A. Alexopoulos. "MRI postprocessing in presurgical evaluation." *Current opinion in neurology* 29:2, 2016, pp. 168–174.

167. A. Bernasconi, F. Cendes, W. H. Theodore, R. S. Gill, M. J. Koepp, R. E. Hogan, G. D. Jackson, P. Federico, A. Labate, A. E. Vaudano, I. Blümcke, P. Ryvlin, and N. Bernasconi. "Recommendations for the use of structural magnetic resonance imaging in the care of

patients with epilepsy: A consensus report from the International League Against Epilepsy Neuroimaging Task Force." *Epilepsia* 73:85, 2019, p. 464.

168. I. Blümcke, A. Arzimanoglou, S. Beniczky, and S. Wiebe. "Roadmap for a competency-based educational curriculum in epileptology: report of the Epilepsy Education Task Force of the International League Against Epilepsy." *Epileptic disorders : international epilepsy journal with videotape* 21:2, 2019, pp. 129–140.

169. S. Beniczky, I. Blümcke, S. Rampp, P. Shisler, E. Biesel, and S. Wiebe. "e-learning comes of age: Web-based education provided by the International League Against Epilepsy." *Epileptic disorders : international epilepsy journal with videotape* 22:3, 2020, pp. 237–244.

170. J. Wagner, B. Weber, H. Urbach, C. E. Elger, and H.-J. Huppertz. "Morphometric MRI analysis improves detection of focal cortical dysplasia type II." *Brain: a journal of neurology* 134:Pt 10, 2011, pp. 2844–2854.

171. Z. I. Wang, S. E. Jones, Z. Jaisani, I. M. Najm, R. A. Prayson, R. C. Burgess, B. Krishnan, A. Ristic, C. H. Wong, W. Bingaman, J. A. Gonzalez-Martinez, and A. V. Alexopoulos. "Voxel-based morphometric magnetic resonance imaging (MRI) postprocessing in MRI-negative epilepsies." *Annals of Neurology* 77:6, 2015, pp. 1060–1075.

172. S. Jayalakshmi, S. K. Nanda, S. Vooturi, R. Vadapalli, P. Sudhakar, S. Madigubba, and M. Panigrahi. "Focal Cortical Dysplasia and Refractory Epilepsy: Role of Multimodality Imaging and Outcome of Surgery." *AJNR. American journal of neuroradiology* 40:5, 2019, pp. 892–898.

173. I. S. N. Berkeley. "The Curious Case of Connectionism". *Open Philosophy* 2:1, 2019, pp. 190–205.

174. C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. "Intriguing properties of neural networks". *arXiv.org*, 2013. arXiv: 1312.6199v4 [cs.CV].

175. A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger. Curran Associates, Inc., 2012.

176. K. Simonyan and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". *arXiv.org*, 2014. arXiv: 1409.1556v6 [cs.CV].

177. Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning". *Nature* 521:7553, 2015, pp. 436–444.

178. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez. "A survey on deep learning in medical image analysis." *Medical Image Analysis* 42, 2017, pp. 60–88.

179. E. J. Topol. "High-performance medicine: the convergence of human and artificial intelligence." *Nature medicine* 25:1, 2019, pp. 44–56.

180. I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

181. R. Aggarwal, V. Sounderajah, G. Martin, D. S. W. Ting, A. Karthikesalingam, D. King, H. Ashrafian, and A. Darzi. "Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis." *NPJ digital medicine* 4:1, 2021, p. 65.

182. A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher. "Deep learning-enabled medical computer vision." *NPJ digital medicine* 4:1, 2021, p. 5.

183. A. L. Beam and I. S. Kohane. "Big Data and Machine Learning in Health Care". *JAMA*, 2018.

184. T. Syeda-Mahmood. "Role of Big Data and Machine Learning in Diagnostic Decision Support in Radiology". *Journal of the American College of Radiology* 15:3, 2018, pp. 569–576.

185. A. Rajkomar, J. Dean, and I. Kohane. "Machine Learning in Medicine." *The New England journal of medicine* 380:14, 2019, pp. 1347–1358.

186. M. L. Martini, A. A. Valliani, C. Sun, A. B. Costa, S. Zhao, F. Panov, S. Ghatan, K. Rajan, and E. K. Oermann. "Deep anomaly detection of seizures with paired stereoelectroencephalography and video recordings." *Scientific reports* 11:1, 2021, p. 7482.

187. K.-O. Cho and H.-J. Jang. "Comparison of different input modalities and network structures for deep learning-based seizure detection." *Scientific reports* 10:1, 2020, p. 122.

188. D. Ahmedt-Aristizabal, C. Fookes, K. Nguyen, S. Denman, S. Sridharan, and S. Dionisio. "Deep facial analysis: A new phase I epilepsy evaluation using computer vision." *Epilepsy & behavior: E&B* 82, 2018, pp. 17–24.

189. R. S. Gill, S.-J. Hong, F. Fadaie, B. Caldairou, B. C. Bernhardt, C. Barba, A. Brandt, V. C. Coelho, L. d'Incerti, M. Lenge, 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, pp. 490–497. DOI: 10.1007/978-3-030-00931-1_56.

190. H. Wang, S. N. Ahmed, and M. Mandal. "Automated detection of focal cortical dysplasia using a deep convolutional neural network." *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 79:2, 2020, p. 101662.

191. E. Gleichgerrcht et al. "Artificial intelligence for classification of temporal lobe epilepsy with ROI-level MRI data: A worldwide ENIGMA-Epilepsy study." *NeuroImage: Clinical* 31, 2021, p. 102765.

192. E. D. Smolyansky, H. Hakeem, Z. Ge, Z. Chen, and P. Kwan. "Machine learning models for decision support in epilepsy management: A critical review." *Epilepsy & behavior: E&B* 123, 2021, p. 108273.

193. J. S. Duncan, G. P. Winston, M. J. Koepp, and S. Ourselin. "Brain imaging in the assessment for epilepsy surgery." *The Lancet. Neurology* 15:4, 2016, pp. 420–433.

194. R. S. Gill, B. Caldairou, N. Bernasconi, and A. Bernasconi. "Uncertainty-informed detection of epileptogenic brain malformations using bayesian neural networks". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2019, pp. 225–233. DOI: 10.1007/978-3-030-32251-9_25.

195. K. M. Bijay Dev, P. S. Jogi, S. Niyas, S. Vinayagamani, C. Kesavadas, and J. Rajan. "Automatic detection and localization of Focal Cortical Dysplasia lesions in MRI using fully convolutional neural network". *Biomedical Signal Processing and Control* 52:2, 2019, pp. 218–225.

196. E. Thomas, S. J. Pawan, S. Kumar, A. Horo, S. Niyas, S. Vinayagamani, C. Kesavadas, and J. Rajan. "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* PP, 2020, pp. 1–1.

197. A. G. Wilson and P. Izmailov. "Bayesian Deep Learning and a Probabilistic Perspective of Generalization". *arXiv.org*, 2020. arXiv: 2002.08791v1 [cs.LG].

198. A. Kendall and Y. Gal. "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" *arXiv.org*, 2017. arXiv: 1703.04977v2 [cs.CV].

199. Z. Ghahramani. "Probabilistic machine learning and artificial intelligence". *Nature* 521:7553, 2015, pp. 452–459.

200. T. Nair, D. Precup, D. L. Arnold, and T. Arbel. "Exploring uncertainty measures in deep networks for Multiple sclerosis lesion detection and segmentation." *Medical Image Analysis* 59, 2020, p. 101557.

201. C. Leibig, V. Allken, M. S. Ayhan, P. Berens, and S. Wahl. "Leveraging uncertainty information from deep neural networks for disease detection." *Scientific reports* 7:1, 2017, p. 17816.

202. A. G. Roy, S. Conjeti, N. Navab, C. Wachinger, and Alzheimer's Disease Neuroimaging Initiative. "Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control." *NeuroImage* 195, 2019, pp. 11–22.

203. R. Tanno, D. E. Worrall, E. Kaden, A. Ghosh, F. Grussu, A. Bizzi, S. N. Sotiropoulos, A. Criminisi, and D. C. Alexander. "Uncertainty Modelling in Deep Learning for Safer Neuroimage Enhancement: Demonstration in Diffusion MRI." *NeuroImage*, 2020, p. 117366.

204. K. Chaudhuri and R. Salakhutdinov, eds. *Direct Uncertainty Prediction for Medical Second Opinions*. PMLR, Long Beach, California, USA, 2019.

205. K. Fabi and J. Schneider. "On Feature Relevance Uncertainty: A Monte Carlo Dropout Sampling Approach". *arXiv.org*, 2020. arXiv: 2008.01468v1 [cs.LG].

206. J. V. Manjón, J. E. Romero, and P. Coupé. "DeepHIPS: A novel Deep Learning based Hippocampus Subfield Segmentation method". *arXiv.org*, 2020. arXiv: 2001.11789v1 [q-bio.QM].

207. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". *arXiv.org*, 2016. arXiv: 1606.06650v1 [cs.CV].

208. M. Zijlmans, W. Zweiphenning, and N. van Klink. "Changing concepts in presurgical assessment for epilepsy surgery." *Nature reviews. Neurology* 27:Suppl. 2, 2019, p. 91.

209. H. Choi, R. L. Sell, L. Lenert, P. Muennig, R. R. Goodman, F. G. Gilliam, and J. B. Wong. "Epilepsy surgery for pharmacoresistant temporal lobe epilepsy: a decision analysis." *JAMA: The Journal of the American Medical Association* 300:21, 2008, pp. 2497–2505.

210. E. L. So and R. W. Lee. "Epilepsy surgery in MRI-negative epilepsies". *Current opinion in neurology* 27:2, 2014, pp. 206–212.

211. W. J. Hader, J. Tellez-Zenteno, A. Metcalfe, L. Hernandez-Ronquillo, S. Wiebe, C.-S. Kwon, and N. Jette. "Complications of epilepsy surgery: a systematic review of focal surgical resections and invasive EEG monitoring." *Epilepsia* 54:5, 2013, pp. 840–847.

212. C. Mellerio, M. A. Labeyrie, F. Chassoux, C. Daumas-Duport, E. Landre, B. Turak, F. X. Roux, J. F. Meder, B. Devaux, and C. Oppenheim. "Optimizing MR Imaging Detection of Type 2 Focal Cortical Dysplasia: Best Criteria for Clinical Practice". *AJNR. American journal of neuroradiology* 33:10, 2012, pp. 1932–1938.

213. K. Noe, V. Sulc, L. Wong-Kisiel, E. Wirrell, J. J. Van Gompel, N. Wetjen, J. Britton, E. So, G. D. Cascino, W. R. Marsh, F. Meyer, D. Horinek, C. Giannini, R. Watson, B. H. Brinkmann, M. Stead, and G. A. Worrell. "Long-term outcomes after nonlesional extratemporal lobe epilepsy surgery." *JAMA Neurology* 70:8, 2013, pp. 1003–1008.

214. Z. I. Wang, A. V. Alexopoulos, S. E. Jones, Z. Jaisani, I. M. Najm, and R. A. Prayson. "The pathology of magnetic-resonance-imaging-negative epilepsy." *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 26:8, 2013, pp. 1051–1058.

215. M. Rados, B. Mouthaan, P. Barsi, D. Carmichael, R. A. Heckemann, A. Kelemen, T. Kobulashvili, G. Kuchukhidze, P. Marusic, K. Minkin, M. Tisdall, E. Trinka, T. Veersema, S. B. Vos, J. Wagner, K. Braun, and P. van Eijsden. "Diagnostic value of MRI in the presurgical evaluation of patients with epilepsy: influence of field strength and sequence selection: a systematic review and meta-analysis from the E-PILEPSY Consortium." *Epileptic disorders : international epilepsy journal with videotape*, 2021.

216. D. Moher. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement". *Annals of Internal Medicine* 151:4, 2009, p. 264.

217. G. Schwarzer. "meta: An R package for meta-analysis". *R news* 7:3, 2007, pp. 40–45.

218. W. Viechtbauer. "Conducting Meta-Analyses in **R** with the metafor Package". *Journal of Statistical Software* 36:3, 2010, pp. 1–48.

219. W. Viechtbauer and M. W.-L. Cheung. "Outlier and influence diagnostics for meta-analysis." *Research synthesis methods* 1:2, 2010, pp. 112–125.

220. H. R. Rothstein, A. J. Sutton, and M. Borenstein. *Publication Bias in Meta-Analysis*. Prevention, Assessment and Adjustments. Wiley-Blackwell, 2005.

221. G. H. Guyatt, A. D. Oxman, G. E. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coello, H. J. Schünemann, and GRADE Working Group. "GRADE: an emerging consensus on rating quality of evidence and strength of recommendations." *BMJ (Clinical research ed.)* 336:7650, 2008, pp. 924–926.

222. H. Balshem, M. Helfand, H. J. Schünemann, A. D. Oxman, R. Kunz, J. Brozek, G. E. Vist, Y. Falck-Ytter, J. Meerpohl, S. Norris, and G. H. Guyatt. "GRADE guidelines: 3. Rating the quality of evidence." *Journal of Clinical Epidemiology* 64:4, 2011, pp. 401–406.

223. M. J. Page, J. A. C. Sterne, J. P. T. Higgins, and M. Egger. "Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review." *Research synthesis methods* 12:2, 2021, pp. 248–259.

224. A. T. Berg, B. G. Vickrey, J. T. Langfitt, M. R. Sperling, T. S. Walczak, S. Shinnar, C. W. Bazil, S. V. Pacia, S. S. Spencer, and Multicenter Study of Epilepsy Surgery. "The multicenter study of epilepsy surgery: recruitment and selection for surgery." *Epilepsia* 44:11, 2003, pp. 1425–1433.

225. G. Alarcon, A. Valentin, C. Watt, R. Selway, M. Lacruz, R. Elwes, J. Jarosz, M. Honavar, F. Brunhuber, N. Mullatti, et al. "Is it worth pursuing surgery for epilepsy in patients with normal neuroimaging?" *Journal of Neurology, Neurosurgery & Psychiatry* 77:4, 2006, pp. 474–480.

226. R. Mosewich, E. L. So, T. O'brien, G. Cascino, F. Sharbrough, W. Mars, F. Meyer, C. Jack, and P. O'Brien. "Factors predictive of the outcome of frontal lobe epilepsy surgery". *Epilepsia* 41:7, 2000, pp. 843–849.

227. M. J. Seong, S. J. Choi, E. Y. Joo, Y.-M. Shon, D.-W. Seo, S. B. Hong, and S. C. Hong. "Surgical outcome and prognostic factors in epilepsy patients with MR-negative focal cortical dysplasia." *PLoS ONE* 16:4, 2021, e0249929.

228. A. Willard, A. Antonic-Baker, Z. Chen, T. J. O'Brien, P. Kwan, and P. Perucca. "Seizure Outcome After Surgery for MRI-Diagnosed Focal Cortical Dysplasia: A Systematic Review and Meta-analysis." *Neurology*, 2021.

229. N. K. Focke, S. B. Bonelli, M. Yogarajah, C. Scott, M. R. Symms, and J. S. Duncan. "Automated normalized FLAIR imaging in MRI-negative patients with refractory focal epilepsy." *Epilepsia* 50:6, 2009, pp. 1484–1490.

230. P. Besson, F. Andermann, F. Dubeau, and A. Bernasconi. "Small focal cortical dysplasia lesions are located at the bottom of a deep sulcus." *Brain: a journal of neurology* 131:Pt 12, 2008, pp. 3246–3255.

231. T. Thesen, B. T. Quinn, C. Carlson, O. Devinsky, J. DuBois, C. R. McDonald, J. French, R. Leventer, O. Felsovalyi, X. Wang, E. Halgren, and R. Kuzniecky. "Detection of Epileptogenic Cortical Malformations with Surface-Based MRI Morphometry". *PLoS ONE* 6:2, 2011, e16430.

232. J. R. Polimeni, B. Fischl, D. N. Greve, and L. L. Wald. "Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1." *NeuroImage* 52:4, 2010, pp. 1334–1346.

233. S.-J. Hong, B. C. Bernhardt, D. Schrader, B. Caldairou, N. Bernasconi, and A. Bernasconi. "MRI-Based Lesion Profiling of Epileptogenic Cortical Malformations". In: *Advanced Lectures on Machine Learning*. Springer International Publishing, Cham, 2015, pp. 501–509.

234. R. Shafee, R. L. Buckner, and B. Fischl. "Gray matter myelination of 1555 human brains using partial volume corrected MRI images." *NeuroImage* 105, 2015, pp. 473–485.

235. K. Im, J. M. Lee, O. Lyttelton, S. H. Kim, A. C. Evans, and S. I. Kim. "Brain size and cortical structure in the adult human brain." *Cerebral cortex* 18:9, 2008, pp. 2181–2191.

236. C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano. "RUSBoost: A Hybrid Approach to Alleviating Class Imbalance". *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 40:1, 2010, pp. 185–197.

237. Y. Freund and R. E. Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". *Journal of Computer and System Sciences* 55:1, 1997, pp. 119–139.

238. P. Geurts, D. Ernst, and L. Wehenkel. "Extremely randomized trees". *Machine Learning* 63:1, 2006, pp. 3–42.

239. S.-J. Hong, B. C. Bernhardt, D. S. Schrader, N. Bernasconi, and A. Bernasconi. "Whole-brain MRI phenotyping in dysplasia-related frontal lobe epilepsy." *Neurology* 86:7, 2016, pp. 643–650.

240. Y.-L. Tan, H. Kim, S. Lee, T. Tihan, L. Ver Hoef, S. G. Mueller, A. J. Barkovich, D. Xu, and R. Knowlton. "Quantitative surface analysis of combined MRI and PET enhances detection of focal cortical dysplasias." *NeuroImage* 166, 2017, pp. 10–18.

241. V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins, and Brain Development Cooperative Group. "Unbiased average age-appropriate atlases for pediatric studies." *NeuroImage* 54:1, 2011, pp. 313–327.

242. S.-J. Hong, B. C. Bernhardt, B. Caldairou, J. A. Hall, M. C. Guiot, D. Schrader, N. Bernasconi, and A. Bernasconi. "Multimodal MRI profiling of focal cortical dysplasia type II." *Neurology* 88:8, 2017, pp. 734–742.

243. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle. "Brain tumor segmentation with Deep Neural Networks". *Medical Image Analysis* 35, 2017, pp. 18–31.

244. S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, A. Oliver, and X. Lladó. "Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach." *NeuroImage* 155, 2017, pp. 159–168.

245. S. Fauser, C. Essang, D.-M. Altenmüller, A. M. Staack, B. J. Steinhoff, K. Strobl, T. Bast, S. Schubert-Bast, U. Stephani, G. Wiegand, M. Prinz, A. Brandt, J. Zentner, and A. Schulze-Bonhage. "Long-term seizure outcome in 211 patients with focal cortical dysplasia." *Epilepsia* 56:1, 2015, pp. 66–76.

246. C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. "On Calibration of Modern Neural Networks". *arXiv.org*, 2017. arXiv: 1706.04599v2 [cs.LG].

247. T. Nair, D. Precup, D. L. Arnold, and T. Arbel. "Exploring Uncertainty Measures in Deep Networks for Multiple Sclerosis Lesion Detection and Segmentation". *arXiv.org*, 2018. arXiv: 1808.01200v1 [cs.CV].

248. Y. Gal and Z. Ghahramani. "Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference". *arXiv.org*, 2015. arXiv: 1506.02158v6 [stat.ML].

249. V. S. Fonov, A. C. Evans, R. C. McKinstry, C. R. Almli, and D. L. Collins. "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood". *NeuroImage* 47, 2009, S102.

250. R. S. Gill, H.-M. Lee, B. Caldairou, S.-J. Hong, C. Barba, F. Deleo, L. D'Incerti, V. C. Mendes Coelho, M. Lenge, M. Semmelroch, D. V. Schrader, F. Bartolomei, M. Guye, A. Schulze-Bonhage, H. Urbach, K. H. 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, 2021, e1571–e1582. DOI: 10.1212/WNL.0000000000012698.

251. K. Snyder, E. P. Whitehead, W. H. Theodore, K. A. Zaghloul, S. J. Inati, and S. K. Inati. "Distinguishing Type II Focal Cortical Dysplasias from Normal Cortex: A Novel Normative Modeling Approach". *NeuroImage: Clinical* 15:2, 2021, p. 102565.

252. G. T. Sica. "Bias in research studies." *Radiology* 238:3, 2006, pp. 780–789.

253. E. J. Ashman and G. S. Gronseth. "Level of evidence reviews: three years of progress." *Neurology* 79:1, 2012, pp. 13–14.

254. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". *Journal of Machine Learning Research* 15, 2014, pp. 1929–1958.

255. R. Guerrini, M. Duchowny, P. Jayakar, P. Krsek, P. Kahane, L. Tassi, F. Melani, T. Polster, V. M. André, C. Cepeda, D. A. Krueger, J. H. Cross, R. Spreafico, M. Cosottini, J. Gotman, F. Chassoux, P. Ryvlin, F. Bartolomei, A. Bernasconi, H. Stefan, I. Miller, B. Devaux, I. Najm, F. Giordano, K. Vonck, C. Barba, and I. Blümcke. "Diagnostic methods and treatment options for focal cortical dysplasia". *Epilepsia* 56:11, 2015, pp. 1669–1686.

256. R. Nieuwenhuys. "The insular cortex: a review." *Progress in brain research* 195, 2012, pp. 123–163.

257. N. Gogolla. "The insular cortex". *Current Biology* 27:12, 2017, R580–R586.

258. J. H. Gilmore, R. C. Knickmeyer, and W. Gao. "Imaging structural and functional brain development in early childhood." *Nature reviews. Neuroscience* 19:3, 2018, pp. 123–137.

259. S. Fauser, S. M. Sisodiya, L. Martinian, M. Thom, C. Gumbinger, H.-J. Huppertz, C. Hader, K. Strobl, B. J. Steinhoff, M. Prinz, J. Zentner, and A. Schulze-Bonhage. "Multifocal occurrence of cortical dysplasia in epilepsy patients." *Brain: a journal of neurology* 132:Pt 8, 2009, pp. 2079–2090.

260. R. J. Slegers and I. Blümcke. "Low-grade developmental and epilepsy associated brain tumors: a critical update 2020." *Acta neuropathologica communications* 8:1, 2020, p. 27.

261. G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren. "Secure, privacy-preserving and federated machine learning in medical imaging". *Nature Machine Intelligence* 2:6, 2020, pp. 305–311.

262. V. Dill, A. R. Franco, and M. S. Pinho. "Automated methods for hippocampus segmentation: the evolution and a review of the state of the art." *Neuroinformatics* 13:2, 2015, pp. 133–150.

263. M. K. Singh and K. K. Singh. "A Review of Publicly Available Automatic Brain Segmentation Methodologies, Machine Learning Models, Recent Advancements, and Their Comparison." *Annals of neurosciences* 28:1-2, 2021, pp. 82–93.

264. N. Bernasconi, A. Bernasconi, F. Andermann, F. Dubeau, W. Feindel, and D. C. Reutens. "Entorhinal cortex in temporal lobe epilepsy: a quantitative MRI study." *Neurology* 52:9, 1999, pp. 1870–1876.

265. F. Cendes, F. Andermann, P. Gloor, A. Evans, M. Jones-Gotman, C. Watson, D. Melanson, A. Olivier, T. Peters, and I. Lopes-Cendes. "MRI volumetric measurement of amygdala and hippocampus in temporal lobe epilepsy." *Neurology* 43:4, 1993, pp. 719–725.

266. H. Kim, P. Besson, O. Colliot, A. Bernasconi, and N. Bernasconi. "Surface-based vector analysis using heat equation interpolation: a new approach to quantify local hippocampal volume changes". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2008, pp. 1008–1015.

267. J. Yang and J. S. Duncan. "3D image segmentation of deformable objects with joint shape-intensity prior models using level sets." *Medical Image Analysis* 8:3, 2004, pp. 285–294.

268. A. Pitiot, H. Delingette, P. M. Thompson, and N. Ayache. "Expert knowledge-guided segmentation system for brain MRI." *NeuroImage* 23 Suppl 1, 2004, S85–96.

269. S. Duchesne, J. Pruessner, and D. L. Collins. "Appearance-based segmentation of medial temporal lobe structures." *NeuroImage* 17:2, 2002, pp. 515–531.

270. A. R. Khan, L. Wang, and M. F. Beg. "FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using Large Deformation Diffeomorphic Metric Mapping." *NeuroImage* 41:3, 2008, pp. 735–746.

271. H. Wang, S. R. Das, J. W. Suh, M. Altinay, J. Pluta, C. Craige, B. Avants, P. A. Yushkevich, and Alzheimer's Disease Neuroimaging Initiative. "A learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex and brain segmentation." *NeuroImage* 55:3, 2011, pp. 968–985.

272. P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. "Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy." *NeuroImage* 46:3, 2009, pp. 726–738.

273. D. Pustina, B. Avants, M. Sperling, R. Gorniak, X. He, G. Doucet, P. Barnett, S. Mintzer, A. Sharan, and J. Tracy. "Predicting the laterality of temporal lobe epilepsy from PET, MRI, and DTI: A multimodal study." *NeuroImage: Clinical* 9, 2015, pp. 20–31.

274. D. L. Collins and J. C. Pruessner. "Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion." *NeuroImage* 52:4, 2010, pp. 1355–1366.

275. J. E. Iglesias, J. C. Augustinack, K. Nguyen, C. M. Player, A. Player, M. Wright, N. Roy, M. P. Frosch, A. C. McKee, L. L. Wald, B. Fischl, K. Van Leemput, and Alzheimer's Disease Neuroimaging Initiative. "A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI." *NeuroImage* 115, 2015, pp. 117–137.

276. H. Zhu, F. Shi, L. Wang, S.-C. Hung, M.-H. Chen, S. Wang, W. Lin, and D. Shen. "Dilated Dense U-Net for Infant Hippocampus Subfield Segmentation." *Frontiers in Neuroinformatics* 13, 2019, p. 30.

277. M. Goubran, E. E. Ntiri, H. Akhavein, M. Holmes, S. Nestor, J. Ramirez, S. Adamo, M. Ozzoude, C. Scott, F. Gao, A. Martel, W. Swardfager, M. Masellis, R. Swartz, B. MacIntosh, and S. E. Black. "Hippocampal segmentation for brains with extensive atrophy using three-dimensional convolutional neural networks." *Human Brain Mapping* 32:4, 2019, p. 288.

278. J. V. Manjón and P. Coupé. "Hippocampus Subfield Segmentation Using a Patch-Based Boosted Ensemble of Autocontext Neural Networks". In: *Advanced Lectures on Machine Learning*. Springer International Publishing, Cham, 2017, pp. 29–36.

279. L. Fang, L. Zhang, D. Nie, X. Cao, I. Rekik, S.-W. Lee, H. He, and D. Shen. "Automatic brain labeling via multi-atlas guided fully convolutional networks." *Medical Image Analysis* 51, 2019, pp. 157–168.

280. A. Buades, B. Coll, and J. M. Morel. "A Review of Image Denoising Algorithms, with a New One". *Multiscale Modeling & Simulation* 4:2, 2005, pp. 490–530.

281. O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". *arXiv.org*, 2015. arXiv: 1505.04597v1 [cs.CV].

282. V. Nair and G. E. Hinton. "Rectified linear units improve restricted boltzmann machines". In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010, pp. 807–814.

283. S. Ioffe and C. Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". *arXiv.org*, 2015. arXiv: 1502.03167v3 [cs.LG].

284. J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu. "Squeeze-and-Excitation Networks". *arXiv.org*, 2017. arXiv: 1709.01507v4 [cs.CV].

285. A. G. Roy, N. Navab, and C. Wachinger. "Concurrent Spatial and Channel 'Squeeze & Excitation' in Fully Convolutional Networks". In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Springer International Publishing, Cham, 2018, pp. 421–429. ISBN: 978-3-030-00928-1.

286. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. "Focal Loss for Dense Object Detection". *arXiv.org*, 2017. arXiv: 1708.02002v2 [cs.CV].

287. K. He, X. Zhang, S. Ren, and J. Sun. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification". *arXiv.org*, 2015. arXiv: 1502.01852v1 [cs.CV].

288. D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". *arXiv.org*, 2014. arXiv: 1412.6980v9 [cs.LG].

289. B. Caldairou, N. A. Foit, C. Mutti, F. Fadaie, R. Gill, H.-M. Lee, T. Demerath, H. Urbach, A. Schulze-Bonhage, A. Bernasconi, and N. Bernasconi. "An MRI-Based Machine Learning Prediction Framework to Lateralize Hippocampal Sclerosis in Patients With Temporal Lobe Epilepsy." *Neurology*, 2021.

290. M. Zaitsev, J. Maclaren, and M. Herbst. "Motion artifacts in MRI: A complex problem with many partial solutions." *Journal of magnetic resonance imaging : JMRI* 42:4, 2015, pp. 887–901.

291. L. Cao, L. Li, J. Zheng, X. Fan, F. Yin, H. Shen, and J. Zhang. "Multi-task neural networks for joint hippocampus segmentation and clinical score regression". *Multimedia Tools and Applications* 77:22, 2018, pp. 29669–29686.

292. M. Crawshaw. "Multi-task learning with deep neural networks: A survey". *arXiv preprint arXiv:2009.09796*, 2020.

293. A. Bernasconi and N. Bernasconi. "The Role of MRI in the Treatment of Drug-Resistant Focal Epilepsy." *European neurology*, 2022, pp. 1–9.

294. G. B. Moody and R. G. Mark. "The impact of the MIT-BIH arrhythmia database." *IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society* 20:3, 2001, pp. 45–50.

295. F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." *Nature methods* 18:2, 2021, pp. 203–211.

296. N. J. Tustison, P. A. Cook, A. J. Holbrook, H. J. Johnson, J. Muschelli, G. A. Devenyi, J. T. Duda, S. R. Das, N. C. Cullen, D. L. Gillen, M. A. Yassa, J. R. Stone, J. C. Gee, and B. B. Avants. "The ANTsX ecosystem for quantitative biological and medical imaging." *Scientific reports* 11:1, 2021, pp. 9068–13.

297. X. Y. Stella and J. Shi. "Multiclass spectral clustering". In: *Computer Vision, IEEE International Conference on*. Vol. 2. IEEE Computer Society. 2003, pp. 313–313.

298. P. J. Rousseeuw. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". *Journal of Computational and Applied Mathematics* 20, 1987, pp. 53–65.

299. C. Shorten and T. M. Khoshgoftaar. “A survey on Image Data Augmentation for Deep Learning”. *Journal of Big Data* 6:1, 2019, p. 1106.

300. M. D. Zeiler. “ADADELTA: An Adaptive Learning Rate Method”. *arXiv.org*, 2012. arXiv: 1212.5701v1 [cs.LG].

301. V. Kuleshov and S. Ermon. “Estimating Uncertainty Online Against an Adversary”. *arXiv.org*, 2016. arXiv: 1607.03594v2 [cs.LG].

302. E. T. Rolls, M. Joliot, and N. Tzourio-Mazoyer. “Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas.” *NeuroImage* 122, 2015, pp. 1–5.