ABSTRACT

Objective. MRI fails to reveal hippocampal pathology in 30-50% of temporal lobe epilepsy (TLE) surgical candidates. To address this clinical challenge, we developed an automated MRI-based classifier that lateralizes the side of covert hippocampal pathology in TLE.

Methods. We trained a surface-based linear discriminant classifier that uses T1-weighted (morphology) and T2-weighted as well as FLAIR/T1 (intensity) features. The classifier was trained on 60 TLE patients (mean age: 35.6; 58% female) with histologically-verified hippocampal sclerosis (HS). Images were deemed as MRI-negative in 42% of cases based on neuroradiological reading (40% based on hippocampal volumetry). The predictive model automatically labelled patients as left or right TLE. Lateralization accuracy was compared to electro-clinical data, including side of surgery. Accuracy of the classifier was further assessed in two independent TLE cohorts with similar demographics and electro-clinical characteristics (n=57; 58% MRI-negative).

Results. The overall lateralization accuracy was 93% (95%; CI 92% - 94%), regardless of HS visibility. In MRI-negative TLE, the combination of T2 and FLAIR/T1 intensities provided the highest accuracy both in the training (84%, area-under-the-curve (AUC): 0.95±0.02) and the validation cohorts (Cohort 1: 90%, AUC: 0.99; Cohort 2: 76%, AUC: 0.94).

Conclusion. This prediction model for TLE lateralization operates on readily available conventional MRI contrasts and offers gain in accuracy over visual radiological assessment. The combined contribution of decreased T1- and increased T2-weighted intensities makes the synthetic FLAIR/T1 contrast particularly effective in MRI-negative HS, setting the basis for broad clinical translation.

INTRODUCTION

Many patients with medically-refractory temporal lobe epilepsy (TLE) present with hippocampal sclerosis (HS) [1]. MRI is instrumental for the identification of this pathology that may form the substrate of the epileptogenic focus, thus streamlining the presurgical evaluation [2]. The main imaging characteristics of HS are loss of hippocampal volume, often associated with hypointense T1- and hyperintense T2-weighted signal. The biological validity of these features has been established through combined MRI-histopathological analyses showing that decreased cell density and gliosis positively correlate with atrophy [3] and T2 hyperintensity [4], respectively. In clinical practice, nevertheless, MRI still fails to reveal hippocampal pathology in 30-50% of surgical candidates with unambiguous electroclinical evidence of TLE [5]. This wide range may be at least partly attributable to suboptimal imaging protocols and limited specialized experience [5]. In addition, while quantitative analyses, including hippocampal volumetry [6], voxel-based morphometry [7], T2 relaxometry [8], and measurements of FLAIR signal intensity [9,10] have been shown to be more sensitive compared to visual evaluation, they remain underused (see [11] for review). Notably, the *in vivo* signature of HS is modulated by the severity of loss of neurons and gliosis [12,13], with subtle forms typified by isolated gliosis [11] often evading detection [1,14]. Many patients with unrevealing MRI may thus undergo intracranial EEG, a procedure that carries risks similar to resective surgery [15] and incurs high costs [16].

Despite the large body of neuroimaging literature assessing hippocampal structural integrity in TLE, the vast majority of studies have addressed group-level changes [5,6,17]. Individual analyses, on the other hand, have commonly exploited single contrasts normalized to the distribution of healthy subjects and rarely addressed the challenge of lateralizing MRI-negative patients [6,18,19]. Our purpose was to implement a machine

learning framework to lateralize HS in individual patients relying on readily available conventional T1- and T2-weighted contrasts [20]. As HS is typically characterized by T1-weighted hypointensity and T2-weighted hyperintensity, we also generated a synthetic contrast by dividing the FLAIR intensity by T1-weighted intensity, thereby maximizing their combined contributions to detect the full HS spectrum. We applied the classifier to MRI features of HS in TLE and assessed generalizability in two independent cohorts.

MATERIALS AND METHODS

We present an algorithm for automated lateralization of the epileptogenic lesion in TLE, with the prediction model labeling patients as left TLE (LTLE) or right TLE (RTLE). In short, we trained a surface-based linear discriminant classifier using volume and signal intensity based on T1- and T2-weighted MRI, and FLAIR/T1 of patients with histologically-verified HS. Lateralization accuracy was compared to electro-clinical data, including side of surgery. To address generalizability, accuracy was tested in two independent validation TLE cohorts with similar electro-clinical and imaging characteristics. Classifiers were cross-validated using a 5-fold scheme with 100 repetitions. Evaluation performance was further assessed via receiver operating characteristics curves (ROC) and area under the curve (AUC).

Subjects

Training cohort. Sixty consecutive patients with TLE and validated HS were collected from 2010 to 2014 retrospectively. All had research-dedicated 3T MRI comprising high-resolution 3D T1-weighted, 3D FLAIR, and 2D T2-weighted images, as well as equivalent clinical MRI sequences as part of the presurgical evaluation. Patients had been investigated for drug-resistant epilepsy with a standard presurgical workup including neurological exam, history of seizures, and EEG telemetry. Clinical neuroradiological diagnosis of HS was based on morphological anomalies of the

hippocampus (atrophy, loss of internal structure, decreased T1 and increased T2). Side-by-side comparison of morphology, including shape, and signal was done on coronal images, while sagittal cuts yielded antero-posterior evaluation, thereby easing the visibility of distribution of signal anomalies.

Validation cohorts. Validation cohorts comprised 56 drug-resistant TLE patients collected from 2015 to 2017. The internal cohort comprised 43 consecutive cases seen at the Montreal Neurological Hospital retrospectively (26 with histologically-validated HS). The external cohort consisted of 14 TLE patients from the Freiburg Epilepsy Center in Germany (13 with histologically-validated HS). Patients in both cohorts underwent the same clinical and MRI evaluation as the Training cohort. Imaging was done either on a 3 Tesla Siemens Trio or Prisma scanners.

Standard protocol approval, registrations, and patients consents

The Research Ethics boards of the Montreal Neurological Hospital and the Freiburg Medical Centre gave approval of the study and all participants gave informed consent.

MR imaging and processing

In the Training cohort and controls, images were acquired on a 3T TimTrio using a 32-arrays coil with the HARNESS-MRI protocol [20], including 3D T1-weighted MPRAGE (TR=2300 ms, TE=2.98 ms, TI =900 ms, flip angle=9˚, matrix=256×256, FOV=256×256mm, yielding 1 cubic mm voxels; 6.35 mins), 3D FLAIR (TR=5000 ms, TE=390 ms, TI=1800 ms, matrix=230×230, FOV=207×207 mm, 0.9 cubic mm; 6.22 mins) and 2D turbo spin echo T2 sequence (TR=10,810 ms, TE=81 ms, flip angle=119°, matrix=512×512, FOV=203×203, 0.4×0.4×2.0 cubic mm; 5.5 mins). The acquisition parameters were similar for the Validation cohorts.

In all subjects, T1-weighted, FLAIR and T2-weighted images underwent field non-uniformity correction, followed by a standardization of signal, as well as alignment to

the ICBM-152 template [21]. FLAIR and T2-weighted images were registered to T1 MRI in MNI-space. A surface-based multi-template algorithm [22] automatically segmented the hippocampus in the CA1-3, CA4-DG and subiculum [23]; this algorithm has shown excellent Dice overlap indices with manual segmentations [22].

Classifier design

Data sampling. We mapped the medial sheet, namely, a surface along the central axis of each hippocampal subfield; this allowed extracting features minimizing effects of partial volume [24]. We obtained the following features on the medial sheets:

a) Columnar Volume. Neuronal loss is associated with atrophy on MRI, which we estimated by calculating columnar volume as previously [24].

b) T2 signal intensity. Gliosis is characterized by increased T2 signal intensity. We computed the relative intensity of T2 at every voxel [25].

c) FLAIR/T1 intensity. To maximize the detection of HS, typically characterized by T1-hypo- and T2-weighted hyper-intensity, we generated a synthetic contrast by dividing the FLAIR by T1-weighted intensity, thereby maximizing their combined contribution.

After z-normalizing each feature with respect to healthy controls, we generated asymmetry maps computed as [2x(left-right)/(left+right)], which served as inputs to the algorithm. The gold standard for training and cross-validation consisted of TLE patients with histologically-validated HS (Training cohort).

Architecture (**Figure 1A**). Our algorithm aims at lateralizing the epileptogenic lesion by assigning patients to either left or right TLE. Lateralization is formulated as a classification task, leveraging Linear Discriminant Analysis (LDA) as the classifier. LDA requires virtually no parameter tuning while operating efficiently in low-dimensional space with limited number of features, as in our case, obviating the need for more complex or nonlinear algorithms with significant time requirement. The

training procedure consists of two components. Firstly, we generate the optimal ROI, namely a spatial constraint for feature sampling and averaging derived from the asymmetry maps of columnar volume, T2 and FLAIR/T1 intensities. More specifically, paired t-tests (or Hotelling's T-squared for feature combinations) contrasted corresponding vertices of the ipsi- and contralateral hippocampal subfields (relative to the epileptogenic focus) to highlight regions exhibiting the largest feature asymmetry. The resulting t-map was then thresholded from zero to the highest absolute t-statistic to generate binarized t-maps. Notably, the choice of the threshold determined the spatial extent of the binarized t-map, and consequently, the discriminative ability of the averaged asymmetry features. Therefore, this binarized t-map (or ROI) served as hyper-parameter optimized through a nested cross-validation procedure. This process selected the most performant model [26] (*i.e.*, the optimal ROI or threshold; **Figure 1B**) while mitigating the propagation of ground truth information across folds (also known as data leakage [27]). Subsequently, for each feature, values sampled on the asymmetry maps (constrained by the optimal ROI) were averaged across the hippocampus. Finally, these values served as inputs to the LDA which determined laterality (*i.e.*, LTLE or RTLE) based on the learned statistical patterns.

Predictors considered for modeling. To evaluate the differential impact of features and their combinations, we tested the lateralization performance of the classifier when using: *i*) columnar volume; *ii*) T2-weighted intensity; *iii*) FLAIR/T1 intensity; and *iv*) a combination of T2 and FLAIR/T1. Notably, volumetry was excluded from the combinatorial analysis since it is not expected to be discriminative for MRI-negative patients, as their hippocampal size is generally within normative range of healthy controls [25].

Performance evaluation. For the Training cohort, performance was assessed through a 5-fold (nested) cross-validation, repeated 100 times. Stratification ensured that each fold had proportional representation of both LTLE and RTLE. Briefly, after randomly splitting the Training cohort into five folds (or partitions), the classifier was trained on four and tested iteratively on the one held-out until all folds had served as a test set; this procedure was repeated 100 times. To assess generalizability, algorithmic performance was tested on two Validation cohorts. To guarantee the highest confidence, we trained the classifier based on a random sample comprising 80% of patients from the Training cohort, repeating the process 100 times.

In designing this classifier, our objective was to determine the side of the pathology (not whether it is present or absent). Our primary performance validation metric was thus accuracy, which reflects the average of two classes (RTLE and LTLE); we did not intend to evaluate the class dichotomy of a TLE patient *versus* a healthy control. We also obtained ROC and AUC curves. Lateralization accuracy and AUC were measured for each 5-fold validation repetition and averaged across them. The Brier score was used as a measure of calibration [28]; values close to zero signify a well-calibrated classifier. Comparisons among experiments were assessed with the Friedman test with Bonferroni correction.

Statistical analysis

Group Analysis. Statistical analysis was carried out with *SurfStat* (Matlab). For each subject, we first z-scored vertex-wise values (columnar volume, normalized T2 intensity, and FLAIR/T1) based on healthy controls. We then sorted the left and right values into ipsilateral and contralateral with respect to the focus. Student test assessed differences between patients and controls, correcting at a family-wise error of $P_{FWE}<0.05$ using random field theory.

Data availability

The source codes for: (1) generating blade surfaces (as described in the *Data Sampling* section); (2) data sampling (intersection of blade surfaces and volumes); (3) computing columnar volumes; and (4) training and testing the classifier are available ~~from the corresponding author.~~ at github.com/NOEL-MNI/Automated_TLE_Lateralization. We also make available the optimal ROI and the pre-trained model (based on the Training cohort) to enable lateralization prediction on a test subject without the need to collect data to train the classifier.

RESULTS

Table 1 details clinical and demographic features of the training and validation cohorts.

Clinical, demographics and imaging characteristics

Training cohort. In 42 patients, the focus side was determined by EEG monitoring with scalp electrodes showing unequivocal temporal lobe seizures onset (and >70% of spikes); in cases with non-localized seizure onset or rapid inter-hemispheric seizure spread (n=18), lateralization was established using stereoencephalography (SEEG). Accordingly, patients were dichotomized into LTLE (n=29; 17 females; mean±SD age=35.6±11 yrs; range=18-59 yrs) and RTLE (n=31; 18 females; 35.5±11 yrs; 17-62 yrs). As per the reading of the neuroradiologist, 35 patients (35/60=58%) had ipsilateral atrophy of the hippocampus together with T2 hypersignal (MRI-positive), while the MRI was reported as unremarkable in 25 (42%; MRI-negative). No other anomalies were seen. Performing volumetry had minimal impact, with only a single MRI-negative patient becoming MRI-positive with an ipsilateral hippocampal volume reduction of -2.5 SD below the mean of healthy controls, bringing the total count to 36 MRI-positive (60%) and 24 (40%) MRI-negative patients.

Based on histopathological review of the resected specimen, 40 patients had severe

neuronal cell loss and astrogliosis: 23 in CA1-3 and CA4 (ILAE HS-1), 9 CA1 predominant (HS-2), 9 CA4 (HS-3); 19 patients showed isolated gliosis [1]. Notably, all MRI-positive patients had HS, while MRI-negative patients presented with both mild HS (5/24=21%; 3 Type-2 and 2 Type-3) and isolated gliosis (19/24=79%). At a follow-up of >2 years, Engel I was reported in 48 (80%) patients, Engel II in 9 (15%), Engel III in 3 (3.3%), Engel IV in one (1.7%). Thirty-six healthy individuals (18 females; 32.2±7.3 yrs; range = 23-53) formed the control group.

Validation cohorts. Based on the same criteria as in the Training cohort, patients were dichotomized into LTLE (n=35; 25 females; mean±SD age=37.2±11 years; range=19-58 years) and RTLE (n=22; 11 females; 36.9±12 years; 18-54 years) based scalp EEG (n=36) and SEEG (n=21). Twenty-four patients (24/57=42%) has ipsilateral hippocampal volume reduction and high T2 (MRI-positive), while the MRI was reported as unremarkable in 33 (58%; MRI-negative). No patient had hippocampal atrophy on volumetry. Thirty-nine patients had surgery. Histopathology showed severe HS in 24 (HS-1=15; HS-2=6; HS-3=3) and isolated gliosis in 15 [1].

Comparisons. We observed no difference among cohorts for age (one-way ANOVA, p=0.73), sex (Chi-Square test, χ^2=0.57, p=0.75), proportion of MRI-positive and MRI-negative patients (χ^2=5.43, p=0.07) and surgical outcome (χ^2=3.59, p=0.17). In all three cohorts, the proportion of isolated gliosis was higher than HS in MRI-negative patients (Training Cohort: χ^2=38.13, p<0.001; Validation Cohort 1: χ^2=13.16, p<0.001; Validation Cohort 2: χ^2=9.12, p=0.002). Proportion of patients who underwent SEEG was higher in MRI-negative patients with respect to MRI-positive patients in the Training Cohort (χ^2=7.62, p=0.006) and in the Validation Cohort 1 (χ^2=4.61, p=0.032). The proportion of LTLE was higher than RTLE in Validation Cohort 1 (χ^2=6.96, p=0.03).

Group analysis

Compared to the healthy subjects (**Figure 2)**, patients exhibited diffuse ipsilateral atrophy across all subfields (CA1-3: t=-2.5, P_{FWE}<0.0001; subiculum: t=-2.2, P_{FWE}<0.0001; CA4-DG: t=2.3, P_{FWE}<0.0001). Moreover, marked ipsilateral T2 hypersignal signal was present in CA1-3 (t=2.8, P_{FWE}<0.0001) and CA4-DG (t=3.9, P_{FWE}<0.0001). FLAIR/T1 was also increased across all subregions (CA1-3: t=3.5, P_{FWE}<0.0001; CA4-DG: t=4.3, P_{FWE}=0.004; subiculum: t=2.4, P_{FWE}<0.0001), with additional subtle increases contralaterally. Anomalies in MRI-positive patients had similar distributions across subfields, albeit more severe (volume: t=-3.9 to -4.4, P_{FWE}<0.0001; T2 signal: t=3.6 to 5.4, P_{FWE}<0.0001; FLAIR/T1: t=2.7 to 6.0, P_{FWE}<0.002). Conversely, MRI-negative patients did not show any volumetric alteration, but subtle T2 increases in the ipsilateral CA4-DG (t=2.5, P_{FWE}<0.0001) and CA1-3 (t=2.1, P_{FWE}=0.014), as well as FLAIR/T1 hyperintensities along all subfields (CA1-3: t=2.5, P_{FWE}<0.0001; CA4-DG: t=2.9, P_{FWE} =0.005; subiculum: t=2.1, P_{FWE}<0.0001).

Performance evaluation

In the Training cohort, the overall lateralization accuracy based on individual features were similar for FLAIR/T1 (85±3%) and T2 signal (86±2%), which were superior to volumetry (77±3%; p<0.05). The combination of T2 and FLAIR/T1 yielded the best overall performance of 93±2%, with an accuracy of 84±5% in MRI-negative and 100±1% in MRI-positive TLE (**Table 2**). Notably, the classifier's lateralization was consistently correct across 90/100 iterations in both MRI-negative and MRI-positive patients. AUC metrics (**Table 3**) confirmed the discriminative capability of FLAIR/T1 (0.80±0.05) and T2 Signal (0.79±0.03) over volumetry (0.46±0.06, $P_{Bonferroni}$<0.05) in MRI-negative patients; the combination of T2 and FLAIR/T1 yielded the highest AUC (0.95±0.03, $P_{Bonferroni}$<0.05). The Brier score demonstrated better calibration of the

classifier when the combination of T2 and FLAIR/T1 is used, relative to other scenarios (0.097±0.016, $P_{Bonferroni}<0.05$). In the Validation cohorts, the combination of T2 and FLAIR/T1 also yielded the best overall performance with >90% lateralization accuracy. Moreover, AUC showed the discriminative capability of this combination in MRI-negative patients (1.00±0.00 and 0.94±0.12 for Validation cohorts 1 and 2, respectively). The Brier score reflected the improved calibration when combining T2 and FLAIR/T1 for Validation cohort 1 relative to single features (0.069±0.015, $P_{Bonferroni}<0.05$), while in Validation cohort 2 similar scores were obtained for combined T2 and FLAIR/T1 (0.152±0.053), and FLAIR/T1 (0.160±0.020). Examples of lateralization predictions are shown in **Figure 3**. **Figure 4** highlights the robustness of combining T2 and FLAIR/T1 to variability in the training dataset, as evidenced by the low variance of the average ROC curve.

DISCUSSION

In epilepsy surgery, converging independent diagnostic tests aim at the lateralization of the epileptic focus based on neurophysiology and MRI localization of the epileptogenic lesion. When either of these two diagnostic pillars are unclear or unrevealing, additional investigations are needed. Invasive EEG recordings may clarify the focus lateralization, for example in cases with non-localized seizure onset or rapid inter-hemispheric seizure spread. Moreover, MRI post-processing and machine learning may clarify lesion location in cases with normal MRI. Here, we present the first decision-support system to lateralize HS in TLE using structural MRI. The classifier relies on automatically-segmented hippocampal subfields and feature sampling derived from conventional T1- and T2-weighted contrasts available on most MR scanners [20]. Our method yields an overall accuracy of >90% regardless of the degree of HS visibility on conventional

MRI. Importantly, it lateralizes MRI-negative TLE with >80% accuracy, offering considerable gain over visual radiological assessment.

Our purpose was to develop a prediction model with the potential to be implemented into clinical practice. Importantly, to ensure a valid biological foundation, we trained the LDA algorithm on well-known, histologically validated features of HS. Notably, when using a conventional z-score comparison with healthy controls, we were able to lateralize pathology in only a fraction (<15%) of MRI-negative TLE patients using either volumetry, T2 or FLAIR/T1. This disparity in performance relative to machine learning can be explained mainly by the use of a spatial ROI allowing for sampling of features relevant to lateralization. This optimal ROI relies on the population-wise feature difference between the ipsi- and contralateral hippocampal subfields, targeting regions with the most marked inter-hemispheric asymmetries. Through the training procedure, the extent of the ROI was adjusted iteratively, thereby boosting lateralization accuracy in single cases. In contrast, a z-score normalization of the features of interest across the whole hippocampus (or each subfield) would likely include non-pertinent information, leading to suboptimal lateralization accuracy. More recent work based on non-conventional contrasts, including diffusion [29] and network parameters [30–32], has targeted whole-brain anomalies in patients with MRI-positive TLE. Only two previous studies have addressed the lateralization challenge in both MRI-positive and MRI-negative TLE [33,34], operating on T1-derived volumetry with groups predefined by side and visibility of hippocampal atrophy on MRI. Notably, in one study [33], besides the lack of histopathological confirmation, anatomical structures identifying TLE groups were mainly outside the mesiotemporal lobe, different across groups and difficult to interpret, particularly in MRI-negative patients. While methodologically appropriate, such a design may be unsatisfactory in a real-world scenario of previously unseen cases.

Our study presents several novelties. Firstly, it uses conventional contrasts [20]. Secondly, we targeted morphological and signal alterations of the hippocampus, the surgically-amenable substrate of mesial TLE. Thirdly, effects of partial volume were minimized by surface-based image processing. For individual prediction, we opted for a LDA, a robust and easy to interpret classifier [35]. Notably, the classifier operated in regions that were identical in both MRI-negative and MRI-positive TLE. From a statistical standpoint, to mitigate optimistic estimates of lateralization accuracy (otherwise known as model overfitting) that would occur using a conventional leave-one-out scheme, we kept the training and testing datasets separate using a cross-validation with 5-folds and 100 repetitions. The effectiveness and generalizability of our algorithm is also evidenced by the high degree of consistency across repetitions, which is further supported by high AUC values in both the training and validation cohorts. Indeed, when applying the classifier to independent datasets of unseen cases including both MRI-positive and MRI-negative patients, accuracy reached 90% correct lateralization. Despite excellent performance, our method presents some limitations. Notably, contrary to previous work [36], our design does not allow discriminating patients from controls. Moreover, our sample size for training may be considered as relatively small with respect to general machine learning standards. Nevertheless, our numbers compare favorably to work in epilepsy. Indeed, prior studies analyzing imaging acquired at 3T were based on a maximum of 80 individuals when both training and testing sets were combined [33,36]. In contrast, our study targeted a significantly larger cohort of patients with 60 individuals for training and 57 for validation.

High performance and generalizability of the classifier across cohorts, scanners and parameters set the basis for wide translation. Nevertheless, successful integration into clinical practice rests on key requirements. In line with recent educational initiatives

[20,37], clinicians should develop competencies in neuroimaging spanning from basic visual diagnostics to the interpretation of advanced post-processing methods and machine learning. The availability of the source code is intended to foster novel synergies between engineers and clinicians within the epilepsy community and should be regarded as a first step towards an online application or a tool integrated into patients' electronic health record.

The imaging correlates of HS associating neuronal loss and gliosis have been long established [3]. At the group-level, all features in our MRI-positive patients were significantly abnormal, including extensive ipsilateral hippocampal volume loss and increased T2-weighted as well as FLAIR/T1 signal intensities compared to healthy controls. Therefore, volumetry was as efficient to map pathology and to lateralize individual patients as any other intensity features. Conversely, MRI-negative cases, the vast majority with isolated gliosis, exhibited only subtle ipsilateral signal anomalies and no volumetric alterations (aside from one with reduced volume). Expectedly, when applying machine learning to volumetry alone, performance was at a chance level, whereas a combination of intensity features derived from T2- and T1-weighted MRI (*i.e.*, T2 together with FLAIR/T1) outperformed any unimodal contrast, thereby offering substantial gain over expert visual assessments. While our results are in agreement with previous studies demonstrating the value of T2 signal in the detection of gliosis [4,19,38], FLAIR/T1 was more effective to lateralize the seizure focus. A possible biological explanation may be that in MRI-negative TLE severe gliosis co-exists with subtle neuronal loss below the 10% sensitivity threshold of qualitative histopathology [39]. In addition, since 30% of these patients had mild HS, it is plausible that 3D FLAIR/T1 maximized the combined sensitivity of both contrasts to detect the full spectrum of hippocampal pathology. From a practical standpoint, our results also

suggest that FLAIR/T1 may be a good alternative to 2D coronal T2-weighted images more prone to movement artifacts.

Non-diagnostic MRIs have led to an increase in invasive studies. By offering a non-invasive decision-support, advanced imaging analysis circumvents some of the limitations and risks related to invasive diagnostics [40], possibly reducing the need for prolonged costly hospitalizations. Conversely, since the presurgical evaluation of drug-resistant TLE is multidisciplinary, an MRI-derived binary lateralization outcome (right *vs.* left) may not be sufficient *per se* for surgical decision-making. However, the increased availability of imaging-derived classification algorithms ought to pave the way for systems that integrate diverse sources of evidence, including other imaging modalities such as PET, as well as electro-clinical data to increase diagnostic yield and certainty. Finally, the methodology presented in this study may be expanded to other epilepsy syndromes associated with HS, including cortical developmental malformations.

REFERENCES

1. Blümcke I, Thom M, Aronica E, et al. International consensus classification of hippocampal sclerosis in temporal lobe epilepsy: A Task Force report from the ILAE Commission on Diagnostic Methods. *Epilepsia*. 2013;54(7):1315-1329. doi:10.1111/epi.12220

2. Jones AL, Cascino GD. Evidence on Use of Neuroimaging for Surgical Treatment of Temporal Lobe Epilepsy: A Systematic Review. *JAMA Neurol*. 2016;73(4):464-470. doi:10.1001/jamaneurol.2015.4996

3. Cascino GD, Jack CR, Parisi JE, et al. Magnetic resonance imaging-based volume studies in temporal lobe epilepsy: pathological correlations. *Ann Neurol*. 1991;30(1):31-36. doi:10.1002/ana.410300107

4. Peixoto-Santos JE, Kandratavicius L, Velasco TR, et al. 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*. 2017;58(1):149-159. doi:10.1111/epi.13620

5. Cendes F, Theodore WH, Brinkmann BH, Sulc V, Cascino GD. Neuroimaging of epilepsy. *Handb Clin Neurol*. 2016;136:985-1014. doi:10.1016/B978-0-444-53486-6.00051-X

6. Vos SB, Winston GP, Goodkin O, et al. Hippocampal profiling: Localized magnetic resonance imaging volumetry and T2 relaxometry for hippocampal sclerosis. *Epilepsia*. 2020;61(2):297-309. doi:10.1111/epi.16416

7. Riederer F, Seiger R, Lanzenberger R, et al. Voxel-Based Morphometry-from Hype to Hope. A Study on Hippocampal Atrophy in Mesial Temporal Lobe Epilepsy. *AJNR Am J Neuroradiol*. 2020;41(6):987-993. doi:10.3174/ajnr.A6545

8. Winston GP, Vos SB, Burdett JL, Cardoso MJ, Ourselin S, Duncan JS. Automated T2 relaxometry of the hippocampus for temporal lobe epilepsy. *Epilepsia*. 2017;58(9):1645-1652. doi:10.1111/epi.13843

9. Focke NK, Bonelli SB, Yogarajah M, Scott C, Symms MR, Duncan JS. Automated normalized FLAIR imaging in MRI-negative patients with refractory focal epilepsy. *Epilepsia*. 2009;50(6):1484-1490. doi:10.1111/j.1528-1167.2009.02022.x

10. Beheshti I, Sone D, Maikusa N, et al. FLAIR-Wise Machine-Learning Classification and Lateralization of MRI-Negative (18)F-FDG PET-Positive Temporal Lobe Epilepsy. *Front Neurol*. 2020;11:580713. doi:10.3389/fneur.2020.580713

11. Martin P, Bender B, Focke NK. Post-processing of structural MRI for individualized diagnostics. *Quant Imaging Med Surg*. 2015;5(2):188-203. doi:10.3978/j.issn.2223-4292.2015.01.10

12. Rudie JD, Colby JB, Salamon N. Machine learning classification of mesial temporal sclerosis in epilepsy patients. *Epilepsy Res*. 2015;117:63-69.

doi:10.1016/j.eplepsyres.2015.09.005

13. Bernhardt BC, Hong S-J, Bernasconi A, Bernasconi N. Magnetic resonance imaging pattern learning in temporal lobe epilepsy: Classification and prognostics. *Ann Neurol*. 2015;77(3):436-446. doi:10.1002/ana.24341

14. Hattingen E, Enkirch SJ, Jurcoane A, et al. Hippocampal "gliosis only" on MR imaging represents a distinct entity in epilepsy patients. *Neuroradiology*. 2018;60(2):161-168. doi:10.1007/s00234-017-1939-3

15. Tanriverdi T, Ajlan A, Poulin N, Olivier A. Morbidity in epilepsy surgery: an experience based on 2449 epilepsy surgery procedures from a single institution. *J Neurosurg*. 2009;110(6):1111-1123. doi:10.3171/2009.8.JNS08338

16. Muhlhofer W, Tan Y-L, Mueller SG, Knowlton R. MRI-negative temporal lobe epilepsy—What do we know? *Epilepsia*. 2017;58(5):727-742. doi:10.1111/epi.13699

17. Caciagli L, Bernasconi A, Wiebe S, Koepp MJ, Bernasconi N, Bernhardt BC. A meta-analysis on progressive atrophy in intractable temporal lobe epilepsy: Time is brain? *Neurology*. 2017;89(5):506-516. doi:10.1212/WNL.0000000000004176

18. Mueller SG, Laxer KD, Barakos J, Cheong I, Garcia P, Weiner MW. Subfield atrophy pattern in temporal lobe epilepsy with and without mesial sclerosis detected by high-resolution MRI at 4 Tesla: Preliminary results. *Epilepsia*. 2009;50(6):1474-1483. doi:10.1111/j.1528-1167.2009.02010.x

19. Kubota BY, Coan AC, Yasuda CL, Cendes F. T2 hyperintense signal in patients with temporal lobe epilepsy with MRI signs of hippocampal sclerosis and in patients with temporal lobe epilepsy with normal MRI. *Epilepsy Behav EB*. 2015;46:103-108. doi:10.1016/j.yebeh.2015.04.001

20. Bernasconi A, Cendes F, Theodore WH, et al. 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*. 2019;60(6):1054-1068. doi:10.1111/epi.15612

21. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*. 2011;54(1):313-327. doi:10.1016/j.neuroimage.2010.07.033

22. Caldairou B, Bernhardt BC, Kulaga-Yoskovitz J, Kim H, Bernasconi N, Bernasconi A. A Surface Patch-Based Segmentation Method for Hippocampal Subfields. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Lecture Notes in Computer Science. Springer International Publishing; 2016:379-387. doi:10.1007/978-3-319-46723-8_44

23. Kulaga-Yoskovitz J, Bernhardt BC, Hong S-J, et al. Multi-contrast submillimetric 3 Tesla hippocampal subfield segmentation protocol and dataset. *Sci Data*. 2015;2:150059. doi:10.1038/sdata.2015.59

24. Kim H, Bernhardt BC, Kulaga-Yoskovitz J, Caldairou B, Bernasconi A, Bernasconi N. Multivariate Hippocampal Subfield Analysis of Local MRI Intensity and Volume: Application to Temporal Lobe Epilepsy. In: Golland P, Hata N, Barillot C, Hornegger J, Howe R, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*. Springer International Publishing; 2014:170-178.

25. Bernhardt BC, Bernasconi A, Liu M, et al. The spectrum of structural and functional imaging abnormalities in temporal lobe epilepsy. *Ann Neurol*. 2016;80(1):142-153. doi:10.1002/ana.24691

26. Hastie T, Tibshirani R, Friedman J. Model Assessment and Selection. In: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer New York; 2009:219-259.

27. Luo W, Phung D, Tran T, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *J Med Internet Res*. 2016;18(12):e323. doi:10.2196/jmir.5870

28. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950;78(1):1-3.

29. Gaizo JD, Mofrad N, Jensen JH, et al. Using machine learning to classify temporal lobe epilepsy based on diffusion MRI. *Brain Behav*. 2017;7(10):e00801. doi:10.1002/brb3.801

30. Morgan VL, Sonmezturk HH, Gore JC, Abou-Khalil B. Lateralization of temporal lobe epilepsy using resting functional Magnetic Resonance Imaging connectivity of hippocampal networks. *Epilepsia*. 2012;53(9):1628-1635. doi:10.1111/j.1528-1167.2012.03590.x

31. Yang Z, Choupan J, Reutens D, Hocking J. Lateralization of Temporal Lobe Epilepsy Based on Resting-State Functional Magnetic Resonance Imaging and Machine Learning. *Front Neurol*. 2015;6:184. doi:10.3389/fneur.2015.00184

32. Barron DS, Fox PT, Pardoe H, et al. Thalamic functional connectivity predicts seizure laterality in individual TLE patients: Application of a biomarker development strategy. *NeuroImage Clin*. 2015;7:273-280. doi:10.1016/j.nicl.2014.08.002

33. Keihaninejad S, Heckemann RA, Gousias IS, et al. Classification and lateralization of temporal lobe epilepsies with and without hippocampal atrophy based on whole-brain automatic MRI segmentation. *PloS One*. 2012;7(4):e33096. doi:10.1371/journal.pone.0033096

34. Hadar PN, Kini LG, Coto C, et al. Clinical validation of automated hippocampal segmentation in temporal lobe epilepsy. *NeuroImage Clin*. 2018;20:1139-1147. doi:10.1016/j.nicl.2018.09.032

35. Wang S, Summers RM. Machine Learning and Radiology. *Med Image Anal*. 2012;16(5):933-951. doi:10.1016/j.media.2012.02.005

36. Mo J, Liu Z, Sun K, et al. Automated detection of hippocampal sclerosis using clinically empirical and radiomics features. *Epilepsia*. 2019;60(12):2519-2529. doi:10.1111/epi.16392

37. Blümcke I, Arzimanoglou A, Beniczky S, Wiebe S. Roadmap for a competency-based educational curriculum in epileptology: report of the Epilepsy Education Task Force of the International League Against Epilepsy. *Epileptic Disord Int Epilepsy J Videotape*. 2019;21(2):129-140. doi:10.1684/epd.2019.1039

38. Goubran M, Rudko DA, Santyr B, et al. In vivo normative atlas of the hippocampal subfields using multi-echo susceptibility imaging at 7 Tesla. *Hum Brain Mapp*. 2014;35(8):3588-3601. doi:10.1002/hbm.22423

39. Wyler AR, Dohan FC, Schweitzer JB, Berry AD. A grading system for mesial temporal pathology (hippocampal sclerosis) from anterior temporal lobectomy. *J Epilepsy*. 1992;5(4):220-225. doi:https://doi.org/10.1016/S0896-6974(05)80120-3

40. Cardinale F, Rizzi M, Vignati E, et al. Stereoelectroencephalography: retrospective analysis of 742 procedures in a single centre. *Brain*. 2019;142(9):2688-2704. doi:10.1093/brain/awz196

Table 1. *Training* and *Validation* cohorts characteristics.

	Age	**Female**	**SEEG**	**TLE(L/R)**	**HS/G**	**Engel-I**
Training						
MRI pos (36)	36±11	53%	17%	17/19	36/0	89%
MRI neg (24)	35±11	67%	50%*	12/12	5/19*	67%
Validation 1						
MRI pos (16)	39±12	75%	19%	10/6*	13/0	92%
MRI neg (27)	36±10	60%	52%*	20/7*	3/10*	70%
Validation 2						
MRI pos (8)	33±13	50%	25%	4/4	8/0	100%
MRI neg (6)	42±11	67%	67%	1/5	0/5*	100%

Age: in years ± standard deviation; (n): sample size; neg: negative; pos: positive; MRI +/-: MRI-positive/negative; L/R: EEG lateralization; HS/G: hippocampal sclerosis/isolated gliosis; *: higher proportions in a given category.

Table 2. Lateralization performance (accuracy) across *Training* and *Validation* cohorts.

	VOLUME	**T2 INTENSITY**	**FLAIR/T1**	**T2 + FLAIR/T1**
Training				
MRI pos (36)	97 ± 2% [97% 98%]	96 ± 2% [96% 96%]	95 ± 2% [94% 95%]	100 ± 1%** [100% 100%]
MRI neg (24)	45 ± 7% [44% 47%]	71 ± 5%* [70% 72%]	71 ± 5%* [70% 73%]	84 ± 5%** [83% 85%]
Validation 1				
MRI pos (16)	95 ± 3% [95% 96%]	96 ± 3% [95% 96%]	100 ± 2%** [99% 100%]	100 ± 0%** [100% 100%]
MRI neg (27)	61 ± 8% [59% 62%]	78 ± 4% [78% 79%]	88 ± 4%** [88% 89%]	90 ± 4%** [90% 91%]
Validation 2				
MRI pos (8)	99 ± 5% [98% 100%]	100 ± 0% [100% 100%]	99 ± 3% [99% 100%]	100 ± 0% [100% 100%]
MRI neg (6)	55 ± 11% [53% 57%]	49 ± 5% [48% 50%]	70 ± 7%** [69% 71%]	76 ± 13%** [73% 78%]

(n): sample size; */**: increased/best lateralization accuracy with respect to at least one/any model (Friedman $p<0.05$, after correction). 95% confidence intervals shown in brackets.

Table 3. Lateralization performance (AUC) across cohorts.

	VOLUME	**T2 INTENSITY**	**FLAIR/T1**	**T2 + FLAIR/T1**
Training				
MRI pos (36)	1.00 ± 0.01 [1.00 1.00]	1.00 ± 0.00 [1.00 1.00]	0.98 ± 0.01 [0.98 0.99]	1.00 ± 0.00 [1.00 1.00]
MRI neg (24)	0.46 ± 0.06 [0.45 0.47]	0.79 ± 0.03* [0.79 0.80]	0.80 ± 0.05* [0.78 0.81]	0.95 ± 0.03** [0.95 0.96]
Validation 1				
MRI pos (16)	1.00 ± 0.00 [1.00 1.00]	1.00 ± 0.00 [1.00 1.00]	1.00 ± 0.00 [1.00 1.00]	1.00 ± 0.00 [1.00 1.00]
MRI neg (27)	0.65 ± 0.06 [0.64 0.66]	0.91 ± 0.02* [0.91 0.92]	0.93 ± 0.04* [0.93 0.94]	1.00 ± 0.00** [1.00 1.00]
Validation 2				
MRI pos (8)	1.00 ± 0.03 [0.98 1.00]	1.00 ± 0.00 [1.00 1.00]	1.00 ± 0.00 [1.00 1.00]	1.00 ± 0.00 [1.00 1.00]
MRI neg (6)	0.61 ± 0.18 [0.57 0.65]	0.65 ± 0.16 [0.62 0.68]	0.76 ± 0.10* [0.74 0.78]	0.94 ± 0.12** [0.91 0.96]

(n): sample size; */**: increased/best lateralization accuracy with respect to at least one/any model (Friedman $p<0.05$, corrected for multiple comparisons). 95% confidence intervals are shown in brackets.

Figure legends

Figure 1. Classifier design. A. Training. The objective of training was to define an optimal region of interest (ROI) used to sample MRI features of hippocampal sclerosis. (1) For each feature in the training set, paired t-tests compared corresponding vertices of the ipsi- and contralateral hippocampal subfields, z-scored with respect to healthy controls (only maps of T2-weighted anomalies shown). (2) The resulting group-level t-map was exhaustively thresholded from 0 to the highest value and binarized. (3) For each threshold, we overlaid the binarized t-map on the asymmetry map of each individual and computed the average across subfields. (4) We then trained one linear discriminant analysis (LDA) classifier per threshold and retained the model yielding the highest lateralization accuracy (here, LDA Model 3, surrounded by the black dotted box) and used it to test the classifier. **B. Statistical parametric anatomical map of optimal ROIs**. For each modality, maps show the vertex-wise group-level probability of anomalies (optimal ROI) over 100 repetitions of the 5-fold cross-validation determined during training.

Figure 2. Group-level findings. Differences in columnar volume, T2 intensity, and FLAIR/T1 intensity between patients and controls are mapped on the hippocampal subfield surfaces. Results are corrected using random field theory. Red and blue indicate increases and decreases, respectively (scaled by Cohen's *d* effect size).

Figure 3. Individual lateralization prediction. Examples of lateralization prediction in two patients with MRI-positive right TLE (A) and MRI-negative left TLE (B). For each case, coronal sections of the T1-weighted and T2-weighted MRI and the synthetic FLAIR/T1 contrast (right is right on images) are shown together with the automatically generated asymmetry maps for columnar volume, T2-weighted and FLAIR/T1 intensities. On each map, the dotted line corresponds to the coronal MRI section and the optimal ROI (see figure 1) is outlined in black.

Figure 4. ROC curves in MRI-negative TLE. The Receiver Operating Characteristic (ROC) curves based on lateralization posterior probabilities from the trained LDA model are shown for the Training and Validation Cohorts. X and Y-axes represent the lateralization false positive rate (FPR) and true positive rate (TPR), respectively. Dotted

blue lines represent individual curves drawn from each validation repetition, thick lines (green: columnar volume; purple: T2-weighted intensity; orange: FLAIR/T1 intensity; red: T2 + FLAIR/T1) show average ROC curves across iterations, and dashed black lines correspond to a random classifier. The average area under the curve (AUC) is indicated. The right-most panels show the overlay of all curves for each cohort to facilitate comparisons.