



More Than Meets the Eye: Human Versus Computer in the Neuroimaging of Temporal Lobe Epilepsy

MRI-Based Deep Learning Can Discriminate Between Temporal Lobe Epilepsy, Alzheimer's Disease, and Healthy Controls

Chang AJ, Roth R, Bougioukli E, Ruber T, Keller SS, Drane DL, Gross RE, Welsh J, Abrol A, Calhoun V, Karakis I, Kaestner E, Weber B, McDonald C, Gleichgerricht E, Bonilha L; Alzheimer's Disease Neuroimaging Initiative. *Commun Med (Lond)*. 2023;3(1):33. doi:10.1038/s43856-023-00262-4.

Background: Radiological identification of temporal lobe epilepsy (TLE) is crucial for diagnosis and treatment planning. TLE neuroimaging abnormalities are pervasive at the group level, but they can be subtle and difficult to identify by visual inspection of individual scans, prompting applications of artificial intelligence (AI) assisted technologies. **Method:** We assessed the ability of a convolutional neural network (CNN) algorithm to classify TLE vs. patients with AD vs. healthy controls using T1-weighted magnetic resonance imaging (MRI) scans. We used feature visualization techniques to identify regions the CNN employed to differentiate disease types. **Results:** We show the following classification results: healthy control accuracy = 81.54% (SD = 1.77%), precision = 0.81 (SD = 0.02), recall = 0.85 (SD = 0.03), and F1-score = 0.83 (SD = 0.02); TLE accuracy = 90.45% (SD = 1.59%), precision = 0.86 (SD = 0.03), recall = 0.86 (SD = 0.04), and F1-score = 0.85 (SD = 0.04); and AD accuracy = 88.52% (SD = 1.27%), precision = 0.64 (SD = 0.05), recall = 0.53 (SD = 0.07), and F1 score = 0.58 (0.05). The high accuracy in identification of TLE was remarkable, considering that only 47% of the cohort had deemed to be lesional based on MRI alone. Model predictions were also considerably better than random permutation classifications ($p < 0.01$) and were independent of age effects. **Conclusions:** AI (CNN deep learning) can classify and distinguish TLE, underscoring its potential utility for future computer-aided radiological assessments of epilepsy, especially for patients who do not exhibit easily identifiable TLE associated MRI features (e.g., hippocampal sclerosis).

Commentary

As neurologists we pride ourselves in our prowess at interpreting neuroimages in a purely visual process of analysis. However, even in focal temporal lobe epilepsy (TLE), approximately 30% of MRIs are lesion negative.¹ We know there is an epileptogenic focus there, but we cannot see it. We can improve by using manual morphometrics or postprocessing techniques, but most of the time in clinical practice we are using our eyes, and the visual processing centers of the human cortex. Can we do better?

Chang and colleagues investigated the use of a form of artificial intelligence (AI) called a convolutional neural network (CNN) to analyze T1-weighted MRI sequences of the human brain to discriminate between TLE, Alzheimer's disease (AD), and healthy controls.² A convolutional neural network is a way the computer can "see" and is a type of deep learning architecture that learns from images or pixel data.

The author's hypothesis was that CNN, optimized for TLE, could use T1-weighted voxel-based whole brain atrophy

patterns to classify TLE versus AD or healthy controls, with a higher accuracy than by chance, and investigate which regions were important to classify.

They used MRI from pure unilateral TLE patients and matched controls from 3 sites, 2 in United States, 1 in Germany, between 2017 and 2020. Like TLE, AD is also associated with limbic atrophy on MRI, and an MR cohort was available through the Alzheimer's Disease Neuroimaging Initiative database.³ This AD cohort had probable AD based on guidelines (and excluded comorbid AD and epilepsy).

The TLE group had adult drug-resistant unilateral TLE based on clinical, neurophysiological (scalp or invasive) and radiological data by surgical conference consensus. No patients had a comorbid diagnosis of AD. Patients were excluded if they did not proceed to epilepsy surgery, those with mass lesions (which can distort anatomy and were readily visible), bilateral TLE, or an additional extratemporal focus. Forty-seven percent of patients had a *visible* radiographic diagnosis of mesial temporal sclerosis.





They recruited 157 patients with TLE, 73 with AD, and 250 controls. The TLE average age was 38.68 years (SD = 12.45), with average age of epilepsy onset 17.05 years (SD = 12.39), with duration of epilepsy 21.62 years (SD = 14.84). The AD average age was 75.71 years (SD = 8.1) and healthy average age was 51.59 years (SD = 20.97).

T1-weighted MRI images were preprocessed into 58 axial slices per participant, with each slice tested independently. Potential age-related effects on gray matter were removed with a technique known as voxel-based linear age regression. Both temporal and extra-temporal structures were involved and contributed to the analysis.

Artificial intelligence models have an input (in this case, axial dimensions of a slice with gray scale), a hidden “convolutional” layer, and an output (TLE, AD, or healthy control). The model was asked to predict the slice level diagnosis and was “tested” by comparing the trained model to a “shuffled” or random model.

Feature visualization analysis (FVA) is a way that one can see what criteria a CNN has used to classify an output. In this study, FVA highlighted areas critical to the classification of each group and inferred to be the distribution of the underlying pathology. Atrophy patterns in temporal and extratemporal regions in TLE were similar to voxel-based morphometry (VBM) performed in this group and in other studies,⁴ and included expected limbic areas, but also thalamic, orbital, olfactory, and precuneus regions, and adds to evidence of TLE as a network disease. Supplemental data in the article contains original data elements such as specific volumes of segmented brain regions for each subject and controls, and for trained and shuffled models.

The CNN model had a mean accuracy (true positive + true negative/total) of 86.84% (SD = 1.33%) and was best for TLE (90.45%, SD = 1.59%). So, 90% for the computer and 47% for the humans.

The authors concluded that the CNN model correctly identified TLE versus controls and TLE versus AD or nonspecific limbic atrophy at a statistically significant level independent of age, or lesional status identified by visual human analysis. The use of a comparison group of AD demonstrated accuracy in differentiating between different patterns of limbic atrophy.

Why are humans less accurate?—only 47% of TLE scans in this study were “lesional” to the human eye. By human nature, we can naturally restrict our focus, such as on the temporal lobe, particularly when seizure semiology and EEG support this, and so can lose some objectivity. It does serve as a reminder to make sure we carefully interrogate the whole brain. We have our fallibilities—distraction experiments showed that 83% of radiologists missed an image of an angry gorilla on a CT chest as they were so focused on finding lung nodules.⁵ The human visual processing system has its limits and cannot instantaneously analyze subtle changes in multiple different regions at once in the same way that CNN can.

The attraction of this AI method for clinical diagnostic purposes is that individual scans could be rapidly evaluated—manual morphometry measurements are time-consuming and

macroscopic, and voxel-based VBM is used as a whole-brain research tool and relies on statistical analysis of *groups* of subjects.

However, we are not out of a job yet—while the concrete objectivity of AI has its advantages, humans must always be there to analyze results, apply to real-world situations, and interpret unexpected outcomes. This study applies only to patients with unilateral TLE. There are other features not analyzable by this algorithm such as subtle signs of focal cortical dysplasia, or that unexpected, enlarged lymph node in the soft tissues outside the brain.

The technology has potential to be extended to study non-lesional extratemporal or even generalized epilepsies and other neurological or neurodegenerative conditions. Detection of an atrophy pattern consistent with generalized epilepsy would be clinically useful to differentiate from nonlesional focal epilepsy. The AI technology uses existing data and is relatively easy to perform and is inexpensive—such is the promise of AI in our everyday lives. But it would be premature to use this pattern of atrophy as lesional to guide surgical evaluation. It would be interesting to know the surgical outcome of the human visible and AI-only visible cases. Is this ready for prime time? Not yet, at least until further clinical validation and reproducibility is demonstrated. Human visual processing abilities will always be necessary.

Peter Widdess-Walsh, MB, FRCPI^{id}
Department of Neurology,
Beaumont Hospital, Dublin

ORCID iD

Peter Widdess-Walsh, MB, FRCPI ^{id} <https://orcid.org/0000-0003-4217-2575>

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Muhlhofer W, Tan Y-L, Mueller SG, Knowlton R. MRI-negative temporal lobe epilepsy—what do we know? *Epilepsia*. 2017;58(5):727-742.
- Chang AJ, Roth R, Bougioukli E, et al. MRI-based deep learning can discriminate between temporal lobe epilepsy, Alzheimer’s disease, and healthy controls. *Commun Med (Lond)*. 2023;3(1):33. doi:10.1038/s43856-023-00262-4
- Alzheimer’s disease neuroimaging initiative. Available at: <https://clinicaltrials.gov/ct2/show/NCT02854033> (accessed: 23 August 2023)
- Keller SS, Roberts N. Voxel-based morphometry of temporal lobe epilepsy: an introduction and review of the literature. *Epilepsia*. 2008;49(5):741-757.
- Drew T, Võ ML, Wolfe JM. The invisible gorilla strikes again: sustained inattention blindness in expert observers. *Psychol Sci*. 2013;24(9):1848-1853. doi:10.1177/0956797613479386