AI and EEG: Should EEGers RIP (Rest in Peace)?

Automated Interpretation of Clinical Electroencephalograms Using Artificial Intelligence

Tveit J, Aurlien H, Plis S, Calhoun VD, Tatum WO, Schomer DL, Arntsen V, Cox F, Fahoum F, Gallentine WB, Gardella E, Hahn CD, Husain AM, Kessler S, Kural MA, Nascimento FA, Tankisi H, Ulvin LB, Wennberg R, Beniczky S. *JAMA Neurol.* 2023;80(8):805-812. doi:10.1001/jamaneurol.2023.1645

Importance: Electroencephalograms (EEGs) are a fundamental evaluation in neurology but require special expertise unavailable in many regions of the world. Artificial intelligence (AI) has a potential for addressing these unmet needs. Previous AI models address only limited aspects of EEG interpretation such as distinguishing abnormal from normal or identifying epileptiform activity. A comprehensive, fully automated interpretation of routine EEG based on AI suitable for clinical practice is needed. Objective: To develop and validate an AI model (Standardized Computer-based Organized Reporting of EEG–Artificial Intelligence [SCORE-AI]) with the ability to distinguish abnormal from normal EEG recordings and to classify abnormal EEG recordings into categories relevant for clinical decision-making: epileptiform-focal, epileptiform-generalized, nonepileptiform-focal, and nonepileptiform-diffuse. Design, Setting, and Participants: In this multicenter diagnostic accuracy study, a convolutional neural network model, SCORE-AI, was developed and validated using EEGs recorded between 2014 and 2020. Data were analyzed from January 17, 2022, until November 14, 2022. A total of 30 493 recordings of patients referred for EEG were included into the development data set annotated by 17 experts. Patients aged more than 3 months and not critically ill were eligible. The SCORE-AI was validated using 3 independent test data sets: a multicenter data set of 100 representative EEGs evaluated by 11 experts, a single-center data set of 9785 EEGs evaluated by 14 experts, and for benchmarking with previously published AI models, a data set of 60 EEGs with external reference standard. No patients who met eligibility criteria were excluded. Main Outcomes and Measures: Diagnostic accuracy, sensitivity, and specificity compared with the experts and the external reference standard of patients' habitual clinical episodes obtained during video-EEG recording. Results: The characteristics of the EEG data sets include development data set (N = 30 493; 14 980 men; median age, 25.3 years [95%CI, 1.3-76.2 years]), multicenter test data set (N = 100; 61 men, median age, 25.8 years [95%CI, 4.1-85.5 years]), single-center test data set (N = 9785; 5168 men; median age, 35.4 years [95%CI, 0.6-87.4 years]), and test data set with external reference standard (N = 60; 27 men; median age, 36 years [95%CI, 3-75 years]). The SCORE-AI achieved high accuracy, with an area under the receiver operating characteristic curve between 0.89 and 0.96 for the different categories of EEG abnormalities, and performance similar to human experts. Benchmarking against 3 previously published AI models was limited to comparing detection of epileptiform abnormalities. The accuracy of SCORE-AI (88.3%; 95%CI, 79.2%-94.9%) was significantly higher than the 3 previously published models (P < .001) and similar to human experts. Conclusions and Relevance: In this study, SCORE-AI achieved human expert level performance in fully automated interpretation of routine EEGs. Application of SCORE-AI may improve diagnosis and patient care in underserved areas and improve efficiency and consistency in specialized epilepsy centers.

Commentary

Artificial intelligence (AI) systems in health care describes the use of machine learning algorithms and software to approximate conclusions based solely on input data. By processing large and diverse data, health-related AI applications can analyze relationships between clinical data and patient outcomes. As an example, deep learning models have been used to analyze rich data sources, such as electroencephalogram (EEG) signals from unresponsive patients to predict recovery from acute brain injuries.[1]

Electroencephalogram is an essential diagnostic tool for neurological care that requires special expertise, is time-consuming, and is not widely available outside of developed economies.[2] Artificial intelligence has the potential to address these features by lowering the time-burden of review and reduce misinterpretation. Hybrid AI-based algorithms that detect and cluster interictal

---

epileptiform discharges in routine EEG studies, and then reviewed by experts, have been shown to have high specificity and good sensitivity. Most of these studies have addressed only limited aspects of EEG interpretation, such as the presence or absence of seizures[3] or epileptiform discharges.[4] Unfortunately, previous fully automated detection methods have suffered from low specificity making them unsuitable for clinical implementation.[5]

In the current diagnostic accuracy study using a convolutional neural network model, Tveit et al[6] report on the use of a fully automated and comprehensive assessment of routine clinical EEG studies compared to expert-level performance (Convolutional neural network, a type of machine learning program, is a feed-forward neural network that learn feature engineering by itself via filters optimization). This model uses the Standardized Computer-based Organized Reporting of EEG (SCORE-AI), a software tool for annotating EEGs using common data elements used by experts to label relevant EEG features, which are then feed into a centralized database.

The model was built on a data set of over 30 000 EEG recordings with a mean duration of 33 minutes, configured to access 19 sensors (10-20 system) and sampling at 256 Hz. These recording were obtained from subjects with a median age 25.3 years (range 1.2-76.2 years) studied at 2 Danish centers. Recordings were classified as normal versus abnormal (focal vs generalized, epileptiform vs non-epileptiform). This data set was then used to determine the model output threshold to enable a probabilistic interpretation. The tool was then integrated with a commercially available EEG system that enabled a fully automated analysis. An independent data set from patients not included in the developmental phase was used for clinical validation. A multicenter test data set, based on the interpretation of 11 experts from 11 different centers, severed as a reference standard. Finally, the model's output was compared with the clinical assessments of a large data set from one center that did not participate in the development of the model.

Outcome measures included the interrater agreement among experts and between the AI model and experts. The majority consensus was used as the reference standard in the multicenter data set to determine the diagnosis accuracy of the model. Impressively, the overall accuracy of SCORE-AI was similar to human expert interpretations (88.3% vs 83.3%, 95% CI, 73%-94.9%) and more accurate (*P* < .001) than previously published AI models.

A major advantage of this model is the ability to integrate with a widely used commercially available clinical EEG system. Additionally, the multicenter test data set was based on EEG recorded with different EEG equipment and reviewed by a panel of European and North American experts. An unexpected and intriguing finding is the low-interrater variability, which is known to be high at least for interpreting interictal epileptiform abnormalities and partly due to the use of short segments of EEG for assessment.[5]

Limitations of the study include exclusion of EEG studies from neonatal and critically ill patients and lack of information on how noise/artifacts affected the EEG interpretation. Nevertheless, this herculean work is to be applauded for the comprehensive review of such a large cohort of EEG studies, for the inclusion of a large international group of experts, and methodological design and analysis.

Real-life medical practice is likely to involve human-in-the-loop setups, where clinicians actively collaborate with the AI systems and provide oversight. This raises questions regarding the regulation of AI in medicine, the ways with which AI may modify responsibilities throughout health care systems as well as ethical concerns about data use, equity in medical AI, and concerns around accountability.[7]

Adoption of this tool in clinical practice will partly depend on winning administrative support and clearing regulatory hurdles to allow for public insurance reimbursement, akin to the use of AI systems form medical image diagnosis.[8] Likewise, having an AT tool that is reliable, convenient to use, and easy to integrate into clinical workflows will partly depend on earning the clinician's trust on the model and its accuracy. This will require studies to be replicable so that the models perform consistently even when trained with different samples of data.

In conclusion, Tviet et al[6] have developed the first fully automated and integrated neural network tool that achieves expert-level performance in interpreting routine clinical EEG studies. EEGers should embrace this powerful tool that can complement their practice, enhancing their productivity and allowing them to concentrate of more challenging diagnostic studies, such as intracranial EEG. EEGers, rest assure, the potential of AI to markedly improve your future will not make your work irrelevant.

David King-Stephens, MD 🆔
Yale School of Medicine
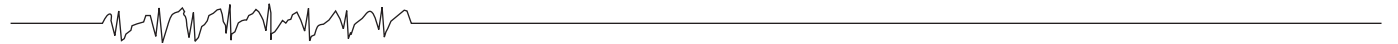*Department of Neurology,*
University of California Irvine

ORCID iD

David King-Stephens 🆔 http://orcid.org/0000-0002-1455-9847

Declaration of Conflicting Interests

References

1. Claassen J, Doyle K, Matory A, et al. Detection of brain activation in unresponsive patients with acute brain injury. *N Engl J Med.* 2019;380(26):2497-2505.
2. Kwon CS, Wagner RG, Carpio A, Jetté N, Newton CR, Thurman DJ. The worldwide epilepsy treatment gap: a systematic review and recommendations for revised definitions—a report from the ILAE Epidemiology Commission. *Epilepsia.* 2022;63(3):551-564.
3. Baumgartner C, Koren JP. Seizure detection using scalp-EEG. *Epilepsia.* 2018;59(suppl 1):14-22.
4. da Silva Lourenço C, Tjepkema-Cloostermans MC, van Putten MJAM. Machine learning for detection of interictal epileptiform discharges. *Clin Neurophysiol.* 2021;132(7):1433-1443.
5. Kuyal MA, Jing J, Fürbass F, et al. Accurate identification of EEG recordings with interictal epileptiform discharges using a hybrid

approach: artificial intelligence supervised by human experts. *Epilepsia*. 2022;63(5):1064-1073.

6. Tveit J, Aurlien H, Plis S, et al. Automated interpretation of clinical electroencephalograms using artificial intelligence. *JAMA Neurol*. 2023;80(8):805-812. doi:10.1001/jamaneurol.2023.1645

7. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31-38.

8. Centers for Medicare & Medicaid Services. Medicare program; hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and final policy changes and fiscal year 2021 rates; quality reporting and Medicare and Medicaid promoting interoperability programs requirements for eligible hospitals and critical access hospitals. *Fed Regist*. 2020;85:58432-59107.