

The ability of artificial intelligence to distinguish abnormal from normal EEG in patients suspected of epilepsy – updated literature review

Marcin Kopka

Centrum Medyczne NeuroProtect, Warsaw

Received June 12, 2024
Accepted for publication October 28, 2024
Published on-line November 6, 2024

Correspondence

Centrum Medyczne NeuroProtect
ul. Klaudyńny 16 C/1 piętro
01-684 Warszawa
e-mail: marcin.kopka@wp.pl

SUMMARY

Introduction. In patients suspected of epilepsy, electroencephalography (EEG) is an essential tool in the diagnostic workup. Currently, visual analysis of interictal epileptiform discharges by experts is the gold standard. Neurophysiologists perform an analysis of EEG through visual inspection. This is very time-consuming and inefficient. There is an increasing need to develop reliable and validated automated EEG analysis methods. Methods based on artificial intelligence can potentially help achieve this task.

Aim. The present review paper aims to present the current state of knowledge regarding the ability of artificial intelligence to distinguish abnormal from normal EEG in patients suspected of epilepsy.

Material and methods. This review covers the most relevant and recent papers identified using the PubMed database.

Results and discussion. Artificial intelligence (AI) has the potential to improve the management of epilepsy. It was shown that AI could distinguish normal from abnormal recordings, detect seizures, or detect interictal epileptiform discharges. The AI model (SCORE-AI) was developed and validated to assess routine clinical EEGs comprehensively. The sensitivity of SCORE-AI (86.7%) was lower than the sensitivity of the human experts (93.3%) and two models, Encevis (96.7%) and Persyst (100%) but higher than the sensitivity of SpikeNet (66.7%). The accuracy of SCORE-AI was similar to that of human experts and higher than that of the three previously published AI models. SCORE-AI achieves high specificity similar to the human raters and significantly higher accuracy than the three previously published AI models.

Conclusion. Methods based on artificial intelligence can potentially be helpful in EEG analysis. SCORE-AI may reduce excessive workloads for experts who interpret high volumes of EEG recordings. The SCORE-AI is the first model capable of completing a fully automated and comprehensive clinically relevant assessment of routine EEGs.

Key words: artificial intelligence • epilepsy • EEG

INTRODUCTION

Epilepsy is defined as a disease of the brain characterised by any of the following conditions: (1) At least two unprovoked (or reflex) seizures occurring >24 h apart; (2) one unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring over the next 10 years; (3) diagnosis of an epi-

lepsy syndrome (Fisher et al., 2014). An epileptic seizure is the clinical manifestation of an abnormal, excessive, and synchronised electrical discharge in the neurons. In patients suspected of epilepsy, electroencephalography (EEG) is an essential tool in the diagnostic workup (Tatum et al., 2018). The general practice of analysing the recorded EEG in patients suspected of epilepsy is to

detect the presence of epileptiform patterns. Their significance is well documented (Tatum et al., 2018; Pillai, Sperling, 2006). They are present in about half of the routine EEG recordings (Lodder et al., 2014). Currently, visual analysis of interictal epileptiform discharges by experts is the gold standard (Halford, 2009). Neurophysiologists perform an EEG analysis through visual inspection (Singh, Trevick, 2016). This is very time-consuming and inefficient, especially during the analysis of long-term EEG recording (Tatum et al., 2018; Pillai, Sperling, 2006). The overlap of symptomatology of epilepsy with other neurological disorders and contamination of EEG signals by artefacts makes this task very challenging, even for an experienced neurophysiologist. Inter-rater agreement of experts on identifying epileptiform EEG discharges is low (Halford et al., 2018). Misinterpretation of EEG is the most common cause of epilepsy misdiagnosis (Benbadis, Lin, 2008; Benbadis, Tatum, 2003). The delayed or incorrect diagnosis could have serious social and health consequences. There is an increasing need to develop reliable and validated automated EEG analysis methods. Methods based on artificial intelligence can potentially help improve the management of epilepsy. For example, automating the annotation and inspection of hours-long EEG records for IDE will decrease the excessive workloads placed on human experts in specialised centres.

AIM

The review paper presents the current knowledge regarding artificial intelligence's ability to distinguish abnormal from normal EEG in patients suspected of epilepsy. Available AI models will be mentioned and compared mainly in terms of accuracy. The complete capture of AI applications in epilepsy is beyond the scope of this article.

MATERIAL AND METHOD

This review covers the most relevant recent papers identified using the PubMed database. The search was conducted in June 2024 and included the terms artificial intelligence, deep learning, interictal epileptiform discharges, and automatic spike detection. It encompasses articles published in English from 2017 to 2024.

RESULTS AND DISCUSSION

Artificial intelligence

Artificial intelligence (AI) is defined as the simulation

of human intelligence processes by computer systems. It is important to note that AI is not a simple methodology but a collection of techniques to solve particular tasks (Copeland, 2009). One of them is the detection of interictal epileptiform discharges (IEDs). Generally, fully automated and hybrid approaches are used. In a hybrid method, IEDs detected by AI algorithms are visually evaluated by experts using the criteria in the operational definition of the International Federation of Clinical Neurophysiology (IFCN) for IEDs, which provide high specificity essential in clinical EEG reading (Kane et al., 2017; Kane et al., 2020).

Artificial intelligence in epileptology

Artificial intelligence (AI) has the potential to improve the management of epilepsy. It can decrease excessive workloads on human experts interpreting EEGs in specialised centres (Rajpurkar et al., 2022; Beniczky et al., 2021; Abbasi, Goldenholz, 2019). It was shown that AI could distinguish normal from abnormal recordings (van Leeuwen et al., 2019), detect seizures (Benbadis, Tatum, 2018; Pavel et al., 2020; Japaridze et al., 2023), or detect interictal epileptiform discharges (da Silva Lourenco et al., 2021). Typically, a large amount of labelled data is required to develop deep learning-based algorithms (Attia et al., 2019; van Leeuwen et al., 2019). The highly annotated large Standardized Computer-based Organized Reporting of EEG (SCORE EEG) database may provide a rich source of data for training an AI model (Beniczky et al., 2013; Beniczky et al., 2017).

SCORE

Despite the introduction of standardised certification measures, its interpretation remains subjective. The reported rates of interobserver agreement in the studies cited above range from as low as 19% (Williams et al., 1985) to as high as 95% (Struve et al., 1975) in the study, which 50 standard EEGs and 61 EEGs after partial sleep deprivation from 93 children were included. The interobserver agreement was moderate for the interpretation of the EEG as normal or abnormal (κ 0.66), strong for the presence of epileptiform discharges (κ 0.83), and moderate for the presence of focal non-epileptiform discharges (κ 0.54) (Stroink et al., 2006). Interobserver agreement was moderate among four neurologists who independently read 50 EEGs from adult patients with untreated idiopathic first seizures regarding whether the EEGs were normal or abnormal (κ 0.45). Slightly better results were found for the presence

or absence of epileptic discharges (kappa 0.5). Agreement rates on paroxysmal vs non-paroxysmal abnormalities were lower (kappa 0.33) (van Donselaar et al., 1982). Interobserver agreement tends to be highest for broad categories, such as normal versus abnormal, and lower for more specific aspects of interpretation, such as identifying focal epileptiform discharges. The interobserver agreement can be consistently higher when well-defined descriptive terms are used instead of routine texts (Gerber et al., 2008).

A software for standardised assessment, interpretation, and reporting of EEG has been developed by an international panel of experts under the auspices of the IFCN (Beniczky et al., 2013; Beniczky et al., 2017). It is called SCORE (Standardised computer-based organised reporting of EEG) and is endorsed by the International League Against Epilepsy. SCORE is a standardised software tool for annotating EEGs using common data elements. SCORE will help bridge the gap between the classical visual analysis method of EEG and the advanced (computerised) analysis methods. A free version of the software is available.

SCORE-AI and comprehensive assessment of EEG

The study published in the JAMA journal aimed to develop and validate an AI model (SCORE-AI) to comprehensively assess routine clinical EEGs (Tveit et al., 2023). A deep learning model was trained using the SCORE EEG system (Beniczky et al., 2013; Beniczky et al., 2017). A large data set of EEG (more than 30,000 recordings) from different centres was used to train the SCORE-AI model. These recordings were annotated by 17 human experts using SCORE-EEG (Beniczky et al., 2013; Beniczky et al., 2017). A test data set with a representative distribution (nearly 10000 EEG recordings) was used to ensure the generalizability of results. The test data set was independent of the development data set. The SCORE-AI achieved high accuracy. There are only minor relative variations related to the recording duration (less or more than 20 minutes). Generally, the average accuracy for the different categories of EEG abnormalities was from 0.89 to 0.96.

SCORE-AI was compared with three previously published AI models (Halford et al., 2017). SCORE-AI demonstrated substantially greater specificity than them (90% vs 3%–63%). Moreover, it was more specific than the majority consensus of the three human experts (73.3%). The sensitivity of SCORE-AI (86.7%) was lower than the sensitivity of the human experts (93.3%)

and two models, Encevis (96.7%) and Persyst (100%), but higher than the sensitivity of SpikeNet (66.7%). The accuracy of SCORE-AI was similar to that of human experts and higher than that of the three previously published AI models.

It is worth highlighting that in SCORE-AI, there is a high interrater agreement among human experts. In the previous studies, it was only moderate (Halford et al., 2017; Kural et al., 2020). However, the new model to train shorter segments of EEG with selected abnormal patterns was used instead of a complete, continuous recording (Tveit et al., 2023). There is no open-source or commercially available software package for comprehensive assessment of routine clinical EEGs. Several AI-based models have been developed to detect epileptiform activity on EEG (da Silva Lourenco et al., 2021). The other major limitation of the previously published models is the high number of false detections (0.73 per minute) precluding their clinical implementation (da Silva Lourenco et al., 2021b). Moreover, recently published data indicate that the fully automated application of three previously published AI models had specificity (3%–63%) (Kural et al., 2022). Only the hybrid approach may be suitable for clinical implementation, where human raters use the operational IFCN definition to confirm IEDs automatically detected and clustered by an AI-based algorithm.

CONCLUSION

SCORE-AI achieves high specificity similar to the human raters and significantly higher accuracy than the three previously published AI models. Moreover, the previously published models typically need human intervention because their specificity is very low, so human experts must confirm the findings. SCORE-AI, without any additional human intervention, the model achieves the same performance as the experts. SCORE-AI may reduce excessive workloads for experts who interpret high volumes of EEG recordings. The SCORE-AI is the first model capable of completing a fully automated and comprehensive clinically relevant assessment of routine EEGs (Tveit et al., 2023). The limitation of SCORE-AI is that it was developed and validated on routine EEG recordings, excluding critically ill patients and neonates. Another significant limitation is that human experts trained the model to find biomarkers visually identified. However, it will be more useful in clinical practice when it can predict diagnosis or therapeutic response. It is worth noting that AI was not invent-

ed to replace humans but to help optimise their performance. Indeed, visually annotating and inspecting hours-long EEG records for IDE is a time-consuming and demoting task for clinicians.

REFERENCES

- Abbasi B., Goldenholz D.M.: *Machine learning applications in epilepsy*. *Epilepsia*, 2019, 60: 2037–2047 .
- Attia Z.I., Noseworthy P.A., Lopez-Jimenez F., Asirvatham S.J., Deshmukh A.J., Gersh B.J. et al.: *An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction*. *Lancet*, 2019, 394: 861–867.
- Benbadis S.R., Lin K.: *Errors in EEG interpretation and misdiagnosis of epilepsy: which EEG patterns are overread?* *Eur. Neurol.*, 2008, 59: 267–271.
- Baumgartner C., Koren J.P.: *Seizure detection using scalp-EEG*. *Epilepsia*, 2018, 59 (suppl. 1): 14–22.
- Benbadis S.R., Tatum W.O.: *Overinterpretation of EEGs and misdiagnosis of epilepsy*. *J. Clin. Neurophysiol.*, 2003, 20: 42–44.
- Beniczky S., Karoly P., Nurse E., Ryvlin P., Cook M.: *Machine learning and wearable devices of the future*. *Epilepsia*, 2021, 62(suppl. 2): S116–S124.
- Beniczky S., Aurlien H., Brogger J.C. et al. : *Standardised Computer-based Organized Reporting of EEG: SCORE*. *Epilepsia*, 2013, 54: 1112–1124.
- Beniczky S., Aurlien H., Brogger J.C., Hirsch L.J., Schomer D.L., Trinka E. et al. : *Standardised computer-based organized reporting of EEG: SCORE – Second version*. *Clin. Neurophysiol.*, 2017, 128: 2334–2346.
- Copeland B.J.: „Artificial intelligence”. *Encyclopedia Britannica*. 27 Sep. 2024, <https://www.britannica.com/technology/artificial-intelligence>. Accessed 30 September 2024.
- van Donselaar C.A., Schimsheimer R.J., Geerts A.T., Declerck A.C.: *Value of the electroencephalogram in adult patients with untreated idiopathic first seizures*. *Arch. Neurol.*, 1992, 49: 231–237.
- Fisher R.S., Acevedo C., Arzimanoglou A., Bogacz A., Cross J.H., Elger C.E., Engel J. Jr, Forsgren L. et al.: *ILAE official report: a practical clinical definition of epilepsy*. *Epilepsia*, 2014, 55: 475–482. doi: 10.1111/epi.12550.
- Gerber P.A., Chapman K.E., Chung S.S., Drees C., Maganti R.K., Ng Y.T. et al.: *Interobserver agreement in interpreting EEG patterns in critically ill adults*. *J. Clin. Neurophysiol.*, 2008, 25: 241–249.
- Halford J.J., Westover M.B., LaRoche S.M., Macken M.P., Kuttuay E., Edwards J.C. et al.: *Interictal Epileptiform Discharge Detection in EEG in Different Practice Settings*. *J. Clin. Neurophysiol.*, 2018, 35: 375–380.
- Halford J.J., Arain A., Kalamangalam G.P., LaRoche S.M., Leonardo B., Basha M. et al.: *Characteristics of EEG Interpreters Associated With Higher Interrater Agreement*. *J. Clin. Neurophysiol.*, 2017, 34: 168–173.
- Halford J.J.: *Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized ECG interpretation*. *Clin. Neurophysiol.*, 2009, 120: 1909–1915.
- Japaridze G., Loeckx D., Buckinx T., Armand Larsen S., Proost R. et al.: *Automated detection of absence seizures using a wearable electroencephalographic device: a phase 3 validation study and feasibility of automated behavioral testing*. *Epilepsia*, 2023, 64 (Suppl. 4): S40–S46.
- Kane N., Acharya J., Beniczky S., Caboclo L., Finnigan S., Kaplan P.W. et al.: *A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017*. *Clin. Neurophysiol. Pract.*, 2017, 2: 170–185.
- Kural M.A., Duez L., Sejer Hansen V., Larsson P.G., Rampp S., Schulz R. et al.: *Criteria for defining interictal epileptiform discharges in EEG: a clinical validation study*. *Neurology*, 2020, 94: e2139–e2147.
- Kural M.A., Jing J., Fürbass F., Perko H., Qerama E., Johnsen B. et al.: *Accurate identification of EEG recordings with interictal epileptiform discharges using a hybrid approach: Artificial intelligence supervised by human experts*. *Epilepsia*, 2022, 63: 1064–1073.
- Kural M.A., Duez L., Sejer Hansen V., Larsson P.G., Rampp S., Schulz R. et al.: *Criteria for defining interictal epileptiform discharges in EEG: A clinical validation study*. *Neurology*, 2020, 94: e2139–e2147.
- Lodder S.S., Askamp J., van Putten M.J.: *Computer-assisted interpretation of the EEG background pattern: a clinical evaluation*. *PloS One*, 2014, 9: e85966.
- van Leeuwen K.G., Sun H., Tabaeizadeh M., Struck A.F., van Putten M.J.A.M., Westover M.B.: *Detecting abnormal electroencephalograms using deep convolutional networks*. *Clin. Neurophysiol.*, 2019, 130: 77–84.
- Pavel A.M., Rennie J.M., de Vries L.S., Blennow M., Foran A., Shah D.K. et al.: *A machine-learning algorithm for neonatal seizure recognition: a multicentre, randomised, controlled trial*. *Lancet Child Adolesc Health*, 2020, 4: 740–749.
- Pillai J., Sperling M.R.: *Interictal EEG and the diagnosis of epilepsy*. *Epilepsia*, 2006, 47 (Suppl. 1): 14–22.
- Rajpurkar P., Chen E., Banerjee O., Topol E.J.: *AI in health and medicine*. *Nat. Med.*, 2022, 28: 31–38.
- Singh A., Trevick S.: *The epidemiology of global epilepsy*. *Neurol. Clin.*, 2016, 34: 837–847.
- da Silva Lourenco C., Tjepkema-Cloostermans M.C., van

Putten M.J.A.M.: *Machine learning for detection of interictal epileptiform discharges.* Clin. Neurophysiol., 2021, 132: 1433–1443.

da Silva Lourenco C., Tjepkema-Cloostermans M.C., van Putten M.J.A.M.: *Efficient use of clinical EEG data for deep learning in epilepsy.* Clin. Neurophysiol., 2021, 132: 1234–1240.

Struve F.A., Becka D.R., Green M.A., Howard A.: *Reliability of clinical interpretation of an electroencephalogram.* Clin. Electroencephalogr., 1975, 6: 54–60.

Stroink H., Schimsheimer R.J., de Weerd A.W., Geerts A.T., Arts W.F., Peeters E.A. et al.: *Interobserver reliability of visual interpretation of electroencephalograms in children with newly diagnosed seizures.* Dev. Med. Child. Neurol., 2006, 48: 374–377.

Tatum W.O., Rubboli G., Kaplan P.W., Mirsatari S.M., Radhakrishnan K., Gloss D. et al.: *Clinical utility of EEG in diagnosing and monitoring epilepsy in adults.* Clin. Neurophysiol., 2018, 129: 1056–1082.

Tveit J., Aurlien H., Plis S., Calhoun V.D., Tatum W.O., Schomer D.L. et al.: *Automated Interpretation of Clinical Electroencephalograms Using Artificial Intelligence.* JAMA Neurol., 2023, 80: 805–812.

Williams G.W., Lüders H.O., Brickner A., Goormastic M., Klass D.W.: *Interobserver variability in EEG interpretation.* Neurology, 1985, 35: 1714–1719.

