HHS Public Access

Author manuscript

J Clin Neurophysiol. Author manuscript; available in PMC 2025 August 10.

Published in final edited form as:

J Clin Neurophysiol. 2026 March 01; 43(3): 237–244. doi:10.1097/WNP.0000000000001185.

Inter-Rater Reliability of EEG-Based Encephalopathy Grading

Ryan A. Tesh[*,†,‡], **Anika Zahoor**[*,†,‡], **Jayme Banks**[*,†,‡], **Kaileigh Gallagher**[*,†,‡], **Christine A. Eckhardt**[†,‡,§], **Haoqi Sun**[*,†,‡], **Ioannis Karakis**[‖,¶], **Roohi Katyal**[#], **Jonathan Williams**[**], **Chetan Nayak**[**], **Aline Herlopian**[††], **Marcus C. Ng**[‡‡], **Adam S. Greenblatt**[**], **Emma Meyers**[†,§], **Mike Westmeijer**[*,†,‡], **Daniel S. Harrison**[†,§], **Wolfgang Ganglberger**[*,†,‡,§§], **Galina Gheihman**[†,§], **Tracey Fan**[†], **Aaron F. Struck**[‖], **Irfan S. Sheikh**[¶¶], **Fábio A. Nascimento**[**], **M. Brandon Westover**[*,†,‡]

[*]Department of Neurology, Beth Israel Deaconess Medical Center (BIDMC), Boston, Massachusetts, U.S.A.;

[†]Department of Neurology, Massachusetts General Hospital (MGH), Boston, Massachusetts, U.S.A.;

[‡]Clinical Data Animation Center (CDAC), BIDMC, Boston, Massachusetts, U.S.A.;

[§]Department of Neurology, Brigham and Women's Hospital (BWH), Boston, Massachusetts, U.S.A.;

[‖]Department of Neurology, Emory University School of Medicine, Atlanta, Georgia, U.S.A.;

[¶]University of Crete School of Medicine, Heraklion, Crete, Greece;

[#]Department of Neurology, Ochsner Louisiana State University Health Shreveport (LSUHS), Shreveport, Louisiana, U.S.A.;

[**]Department of Neurology, Washington University School of Medicine in St. Louis, St. Louis, Missouri, U.S.A.;

[††]Department of Neurology, Yale University School of Medicine, New Haven, Connecticut, U.S.A.;

[‡‡]Department of Neurology, University of Manitoba, Winnipeg, MB, Canada;

[§§]Sleep & Health Zurich, University of Zurich, Zurich, Switzerland;

Address correspondence and reprint requests to M. Brandon Westover, MD, PhD, Department of Neurology, Beth Israel Deaconess Medical Center (BIDMC), 330 Brookline Avenue, Boston, MA 02115, U.S.A.; bwestove@bidmc.harvard.edu.

I. S. Sheikh, F. A. Nascimento, and M. B. Westover are co-senior authors.

I. S. Sheikh, F. A. Nascimento, and M. B. Westover conceived the study. R. A. Tesh, A. Zahoor J. Banks, I. S. Sheikh, F. A. Nascimento, and M. B. Westover created the EEG training slides and assessment. F. A. Nascimento and M. B. Westover recorded the EEG training videos. A. Zahoor, J. Banks, H. Sun, I. Karakis, R. Katyal, J. Williams, C. Nayak, A. Herlopian, M. C. Ng, A. S. Greenblatt, E. Meyers, M. Westmeijer, D. S. Harrison, W. Ganglberger, G. Gheihman, T. Fan, I. S. Sheikh, F. A. Nascimento, and M. B. Westover completed the EEG training assessment. I. S. Sheikh, F. A. Nascimento, and M. B. Westover analyzed the data. R. A. Tesh and Anika Zahoor drafted the initial article. R. A. Tesh and M. B. Westover have verified the underlying data, had full access to all the data in the study, and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors were involved in data interpretation and revising the article for intellectual content.

M. B. Westover is a cofounder, scientific advisor, consultant to, and has personal equity interest in Beacon Biosignals. The remaining authors have disclosed that they do not have any conflicts of interest.

The data and code supporting the results reported in this article (text, tables, figures) are available on GitHub (https://github.com/bdsp-core/EGG-IRR-1.0).

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.clinicalneurophys.com).

‖ ‖Department of Neurology, University of Wisconsin School of Medicine, William S. Middleton Veterans Memorial Hospital, Madison, Wisconsin, U.S.A.;

¶¶Department of Neurology, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A.

Abstract

Purpose: Visual EEG Confusion Assessment Method-Severity (VE-CAM-S) quantifies encephalopathy severity based on electroencephalography features. This study evaluated inter-rater reliability among experts using the VE-CAM-S scale.

Methods: Nine experts from six institutions independently reviewed 32 15-second electroencephalography samples in an online test, assessing 29 features (16 in the VE-CAM-S and 13 additional, or "VE-CAM-S+"). A consensus of three experts served as the gold standard. Performance was measured by the median Matthews correlation coefficient between expert and gold-standard VE-CAM-S+ scores, along with average sensitivity and specificity. Qualitative analysis identified common feature-recognition errors affecting scores.

Results: Experts achieved a median Matthews correlation coefficient of 0.82 [95% CI: 0.74–0.99]. Specificity exceeded 90% for most features except background β (87%) and generalized delta (71%). Sensitivity was ≥65% except for burst suppression with epileptiform activity (61%), extreme delta brush (EDB; 61%), posterior dominant rhythm (50%), background α (59%) and β (42%). Common errors included missing subtle findings, confusing features, and misidentifying extreme delta brush.

Conclusions: This pilot study offers some initial support for the reliability of VE-CAM-S+ scoring. The largest errors occurred when experts missed or falsely identified features with higher weight in the VE-CAM-S. Encephalopathy grading through VE-CAM-S may be improved by breaking high-stakes features into smaller parts, creating a "cheat sheet" with scored examples, and designing teaching materials.

Keywords

Electroencephalography (EEG); Encephalopathy; Pilot study; Reliability (inter-rater reliability, IRR); Critical care; Teaching

---

Acute encephalopathy refers to diffuse cerebral dysfunction that ranges in severity. Clinically, it is characterized by alterations in attention, cognition, and/or level of consciousness. It can result from various causes, including primary neurologic conditions (e.g., epileptic seizures) and systemic disorders.[1] In recent work,[2] we developed the Visual EEG Confusion Assessment Method–Severity (VE-CAM-S), an interpretable neurophysiologic grading scale to quantify acute encephalopathy based on EEG features. We found that VE-CAM-S has excellent discrimination between levels of delirium and correlated strongly with the well-validated CAM-S Long Form[3] delirium severity score. In addition, VE-CAM-S scores were strongly associated with clinical outcomes, including in-hospital mortality, 3-month mortality, and Glasgow Outcome Score[4] at discharge.

Before the VE-CAM-S scale can be recommended for systematic clinical use, it is important to measure how reliable EEG experts are in assessing encephalopathy severity using this grading system. Therefore, we designed an inter-rater reliability study to (1) evaluate expert agreement in assigning VE-CAM-S scores and identify potential sources of error that could affect score validity and (2) evaluate nonexpert agreement to explore the feasibility of training nonexperts in EEG feature recognition. We created an online test featuring a range of EEG examples, including normal and abnormal EEG findings. EEG teaching materials, such as slides and videos, were also provided alongside the test. Overall, this pilot study aims to gather insights on improving the teaching of encephalopathic EEG grading and the VE-CAM-S scale for use in future studies and, ultimately, clinical care. We hypothesized that agreement in the VE-CAM-S scale among experts would be high and that using EEG educational materials among nonexperts could achieve an inter-rater reliability comparable with that of experts when quantifying encephalopathic EEGs.

METHODS

Participant Selection

This pilot study, conducted from September to November 2022, enlisted participants across 10 institutions (see Figure 1, Supplemental Digital Content 1, http://links.lww.com/JCNP/A343), including (1) epileptologists or clinical neurophysiologists (neurologists with at least 1 year of fellowship training in epilepsy or clinical neurophysiology), (2) neurologists or neurology residents with minimal (<1 year) or no fellowship training in epilepsy or clinical neurophysiology, and (3) non-physicians (researchers with EEG experience). Written informed consent was not required because the participants were the investigators. Data for the study were used under a Mass General Brigham Institutional Review Board-approved protocol (IRB # 2013P001024).

EEG Training Materials

Participants were invited to review training materials, including slides and/or videos, covering a range of EEG features. The slides were organized into four modules (A–D): (A) background EEG; (B) sporadic interictal epileptiform discharges (IEDs); (C) seizures, rhythmic, and periodic patterns; and (D) voltage and burst suppression. Each module followed a structured format: (1) introduction: background on the EEG category (1–4 slides); (2) examples: 15-second EEG samples with colored annotations highlighting key features (variable number of slides); and (3) summary: a wrap-up slide listing all features introduced in the module. The content was based on the American Clinical Neurophysiology Society Standardized Critical Care EEG Terminology (2021).[5] In addition, two board-certified experts (F.N., M.B.W.) recorded videos explaining the slides. The teaching slides are available on GitHub (https://github.com/bdsp-core/EGG-IRR-1.0), and the encephalopathy videos can be accessed at eegtalk.com (playlist: https://bit.ly/encephalopathy).

EEG Training Assessment

The test was developed using a web-based survey platform (Google Forms) and included an introductory page followed by 32 EEG samples (15-second segments in longitudinal

bipolar montage). The samples were carefully selected to illustrate 29 EEG features relevant to encephalopathy (see Table 1, Supplemental Digital Content 2, http://links.lww.com/JCNP/A343): 14 features in VE-CAM-S (scored points), 2 features in VE-CAM-S (did not score points but were inputs in the final model used to create the scale), and 13 additional features not in VE-CAM-S. Two other features (generalized nonconvulsive status epilepticus and unreactive EEG) that scored points in VE-CAM-S were excluded from the test, because they generally require clinical correlation rather than EEG review alone. The additional features were based on an informal group discussion conducted through video conference and further 1:1 communication through email with EEG experts affiliated with the Critical Care EEG Monitoring Research Consortium (CCEMRC). Though the focus of the study was VE-CAM-S, we also chose to include additional features (e.g. background and sleep-related features) that might be useful to include in future improvements of the VE-CAM-S."

Answer options were categorized into six subsections: (1) background, (2) voltage and burst suppression, (3) sleep indicators, (4) sporadic IEDs, (5) rhythmic and periodic patterns, and (6) seizures and brief potentially ictal rhythmic discharges. For each EEG sample, participants were asked to select all applicable features, where selecting a feature indicated its presence and leaving it unselected indicated its absence. Participants were instructed to consider the following guidelines: posterior dominant rhythm and background α activity are not mutually exclusive; sleep spindles and background β activity are not mutually exclusive; eye blinks do not include horizontal eye movements; do not select generalized rhythmic delta activity (GRDA) or generalized periodic discharges if extreme delta brush (EDB) is selected.[5] The gold standard was established by consensus among three epileptologists/clinical neurophysiologists who designed the test (I.S., F.N., M.B.W). Consensus was reached through discussion over video conference. Participants could complete the test in multiple sittings and were allowed to use reference materials, including the training slides/videos, during the test ("open book" format).

After completing the test, participants provided feedback on several aspects, including their overall rating of the assessment (scale: 1 = poor, 10 = outstanding), the country where they trained or practiced, their training level and profession, the number of EEGs (routine, long-term, and epilepsy monitoring unit studies) they had independently reviewed during training or practice, the number of times they completed the test, and suggestions for additional EEG features relevant to encephalopathy that were not included in the test. Participants were also asked whether they reviewed the teaching materials (slides and/or videos) before taking the test and, if so, to comment on their quality. In addition, they reported their confidence level in grading encephalopathy EEGs at baseline/before and after reviewing the materials, using a 5-point scale: not confident (1), slightly confident (2), somewhat confident (3), moderately confident (4), or very confident (5).

Statistical Analysis

Test performance was quantified by the median Matthews correlation coefficient (MCC) between participant and gold-standard VE-CAM-S+ scores, along with average sensitivity and specificity. The 95% confidence intervals were calculated through bootstrapping.

Confidence levels in grading encephalopathy EEGs are given as median inter-quartile range [IQR].

Qualitative analysis was performed by three experts (I.S., F.N., M.B.W.) to identify feature-recognition errors that had the greatest impact on VE-CAM-S+ scores, and involved (1) reviewing all false positives and false negatives for each EEG sample, (2) flagging cases with either at least one feature misidentified by >50% of experts or errors in high-stakes features, and (3) cross-referencing errors to uncover potential explanations.

RESULTS

Overall, 37 individuals from 10 institutions were invited to participate (see Figure 1, Supplemental Digital Content 3, http://links.lww.com/JCNP/A343), including 15 experts (epileptologists/clinical neurophysiologists), 14 neurologists/neurology residents with minimal (<1 year) or no fellowship training in epilepsy or clinical neurophysiology, and 8 nonphysician EEG researchers. Of these, nine experts from six institutions, along with five neurologists/neurology residents (including two neurology residents, two neurocritical care fellows, and one epilepsy fellow), and five nonphysician EEG researchers, completed the online test. Participant details are given in Table 1. Most EEG samples (20/32, 63%) were from patients in intensive care units undergoing long-term continuous monitoring; Table 2 provides patient characteristics for the EEG samples.

Expert Agreement

Experts achieved a median MCC of 0.82 [95% CI: 0.74–0.99]. Specificity was very high (>90%) for most EEG features (27/29), except for background β activity (excluding sleep spindles) (87%) and generalized irregular delta activity (71%). Sensitivity was moderate to high (>70%) for many EEG features (12/29) but was lower for focal irregular delta activity (67%), bilateral independent periodic discharges (67%), focal IEDs (67%), vertex waves (67%), generalized irregular θ activity (66%), generalized periodic discharges without triphasic morphology (65%), burst suppression with epileptiform activity (61%), EDB (61%), background α activity (59%), posterior dominant rhythm (50%), and background β activity (excluding sleep spindles) (42%). The performance of experts is given in Supplemental Digital Content 4 (see Table 2, http://links.lww.com/JCNP/A343) (gold standard) and Table 3 (all).

Nonexpert Agreement

Neurologists/neurology residents (without 1 year of formal EEG training) achieved a median MCC of 0.60 [95%-CI: 0.39–1.00]. Specificity was very high (>90%) for most EEG features (27/29), except for attenuation/low voltage (10–20 uV) (88%) and generalized irregular delta activity (84%). Sensitivity was low to moderate (≤60%) for about half the EEG features (15/29) but was higher for eye blinks (100%), GRDA (100%), generalized seizure (100%), K-complexes (80%), suppression/very low voltage (<10 uV) (80%), burst suppression without epileptiform activity (80%), lateralized rhythmic delta activity (70%), and EDB (70%). The performance of neurologists/neurology residents is given in Supplemental Digital Content 4 (see Table 3, http://links.lww.com/JCNP/A343).

Nonphysician EEG researchers achieved a median MCC of 0.56 [95% CI: 0.38–1.00]. Specificity was very high (>90%) for most EEG features (26/29), except attenuation/low voltage (10–20 uV) (89%), background β activity (excluding sleep spindles) (87%), and generalized irregular delta activity (80%). Sensitivity was low to moderate (≤60%) for about half the EEG features (14/29) but was higher for burst suppression without epileptiform activity (100%), EDB (100%), generalized seizure (100%), lateralized rhythmic delta activity (100%), GRDA (83%), bilateral independent periodic discharges (67%), suppression/very low voltage (<10 uV) (67%), eye blinks (67%), and K-complexes (67%). Results from two nonphysician EEG researchers were excluded, because they were involved in test creation. The performance of nonphysician EEG researchers is shown in Supplemental Digital Content 5 (see Table 4, http://links.lww.com/JCNP/A343).

Error Analysis

More than 50% of experts misidentified the following nine EEG features in at least one question (q): background α activity (4q), background β activity (excluding sleep spindles) (3q), generalized irregular delta activity (5q), generalized irregular θ activity (7q), burst suppression with epileptiform activity (1q), focal IEDs (1q), generalized periodic discharges without triphasic morphology (2q), attenuation/low voltage (10–20 uV) (1q), and GRDA (1q). In addition, at least 30% of experts misidentified the following high-stakes features in at least one question: EDB (2q), bilateral independent periodic discharges (1q), and generalized brief attenuations (3q). Examples of challenging cases are illustrated in Fig. 1A–D.

During qualitative analysis, several areas of confusion were noted among experts. Some experts failed to identify background α activity while correctly recognizing the presence or absence of a posterior dominant rhythm, and vice versa. Despite initial guidelines, this suggests some may have misunderstood these features as mutually exclusive. In one instance, superimposed intermittent α was not identified as primarily α activity. Background β activity was occasionally confused with myogenic artifact. In some cases, experts identified EDB but neglected to also note background β activity. Generalized irregular delta activity was missed in some cases with EDB present. In one instance, experts struggled to differentiate between regular and irregular delta activity; this was a case where the pattern was borderline regular. Subtle generalized irregular θ activity was sometimes overshadowed by focal irregular delta activity or lateralized rhythmic delta activity. In other cases, the inability to adjust test image sensitivity hindered waveform counting. Low image resolution made interpreting burst suppression with epileptiform activity challenging in one borderline case. Burst suppression was mistaken for generalized brief attenuations by one expert and some neurologists/neurology residents. Focal IEDs were confused with breach artifact in one instance. Generalized periodic discharges were sometimes mistaken for generalized or focal/multifocal IEDs. Generalized periodic discharges without triphasic morphology were missed in two cases with uncertainty about their periodicity (e.g., fewer than six consecutive cycles). In one case, two-thirds of experts incorrectly endorsed borderline attenuation/low voltage (10–20 uV). Extreme delta brush was frequently mislabeled as GRDA, and experts failed to recognize this less familiar pattern in some instances. A third of experts missed bilateral independent periodic discharges in one case where the right-side pattern was clear,

but the left-side pattern was less distinct. Some experts missed generalized brief attenuations in two cases while incorrectly endorsing them in another.

EEG Training Materials

Baseline confidence levels in grading encephalopathy EEGs (1 = not confident, 5 = very confident) for each group (overall, experts, neurologists/neurology residents, nonphysician EEG researchers) were 2.0 [1.0–4.0], 4.5 [4.0–5.0], 1.0 [1.0–1.5], and 1.0 [1.0–1.0], respectively. Nearly 75% of participants (14/19) reviewed the EEG training materials (both slides and videos), including 56% of experts (5/9), 80% of neurologists/neurology residents (4/5), and all nonphysician EEG researchers (5/5). Participant confidence levels are given in Table 1.

Among experts who reviewed the training materials, 60% (3/5) reported increased confidence in grading encephalopathy EEGs (two with a 1-level increase, one with a 3-level increase), while 40% (2/5) reported no change, remaining highly confident both before and after the review. All neurologists/neurology residents reported increased confidence (three with a 1-level increase, one with a 2-level increase). Among nonphysician EEG researchers, 80% (4/5) reported increased confidence (one with a 1-level increase, two with a 2-level increase, and one with a 3-level increase), while one participant did not report postreview confidence. For the 13 participants who provided both pre- and postconfidence levels, preconfidence levels by group were 2.0 [1.0–4.0], 4.0 [4.0–5.0], 1.0 [1.0–1.5], and 1.0 [1.0–1.3], respectively; postconfidence levels by group were 4.0 [3.0–5.0], 5.0 [5.0–5.0], 2.5 [2.0–3.3], and 3.5 [2.8–4.0], respectively.

Participant Feedback

Participants rated the EEG training assessment on a scale of 1 to 10 (1 = poor, 10 = outstanding). The mean (SD) scores were as follows: 8.8 (1.3) for all participants ($n = 17$), 9.0 (1.3) for experts ($n = 9$), 8.2 (1.1) for neurologists/neurology residents ($n = 5$), and 9.0 (1.7) for nonphysician EEG researchers ($n = 3$). Additional feedback on EEG features relevant to encephalopathy that could be included in the training assessment and comments on the quality of the teaching materials are summarized in Supplemental Digital Content 6 (see Table 5, http://links.lww.com/JCNP/A343).

DISCUSSION

This study demonstrates that the VE-CAM-S+ scoring rubric provides a reliable framework for grading encephalopathy severity based on EEG features, with high agreement among experts across multiple institutions. The strong inter-rater reliability (median MCC of 0.82) and high specificity for most features (>90%) underscore the robustness of the VE-CAM-S+ in expert hands. This is significant because a reliable grading system for encephalopathy may improve diagnosis, guide treatment, support research, optimize resources, and standardize care, leading to better patient outcomes and a deeper understanding of encephalopathy. However, challenges with sensitivity for certain high-stakes features, such as posterior dominant rhythm (50%), background α (59%) and β activity (42%), generalized irregular delta (72%) and θ (66%) activity, and burst suppression with epileptiform activity

(61%), highlight areas for refinement (e.g., including more examples, enhancing teaching materials).

Our findings align with prior studies demonstrating variability in EEG interpretation, even among highly trained professionals. For example, misinterpretation rates have been at least 21% of EEG patterns, and when categories condensed to normal, ictal, and nonictal abnormalities, misinterpretation rates were 11.5%.[6] These are in addition to the EEG patterns misinterpreted because of subtle or artifact-prone features.[7,8] In our analysis, lower sensitivity was observed for complex features, such as generalized periodic discharges and EDB, consistent with previous findings highlighting these patterns as common sources of error.[9–13] This reflects the inherent difficulty of distinguishing subtle EEG findings, especially in cases with overlapping or borderline patterns.[6,14–17] Experts had high baseline confidence and were less likely to review the teaching materials (56%). However, difficulties in accurately characterizing certain fundamental features persisted, suggesting a need for tailored resources and underscoring the potential risks of overconfidence in medical practice.

Nonexpert performance revealed reduced sensitivity compared with experts, although specificity remained comparable. This emphasizes the need for targeted training to enhance the recognition of nuanced EEG features. Educational initiatives have historically focused on high-yield abnormalities, such as seizures, while background and milder features receive less attention.[18,19] Our tailored training materials, which included expert-guided slides and videos based on American Clinical Neurophysiology Society guidelines, were particularly valuable for nonexperts, who reported significant improvements in confidence after review. These findings suggest that structured and standardized educational resources can help improve nonexpert performance.

Error analysis revealed key areas for improvement in VE-CAM-S+ scoring. Misidentifications were most frequent for features with significant weight in the scoring rubric, such as EDB and burst suppression. These errors often stemmed from overlapping morphologies, low-resolution images, or misinterpretation of periodicity. To address these challenges, future iterations of VE-CAM-S+ could incorporate more granular definitions of high-stakes features, breaking them into smaller, less error-prone components, that is, “cheat-sheets” with scored examples. In addition, interactive training materials that allow manipulation of EEG parameters (e.g., sensitivity, filtering) and enhanced resolution of test images could further improve accuracy.

Despite the strengths of the VE-CAM-S+ scale, the limitations of this study must be acknowledged. The sample size of EEGs and participants, while sufficient for a pilot study, does not fully represent the diversity of clinical settings. Moreover, reliance on consensus among a subset of experts to establish the gold standard introduces potential bias. Also, we allowed for an “open-book” format where reference materials and training slides/videos could be used, which potentially diminishes known expertise. Future studies should aim to validate these findings in larger, more diverse populations and explore the impact of advanced teaching tools, such as case-based learning and automated feedback systems. In addition, broader validation efforts and refinements are needed before such a tool can be applied in a clinical or high-fidelity research setting.

CONCLUSIONS

This pilot study offers some initial support for the reliability of VE-CAM-S+ scoring, while identifying key areas for improvement. Refinements to the scoring system and training materials, alongside broader validation efforts, could enhance the utility of VE-CAM-S+ for standardizing encephalopathy grading in clinical and research settings. By addressing these challenges, VE-CAM-S+ has the potential to bridge the gap between expert and nonexpert EEG interpretation in cases of encephalopathy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by the National Institutes of Health (NIH grant numbers RF1AG064312, RF1NS120947, R01AG073410, R01HL161253, R01NS126282, R01AG073598, R01NS131347, R01NS130119) and the National Science Foundation (NSF 2014431). NIH (RF1AG064312, RF1NS120947, R01AG073410, R01HL161253, R01NS126282, R01AG073598, R01NS131347, R01NS130119), and NSF (2014431).

Funding did not play a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the article; and decision to submit the article for publication.

REFERENCES

1. Erkkinen MG, Berkowitz AL. A clinical approach to diagnosing encephalopathy. Am J Med 2019;132:1142–1147. [PubMed: 31330129]
2. Tesh RA, Sun H, Jing J, et al. VE-CAM-S: visual EEG-based grading of delirium severity and associations with clinical outcomes. Crit Care Explor 2022;4:e0611. [PubMed: 35072078]
3. Inouye SK, Kosar CM, Tommet D, et al. The CAM-S: development and validation of a new scoring system for delirium severity in 2 cohorts. Ann Intern Med 2014;160:526–533. [PubMed: 24733193]
4. Jennett B, Bond M. Assessment of outcome after severe brain damage. Lancet 1975;1:480–484. [PubMed: 46957]
5. Hirsch LJ, Fong MWK, Leitinger M, et al. American clinical neurophysiology society's standardized critical care EEG Terminology: 2021 version. J Clin Neurophysiol 2021;38:1–29. [PubMed: 33475321]
6. Grant AC, Abdel-Baki SG, Weedon J, et al. EEG interpretation reliability and interpreter confidence: a large single-center study. Epilepsy Behav 2014;32:102–107. [PubMed: 24531133]
7. Amin U, Nascimento FA, Karakis I, Schomer D, Benbadis SR. Normal variants and artifacts: importance in EEG interpretation. Epileptic Disord 2023;25:591–648. [PubMed: 36938895]
8. Mader EC, Miller D, Toler JM, Olejniczak PW. Focal epileptiform discharges can mimic electrode artifacts when recorded on the scalp near a skull defect. J Invest Med High Impact Case Rep 2018;6:2324709618795305.
9. Hartshorn JA, Foreman B. Generalized periodic discharges with triphasic morphology. J Neurocrit Care 2019;12:1–8.
10. Gerber PA, Chapman KE, Chung SS, et al. Interobserver agreement in the interpretation of EEG patterns in critically ill adults. J Clin Neurophysiol 2008;25:241–249. [PubMed: 18791475]
11. Jeannin-Mayer S, André-Obadia N, Rosenberg S, et al. EEG analysis in anti-NMDA receptor encephalitis: description of typical patterns. Clin Neurophysiol 2019;130:289–296. [PubMed: 30611120]

12. Abend NS, Gutierrez-Colina A, Zhao H, et al. Interobserver reproducibility of electroencephalogram interpretation in critically ill children. J Clin Neurophysiol 2011;28:15–19. [PubMed: 21221016]
13. Jing J, Ge W, Struck AF, et al. Interrater reliability of expert electro-encephalographers identifying seizures and rhythmic and periodic patterns in EEGs. Neurology 2023;100:e1737–e1749. [PubMed: 36460472]
14. Gélisse P, Benbadis SR, Crespel A, Tatum WO. Overcoming traps and pitfalls leading to misinterpretation of normal EEG variants and variation of the background activity. J Neurol 2024;271:3869–3878. [PubMed: 38761192]
15. Britton JW, Frey LC, Hopp JL etal, authors St, Louis EK, Frey LC, eds. Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants [Internet]. Chicago: American Epilepsy Society, 2016. The Abnormal EEG. Available at https://www.ncbi.nlm.nih.gov/books/NBK390357/
16. Fürbass F, Herta J, Koren J, et al. Monitoring burst suppression in critically ill patients: multi-centric evaluation of a novel method. Clin Neurophysiol 2016;127:2038–2046. [PubMed: 26971487]
17. Dhakar MB, Sheikh ZB, Desai M, et al. Developing a standardized approach to grading the level of brain dysfunction on EEG. J Clin Neurophysiol 2023;40:553–561. [PubMed: 35239553]
18. Pan Y, Laohathai C, Weber DJ. The effectiveness of neurology resident EEG training for seizure recognition in critically ill patients. Epilepsy Behav Rep 2021;15:100408. [PubMed: 33458646]
19. Taran S, Ahmed W, Pinto R, et al. Educational initiatives for electroencephalography in the critical care setting: a systematic review and meta-analysis: revue systématique et méta-analyse. Can J Anaesth 2021;68:1214–1230. [PubMed: 33709264]

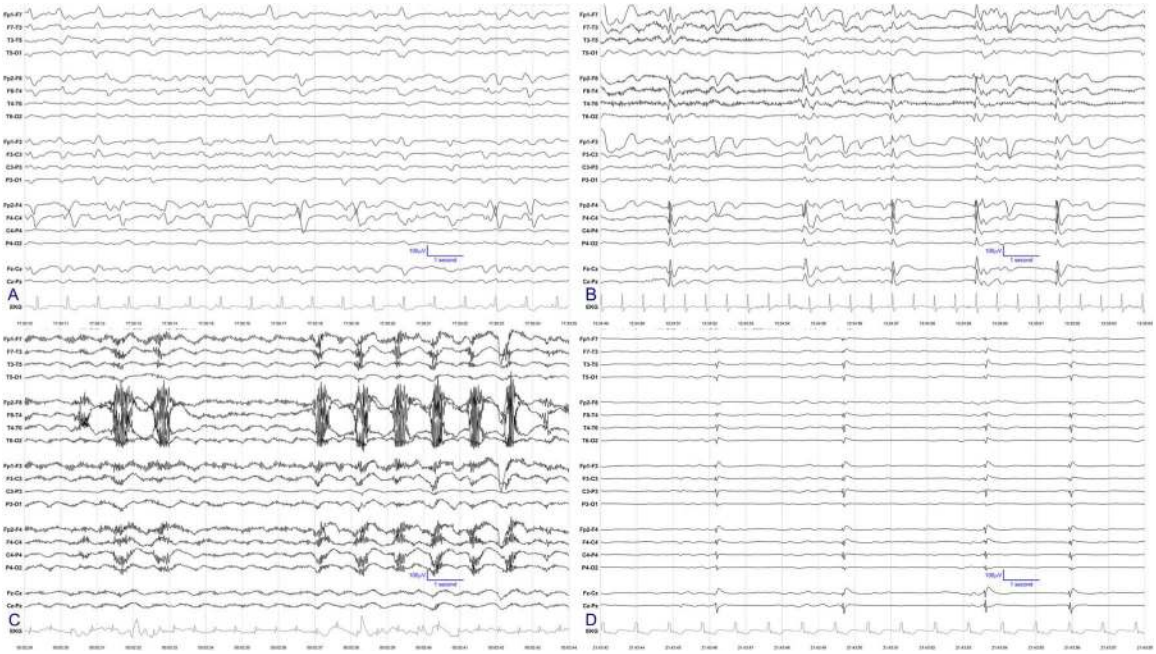

FIG. 1.

Examples of challenging EEG samples from the training assessment (**A**) (top left) The EEG sample includes generalized irregular delta and θ activity with BIPDs. Among experts, 30% (3/9) misclassified BIPDs (clearly visible on the right but unclear on the left) as either LPDs (2/3) or GPDs with triphasic morphology (1/3). In addition, 56% (5/9) missed the generalized irregular θ activity. (**B**) (top right) The EEG sample includes generalized irregular delta and θ activity, generalized brief attenuations, and GPDs without triphasic morphology. Among experts, 56% (5/9) confused GPDs (not fully periodic) with generalized IEDs (4/5) or multifocal IEDs (1/5). (**C**) (bottom left) The EEG sample includes background β activity (excluding sleep spindles), generalized irregular delta activity, and EDB. These were missed by 56% (5/9), 67% (6/9), and 44% (4/9) of experts, respectively. Furthermore, 30% (3/9) mislabeled EDB as GRDA, with 67% (6/9) incorrectly endorsing GRDA overall. (**D**) (bottom right) The EEG sample includes burst suppression with epileptiform activity and GPDs without triphasic morphology. This borderline case of burst suppression was missed by 56% (5/9) of experts, who instead identified it as suppression/very low voltage (<10 μV) (3/5), attenuation/low voltage (10–20 μV) (2/5), or generalized brief attenuations (1/5). In addition, 67% (6/9) incorrectly endorsed generalized irregular delta activity. BIPDs, bilateral independent periodic discharges; EDB, extreme delta brush; EEG, electroencephalography; GPDs, generalized periodic discharges; GRDA, generalized rhythmic delta activity; IEDs, interictal epileptiform discharges; LPDs, lateralized periodic discharges.

J Clin Neurophysiol. Author manuscript; available in PMC 2025 August 10.

TABLE 1.

Participant Characteristics

Background	Total (n = 19)	Expert (Epileptologists/Clinical Neurophysiologists) (n = 9)	Nonexpert (Neurologists/ Neurology Residents) (n = 5)	Nonexpert (EEG Researchers) (n = 5)
Country of training/practice: n (%)				
United States	16 (84.2)	7 (77.8)	5 (100.0)	4 (80.0)
International	3 (15.8)	2 (22.2)	0 (0.0)	0 (0.0)
Canada	1 (5.3)	1 (11.1)	0 (0.0)	0 (0.0)
Netherlands	1 (5.3)	0 (0.0)	0 (0.0)	1 (20.0)
Antarctica	1 (5.3)	1 (11.1)	0 (0.0)	0 (0.0)
EEG experience (No. of EEGs read independently in training/ clinical practice)				
0	6 (31.6)	0 (0.0)	3 (60.0)	3 (60.0)
1–20	3 (15.8)	0 (0.0)	1 (20.0)	2 (40.0)
21–50	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
51–100	1 (5.3)	0 (0.0)	1 (20.0)	0 (0.0)
101–200	1 (5.3)	1 (11.1)	0 (0.0)	0 (0.0)
>200	8 (42.1)	8 (88.9)	0 (0.0)	0 (0.0)
Years of EEG experience (including training)				
<1	N/A	0 (0.0)	N/A	N/A
1–3	N/A	4 (44.4)	N/A	N/A
4–6	N/A	0 (0.0)	N/A	N/A
7–9	N/A	2 (22.2)	N/A	N/A
10 or more	N/A	3 (33.3)	N/A	N/A
Read ICU EEGs as part of clinical practice	N/A	9 (100.0)	N/A	N/A
Board certified at time of study assessment	N/A	7 (77.8)	N/A	N/A
Taken and passed the CCEMRC terminology open-book self-assessment * before study	N/A	4 (44.4)	N/A	N/A
Confidence grading encephalopathy EEGs (baseline): n (%)				
Not confident	7 (36.8)	0 (0.0)	3 (60.0)	4 (80.0)
Slightly confident	2 (10.5)	1 (11.1)	0 (0.0)	1 (20.0)
Somewhat confident	1 (5.3)	0 (0.0)	1 (20.0)	0 (0.0)
Moderately confident	3 (15.8)	3 (33.3)	0 (0.0)	0 (0.0)

Background	Total (n = 19)	Expert (Epileptologists/Clinical Neurophysiologists) (n = 9)	Nonexpert (Neurologists/ Neurology Residents) (n = 5)	Nonexpert (EEG Researchers) (n = 5)
Very confident	4 (21.1)	4 (44.4)	0 (0.0)	0 (0.0)
Unanswered	2 (10.5)	1 (11.1)	1 (20.0)	0 (0.0)
Reviewed EEG training materials?				
No (neither slides nor videos)	5 (26.3)	4 (44.4)	1 (20.0)	0 (0.0)
Yes (both slides and videos)	14 (73.6)	5 (55.6)	4 (80.0)	5 (100.0)
Confidence grading encephalopathy EEGs (prereview): n (%)				
Not confident	7 (50.0)	0 (0.0)	3 (75.0)	4 (80.0)
Slightly confident	2 (14.3)	1 (20.0)	0 (0.0)	1 (20.0)
Somewhat confident	1 (7.1)	0 (0.0)	1 (25.0)	0 (0.0)
Moderately confident	2 (14.3)	2 (40.0)	0 (0.0)	0 (0.0)
Very confident	2 (14.3)	2 (40.0)	0 (0.0)	0 (0.0)
Confidence grading encephalopathy EEGs (postreview): n (%)				
Not confident	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Slightly confident	3 (21.4)	0 (0.0)	2 (50.0)	1 (20.0)
Somewhat confident	2 (14.3)	0 (0.0)	1 (25.0)	1 (20.0)
Moderately confident	3 (21.4)	0 (0.0)	1 (25.0)	2 (40.0)
Very confident	5 (35.7)	5 (100.0)	0 (0.0)	0 (0.0)
Unanswered	1 (7.1)	0 (0.0)	0 (0.0)	1 (20.0)

*CCEMRC Self-Assessment Examination on the American Clinical Neurophysiology Society's 2021 Critical Care EEG terminology. CCEMRC, Critical Care EEG Monitoring Research Consortium; EEG, electroencephalography; ICU, intensive care unit.

TABLE 2.

Patient Characteristics for EEG Samples Included in the Online Test

Characteristic	Total (*n* = 32)
Age, years: mean (SD) [*]	48.2 (23.9)
Sex [†]: *n* (%)	
Female	17 (53.1)
Male	12 (37.5)
Unknown	3 (9.4)
Race [†]: *n* (%)	
Asian	3 (9.4)
Black	3 (9.4)
Native American or other Pacific Islander	0 (0.0)
White	21 (65.6)
Other or unknown	5 (15.6)
Ethnicity [†]: *n* (%)	
Hispanic	2 (6.3)
Non-Hispanic	25 (78.1)
Unknown	5 (15.6)
EEG type	
Routine EEG (<60 min)	6 (18.8)
LTM	23 (71.9)
Unknown: *n* (%)	3 (9.4)
Location of EEG: *n* (%)	
ED	4 (12.5)
Outpatient EEG laboratory	2 (6.3)
Inpatient hospital floor (neurosciences)	2 (6.3)
RACU	1 (3.1)
ICU	20 (62.5)
NICU	16 (50.0)
MICU	2 (6.3)
CSICU	1 (3.1)
CCU	1 (3.1)
Unknown	3 (9.4)
Primary diagnosis [‡]: *n* (%)	
Altered mental status	1 (3.1)
Seizure	7 (21.9)
Neurovascular	6 (18.8)
Neuro-oncology	0 (0.0)
Neurology (other)	6 (18.8)
Psychiatric disorders	1 (3.1)
Infection	2 (6.3)
Cardiovascular	3 (9.4)

Characteristic	Total (*n* = 32)
Hematology/oncology	0 (0.0)
Gastrointestinal	0 (0.0)
Respiratory	3 (9.4)
Renal disease	0 (0.0)
Elective surgery	1 (3.1)
Trauma	0 (0.0)
Other	0 (0.0)
Unknown	3 (9.4)

[*] Age, mean (SD) calculated for those with age information (*N* = 29); excludes three patients whose age was unknown/unavailable.

[†] Demographic data (age, sex, race, and ethnicity) reported based on information obtained from the electronic health record.

[‡] Primary diagnosis percentages add up to >100% because patients could have >1 primary diagnosis.

CCU, coronary care unit; CSICU, cardiac surgery ICU; ED, emergency department; EEG, electroencephalography; ICU, intensive care unit; LTM, long-term monitoring; MICU, medical ICU; NICU, neurologic ICU; RACU, respiratory acute care unit.

TABLE 3.

Performance of Epileptologists/Clinical Neurophysiologists (n = 9)

EEG Features (29)	MCC (Gold std)f	MCC (all experts)*,†	Sensitivity*,†	Specificity*,†	PPV*,†	NPV*,†
Background						
PDR (y/n)	0.58 (0.38–1.00)	0.58 (0.38–1.00)	0.50 (0.00–1.00)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.67 (0.50–1.00)
Background α activity (y/n)	0.84 (0.73–0.95)	0.58 (0.44–0.72)	0.59 (0.43–0.74)	0.94 (0.91–0.97)	0.92 (0.85–0.96)	0.72 (0.65–0.81)
Background β activity (excluding sleep spindles) (y/n)	0.61 (0.29–0.94)	0.31 (−0.12–0.67)	0.42 (0.11–0.80)	0.87 (0.82–0.92)	0.70 (0.14–0.80)	0.63 (0.52–0.85)
Generalized irregular delta activity (y/n)†	0.61 (0.19–0.98)	0.48 (0.15–0.76)	0.72 (0.65–0.79)	0.71 (0.44–1.00)	0.78 (0.66–1.00)	0.76 (0.41–0.83)
Generalized irregular θ activity (y/n)†	0.86 (0.69–0.97)	0.60 (0.42–0.74)	0.66 (0.58–0.74)	0.91 (0.77–1.00)	0.88 (0.75–1.00)	0.76 (0.69–0.81)
Focal irregular delta activity (y/n)	0.80 (0.55–0.98)	0.61 (0.42–0.78)	0.67 (0.56–0.89)	0.91 (0.85–0.96)	0.88 (0.54–0.94)	0.76 (0.70–0.94)
Focal irregular θ activity (y/n)	N/A	N/A	N/A	0.98 (0.95–1.00)	N/A	N/A
Voltage and burst suppression						
Attenuation/low voltage (10–20 uV) (y/n)†	N/A	N/A	N/A	0.91 (0.83–0.97)	N/A	N/A
Suppression/very low voltage (<10 uV) (y/n)*	1.00 (1.00–1.00)	0.95 (0.92–0.98)	1.00 (1.00–1.00)	0.95 (0.91–0.98)	0.95 (0.93–0.98)	1.00 (1.00–1.00)
Generalized brief attenuations (y/n)*	0.87 (0.77–1.00)	0.82 (0.74–0.93)	0.76 (0.67–0.84)	0.95 (0.92–0.98)	0.95 (0.83–0.98)	0.84 (0.81–0.91)
Burst suppression without epileptiform activity (y/n)*	1.00 (1.00–1.00)	0.99 (0.98–1.00)	1.00 (1.00–1.00)	0.99 (0.98–1.00)	0.99 (0.98–1.00)	1.00 (1.00–1.00)
Burst suppression with epileptiform activity (y/n)*	0.72 (0.72–1.00)	0.73 (0.66–1.00)	0.61 (0.44–0.78)	1.00 (0.99–1.00)	1.00 (0.98–1.00)	0.76 (0.72–0.89)
Sleep indicators						
Eye blinks (y/n)	0.99 (0.97–1.00)	0.96 (0.92–0.99)	1.00 (1.00–1.00)	0.95 (0.92–0.98)	0.96 (0.93–0.99)	1.00 (1.00–1.00)
Sleep spindles (y/n)	0.90 (0.80–1.00)	0.85 (0.74–1.00)	0.74 (0.56–0.89)	1.00 (0.99–1.00)	0.99 (0.83–1.00)	0.83 (0.78–0.94)
K-complexes (y/n)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.78 (0.78–0.78)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.89 (0.89–0.89)
Vertex waves (y/n)	1.00 (1.00–1.00)	0.87 (0.83–1.00)	0.67 (0.56–0.78)	1.00 (0.99–1.00)	1.00 (0.98–1.00)	0.81 (0.78–0.89)
Sporadic IEDs						
Generalized IEDs (y/n)	N/A	N/A	N/A	0.95 (0.91–0.98)	N/A	N/A
Focal IEDs (y/n)	0.83 (0.58–1.00)	0.75 (0.46–0.98)	0.67 (0.44–0.89)	0.97 (0.93–0.99)	0.95 (0.54–0.99)	0.79 (0.72–0.94)
Multifocal IEDs (y/n)	0.99 (0.96–1.00)	0.85 (0.82–1.00)	0.78 (0.78–0.78)	0.99 (0.97–1.00)	0.86 (0.85–1.00)	0.89 (0.89–0.89)
Rhythmic and periodic patterns						
GRDA (y/n)*	0.98 (0.95–1.00)	0.79 (0.66–0.97)	0.83 (0.78–0.89)	0.94 (0.89–0.99)	0.93 (0.72–0.98)	0.88 (0.87–0.94)
LRDA (y/n)*	0.99 (0.97–1.00)	0.87 (0.82–0.99)	0.89 (0.89–0.89)	0.97 (0.93–1.00)	0.96 (0.84–0.99)	0.92 (0.92–0.94)
GPDs without triphasic morphology (y/n) *	0.81 (0.72–1.00)	0.77 (0.69–0.95)	0.65 (0.50–0.78)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.78 (0.73–0.86)

EEG Features (29)	MCC (Gold std)f	MCC (all experts)*,†	Sensitivity*,†	Specificity*,†	PPV*,†	NPV*,†
GPDs with triphasic morphology (i.e., TW) (y/n)*	N/A	N/A	N/A	0.99 (0.98–1.00)	N/A	N/A
LPDs (y/n)*	0.99 (0.96–1.00)	0.97 (0.95–0.99)	1.00 (1.00–1.00)	0.97 (0.94–0.99)	0.97 (0.95–0.99)	1.00 (1.00–1.00)
BIPDs (y/n)*	1.00 (1.00–1.00)	0.98 (0.96–1.00)	0.67 (0.67–0.67)	0.99 (0.97–1.00)	0.98 (0.96–1.00)	0.83 (0.83–0.83)
EDB (y/n)*	1.00 (1.00–1.00)	0.70 (0.67–1.00)	0.61 (0.56–0.67)	1.00 (0.99–1.00)	0.88 (0.83–1.00)	0.78 (0.77–0.83)
Seizures and BIRDs						
BIRDs (y/n)	N/A	N/A	N/A	0.98 (0.94–1.00)	N/A	N/A
Focal seizure (y/n)	N/A	N/A	N/A	1.00 (0.99–1.00)	N/A	N/A
Generalized seizure (y/n)	1.00 (1.00–1.00)	1.00 (0.99–1.00)	1.00 (1.00–1.00)	1.00 (0.99–1.00)	1.00 (0.99–1.00)	1.00 (1.00–1.00)

*EEG features in VE-CAM-S (scored points) and included on the EEG training assessment.

†Values given as *n* (95% confidence intervals); N/A (not available) indicates areas that could not be calculated (because no examples of that EEG feature were given on test). BIPDs, bilateral independent periodic discharges; BIRDs, brief potentially ictal rhythmic discharges; EDB, extreme delta brush; EEG, electroencephalography; GPDs, generalized periodic discharges; GRDA, generalized rhythmic delta activity; IEDs, interictal epileptiform discharges; LPDs, lateralized periodic discharges; LRDA, lateralized rhythmic delta activity; MCC, Matthews correlation coefficient; NPV, negative predictive value; PDR, posterior dominant rhythm; PPV, positive predictive value; TW, triphasic waves.