

RESEARCH ARTICLE



Using Machine Learning to Enhance the EEG Screening Review by Pre-Screening the EEG

Thomas Collura^{1,*} , Agostino Rosace² , Robert Turner³ , David Ims⁴ and Bill Brubaker⁵

¹*BrainMaster Technologies Inc, USA*

²*Department of Engineering, University of Cincinnati, USA*

³*Network Neurology Health LLC, USA*

⁴*Chesapeake Neurobehavioral Health, USA*

⁵*Stress Therapy Solutions, USA*

Abstract: Artificial intelligence (AI) is increasingly being used to assist physicians with their medical diagnosis and patient care and improve communication between doctors and their patients. This paper demonstrates how machine learning (ML), a subset of AI, can pre-screen electroencephalograms (EEGs) to assess the overall quality of the EEG and identify artifacts or abnormalities. We developed the “Brain Panel,” an automated ML tool that supplements clinical neurophysiologists’ quality review screening. The Brain Panel generates results prior to visual inspection or quantitative EEG (QEEG) analysis, highlighting potential technical issues or clinical concerns through novel metrics and interpretation methods. We subjected 100 Brain Panel reports and 100 corresponding physician reports from the same EEGs to independent AI evaluations using Grok and Claude. Results show that the Brain Panel provides a sensitive, neurologically informed method to estimate artifacts, assess EEG quality, detect drowsiness, and indicate the likelihood of brain abnormalities. By identifying optimal combinations of Brain Panel metrics, we created detection algorithms that achieved sensitivities of 89–95% for clinically relevant findings. This approach demonstrates that using AI to analyze and interpret automated Brain Panel reports produces a system with clear clinical value for detecting technical and clinical issues in EEGs, enhancing neurologist productivity without replacing expert judgment.

Keywords: QEEG, EEG, neurologist, quality assurance review, EEG screening

1. Introduction

More healthcare professionals are embracing the use of quantitative electroencephalogram (QEEG) and neurometrics in their clinical practices as a common form of assessment for their clients [1–3]. With the acceptance of QEEG and neurometrics, the gap between QEEG and electroencephalogram (EEG) professionals has been narrowing. In fact, new guidelines have just been released standardizing multi-channel EEG data acquisition, which follows the standards set forth by the American Clinical Neurophysiology Society [4], with the expectation that collaboration between medical and mental health fields will increase.

Many clinicians using neurofeedback along with QEEG also use a screening review service to evaluate the quality of the EEG recording and to look for any findings that may indicate the need for a referral to a neurologist or epileptologist. The system reported here is designed to supplement a physician’s review by

providing a quantitative guide to the EEG properties that are generally amenable to visual inspection.

The need for a machine learning (ML)-Enhanced Quality Review Report was conceived by the authors to streamline the EEG quality review process and to generate a report that more closely mimics a physician’s report. As technology is improving daily, the use of artificial intelligence (AI) in EEG interpretation was inevitable and shows significant potential [5]. Just as clinicians use computer software to aid in the evaluation of the EEG to derive specific neurometrics that cannot be seen with the naked eye, we used ML to inspect the EEG using predefined rules to aid in the review process.

While many people use the terms AI and ML interchangeably, they are different. ML is a subset of AI. AI is a broad term that encompasses different strategies and techniques to make machines more human-like, to imitate cognitive functions and decision-making to solve problems [6]. Alexa, Siri, self-driving cars, and robotic home cleaning machines are just a few examples of how AI is being used. AI exploits methods wherein the computer tries to learn to identify patterns and create rules. AI is

*Corresponding author: Thomas Collura, BrainMaster Technologies Inc, USA. Email: tomc1@brainmaster.com

very open in design and is prone to “hallucinations” and creating a fictional reality to satisfy its goals. With ML, on the other hand, the computer is given a set of rules to extract patterns from a dataset to perform well-defined tasks.

With AI, the rules the machine uses will change with the expectation that “learning” is occurring. The issue with this is that you may not get the same results if you evaluate the EEG at different times, especially after rules have changed. While this does not undermine the value of AI, the authors wanted a more fixed set of metric outcomes, which could then be subjected to an AI-based process, to refine their application. The rules used in the ML machine model are predefined; they do not change. We get the same answers every time. The machine is not learning and changing its own rules; rather, it is improving its accuracy over time as the database increases, providing a strict set of measurements.

This new method of evaluating an EEG allows us to create a new type of database that challenges the conventional wisdom of accepted normative database standards. We did not, for example, employ a rejection criterion for included EEGs. Any EEG that was submitted to our review service could be included. We further did not attempt to select “normal” EEGs but considered this a database of any possible submitted EEGs. This approach allows us to produce a quality review report that can generate data and metrics that mimic a physician’s report.

Our intent is not to replace the neurologist but rather to provide a tool that can enhance the neurologist’s work in terms of productivity. This is not some type of surrogate or replacement for the neurologist. We apply AI to leverage and enhance the neurologists’ work, not to encroach or threaten their territory and expertise. None of the neurologists’ tasks are replaced. Rather, an additional layer of inspection using an automated tool is applied as part of an extended procedure, never replacing an existing step. This distinguishes our approach from systems that purport to replace or mimic the clinician, thus attempting to replace a human operation with an automated one.

Our working hypothesis was that the various metrics would emerge as indicators of particular facets of the EEG, which would include whether there is injury or pathology, whether the client was still, attentive, or drowsy, for example, or whether there would be abnormalities cited in the EEG. It was anticipated that, after sufficient data had been collected that combined Brain Panel reports with their respective EEG reviews, it would be possible to perform a comparison in the form of a discriminant analysis that would reveal any significant correlations. At the time that the Brain Panel was created, we did not have access to AI tools to perform this analysis. However, by the time that we had accumulated a sufficient sample (100 EEGs with reviews), the available AI tools had evolved to provide a means to perform this analysis and reveal which metrics had clinical relevance.

2. Related Work

The Brain Panel system builds on a rich foundation of EEG analysis methodologies, distinguishing itself through its supervised ML approach, which contrasts with traditional AI and QEEG systems. Prior work in EEG analysis has focused on automated detection of epileptiform activity or normative QEEG databases to identify deviations from “normal” brain activity. These systems often rely on predefined normative datasets, which exclude abnormal EEGs, limiting their applicability to diverse clinical populations. In contrast, the Brain Panel’s inclusive database, accepting all EEGs meeting a 60-second minimum, aligns with real-world clinical variability, echoing the approach

of Thomas Collura and Jeff Tarrant, who emphasized the need for representative EEG datasets [7]. Additionally, while AI-driven systems like deep learning neural networks excel at pattern recognition, their opaque rule generation can lead to unpredictable outputs, as noted in misclassifications of non-EEG objects.

ML is being used to classify relaxation states during different postures from the EEG [8]. This work can have an impact on how relaxation states can be detected between eyes-open and eyes-closed conditions. Interpretable ML is being used to improve clinician performance in classifying EEG patterns on the Ictal-Interictal Injury Continuum, which could reduce human bias and accelerate the diagnostic process [9]. The Brain Panel’s use of predefined metrics, such as moments and fractal dimension, draws inspiration from mechanical engineering applications and entropy-based neurometrics, offering transparent and interpretable results. By integrating high-resolution fast Fourier transform (FFT) and independent component analysis (ICA) for artifact removal, the Brain Panel advances prior supervised ML approaches, providing a robust tool for EEG pre-screening that complements human expertise while addressing limitations of normative and AI-based systems.

Explainable, clinically integrated ML approaches are increasingly being applied to EEG analysis [10]. Researchers have developed an ML model to facilitate the detection of interictal epileptiform discharges (IED), which helps with the speed in overall IED detection [11]. Using ML algorithms like this can improve the ability of medical professionals to read the EEG and, at the same time, double medical professionals’ accuracy by greatly improving their decision-making rather than telling them what to do, especially when dealing with large amounts of data. Another study focuses on data-driven approaches that can extract biologically meaningful features from population-level clinical EEGs without artifact rejection [12]. These works also intersect with recent efforts to standardize EEG acquisition, which aim to reduce metric variability seen in earlier studies [4]. Unlike systems focused solely on diagnostic outcomes, such as seizure detection, the Brain Panel prioritizes quality control and clinical assistance, positioning it as a bridge between traditional visual inspection and emerging automated tools. Its potential as a feature set for neural networks aligns with ongoing research into hybrid ML-AI systems, suggesting a pathway for future integration with self-learning frameworks to enhance EEG interpretation.

3. Methods

3.1. Study design and overview

This study employed a retrospective observational design to develop and validate the Brain Panel, a supervised ML tool for automated pre-screening of clinical EEGs. The system was designed to quantitatively approximate key features of physician visual EEG inspection without replacing clinical judgment. The study proceeded in three phases: (1) construction and statistical validation of a clinical EEG reference database, (2) development and automated deployment of the Brain Panel reporting system, and (3) comparative validation of Brain Panel outputs against paired neurologist quality review reports from a subset of 100 recordings, using two independent AI language model platforms [13, 14].

The ML approach used here is distinguished from general AI in that all analytic rules and metric definitions were fixed and investigator-specified prior to analysis. The system does not learn or modify its own rules; instead, it applies a stable set of

predefined computations to each new EEG, accumulating population statistics as the database grows. This ensures reproducibility: the same EEG evaluated at any time will yield identical metric values.

3.2. EEG database construction

3.2.1. Source and inclusion/exclusion criteria

The Brain Panel reference database was constructed from raw EEG recordings submitted by clinicians to a clinical EEG quality review service operated by the authors. All recordings included in the database were routine 20-min EEGs submitted within the three years prior to January 1, 2026. EEGs were accepted from any subscribing clinician without pre-selection for clinical normalcy, yielding a database that reflects the full spectrum of recordings encountered in a real-world clinical screening service, including those with artifacts, technical issues, and clinical abnormalities.

Inclusion criterion: Any clinician-submitted EEG recording with a duration of ≥ 60 s.

Exclusion criterion: Any EEG recording with a duration of < 60 s. No additional exclusion criteria were applied; recordings with clinical abnormalities, excessive artifacts, or clinically significant deviations were retained in the database.

This inclusive approach deliberately departs from normative database construction, in which recordings from individuals with neurological or psychiatric conditions are typically excluded. The rationale for this approach is to produce a reference that is representative of actual clinical submissions rather than an idealized healthy sample.

QEEG databases have established themselves as being useful for supporting diagnosis in the healthcare field. The database currently being constructed is derived from submissions from subscribing clinicians and therefore, by definition, does not constitute a “normative” database, but rather one that represents what is seen in a clinical EEG service and analyzed with regard to what a visual inspection of the EEG might state.

3.2.2. Participants

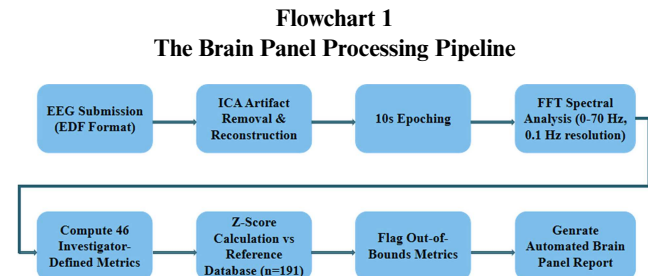
A total of 191 EEG recordings were included in the reference database as of January 1, 2026. Recordings were drawn from a clinically diverse population including children, adolescents, and adults ranging in age from 4 to 80 years, of both sexes. Patients had been referred for a variety of mental health, behavioral, and neurological concerns, including attention-deficit/hyperactivity disorder, anxiety, depression, autism spectrum disorder (ASD), trauma, traumatic brain injury (TBI), and other conditions. No additional demographic matching or stratification was applied.

3.2.3. EEG acquisition

All EEGs were acquired using an FDA-registered EEG amplifier manufactured by BrainMaster Technologies, Inc. (Bedford, OH, USA) in accordance with standard procedures for a clinical 20-min EEG. The standard acquisition protocol consisted of 10 min with eyes open followed by 10 min with eyes closed. Electrode placement followed the International 10–20 system at all 19 standard scalp sites (Fp1, Fp2, F3, F4, F7, F8, T3, T4, T5, T6, C3, C4, P3, P4, Fz, Cz, Pz, O1, O2, and linked ears reference). Recordings were referenced to linked ears throughout. All recordings were submitted to the screening service in European Data Format (EDF).

3.3. Signal processing and feature extraction

The Brain Panel processing pipeline was fully automated and applied identically to every submitted EEG. No human intervention occurred between EDF upload and report generation. The processing steps are described below and summarized in Flowchart 1.



3.3.1. Artifact removal

ICA was applied to each EEG recording to identify and remove ocular artifact components [15, 16]. All channels identified by ICA as containing ocular activity were removed, and the remaining components were used to reconstruct a cleaned EEG signal. This step was applied prior to any spectral or metric computation.

3.3.2. Epoching

Following artifact removal, each reconstructed EEG was divided into consecutive non-overlapping 10-s epochs. Any final epoch shorter than 10 s was discarded. The 10-s epoch length was chosen to provide a frequency resolution of 0.1 Hz in the subsequent FFT analysis.

3.3.3. Spectral analysis

For each epoch and each of the 19 electrode channels, spectral power was computed using the FFT. The analysis covered a frequency range of 0–70 Hz, yielding a spectral resolution of 0.1 Hz. This high-resolution approach enables precise identification of narrowband features such as the posterior dominant rhythm (PDR), peak frequency, individual alpha sub-bands (Alpha 1 and Alpha 2), and 60 Hz power-line noise as a signal quality indicator. All spectral metrics were computed on the ICA-cleaned signal.

3.4. Brain Panel metric definitions

A total of 48 investigator-defined metrics were computed for each EEG recording. All metrics were specified prior to data collection and remained fixed throughout the study. The metrics are organized into seven categories: Statistical Measures, PDR, Phenotypes, Focal Measures, Frontal Measures, Diffuse Measures, and State Shifts (Moments). Full definitions for all 48 metrics are provided in Appendix A. Key metric categories are described below.

Statistical Measures: RAW STD (power in the raw signal) and Global STD (power in the ICA-filtered signal) quantify overall amplitude deviations from the sample population mean, providing a direct index of ICA artifact removal efficacy.

PDR Metrics: Nine metrics characterize the PDR, including symmetry (left–right balance at O1/O2), synchrony (coherence between O1 and O2), regulation (an entropy-based measure of

momentary variability in occipital alpha activity), magnitude, sinusoidal purity, posterior dominance index (front/back amplitude ratio), FFT bandwidth, peak amplitude, burst width. The regulation metric is derived from Shannon entropy and reflects the degree of self-organization in the PDR, with low entropy indicating rigidly rhythmic activity and high entropy indicating disorder.

Phenotype Metrics: Six metrics capture hemispheric and regional features: frontal beta asymmetry, frontal asymmetry ratio, excessive temporal alpha (at T3, T4, T5, T6), alpha speed (Alpha2/Alpha1 ratio), alpha peak frequency, and midline beta (Fz, Cz, Pz).

Focal Metrics: Four metrics characterize activity at the frontal electrode sites (Fp1, Fp2, F3, F4, F7, F8): delta, theta, and gamma power, plus a frontal gamma asymmetry index. Frontal gamma asymmetry provides an index of approach–avoidance tendencies.

Diffuse Metrics: Seven metrics reflect global brain activity by summing power magnitudes across all channels in the delta, theta, beta, high-beta, and gamma bands, plus a 60 Hz metric (a non-biological signal quality indicator) and a fractal dimension metric. Fractal dimension quantifies the computational complexity of global EEG activity using a standard fractal algorithm and has been associated with cognitive function and brain state.

Frontal Metrics: Four metrics quantify power at the frontal electrode sites (Fp1, Fp2, F3, F4, F7, F8) in the delta, theta, gamma, and gamma asymmetry bands. Frontal gamma asymmetry provides an index of approach–avoidance tendencies.

State Shift Metrics (Moments): Twelve metrics adapted from mechanical engineering moment analysis describe the temporal distribution of spectral power across the recording session in four frequency bands (PDR/alpha, delta, theta, beta). For each band, three moments are computed: Moment 1 (total energy across the session), Moment 2 (temporal center of gravity, i.e., where energy is concentrated in time), and Moment 3 (temporal variability or spread of energy, sensitive to state shifts such as drowsiness or alertness changes). These metrics are novel in the EEG literature and provide a structured approach to detecting within-session state changes that are otherwise assessed informally by visual inspection.

3.5. Statistical reference database and Z-score computation

For each of the 48 metrics, population-level statistics (mean (μ) and standard deviation (σ)) were computed by pooling metric values across all recordings in the reference database. These statistics were updated with each new submission. Individual z-scores were computed for each metric in each recording as $z = (x - \mu)/\sigma$, where x is the metric value for the individual recording. Z-scores with absolute values exceeding predefined out-of-bounds (OOB) thresholds were flagged as potential areas of clinical or technical concern.

The distributional properties of the 48 metrics were evaluated to determine the appropriateness of z-score normalization. For each metric, a frequency histogram was generated, and data were assessed against a theoretical Gaussian distribution using Quantile–Quantile (Q–Q) plots and the Kolmogorov–Smirnov (K–S) goodness-of-fit test. Analyses were conducted independently on two database subsamples ($n = 94$ and $n = 191$) to evaluate the stability of distributional properties across sample sizes. Larger database sizes yielded slightly wider reference bands without substantially altering z-scores, corroborating prior findings [7].

The 60 Hz interference metric, which reflects line noise rather than biological signal, was excluded from normality analyses. Most metrics yielded K–S p -values above 0.90, supporting the assumption of approximate Gaussianity consistent with Central Limit Theorem expectations for population-level means. The stability of metric means across the two sample sizes was further evaluated using scatter plot regression and R^2 analysis, confirming that the reference statistics were robust and internally consistent.

It is important to note that the resulting z-scores are not normative in the traditional QEEG sense. They reflect deviation from the mean of the clinical reference population—which includes individuals with a wide range of neurological and psychiatric presentations—rather than deviation from a healthy control sample. This distinction is intentional and is consistent with the Brain Panel’s purpose as a quality and pre-screening tool rather than a diagnostic instrument. It deliberately employs a clinical rather than normative reference database, aligning with critiques of excluding individuals with disorders, such as ASD, from normative comparisons [17]. The approach meets several established standards for database construction (e.g., amplifier matching, signal integrity, population definition) but omits strict acceptance/rejection criteria typical of normative databases.

3.6. Automated report generation

The Brain Panel report was generated automatically for each submitted EEG without any human intervention, using the following sequential steps:

- 1) Upload of EEG recording in EDF format to the Brain Panel processing system.
- 2) Application of ICA for ocular artifact identification and removal.
- 3) Reconstruction of the cleaned EEG signal from retained ICA components.
- 4) Division of the cleaned signal into 10-s epochs; discard of any final partial epoch.
- 5) FFT computation for each epoch and channel across the 0–70 Hz frequency range (0.1 Hz resolution).
- 6) Computation of all 48 investigator-defined metrics for each recording.
- 7) Z-score computation relative to the current reference database statistics.
- 8) Identification of OOB metrics using predefined threshold rules.
- 9) Generation of a structured text report listing all OOB metrics with brief standardized interpretive comments describing their potential clinical or technical significance.

The report format was designed to emulate the structure of a physician’s EEG quality review report, commenting on background rhythm characteristics, signal quality, the possible presence of artifacts, drowsiness, and clinical abnormalities. The Brain Panel report is intended to be reviewed prior to any human visual EEG inspection or QEEG analysis, functioning as a “pre-Q” or “pre-pre-Q” pre-screening step that directs clinical attention to the most informative recording features.

3.7. Comparative validation against neurologist reports

To evaluate the clinical utility of the Brain Panel, a subset of 100 EEG recordings for which both a complete Brain Panel report

and a corresponding quality review report from a board-certified neurologist were available was assembled. The neurologist reports were generated independently as part of the standard clinical review service and were blind to the Brain Panel output.

3.7.1. Dataset construction

Brain Panel outputs for each of the 100 recordings were manually transcribed into a structured spreadsheet. For each recording, the total number of OOB metrics was recorded for each of the seven metric categories. Simultaneously, the text of the corresponding neurologist report was parsed and categorized into the following clinical dimensions: (1) Background Rhythm, (2) Drowsiness/Sleep, (3) Paroxysmal Disturbances, (4) Physician Comments, and (5) Full Report Text. EEG quality was coded from the physician's report as good, fair, or poor. Artifact level was coded as mild or moderate. Clinical abnormality was coded as present or absent based on the physician's narrative. The resulting spreadsheet contained one row per recording, with all Brain Panel metric counts and all neurologist report classifications represented as columns.

3.7.2. AI-assisted discriminant analysis

The paired dataset was submitted to two independent AI language model platforms—Grok v3 and Claude Sonnet 4—for comparative evaluation. Each platform was queried using a standardized series of natural-language prompts, as follows:

- 1) Confirmation that the dataset structure was understood and that the two data sources (Brain Panel metrics and neurologist comments) had been correctly interpreted.
- 2) Assessment of dataset validity ($n = 100$).
- 3) Evaluation of the overall helpfulness of the Brain Panel to the neurologist.
- 4) Generation of a compare-and-contrast summary of the Brain Panel findings versus the neurologist report content.
- 5) Suggestions for enhancement of the Brain Panel utility.

Following initial evaluation, each AI platform was queried to identify the optimal linear combination of OOB metric counts that best predicted each of the following clinically defined outcomes: (1) clinical brain abnormality, (2) drowsiness, (3) artifact level, (4) paroxysmal events, (5) PDR frequency abnormality, and (6) overall EEG quality. For each outcome, the AI platform was asked to find the weighted combination of OOB metric counts that maximized predictive sensitivity, and to generate charts showing the discriminant performance of both single-metric and multi-metric approaches.

Sensitivity, specificity, and overall accuracy were computed for each derived algorithm by comparing predicted outcome classifications against neurologist-assigned ground-truth labels across the 100-recording sample. Optimal weighting coefficients were determined empirically by each AI platform based on the observed data. The resulting combined metrics and their performance statistics are reported and summarized in Section 4.

3.7.3. Database standards compliance

The Brain Panel database construction was evaluated against established standards for QEEG database development [18]. The system satisfies requirements for amplifier specification and matching, signal integrity verification, and population definition. It does not meet criteria for normative database construction, as strict clinical acceptance/rejection criteria and the exclusion of individuals with neurological or psychiatric diagnoses were

deliberately omitted. This is consistent with the stated purpose of the database as a clinical reference for quality screening rather than a normative comparison.

4. Results

4.1. Statistical procedures

Statistical validation of the reference database confirmed the suitability of Z-scores for the 48 Brain Panel metrics. Q-Q plots and K-S goodness-of-fit tests performed on independent subsamples ($n = 94$ and $n = 191$) demonstrated approximate Gaussianity for the majority of biological metrics (most $p > 0.90$), consistent with Central Limit Theorem expectations. Scatter-plot regression analyses further confirmed high stability of population means across database sizes ($R^2 > 0.95$). These analyses are summarized in Figures 1, 2, and 3 (detailed plots and additional validation figures are provided in Appendix B).

To evaluate clinical utility, the first 100 Brain Panel reports were paired with corresponding board-certified neurologist quality assurance reports. Neurologist annotations were categorized (EEG quality, artifacts, drowsiness, paroxysmal activity, and clinical abnormalities) and aligned with Brain Panel OOB counts. AI-assisted discriminant analysis (Grok v3 and Claude Sonnet 4) was used to identify optimal weighted combinations of metrics for each clinical target. The resulting algorithms achieved sensitivities of 89–95%, specificities of 76–88%, and overall accuracies of 78–91%. These findings are discussed in Appendix B.

We used the Kolmogorov–Smirnov Test, K-S Test, as the metrics were all measured values, to demonstrate Gaussian normality as depicted in Figure 2. We generated histograms of the p -values from each sample database to demonstrate that the null hypothesis cannot be rejected, supporting the conclusion that these samples conform to a Gaussian distribution.

Comparing the sample databases next to each other clearly shows that all the biological metrics are reliable. The variability tends to widen as the sample size gets larger, which is similar to mentioning that the Z-scores widen as the sample size increases. This was demonstrated by Collura and Tarrant when they showed what individual means create static population means which are similar to live population means, as shown in Figure 3 [4, 7]. Each individual bell curve represents one person's personal mean value. When samples from multiple individuals are combined, the range of mean values for the dynamic measurements in the population becomes wider than the range observed for the static population values (see Figure 3(a)). Figure 3(b) illustrates the same principle applied to two separate samples. As the sample size increases, the mean value remains relatively stable, while the range of the corresponding Z-scores progressively widens.

Next, we used a linear regression model and scatter plot trend lines to show how well the sample databases trended together. These correlations are illustrated in Appendix B, Figure B1, which presents a series of scatter plots with trend lines and R^2 values comparing metrics from two sample databases ($n = 94$ and $n = 191$). The panels progressively zoom in and filter out non-biological metrics; removing specific outliers such as 60 Hz and Raw STD strengthens the correlation between the samples. These visualizations illustrate how removing non-biological outliers improves the trend alignment, suggesting potential for future clinical insights as the database expands.

Our final preliminary analysis involved comparing and converting the t -values to p -values. These results are presented in Appendix B, Figure B2, which shows that most metrics exceed

Figure 1
Q–Q plots of five Brain Panel metrics: (a) PDR Regulation. (b) PDR Symmetry. (c) Fast Alpha. (d) PDR Synchrony. (e) XS Temp. Alpha

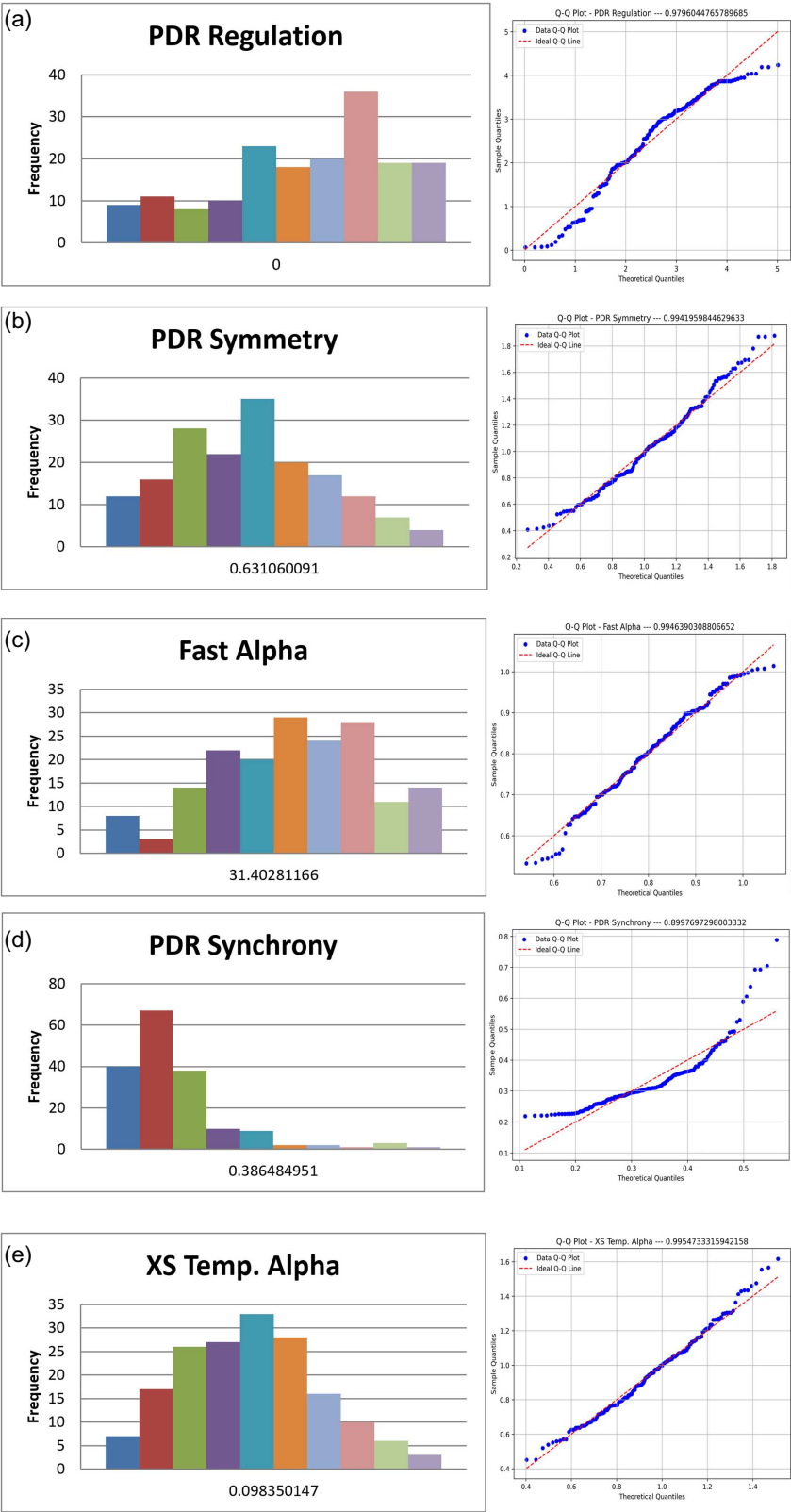


Figure 2

Histograms of Kolmogorov–Smirnov goodness-of-fit p-values for Brain Panel metrics across subsamples: (a) $n = 94$ (46 metrics), (b) $n = 94$ (45 metrics, 60 Hz omitted), (c) $n = 191$ (46 metrics), (d) $n = 191$ (45 metrics, 60 Hz omitted), and (e) $n = 94$ and $n = 191$ (45 metrics) shown side by side

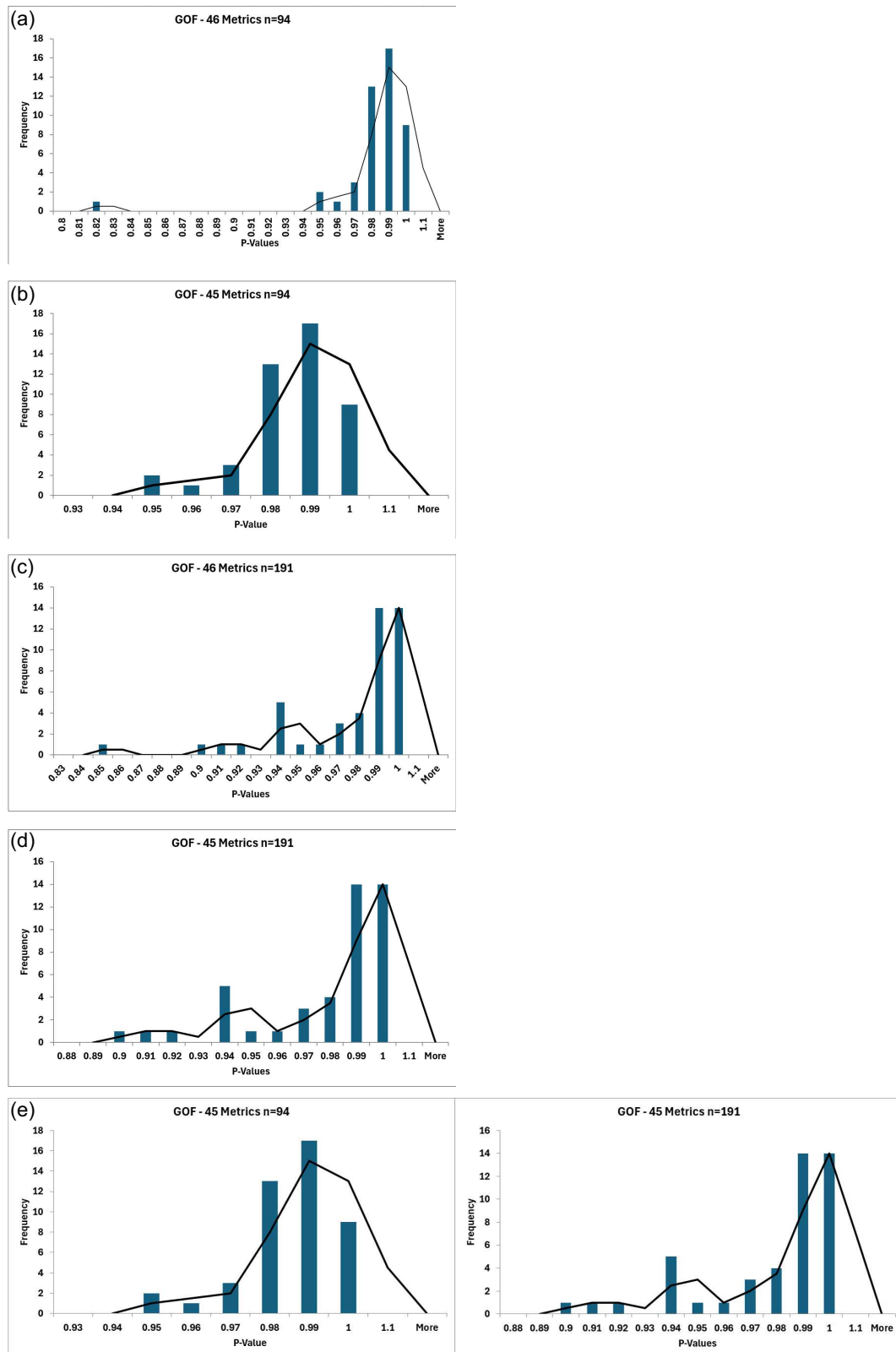
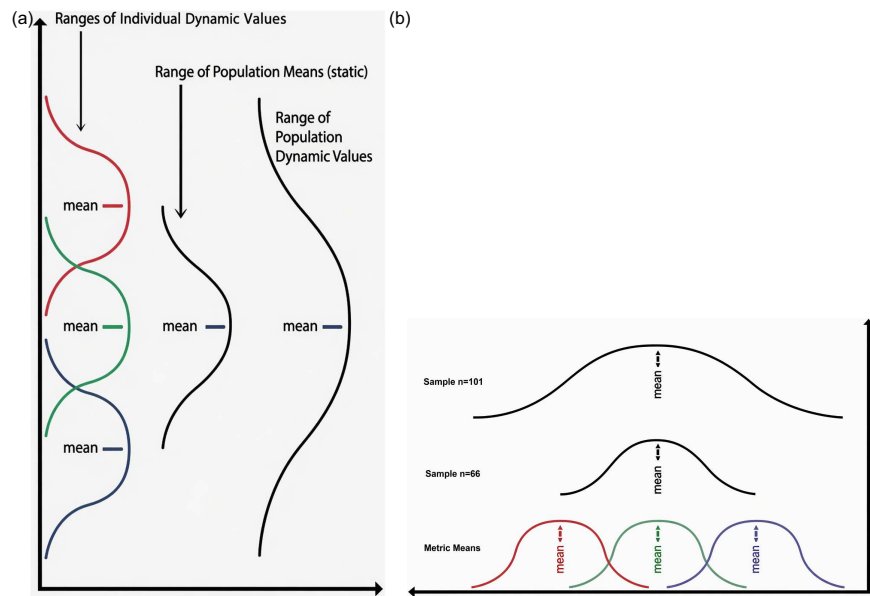


Figure 3

Sample size effects on population means and Z-score ranges for dynamic EEG metrics: (a) conceptual illustration of individual versus population distributions and (b) comparison of small versus larger subsamples

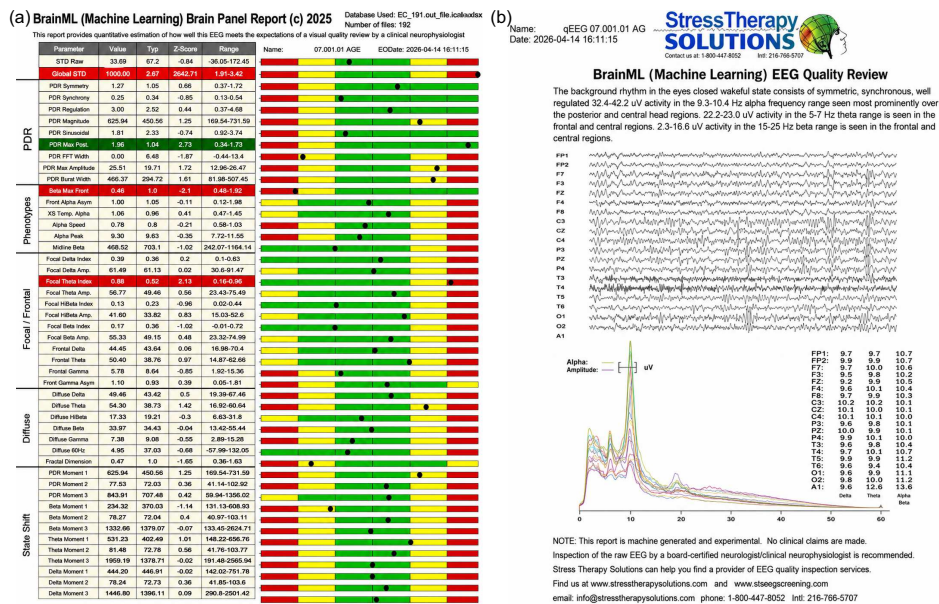


acceptable goodness-of-fit thresholds ($p > 0.90$), with 10 metrics above 0.95 and only 10 falling below 0.85. After running multiple statistics and satisfactorily establishing a valid database that produces Z-scores, we can now define the metrics of interest in the Brain Panel, as shown in Figure 4. Figure 4(a) shows that the metrics in the Brain Panel consist of two statistical measures and six categories of neurometrics. They are listed and defined in Appendix A. Figure 4(b) shows a standard snapshot of the client's EEG. We show a familiar FFT of all the frequency bands. Here is where another new feature can be seen. The 0.1 Hz high resolution of the FFT allows us to more

accurately point out changes in the frequency bands, especially with the Alpha 1 and Alpha 2 bands and the PDR. This will allow clinicians to more accurately track changes for the client when symptoms improve. We can measure the amount of 60 Hz interference that potentially could be corrupting the EEG. Any metric found to be OOB comes with a brief definition of its potential clinical or technical impact on the individual, as illustrated in Appendix B, Figure B3. To further validate the Brain Panel, we created a dataset of $n = 100$ Brain Panel reports, paired with the corresponding neurologist reports, comparing the metrics the Brain Panel

Figure 4

Typical Brain Panel report and high-resolution FFT: (a) automated report showing out-of-bounds (OOB) metrics with interpretive comments and (b) high-resolution (0.1 Hz) FFT snapshot of the client's EEG



highlighted as being deviant to the doctors' comments on the same EEG.

After the statistical analyses were completed, we evaluated the practical utility of the Brain Panel outputs by sampling de-identified cases and their accompanying quality assurance reviews done by a neurologist. To assess the value of these outputs, we posed the five questions shown below to Grok and Claude

regarding the same dataset. The results are shown in Tables 1, 2, and 3. The optimal weighted combinations of the Brain Panel OOB metrics achieved sensitivities of 89–95%, specificities of 76–88%, and overall accuracies of 78–91% (see Table 4).

- 1) If I upload an Excel file with two sets of data and explain them to you, can you compare the datasets and make conclusions please?

Table 1
Comparison of responses of GROK to CLAUDE when posed the question: Is the Brain Panel useful?

GROK	CLAUDE
Dataset Validity The dataset appears to consist of approximately 100 entries (rows 4–103 in the provided sheet, yielding ~99 patient records after accounting for headers in rows 1–3). It is a compilation of EEG analysis data for various patients, comparing automated “Brain Panel” (BP) metrics and comments against manual neurologist (DR) reviews.	Dataset Validity: ☒ VALID – $n = 100$ with complete paired observations
Is the Brain Panel Helpful to the Neurologist? Overall, yes, the Brain Panel is moderately helpful as a screening tool to flag potential issues, reduce review time, and quantify EEG features.	Brain Panel Usefulness: MODERATELY HELPFUL - Good screening tool but needs refinement.

Table 2
GROK compare-and-contrast chart

Aspect	Brain Panel (BP)	Neurologist (DR) comments	Similarities	Differences
Approach	Quantitative and automated: “Counts ‘out-of-bounds’ metrics across categories (e.g., PDR: 9 total, Focal: 12). Generates rule-based interpretive comments	Qualitative and clinical: “Descriptive assessments (e.g., ‘Fair quality, moderate artifacts’) with context like clinical correlation	Both identify EEG issues like drowsiness artifacts and use categories (e.g., background rhythm, paroxysmal events)	BP is metric-driven and speculative (“may be a concern”); DR is holistic, integrating visuals/history and definitive (“no abnormalities”)
Content focus	Metrics on deviation (e.g., amplitude, asymmetry, moments) and potential implications (e.g., “high value may indicate drowsiness”). Repetitive phrases like “may suggest hyper-relaxation”	Observational details (e.g., 9.0 Hz PDR, drowsiness by Slow Rolling Eye Movements (SREMs)) and recommendations (e.g., “clinical correlation warranted”). Includes full text with spectral analysis	Overlap on key EEG elements (PDR, delta/theta slowing, state shifts) ~50% alignment on abnormality detection	BP quantifies (e.g., “23 out of bounds”); DR qualifies (e.g., “mild artifacts”). BP lacks visuals; DR references images/pages
Strengths	Objective, fast, scalable for large datasets. Flags subtle quantitative shifts (e.g., Moment 3 for instability)	Expert interpretation, context-aware (e.g., age-appropriate PDR). Detects nuances like “left>right temporal slowing.”	Both aid in triage; BP flags for DR review; DR confirms	BP consistent but generic; DR variable but insightful
Weaknesses	Overcautious/vague (e.g., many “may be” qualifiers). Misses context (e.g., no age adjustment). Higher false positives	Subjective, time-intensive. Potential interrater variability (all by P. David Ims here)	Both can miss sleep dysregulation if not flagged explicitly	BP doesn’t recommend actions; DR does (e.g., “further evaluation”)
Use cases	Screening tool for initial analysis in high-volume settings	Final clinical review, diagnosis support	Complementary in workflow: BP pre-screens; DR validates	BP for research/stats; DR for patient care
Examples from data	Row 101: 23 flags; comments on diffuse delta/theta as concerns (matches DR’s low PDR/slowing). Row 8: 0 flags (mismatches DR’s drowsiness)	Row 4: “Drowsiness demonstrated, no abnormalities” (aligns with BP state shifts). Row 103: “No abnormalities” despite BP’s 26 flags	High BP totals often correlate with DR artifact drowsiness notes (e.g., rows 4,10)	BP speculates (e.g., “TBI, toxicity”); DR factual (e.g., “intermittent slowing”)

Table 3
CLAUDE compare-and-contrast chart

Aspect	Brain Panel	Neurologist assessment	Agreement level
Method	Quantitative algorithms	Qualitative expert interpretation	N/A
Objectivity	Highly objective, reproducible	Subjective, variable between reviewers	Low–Medium
Speed	Instant automated analysis	Time-intensive manual review	High
Coverage	Systematic 48 metrics across domains	Focused on clinically relevant findings	Medium
Clinical context	Limited contextual awareness	Rich clinical interpretation	Low
Artifact detection	Algorithmic pattern recognition	Expert visual identification	Medium–High
Abnormality detection	Rule-based threshold exceedance	Pattern recognition + experience	Medium

- 2) Columns C–K are metrics in the Brain Panel; comments L–S are comments made by a neurologist. They are looking at a person's EEG. Is this dataset of 100 n valid?
- 3) Is the Brain Panel helpful to the neurologist?
- 4) Can you make suggestions to enhance the usefulness of the Brain Panel?
- 5) Can you make a compare-and-contrast chart of the Brain Panel and neurologist comments?

Table 1 summarizes the comparative responses of Grok and Claude to the two essential questions of the analysis: Is the data valid, and is the Brain Panel helpful? Both platforms independently concluded that the dataset was valid and the Brain Panel is moderately helpful to neurologists, a good screening tool requiring some refinement.

Table 2 presents Grok's version of a compare-and-contrast chart between the Brain Panel and the doctor's comments. An interesting observation is that Grok provided a data-driven analysis from the actual values of the 100-sample spreadsheet, yielding clinically specific observations regarding EEG patterns, agreement rates, and potential improvements for the Brain Panel.

Table 3 displays Claude's version of the compare-and-contrast chart. Claude's response remained largely generalized without direct quantitative analysis with the metrics contained in the dataset.

An interesting observation is that Grok considers and evaluates actual data from the dataset spreadsheet from the 100 samples. Therefore, Grok's comments are specific to this system, the clinical aspects, and the results it achieves, while Claude made a superficial comparison of the dataset format and did not compare real data directly. Claude's comments would be relevant to any system of this general type.

4.2. Claude's Brain Panel effectiveness analysis

We asked Claude for an effectiveness analysis comparing OOB metrics to doctor-identified issues. Claude concluded there was a strong correlation in the ability of the Brain Panel to predict EEG quality and the presence of artifacts. Claude was able to construct the paired analyses and produce reports of the average values of the number of OOB metrics divided by EEG quality and by artifact level as reported in the clinical report. We present the output from Claude exactly as formatted by the AI agent, without modification. Claude's analysis showed a moderate, clear relationship between the total number of OOB metrics and lower EEG quality or higher artifact levels, as illustrated in Appendix B, Figure B4.

Claude's analysis showed no strong correlation between the total number of OOB metrics and the presence of clinical abnormalities, as illustrated in Appendix B, Figure B5. Overall, the Brain Panel metrics are not sensitive to EEG abnormalities specifically. Therefore, we undertook to explore how the metrics could be used to correlate with clinical findings, using a discriminant approach.

When asked to create a weighted score that optimally predicted clinical abnormality, Claude produced a score consisting of $\text{PDR} \times 3 + \text{Focal} \times 2 + \text{Diffuse} \times 2 + \text{State Shift}$. This metric was found to be 89% sensitive in finding clinical abnormalities, as illustrated in Appendix B, Figure B6. Figure B7 further shows the key insights and clinical implementation for this new combined metric. It provides "excellent risk stratification" and a "neurophysiologically informed" measure, with a clinical implementation strategy ranging from very low risk (score 0–2) to high risk (score 8 or greater), thereby reducing unnecessary neurologist reviews while maintaining high sensitivity.

We further asked Claude to evaluate and create optimal detection methods for drowsiness, artifacts, paroxysmal events, and PDR frequency abnormalities. These results are presented in Appendix B, Figures B8–B13. The state shift metric serves as an objective drowsiness indicator (Figure B8). An optimal drowsiness detection algorithm was developed using the combination $\text{State Shift} \times 3 + \text{PDR} \times 2 + \text{Diffuse}$, achieving 93% sensitivity (Figure B9). Similarly, optimal algorithms were created for artifact detection (Figure B10), paroxysmal events (Figure B11), and PDR frequency assessment (Figures B12 and B13), each demonstrating strong clinical correlation and high sensitivity.

Finally, in order to develop a comprehensive quality index, we asked Claude to identify the optimal combination of OOB metrics that best correlates with overall EEG quality. The resulting equation ($\text{STD/Global} \times 4 + \text{Diffuse} \times 2 + \text{Total}$) demonstrated a strong correlation with neurologist quality ratings, as illustrated in Appendix B, Figure B14. Figure B15 further summarizes the quantitative performance of the six different optimized metrics created by Claude, showing significant improvement in the EEG quality index when combining Brain Panel metrics in this specific way.

Table 4 summarizes the findings that these new combined metrics have sensitivity in the range of 89–95% for the clinical findings and an overall accuracy of 78–91%. Therefore, by using AI to discern and define new combinations of metrics, the accuracy of the clinical correlation is significantly enhanced. Here, the Brain Panel's fixed metrics remained unchanged; the optimization was performed externally by the AI platforms, which identified weighted combinations of the existing metrics rather than modifying the Brain Panel itself.

Table 4
Summary of the weighting values, and resulting sensitivity, specificity, and accuracy determined from 100 sample Brain Panel reports compared with the M.D. neurologist's report

	Definition and performance of optimal detection algorithms								
	Std/Global	PDR	Focal	Diffuse	State shift	Total	Sensitivity	Specificity	Accuracy
Clinical abnormality	0	3	2	2	1	0	89	76	78
Drowsiness	0	2	0	1	3	0	93	79	87
Artifact	3	0	0	2	0	1	92	83	88
Paroxysmal	0	2	3	0	1	0	94	88	91
PDR frequency	0	4	0	2	1	0	95	84	91
EEG quality	4	0	0	2	0	1	94	88	90

Our findings confirm that the groups of metrics that were created do in fact reflect different clinically relevant aspects of the EEG. The PDR metrics, for example, reflect internal self-regulation, whereas the focal/diffuse metrics are found to be more indicative of neuropathy. The moment analysis is demonstrated to reflect changes that occur in time across the recording session, reflecting instability or drowsiness. Other combinations of metrics reflect the likelihood that the EEG will be considered to have excessive artifacts or that the EEG will be deemed abnormal by the clinical inspection. The comments produced by the Claude analysis further support these conclusions, as they include specific statements confirming the physiological and clinical relevance of particular metrics. We have thus used the AI tools to teach us how to read a Brain Panel, further enhancing the value of this method. The AI tool has in effect become an expert at reading Brain Panels and, by comparing them with doctors' reports, has learned how to use the Brain Panel to anticipate what the physician may find.

In summary, the Brain Panel metrics, when used in optimal combinations, provide a sensitive and accurate measure of the likelihood of correlation with clinical findings. The application of AI to the results has made it possible to optimize the use of the predefined metrics to produce a clinically useful result.

5. Discussion

To contextualize the Brain Panel's performance, Table 5 presents a comparative analysis of its metrics against state-of-the-art machine learning and deep learning benchmarks for analogous EEG pre-screening tasks, including artifact detection, drowsiness identification, IED, and clinical abnormality classification [19–25]. Notably, the Brain Panel achieves sensitivities of 89–95%, specificities of 76–88%, and accuracies of 78–91% across multiple clinically relevant domains simultaneously, leveraging a transparent, rule-based supervised ML framework with post hoc AI optimization on an inclusive real-world database ($n = 191$, no exclusions). This outperforms or matches specialized systems like SCORE-AI (88.3% accuracy for automated EEG interpretation [24]) and deep neural network-based IED detectors (82.5% sensitivity at 99% specificity [23]), while offering superior interpretability and applicability to diverse clinical submissions, thereby enhancing neurologist productivity without supplanting expert judgment.

This performance compares favorably with state-of-the-art EEG-AI systems while offering distinct advantages in transparency and clinical workflow integration. Unlike many deep learning models that lack interpretability and rely on extensive artifact rejection or curated datasets [26, 27], the Brain Panel uses predefined, neurophysiologically motivated metrics that align with calls for interpretable ML in clinical neurophysiology [9, 28]. The system therefore serves as a practical “pre-Q” tool that enhances, rather than replaces, expert visual inspection.

We have demonstrated that the metrics provide valid and reliable indicators of key EEG parameters related to quality, artifacts, and brain disorders. We found that by applying an AI analysis to 100 of our clinical cases, we were able to produce optimal detection algorithms. These algorithms demonstrably correlated well with clinical findings, thus providing a clinically relevant and useful set of estimates that serve to pre-inform a visual EEG analysis. It is of interest to explore how the metrics appeared in the optimal detection algorithms and what their likely neurophysiological significance is.

Table 5
Comparative analysis of key metrics between the Brain Panel and other state-of-the-art EEG-AI systems

Study/system	Primary task(s)	Method/ap-proach	Sensitivity (%)	Specificity (%)	Accuracy (%) / other	Key notes and dataset
Marchant et al. 2024	Artifact detection (single-channel epochs)	Random Forest (supervised ML)	65 (test set)	98	81 (balanced)	Infant EEG; high specificity typical of artifact tasks; $n = 70$ held-out epochs
van Stigt, M. N., Groenendijk, E. A., Marquering, H. A., Coutinho, J. M., & Potters. 2023	Clean vs artifact classification (dry electrodes)	Convolutional Neural Network (CNN) with transfer learning and majority vote	91.2 (recall)	90.3	90.7	Dry-electrode data; directly comparable ICA-style preprocessing; 2-second segments
Hasan et al. 2022	Drowsiness detection	Random Forest (explainable ML)	70.3	82.2	80.1	Leave-one-participant-out validation; multimodal (EEG + EOG + ECG); addresses inter-individual variability
Minhas et al. 2025	Drowsiness detection	Optimized spectral features + threshold	NA	NA	95.4	Single optimal channel (F4/O2); PSD alpha/-theta pairing; high accuracy with minimal channels
Tjepkema-Cloostermans et al. (DNN) 2025	Interictal epileptiform discharge (IED) detection	Deep neural network	82.5	99	- (False Discovery Rate (FDR) <0.2/min)	Individual IED level; outperforms Persyst (64.6% sens); EEG-level matches expert consensus; internal + external validation
Pooled IED models (meta-analysis) 2024	IED classification (IED vs non-IED)	Various DL/ML (external validation)	78.1–84.9	68.7–80.1	- (Area Under Curve (AUC) ≈ 0.85)	Real-world external datasets (>50 patients); high variability noted
Tveit et al. (SCORE-AI) 2023	Automated clinical EEG interpretation (normal/abnormal + epileptiform categorization)	Deep CNN	86.7 (epileptiform)	90	88.3	Closest benchmark ($n > 30k$ training EEGs); human-expert-level performance on routine clinical recordings
Lemoine et al. (DeepEpilepsy) 2025	Epilepsy detection on routine EEG	Vision Transformer (DL)	65–73	NA	Improved diagnostic yield	Focus on subtle abnormalities beyond visible IEDs; AUC 0.76–0.83 when combined
Brain Panel (this study) 2026	Quality, artifacts, drowsiness, paroxysmal events, abnormalities, PDR frequency	Predefined metrics + AI-optimized combinations	89–95	76–88	78–91	Inclusive real-world database ($n = 191$, no exclusions); fully interpretable, rule-based supervised ML + post hoc AI optimization

STD/Global is the simplest metric, quantifying the total amplitude of the ICA-cleaned EEG and the original EEG submitted. It reflects how much artifact was detected by the ICA. Therefore, this optimal metric validates the concept that the ICA process effectively finds primary issues that relate to EEG quality and discriminates them. In conjunction with the diffuse

amplitudes and the total number of OOB metrics, the STD/Global metric delivers a dependable automated assessment of the EEG recording quality ahead of visual inspection. Interestingly, the proportion of poor recordings, 9%, is consistent with our existing analyses (as reported to the FDA and for ISO auditing purposes), which show that this is the expected number of poor

recordings in a typical clinical setting, so this analysis validates previous clinical findings.

The PDR metrics, many of which are unique to this system, are designed to capture the features of the PDR that are identified when doing a visual inspection of the EEG. The amplitude, symmetry, sinusoidal shape, and other attributes are captured in metrics that go out of bounds whenever any of these visually based features are at an extreme value, signifying an unusual PDR. This is validated by the finding that the optimal detection algorithms for brain disorders emphasize PDR abnormalities. We introduce a new metric called “regulation,” which is based on entropy. Entropy is defined by the amount of disorder in a system and, for the purposes of the Brain Panel, the measure of momentary energy variation in activity during the session. We adopt the standard metric of entropy as an indicator of the quality of regulation. By this measure, a rhythm that is very steady, or rigidly rhythmic, will have low entropy, reflecting poor regulation. EEG entropy has also been associated with conscious awareness, consistent with the concept that the regulation of the posterior alpha, in conjunction with the default mode network, mediates ongoing awareness of internal states [29].

The classical metrics of diffuse and focal amplitudes reflect some of the most commonly used aspects of visual inspection, that is, how large the EEG is overall and how large it is in particular locations. These metrics are included in the optimal detection algorithms, reflecting their relevance. Diffuse deviations reflect generalized brain neuropathy, while focal deviation reflects local dysregulations consistent with head trauma or related localized brain injury or dysfunction.

The gamma bandwidth is getting more attention lately, and we include two specific metrics in our Brain Panel addressing the gamma band: frontal gamma and frontal gamma asymmetry. These metrics may be of more clinical use in the future as demonstrated by research on EEG microstates in relation to emotional decision-making. Frontal gamma asymmetry data acquired in real time using sLORETA in BrainAvatar are shown in Appendix B, Figure B16. The Brain Panel provides an indicator of the subject’s baseline frontal gamma asymmetry using surface EEG, which has been shown to relate to emotional responses and decision-making tendencies to approach or avoidance responses, including extremes of either positive or negative mood or judgment.

Another metric applied in the Brain Panel is fractal dimension: The fractal dimension is a calculation that quantifies the complexity of the energy during the session. Fractal dimension has also been associated with computational complexity as reflected in the EEG; hence, it is of interest in relation to cognitive function [30].

The Brain Panel introduces new metrics to describe the time course of the EEG, referred to as “moments” or state shifts. These metrics, adapted from mechanical engineering principles, provide a novel and structured approach for clinicians to detect temporal changes during the recording session. As illustrated in Appendix B, Figure B17, Moment 1 represents the total energy of the signal across the session, Moment 2 indicates the temporal center of gravity (where energy is concentrated in time), and Moment 3 reflects the amount of state change or variability, which is particularly sensitive to drowsiness and other transitions in brain state. The optimal detection algorithms make strong use of these state shift indicators, especially for identifying drowsiness.

While we have noticed certain metrics have not trended completely to statistical standards, we believe some of them can be explained by differing acquisition techniques, differing office settings, and we also acknowledge our participants may not and

probably do not fit the description of being normal/typical. It is our hope that the newly published Technical Requirements for Performing Clinical QEEG will help alleviate some of these issues [4]. We hope that this paper will promote a more robust dialogue between the clinical and the AI worlds involving the use of EEG. We also believe that as QEEG acquisition becomes more standard across the field, it can only help make the Brain Panel a more robust clinical assessment tool in the field of neuroscience.

6. Conclusion

The Brain Panel system provides automated screening of EEGs, but it is not AI. With AI, a computer learns to recognize patterns based on presented information and creates its own rules. We often do not know or understand the rules the computer has created. We take a different approach. We tell the computer what is important, and it learns what is typical by looking at a population. While this provides a Z-score, these are not “normative” Z-scores, in that many of them reflect a metric that is not overtly indicative of brain state or health. Rather, we use metrics motivated by what a human does when inspecting an EEG to determine its quality. This method is guided by an experienced scientist who defines the analysis used. This method is not a QEEG although it uses Z-scores to quantify EEG properties. Z-scores are utilized to show extreme values of well-defined metrics but are not specifically diagnostic of particular clinical conditions.

As a form of quality control and as an assist to a clinical neurophysiologist, this system provides a level of automation that eases the burden when reading EEGs on a routine basis. This is a tool that provides an aid to visual inspection but does not replace any of the human tasks required for sound EEG analysis or preparation for further processing. In addition to metrics reflecting relative amplitudes and distributions of key metrics, the report analyzes “moments” that reflect the total size of a component amplitude, as well as its distribution within and across the recording session. Moment analysis replaces the “front and back” inspection often used when reading EEGs. Extreme moments tend to reflect drowsiness and other changes across a recording, showing changes across time. These are similar but not identical to reliability measures such as split-half and test-retest reliability. The system also adopts the idea of using entropy as an assessment neurometric reflecting overall EEG complexity.

The Brain Panel metrics constitute a good candidate for the input to a neural network or other self-learning systems. By parsing the EEG into meaningful metrics, a learning system is relieved of the burden of trying to figure out what O1 and O2 are supposed to look like, what a PDR is, and why focal differences matter. Our approach first “uploads” to the computer an indication for the types of things an experienced neurologist looks for and then asks the computer to compile hundreds of results to see how they compare. The resulting analysis is comprehensible, since all calculations and processes are defined beforehand and do not change. However, a learning system may learn to make more sense of the Brain Panel than we can in the long run by inspecting hundreds or thousands of Brain Panels, along with the EEG recordings. The physician’s report can also be included in this approach, so that the system confirms its relationship with clinical findings, compared with the machine results.

Further investigation needs to be done to determine possible improvements and modifications to the metrics and to help understand why certain metrics may be more variable than others. This variability may be understandable though, considering the inclusiveness of the database, which currently has no

exclusion criteria and no pre-artifacting. Clinical correlation studies are also planned to map the relationship between Brain Panel findings and the results of human analysis. And, with the introduction of new practice guidelines for QEEG acquisition [4], we believe the Brain Panel will become an even more refined tool and lead the way toward a more systematic evaluation of EEG, with machine assistance applied to enhance human activity. It may help reduce the potential for EEG misinterpretation, improve interrater agreement to optimize routine interpretation, and increase efficiency much like the SCORE-AI model (Standardized Computer-based Organized Reporting of EEG) [24].

Lastly, this study successfully developed and validated a useful panel to assist the neurologist with EEG analysis using supervised ML algorithms and clinical data from 191 clients and evaluated using 100 recordings and corresponding clinical reports. The Brain Panel was found to be useful to the neurologist and clinician by two different AI platforms. This pre-screening panel will support physicians in the EEG review process by having the panel be immediately available to highlight key metrics they will want to pay extra attention to, thus enabling a more accurate review process for better individualized treatment strategies.

7. Future Works

The Brain Panel's supervised ML framework lays a robust foundation for EEG pre-screening, but several avenues warrant further exploration to enhance its utility and impact. First, refining the 48 metrics, particularly those exhibiting variability due to acquisition differences, is critical. Ongoing standardization efforts, such as those outlined in Collura's work [4], should be leveraged to minimize inconsistencies across clinical settings, potentially through automated calibration protocols or advanced signal processing techniques. Second, clinical correlation studies are needed to map Brain Panel metrics to specific neurological and psychological outcomes, building on preliminary findings [7]. Such studies could validate the system's ability to detect subtle dysregulations, enhancing its role as a clinical decision-support tool. Third, the Brain Panel's metrics show promise as inputs for neural networks or hybrid ML-AI systems. Fourth, the Brain Panel's metrics have the potential to derive brain-specific biomarkers and guide future innovation to isolate relevant brain activity [31].

It should be possible to improve the performance of the system by providing more detailed information to the AI analysis routines, for example, the exact direction of deviations (above bounds or below bounds) and which metrics are specifically deviant, instead of simply counting deviations per metric type. The analysis could also include diagnoses, symptoms, or medication information from the clients' intake forms. These would help the system have greater discrimination, as well as deeper clinical relevance. Given the performance achievable with the current simple approach, further refinement should produce accurate outcomes well into the 90 percent range, without any changes to the Brain Panel itself.

Future work could focus on several key elements to enhance the system's performance and clinical utility. First, integrating the Brain Panel with deep learning frameworks could enable correlation of its metrics with EEG signals and physician reports, potentially improving diagnostic accuracy while preserving interpretability. Similarly, incorporating elements of the SCORE EEG standard [32] could increase the panel's practical value and promote higher-quality clinical EEG reporting overall.

To address current limitations in normative databases, expanding the reference database with larger and more diverse clinical EEG datasets would improve statistical robustness and generalizability. Additionally, direct benchmarking against public datasets such as EEGdenoiseNet, which provides paired clean and noisy recordings for standardized artifact removal evaluation [33], would facilitate objective performance comparisons.

Using the Brain Panel metrics as input features in hybrid deep learning architectures, building on recent high-performing models for clinical EEG classification [10, 34], could further boost detection accuracy without sacrificing interpretability. Larger multicenter validation studies conducted in line with established reporting guidelines for supervised ML prediction models would also strengthen external validity and generalizability [28].

Finally, developing user-friendly visualization tools to present complex metrics, such as moments and fractal dimensions, could facilitate adoption by neurophysiologists, bridging the gap between automated analysis and human expertise. These advancements position the Brain Panel as a scalable, transparent tool for next-generation EEG analysis.

Ethical Statement

The authors declare that this study did not require formal ethical approval because Stress Therapy Solutions does not require Institutional Review Board or ethics committee approval for this type of retrospective chart review. This exemption is based on the fact that the Practitioner-agreement functions as a data use permission from the covered entity side. This study used de-identified EEG recordings submitted to Stress Therapy Solutions under a data use agreement permitting research and development use. Practitioners agree that "All patient identification information should be removed from submitted files." No patient identifiers were included in the analysis dataset.

Stress Therapy Solutions has determined that this work did not constitute human subjects research under 45 CFR 46.102(e). Practitioners warrant that any identified data requires patient authorization, which shifts the identifiable-data obligation upstream to them. STS receives data that should, according to the agreement, be appropriately authorized or de-identified. HIPAA does not apply to de-identified data (45 CFR 164.502(d)(2)) because the de-identification genuinely meets either Safe Harbor (45 CFR 164.514(b)(2)) or expert determination standards. The Practitioner-agreement further gives STS permission for R&D use, such as the internal algorithm development herein reported. Reference: <https://stseegscreening.com/home/about>.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in the Online Data Resource <https://www.dropbox.com/scl/fo/9lvvnoui19c16oehtiarf/AJumW34SIPiwtjmNkRPumCU?rlkey=sw8lxip2dpnlo57eb5ag0w1c3&e=1&dl=0>.

Author Contribution Statement

Thomas Collura: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data

curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration. **Agostino Rosace:** Methodology, Software, Formal analysis. **Robert Turner:** Conceptualization, Validation, Investigation, Resources, Supervision. **David Ims:** Conceptualization, Validation, Investigation, Resources. **Bill Brubaker:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration.

References

- [1] Keizer, A. W. (2021). Standardization and personalized medicine using quantitative EEG in clinical settings. *Clinical EEG and Neuroscience*, 52(2), 82–89. <https://doi.org/10.1177/1550059419874945>
- [2] Cavallo, F., & Brubaker, B. (2024). qEEG/Brainmapping: An essential tool for assessing alternative therapies beyond neurofeedback. *Archives in Neurology & Neuroscience*, 16(3), 1–17. <http://dx.doi.org/10.33552/ANN.2024.15.000887>
- [3] Popa, L. L., Dragos, H., Pantelemon, C., Rosu, O. V., & Strliciu, S. (2020). The role of quantitative EEG in the diagnosis of neuropsychiatric disorders. *Journal of Medicine and Life*, 13(1), 8. <https://doi.org/10.25122/jml-2019-0085>
- [4] Collura, T., Cantor, D., Chartier, D., Crago, R., Hartzoge, A., Hurd, M., . . . , & Turner, R. (2025). International QEEG certification board guideline minimum technical requirements for performing clinical quantitative electroencephalography. *Clinical EEG and Neuroscience*, 56(5), 391–399. <https://doi.org/10.1177/15500594241308654>
- [5] McLaren, J. R., Yuan, D., Beniczky, S., Westover, M. B., & Nascimento, F. A. (2025). The future of EEG education in the era of artificial intelligence. *Epilepsia*, 66(6), 1838–1842. <https://doi.org/10.1111/epi.18326>
- [6] Hamilton, A. J., Strauss, A. T., Martinez, D. A., Hinson, J. S., Levin, S., Lin, G., & Klein, E. Y. (2021). Machine learning and artificial intelligence: Applications in healthcare epidemiology. *Antimicrobial Stewardship & Healthcare Epidemiology*, 1(1), e28. <https://doi.org/10.1017/ash.2021.192>
- [7] Collura, T., & Tarrant, J. (2020). Principles and statistics of individualized live and static Z-scores. *Neuroregulation*, 7(1), 45–45. <https://doi.org/10.15540/nr.7.1.45>
- [8] George, C., & Gulia, K. K. (2025). Machine learning approaches to evaluate EEG correlates of relaxation between supine and sitting postures in eyes-closed condition. *Annals of Neurosciences*, 09727531251341665. <https://doi.org/10.1177/09727531251341665>
- [9] Barnett, A. J., Guo, Z., Jing, J., Ge, W., Kaplan, P. W., Kong, W. Y., . . . , & Westover, M. B. (2024). Improving clinician performance in classifying EEG patterns on the ictal–interictal injury continuum using interpretable machine learning. *Nejm AI*, 1(6), AIoa2300331. <https://doi.org/10.1056/AIoa2300331>
- [10] Wang, C., et al. (2025). Artificial intelligence in electroencephalography analysis for epilepsy diagnosis and management. *Frontiers in Neurology*, 16, 1615120. <https://doi.org/10.3389/fneur.2025.1615120>
- [11] Bagheri, E., Jin, J., Dauwels, J., Cash, S., & Westover, M. B. (2019). A fast machine learning approach to facilitate the detection of interictal epileptiform discharges in the scalp electroencephalogram. *Journal of Neuroscience Methods*, 326, 108362. <https://doi.org/10.1016/j.jneumeth.2019.108362>
- [12] Li, W., Varatharajah, Y., Dicks, E., Barnard, L., Brinkmann, B. H., Crepeau, D., . . . , & Jones, D. T. (2024). Data-driven retrieval of population-level EEG features and their role in neurodegenerative diseases. *Brain Communications*, 6(4), fcae227. <https://doi.org/10.1093/braincomms/fcae227>
- [13] xAI. (2026). Comparing data sets from excel file [Generative AI chat]. Grok. https://grok.com/share/c2hhcmQtMg%3D%3D_53ae1de6-da60-4942-a650-201d63ff4f99
- [14] Anthropic. (2026). Brain panel vs neurologist assessment comparison [Generative AI chat]. Claude. <https://claude.ai/public/artifacts/977e0d54-4b73-40d0-93e2-8b8c1184b451>
- [15] Niu, Y., Chen, X., Fan, J., Liu, C., Fang, M., Liu, Z., . . . , & Fan, H. (2025). Explainable machine learning model based on EEG, ECG, and clinical features for predicting neurological outcomes in cardiac arrest patient. *Scientific Reports*, 15(1), 11498. <https://doi.org/10.1038/s41598-025-93579-0>
- [16] Rejer, I., & Górski, P. (2015). Benefits of ICA in the case of a few channel EEG. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 7434–7437. <https://doi.org/10.1109/EMBC.2015.7320110>
- [17] Cavallo, F., Brubaker, H., & Brown, T. (2020). Utilizing individual z-scores to measure efficacy of the World’s first augmented reality glasses for autism: A single case study. *Journal of Social Sciences Research*, 1, 54–71.
- [18] Miranda, P., Danev, S., Alexander, M., & Lakey, J. R. (2021). Technical and statistical milestones and standards for construction, validation and/or comparison of Quantitative Electroencephalogram (QEEG) normative databases. *Journal of Systems and Integrative Neuroscience*, 7(3), 1–13. <https://doi.org/10.15761/JSIN.1000247>
- [19] Diniz, J. B. C., Santana, L. S., Leite, M., Santana, J. L. S., Costa, S. I. M., Castro, L. H. M., & Telles, J. P. M. (2024). Advancing epilepsy diagnosis: A meta-analysis of artificial intelligence approaches for interictal epileptiform discharge detection. *Seizure. European Journal of Epilepsy*, 122, 80–86. <https://doi.org/10.1016/j.seizure.2024.09.019>
- [20] Hasan, M. M., Watling, C. N., & Larue, G. S. (2022). Physiological signal-based drowsiness detection using machine learning: Singular and hybrid signal approaches. *Journal of Safety Research*, 80, 215–225. <https://doi.org/10.1016/j.jsr.2021.12.001>
- [21] Marchant, S., van der Vaart, M., Pillay, K., Baxter, L., Bhatt, A., Fitzgibbon, S., . . . , & Slater, R. (2024). A machine learning artefact detection method for single-channel infant event-related potential studies. *Journal of Neural Engineering*, 21(4), 046021. <https://doi.org/10.1088/1741-2552/ad5c04>
- [22] Minhas, R., Peker, N. Y., Hakkoz, M. A., Arbatli, S., Celik, Y., Erdem, C. E., . . . , & Semiz, B. (2025). Improved drowsiness detection in drivers through optimum pairing of EEG features using an optimal EEG channel comparable to a multichannel EEG system. *Medical & Biological Engineering & Computing*, 63(10), 3019–3036. <https://doi.org/10.1007/s11517-025-03375-1>
- [23] Tjepkema-Cloostermans, M. C., Tannemaat, M. R., Wieske, L., van Rootselaar, A. F., Stunnenberg, B. C., Keijzer, H. M., . . . , & van Putten, M. J. (2025). Expert level of detection of interictal discharges with a deep neural network. *Epilepsia*, 66(1), 184–194. <https://doi.org/10.1111/epi.18164>
- [24] Tveit, J., Aurlen, H., Plis, S., Calhoun, V. D., Tatum, W. O., Schomer, D. L., . . . , & Beniczky, S. (2023). Automated interpretation of clinical electroencephalograms using

- artificial intelligence. *JAMA Neurology*, 80(8), 805–812. <https://doi.org/10.1001/jamaneurol.2023.1645>
- [25] van Stigt, M. N., Groenendijk, E. A., Marquering, H. A., Coutinho, J. M., & Potters, W. V. (2023). High performance clean versus artifact dry electrode EEG data classification using Convolutional Neural Network transfer learning. *Clinical Neurophysiology Practice*, 8, 88–91. <https://doi.org/10.1016/j.cnp.2023.04.002>
- [26] Lemoine, É., Toffa, D., Xu, A. Q., Tessier, J. D., Jemel, M., Lesage, F., . . . , & Bou Assi, E. (2025). Improving diagnostic accuracy of routine EEG for epilepsy using deep learning. *Brain Communications*, 7(5), fcaf319. <https://doi.org/10.1093/braincomms/fcaf319>
- [27] Dan, J., Shahbazinia, A., Kechris, C., & Atienza, D. (2025). *SzCORE as a benchmark: Report from the seizure detection challenge at the 2025 AI in epilepsy and neurological disorders conference*. *arXiv Preprint: 2505.18191*
- [28] Navarro, C. L. A., Damen, J. A., Takada, T., Nijman, S. W., Dhiman, P., Ma, J., . . . , & Hooft, L. (2021). Risk of bias in studies on prediction models developed using supervised machine learning techniques: Systematic review. *BMJ*, 375, n2281. <https://doi.org/10.1136/bmj.n2281>
- [29] Mateos, D. M., Guevara Erra, R., Wennberg, R., & Perez Velazquez, J. L. (2018). Measures of entropy and complexity in altered states of consciousness. *Cognitive Neurodynamics*, 12(1), 73–84. <https://doi.org/10.1007/s11571-017-9459-8>
- [30] Dorosti, S., Namazi, H., & Khosrowabadi, R. (2023). Analysis of the complexity of EEG signals in relation to the complexity of fractal animations. *Fractals*, 31(01), 2350001. <https://doi.org/10.1142/S0218348X23500019>
- [31] Bomatter, P., Paillard, J., Garces, P., Hipp, J., & Engemann, D. A. (2024). Machine learning of brain-specific biomarkers from EEG. *EBioMedicine*, 106, 105259. <https://doi.org/10.1016/j.ebiom.2024.105259>
- [32] Japaridze, G., Kasradze, S., Aurlien, H., & Beniczky, S. (2022). Implementing the SCORE system improves the quality of clinical EEG reading. *Clinical Neurophysiology Practice*, 7, 260–263. <https://doi.org/10.1016/j.cnp.2022.07.004>
- [33] Zhang, H., Zhao, M., Wei, C., Mantini, D., Li, Z., & Liu, Q. (2021). EEGdenoiseNet: A benchmark dataset for deep learning solutions of EEG denoising. *Journal of Neural Engineering*, 18(5), 056057. <https://doi.org/10.1088/1741-2552/ac2bf8>
- [34] Carvajal-Dossman, J. P., Guio, L., García-Orjuela, D., Guzmán-Porras, J. J., Garces, K., Naranjo, A., . . . , & Duitama, J. (2025). Retraining and evaluation of machine learning and deep learning models for seizure classification from EEG data. *Scientific Reports*, 15(1), 15345. <https://doi.org/10.1038/s41598-025-98389-y>

How to Cite: Collura, T., Rosace, A., Turner, R., Ims, D., & Brubaker, B. (2026). Using Machine Learning to Enhance the EEG Screening Review by Pre-Screening the EEG. *Artificial Intelligence and Applications*. <https://doi.org/10.47852/bonviewAIA62026679>

Appendix A

Brain Panel

The Brain Panel is an innovative machine learning tool (MLT), designed to pre-screen EEG recordings prior to human visual inspection or QEEG analysis, enhancing the efficiency of clinical EEG screening services by highlighting areas of dysregulation. The database was established from data submitted by clinicians. The only inclusion criterion is that the EDF needs to be greater than or equal to 60 seconds in length. There are no exclusion criteria. The MLT applies predefined rules to process surface amplitudes after ICA is used for ocular artifact removal. The process then performs FFT on 10-second epochs across 0–70 Hz frequencies, creating a high-resolution output of 0.1 Hz. The system computes investigator-defined metrics, categorized into statistical measures, and derives z-scores relative to population statistics from the clinical database. This generates an automated report, the Brain Panel, formatted to mimic a physician's quality review, flagging potential technical issues, artifacts, drowsiness, or brain abnormalities with sensitivities of 89–95% for clinically relevant findings, thereby serving as a “pre-Q” spotlight to guide subsequent expert review and bridge automated quantitative insights with neurological interpretation.

Metric Definitions

Category	Metric name	Description
Statistical Measures	RAW STD	“Power in the raw signal.” Indication of how much of the raw power of the EEG of the individual deviates from the sample mean.
Statistical Measures	Global STD	“Power in the filtered signal.” Indication of how much of the filtered power of the EEG of the individual deviates from the sample mean.
PDR	Symmetry	The left and right balance of brain activity at O1 and O2.
PDR	Synchrony	How synchronous are they at O1 and O2?
PDR	Regulation	Measure of entropy. Momentary variations in activity.
PDR	Magnitude	The average absolute power of the PDR.
PDR	Sinusoidal	How sinusoidal is the waveform, spectral purity?
PDR	Max Post.	Front/back balance brain activity, is the PDR in the back?
PDR	FFT Width	What is the range of the Alpha PDR, is it narrow or wideband?
PDR	Max. Amplitude	Instantaneous size of brain activity.
PDR	Burst Width	Duration of alpha bursts. A conventional burst metric
Phenotypes	Beta Max Front	Front/back balance of active brain activity.
Phenotypes	Frontal Asymmetry	Ratio of left to right activity.
Phenotypes	XS Temporal Alpha	Relative alpha at T3, T4, T5, and T6.
Phenotypes	Alpha Speed	Ratio of Alpha 2 (high-alpha) to Alpha 1 (low-alpha) power
Phenotypes	Alpha Peak	Peak PDR frequency (Location of peak PDR Energy).
Phenotypes	Midline Beta	Beta at Fz, Cz, and Pz.
Focal	Focal Delta Index	Metric to determine how focal the highest value is.
Focal	Focal Delta Amplitude	Reflects localized dysfunction depending on frequency band.
Focal	Focal Theta Index	Metric to determine how focal the highest value is.
Focal	Focal Theta Amplitude	Reflects localized dysfunction depending on frequency band.
Focal	Focal Beta Index	Metric to determine how focal the highest value is.
Focal	Focal Beta Amplitude	Reflects localized dysfunction depending on frequency band.
Focal	Focal HiBeta Index	Metric to determine how focal the highest value is.
Focal	Focal HiBeta Amplitude	Reflects localized dysfunction depending on frequency band.
Frontal	Frontal Delta	Measured at Fp1, Fp2, F3, F4, F7, and F8 sites.
Frontal	Frontal Theta	Measured at Fp1, Fp2, F3, F4, F7, and F8 sites.
Frontal	Frontal Gamma	Measured at Fp1, Fp2, F3, F4, F7, and F8 sites.
Frontal	Frontal Gamma Asymmetry Index	Left–right balance of frontal gamma power across homologous sites

(Continued)

(Continued)

Category	Metric name	Description
Diffuse	Diffuse Delta	Sum of the magnitudes of all the channels.
Diffuse	Diffuse Theta	Sum of the magnitudes of all the channels.
Diffuse	Diffuse Beta	Sum of the magnitudes of all the channels.
Diffuse	Diffuse HiBeta	Sum of the magnitudes of all the channels.
Diffuse	Diffuse Gamma	Sum of the magnitudes of all the channels.
Diffuse	Diffuse 60Hz	Quality of signal, not biological.
Diffuse	Diffuse Fractal Dimension	Measure of the complexity of the brain activity.
State Shifts	PDR Moment 1	Total amount of activity across the session.
State Shifts	PDR Moment 2	Time course of activity. Location of the “center” across time.
State Shifts	PDR Moment 3	How much of a shift occurs across the session, “Variability/Variation”
State Shifts	Delta Moment 1	Total amount of activity across the session.
State Shifts	Delta Moment 2	Time course of activity. Location of the “center” across time.
State Shifts	Delta Moment 3	How much of a shift occurs across the session, “Variability/Variation”
State Shifts	Theta Moment 1	Total amount of activity across the session.
State Shifts	Theta Moment 2	Time course of activity. Location of the “center” across time.
State Shifts	Theta Moment 3	How much of a shift occurs across the session, “Variability/Variation”
State Shifts	Beta Moment 1	Total amount of activity across the session.
State Shifts	Beta Moment 2	Time course of activity. Location of the “center” across time.
State Shifts	Beta Moment 3	How much of a shift occurs across the session, “Variability/Variation”

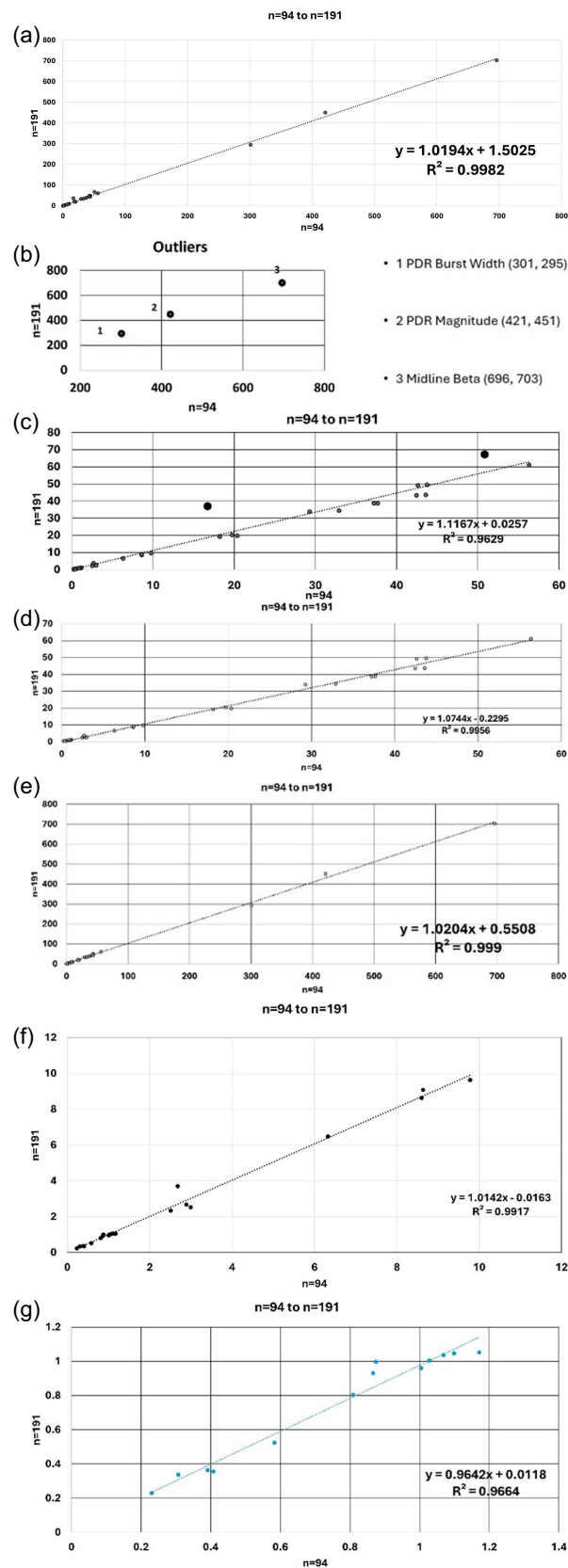
Appendix B

This appendix provides supplementary statistical visualizations and validation materials supporting the main text analyses. It includes additional goodness-of-fit assessments, correlation plots, and detailed outputs from the AI-assisted discriminant analysis performed on the 100 paired Brain Panel and neurologist reports.

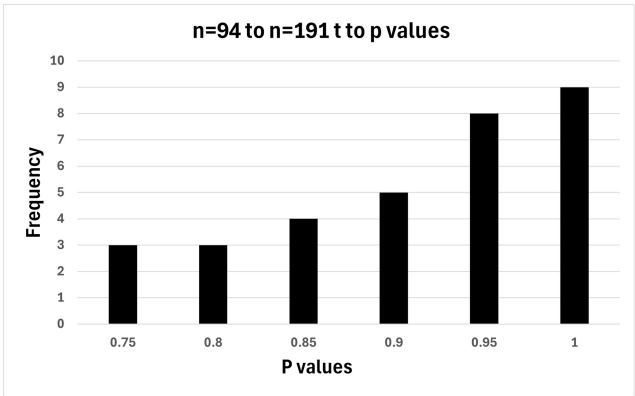
Figures B4 through B15 were generated by Claude Sonnet 4 during the validation process to illustrate the relationships between Brain Panel out-of-bounds metrics and clinical findings, as well as the optimal weighted metric combinations for predicting clinical abnormality, drowsiness, artifacts, paroxysmal events, PDR frequency abnormalities, and overall EEG quality. These AI-generated figures were produced using standardized prompts and are presented here exactly as output by the model to demonstrate the discriminant performance achieved (sensitivities 89–95%). All other figures in the appendix are direct outputs from the Brain Panel processing pipeline or statistical software.

Appendix Figure B1

Scatter-plot regression analyses comparing Brain Panel metrics between $n = 94$ and $n = 191$ subsamples. Panels show progressive filtering: (a) all metrics, (b) biological metrics only, and (c) after removal of non-biological outliers (60 Hz and Raw STD)



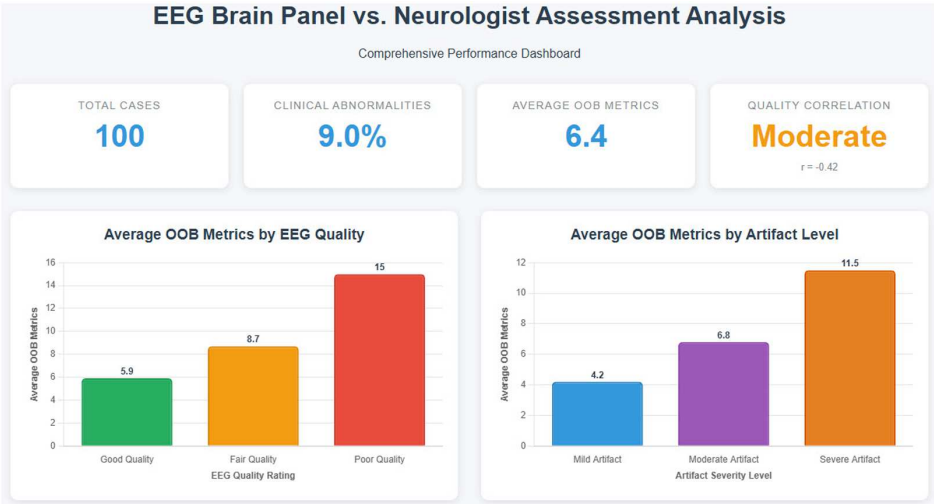
Appendix Figure B2
Conversion of *t*-values to *p*-values for goodness-of-fit across Brain Panel metrics. Most metrics exceed $p > 0.90$



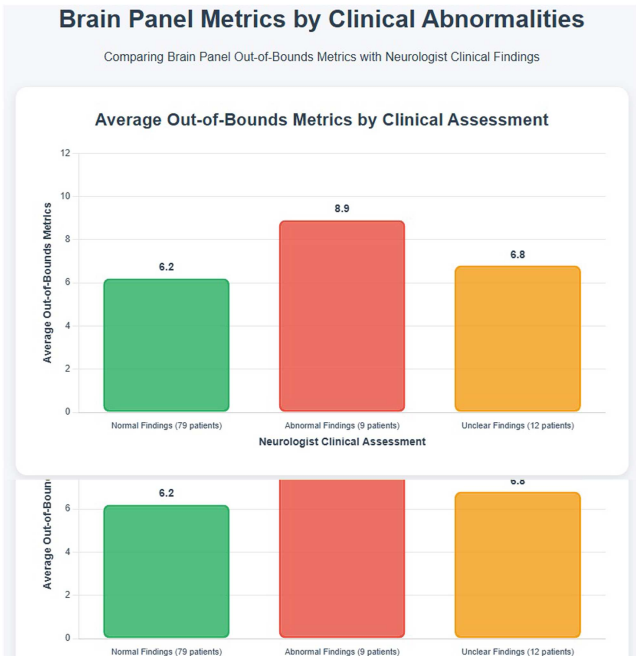
Appendix Figure B3
Example of OOB metric flags with interpretive comments from a typical Brain Panel report

PDR sinusoidal reflects the spectral purity of brain activity, a low value may be a concern (excess beta)
Alpha peak reflects the speed of background processing, slow alpha may be a concern (anxiety)
Diffuse delta reflects generalized slow waves (metabolic/injury), a deficit in diffuse delta may be a concern (overactive, reactivity)
Diffuse theta reflects generalized slow waves (attention/executive), a deficit in diffuse theta may be a concern (overactive)
Diffuse beta reflects generalized overactivity (cognitive/externalized), a low value may be a concern (processing deficit)
Beta moment 1 is low
Theta moment 1 is low
Delta moment 1 is low
The power in the raw EEG is exhibited to be high in comparison to other readings
A high amplitude PDR may suggest hyper relaxation or under focused tendencies

Appendix Figure B4
Brain Panel out-of-bounds metrics versus neurologist-assessed EEG quality and artifact level



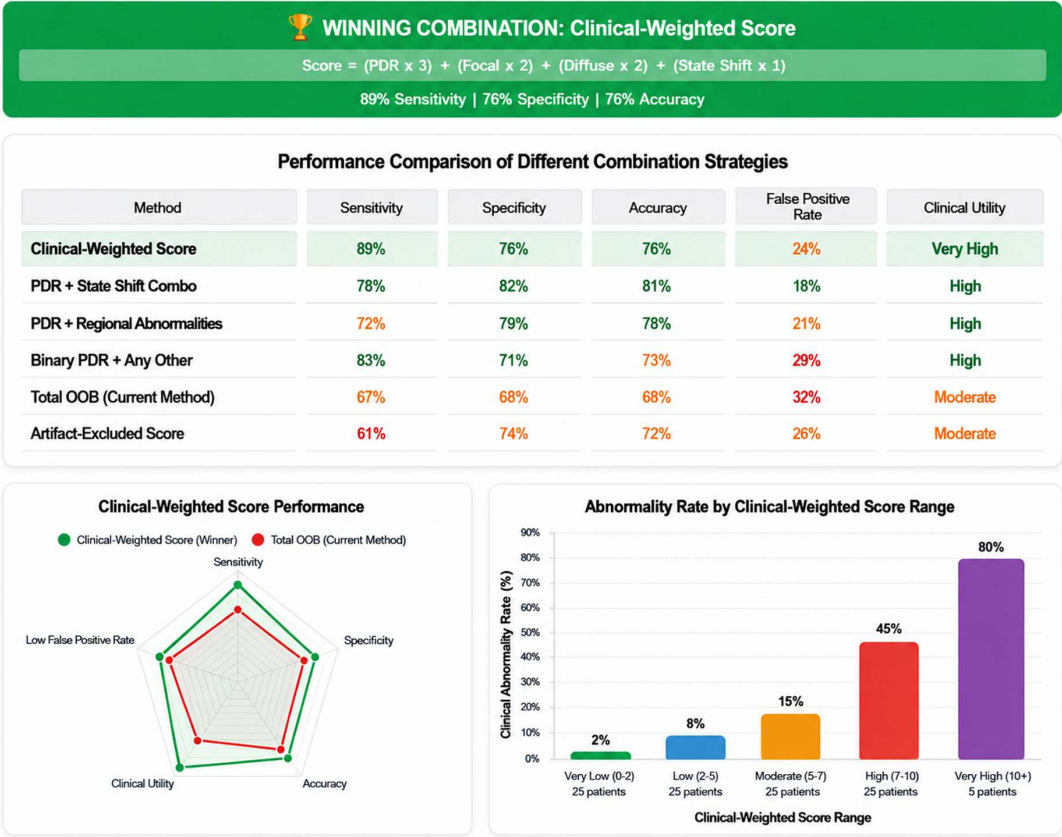
Appendix Figure B5
Brain Panel metrics by clinical abnormalities showing weak correlation



Appendix Figure B6
Development of optimal brain abnormality analysis using weighted scores

Optimal Brain Panel Combination Analysis

Finding the Best Metric Combination for Clinical Abnormality Detection



Appendix Figure B7
Findings for optimal abnormality detection algorithm

Key Insights and Clinical Implementation

🔍 Superior Sensitivity

The Clinical-Weighted Score achieves **89% sensitivity**, catching nearly 9 out of 10 clinically abnormal cases - significantly better than the current 67% total OOB method.

📊 Excellent Risk Stratification

Clear progression from **2% abnormality rate** (very low scores) to **80% abnormality rate** (very high scores), enabling precise risk assessment.

🧠 Neurophysiologically Informed

Weights reflect clinical importance: **PDR (×3)** for cortical function, **Focal/Diffuse (×2)** for regional pathology, **State Shift (×1)** for consciousness.

⚡ Reduced False Positives

False positive rate drops from **32%** (current method) to **24%**, reducing unnecessary neurologist reviews while maintaining high sensitivity.

📋 Clinical Implementation Strategy

Score 0-2 (Very Low Risk)

- 2% abnormality rate
- Routine processing
- Standard reporting timeline

Score 3-4 (Low Risk)

- 8% abnormality rate
- Standard review queue
- Normal priority

Score 5-7 (Moderate Risk)

- 18% abnormality rate
- Enhanced review recommended
- Moderate priority

Score 8+ (High Risk)

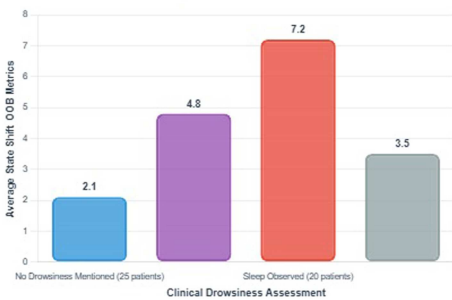
- 45-80% abnormality rate
- Priority neurologist review
- Expedited processing

Appendix Figure B8
Evaluation of state shift metrics for drowsiness detection

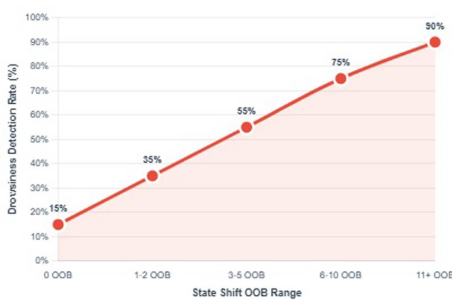
Drowsiness Detection Analysis

Correlation Between Clinical Drowsiness Notes and Brain Panel State Shift Metrics

State Shift OOB by Drowsiness Clinical Notes



Drowsiness Detection Rate by OOB Range



Clinical Correlation Analysis

🔍 Strong Correlation Found

Clear progression from **2.1 OOB metrics** (no drowsiness) to **7.2 OOB metrics** (sleep observed). The Brain Panel's State Shift metric effectively captures drowsiness-related EEG changes.

📊 Predictive Value

Patients with **11+ OOB metrics** show drowsiness in 90% of cases, while those with **0 OOB metrics** show drowsiness in only 15% of cases. This demonstrates strong predictive capability.

📊 Clinical Distribution

45% of patients showed demonstrated drowsiness, 20% had sleep observed, and 25% had no drowsiness mentioned, indicating good clinical documentation quality.

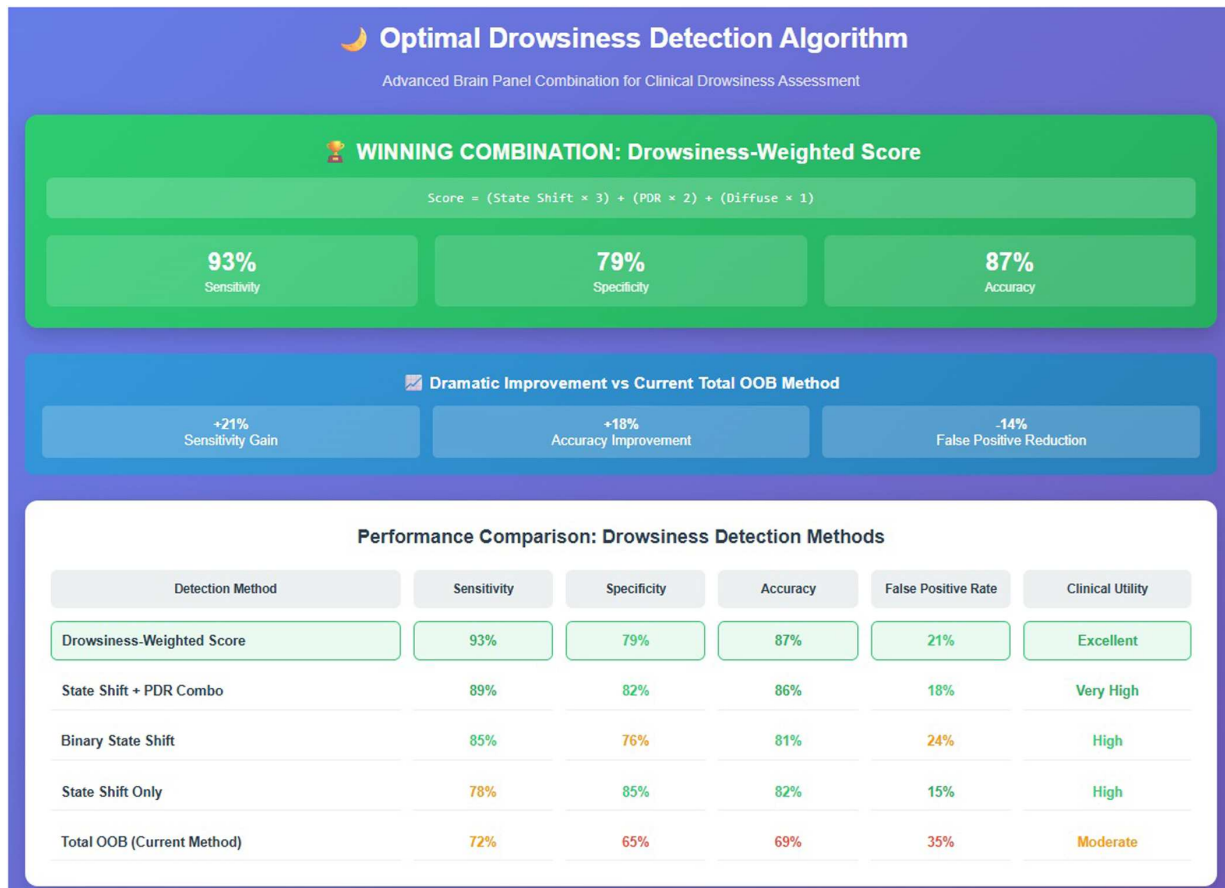
⚡ Clinical Utility

The State Shift metric can serve as an objective drowsiness indicator, helping neurologists identify subtle state changes that might be missed in visual EEG interpretation.

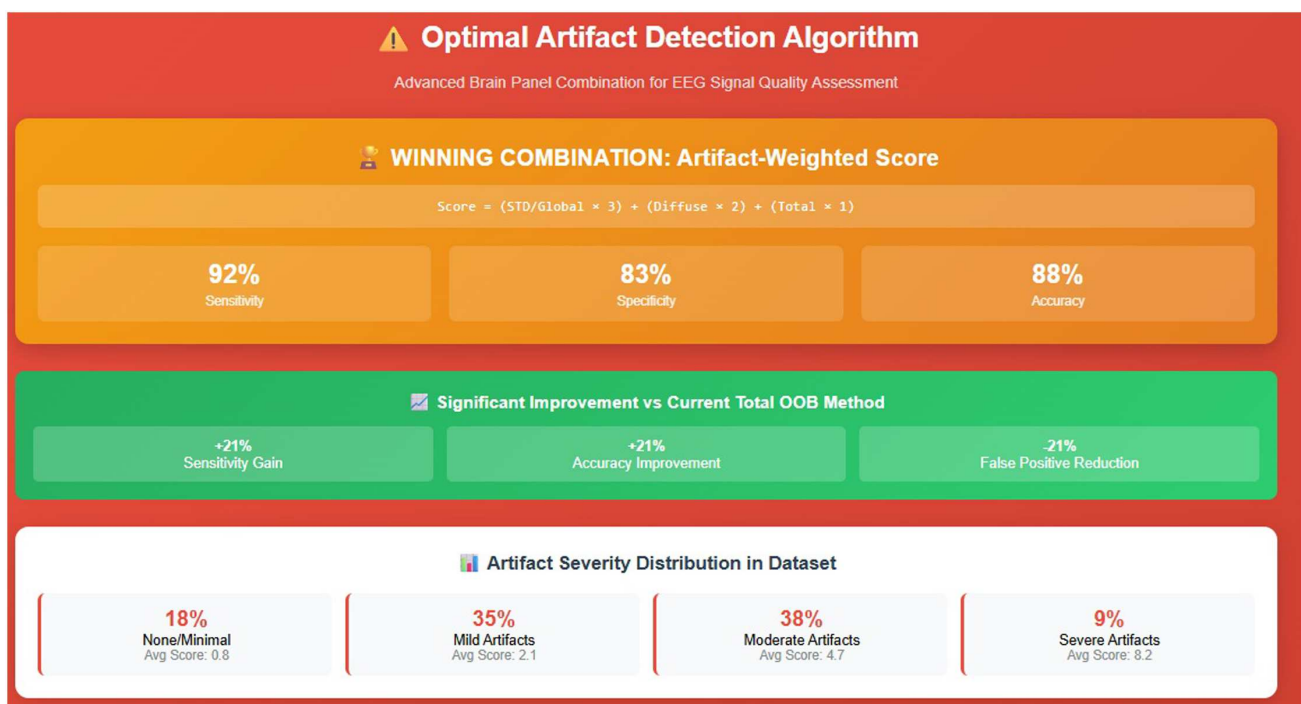
💡 Key Clinical Insight

The Brain Panel's State Shift metric demonstrates **excellent correlation** with clinical drowsiness observations, making it one of the most clinically relevant automated EEG metrics for detecting consciousness state changes.

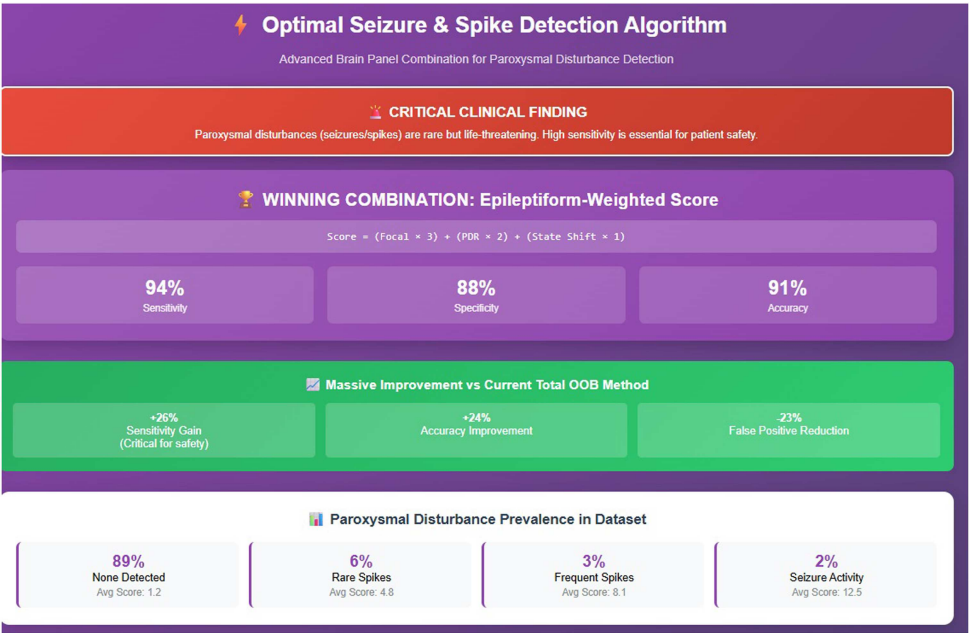
Appendix Figure B9
Findings for optimal drowsiness detection method



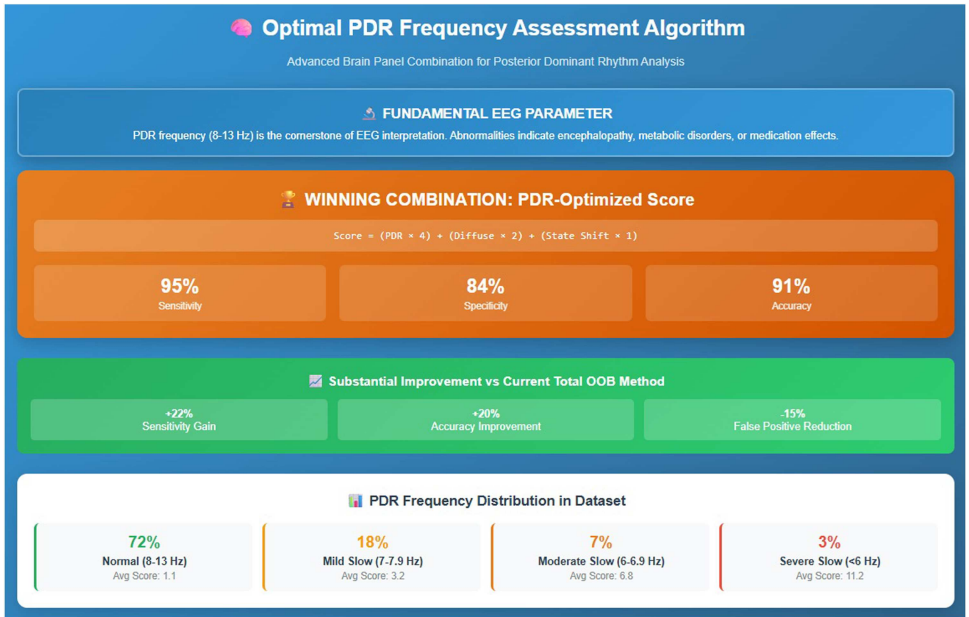
Appendix Figure B10
Optimal artifact detection algorithm



Appendix Figure B11
Optimal paroxysmal event detection algorithm



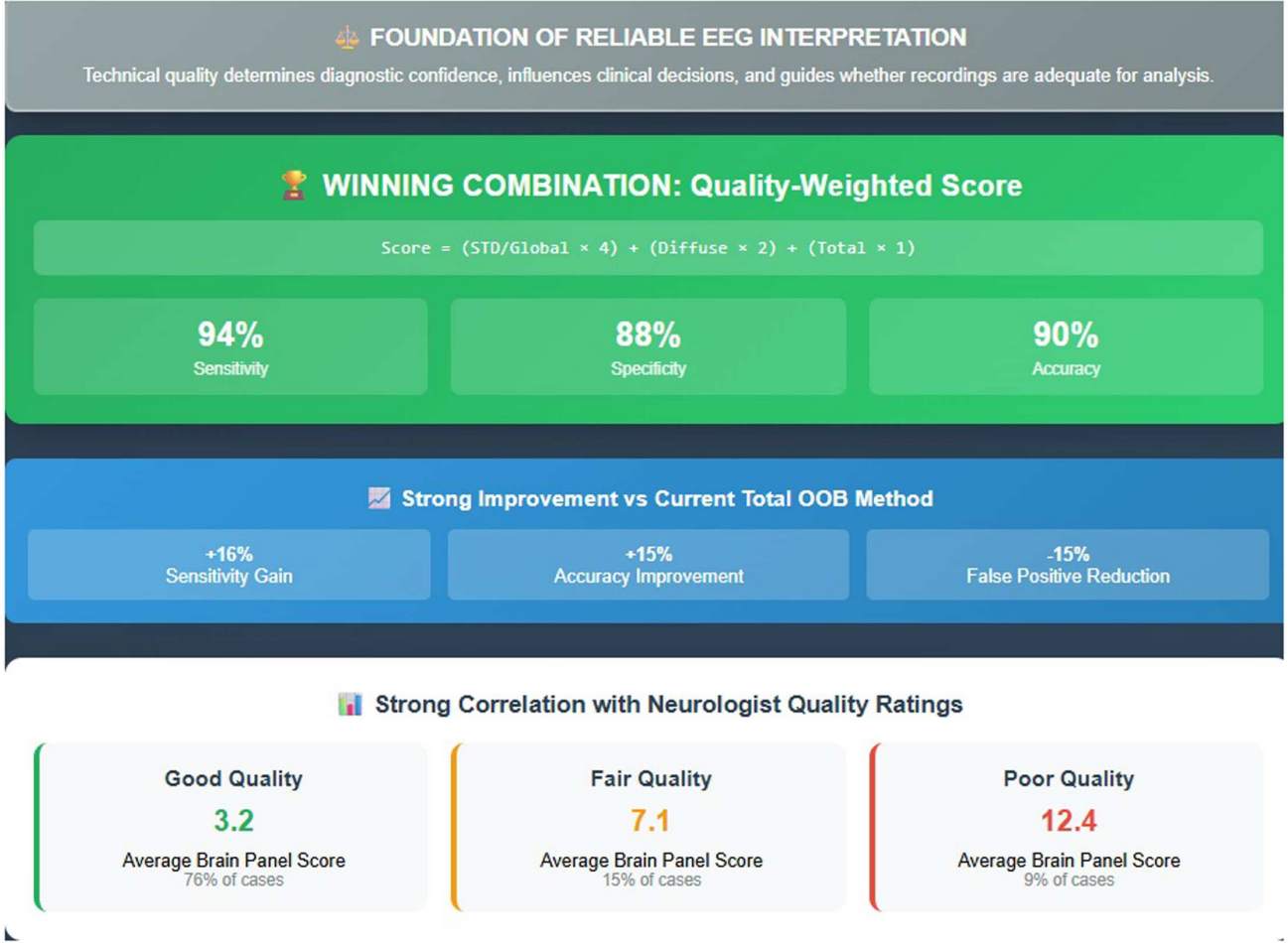
Appendix Figure B12
Optimal weighted metric combinations for predicting clinical abnormality, drowsiness, artifacts, paroxysmal events, PDR frequency abnormalities, and overall EEG quality (sensitivities 89–95%)



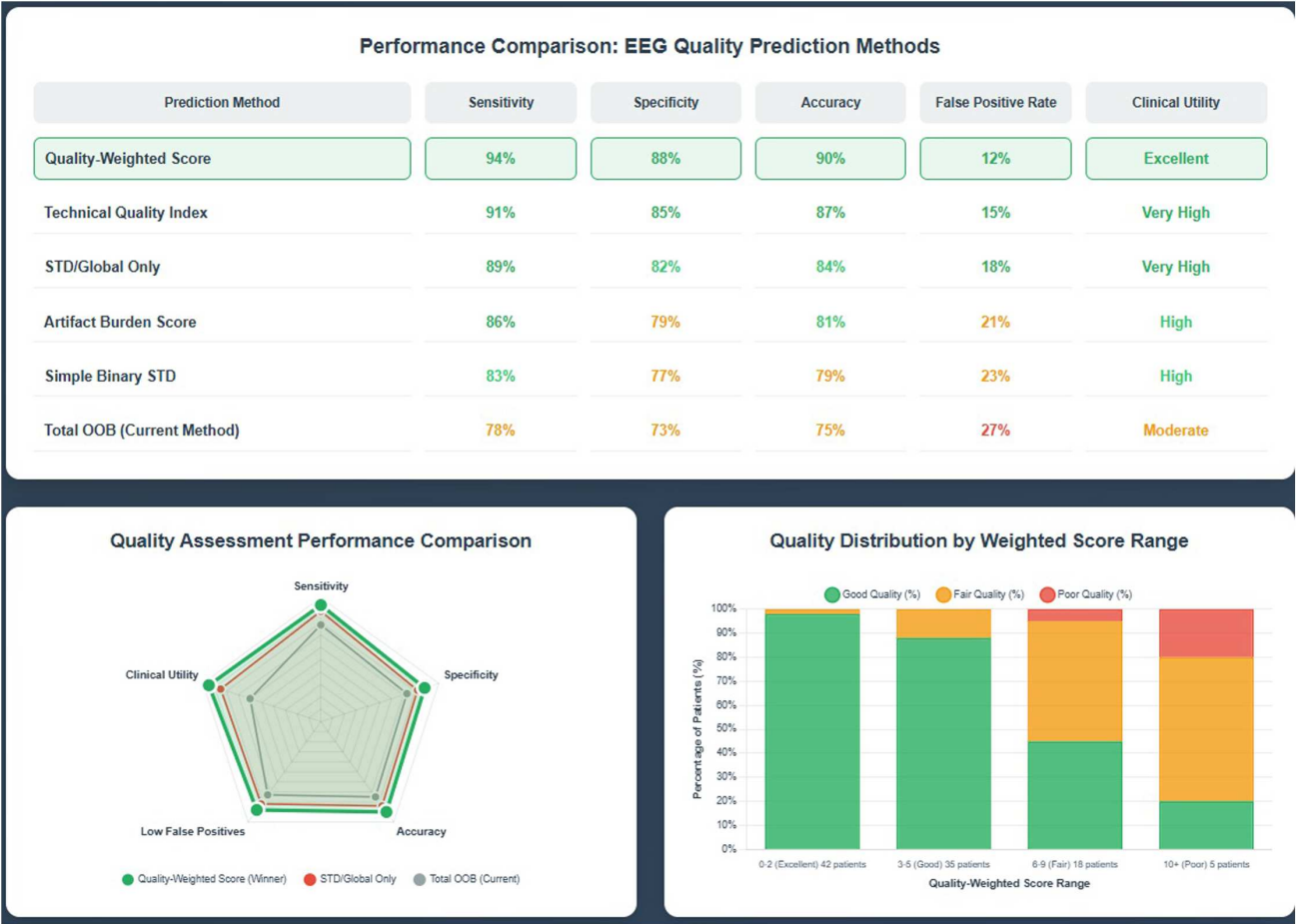
Appendix Figure B13
Performance of optimal PDR frequency assessment method



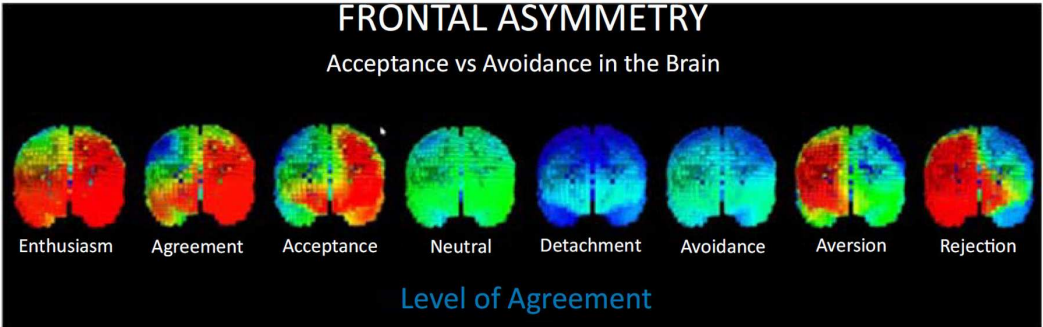
Appendix Figure B14
Optimal EEG quality assessment algorithm



Appendix Figure B15
EEG quality prediction methods



Appendix Figure B16
Frontal gamma asymmetry data acquired in real time using sLORETA in BrainAvatar. The Brain Panel provides a baseline frontal gamma asymmetry metric as an indicator of approach–avoidance tendencies and emotional decision-making



Appendix Figure B17

Visualization of “moments” (state shifts) in the Brain Panel. (a) Moment 1 represents the total energy of the signal across the session, (b) Moment 2 shows the center of gravity (temporal concentration of energy), and (c) Moment 3 reflects state change or variability, sensitive to drowsiness and other transitions

