Brain Research Bulletin 231 (2025) 111546

Contents lists available at ScienceDirect

Brain Research Bulletin

journal homepage: www.elsevier.com/locate/brainresbull

Research report

A normative model–based assessment framework for large-scale, multi-site EEG data

Qiwei Dong [a,c,d], Yuxi Zhou [a,b], Xiaoyu Xiong [a,b], Pengyu Liu [a,b], Jianfu Li [a,b], Cheng Luo [a,b,c,d], Diankun Gong [a,b,c], Li Dong [a,b,c,*], Dezhong Yao [a,b,c,d,*]

[a] The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China
[b] School of Life Science and Technology, Center for information in medicine, University of Electronic Science and Technology of China, Chengdu, China
[c] Sichuan Institute for Brain Science and Brain-Inspired Intelligence, Chengdu, China
[d] Research Unit of NeuroInformation, Chinese Academy of Medical Sciences, Chengdu 2019RU035, China

ARTICLE INFO

Keywords:
EEG
Attention assessment
Normative modeling
Multi-site data
Regression models

ABSTRACT

Background: Electroencephalography (EEG) overcomes the subjectivity inherent in questionnaire-based and observational assessments. However, most existing EEG-based evaluation methods still impose discrete categorical states onto continuously varying neural dynamics, thereby neglecting the continuity of states. With the rise of neuroscience alliances, challenges such as batch-effects across datasets and inconsistencies introduced by diverse EEG electrode montages have become increasingly prominent. Therefore, a robust assessment framework that accommodates large-scale, multi-site EEG data is expected.

Methods: A normative model-based assessment framework was developed for large-scale, multi-site EEG data, with attention assessments used as illustrative examples. Normative models are first constructed using EEG features from 1212 young individuals, and quantile ranks are computed. Next, feature selection is performed, and elastic net regression and support vector regression are used to model distributed attention (DA) and focused attention (FA). The results from normative model-based features are compared with original features to demonstrate the advantage of quantile rank features. Finally, the model's test-retest reliability and generalizability are assessed.

Results: The framework identifies statistical differences ($q < 0.05$) in attention performance between the top and bottom 20 % participants on attention scales. EEG features demonstrated specific patterns related to accuracy and reaction time in both DA and FA tasks. The normative model outperformed in predictive tasks, showing enhanced stability and interpretability. Additionally, the framework demonstrates strong test-retest reliability and robust generalizability (ICC > 0.9).

Conclusion: In conclusion, we proposed a normative model–based framework that harmonizes large-scale, multi-site EEG data, enabling efficient and reliable attention assessment while demonstrating promise for broader EEG-based applications.

1. Introduction

Electroencephalography (EEG)-based assessment methods have been extensively applied in various fields such as cognition and emotion with the continuous advancement of neuroscientific techniques (Li et al., 2022; Wan et al., 2021; Kant et al., 2022). Traditional subjective assessment tools, which rely heavily on questionnaires and observational records, often exhibit limitations including insufficient quantitative accuracy and vulnerability to individual state variations (Price et al., 2003; Van Voorhees et al., 2010). In contrast, as a non-invasive, portable, and cost-effective neuroimaging technique with high temporal resolution, EEG provides objective and refined physiological signal

* Corresponding authors at: The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China.
E-mail addresses: Lidong@uestc.edu.cn (L. Dong), dyao@uestc.edu.cn (D. Yao).

https://doi.org/10.1016/j.brainresbull.2025.111546
Received 24 May 2025; Received in revised form 1 September 2025; Accepted 9 September 2025
Available online 12 September 2025

data to effectively overcome uncertainties introduced by subjective feedback methods (Yao, 2024).

Current EEG-based assessment approaches commonly segment continuously varying neural activities into predefined fixed categories or states. For instance, dividing attention into different levels (Wan et al., 2021), evaluating cognitive load based on predetermined tasks (Zanetti et al., 2022), or defining emotion as discrete states (Li et al., 2022). Although such discretization strategies simplify data processing, their limitations are evident: fixed cutoff values frequently fail to adequately capture inter-individual physiological differences and nonlinear characteristics, potentially neglecting subtle changes of states (Hany et al., 2024; Chervyakov et al., 2016; Rosenberg et al., 2013). This consequently limits the accuracy and sensitivity of assessment outcomes. In comparison, regression methods are capable of comprehensively capturing data features, yet they face significant challenges in application, including the nonlinear age dependency of raw EEG features, individual heterogeneity obscuring biologically meaningful deviations, and model sensitivity to extreme values (Smit et al., 2016; Verdonck et al., 2024).

The emergence of neuroscience alliances has ushered in a new era of large-scale multisite cohort studies, exemplified by international collaborative initiatives such as the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) consortium spanning 43 countries/regions (Thompson et al., 2020), the Adolescent Brain Cognitive Development (ABCD) study encompassing over 10,000 participants across 21 research sites (Casey et al., 2018), and major open-access databases like the Temple University Hospital EEG (TUH-EEG) corpus containing more than 30,000 clinical electroencephalography recordings (Obeid and Picone, 2016). These efforts have significantly advanced neuroimaging research through standardized protocols. However, they remain challenged by batch-effects, particularly differences in EEG recording device configurations and electrode montages—a key confounding factor that not only affects multi-site studies but also poses a formidable barrier to individual researchers attempting cross-database integration (Markiewicz et al., 2021). A more effective framework for mining these big EEG data is expected for assessing outcomes in applications.

Normative modeling is a statistical framework that enables statistical inferences at the individual level concerning expected patterns (Rutherford et al., 2022). By mapping quantitative measurements to relevant variables, normative models can identify variations within large-scale cohorts, providing a method for quantifying and describing deviations between individuals and expected patterns, as well as among individuals themselves. Brain charts developed from normative models serve as essential tools for analyzing brain development and aging processes, with quantile rank mapping being a key analytical approach (Zhang et al., 2022). Currently, quantile rank features derived from normative models have been widely applied to MRI data (Habes et al., 2021; Frangou et al., 2022). Although EEG-based normative modeling studies are relatively limited, existing research suggests that quantile rank features in EEG data hold promising application potential (Taylor et al., 2022; Janiukstyte et al., 2023). Furthermore, despite the availability of standardized EEG processing pipelines (e.g., EEGLAB (Delorme and Makeig, 2004) and WeBrain (Dong et al., 2021a)), unified electrode montages and batch-effect correction methods, systematic integration of these preprocessing techniques with normative modeling assessments remains limited. This lack of integration may contribute to inconsistent outcomes across different laboratories and complicate the integration of multi-center datasets. Therefore, developing a unified framework that integrates EEG preprocessing techniques with normative modeling assessments is critically needed.

In this study, a normative model-based assessment framework was developed for large-scale, multi-site EEG data, with distributed attention (DA) and focused attention (FA) serving as illustrative examples. To our knowledge, this is the first large-scale, multi-site assessment framework based on EEG normative modeling in the healthy population. The framework consists of several key steps. First, EEG electrode montages were unified, and batch effects across multiple sites were corrected. Next, normative models were constructed using resting-state EEG brain network metrics from 1212 healthy young adults, and quantile ranks of the EEG dataset under assessment were evaluated. Finally, the effectiveness and robustness of the framework in predicting DA and FA were tested through feature selection and regression models. Overall, this framework provides a novel approach for EEG-based assessment, facilitating efficient and reliable attention assessments in large-scale, multi-site EEG studies.

2. Materials and methods

2.1. Participants

In this study, resting-state eyes-closed EEG data were collected from 1212 and 189 right-handed healthy young adults to construct the normative model database (NMD, Table 1) and the mapping relationship database (MRD, Table 2), respectively. Databases 1–8 of NMD and the MRD were sourced from the University of Electronic Science and Technology of China (UESTC), while Databases 9–16 of NMD were obtained from public datasets. Prior to the collection of MRD, participants completed DA tests to measure reaction time (DA-RT) and accuracy (DA-AC), as well as FA tests to measure reaction time (FA-RT) and accuracy counts (FA-AC). The experimental paradigm is illustrated in Supplementary Figure S1, and more detailed information can be found at http://www.dweipsy.com:8082/lattice/latticeUI/index.html#/home. The study was approved by the Ethics Committee of the Department of Life Science and Technology at the UESTC. All participants provided written informed consent prior to participation.

2.2. EEG acquisition and preprocessing

The EEG acquisition parameters and channel location diagrams are provided in Supplementary Table S1 and Supplementary Figure S2, respectively. All data were quality-assessed and preprocessed using the preprocessing pipeline of the WeBrain platform (https://webrain.uestc.

Table 1
Demographic Information of NMD.

Dataset	Sample	Age Range	Mean Age	Age SD	Source
Database 1	66	17 – 20	18.71	0.65	UESTC
Database 2	39	21 – 28	23.90	1.64	UESTC
Database 3	86	17 – 26	20.95	2.29	UESTC
Database 4	33	18 – 24	20.30	1.72	UESTC
Database 5	98	17 – 20	18.72	0.67	UESTC
Database 6	245	18 – 24	19.61	0.99	UESTC
Database 7	146	18 – 27	22.46	1.83	UESTC
Database 8	203	18 – 27	21.30	2.26	UESTC
Database 9	30	18 – 22	19.57	1.14	Alexander et al., 2017
Database 10	52	18 – 28	19.92	1.75	Duan et al., 2021
Database 11	62	18 – 25	19.79	1.48	Pavlov et al., 2022
Database 12	60	18 – 28	20.02	1.88	Wang et al., 2022
Database 13	37	18 – 23	20.38	1.46	Xiang et al., 2024
Database 14	19	18 – 28	20.63	2.61	Kasanov et al., 2024
Database 15	19	18 – 22	18.79	1.27	Cavanagh et al., 2019a
Database 16	17	18 – 30	23.00	3.77	Cavanagh et al., 2019b
Total	1212	17 – 30	20.54	2.12	-

NMD: Normative Model Database, UESTC: EEG data collected by the University of Electronic Science and Technology of China.

Table 2
Demographic Information of MRD.

Group	Sample	Age Range	Mean Age	Age SD	Source
MRD	189	18 – 22	19.59	0.97	UESTC

MRD: Mapping Relationship Database, UESTC: EEG data collected by the University of Electronic Science and Technology of China.

edu.cn/) (Dong et al., 2021a). The main procedures included: 1) A quality assessment (QA) method (Zhao et al., 2023) was used to detect bad channels with different types of artifacts (detecting constant or NaN/Inf signals, unusually high or low amplitudes, high or power frequency noises, and low correlation signals); 2) EEG signals were subjected to bandpass filtering in the range of 1–60 Hz, with the addition of 50 Hz or 60 Hz notch filters to eliminate power-line interference; 3) the ICA-based MARA algorithm was used to remove artifacts, with a conservative threshold of 0.7 (Auger et al., 2022); 4) Reference Electrode Standardization Interpolation Technique (RESIT) (Dong et al., 2021b) was applied to interpolate the bad channels and to transform the reference electrode to an idealized zero potential point at infinity using the reference electrode standardization technique (REST) (Dezhong, 2001); 5) The channel distributions were harmonized to the 32-channel electrode montage (Supplementary Figure S2(a)) by REST (Dong et al., 2024).

2.3. Assessment framework

The framework includes the following steps: 1) Functional connectivity networks were computed using clean EEG data; 2) To mitigate batch effects arising from multicenter data acquisition, the ComBat harmonization method was applied to the functional connectivity networks while effectively preserving inter-individual variability associated with age; 3) Normative models were established using functional connectivity features from the NMD. For each MRD participant, connectivity measures were transformed into quantile ranks within these models; 4) Regression models were constructed by replacing the original EEG-based functional connectivity features with their quantile rank values. These quantile features were then used for training and evaluation, enabling a normative-model-based regression analysis and prediction.

2.3.1. Phase Synchronization Index (PSI)

Functional connectivity networks were derived from the PSI after channel harmonization. PSI were calculated as follows (Jose et al., 2004):

$$PSI = \sqrt{< \cos(\Delta\varphi(t)) >^2 + < \sin(\Delta\varphi(t)) >^2}$$

where $\Delta\varphi(t)$ is the instantaneous phase difference between two EEG signals for a particular frequency, and $< \bullet >$ means the temporal average. Analyses were performed within the delta (1–4 Hz), theta (4–8 Hz), alpha (8–12.5 Hz), beta (12.5–25 Hz), high beta (25–30 Hz), gamma1 (30–40 Hz), gamma2 (40–60 Hz).

2.3.2. Batch effect removal

Batch-related differences introduce non-biological variations in multicenter studies. To remove these effects while retaining age-related effects, the ComBat method, a statistical method used to correct for batch effects in data, particularly in genomics and imaging studies, was used (Fortin et al., 2017):

$$y_{ijc} = \alpha_c + \boldsymbol{X}_{ij}\boldsymbol{\beta}_c + \gamma_{ic} + \delta_{ic}\varepsilon_{ijc}$$

where α_c is the overall PSI, $\boldsymbol{X}$ includes age covariates, γ_{ic}, δ_{ic}, ε_{ijc} represent additive and multiplicative batch effects on PSI, respectively. The ComBat employs least squares for initial parameter estimation and uses empirical Bayesian techniques to adjust for batch effects. To evaluate the effect of removing batch effects, 20 participants were randomly selected from 4 independent datasets to compare alpha-band PSI before and after removal, and a one-way ANOVA was subsequently performed to test the statistical differences among the groups.

2.3.3. Normative modeling

EEG normative models were developed using the Generalized Additive Models for Location, Scale and Shape (GAMLSS) package in RStudio. This regression method fits distributions with up to four parameters (location, scale, skewness, kurtosis), allowing each parameter to be modeled as a function of explanatory variables or nonparametric smoothing functions. The model is expressed as:

$$Y_i = D(\mu_i, \sigma_i, \nu_i, \tau_i)$$

$$g_1(\mu) = \eta_1 = X_1\beta_1 + \sum_j^{J_1} s_{1j}(x_{1j})$$

$$g_2(\sigma) = \eta_2 = X_2\beta_2 + \sum_j^{J_2} s_{2j}(x_{2j})$$

$$g_3(\nu) = \eta_3 = X_3\beta_3 + \sum_j^{J_3} s_{3j}(x_{3j})$$

$$g_4(\tau) = \eta_4 = X_4\beta_4 + \sum_j^{J_4} s_{4j}(x_{4j})$$

where Y_i denotes the PSI; μ, σ, ν, and τ represent the mean, standard deviation, skewness, and kurtosis of PSI, respectively; X_k are the design matrices; β_k are the coefficients; $g_k(.)$ are monotonic link functions; $s_{.j}$ are nonparametric smooth functions applied to X_{kj}. Model fitting was conducted using the Box–Cox t (BCT) distribution with cubic spline smoothing for all four parameters, and the final model parameters were selected according to the Akaike Information Criterion (AIC).

2.3.4. Feature selection and predictive models

The framework employed filter-based feature stability feature selection before constructing predictive models to enhance the model's robustness in applications. Specifically, within the MRD dataset, random subsets of 152 samples (matching the training size for five-fold cross-validation) were selected and resampled 50 times using the bootstrap method. For each resample, Spearman rank correlation coefficients between features and behavioral indicators were computed. Features surpassing a predefined correlation threshold and appearing in more than 90 % of the resamples were retained. This process was executed 50 times over 31 thresholds (defined as the mean absolute correlation ±1.5 standard deviations (SD), with a step size of 0.1), and the normalized Percentage of Overlapping Genes (nPOG) index was used to determine the optimal threshold (Gopakumar et al., 2015; Sen et al., 2021). In addition, potential multicollinearity was addressed by clustering highly correlated features and averaging the features within each cluster (Ge et al., 2019).

For predictive models, linear regression based on the elastic net regression (ENR) algorithm was primarily employed (Shen et al., 2021). Hyperparameters were optimized using the grid search method on the MRD. Specifically, the α value ranged from 0.1 to 1 in increments of 0.1, and the λ value ranged from 0.001 to 1 in increments of 0.001. Model performance was initially assessed using the mean squared error (MSE) from five-fold cross-validation. Subsequently, 20 iterations of five-fold cross-validation assessed MSE and correlation coefficient. Additionally, to evaluate the adaptability of nonlinear methods within the framework proposed in this study, support vector regression (SVR) with a radial basis function (RBF) kernel was tested using similar grid search tuning (Quitadamo et al., 2017). It is worth noting that, unless otherwise specified, the regression models used in this study are all based on the

ENR method.

2.4. Framework validation

The framework's validity was assessed by comparing the top 20 % and bottom 20 % performers on each scale using t-tests. Additionally, the evaluation method was applied to all participants in the NMD as a reference. A comprehensive analysis compared top and bottom 20 % performers across all four scales, and t-tests (based on Pearson correlation and MSE) were used to contrast results from quantile rank features and PSI features. To further assess generalizability, the framework was applied to various connectivity metrics, including the Phase Lag Index (PLI), Pearson correlation (COR), and Coherence (COH) for attention evaluation (Lee and Hsieh, 2014; Stam et al., 2007).

Test-retest reliability was examined using 50 randomly selected MRD participants, each with over 6 min of recording. Each recording was divided into five overlapping 2-minute segments (with a maximum 1-minute overlap). For each segment, the evaluation framework computed DA-AC, DA-RT, FA-AC, and FA-RT scores. Reliability was quantified by the intra-class correlation coefficient (ICC) and the correlation between segment scores and the original evaluation scores (Suárez-Revelo et al., 2015). Normative models were generated using RStudio (version 2024.04.1), and additional analyses were performed using MATLAB (version 2020b). Multiple comparisons were controlled using the Benjamini–Hochberg false discovery rate correction, with q values representing FDR-adjusted p values and significance defined at $q < 0.05$.

3. Results

3.1. Assessment framework

3.1.1. Batch effect removal

The Combat algorithm was applied to harmonize the PSI metric across datasets. Figs. 2(a) and 2(b) illustrate this process using the alpha band from four randomly selected NMD datasets. Fig. 2(a) (left) shows that PSI distributions among 20 participants varied significantly across datasets, with statistical differences evident (right). After harmonization (Fig. 2(b)), the PSI distributions became more consistent with no significant differences among datasets.

3.1.2. Normative models

Fig. 3 presents normative models for the C3-CP2 theta and C3-CP5 beta bands. In the C3-CP2 theta model, PSI values decrease with age, with a maximum value below 0.8. Conversely, the C3-CP5 beta model shows PSI values increasing with age, with a maximum exceeding 1.1. Additionally, regions with denser data points display narrower percentile intervals, whereas regions with sparser data exhibit wider intervals.

3.1.3. Feature selection

As shown in Fig. 4(a), when the correlation threshold falls within the range of the mean ± 1.5 SD, both the average feature stability and the number of selected features first increase and then decrease as the threshold changes. The nPOG index reached its peak when the threshold was set at the mean - 0.4 SD, after which it plateaued and began to decline around the mean. Considering that fewer features are preferred when stability levels are comparable, the mean value was ultimately selected as the optimal threshold for feature filtering. Fig. 4(b) illustrates the distribution of selected PSI features in all MRD samples using the mean as the threshold. In DA-RT, the majority of features were identified in the high beta band, primarily located between the fronto-parietal and

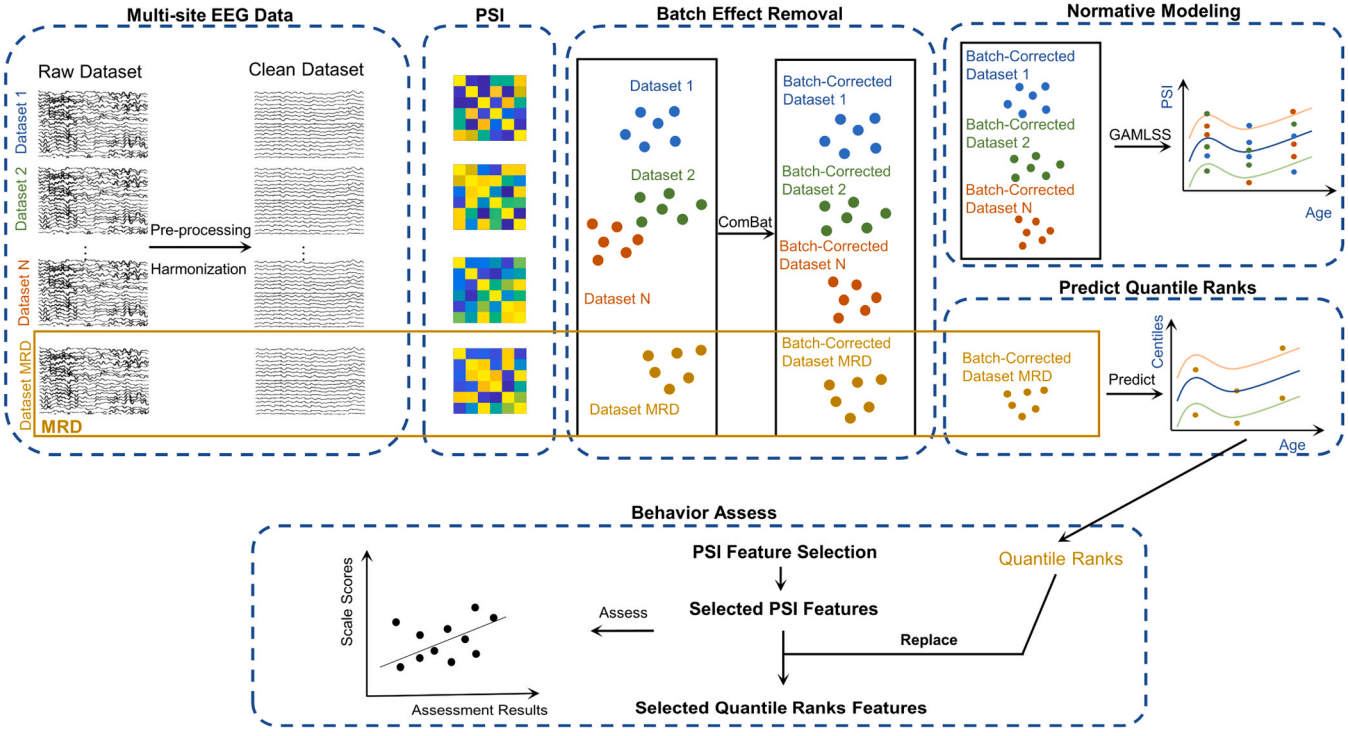

Fig. 1. Assessment framework for large-scale multi-site EEG datasets. Raw EEG data from multiple sites were first preprocessed and standardized to a unified electrode montage; functional connectivity networks were calculated; batch effects were removed using the ComBat method; NMD was used to establish normative models, and MRD was utilized to calculate quantile ranks based on normative models; features from the MRD were selected according to functional connectivity metrics and replaced with corresponding quantile ranks; regression models were trained at last, MRD: mapping relationship database, NMD: normative model database, PSI: phase synchronization index.

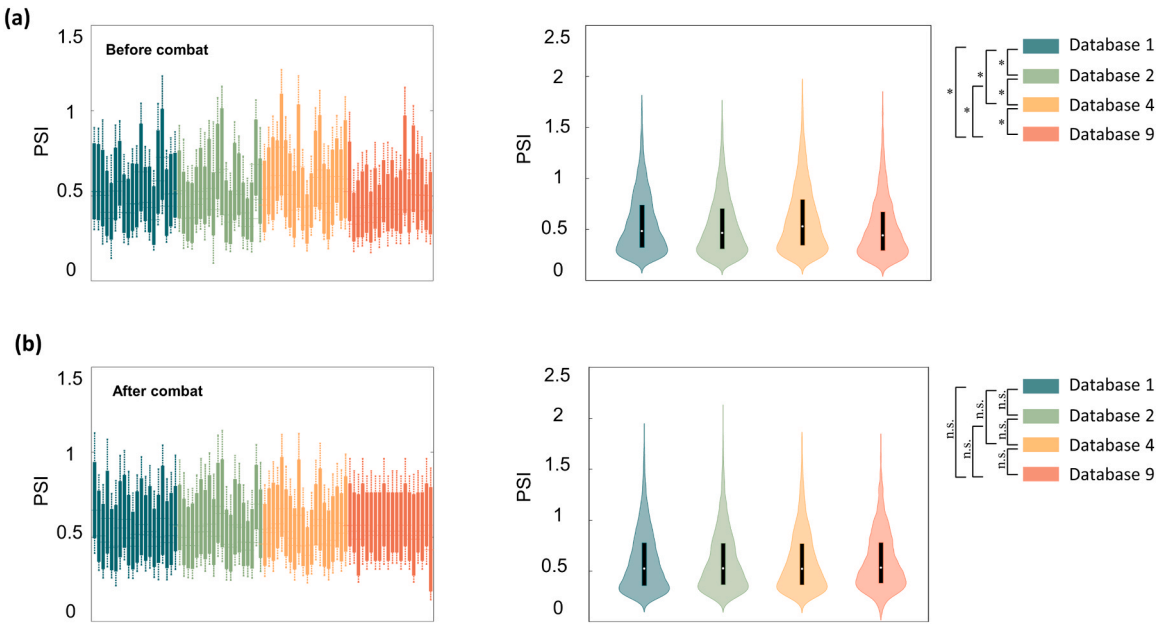

Fig. 2. Batch Effect Removal Results. (a) Visualization of PSI data before batch effect removal and its ANOVA test results; (b) Visualization of PSI data after batch effect removal and the ANOVA results, PSI: phase synchronization index, *: $q < 0.05$, n.s.: no statistical significance.

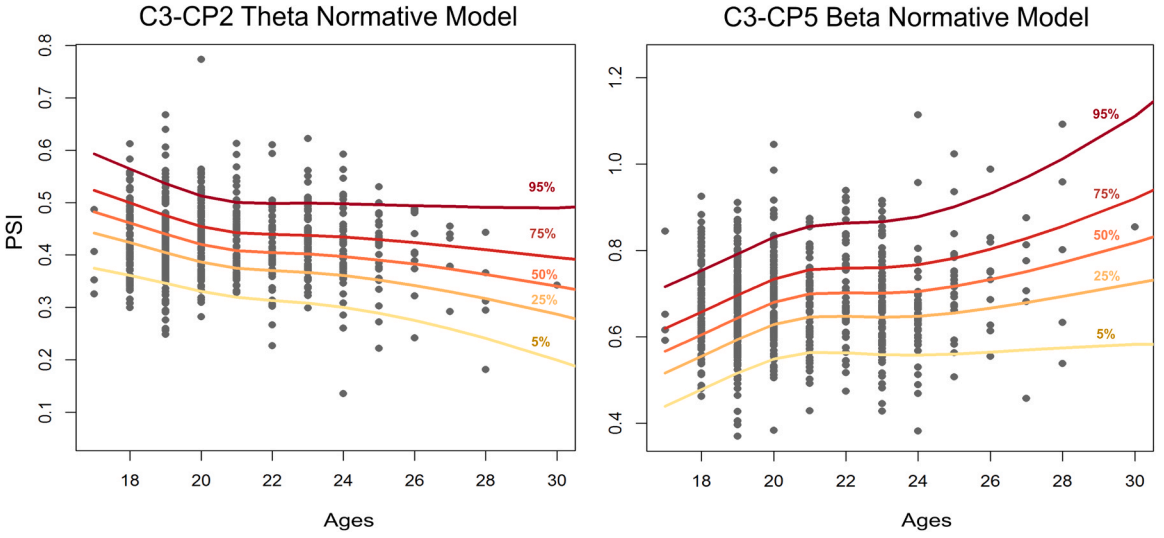

Fig. 3. Normative models. The C3-CP2 theta band normative model, the C3-CP5 beta band normative model, and their corresponding PSI values at different percentiles, PSI: phase synchronization index.

fronto-occipital regions, and these features exhibited a negative correlation with the scale. For DA-AC, most features were concentrated in the alpha band, with their functional connectivity strengths generally showing negative correlations with the scale. In FA-AC, a large proportion of features were derived from the gamma1 and gamma2 bands, which displayed positive correlations with the scale. As for FA-RT, most features were negatively correlated with the scale, predominantly distributed in the beta band, and located between the frontal and parietal regions.

3.2. Framework validation

Fig. 5(a) displays ENR scatter plots, revealing a positive correlation between AC and scale scores and a negative correlation between RT and scale scores. Specifically, DA-AC showed a correlation of 0.5697, while FA-RT had –0.4434, and the regression results based on PSI features are provided in Supplementary Figure S3. Regression models using SVR (Supplementary Figures S4 and S5) confirmed these significant correlations. Moreover, assessment results for the top 20 % versus the bottom 20 % of scale scores (Fig. 5(b)) indicate that individuals with higher scores generally achieve superior evaluations, with statistical tests supporting this trend (Fig. 5(c)). Radar charts (Fig. 5(d)) demonstrate that although DA and FA accuracy did not significantly differ between high and low performers, RT scores were significantly higher in high performers. Finally, assessments based on connectivity metrics COR and COH paralleled the PSI results, whereas evaluations using PLI features were significantly lower (Supplementary Figure S6).

Fig. 6 compares assessment performance based on quantile rank and PSI features. For both ENR (Figs. 6(a) and 6(b)) and SVR (Figs. 6(c) and 6(d)) evaluations, quantile rank features consistently outperformed PSI features; however, the differences were not statistically significant ($q > 0.05$), likely due to limited sample size. When combining ENR and SVR

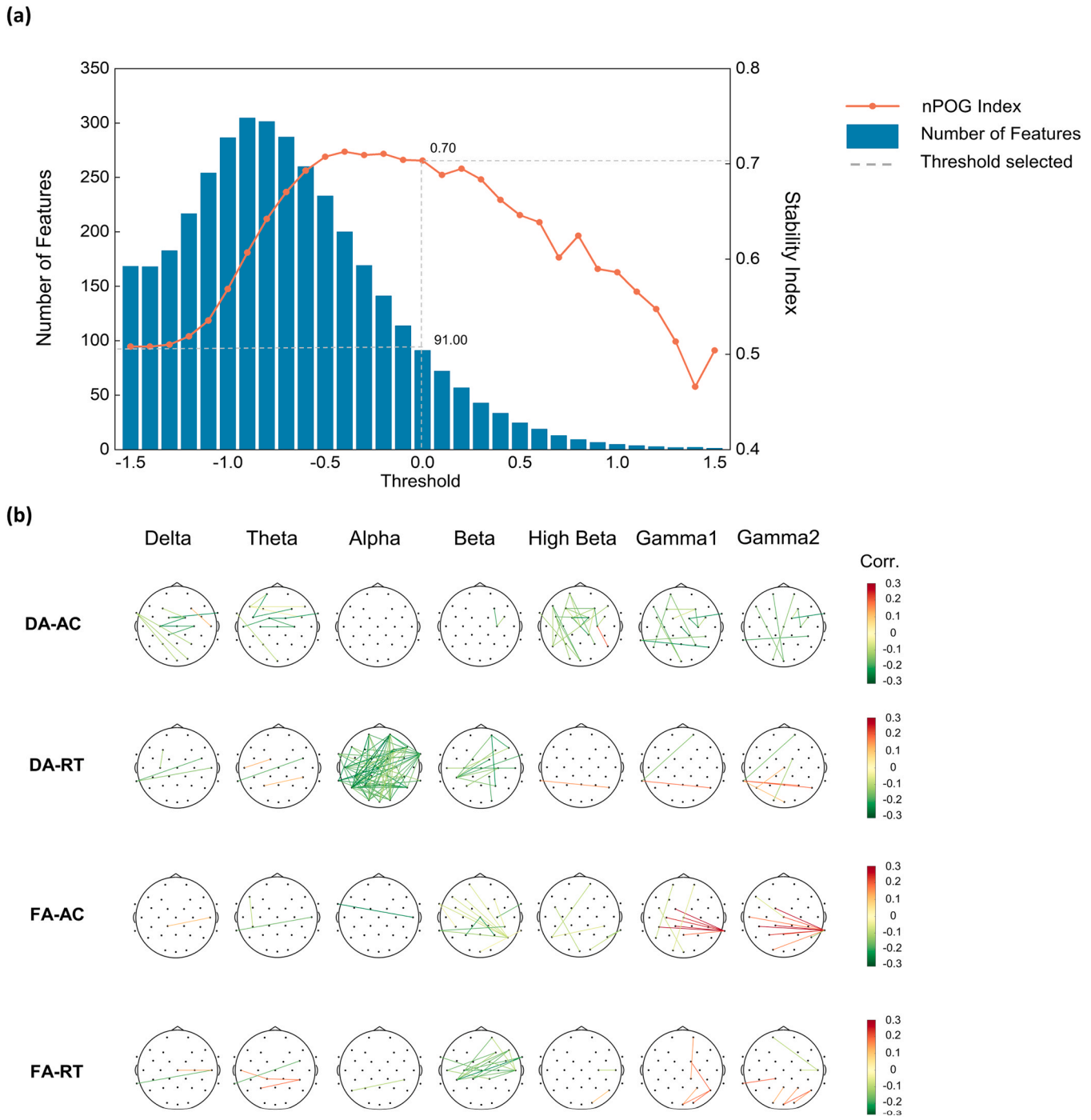

Fig. 4. Feature selection. (a) The number of features selected and their stability evaluation based on a threshold set from the mean - 1.5 SD to + 1.5 SD; (b) Feature selection results with the threshold set as the mean in MRD, MRD: mapping relationship database, DA-AC: distributed attention accuracy, DA-RT: distributed attention reaction time, FA-AC: focused attention accuracy counts, FA-RT: focused attention reaction time, Corr.: correlation between PSI and the scales.

outcomes, the correlations for quantile rank features were significantly higher ($q < 0.05$), and the MSE values were significantly lower (Figs. 6 (f) and 6(g)).

In addition, Supplementary Figure S7(a) illustrates that, across all scales, assessment results for most individuals remained stable with minimal fluctuations across different segments. The ICC calculated in combination with the initial results indicates that the ICC values for all scales exceeded 0.9, reaching an excellent level. Additionally, the average Pearson correlation between each segment and the initial results was greater than 0.65.

4. Discussion

4.1. Framework methodology

This study proposes a normative model-based attention assessment framework for large-scale, multi-site EEG data. First, electrode coordinate transformation was performed using the REST method to unify channel distribution, followed by employing the ComBat algorithm to eliminate batch effects. Next, EEG-derived normative models were established, and quantile ranks of MRD within these models were calculated. Finally, a regression model was developed using selected

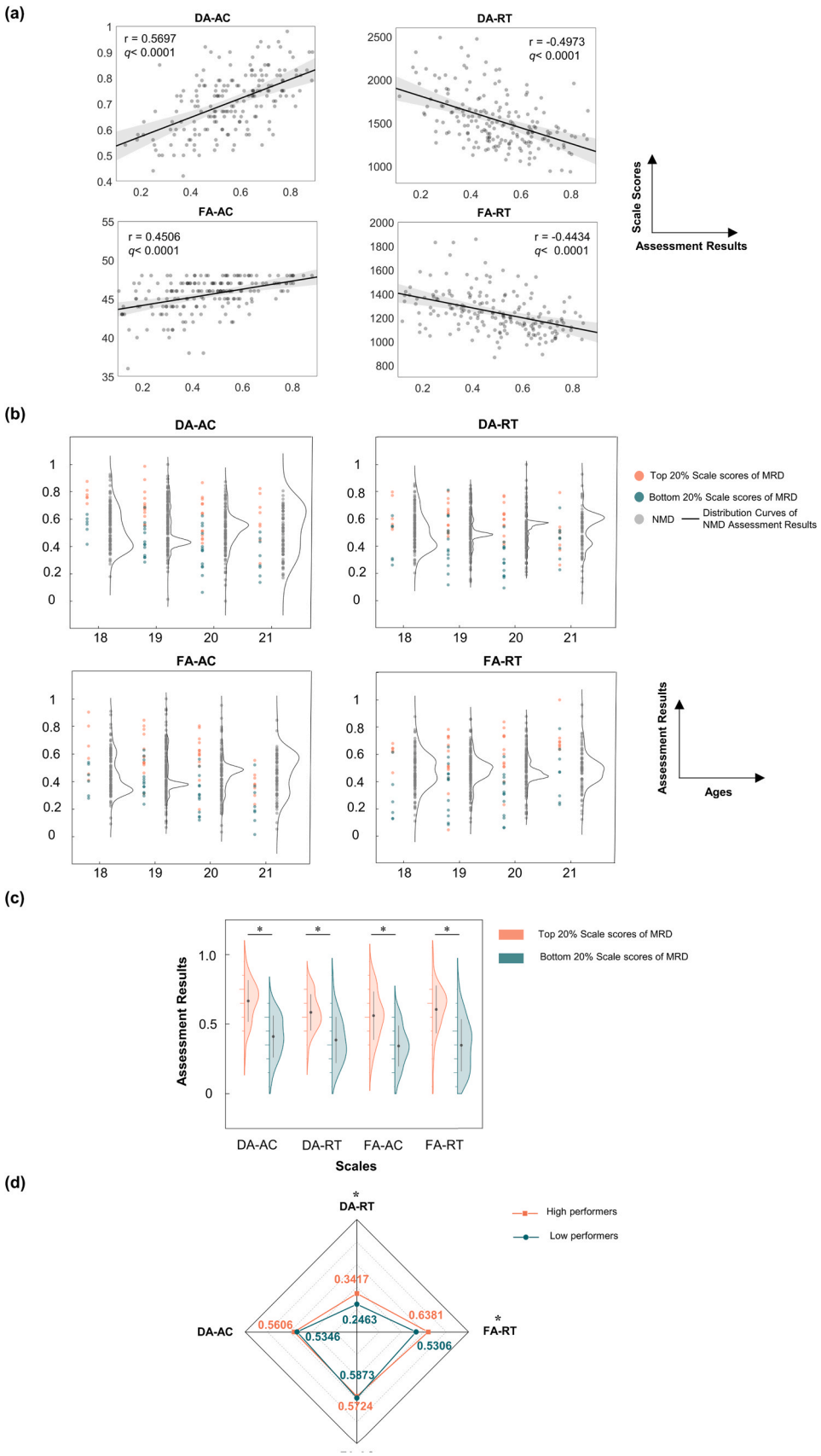

Fig. 5. Framework validation. (a) scatter plots of ENR assessment results based on quantile rank features, (b) assessment results of the top 20 % and bottom 20 % of each scale, (c) *t*-test results for assessment results of the top 20 % and bottom 20 % of each scale, (d) radar plot of comprehensive assessment results for the top 20 % and bottom 20 %. ENR: elastic net regression, DA-AC: accuracy of distributed attention, DA-RT: reaction time of distributed attention, FA-AC: accuracy counts of focused attention, FA-RT: reaction time of focused attention, *: $q < 0.05$.

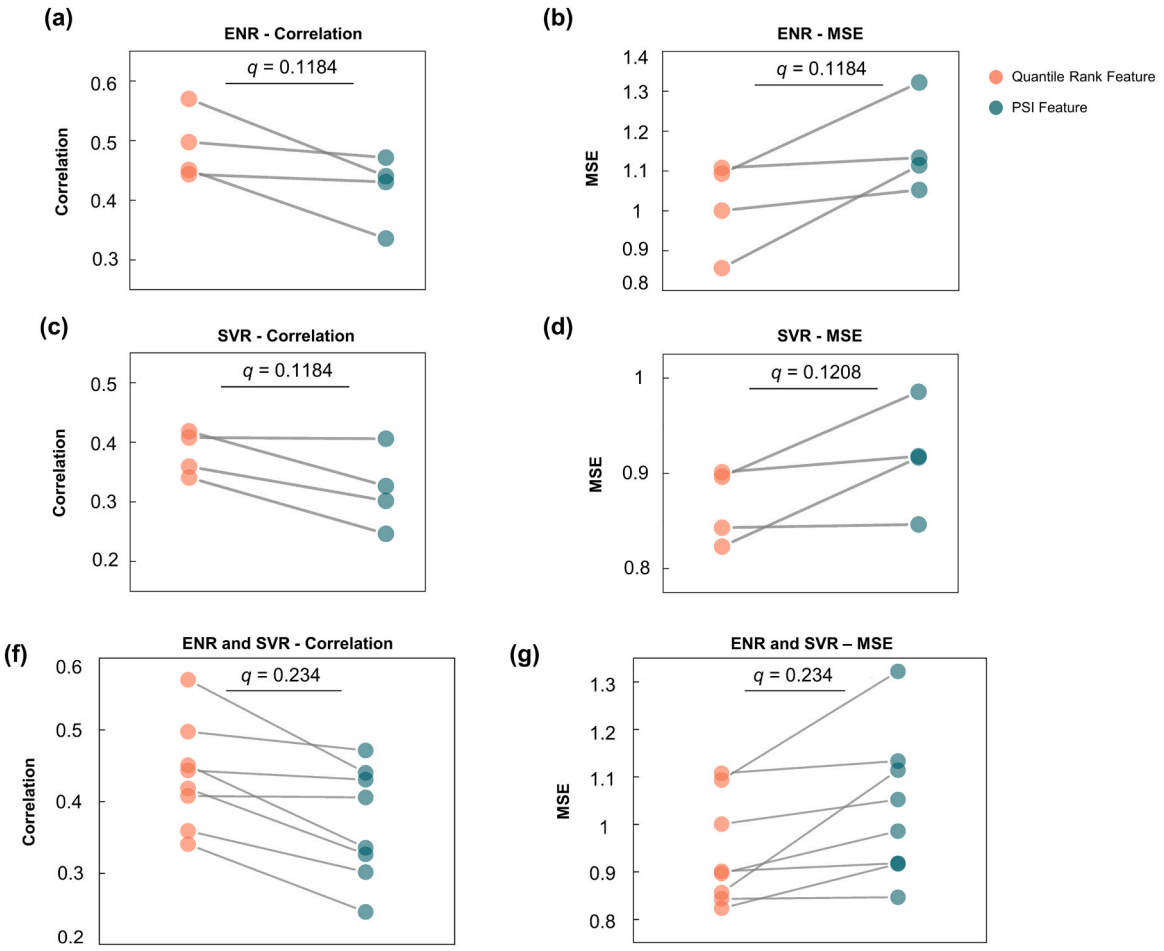

Fig. 6. Comparison of assessment performance based on PSI features and quantile rank features. (a) correlation results of ENR, (b) MSE results of ENR, (c) correlation results of SVR, (d) MSE results of SVR, (e) correlation results of ENR and SVR, (f) MSE results of ENR and SVR. PSI: phase synchronization index, ENR: elastic net regression, SVR: support vector regression, MSE: mean square error.

features to assess attention.

Considering the high-dimensional nature of EEG feature spaces, feature selection is crucial to achieve dimensionality reduction, which directly impacts the interpretability and generalizability of regression models. To assess this, the study evaluated the stability of feature selection across different threshold parameters using the nPOG index, based on 50 bootstrap resampling iterations. As illustrated in Fig. 3(a), the stability metrics displayed a pronounced non-monotonic pattern with increasing threshold levels, first rising and then declining. These results were attributed to the inherent high noise levels and redundancy issues present in EEG signals. At lower threshold levels, more redundant or noise-related features were preserved; although this permitted the capture of additional information to some extent, it also introduced unnecessary instability. Conversely, when excessively high thresholds were applied, some critically contributive features were lost, which led to a decrease in stability. These findings are consistent with the stability-accuracy trade-off principle in feature selection theory (Khaire and Dhanalakshmi, 2022). Notably, a stability index between 0.4 and 0.75 is generally considered intermediate to good (Nogueira et al., 2018). In this study, we selected a threshold of 0.70 based on the dimensionality of the features, suggesting that the chosen feature set exhibits high stability.

Quantile rank features constructed using the GAMLSS normative model demonstrated superior predictive performance compared to raw PSI features, as presented in Fig. 6. Specifically, in separate ENR and SVR analyses, quantile rank features yielded stronger correlations with scale scores and lower MSE. When the two regression models were combined, this advantage reached statistical significance. Several mechanisms underlie these findings. First, by mapping raw features to percentile space, the GAMLSS model dynamically adjusts for age-related nonlinear effects while constraining extreme values, thereby attenuating the impact of outliers (Borghi et al., 2006). Second, quantile ranks capture individual variability as interpretable deviations by reflecting each individual's relative position within the population, enhancing the quantification of heterogeneity (Marquand et al., 2019). Third, using percentiles improves comparability across disparate features or groups, enabling the model to detect latent patterns more effectively (Agelink van Rentergem et al., 2017). Critically, expressing features as age-adjusted percentiles links deviations to lifespan trajectories (spanning development through aging): values near the 10th percentile indicate a level below age-expected norms—consistent with delayed maturation in youth or accelerated decline in older adults—whereas values near the 90th percentile indicate advanced or preserved expression relative to age-matched peers. This converts raw numerical differences into age-anchored, clinically interpretable deviations along normative brain trajectories (Marquand et al., 2016; Bethlehem et al., 2022).

The framework's test-retest reliability was also evaluated, with results indicating a high level of consistency. All scales surpassed an ICC value of 0.9, meeting the benchmark for excellent reliability (Liljequist et al., 2019). The average correlation between the assessment results from all segmented data and the original data exceeded 0.65, demonstrating robust reproducibility. Moreover, the minimum sample size for normative modeling analysis (Supplementary Figure S8) indicated that,

relative to a 150-participant reference per age group in healthy young adults, models constructed with only 50 individuals per age group still achieved 85.0 % consistency with the reference evaluation, thereby demonstrating the robustness of GAMLSS-based normative models and underscoring their feasibility for practical application in large-scale research and clinical contexts.

4.2. Attention-based framework validation

EEG resting-state activity can be regarded as a baseline state that reflects the brain's underlying information-processing capacity (Ramos-Loyo et al., 2004). The neurophysiological patterns identified through our framework corroborate established attention mechanisms. For the DA scale, a larger number of features were selected in the high beta frequency band, which were negatively correlated with accuracy. This finding is consistent with previous studies, suggesting that weaker intrinsic long-range connections enhance the network's reconfiguration ability, thereby better meeting the demands of attention tasks (Rogala et al., 2020; Santarnecchi et al., 2014). Additionally, the RT-related features of the DA scale were primarily concentrated in the alpha frequency band and were inversely proportional to reaction time, indicating that better performance is associated with weaker functional connectivity in the alpha band. This may reflect a decrease in resting-state alpha connectivity, signifying a reduction in the topological properties of default mode network (DMN) nodes, which affects information processing speed and leads to decreased attention levels (Wu et al., 2021; Wantzen et al., 2022). For the FA scale, accuracy was primarily positively correlated with functional connectivity in the gamma1 and gamma2 bands, which may be associated with the link between gamma band connectivity and the ventral attention network (VAN) (Titone et al., 2022). The reaction time of FA was mainly negatively correlated with functional connectivity in the beta frequency band, indicating that better performance is associated with weaker beta band connectivity. This result aligns with previous studies and further supports the notion that individuals with higher network reconfiguration ability in the beta band perform better (Wu et al., 2021; Gross et al., 2004). Therefore, the main patterns identified in this study are consistent with findings from prior research, further validating the effectiveness of the feature selection method used.

The assessment results showed systematic stratification between participants scoring in the top 20 % and bottom 20 % on the scales. In both ENR and SVR analyses, individuals in the top 20 % exhibited significantly stronger AC performance, while those in the bottom 20 % had significantly poorer RT performance—both differences were statistically significant. Among participants with higher overall scores, no notable differences in AC were observed between DA and FA. However, the RT measure displayed highly significant differences, aligning with the clinical neuropsychological consensus that reaction time is a pivotal indicator of cognitive processing efficiency (Miller and Ulrich, 2013). Consequently, these findings support the framework's effectiveness in assessing individual attention.

Beyond validation in healthy adults, this framework also holds promise for clinical applications. Patients with psychiatric or neurological disorders often show systematic alterations in resting-state EEG, such as elevated slow-wave power, reduced alpha coherence, or atypical long-range connectivity. If such deviations shift the underlying distribution relative to the healthy reference, direct use of the existing normative percentiles may misrepresent an individual's position within the population. To address this, the framework can be adapted by recalibrating percentiles using age- and sex-matched control samples, or by building disorder-specific extensions of the normative model (Rutherford et al., 2022). This strategy has been applied in previous EEG test–retest and normative modeling studies, and could improve sensitivity to clinically relevant deviations while preserving comparability with the healthy baseline.

4.3. Limitations

This study has several limitations. First, the sample was mainly healthy individuals, leaving its generalizability to clinical populations uncertain. Second, the analysis focused on specific frequency bands and connectivity metrics; future research should consider additional EEG features (e.g., phase-amplitude coupling, microstates) or advanced signal processing methods. Finally, although the GAMLSS model accounted for age-related effects, covariates such as gender, education, and medication use were not integrated. More comprehensive modeling will be necessary to further elucidate attention processes across diverse populations.

5. Conclusion

In conclusion, this study proposes and validates a novel normative model-based attention assessment framework for large-scale, multi-site EEG data. By combining REST coordinate transformation, ComBat-based harmonization, GAMLSS-based normative modeling, and stability-oriented feature selection, the framework achieved strong predictive performance and reliable reproducibility. Furthermore, our findings highlight the potential of EEG-based quantile rank features for accurate, stable, and scalable attention assessment. These findings establish the framework as a reliable approach for attention assessment, with potential generalization to other EEG-based evaluations and clear value for future research and clinical applications.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Department of Life Science and Technology at the University of Electronic Science and Technology of China (UESTC). All participants provided written informed consent prior to participation.

Funding

The present study was financially supported by was financially supported by the STI 2030-Major Projects (Grant No. 2021ZD0200801) and the CAMS Innovation Fund for Medical Sciences (2019-I2M-5–039).

Authors contribution

Conceived and designed the work: LD and DY. Wrote code: QD, LD, and JL. Acquired the data: YZ, XX, DG, and CL. Analyzed the data: LD, QD, YZ, and PL. Wrote and revised the manuscript: QD, LD, CL, and DY. All authors revised the work for important intellectual content. All of the authors have read and approved the manuscript. This manuscript is original, has not been published before, and is not under consideration elsewhere. All authors have read and approved the final manuscript and agree with the order of authorship.

CRediT authorship contribution statement

Li Dong: Writing – review & editing, Writing – original draft, Validation, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Dezhong Yao:** Validation, Supervision, Funding acquisition, Conceptualization. **Cheng Luo:** Writing – review & editing, Data curation. **Diankun Gong:** Data curation. **Qiwei Dong:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Pengyu Liu:** Investigation, Data curation. **Jianfu Li:** Software. **Yuxi Zhou:** Investigation, Data curation. **Xiaoyu Xiong:** Investigation, Data curation.

Declaration of Competing Interest

All authors have no conflict of interest to disclose.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.brainresbull.2025.111546.

Data availability

Data will be made available on request.

References

Agelink van Rentergem, J.A., Murre, J.M.J., Huizenga, H.M., 2017. Multivariate normative comparisons using an aggregated database. PLOS ONE 12 (3), e0173218.

Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., et al., 2017. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci. Data 4, 170181.

Auger, E., Berry-Kravis, E.M., Ethridge, L.E., 2022. Independent evaluation of the harvard automated processing pipeline for electroencephalography 1.0 using multi-site EEG data from children with fragile x syndrome. J. Neurosci. Methods 371, 109501.

Bethlehem, R.A.I., Seidlitz, J., White, S.R., Vogel, J.W., Anderson, K.M., Adamson, C., et al., 2022. Brain charts for the human lifespan. Nature 604 (7906), 525–533.

Borghi, E., de Onis, M., Garza, C., Van den Broeck, J., Frongillo, E.A., Grummer-Strawn, L., et al., 2006. Construction of the world health organization child growth standards: selection of methods for attained growth curves. Stat. Med. 25 (2), 247–265.

Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., et al., 2018. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54.

Cavanagh, J.F., Bismark, A.W., Frank, M.J., Allen, J.J.B., 2019a. Multiple dissociations between comorbid depression and anxiety on reward and punishment processing: evidence from computationally informed EEG. Comput. Psychiatr. 3, 1–17.

Cavanagh, J.F., Wilson, J.K., Rieger, R.E., Gill, D., Broadway, J.M., Story Remer, J.H., et al., 2019b. ERPs predict symptomatic distress and recovery in sub-acute mild traumatic brain injury. Neuropsychologia 132, 107125.

Chervyakov, A.V., Sinitsyn, D.O., Piradov, M.A., 2016. Variability of neuronal responses: types and functional significance in neuroplasticity and neural darwinism. Front. Hum. Neurosci. 10, 603.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21.

Dezhong, Y., 2001. A method to standardize a reference of scalp EEG recordings to a point at infinity. Physiol. Meas. 22 (4), 693.

Dong, L., Li, J., Zou, Q., Zhang, Y., Zhao, L., Wen, X., et al., 2021a. WeBrain: a web-based brainformatics platform of computational ecosystem for EEG big data analysis. Neuroimage 245, 118713.

Dong, L., Yang, R., Xie, A., Wang, X., Feng, Z., Li, F., et al., 2024. Transforming of scalp EEGs with different channel locations by REST for comparative study. Brain Res. Bull. 217, 111064.

Test-retest reliability in electroencephalographic recordings. In: Suárez-Revelo, J.X., Ochoa-Gómez, J.F., Duque-Grajales, J., Montoya-Betancur, A., Sánchez-López, S. (Eds.), 2015. 2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA), pp. 2–4.

Dong, L., Zhao, L., Zhang, Y., Yu, X., Li, F., Li, J., et al., 2021b. Reference electrode standardization interpolation technique (RESIT): a novel interpolation method for scalp EEG. Brain Topogr. 34 (4), 403–414.

Duan, W., Chen, X., Wang, Y.J., Zhao, W., Yuan, H., Lei, X., 2021. Reproducibility of power spectrum, functional connectivity and network construction in resting-state EEG. J. Neurosci. Methods 348, 108985.

Fortin, J.P., Parker, D., Tunc, B., Watanabe, T., Elliott, M.A., Ruparel, K., et al., 2017. Harmonization of multi-site diffusion tensor imaging data. Neuroimage 161, 149–170.

Frangou, S., Modabbernia, A., Williams, S.C.R., Papachristou, E., Doucet, G.E., Agartz, I., et al., 2022. Cortical thickness across the lifespan: data from 17,075 healthy individuals aged 3–90 years. Hum. Brain Mapp. 43 (1), 431–451.

Ge, C., Wang, Y., Xie, X., Liu, H., Zhou, Z., 2019. Editors. An integrated model for urban subregion house price forecasting: a Multi-source data perspective. 2019 IEEE International Conference on Data Mining (ICDM), pp. 8–11. Nov. 2019.

Gopakumar, S., Tran, T., Nguyen, T.D., Phung, D., Venkatesh, S., 2015. Stabilizing high-dimensional prediction models using feature graphs. IEEE J. Biomed. Health Inf. 19 (3), 1044–1052.

Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., et al., 2004. Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. Proc. Natl. Acad. Sci. 101 (35), 13050–13055.

Habes, M., Pomponio, R., Shou, H., Doshi, J., Mamourian, E., Erus, G., et al., 2021. The brain chart of aging: Machine-learning analytics reveals links between brain aging, White matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. Alzheimer's. Dement. 17 (1), 89–102.

Hany, N., Sherif, R., Emad, K., Emad, A., Elsayed, M., Abdelrahman, H., 2024. Editors. Nexia tutor: an AI-Powered language personalized learning system for kids with dyslexia and Reading challenges. 2024 Int. Mob. Intell. Ubiquitous Comput. Conf. (MIUCC) 13–14. Nov. 2024.

Janiukstyte, V., Owen, T.W., Chaudhary, U.J., Diehl, B., Lemieux, L., Duncan, J.S., et al., 2023. Normative brain mapping using scalp EEG and potential clinical application. Sci. Rep. 13 (1), 13442.

Jose, M.H., Leonid, L.R., Karen, A.S., 2004. Statistical method for detection of Phase-Locking episodes in neural oscillations. J. Neurophysiol. 91 (4), 1883–1898.

Kant, P., Laskar, S.H., Hazarika, J., 2022. Transfer learning-based EEG analysis of visual attention and working memory on motor cortex for BCI. Neural Comput. Appl. 34 (22), 20179–20190.

Kasanov, D., Dorogina, O., Mushtaq, F., Pavlov, Y.G., 2024. Theta transcranial alternating current stimulation is not effective in improving working memory performance. bioRxiv 2024, 585954, 03.20.

Khaire, U.M., Dhanalakshmi, R., 2022. Stability of feature selection algorithm: a review. J. King Saud. Univ. Comput. Inf. Sci. 34 (4), 1060–1073.

Lee, Y.-Y., Hsieh, S., 2014. Classifying different emotional states by means of EEG-Based functional connectivity patterns. PLOS ONE 9 (4), e95415.

Li, X., Zhang, Y., Tiwari, P., Song, D., Hu, B., Yang, M., et al., 2022. EEG based emotion recognition: a tutorial and review. ACM Comput. Surv. 55 (4), 1–57.

Liljequist, D., Elfving, B., Skavberg Roaldsen, K., 2019. Intraclass correlation – a discussion and demonstration of basic features. PLOS ONE 14 (7), e0219854.

Markiewicz, C.J., Gorgolewski, K.J., Feingold, F., Blair, R., Halchenko, Y.O., Miller, E., et al., 2021. The OpenNeuro resource for sharing of neuroscience data. eLife 10, e71774.

Marquand, A.F., Kia, S.M., Zabihi, M., Wolfers, T., Buitelaar, J.K., Beckmann, C.F., 2019. Conceptualizing mental disorders as deviations from normative functioning. Mol. Psychiatry 24 (10), 1415–1424.

Marquand, A.F., Rezek, I., Buitelaar, J., Beckmann, C.F., 2016. Understanding heterogeneity in clinical cohorts using normative models: beyond Case-Control studies. Biol. Psychiatry 80 (7), 552–561.

Miller, J., Ulrich, R., 2013. Mental chronometry and individual differences: modeling reliabilities and correlations of reaction time means and effect sizes. Psychon. Bull. Rev. 20 (5), 819–858.

Nogueira, S., Sechidis, K., Brown, G., 2018. On the stability of feature selection algorithms. J. Mach. Learn. Res. 18, 1–54.

Obeid, I., Picone, J., 2016. The temple university hospital EEG data corpus. Front Neurosci. 10, 196.

Pavlov, Y.G., Kasanov, D., Kosachenko, A.I., Kotyusov, A.I., Busch, N.A., 2022. Pupillometry and electroencephalography in the digit span task. Sci. Data 9 (1), 325.

Price, K.J., Joschko, M., Kerns, K., 2003. The ecological validity of pediatric neuropsychological tests of attention. Clin. Neuropsychol. 17 (2), 170–181.

Quitadamo, L.R., Cavrini, F., Sbernini, L., Riillo, F., Bianchi, L., Seri, S., et al., 2017. Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review. J. Neural Eng. 14 (1), 011001.

Ramos-Loyo, J., Gonzalez-Garrido, A.A., Amezcua, C., Guevara, M.A., 2004. Relationship between resting alpha activity and the ERPs obtained during a highly demanding selective attention task. Int. J. Psychophysiol. 54 (3), 251–262.

Rogala, J., Kublik, E., Krauz, R., Wróbel, A., 2020. Resting-state EEG activity predicts frontoparietal network reconfiguration and improved attentional performance. Sci. Rep. 10 (1), 5064.

Rosenberg, M., Noonan, S., DeGutis, J., Esterman, M., 2013. Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task. Atten. Percept. Psychophys. 75 (3), 426–439.

Rutherford, S., Kia, S.M., Wolfers, T., Fraza, C., Zabihi, M., Dinga, R., et al., 2022. The normative modeling framework for computational psychiatry. Nat. Protoc. 17 (7), 1711–1734.

Santarnecchi, E., Galli, G., Polizzotto, N.R., Rossi, A., Rossi, S., 2014. Efficiency of weak brain connections support general cognitive functioning. Hum. Brain Mapp. 35 (9), 4566–4582.

Sen, R., Mandal, A.K., Chakraborty, B., 2021. A critical study on stability measures of feature selection with a novel extension of lustgarten index. Mach. Learn. Knowl. Extr. 3 (4), 771–787.

Shen, F., Liu, J., Wu, K., 2021. Multivariate time series forecasting based on elastic net and High-Order fuzzy cognitive maps: a case study on human action prediction through EEG signals. IEEE Trans. Fuzzy Syst. 29 (8), 2336–2348.

Smit, D.J.A., de Geus, E.J.C., Boersma, M., Boomsma, D.I., Stam, C.J., 2016. Life-Span development of brain network integration assessed with phase lag index connectivity and minimum spanning tree graphs. Brain Connect. 6 (4), 312–325.

Stam, C.J., Nolte, G., Daffertshofer, A., 2007. Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Hum. Brain Mapp. 28 (11), 1178–1193.

Taylor, P.N., Papasavvas, C.A., Owen, T.W., Schroeder, G.M., Hutchings, F.E., Chowdhury, F.A., et al., 2022. Normative brain mapping of interictal intracranial EEG to localize epileptogenic tissue. Brain 145 (3), 939–949.

Thompson, P.M., Jahanshad, N., Ching, C.R.K., Salminen, L.E., Thomopoulos, S.I., Bright, J., et al., 2020. ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. Transl. Psychiatry 10 (1), 100.

Titone, S., Samogin, J., Peigneux, P., Swinnen, S., Mantini, D., Albouy, G., 2022. Connectivity in Large-Scale Resting-State brain networks is related to motor learning: a High-Density EEG study. Brain Sci. [Internet] 12 (5).

Van Voorhees, E.E., Hardy, K.K., Kollins, S.H., 2010. Reliability and validity of Self- and Other-Ratings of symptoms of ADHD in adults. J. Atten. Disord. 15 (3), 224–234.

Verdonck, T., Baesens, B., Óskarsdóttir, M., vanden Broucke, S., 2024. Special issue on feature engineering editorial. Mach. Learn. 113 (7), 3917–3928.

Wan, W., Cui, X., Gao, Z., Gu, Z., 2021. Frontal EEG-Based Multi-Level attention states recognition using dynamical complexity and extreme gradient boosting. Front Hum. Neurosci. 15, 673955.

Wang, Y., Duan, W., Dong, D., Ding, L., Lei, X., 2022. A test-retest resting, and cognitive state EEG dataset during multiple subject-driven states. Sci. Data 9 (1), 566.
Wantzen, P., Clochon, P., Doidy, F., Wallois, F., Mahmoudzadeh, M., Desaunay, P., et al., 2022. EEG resting-state functional connectivity: evidence for an imbalance of external/internal information integration in autism. J. Neurodev. Disord. 14 (1), 47.
Wu, J., Zhou, Q., Li, J., Chen, Y., Shao, S., Xiao, Y., 2021. Decreased resting-state alpha-band activation and functional connectivity after sleep deprivation. Sci. Rep. 11 (1), 484.
Xiang, C., Fan, X., Bai, D., Lv, K., Lei, X., 2024. A resting-state EEG dataset for sleep deprivation. Sci. Data 11 (1), 427.
Yao, D., 2024. Neuromodulation 3.0: brain-apparatus communication-based neuromodulation. Brain Appar. Commun. A J. Bacomics 3 (1), 2432311.
Zanetti, R., Arza, A., Aminifar, A., Atienza, D., 2022. Real-Time EEG-Based cognitive workload monitoring on wearable devices. IEEE Trans. Biomed. Eng. 69 (1), 265–277.
Zhang, H., Li, J., Su, X., Hu, Y., Liu, T., Ni, S., et al., 2022. Growth charts of brain morphometry for preschool children. NeuroImage 255, 119178.
Zhao, L., Zhang, Y., Yu, X., Wu, H., Wang, L., Li, F., et al., 2023. Quantitative signal quality assessment for large-scale continuous scalp electroencephalography from a big data perspective. Physiol. Meas. 44 (3).