Received: 11 January 2026 | Revised: 6 March 2026 | Accepted: 9 May 2026 | Published online: 3 June 2026

RESEARCH ARTICLE

Medinformatics
2026, Vol. 00(00) 1–9
DOI: 10.47852/bonviewMEDIN62029078

BON VIEW PUBLISHING

Resting-State EEG Complex-Network Descriptors for Predicting Mental Arithmetic Performance

Miguel Angel Vargas Cruz[1,*]

[1] *Grupo Alianza Empresarial, Mexico*

Abstract: Interindividual differences in mental arithmetic performance may be partly constrained by pre-task (resting-state) functional organization measurable from EEG (electroencephalogram). We operationalize organization as a windowed Pearson correlation connectivity graph and construct compact descriptors by combining a discrete Fourier transform–conjugated adjacency representation with eigen-spectral summaries, thereby capturing both local coupling patterns and global network structure in a unified feature space. Using a strictly subject-level, balanced leave-pair-out evaluation on a curated subset of the PhysioNet EEGMAT cohort (n = 20, 10 low- vs 10 high-performance; labels from the dataset's behavioral count-quality annotation), the resting-state pipeline produced perfect separation of performance groups (20/20 correct), indicating that baseline connectivity alone may encode sufficiently discriminative information for between-subject stratification. A complementary task-state setting reached 95% accuracy (19/20 correct), consistent with the repository's short arithmetic segments and their associated temporal constraints. The analysis explicitly avoids within-subject leakage by enforcing subject-wise partitioning, and performance estimates are accompanied by exact confidence intervals appropriate for small-sample evaluation. Given the limited cohort size and electrode density, these results should be interpreted cautiously and require replication on larger samples and higher-density montages; nevertheless, they provide evidence that performance-relevant information is detectable in baseline connectivity structure under rigorous subject-level evaluation.

Keywords: EEG, complex networks, machine learning, resting state, mental arithmetic performance

1. Introduction

Mental arithmetic is a standard experimental probe of numerical cognition because it induces measurable load on attention, working memory, and executive control while remaining behaviorally quantifiable [1, 2]. A persistent open issue is whether interindividual variability in arithmetic efficiency is explained primarily by task-evoked dynamics or whether it is partly constrained by pre-task (baseline) large-scale organization that is already present before engagement. If baseline organization contains performance-relevant information, then pre-task electrophysiology would provide a principled substrate for predicting performance strata without requiring access to extended task recordings.

Positioning relative to prior electroencephalogram (EEG) mental arithmetic literature, most EEG work on mental arithmetic focuses on state discrimination (rest vs arithmetic) or within-subject recognition using spectral and connectivity features, including effective-connectivity and graph-based frameworks [3, 4]. In contrast, the present aim is between-subject stratification of arithmetic performance using only pre-task resting-state recordings, aligning with broader evidence that resting-state EEG connectivity and coherence track individual differences in cognitive functions and mathematical performance [5]. This distinction is made explicit below to avoid conflating performance prediction with task-state decoding.

Functional-connectivity modeling offers an operational route from multichannel EEG time series to network-level descriptors by treating electrodes as nodes and statistical coupling as weighted edges, thereby enabling the use of complex-network constructs to summarize organization beyond local amplitude dynamics [6–9]. Within this perspective, the methodological question becomes whether compact graph-derived features can capture stable, subject-specific organization that correlates with behavioral performance and generalizes under strict subject-level evaluation.

The present study adapts the complex-network EEG classification framework introduced in [10, 11] (originally developed for discriminating neuropsychiatric and neurodegenerative conditions) to a non-pathological stratification problem using mental arithmetic recordings [1]. The adaptation was motivated by an exploratory observation in pre-task connectivity graphs: community-structure-related organization (modularity-linked structure) exhibited a systematic pattern among lower-performing participants, whereas higher-performing participants showed substantially greater heterogeneity, rendering modularity alone insufficient as a single-marker descriptor. This asymmetry suggested that a multi-feature network representation (rather than reliance on a single scalar index) would be necessary, while still

***Corresponding author:** Miguel Angel Vargas Cruz, Grupo Alianza Empresarial, Mexico. Email: miguelangel@grupoalianzaempresarial.com

allowing modularity-linked structure to serve as an empirical cue guiding feature design. This pattern-driven workflow parallels the development logic in [10], but here, it is tested in the context of performance stratification rather than pathology discrimination.

Accordingly, we evaluate the hypothesis that multilevel network descriptors extracted from pre-task EEG can discriminate arithmetic performance groups with high fidelity and that task-state EEG can be classified reliably even when the available arithmetic segments are temporally constrained [1]. The implementation emphasizes (i) windowed correlation graphs [9, 12], (ii) a node-space discrete Fourier transform (DFT)-based change of basis to re-express adjacency structure in a spectral coordinate system [13], (iii) eigen-spectrum summaries as global descriptors [7], and (iv) a regularized learning pipeline combining robust scaling, low-dimensional linear compression, and a nonlinear ensemble classifier [14–17]. Generalization is assessed at the subject level via a balanced leave-pair-out protocol, so reported performance corresponds to out-of-sample predictions for every participant under class-balanced outer evaluation.

2. Methods

2.1. Study design, reporting guideline, and ethics

This work is a secondary analysis of a publicly available, de-identified dataset (PhysioNet EEGMAT) [1, 2]. The study is a prediction-model development and internal validation exercise at the subject level, and reporting is structured in accordance with applicable TRIPOD recommendations [18]. No new human-subject data were collected; therefore, no additional institutional ethics approval was required for this re-analysis beyond the approvals and consent procedures described by the original data providers [2].

2.2. Participants, labels, recordings, and channel selection

A balanced cohort of $n = 20$ participants (Table 1) from the PhysioNet mental arithmetic EEG resource [1] was analyzed at the subject level. Subjects were partitioned into low-performance ($y = 0$, $n_0 = 10$) and high-performance ($y = 1$, $n_1 = 10$) classes by fixed identifier lists (defined a priori in code and held constant across all experiments). The analysis is implemented via two scripts that share the same processing and modeling operators but differ in condition-specific file suffix, channel subset, window duration, and (in one script) the Random Forest (RF) seed.

We index the condition by $\kappa \in$ {rest, task}, with EDF (European Data Format) naming convention

$$\begin{aligned} \kappa &= \text{rest} \Rightarrow \text{Subject} < \text{ID} > \_1.\text{edf}, \\ \kappa &= \text{task} \Rightarrow \text{Subject} < \text{ID} > \_2.\text{edf}. \end{aligned} \tag{1}$$

For each subject $s \in \{1, \ldots, n\}$ and condition κ, a single EDF recording is loaded with MNE (Python package using `read_raw_edf(..., preload=True)`) and then reduced by channel picking to a two-node montage

$$\mathscr{C}_\kappa = \begin{cases} \{EEGC3, EEGPz\}, & \kappa = rest, \\ \{EEGF8, EEGPz\}, & \kappa = task, \end{cases} \quad N = |\mathscr{C}_\kappa| = 2. \tag{2}$$

No additional preprocessing operators (filtering, re-referencing, artifact rejection/ICA, resampling) are applied in the provided scripts beyond the explicit standardization step described below; downstream computations therefore operate on the channel time series returned by `raw.get_data()` for the picked channels. Let $x_{s,c}[n]$ denote the discrete-time EEG sample for subject s and retained channel $c \in \{1, \ldots, N\}$ at time index $n \in \{1, \ldots, T_{s,\kappa}\}$, sampled at subject- and recording-specific sampling frequency $f_{s,\kappa}$ (Hz) obtained from `raw.info ['sfreq']`.

Operationally, the pipeline is executed inside an outer subject-level evaluation loop; in each fold, each required subject-condition EDF is processed independently. If an EDF file is missing and unreadable or yields no valid windows/features after the retention criteria, the corresponding subject instance returns `None` and is dropped from that fold's aggregation by the executor-collection logic (this behavior is implementation-defined by the code path and not corrected by subject-level imputation).

2.3. Global channel standardization (subject-wise; implementation-equivalent)

For each subject-condition recording (s, κ), channel-wise standardization is applied to the full retained-channel time series prior to windowing using the transformation implemented by `StandardScaler` (in code: `StandardScaler().fit_transform(data.T).T`). Writing $T_{s,\kappa}$ for the number of samples in the recording and using the population-variance convention (`ddof=0`), the transformation is

$$\begin{aligned} \mu_{s,\kappa,c} &= \frac{1}{T_{s,\kappa}} \sum_{n=1}^{T_{s,\kappa}} x_{s,\kappa,c}[n], \\ \sigma_{s,\kappa,c} &= \sqrt{\frac{1}{T_{s,\kappa}} \sum_{n=1}^{T_{s,\kappa}} \left(x_{s,\kappa,c}[n] - \mu_{s,\kappa,c}\right)^2}, \\ \tilde{x}_{s,\kappa,c}[n] &= \frac{x_{s,\kappa,c}[n] - \mu_{s,\kappa,c}}{\sigma_{s,\kappa,c}}. \end{aligned} \tag{3}$$

Let $\tilde{\mathbf{x}}_{s,\kappa}[n] = \left(\tilde{x}_{s,\kappa,1}[n], \ldots, \tilde{x}_{s,\kappa,N}[n]\right)^\top \in \mathbb{R}^N$ and $\tilde{\mathbf{X}}_{s,\kappa} = \left[\tilde{\mathbf{x}}_{s,\kappa}[1], \ldots, \tilde{\mathbf{x}}_{s,\kappa}[T_{s,\kappa}]\right] \in \mathbb{R}^{N \times T_{s,\kappa}}$. This step enforces global (record-level) centering and unit-variance scaling per channel and is applied before any segmentation.

Because Pearson correlation is invariant under independent affine transformations per channel, this standardization does not change idealized correlation values (up to floating-point effects) but does fix numerical scale for (i) the window variance degeneracy check and (ii) the stability of subsequent linear-algebraic transforms. The standardization is unsupervised (label-independent) and is computed independently for each subject-condition recording.

2.4. Windowing operator and retention criteria

Each standardized multichannel signal $\tilde{\mathbf{X}}_{s,\kappa}$ is segmented into fixed-length, non-overlapping windows (`overlap=0`). The window duration is condition-specific in the provided scripts:

$$W_\kappa = \begin{cases} 170s, & \kappa = rest, \\ 60s, & \kappa = task. \end{cases} \tag{4}$$

The window length in samples is

$$L_{s,\kappa} = \lfloor W_\kappa \, f_{s,\kappa} \rfloor, \tag{5}$$

and the step size equals $L_{s,\kappa}$. Index windows by $k \in \{1, \ldots, K_{s,\kappa}^{cand}\}$ with start index $a_k = 1 + (k - 1) L_{s,\kappa}$, and define the window matrix

$$\begin{aligned} \boldsymbol{X}_{s,\kappa}^{(k)} &= \left[\tilde{\boldsymbol{x}}_{s,\kappa}[a_k], \tilde{\boldsymbol{x}}_{s,\kappa}[a_k + 1], \ldots, \tilde{\boldsymbol{x}}_{s,\kappa}[a_k + L_{s,\kappa} - 1]\right] \\ &\in \mathbb{R}^{N \times L_{s,\kappa}}, \end{aligned} \tag{6}$$

Table 1. Participant metadata and binary performance labels (y) for the balanced cohort (10 high-performance, 10 low-performance) derived from the PhysioNet EEGMAT subject-info.csv file. Count quality is the dataset-provided good/bad indicator (0 = bad, 1 = good)

Subject	Age	Sex	Year	#Subtractions	Count quality	y
Subject27	19	F	2010	34.59	1	1
Subject13	24	M	2012	34	1	1
Subject03	17	F	2010	31	1	1
Subject34	18	F	2010	31	1	1
Subject25	17	M	2012	30.53	1	1
Subject01	18	F	2011	29.35	1	1
Subject17	17	F	2010	28.7	1	1
Subject23	16	F	2010	27.47	1	1
Subject28	19	F	2010	27	1	1
Subject12	17	F	2010	26.36	1	1
Subject30	17	M	2011	10	0	0
Subject00	21	F	2011	9.7	0	0
Subject14	17	F	2010	9	0	0
Subject04	17	F	2010	8.6	0	0
Subject19	22	M	2010	7.06	0	0
Subject09	16	F	2010	7	0	0
Subject22	19	F	2011	4.47	0	0
Subject06	18	M	2011	4.35	0	0
Subject10	17	F	2010	1	0	0
Subject21	20	F	2010	1	0	0

whenever $a_k + L_{s,\kappa} - 1 \le T_{s,\kappa}$, where

$$K_{s,\kappa}^{cand} = \left\lfloor \frac{T_{s,\kappa}}{L_{s,\kappa}} \right\rfloor. \tag{7}$$

Equivalently, the implementation iterations `start` over

$$\text{start} \in \{0, L_{s,\kappa}, 2L_{s,\kappa}, \ldots\} \text{ with start} + L_{s,\kappa} \le T_{s,\kappa}, \tag{8}$$

so no padding is performed, and any trailing remainder shorter than one full window is discarded.

To exclude numerically degenerate segments, a candidate window is retained only if its empirical variance over all entries exceeds a threshold $\varepsilon = 10^{-6}$:

$$Var\left(\mathbf{X}_{s,\kappa}^{(k)}\right) > \varepsilon. \tag{9}$$

Let $\mathcal{K}_{s,\kappa}$ denote the retained window index set and $K_{s,\kappa} = |\mathcal{K}_{s,\kappa}| \ge 1$. In addition to this variance gate, windows yielding NaNs at later stages (correlation or transformed representation) are discarded by explicit checks in the code.

2.5. Functional connectivity as a windowed correlation adjacency

For each retained window $\mathbf{X}_s^{(k)}$, functional connectivity is operationalized by the Pearson correlation matrix [12]. Writing $X_{s,i,:}^{(k)} \in \mathbb{R}^{L_s}$ for row i of $\mathbf{X}_s^{(k)}$, define the within-window mean

$$\overline{X}_{s,i}^{(k)} = \frac{1}{L_s} \sum_{t=1}^{L_s} X_{s,i,t}^{(k)}. \tag{10}$$

Then the adjacency matrix $\mathbf{A}_s^{(k)} \in \mathbb{R}^{N \times N}$ has entries

$$A_{s,ij}^{(k)} = \frac{\sum_{t=1}^{L_s} \left(X_{s,i,t}^{(k)} - \overline{X}_{s,i}^{(k)}\right)\left(X_{s,j,t}^{(k)} - \overline{X}_{s,j}^{(k)}\right)}{\sqrt{\sum_{t=1}^{L_s} \left(X_{s,i,t}^{(k)} - \overline{X}_{s,i}^{(k)}\right)^2} \sqrt{\sum_{t=1}^{L_s} \left(X_{s,j,t}^{(k)} - \overline{X}_{s,j}^{(k)}\right)^2}}. \tag{11}$$

This is exactly the object computed by `np.corrcoef` when the input matrix is shaped as $(N \times L_s)$ (variables as rows, observations as columns). By construction, $\mathbf{A}_s^{(k)}$ is symmetric, includes self-correlations $A_{s,ii}^{(k)} = 1$ under non-degenerate variance, and yields a weighted undirected functional graph per window in the standard brain network formalism [8, 9]. Windows producing NaNs in $\mathbf{A}_s^{(k)}$ (typically due to near-zero within-window variance) are discarded.

2.6. Node-space DFT conjugation and real-part projection

Each window adjacency $\mathbf{A}_{s,\kappa}^{(k)}$ is transformed by conjugation with the unitary DFT matrix $\mathbf{F}_N \in \mathbb{C}^{N \times N}$:

$$(\boldsymbol{F}_N)_{pq} = \frac{1}{\sqrt{N}} \omega_N^{pq},\ \omega_N = \exp\left(\frac{2\pi i}{N}\right),\ p,\ q \in \{0,\ \ldots,\ N-1\}\ . \tag{12}$$

Define

$$\mathbf{B}_{s,\kappa}^{(k)} = \mathbf{F}_N\, \mathbf{A}_{s,\kappa}^{(k)}\, \mathbf{F}_N^*,\ \mathbf{T}_{s,\kappa}^{(k)} = \mathfrak{R}\left(\mathbf{B}_{s,\kappa}^{(k)}\right) \in \mathbb{R}^{N \times N}, \tag{13}$$

where $(\cdot)^*$ denotes conjugate transpose and $\mathfrak{R}(\cdot)$ the elementwise real part. In the scripts, this corresponds to `transformed_`

`matrix = np.real(dft@matrix@dft.conj().T)`. Matrices containing NaNs after transformation are discarded. For $\overline{N} = 2$, the constructed $\mathbf{F}_2$ is real, and the real-part projection is algebraically redundant (up to floating-point roundoff); for general N, it enforces a real-valued feature representation.

Closed form for the present montage ($N = 2$). For $N = 2$,

$$\mathbf{F}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix},\ \mathbf{A}\begin{bmatrix}1 & r\\ r & 1\end{bmatrix}, r \in [-1,1], \tag{14}$$

and direct multiplication yields

$$\mathbf{T} = \Re\left(\mathbf{F}_2\mathbf{A}\mathbf{F}_2^*\right) = \begin{bmatrix}1+r & 0\\ 0 & 1-r\end{bmatrix}. \tag{15}$$

Under NumPy's default row-major flattening (`.flatten()`), $\mathrm{vec}_C(\mathrm{T}) = [1 + r, 0, 0, 1 - \mathrm{r}]^{\mathrm{T}}$. Since $\mathbf{T}$ is diagonal, `eigvalsh` returns the diagonal entries sorted in ascending order:

$$\begin{aligned}\boldsymbol{\lambda} &= \mathrm{eigvalsh}(\mathrm{T})\\ &= [\min(1+r, 1-r), \max(1+r, 1-r)^{\mathrm{T}}\\ &= [1-|r|, 1+|r|]]^{\mathrm{T}}.\end{aligned} \tag{16}$$

This sorting convention is material when aggregating eigenvalues across windows because sign changes of r across windows affect the mapping between "first" / "second" eigenvalue and the diagonal elements of $\mathbf{T}$.

2.7. Window-level feature construction

For each retained window $k \in \mathcal{K}_s$, two feature families are computed from $\boldsymbol{T}_s^{(k)}$ (this matches the provided implementation):

1. Flattened transformed adjacency:

$$\mathbf{t}_s^{(k)} = vec_C\left(\mathbf{T}_s^{(k)}\right) \in \mathbb{R}^{N^2}. \tag{17}$$

2. Eigen-spectrum of the transformed matrix:

$$\boldsymbol{\lambda}_s^{(k)} = eig\left(\mathbf{T}_s^{(k)}\right) \in \mathbb{R}^N,\ \lambda_{s,1}^{(k)} \le \cdots \le \lambda_{s,N}^{(k)}, \tag{18}$$

computed via a symmetric/Hermitian solver (`eigvalsh`). Eigen-spectral summaries are standard global descriptors in complex-network analysis [7, 9].

In the present $N = 2$ montage, both $\mathbf{t}_s^{(k)}$ and $\boldsymbol{\lambda}_s^{(k)}$ reduce to deterministic functions of the single correlation parameter $r_s^{(k)}$, so the resulting feature vector is a redundant embedding; this redundancy is preserved to maintain compatibility with the general-N template, where vectorized transformed operators and spectra are not generally equivalent.

2.8. Temporal aggregation to a subject-level feature vector

Let $\mathcal{K}_{s,\kappa}$ be the retained window set for subject s in condition κ, with $K_{s,\kappa} = |\mathcal{K}_{s,\kappa}| \ge 1$. Window-level features are aggregated by arithmetic means, consistent with the code (`mean(axis=0)`):

$$\bar{\boldsymbol{t}}_{s,\kappa} = \frac{1}{K_{s,\kappa}}\sum_{k\in\mathcal{K}_{s,\kappa}} \boldsymbol{t}_{s,\kappa}^{(k)} \in \mathbb{R}^{N^2},\ \bar{\boldsymbol{\lambda}}_{s,\kappa} = \frac{1}{K_{s,\kappa}}\sum_{k\in\mathcal{K}_{s,\kappa}} \boldsymbol{\lambda}_{s,\kappa}^{(k)} \in \mathbb{R}^N. \tag{19}$$

The subject-level descriptor concatenates both:

$$\boldsymbol{f}_{s,\kappa} = \begin{bmatrix}\bar{\boldsymbol{t}}_{s,\kappa}\\ \bar{\boldsymbol{\lambda}}_{s,\kappa}\end{bmatrix} \in \mathbb{R}^{N^2+N}. \tag{20}$$

For $N = 2$, $\dim(\mathbf{f}_{s,\kappa}) = 6$. In closed form, writing $r_{s,\kappa}^{(k)}$ for the off-diagonal correlation in window k and defining

$$\overline{r_{s,\kappa}} = \frac{1}{K_{s,\kappa}}\sum_{k\in\mathcal{K}_{s,\kappa}} r_{s,\kappa}^{(k)},\ \overline{|r|}_{s,\kappa} = \frac{1}{K_{s,\kappa}}\sum_{k\in\mathcal{K}_{s,\kappa}} \left|r_{s,\kappa}^{(k)}\right|, \tag{21}$$

one obtains

$$\bar{\boldsymbol{t}}_{s,\kappa} = \left[1+\bar{r}_{s,\kappa},\ 0,\ 0,\ 1-\bar{r}_{s,\kappa}\right]^{\mathrm{T}},\ \bar{\boldsymbol{\lambda}}_{s,\kappa} = \left[1-\overline{|r|}_{s,\kappa},\ 1+\overline{|r|}_{s,\kappa}\right]^{\mathrm{T}}, \tag{22}$$

because `eigvalsh` sorts eigenvalues within each window prior to averaging. If $r_{s,\kappa}^{(k)}$ does not change sign across retained windows, then $\overline{|r|}_{s,\kappa} = |\bar{r}_{s,\kappa}|$ and the two feature families become algebraically coupled; if sign changes occur, the eigenvalue aggregation is governed by the mean absolute correlation magnitude rather than the signed mean.

2.9. Unsupervised cluster-label augmentation (as implemented)

To augment $\boldsymbol{f}_s$ with a coarse discretization of feature space, a K-means model with $K = 2$ clusters is fit and used to generate assignments [19, 20]. Given a design matrix $\boldsymbol{F} \in \mathbb{R}^{\mathrm{m}\times\mathrm{d}}$ collecting m subjects with row vectors $\mathbf{f}_s^{\mathrm{T}}$ $(d = N^2 + N)$, K-means computes

$$(\hat{z}_1, \ldots, \hat{z}_m) = \arg\min_{z_1,\ldots,z_m}\ \min_{\mu_0,\mu_1}\ \sum_{s=1}^{m}\left|\boldsymbol{f}_s - \boldsymbol{\mu}_{z_s}\right|_2^2,\ \hat{z}_s \in \{0,1\}. \tag{23}$$

The assignment is appended as an additional feature:

$$\boldsymbol{f}'_s = \begin{bmatrix}\boldsymbol{f}_s\\ \hat{z}_s\end{bmatrix} \in \mathbb{R}^{d+1}. \tag{24}$$

Train/test fitting behavior (declaration). In the released implementation, cluster features are computed by *independent* `fit_predict` calls on the training subjects and on the two-subject test pair in each outer iteration (with `random_state=1`). Denoting the outer-iteration training and test sets by $\mathcal{S}^{(i)}_{train}$ and $\mathcal{S}^{(i)}_{test}$,

$$\begin{aligned}(\hat{z}_s)_{s\in\mathcal{S}^{(i)}_{train}} &\leftarrow KMeansFitPredict\left(\{\boldsymbol{f}_s\}_{s\in\mathcal{S}^{(i)}_{train}}\right),\\ (\hat{z}_u)_{u\in\mathcal{S}^{(i)}_{test}} &\leftarrow KMeansFitPredict\left(\{\boldsymbol{f}_u\}_{u\in\mathcal{S}^{(i)}_{test}}\right).\end{aligned} \tag{25}$$

Since $\left|\mathcal{S}^{(i)}_{test}\right| = 2$ and $K = 2$, the test-set objective is minimized by assigning one centroid to each test point (zero within-cluster sum of squares), so the appended cluster labels on the test pair constitute an effectively arbitrary binary coding of the two held-out subjects (up to label permutation and initialization). This is reported explicitly because it is the exact procedure executed for the presented results.

2.10. Supervised learning pipeline and hyperparameters

Classification is performed with a scikit-learn pipeline applied to $\mathbf{f}'_{s,\kappa}$:

$$\mathbf{f}'_{s,\kappa} \rightarrow \textit{mean impute} \rightarrow \textit{Robust Scaler} \rightarrow \textit{PCA}\,(r=2) \rightarrow \textit{RF} \rightarrow \hat{y}_{s,k}. \tag{26}$$

The imputer uses feature-wise means. Robust scaling uses the median and IQR per feature (estimated from the training data used to fit the scaler). PCA retains $r = 2$ principal components [21]. The classifier is an RF [10, 12] with a condition-specific seed:

$$\textit{random\_state} = \begin{cases} 3, & \kappa = \textit{rest}, \\ 1, & \kappa = \textit{task}, \end{cases} \tag{27}$$

and a singleton grid

$$n_{\text{trees}} = 100,\ \text{max\_depth} = \text{None},\ \text{min\_samples\_split} = 10. \tag{28}$$

Model selection is therefore degenerate (no competing hyperparameters), but the pipeline is embedded in an inner cross-validation procedure via `GridSearchCV` for standardized fitting/refit behavior.

2.11. Outer evaluation: Paired-index balanced leave-pair-out with inner CV

Generalization is evaluated at the subject level by 10 outer iterations formed by pairing subjects by index position in the two identifier lists. Let $\left\{s_0^{(i)}\right\}_{i=1}^{10}$ be the ordered low-performance list and $\left\{s_1^{(i)}\right\}_{i=1}^{10}$ the ordered high-performance list. In outer iteration i,

$$\mathscr{S}^{(i)}_{test} = \left\{s_0^{(i)}, s_1^{(i)}\right\}, \mathscr{S}^{(i)}_{train} = \{1, \ldots, 20\} \setminus \mathscr{S}^{(i)}_{test}, \tag{29}$$

so exactly one low- and one high-performance subject are held out per iteration (18 training subjects). An inner 9-fold cross-validation is run via `GridSearchCV` on $\mathscr{S}^{(i)}_{train}$ using accuracy as the scoring metric. The call `fit(train_features, train_labels)` ensures that, within the supervised pipeline, imputation, robust scaling, PCA, and RF are fit using only the training data of each inner fold during cross-validation and then refit on the full outer training set for the final estimator before predicting the held-out pair. Test predictions are generated only once per outer iteration by `predict(test_features)` of the refit estimator.

2.12. Prediction metrics and reporting

Aggregating predictions over outer iterations yields $\{(y_s, \hat{y}_s)\}_{s=1}^{20}$ with $y_s \in \{0, 1\}$. In the provided scripts, the reported outputs are (i) `classification_report` (precision, recall, F1-score, and support per class, plus macro/weighted averages), (ii) `confusion_matrix`, and (iii) scalar accuracies computed explicitly in code: overall accuracy and per-class conditional accuracies defined as the fraction of correctly predicted subjects within each true label.

Additional summary indices such as balanced accuracy and Matthews correlation coefficient (MCC) are not computed explicitly in the provided scripts but can be derived deterministically from the confusion matrix if computed in external post-processing. For reference, with counts (TP, TN, FP, FN), MCC is

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \tag{30}$$

2.13. Parallelization and computational implementation

Subject-level preprocessing and feature extraction (EDF loading, windowing, correlation computation, DFT transformation, and window averaging) are parallelized across subjects using process-based execution (`ProcessPoolExecutor`). For each subject identifier in a given fold's training or test list, a worker executes the full feature-extraction chain and returns either a tuple (`features, label`) or `None` if file existence checks fail, EDF loading raises an exception, no windows survive the retention criteria, or NaNs arise in correlation/transformed matrices. The fold-level collector appends only non-`None` results. Random seeds are fixed for K-means (`random_state = 1`) and for RF (condition-specific as stated above), yielding deterministic behavior for those components conditional on library versions and runtime environment.

In addition to printing the console report and confusion matrix, each script writes per-subject predicted labels and accuracy summaries to a results file. In the resting/baseline script, the per-subject listing strings reference `Subject<ID>_1.edf` consistently with the processed files. In the task script, the per-subject listing strings are constructed with the literal `suffix _1.edf` despite processing `Subject<ID>_2.edf`; this affects only textual provenance in the output log (not the actual data processing or predictions).

2.14. Average activity visualizations

For Figures 1, 2, 3, and 4 reported in this work (topographic arrow maps and averaged graphs per condition), the averaging operator is applied across subjects within the corresponding condition after mapping each subject to a window-aggregated connectivity estimate. These visual summaries are included to illustrate interindividual variability in activity and coupling despite shared performance labels and to emphasize that ceiling-level discrimination (when observed) is not driven by trivially visible, condition-averaged graph patterns. The two provided classification scripts do not generate these figures and do not consume them as inputs; they are produced by separate visualization routines external to the predictive pipeline described above.

For the figures reported in this work: (a) depicts the average topographic arrow map for resting state, (b) the average resting-state graph, (c) the average topographic arrow map for mental arithmetic, and (d) the average mental arithmetic graph.

2.14.1. High performance

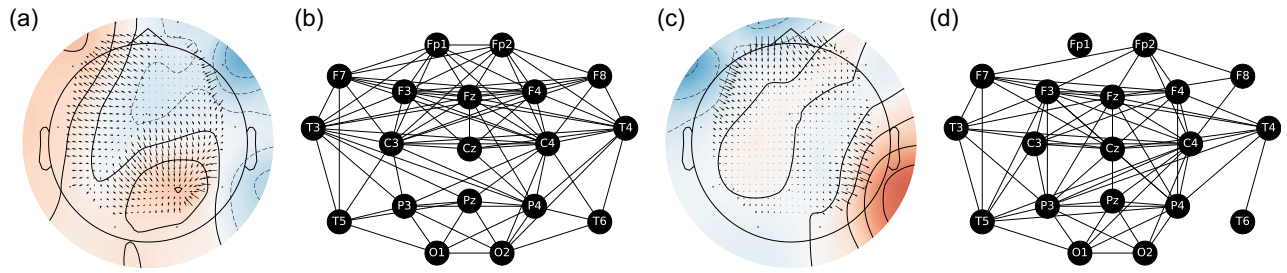

Figure 1. Subject 03. Age: 17. Gender: F. Year: 2010. Operations: 31. Quality: 1. Resting-state modularity: 0.190976350905007. Task modularity: 0.141583786826797

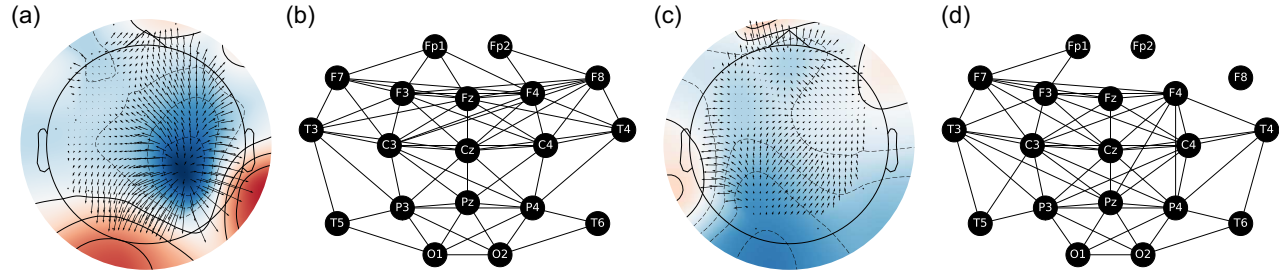

Figure 2. Subject 34. Age: 18. Gender: F. Year: 2010. Operations: 31. Quality: 1. Resting-state modularity: 0.255394857667585. Task modularity: 0.2469482421875

2.14.2. Low performanc

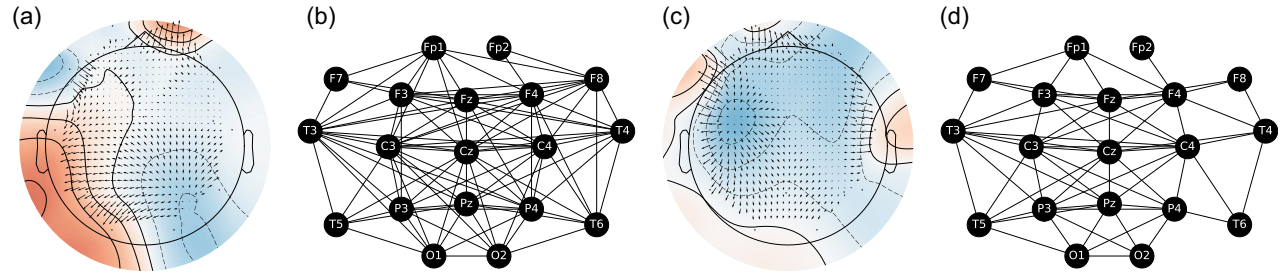

Figure 3. Subject 10. Age: 17. Gender: F. Year: 2010. Operations: 1. Quality: 0.Resting state modularity: 0.140278406518787. Task modularity: 0.2384033203125.

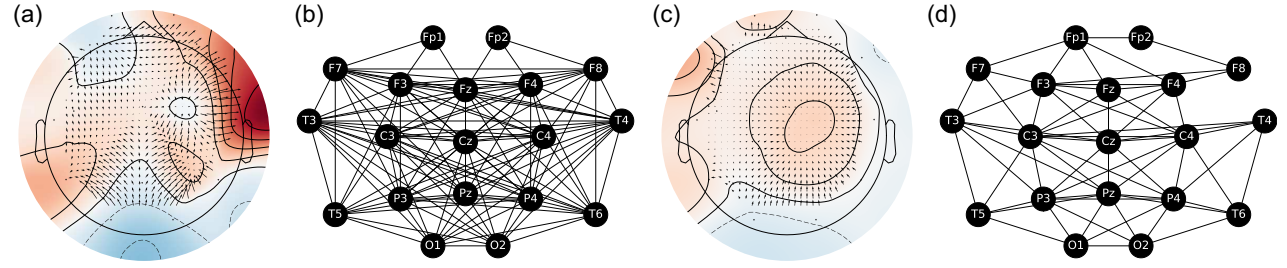

Figure 4. Subject 21. Age: 20. Gender: F. Year: 2010. Operations: 1. Quality: 0. Resting-state modularity: 0.0824738747353323. Task modularity: 0.2303466796875

Table 2. Summary performance metrics under the balanced leave-pair-out evaluation. Confidence intervals are exactbinomial (Clopper–Pearson) intervals for the observed accuracy

Condition	Accuracy	95% CI (Clopper–Pearson)	Balanced accuracy	MCC
Resting state	1.00 (20/20)	[0.832, 1.000]	1.00	1.00
Task state	0.95 (19/20)	[0.751, 0.999]	0.95	0.905

2.15. Code availability

The complete codebase and full subject-level performance outputs are available at the referenced repository link [22], and the summary of the performance metrics is available in Table 2.

3. Results

3.1. Performance-group prediction (high vs low)

Under the subject-level, balanced leave-pair-out evaluation (10 outer iterations; one low- and one high-performance subject held out per iteration), the classifier produced perfect out-of-sample separation between low- and high-performance participants. Across the full cohort (10 subjects per class), overall accuracy and balanced accuracy were both 1.00, with no observed false positives or false negatives. The resulting confusion matrix was:

$$\begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}.$$

Equivalently, the class-conditional recalls were 1.00 for both low- and high-performance labels, and the MCC was 1.00, indicating complete agreement between predicted and true class assignments under the adopted protocol.

3.2. Task-state classification

In a complementary task-state setting derived from the arithmetic recordings [1], the model achieved 95% overall accuracy (19/20 correct), with a single error involving one low-performance subject misclassified as high-performance. The confusion matrix was:

$$\begin{bmatrix} 9 & 1 \\ 0 & 10 \end{bmatrix}.$$

This error pattern implies perfect sensitivity for the high-performance class (recall = 1.00) and a 0.90 recall for the low-performance class, with high-performance precision of 10/11 $\approx$ 0.91 and an MCC of $\approx$ 0.90. The deviation from perfect classification is consistent with the limited available arithmetic segment duration in the dataset relative to the originally intended task duration [1], which can reduce the stability of windowed connectivity estimates and thereby disproportionately affect borderline subjects.

4. Discussion

Two primary findings emerge. First, performance-group prediction reached ceiling accuracy, supporting the claim that resting-state (pre-task) functional organization can encode stable individual differences in arithmetic efficiency, in line with the broader view that graph-theoretic descriptors can capture behaviorally relevant aspects of brain network organization [5, 6]. Second, task-state classification remained near-ceiling with one error, consistent with the expectation that arithmetic engagement induces network-level signatures detectable through connectivity-based representations even when data are temporally constrained [1].

Interpretation and contribution. The main contribution is empirical, under a subject-level protocol in a public dataset, baseline (pre-task) connectivity descriptors were sufficient to stratify arithmetic performance groups, complementing prior work that primarily targets rest–task discrimination rather than performance prediction [3, 4]. The result is also mechanistically consistent with evidence linking resting-state EEG coherence/connectivity to mathematical improvement and other cognitive traits [5]. Given the small sample, ceiling accuracy should be interpreted as a dataset-specific observation rather than a population-level guarantee; to contextualize, the probability of achieving 20/20 correct by independent random guessing is $2^{-20} \approx 9.5 \times 10^{-7}$, while the exact 95% lower bound for accuracy is 0.832 (Table 2).

Conceptually, the feature construction strategy is a direct adaptation of the complex-network EEG classification approach of [10]: direct pairwise coupling (correlation adjacency) encodes local channel-to-channel coordination [12], while transformed-matrix statistics and eigen-spectral summaries provide global descriptors of graph organization [7, 9]. The initial modularity-linked pattern observed in resting-state graphs served as the empirical impetus for implementing this pipeline in a non-pathological context, and the present results indicate that the same methodological template can be informative beyond diagnostic discrimination.

Limitations and validity threats remain material:

1) Segment availability: the arithmetic condition is short in the available data, which can destabilize connectivity estimates and downstream features [1].
2) Montage restriction: only two electrodes were used, constraining the graph to $N = 2$ nodes and limiting neuroanatomical interpretability and robustness to channel choice [8, 9].
3) Small sample size: with $n = 20$, ceiling accuracy warrants conservative interpretation and should be complemented by uncertainty quantification, permutation testing, and external validation.
4) Pipeline audits: in small-n settings, strict separation of training and test transformations is essential, particularly for any unsupervised augmentation [19, 20].
5) Exceptional phenotypes: individuals exhibiting extraordinary or atypical neurocognitive configurations (e.g., extreme compensatory reorganization or rare calculation abilities) would be difficult to categorize within the present framework and may violate implicit population assumptions; however, should such cases be identified, they would warrant dedicated, case-specific experimental designs and higher-density

electrophysiological recordings rather than inclusion in a normative, pathology-free predictive cohort [23–42].

5. Conclusion

A complex-network EEG feature-engineering pipeline originally developed for mental pathology discrimination [10] was adapted to mental arithmetic recordings [1] and achieved ceiling-level separation of low- versus high-performance participants, with near-ceiling classification in a task-state setting. The results motivate replication with larger cohorts, longer task recordings, and higher-density montages, alongside stricter out-of-sample handling of any unsupervised components and formal uncertainty quantification to establish robustness and generalizability.

Ethical Statement

No new human-subject data were collected; therefore, no additional institutional ethics approval was required for this reanalysis beyond the approvals and consent procedures described by the original data providers [2].

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

The EEG recordings analyzed in this study are publicly available from PhysioNet (EEGMAT) [1, 2]. To facilitate reproducibility, the analysis scripts used for feature extraction and model evaluation are provided as supplementary code [22]; all subject identifiers and performance labels used for the balanced cohort are listed explicitly in Table 1, and no additional human data were generated.

Author Contribution Statement

Miguel Angel Vargas Cruz: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Zyma, I., Seleznov, I., Popov, A., Chernykh, M., & Shpenkov, O. (2018). EEG During mental arithmetic tasks [data set]. *PhysioNet*. https://doi.org/10.13026/C2JQ1P

[2] Zyma, I., Tukaev, S., Seleznov, I., Kiyono, K., Popov, A., Chernykh, M., & Shpenkov, O. (2019). Electroencephalograms during mental arithmetic task performance. *Data*, *4*(1), 14. https://doi.org/10.3390/data4010014

[3] Maghsoudi, A., & Shalbaf, A. (2021). Mental arithmetic task recognition using effective connectivity and hierarchical feature selection from EEG signals. *Basic and Clinical Neuroscience*, *12*(6), 817–826. https://doi.org/10.32598/bcn.2021.2034.1

[4] Gangapuram, H., & Manian, V. (2024). Electroencephalogram functional connectivity analysis and classification of mental arithmetic working memory task. *Signals*, *5*(2), 296–325. https://doi.org/10.3390/signals5020016

[5] Zhang, D.-W., Zaphf, A., & Klingberg, T. (2021). Resting state EEG related to mathematical improvement after spatial training in children. *Frontiers in Human Neuroscience*, *15*, 698367. https://doi.org/10.3389/fnhum.2021.698367

[6] Kim, J., & Wilhelm, T. (2008). What is a complex graph? *Physica A: Statistical Mechanics and Its Applications*, *387*(11), 2637–2652. https://doi.org/10.1016/j.physa.2008.01.015

[7] Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256. https://doi.org/10.1137/S003614450342480

[8] Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198. https://doi.org/10.1038/nrn2575

[9] Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

[10] Vargas Cruz, M. A. (2025). Paranoid schizophrenia diagnosis via complex network analysis on EEG data. *Precision Nanomedicine*, *8*(1). https://doi.org/10.33218/001c.128586

[11] Vargas Cruz, M. A. (2025). Diagnosis of Alzheimer's disease, Parkinson's disease, frontotemporal dementia, and paranoid schizophrenia via complex network analysis of EEG data. *Precision Nanomedicine*, *8*(2). https://doi.org/10.33218/001c.133823

[12] Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*(347-352), 240–242. https://doi.org/10.1098/rspl.1895.0041

[13] Bruns, A. (2004). Fourier-, Hilbert- and wavelet-based signal analysis: Are they really different approaches? *Journal of Neuroscience Methods*, *137*(2), 321–332. https://doi.org/10.1016/j.jneumeth.2004.03.002

[14] Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *61*(3), 611–622. https://doi.org/10.1111/1467-9868.00196

[15] Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (*1*, pp. 278–282). IEEE. https://doi.org/10.1109/ICDAR.1995.598994

[16] Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

[17] Smith, P. F., Ganesh, S., & Liu, P. (2013). A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *Journal of Neuroscience Methods*, *220*(1), 85–91. https://doi.org/10.1016/j.jneumeth.2013.08.024

[18] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Journal of British Surgery*, *102*(3), 148–158. https://doi.org/10.7326/M14-0697

[19] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 281–297. University of California Press.

[20] Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

[21] Jolliffe, I. (2002). *Principal component analysis*. Springer. https://doi.org/10.1007/b98835

[22] Vargas, M. (2026). *Complex network analysis of EEG to predict mental arithmetic performance*. https://www.miguelangelvargascruz.com/complexnetworkanalysisofeegtopredictmentalarithmeticperformance_blog_6241.html

[23] Borne, A., Perrone-Bertolotti, M., Jambaqué, I., Castaignède, C., Dorfmüller, G., Ferrand-Sorbets, S., . . . , & Bulteau, C. (2022). Cognitive outcome after left functional hemispherectomy on dominant hemisphere in patients with Rasmussen encephalitis: Beyond the myth of aphasia. *Patient series. Journal of Neurosurgery: Case Lessons*, *4*(22). https://doi.org/10.3171/CASE22410

[24] Colombo, M. A., Favaro, J., Mikulan, E., Pigorini, A., Zauli, F. M., Sartori, I., . . . , & Massimini, M. (2025). Hemispherotomy leads to persistent sleep-like slow waves in the isolated cortex of awake humans. *PLOS Biology*, *23*(10), e3003060. https://doi.org/10.1371/journal.pbio.3003060

[25] Wang, J., Balu, R., Chitale, A., Waziri, A., & Altaweel, L. R. (2024). Continuous quantitative electroencephalography for early detection of acute low-pressure obstructive hydrocephalus in aneurysmal subarachnoid hemorrhage: Illustrative case. *Journal of Neurosurgery: Case Lessons*, *8*(27), CASE24612. https://doi.org/10.3171/CASE24612

[26] Martinez, E., Brandt, A., Dümpelmann, M., & Schulze, A. (2024). Resting state connectivity biomarkers of seizure freedom after epilepsy surgery. *NeuroImage: Clinical*, *44*, 103673. https://doi.org/10.1016/j.nicl.2024.103673

[27] Kalinova, M., Kerkova, B., Kalina, A., Pytelova, V., Amlerova, J., Janca, R., . . . , & Hammer, J. (2026). Temporal order of activations and interactions during arithmetic calculations measured by intracranial electrophysiological recordings in the human brain. *Scientific Reports*, *16*(1), 5587. https://doi.org/10.1038/s41598-026-36122-z

[28] Jafari Malali, M., Sarbaz, Y., Zolfaghari, S., & Khodayarlou, A. (2024). The influence of mental calculations on brain regions and heart rates. *Scientific Reports*, *14*(1), 18846. https://doi.org/10.1038/s41598-024-69919-x

[29] Liu, H., & Liu, H. (2025). A comparative study of EEG functional and effective connectivity patterns in children with learning difficulties during reading and math tasks. *Frontiers in Neuroscience*, *19*, 1612884. https://doi.org/10.3389/fnins.2025.1612884

[30] Chao, W., Wang, E., Yuan, T., He, Q., Zhang, E., & Zhao, J. (2022). Characteristics inhibition defects of children with developmental dyscalculia: Evidence from the ERP. *Frontiers in Psychiatry*, *13*, 877651. https://doi.org/10.3389/fpsyt.2022.877651

[31] Desrosiers, J., Gagnon, D., Ostrolenk, A., Boutros, A., Bernhardt, B. C., Courchesne, V., & Mottron, L. (2025). Calendar calculation: A systematic review of 100 years of research. *Neuroscience & Biobehavioral Reviews*, *178*, 106376. https://doi.org/10.1016/j.neubiorev.2025.106376

[32] Park, H. O. (2023). Autism spectrum disorder and savant syndrome: A systematic literature review. *Journal of the Korean Academy of Child and Adolescent Psychiatry*, *34*(2), 76–92. https://doi.org/10.5765/jkacap.230003

[33] Bjørke, A. B., Østby, Y., Grahl, S. G., Larsson, P. G., Olsen, K. B., Nævra, M. C. J., . . . , & Heuser, K. (2021). Cognition in adult patients with newly diagnosed non-lesional temporal lobe epilepsy. *Epilepsy & Behavior*, *116*, 107771. https://doi.org/10.1016/j.yebeh.2021.107771

[34] Kwok, F. Y., Wilkey, E. D., Peters, L., Khiu, E., Bull, R., Lee, K., & Ansari, D. (2023). Developmental dyscalculia is not associated with atypical brain activation: A univariate fMRI study of arithmetic, magnitude processing, and visuospatial working memory. *Human Brain Mapping*, *44*(18), 6308–6325. https://doi.org/10.1002/hbm.26495

[35] Nobukawa, S., Ikeda, T., Kikuchi, M., & Takahashi, T. (2024). Atypical instantaneous spatio-temporal patterns of neural dynamics in Alzheimer's disease. *Scientific Reports*, *14*(1), 88. https://doi.org/10.1038/s41598-023-50265-3

[36] Wei, Z., Wang, X., Liu, C., Feng, Y., Gan, Y., Shi, Y., . . . , & Deng, Y. (2024). Microstate-based brain network dynamics distinguishing temporal lobe epilepsy patients: A machine learning approach. *NeuroImage*, *296*, 120683. https://doi.org/10.1016/j.neuroimage.2024.120683

[37] Isenstein, E. L., Freedman, E. G., Rico, G. A., Brown, Z., Tadin, D., & Foxe, J. J. (2025). Adults on the autism spectrum differ from neurotypical peers when self-generating but not passively-experiencing somatosensation: A high-density electrophysiological (EEG) mapping and virtual reality study. *NeuroImage*, *311*, 121215. https://doi.org/10.1016/j.neuroimage.2025.121215

[38] Chen, C., Xu, S., Zhou, J., Yi, C., Yu, L., Yao, D., . . . , & Xu, P. (2025). Resting-state EEG network variability predicts individual working memory behavior. *NeuroImage*, *310*, 121120. https://doi.org/10.1016/j.neuroimage.2025.121120

[39] Lin, P., Zhou, X., Zang, S., Zhu, Y., Zhang, L., Bai, Y., & Wang, H. (2023). Early neural markers for individual difference in mathematical achievement determined from rational number processing. *Neuropsychologia*, *181*, 108493. https://doi.org/10.1016/j.neuropsychologia.2023.108493

[40] Sabaghypour, S., Navi, F. F. T., Basiri, N., Shakibaei, F., & Zirak, N. (2024). Differential roles of brain oscillations in numerical processing: Evidence from resting-state EEG and mental number line. *Frontiers in Human Neuroscience*, *18*, 1357900. https://doi.org/10.3389/fnhum.2024.1357900

[41] Marinova, M., & Schiltz, C. (2025). Numerical format integration in primary school children examined with frequency-tagged electroencephalography. *Scientific Reports*, *15*(1), 28641. https://doi.org/10.1038/s41598-025-11281-7

[42] Van Rinsveld, A., & Schiltz, C. (2025). Automatic extraction of meaning from visual number symbols detected by frequency-tagged EEG in children. *Child Development*, *96*(5), 1777–1786. https://doi.org/10.1111/cdev.70002

How to Cite: Cruz, M. A. V. (2026). Resting-State EEG Complex-Network Descriptors for Predicting Mental Arithmetic Performance. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN62029078