QUANTUM-ENHANCED RIEMANNIAN-CONDITIONED GRAPH REPRESENTATION LEARNING FOR ALS-EEG DECODING

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-patient electroencephalography (EEG) decoding for amyotrophic lateral sclerosis (ALS) requires learning from small clinical cohorts while capturing both local EEG evidence and trial-specific inter-component relations. Compact deep encoders capture local temporal–spatial patterns, whereas Riemannian and graph methods explicitly represent relational structure, but these two information sources are typically compressed or aggregated without preserving their component-wise correspondence. We introduce **QRG-EEG**, a quantum-enhanced Riemannian-conditioned graph model that preserves component identity across deep, graph, and quantum representations. Local evidence from component i controls a single-qubit operation on qubit i, while its Riemannian relation (i, j) directly conditions the interaction between qubits (i, j). We analytically show how this edge conditioning becomes observable and induces a joint node–node–edge response. We evaluate QRG-EEG on four public ALS-EEG datasets comprising 30 patient-level LOSO folds, using strict source-only selection, matched classical and quantum controls, and targeted edge interventions. Extensive experiments demonstrate that QRG-EEG achieves the strongest dataset-level Macro-F1 together with secondary metrics (AUROC for P300, balanced accuracy for MI) among the representative competing methods, while removing or misaligning trial-specific edges consistently degrades performance. These results support component-aligned node–edge conditioning as a structured inductive bias for scarce-data cross-patient EEG decoding.

1 INTRODUCTION

Electroencephalography (EEG)-based brain–computer interfaces (BCIs) provide an important pathway for communication and interaction as amyotrophic lateral sclerosis (ALS) progressively restricts voluntary motor control (Chaudhary et al., 2016; Geronimo et al., 2016). Yet reliable EEG decoding in ALS remains difficult because high-quality patient data are intrinsically scarce: existing public ALS-EEG datasets typically contain only a small number of patients, while substantial cross-patient variability further complicates generalization (Mainsah et al., 2025; Riccio et al., 2013; Ngo et al., 2024; Patel et al., 2025a). Compact architectures are therefore attractive in this low-sample regime. Lawhern et al. (2018) proposed EEGNet, whose temporal, depthwise-spatial, and separable convolutions constrain model size while extracting temporal–spatial EEG patterns; the resulting network, however, ultimately maps these patterns to a trial-level classifier and does not expose a separately indexed relation for each latent-component pair.

Riemannian representations provide a natural complement by making trial-specific inter-component structure explicit through symmetric positive-definite (SPD) covariance matrices and tangent-space representations (Barachant et al., 2012; Congedo et al., 2017). Gao et al. (2022) propose to combine convolutional feature extraction with Riemannian processing, while Ju & Guan (2024) develop geometric graph models to preserve structured covariance information on SPD representations. Together, these methods show that covariance geometry can complement local deep features, but their interaction is ultimately resolved through classical feature fusion or graph aggregation. This leaves a **key representational question**: how can local component evidence and pairwise relations remain separately addressable while interacting according to their component identities?

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

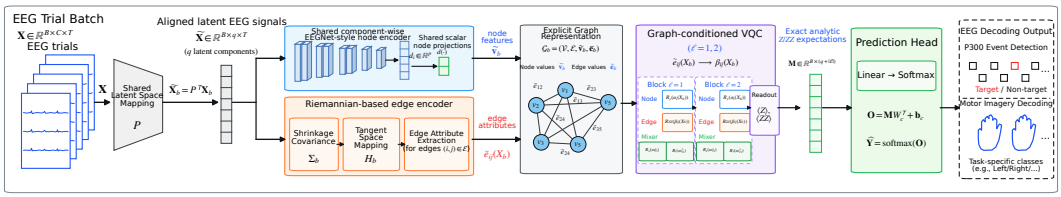

Figure 1: **Overview of QRG-EEG.** A source-fitted projection maps each EEG trial into q shared latent components. A component-wise EEGNet encoder extracts local node features, while a Riemannian branch derives trial-specific relations between the same components, forming a component-aligned EEG graph. The graph-conditioned VQC preserves this correspondence by mapping node i to single-qubit operations on qubit i and edge (i, j) to a two-qubit interaction between qubits (i, j). The resulting Z/ZZ measurements are passed to a linear prediction head. All fitted quantities are estimated exclusively from source patients within each LOSO fold.

Variational quantum circuits (VQCs) offer an alternative interaction primitive for this purpose (Cerezo et al., 2021). QEEGNet (Chen et al., 2024) shows that compressed EEGNet representations can be processed by trainable quantum circuits, while quantum graph models show that graph structure and continuous edge weights can condition pairwise quantum interactions (Mernyei et al., 2022; Skolik et al., 2023). This motivates a component-aligned construction in which each latent EEG component retains a shared identity as a deep feature, graph node, and qubit: local evidence from component i controls a single-qubit operation on qubit i, whereas the Riemannian relation between components (i, j) directly conditions the corresponding two-qubit interaction. We therefore introduce **QRG-EEG**, a quantum-enhanced Riemannian-conditioned graph representation that jointly preserves local temporal–spatial evidence and trial-specific relational geometry for data-limited cross-patient ALS-EEG decoding.

Fig. 1 and 2 illustrate our method's model architecture and the graph-conditioned VQC. In QRG-EEG, we construct a shared latent space in which component $i \equiv$ graph node $i \equiv$ qubit i: component-wise deep features provide local node evidence, while Riemannian features encode trial-specific relations between the same components. The graph-conditioned VQC maps node values to single-qubit rotations and each relation (i, j) to a data-conditioned two-qubit interaction on the corresponding qubits. This component-aligned design keeps local evidence and pairwise geometry separately identifiable while allowing them to interact, rather than collapsing both into a single embedding before classification. We evaluate this mechanism on four public ALS-EEG datasets comprising 30 held-out ALS patients under strict patient-level LOSO evaluation, using embedding-only quantum models, matched classical controls, and targeted interventions that remove, fix, or misalign the Riemannian edge pathway. Across the evaluated folds, QRG-EEG achieves the strongest dataset-level performance among representative competing methods, while the controlled interventions support the joint importance of local node evidence, trial-specific Riemannian relations, and their component-aligned interaction. Our contributions are threefold:

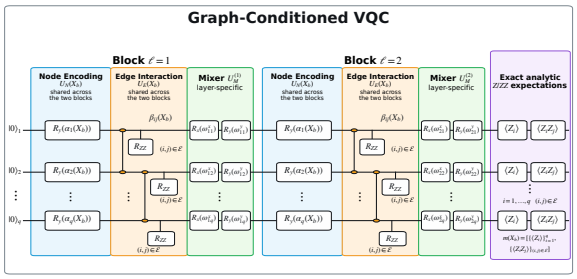

Figure 2: **Graph-conditioned VQC in QRG-EEG.** Each block maps node feature v_i to a single-qubit rotation and aligned Riemannian edge e_{ij} to an R_{ZZ} interaction on qubits (i, j). The two blocks share node/edge maps but use block-specific noncommuting mixers to expose edge-conditioned phases to Z/ZZ readout, yielding $\mathbf{m}_b \in \mathbb{R}^{q+|\mathcal{E}|}$.

- We propose **QRG-EEG**, which represents local EEG features as graph nodes and trial-specific Riemannian relations as graph edges, and maps them to single-qubit and corresponding two-qubit operations, respectively. This preserves which EEG components each feature and relation belongs to throughout the quantum interaction.

- We theoretically characterize **why this node–edge quantum interaction is meaningful**: Riemannian edge information can directly affect the measured quantum representation and interact jointly with the features of its connected nodes, rather than being treated as an independent input vector.
- We evaluate QRG-EEG on **four ALS-EEG datasets** under strict cross-patient LOSO evaluation. Comparisons with quantum and classical baselines, together with controlled ablations, show that both trial-specific Riemannian relations and their correct node–edge alignment contribute to the observed improvements.

2 Related Work

ALS-EEG decoding and structured representations. ALS-EEG decoding is challenging under small patient cohorts and substantial cross-patient variability (Riccio et al., 2013; Geronimo et al., 2016). Compact models such as EEGNet (Lawhern et al., 2018) and FBCNet (Mane et al., 2021) address the low-sample setting through parameter-efficient temporal–spatial modeling, while pretrained models such as LaBraM (Jiang et al., 2024) and CBraMod (Wang et al., 2025) exploit large external EEG corpora. Complementarily, Riemannian methods represent EEG covariance on the SPD manifold and support cross-subject classification and alignment (Barachant et al., 2012; Congedo et al., 2017; Zanini et al., 2018). Geometric deep models further combine learned features with SPD representations, including CRGNet (Gao et al., 2022), Tensor-CSPNet (Ju & Guan, 2022), Graph-CSPNet (Ju & Guan, 2024), MAtt (Pan et al., 2022), and TSMNet (Kobler et al., 2022), while EEG graph networks model spatial or functional relations explicitly (Zhong et al., 2022; Ding et al., 2023; Patel et al., 2025b). QRG-EEG differs operationally by keeping local node evidence and pairwise Riemannian geometry separately addressable throughout their interaction.

Quantum EEG classifiers and relation-conditioned quantum circuits. QEEGNet (Chen et al., 2024) applies a VQC to compressed EEGNet features, while pyRiemann-qiskit (Andreev et al., 2023) and related ERP experiments (Cattan & Quemy, 2023) feed covariance or tangent-space features into generic quantum classifiers. Separately, quantum graph models encode graph topology or continuous edge weights through structured pairwise interactions (Verdon et al., 2019; Mernyei et al., 2022; Skolik et al., 2023). QRG-EEG connects these directions through an EEG-specific alignment: node i controls qubit i, while the Riemannian relation (i, j) conditions the interaction of the corresponding qubit pair. The contribution is therefore not the generic node-to-qubit or edge-to-interaction idea, but preserving EEG component identity across deep, Riemannian, graph, and quantum representations.

Evaluation of variational quantum models. Quantum trainability alone does not establish a useful quantum model; barren plateaus, classical simulability, and classical surrogates remain important concerns (Schuld et al., 2019; Larocca et al., 2025; Cerezo et al., 2025; Landman et al., 2023; Schreiber et al., 2023). We therefore evaluate QRG-EEG against embedding-only quantum, vectorized Riemannian-quantum, same-graph, and parameter-matched classical controls, and restrict our claims to predictive and mechanistic evidence under exact simulation.

3 Methodology

In this section, we present QRG-EEG, which constructs a component-aligned graph from deep node embeddings and Riemannian relations, encodes it into a graph-conditioned VQC, and trains the model with subject-aware objectives under strict source-only LOSO evaluation.

Problem setting and source-only evaluation. Let patient s contribute $\mathcal{D}_s=\{(\mathbf{X}_{s,n}, y_{s,n})\}_{n=1}^{|\mathcal{D}_s|}$, where $\mathbf{X}_{s,n} \in \mathbb{R}^{C\times T}$ denotes an EEG trial and $y_{s,n} \in \{1, \ldots, K\}$ its label. For held-out patient h, we define $\mathcal{D}_{\text{test}}^{(h)} = \mathcal{D}_h$ and $\mathcal{D}_{\text{src}}^{(h)} = \bigcup_{s\neq h} \mathcal{D}_s$, and learn $f_{\boldsymbol{\Theta}} : \mathbb{R}^{C\times T} \to \Delta^{K-1}$ exclusively from $\mathcal{D}_{\text{src}}^{(h)}$. All fitted quantities, including preprocessing statistics, representations, graph construction, model selection, and calibration, are determined from source patients only. We omit the fold superscript below; Appendix B.2 enumerates all fold-local quantities.

Within this source-only setting, QRG-EEG first maps each trial into a shared latent component space that provides common indexing across the deep, Riemannian, graph, and quantum representations

(Fig. 1). For a batched input $\mathbf{X} \in \mathbb{R}^{B\times C\times T}$, a source-fitted map $\mathbf{P} \in \mathbb{R}^{C\times q}$ gives, for trial b,

$$\widetilde{\mathbf{X}}_b = \mathbf{P}^\top \mathbf{X}_b \in \mathbb{R}^{q\times T}. \tag{1}$$

Thus, latent component i, deep node i, covariance coordinate i, graph node i, and qubit i share the same identity. This indexing ensures that a relation computed between components (i, j) remains associated with the same pair throughout graph construction and quantum interaction.

Within each fold, source-patient validation selects $q \in \{6, 8\}$ and the task-specific representation: CSP or PCA for MI and xDAWN for P300 (Ramoser et al., 2000; Rivet et al., 2009). The primary QRG-EEG circuit (L=2, graph multiplier 1.5) is fixed before outer-fold evaluation and is never selected using the held-out patient.

Component-aligned graph construction. We construct, for each trial, a graph in which node evidence and edge relations are defined over the same q latent components, preserving their identity so that the downstream quantum circuit can act on structurally corresponding qubits and qubit pairs.

▷ **Node evidence.** For each latent component, we apply a shared EEGNet-style temporal–separable encoder (Lawhern et al., 2018):

$$\mathbf{d}_{b,i} = \eta_{\boldsymbol{\phi}}(\widetilde{\mathbf{X}}_{b,i,:}) \in \mathbb{R}^{d_{\text{node}}}, \qquad v_{b,i} = \mathbf{w}_v^\top \mathbf{d}_{b,i} + b_v^{\text{node}}. \tag{2}$$

The encoder and scalar projection are shared across components, while the component index i is retained without pooling before the quantum interaction. Source-fitted median/MAD standardization and clipping yield $\widetilde{v}_{b,i} \in [-3, 3]$.

▷ **Riemannian relations.** The geometric branch operates on a source-scaled copy of the same q projected components, preserving their shared identities. The task-specific geometric input construction is detailed in Appendix B.2. From the resulting shrinkage covariance $\boldsymbol{\Sigma}_b \in \mathbb{S}^q_{++}$ (Chen et al., 2010), we compute the source-referenced whitened tangent representation (Barachant et al., 2012; Congedo et al., 2017):

$$\mathbf{H}_b = \log\Big(\overline{\boldsymbol{\Sigma}}_{\text{src}}^{-1/2} \boldsymbol{\Sigma}_b \overline{\boldsymbol{\Sigma}}_{\text{src}}^{-1/2}\Big), \qquad e_{b,ij} = [\mathbf{H}_b]_{ij}, \tag{3}$$

where we denote N_{src} as the number of source trials, $\boldsymbol{\Sigma}_r$ as source trial r's covariance, and $\overline{\boldsymbol{\Sigma}}_{\text{src}}$ as source covariances' affine-invariant Riemannian mean $\overline{\boldsymbol{\Sigma}}_{\text{src}} = \arg\min_{\boldsymbol{\Upsilon}\in\mathbb{S}^q_{++}} \sum_{r=1}^{N_{\text{src}}} \delta^2_{\text{AIRM}}(\boldsymbol{\Upsilon}, \boldsymbol{\Sigma}_r)$. Hence, $e_{b,ij}$ is a trial-specific geometric relation between the same latent components represented by nodes i and j.

▷ **Sparse graph construction.** We define the fold-specific edge importance as

$$\gamma_{ij} = \frac{1}{N_{\text{src}}} \sum_{r=1}^{N_{\text{src}}} |[\mathbf{H}_r]_{ij}|. \tag{4}$$

A maximum-weight spanning tree first ensures connectivity, after which the strongest remaining edges are added until $|\mathcal{E}| = \text{round}(1.5q)$. The topology $\mathcal{E}$ is fixed within each LOSO fold, whereas the edge values $e_{b,ij}$ remain trial-specific. Source-fitted robust standardization and clipping yield $\widetilde{e}_{b,ij} \in [-3, 3]$. Each trial is therefore represented as

$$\mathcal{G}_b = (\mathcal{V}, \mathcal{E}; \widetilde{\mathbf{v}}_b, \widetilde{\mathbf{e}}_b), \qquad \mathcal{V} = \{1, \ldots, q\}. \tag{5}$$

By construction, node i, edge (i, j), and their underlying EEG components retain the same indices, enabling the graph-conditioned circuit to apply each node and relation to the corresponding qubit or qubit pair.

Graph-conditioned VQC. Starting from $|0\rangle^{\otimes q}$, each of the two blocks re-encodes the aligned node and edge attributes using trainable scales shared across blocks:

$$\alpha_{b,i} = \pi \tanh(\theta_i^N \widetilde{v}_{b,i}), \quad U_N(\mathbf{X}_b) = \textstyle\prod_{i=1}^q R_y^{(i)}(\alpha_{b,i}), \quad \beta_{b,ij} = \tfrac{\pi}{2} \tanh(\theta_{ij}^E \widetilde{e}_{b,ij}), \quad U_E(\mathbf{X}_b) = \textstyle\prod_{(i,j)\in\mathcal{E}} R_{ZZ}^{(i,j)}(\beta_{b,ij}), \tag{6}$$

where $R_{ZZ}^{(i,j)}(\beta) = \exp(-\mathrm{i}\beta Z_i Z_j/2)$, and θ_i^N and θ_{ij}^E are trainable node- and edge-specific scales. Thus, node i controls qubit i, while the trial-specific Riemannian relation $\widetilde{e}_{b,ij}$ directly conditions the interaction between the corresponding qubits (i, j). This preserves the component alignment

in the constructed graph and specializes weighted-edge quantum circuits to EEG relational geometry (Mernyei et al., 2022; Skolik et al., 2023). Each block is followed by a block-specific noncommuting mixer,

$$U_M^{(\ell)} = \prod_{i=1}^{q} R_y^{(i)}(\omega_{\ell i}^y) R_x^{(i)}(\omega_{\ell i}^x), \qquad U^{(\ell)}(\mathbf{X}_b) = U_M^{(\ell)} U_E(\mathbf{X}_b) U_N(\mathbf{X}_b), \tag{7}$$

where the implementation executes R_x followed by R_y. The full two-block circuit is

$$U_{\mathbf{\Theta}}(\mathbf{X}_b) = U^{(2)}(\mathbf{X}_b) U^{(1)}(\mathbf{X}_b). \tag{8}$$

The node and edge scales are shared across blocks, whereas mixer angles are block-specific. Because R_{ZZ} is diagonal in the Z basis, the post-edge noncommuting mixer provides a path for edge-conditioned phases to affect Z-basis observables; the conditions for nonzero sensitivity are characterized in Sec. 4.

After the second block, we measure

$$\mathbf{m}_b = \left[\{\langle Z_i \rangle\}_{i=1}^{q}, \{\langle Z_i Z_j \rangle\}_{(i,j)\in\mathcal{E}}\right]^\top \in \mathbb{R}^{q+|\mathcal{E}|}. \tag{9}$$

For a batch, stacking $\mathbf{m}_b^\top$ gives $\mathbf{M} \in \mathbb{R}^{B\times(q+|\mathcal{E}|)}$, and the linear prediction head computes $\mathbf{O}{=}\mathbf{M}\mathbf{W}_c{+}\mathbf{1}\mathbf{b}_c^\top$ and $\widehat{\mathbf{Y}}{=}\,\mathrm{softmax}(\mathbf{O})$ row-wise. All quantum measurements use exact analytic statevector expectations.

Subject-aware optimization. Training proceeds in three stages. We first pretrain the component encoder with a temporary linear classifier on the scalar node vector $\mathbf{v}_b$. We then freeze the encoder and train the graph-conditioned circuit and readout, before jointly fine-tuning the final encoder block, scalar projection, circuit, and readout at lower learning rates. This schedule first establishes discriminative local EEG evidence, then learns node–edge interactions, and finally adapts both jointly.

To account for heterogeneous patient sizes and class imbalance, let ε_n denote the cross-entropy for sample n, $\mathcal{S}_\mathcal{I}$ the patients represented in minibatch $\mathcal{I}$, $\mathcal{I}_s$ the samples from patient s, $\mathcal{I}_{s,k}$ the samples from patient s and class k, and $\mathcal{K}_s$ the classes represented for patient s in the minibatch. We define

$$\mathcal{J}_{\mathrm{nat}} = \frac{1}{|\mathcal{S}_\mathcal{I}|} \sum_{s\in\mathcal{S}_\mathcal{I}} \frac{1}{|\mathcal{I}_s|} \sum_{n\in\mathcal{I}_s} \varepsilon_n, \qquad \mathcal{J}_{\mathrm{pc}} = \frac{1}{|\mathcal{S}_\mathcal{I}|} \sum_{s\in\mathcal{S}_\mathcal{I}} \frac{1}{|\mathcal{K}_s|} \sum_{k\in\mathcal{K}_s} \frac{1}{|\mathcal{I}_{s,k}|} \sum_{n\in\mathcal{I}_{s,k}} \varepsilon_n. \tag{10}$$

Here, $\mathcal{J}_{\mathrm{nat}}$ gives each patient equal weight while preserving its natural class distribution, whereas $\mathcal{J}_{\mathrm{pc}}$ additionally balances classes within each patient.

We further encourage within-patient class separation using a ranking loss. For one-vs-rest log-odds $\xi_{n,k} = o_{n,k} - \log\sum_{a\neq k}\exp(o_{n,a})$, $\mathcal{J}_{\mathrm{rank}}$ averages the margin loss $\mathrm{softplus}(m - \xi_{n^+,k} + \xi_{n^-,k})$ over within-patient positive–negative pairs and eligible classes. Therefore, the final objective is

$$\mathcal{J} = w_{\mathrm{nat}}\mathcal{J}_{\mathrm{nat}} + w_{\mathrm{pc}}\mathcal{J}_{\mathrm{pc}} + w_{\mathrm{rank}}\mathcal{J}_{\mathrm{rank}}, \tag{11}$$

where w_{nat}, w_{pc}, and w_{rank} are loss weights.

After source-only model and checkpoint selection, the selected configuration is refit on all source patients before held-out evaluation. The complete selection, calibration, refitting, ranking-loss definition, and optimization settings are given in Appendix B.2.

4 Theoretical Analysis

In this section, we characterize our mechanism by first analyzing a one-block, two-qubit subsystem exactly in encoded-angle coordinates, then extending the argument to the constructed sparse graph and two-block circuit in Appendices B.8 and B.9. Raw-feature derivatives additionally require the nonsaturated, unclipped regime of the encoding maps. The results establish functional dependencies available to the circuit, not that training must select them or that they imply predictive gains. Appendix B.3 states the complete assumptions and claim boundary.

Consider

$$|\psi\rangle = (\mathscr{R}_1\otimes\mathscr{R}_2) R_{ZZ}(\beta)[R_y(\alpha_1)\otimes R_y(\alpha_2)]|00\rangle, \qquad \mathscr{R}_i = R_y(\omega_i^y) R_x(\omega_i^x), \tag{12}$$

with $\lambda_i^z = \cos\omega_i^x\cos\omega_i^y$, $\lambda_i^x = -\sin\omega_i^y$, and $\lambda_i^y = \sin\omega_i^x\cos\omega_i^y$.

Lemma 4.1 (Edge observability). *For Eq. (12),*

$$\langle Z_1 \rangle = \lambda_1^z \cos\alpha_1 + \lambda_1^x \sin\alpha_1 \cos\beta + \lambda_1^y \sin\alpha_1 \cos\alpha_2 \sin\beta, \tag{13}$$

and, when $\beta = \beta(e_{12})$,

$$\frac{\partial \langle Z_1 \rangle}{\partial e_{12}} = \beta'(e_{12}) \sin\alpha_1 \left[-\lambda_1^x \sin\beta + \lambda_1^y \cos\alpha_2 \cos\beta \right]. \tag{14}$$

Thus the aligned edge is locally observable whenever the displayed factors do not cancel. Without a noncommuting operation after the edge layer, $\lambda_1^x = \lambda_1^y = 0$, so immediate local-$Z$ readout is edge invariant.

Proof sketch. Pulling Z_1 backward through $\mathscr{R}_1$ produces $Z/X/Y$ Pauli components. Conjugating the X/Y terms by R_{ZZ} and evaluating them on the node-encoded product state gives Eq. (13); differentiation gives Eq. (14). Appendix B.4 gives the full calculation, including ZZ readout. The corresponding full-mixer/R_y-only/no-mixer diagnostic for the implemented two-block circuit is reported in Appendix E.

Theorem 4.2 (Mixed node–node–edge term). *If $\lambda_1^y \neq 0$, Eq. (13) contains*

$$\lambda_1^y \sin\alpha_1(v_1) \cos\alpha_2(v_2) \sin\beta(e_{12}). \tag{15}$$

Moreover,

$$\frac{\partial^3 \langle Z_1 \rangle}{\partial\alpha_1 \partial\alpha_2 \partial\beta} = -\lambda_1^y \cos\alpha_1 \sin\alpha_2 \cos\beta, \tag{16}$$

which is nonzero on an open subset of angle space. Under Appendix B.3's nonsaturation assumptions, the corresponding raw-feature mixed derivative is likewise nonzero on open subsets where the feature-to-angle derivatives and the trigonometric factor in Eq. (16) are nonzero.

The edge-conditioned two-body generator therefore induces a measurable term dependent on both incident nodes and their relation. With theory-only linear angle maps, this term has joint Fourier modes in all three coordinates (Schuld et al., 2021); with the implemented nonlinear maps, the finite trigonometric representation is in encoded angles rather than raw EEG features. Node-only and Edge-only VQC each remove one argument of this term, so their comparison with QRG-EEG tests the utility of coupled node–relation information, although a performance gap does not establish that training selected this derivative. Appendix B.5 gives the chain rule and Fourier expansion.

To characterize what trial-specific edge conditioning adds, let $\mathcal{F}_{\text{node}}$ contain circuits whose complete input-dependent path receives $\mathbf{v}$ but not $\mathbf{e}$, and let $\mathcal{F}_{\text{fixed}}$ contain otherwise comparable circuits whose entangling angles do not vary with trial-specific $\mathbf{e}$.

Proposition 4.3 (Restricted edge-conditioned separation). *$\forall g \in \mathcal{F}_{\text{node}} \cup \mathcal{F}_{\text{fixed}}$, $\partial g(\mathbf{v}, \mathbf{e})/\partial e_{ij} = 0$ with $\mathbf{v}$ fixed. The QRG-EEG circuit contains settings and a linear readout for which this derivative is nonzero on an open set; for example, $\omega_1^x = \pi/2$ and $\omega_1^y = 0$ yield $\langle Z_1 \rangle = \sin\alpha_1 \cos\alpha_2 \sin\beta(e_{12})$. Hence QRG-EEG contains edge-varying functions absent from these edge-blind control families.*

This proposition is a coordinate-level existence result. Node-only VQC removes the trial-specific edge path, while Fixed-entanglement VQC retains entanglement but removes its trial dependence. The result neither excludes implicit relational information in learned nodes nor separates QRG-EEG from arbitrary classical models receiving $\mathbf{e}$. Appendix B.6 gives the formal proof.

Finally, enumerate the selected edges and let $\Pi_{\mathcal{E}}$ permute their values while leaving nodes, generators, and readout coordinates fixed.

Corollary 4.4 (Local sensitivity to edge misalignment). *If $g(\mathbf{v}, \mathbf{e}) = g(\mathbf{v}, \Pi_{\mathcal{E}}\mathbf{e})$ throughout an open neighborhood, then necessarily*

$$\nabla_{\mathbf{e}} g(\mathbf{v}, \mathbf{e}) = \Pi_{\mathcal{E}}^{\top} \nabla_{\mathbf{e}} g(\mathbf{v}, \Pi_{\mathcal{E}}\mathbf{e}), \tag{17}$$

throughout that neighborhood. Violating this relation at one point proves that aligned and permuted assignments are not identical local functions.

The identity follows by differentiating the assumed local functional equality and applying the chain rule; its contrapositive gives the stated distinction. Shuffled-edge QRG preserves each trial's edge-value multiset while breaking $e_{ij} \leftrightarrow Z_i Z_j$, so a performance change is consistent with pair identity being relevant. A null result can instead arise from unused conditioning, coincident edge values or sensitivities, or graph/circuit symmetry. Appendix B.7 gives the proof and these exceptions.

5 EXPERIMENTS

We evaluate QRG-EEG under patient-level LOSO to address two research questions: whether graph-conditioned quantum interaction improves cross-patient decoding over embedding-based quantum and matched classical alternatives (**RQ1**), and whether trial-specific Riemannian edges and their component-aligned conditioning provide information beyond local node evidence alone (**RQ2**).

Data and protocol. We evaluate QRG-EEG on four public ALS–EEG cohorts: bigP3BCI Study N and BNCI2014-008 for P300 Target/NonTarget detection, and EEGET-ALS-MI and UCL-ALS-MI for three-class Left/Right/Rest motor imagery (Mainsah et al., 2025; Riccio et al., 2013; Ngo et al., 2024; Patel et al., 2025a). Their 8, 8, 6, and 8 patients define 30 patient-level LOSO folds, with each patient's sessions kept together. Within each fold, all fitted quantities and model selection use source patients only: source validation selects $q \in \{6, 8\}$ and the task representation, and the selected configuration is refit on all source patients. The circuit depth (L=2) and graph multiplier 1.5 stay fixed. Appendices C and B.2 detail preprocessing, nested selection, and implementation settings.

Comparators. We compare against the baselines: tangent-space logistic regression (B1) (Barachant et al., 2012), EEGNet (B2) (Lawhern et al., 2018), xDAWN–LDA (P300) or FBCNet-style decoding (MI) (B3) (Rivet et al., 2009; Mane et al., 2021), QEEGNet-style VQC (B4) (Chen et al., 2024), and xDAWN–BLDA (P300) or Graph-CSPNet-style decoding (MI) (B5) (Rivet et al., 2009; Hoffmann et al., 2008; Ju & Guan, 2024). Same-Graph GAT (B6) (Veličković et al., 2018) receives the same nodes, Riemannian edges, and topology, while Matched Classical Interaction (B7) receives the same aligned node–edge information with approximately matched capacity; both share QRG-EEG's objective and source-validation budget. RQ2 adds Node-only, Fixed-entanglement, Shuffled-edge, Edge-only, and Riemannian-VQC controls to isolate node evidence, trial-specific edges, pair alignment, and explicit graph conditioning. Comparator and intervention definitions are in Appendix C; named neural baselines are fold-compatible adaptations.

Training and evaluation. We use $(w_{\mathrm{nat}}, w_{\mathrm{pc}}, w_{\mathrm{rank}})$=$(0.65, 0.35, 0.08)$ with ranking margin $m = 0.25$. For three fixed seeds $(11, 22, 33)$, we first average scores within each patient-level LOSO fold, then report mean $\pm$ standard deviation across held-out patients, keeping the patient as the unit of analysis. Macro-F1 is the primary metric, supplemented by AUROC for P300 and balanced accuracy for MI. Accuracy, Cohen's κ, task-specific ranking metrics, fold-level results, and paired analyses are given in Appendices D and C. Quantum models use exact differentiable statevector simulation with analytic Z/ZZ expectations; remaining optimization settings are in Appendix B.2.

Table 1: Source-only patient-wise LOSO performance. P300 uses Macro-F1 and AUROC; MI uses Macro-F1 and balanced accuracy (BAcc).

	bigP3BCI-ALS		BNCI2014-008		EEGET-ALS-MI		UCL-ALS-MI	
Model	Macro-F1	AUROC	Macro-F1	AUROC	Macro-F1	BAcc	Macro-F1	BAcc
B1 TS–LR	.417±.065	.520±.026	.462±.080	.553±.039	.331±.077	.405±.045	.354±.143	.420±.074
B2 EEGNet	.475±.006	.614±.064	.616±.053	.791±.032	.344±.106	.406±.060	.326±.050	.394±.070
B3 Task-specific	.475±.004	.585±.060	.591±.051	.776±.040	.293±.119	.391±.091	.308±.032	.360±.023
B4 QEEGNet	.507±.048	.610±.073	.627±.044	.791±.032	.299±.062	.387±.045	.351±.066	.410±.057
B5 Task-specific	.471±.001	.589±.066	.517±.058	.779±.042	.222±.001	.333±.000	.208±.010	.333±.000
B6 Same-Graph GAT	.471±.000	.531±.029	.460±.008	.607±.074	.282±.064	.374±.065	.260±.048	.360±.055
B7 Matched Classical	.471±.001	.593±.057	.618±.034	.789±.034	.404±.069	.455±.071	.335±.063	.391±.059
QRG-EEG	**.550±.044**	**.653±.084**	**.687±.023**	**.807±.030**	**.516±.077**	**.542±.072**	**.470±.082**	**.519±.077**

Values are mean $\pm$ sample standard deviation across held-out patients after averaging three seeds per patient. Bold and underline indicate the best and second-best results, respectively, based on unrounded values. B1: TS–LR (Barachant et al., 2012); B2: EEGNet (Lawhern et al., 2018); B3: xDAWN–LDA / FBCNet (Rivet et al., 2009; Mane et al., 2021); B4: QEEGNet (Chen et al., 2024); B5: xDAWN–BLDA / Graph-CSPNet (Hoffmann et al., 2008; Ju & Guan, 2024); B6: Same-Graph GAT (Veličković et al., 2018); B7: matched classical interaction.

Cross-patient decoding performance (RQ1). Tab. 1 reports the main source-only LOSO comparison, and Fig. 3 shows the between-patient dispersion hidden by dataset averages. QRG-EEG ranks first in Macro-F1 and in the paradigm-appropriate secondary metric on all four datasets. For P300, the joint gain in Macro-F1 and AUROC indicates improvement in both class-balanced decisions and threshold-independent ranking, rather than gains from the majority NonTarget class. For MI, Macro-F1 and balanced accuracy agree, ruling out an ordering driven by a single dominant class. Among the seven main-matrix models, the strongest Macro-F1 comparator differs by task.

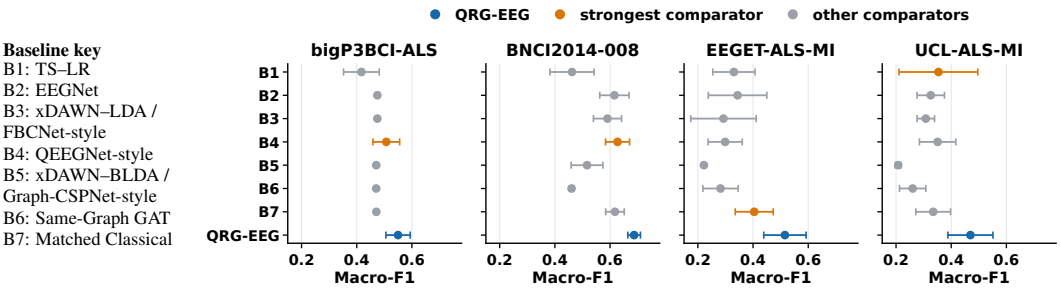

Figure 3: Macro-F1 after averaging seeds within each held-out-patient fold. Points and bars show dataset means and sample standard deviations across patients. Blue denotes QRG-EEG, orange the strongest non-QRG comparator in each dataset, and gray the remaining comparators.

On the P300 datasets it is QEEGNet (B4): the event-locked response provides a stable local signal for a compact encoder, and its VQC adds a nonlinear decision map. QRG-EEG keeps this pathway while conditioning each pairwise interaction on the corresponding trial-specific tangent relation. On EEGET-ALS-MI, Matched Classical Interaction (B7) is strongest, showing that aligned node and edge variables are informative even under classical fusion. On UCL-ALS-MI, tangent-space logistic regression leads the alternatives, consistent with covariance geometry remaining useful under patient and session shift. QRG-EEG combines this geometry with component-wise temporal evidence and thus stays ahead of whichever baseline is strongest in each cohort.

Mechanism-matched controls (RQ1). Fig. 4 tests three alternative explanations for the main ordering. QEEGNet (B4) supplies quantum nonlinearity without pair-aligned Riemannian conditioning; Same-Graph GAT (B6) receives the identical sparse graph and tests whether standard neighborhood aggregation suffices; and Matched Classical Interaction (B7) keeps the aligned node–edge products, training objective, validation budget, and comparable parameter capacity while replacing the circuit. QRG-EEG outperforms all three on every dataset and metric, so neither a VQC alone, graph access alone, nor parameter-matched classical fusion recovers the ordering. The margins also vary by task in an interpretable way. B4 stays more competitive on event-locked P300, whereas B7 narrows the gap more on EEGET-ALS-MI, where relational structure is prominent. QRG-EEG covers both regimes by preserving component identity from the projected signal through the graph and qubits to the readout. This agrees with Section 4: the noncommuting mixer makes edge-conditioned phases observable, and the Z/ZZ readout retains local and pairwise terms, although this evidence is specific to the implemented circuit. Appendix I extends the EEGET-ALS-MI comparison to EEG Conformer-style and EA+EEGNet across all six patient folds.

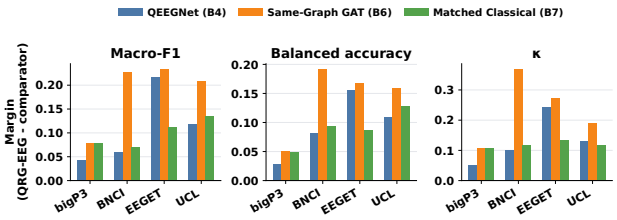

Figure 4: **RQ1 controlled comparisons.** QRG-EEG margins over QEEGNet-style (B4), Same-Graph GAT (B6), and Matched Classical Interaction (B7) from seed-averaged patient folds; positive margins favor QRG-EEG.

Riemannian-edge controls (RQ2). Tab. 2 tests whether performance depends on both information sources and their assignment. CRGNet is the strongest relational alternative, confirming that learned local evidence and Riemannian features are useful together before the interaction operator is considered. QRG-EEG remains first, so the ordering is not explained by a second feature branch alone. Node-only removes trial-specific relations; edge-only removes

Table 2: RQ2 interventions (evaluated by Macro-F1).

Model/control	bigP3BCI-ALS	BNCI2014-008	EEGET-ALS-MI	UCL-ALS-MI
Riemannian-VQC	.471±.000	.455±.001	.272±.033	.335±.035
CRGNet	.515±.050	.634±.042	.459±.061	.385±.104
Node-only VQC	.471±.000	.615±.039	.264±.063	.321±.049
Fixed-entanglement VQC	.471±.001	.623±.038	.292±.051	.318±.046
Shuffled-edge QRG	.471±.000	.615±.036	.274±.056	.334±.047
Edge-only VQC	.471±.000	.455±.000	.244±.034	.284±.043
PLV-GAT (UCL only)	–	–	–	.215±.017
QRG-EEG	**.550±.044**	**.687±.023**	**.516±.077**	**.470±.082**

Aggregation follows Tab. 1. Riemannian-VQC: vectorized geometric-quantum setting (Andreev et al., 2023; Cattan & Quemy, 2023). CRGNet (Gao et al., 2022): adapted convolutional–Riemannian fusion. PLV-GAT (Patel et al., 2025b): UCL cohort only. Remaining rows: controlled interventions (Appendix C).

the local temporal evidence required to interpret them. Fixed entanglement retains two-qubit operations but makes their angles trial independent. Shuffling preserves each trial's edge-value multiset while breaking the edge-to-pair map, and Riemannian-VQC keeps the geometry but vectorizes it before the circuit. Their ordering supports the joint use of node evidence, trial-specific edges, and pair identity. The cohort-compatible PLV-GAT result supplies an additional published graph comparison on UCL-ALS-MI; it is not extrapolated to incompatible cohorts.

Fig. 5 summarizes the RQ2 controls as Macro-F1 margins relative to QRG-EEG. All margins are positive but vary across datasets, showing task-dependent contributions from local node evidence, trial-specific edges, and pair alignment. On the MI datasets, removing either information source or disrupting their structured interaction produces substantial deficits. Together, these patterns support a narrower conclusion: the useful Riemannian signal is relational, and its contribution depends on combining it with local evidence while preserving which latent components each edge connects.

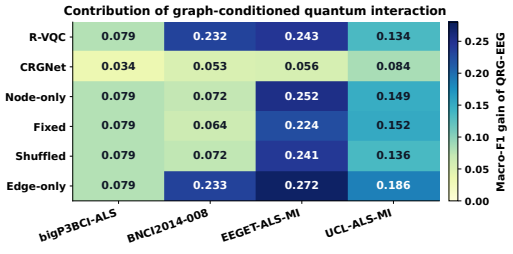

Figure 5: Macro-F1 margin of QRG-EEG over each RQ2 control, by dataset; larger values indicate greater loss when that constraint is removed.

Interaction-design diagnostics. Additional diagnostics support the main design choices. The full R_x–R_y mixer outperforms the R_y-only and no-mixer variants, consistent with the post-edge non-commuting mechanism analyzed in Sec. 4. Source-side preferences for $q \in \{6, 8\}$ vary across datasets, while $L{=}2$ consistently exceeds $L{=}4$, supporting selection between compact widths and the shallow primary circuit (Appendix E, F). Finally, QRG-EEG remains ahead of random-Fourier-feature, and RBF-SVM controls given the same frozen node–edge representation, indicating that these classical surrogates do not reproduce the observed ordering under the evaluated settings (Appendix G). Appendix H ablates each objective component and examines local sensitivity to the loss coefficients, ranking margin, and graph density on EEGET-ALS-MI.

6 Conclusion

Cross-patient ALS–EEG decoding requires learning from scarce patient data while capturing both local EEG evidence and trial-specific relational structure. We introduced QRG-EEG to address this problem by preserving EEG component identity and mapping local node evidence and pairwise Riemannian relations to aligned single- and two-qubit operations. Our theoretical analysis characterizes how edge-conditioned interactions become observable and couple information from the corresponding nodes, while controlled experiments across four ALS–EEG datasets show consistent gains over embedding-based quantum and matched classical alternatives. Targeted interventions further demonstrate that the observed improvements depend on both trial-specific relational information and its correct alignment with local node evidence. Together, these results support component-aligned node–edge conditioning as a structured inductive bias for data-limited cross-patient EEG decoding, while keeping the claim focused on the predictive and representational role of the proposed interaction mechanism.

Limitations and Future Work

Our conclusions are bounded by the scope of the questions studied here. QRG-EEG provides evidence for component-aligned node–edge interaction in data-limited, cross-patient ALS–EEG decoding across P300 and MI; we do not claim that the current Riemannian representation, sparse graph, or latent decomposition is optimal for general EEG learning. Likewise, our exact statevector experiments evaluate a structured quantum interaction mechanism rather than quantum computational advantage, hardware efficiency, or noise robustness. The current results constitute target-isolated development evidence under patient-level LOSO, rather than independent clinical validation or evidence of online BCI performance. Future work should therefore test the frozen protocol on untouched multicenter ALS cohorts, extend the alignment principle to alternative connectivity measures and adaptive graphs, evaluate longitudinal and online BCI settings, and examine finite-shot and noisy quantum implementations against strong classical surrogates.

REPRODUCIBILITY STATEMENT

Section 3 defines the final patient-level LOSO boundary, and Appendix B.2 specifies fold-local preprocessing, source-only candidate screening, validation, calibration, all-source refitting, optimization constants, and averaging of seeds 11, 22, and 33. Within every final run, the held-out patient is used only for final scoring. Table 9 identifies which comparators are fold-compatible reimplementations, and Table 5 defines the controlled interventions. Appendices B.3–B.9 provide the assumptions, proofs, and graph/depth extensions. The anonymous supplementary package contains implementations of QRG-EEG and the reported comparisons, dataset acquisition and schema-audit routines, configuration and environment records, fixed seed lists, and scripts for regenerating fold–seed outputs, tables, and figures. Quantum models use an exact analytic PyTorch statevector simulator without finite-shot sampling or a hardware-noise model; the recorded software environment is reported in Table 8.

AI USE STATEMENT

We used OpenAI's ChatGPT to assist with language editing, manuscript organization, and LaTeX formatting. The authors verified all scientific claims, analyses, results, citations, and final text.

REFERENCES

Hamdi Altaheri, Ghulam Muhammad, and Mansour Alsulaiman. Physics-informed attention temporal convolutional network for EEG-based motor imagery classification. *IEEE Transactions on Industrial Informatics*, 19(2):2249–2258, 2023. doi: 10.1109/TII.2022.3197419. URL https://doi.org/10.1109/TII.2022.3197419.

Anton Andreev, Grégoire Cattan, Sylvain Chevallier, and Quentin Barthélemy. pyRiemann-qiskit: A sandbox for quantum classification experiments with riemannian geometry. *Research Ideas and Outcomes*, 9:e101006, 2023. doi: 10.3897/rio.9.e101006. URL https://doi.org/10.3897/rio.9.e101006.

Alexandre Barachant and Marco Congedo. A Plug&Play P300 BCI using information geometry. *arXiv preprint arXiv:1409.0107*, 2014. doi: 10.48550/arXiv.1409.0107. URL https://arxiv.org/abs/1409.0107.

Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012. doi: 10.1109/TBME.2011.2172210. URL https://doi.org/10.1109/TBME.2011.2172210.

Grégoire H Cattan and Alexandre Quemy. Case-based and quantum classification for erp-based brain–computer interfaces. *Brain Sciences*, 13(2):303, 2023.

M. Cerezo, Martín Larocca, Diego García-Martín, N. L. Diaz, Paolo Braccia, Enrico Fontana, Manuel S. Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, Eric R. Anschuetz, and Zo"e Holmes. Does provable absence of barren plateaus imply classical simulability? *Nature Communications*, 16:7907, 2025. doi: 10.1038/s41467-025-63099-6. URL https://doi.org/10.1038/s41467-025-63099-6.

Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

Ujwal Chaudhary, Niels Birbaumer, and Ander Ramos-Murguialday. Brain–computer interfaces for communication and rehabilitation. *Nature Reviews Neurology*, 12(9):513–525, 2016. doi: 10.1038/nrneurol.2016.113. URL https://doi.org/10.1038/nrneurol.2016.113.

Chi-Sheng Chen, Samuel Yen-Chi Chen, Aidan Hung-Wen Tsai, and Chun-Shu Wei. QEEGNet: Quantum machine learning for enhanced electroencephalography encoding. In *2024 IEEE Workshop on Signal Processing Systems (SiPS)*, pp. 153–158, 2024. doi: 10.1109/SiPS62058.2024.00035. URL https://doi.org/10.1109/SiPS62058.2024.00035.

Yilun Chen, Ami Wiesel, Yonina C. Eldar, and Alfred O. Hero. Shrinkage algorithms for mmse covariance estimation. *IEEE Transactions on Signal Processing*, 58(10):5016–5029, 2010. doi: 10.1109/TSP.2010.2053029.

Marco Congedo, Alexandre Barachant, and Rajendra Bhatia. Riemannian geometry for EEG-based brain–computer interfaces; a primer and a review. *Brain-Computer Interfaces*, 4(3):155–174, 2017. doi: 10.1080/2326263X.2017.1297192. URL https://doi.org/10.1080/2326263X.2017.1297192.

Yi Ding, Neethu Robinson, Chengxuan Tong, Qiuhao Zeng, and Cuntai Guan. LGGNet: Learning from local–global–graph representations for brain–computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. doi: 10.1109/TNNLS.2023.3236635. URL https://doi.org/10.1109/TNNLS.2023.3236635. Early access.

Chang Gao, Wenchao Liu, and Xian Yang. Convolutional neural network and riemannian geometry hybrid approach for motor imagery classification. *Neurocomputing*, 507:180–190, 2022. doi: 10.1016/j.neucom.2022.08.024. URL https://doi.org/10.1016/j.neucom.2022.08.024.

Andrew Geronimo, Zachary Simmons, and Steven J. Schiff. Performance predictors of brain–computer interfaces in patients with amyotrophic lateral sclerosis. *Journal of Neural Engineering*, 13(2):026002, 2016. doi: 10.1088/1741-2560/13/2/026002. URL https://doi.org/10.1088/1741-2560/13/2/026002.

He He and Dongrui Wu. Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2020. doi: 10.1109/TBME.2019.2913914. URL https://doi.org/10.1109/TBME.2019.2913914.

Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi, and Karin Diserens. An efficient P300-based brain–computer interface for disabled subjects. *Journal of Neuroscience Methods*, 167(1):115–125, 2008. doi: 10.1016/j.jneumeth.2007.03.005.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous EEG data in BCI. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=QzTpTRVtrP. Spotlight.

Ce Ju and Cuntai Guan. Tensor-CSPNet: A novel geometric deep learning framework for motor imagery classification. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. doi: 10.1109/TNNLS.2022.3172108. URL https://doi.org/10.1109/TNNLS.2022.3172108. Early access.

Ce Ju and Cuntai Guan. Graph neural networks on SPD manifolds for motor imagery classification: A perspective from the time–frequency analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 35(12):17701–17715, 2024. doi: 10.1109/TNNLS.2023.3307470. URL https://doi.org/10.1109/TNNLS.2023.3307470.

Bruna Junqueira, Bruno Aristimunha, Sylvain Chevallier, and Raphael Y. de Camargo. A systematic evaluation of euclidean alignment with deep learning for EEG decoding. *arXiv preprint arXiv:2401.10746*, 2024. doi: 10.48550/arXiv.2401.10746. URL https://arxiv.org/abs/2401.10746.

Reinmar J. Kobler, Jun-ichiro Hirayama, Qibin Zhao, and Motoaki Kawanabe. SPD domain-specific batch normalization to crack interpretable unsupervised domain adaptation in EEG. In *Advances in Neural Information Processing Systems*, volume 35, 2022. doi: 10.52202/068431-0450. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/28ef7ee7cd3e03093acc39e1272411b7-Abstract-Conference.html.

Jonas Landman, Slimane Thabet, Constantin Dalyac, Hela Mhiri, and Elham Kashefi. Classically approximating variational quantum machine learning with random fourier features. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=ymFhZxw70uz.

Martín Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J. Coles, Lukasz Cincio, Jarrod R. McClean, Zo"e Holmes, and M. Cerezo. Barren plateaus in variational quantum computing. *Nature Reviews Physics*, 7:174–189, 2025. doi: 10.1038/s42254-025-00813-9. URL https://doi.org/10.1038/s42254-025-00813-9.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018. doi: 10.1088/1741-2552/aace8c. URL https://doi.org/10.1088/1741-2552/aace8c.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 3538–3545, 2018. doi: 10.1609/aaai.v32i1.11604. URL https://doi.org/10.1609/aaai.v32i1.11604.

Boyla Mainsah, Chance Fleeting, Thomas Balmat, Eric Sellers, and Leslie Collins. bigP3BCI: An open, diverse and machine learning ready P300-based brain–computer interface dataset. *PhysioNet*, May 2025. doi: 10.13026/0byy-ry86. URL https://doi.org/10.13026/0byy-ry86. Version 1.0.0.

Ravikiran Mane, Effie Chew, Karen Chua, Kai Keng Ang, Neethu Robinson, A. P. Vinod, Seong-Whan Lee, and Cuntai Guan. FBCNet: A multi-view convolutional neural network for brain–computer interface. *arXiv preprint arXiv:2104.01233*, 2021. doi: 10.48550/arXiv.2104.01233. URL https://arxiv.org/abs/2104.01233.

Peter Mernyei, Konstantinos Meichanetzidis, and Ismail Ilkan Ceylan. Equivariant quantum graph circuits. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 15401–15420. PMLR, 2022. URL https://proceedings.mlr.press/v162/mernyei22a.html.

Thi Duyen Ngo, Hai Dang Kieu, Minh Hoa Nguyen, The Hoang-Anh Nguyen, Van Mao Can, Ba Hung Nguyen, and Thanh Ha Le. An EEG & eye-tracking dataset of ALS patients & healthy people during eye-tracking-based spelling system usage. *Scientific Data*, 11:664, 2024. doi: 10.1038/s41597-024-03501-y. URL https://doi.org/10.1038/s41597-024-03501-y.

Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. MAtt: A manifold attention network for EEG decoding. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/c981fd12b1d5703f19bd8289da9fc996-Abstract-Conference.html.

Rishan Patel, Dai Jiang, Barney Bryson, Tom Carlson, Andreas Demosthenous, and Andrew Geronimo. Longitudinal ALS EEG dataset for motor imagery studies, 2025a. URL https://doi.org/10.5522/04/28156016.

Rishan Patel, Ziyue Zhu, Barney Bryson, Tom Carlson, Dai Jiang, and Andreas Demosthenous. Advancing EEG classification for neurodegenerative conditions using BCI: A graph attention approach with phase synchrony. *Neuroelectronics*, 2(1):0001, 2025b. doi: 10.55092/neuroelectronics20250001. URL https://doi.org/10.55092/neuroelectronics20250001.

Adrià Pérez-Salinas, Alba Cervera-Lierta, Elies Gil-Fuster, and José I. Latorre. Data re-uploading for a universal quantum classifier. *Quantum*, 4:226, 2020. doi: 10.22331/q-2020-02-06-226. URL https://doi.org/10.22331/q-2020-02-06-226.

Herbert Ramoser, Johannes Müller-Gerking, and Gert Pfurtscheller. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, 8(4):441–446, 2000. doi: 10.1109/86.895946.

Angela Riccio, Laura Simione, Francesca Schettini, Antonio Pizzimenti, Maurizio Inghilleri, Maria O. Belardinelli, Donatella Mattia, and Febo Cincotti. Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. *Frontiers in Human Neuroscience*, 7:732, 2013. doi: 10.3389/fnhum.2013.00732. URL https://doi.org/10.3389/fnhum.2013.00732.

Bertrand Rivet, Antoine Souloumiac, Virginie Attina, and Guillaume Gibert. xdawn algorithm to enhance evoked potentials: application to brain–computer interface. *IEEE Transactions on Biomedical Engineering*, 56(8):2035–2043, 2009. doi: 10.1109/TBME.2009.2012869.

Franz J. Schreiber, Jens Eisert, and Johannes Jakob Meyer. Classical surrogates for quantum learning models. *Physical Review Letters*, 131(10):100803, 2023. doi: 10.1103/PhysRevLett.131.100803. URL https://doi.org/10.1103/PhysRevLett.131.100803.

Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, 2019. doi: 10.1103/PhysRevA.99.032331. URL https://doi.org/10.1103/PhysRevA.99.032331.

Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3):032430, 2021. doi: 10.1103/PhysRevA.103.032430.

Andrea Skolik, Michele Cattelan, Sheir Yarkoni, Thomas B"ack, and Vedran Dunjko. Equivariant quantum circuits for learning on weighted graphs. *npj Quantum Information*, 9: 47, 2023. doi: 10.1038/s41534-023-00710-y. URL https://doi.org/10.1038/s41534-023-00710-y.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. EEG Conformer: Convolutional transformer for EEG decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2023. doi: 10.1109/TNSRE.2022.3230250. URL https://doi.org/10.1109/TNSRE.2022.3230250.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Guillaume Verdon, Trevor McCourt, Enxhell Luzhnica, Vikash Singh, Stefan Leichenauer, and Jack Hidary. Quantum graph neural networks. *arXiv preprint arXiv:1909.12264*, 2019. doi: 10.48550/arXiv.1909.12264. URL https://arxiv.org/abs/1909.12264.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A criss-cross brain foundation model for EEG decoding. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=NPNUHgHF2w.

Florian Yger, Marie Berar, and Fabien Lotte. Riemannian approaches in brain–computer interfaces: A review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25 (10):1753–1762, 2017. doi: 10.1109/TNSRE.2016.2627016. URL https://doi.org/10.1109/TNSRE.2016.2627016.

Paolo Zanini, Marco Congedo, Christian Jutten, Salem Said, and Yannick Berthoumieu. Transfer learning: A riemannian geometry framework with applications to brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, 2018. doi: 10.1109/TBME.2017.2742541. URL https://doi.org/10.1109/TBME.2017.2742541.

Peixiang Zhong, Di Wang, and Chunyan Miao. EEG-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, 13(3):1290–1301, 2022. doi: 10.1109/TAFFC.2020.2994159. URL https://doi.org/10.1109/TAFFC.2020.2994159.

A Supplementary Overview

Appendix B specifies the notation, executable model, assumptions, and proofs. Appendix C defines the cohorts, comparison models, source-only selection procedure, runtime environment, and patient-level estimands. Appendix D reports secondary metrics and paired patient-fold distributions. Appendices E and F examine the mixer, qubit count, and circuit depth. Appendix G evaluates classical functions on the same frozen node–edge representation. Appendix H reports objective and structural sensitivity, Appendix I compares a modern EEG architecture and a cross-subject alignment method, and Appendix J provides the extended related-work discussion.

B Method and Theory Details

B.1 Notation ledger

Tab. 3 fixes the correspondence used by the architecture, circuit equations, and proofs. In particular, b indexes a trial, i and j identify aligned latent components, graph nodes, and qubits, and $\mathcal{E}$ denotes the source-fitted edge positions. Suppressing b in the proofs shortens the expressions but does not make the node or edge values trial independent. Bold symbols denote trial-level vectors or matrices, whereas scalar symbols denote one component, edge, angle, or measurement coordinate. The encoded angles $\alpha_{b,i}$ and $\beta_{b,ij}$ remain distinct from their underlying node and edge values, $v_{b,i}$ and $e_{b,ij}$, throughout the analysis. The node bias b_v^{node} and ranking log-odds $\xi_{n,k}$ have separate symbols.

Table 3: Notation used across the architecture figures, main method, and theoretical analysis. Bold capitals denote matrix- or tensor-valued quantities, bold lowercase symbols denote vectors, and unbold symbols denote components, edges, angles, or scalars. The proof sections suppress trial index b for readability.

Symbol	Type	Meaning
$\mathbf{X}$, $\mathbf{X}_b$	$\mathbb{R}^{B\times C\times T}$, $\mathbb{R}^{C\times T}$	EEG trial batch and its bth trial
$\mathbf{P}$	$\mathbb{R}^{C\times q}$	Source-fitted spatial map
$\widetilde{\mathbf{X}}$, $\widetilde{\mathbf{X}}_b$	$\mathbb{R}^{B\times q\times T}$, $\mathbb{R}^{q\times T}$	Aligned latent batch and its bth trial
$\mathbf{d}_{b,i}$, $v_{b,i}$, $\widetilde{v}_{b,i}$	$\mathbb{R}^p$, scalar, scalar	Deep component feature, node value, and source-normalized node value
b_v^{node}	scalar	Bias of the shared scalar node projection
$\mathbf{\Sigma}_b$, $\overline{\mathbf{\Sigma}}_{\mathrm{src}}$, $\mathbf{H}_b$	SPD, SPD, symmetric	Trial covariance, source Riemannian reference, and tangent matrix
$e_{b,ij}$, $\widetilde{e}_{b,ij}$, $\mathcal{E}$	scalar, scalar, edge set	Trial relation, normalized relation, and source-fitted sparse topology
$\mathcal{G}_b$	attributed graph	$(\mathcal{V}, \mathcal{E}; \widetilde{\mathbf{v}}_b, \widetilde{\mathbf{e}}_b)$
$\alpha_{b,i}$, $\beta_{b,ij}$	angles	Data-conditioned node and pair-aligned edge angles
U_N, U_E, $U_M^{(\ell)}$	unitaries	Shared node encoding, shared edge interaction, and block-specific mixer
$\mathbf{m}_b$, $\mathbf{M}$	$\mathbb{R}^{q+\lvert\mathcal{E}\rvert}$, $\mathbb{R}^{B\times(q+\lvert\mathcal{E}\rvert)}$	Trial measurement vector and its batch stack
$\mathbf{O}$, $\widehat{\mathbf{Y}}$	$\mathbb{R}^{B\times K}$, $\mathbb{R}^{B\times K}$	Batch logits and row-wise class probabilities
$\xi_{n,k}$	scalar	One-vs-rest log-odds for trial n and class k
$\mathcal{J}$	scalar	Declared subject-aware training objective; distinct from block count L
$\mathscr{R}_i$, $\lambda_i^{z/x/y}$	unitary, scalars	Local mixer in the two-qubit proof and its Pauli-basis coefficients
$\Pi_{\mathcal{E}}$	permutation	Edge-coordinate permutation used only in the alignment analysis

B.2 Implementation details

B.2.1 Fold-local preprocessing and representation fitting

All fitted quantities are reconstructed inside each source split, including preprocessing statistics, the spatial projection $\mathbf{P}$, node/edge scalers, covariance references, sparse graph topology, representation and q selection, checkpoints, training duration, and decision calibration. Filtering and epoching use fixed task rules and do not depend on labels from the held-out patient. The reported implementation resamples every dataset to 128 Hz, uses 0.5–20 Hz and 0–0.8 s for P300, and uses 4–40 Hz for MI. The UCL and EEGET MI windows contain 4 s and 2 s, respectively. A fixed $120\,\mu$V peak gate is applied only to datasets whose stored values have a validated voltage interpretation. EEGET follows a predeclared, label-blind dataset-level bypass because its stored units do not support that fixed gate.

For a source-training array $\mathbf{X}_{\mathrm{src}}$, the component scaler first removes each trial's temporal mean and divides component i by a source-fitted robust scale. It does not impose trial-wise unit variance, which would remove trial-level amplitude and band-power differences. Independent copies of this source-fitted transform feed the node and covariance branches. The copies share the projection $\mathbf{P}$, so scaling does not change the identity relation

$$\begin{aligned}\text{latent component } i &= \text{deep node } i \\ &= \text{covariance row/column } i \\ &= \text{graph node } i = \text{qubit } i.\end{aligned} \tag{18}$$

For multiclass MI, the CSP candidate is fitted one class versus the remaining classes. Candidate filters are ranked by source-only Fisher separation of log-variance and orthogonalized before retaining q filters. The PCA candidate retains the leading eigenvectors of the mean source covariance. For P300, xDAWN solves a source-only generalized eigenproblem between target-ERP and overall covariance and retains q orthonormal filters. Column signs are fixed deterministically.

For the Riemannian branch, MI uses the projected components directly. P300 source-only validation compares a plain geometric input $\breve{\mathbf{X}}_b = \widetilde{\mathbf{X}}_b$ with an ERP-augmented candidate that concatenates a source-derived target-ERP residual along the temporal dimension,

$$\breve{\mathbf{X}}_b = \left[\widetilde{\mathbf{X}}_b, \widetilde{\mathbf{X}}_b - \overline{\mathbf{X}}_{\mathrm{ERP,src}}\right]. \tag{19}$$

The ERP template is estimated exclusively from source patients in the current outer fold; the construction preserves the q latent-component identities.

Given a projected covariance $\mathbf{\Sigma}_b$, the Riemannian mean is computed by the fixed-point iteration

$$\Delta_r = \frac{1}{N}\sum_{n=1}^{N} \log\left(M_r^{-1/2}\mathbf{\Sigma}_n M_r^{-1/2}\right), \tag{20}$$

$$M_{r+1} = M_r^{1/2}\exp(\Delta_r)M_r^{1/2}, \tag{21}$$

until $\|\Delta_r\|_F < 10^{-8}$ or 30 iterations have elapsed. Every eigenvalue used by square root, inverse square root, or logarithm is lower bounded for numerical stability. The full tangent vector used by geometric baselines takes the upper triangle of $\mathbf{H}_b$ and multiplies off-diagonal entries by $\sqrt{2}$; QRG edges use the corresponding entries $[\mathbf{H}_b]_{ij}$ directly.

Node values and tangent-edge values are standardized with source-fitted robust statistics and clipped to $[-3, 3]$. For each candidate pair, the source-only edge importance is

$$\gamma_{ij} = \frac{1}{N_{\mathrm{src}}}\sum_{r=1}^{N_{\mathrm{src}}} |[\mathbf{H}_r]_{ij}|. \tag{22}$$

A maximum-weight spanning tree first ensures connectivity, after which the strongest remaining edges are added until $|\mathcal{E}| = \mathrm{round}(1.5q)$. Thus $\mathcal{E}$ is source-fitted and fixed within an outer fold, whereas $e_{b,ij} = [\mathbf{H}_b]_{ij}$ remains trial-specific.

Table 4: Frozen QRG-EEG settings used by the full model protocol.

Item	Setting
Primary circuit	the QRG-EEG circuit; $L = 2$; graph multiplier 1.5
Source-selected choices	$q \in \{6, 8\}$ and task representation
Graph size	$\lvert\mathcal{E}\rvert = \mathrm{round}(1.5q)$
Node/edge bounds	$\alpha_{\max} = \pi$, $\beta_{\max} = \pi/2$
Node feature width	$p = 12$; 12 temporal filters
Dropout / robust clip	0.25 / 3.0
Mixer initialization	$\omega^x \sim \pi/4$, $\omega^y \sim 0.15$ with s.d. 0.05
Maximum stage budgets	pretrain 60, circuit 60, fine-tune 30 epochs
Stage-2 learning rates	circuit 8×10^{-4}; readout 1.5×10^{-3}
Fine-tuning learning rates	backbone 8×10^{-5}; circuit 3×10^{-4}; readout 8×10^{-4}
Optimizer / regularization	AdamW; weight decay 10^{-4}; gradient norm 5
Schedule / averaging	deterministic cosine; EMA decay 0.99
QRG minibatch	64 on GPU (32 on CPU)
Loss weights	$w_{\mathrm{nat}} = 0.65$; $w_{\mathrm{pc}} = 0.35$; $w_{\mathrm{rank}} = 0.08$
Ranking margin	$m = 0.25$
Seeds	11, 22, 33

B.2.2 Fixed primary architecture and optimization

Tab. 4 collects the frozen architecture and optimization settings used in every primary QRG-EEG run. Only the qubit count $q \in \{6, 8\}$ and task representation are selected from source patients within each outer LOSO fold.

The shared temporal encoder consists of a length-31 temporal convolution, average pooling, a length-15 depthwise convolution, pointwise projection, a second length-7 depthwise-separable block, and adaptive temporal pooling. The temporary pretraining head consumes the same scalar vector $(v_1, \ldots, v_q)$ later used by the circuit; it does not pretrain an unrelated flattened feature. During circuit training the complete node encoder is frozen. During fine-tuning, the first temporal block stays frozen while the second block and scalar projection are updated.

The implemented two-block circuit uses the bounded maps $\alpha_{b,i} = \pi \tanh(\theta^N_i \widetilde{v}_{b,i})$ and $\beta_{b,ij} = (\pi/2)\tanh(\theta^E_{ij}\widetilde{e}_{b,ij})$. The node and edge scales θ^N, θ^E are trainable and shared across the two blocks, whereas mixer angles are block-specific. The code executes R_x followed by R_y on each qubit, so the corresponding operator product is $R_y R_x$.

The population expressions describe the objective. In a stratified minibatch, the implementation applies inverse sampling weights only inside $\mathcal{J}_{\mathrm{nat}}$ to recover each patient's natural trial distribution. It forms $\mathcal{J}_{\mathrm{pc}}$ explicitly from the observed patient–class cells and therefore does not apply a second inverse class multiplier. For binary P300, the ranking term uses class 1 once because the two one-vs-rest rankings are symmetric; multiclass MI averages all one-vs-rest classes. If a minibatch lacks a positive–negative pair within a patient, that patient/class contributes no ranking term. Cross-patient pairs are never substituted because they could reward subject identity.

Using $\xi_{n,k} = o_{n,k} - \log \sum_{a \neq k} \exp(o_{n,a})$, let $\mathcal{I}^+_{s,k} = \mathcal{I}_{s,k}$, $\mathcal{I}^-_{s,k} = \mathcal{I}_s \setminus \mathcal{I}_{s,k}$, and let $\mathcal{K}^\star_s$ contain only the one-vs-rest classes having both sets nonempty (class 1 only for binary P300; every eligible class for multiclass MI). The implemented nested reduction is

$$\mathcal{J}_{\mathrm{rank}} = \frac{1}{|\mathcal{S}^\star_{\mathcal{I}}|} \sum_{s \in \mathcal{S}^\star_{\mathcal{I}}} \frac{1}{|\mathcal{K}^\star_s|} \sum_{k \in \mathcal{K}^\star_s} \frac{1}{|\mathcal{I}^+_{s,k}|\,|\mathcal{I}^-_{s,k}|} \sum_{n^+ \in \mathcal{I}^+_{s,k}} \sum_{n^- \in \mathcal{I}^-_{s,k}} \mathrm{softplus}\big(m - \xi_{n^+,k} + \xi_{n^-,k}\big), \quad (23)$$

where $\mathcal{S}^\star_{\mathcal{I}}$ contains subjects with at least one eligible class. This ordering gives equal weight to eligible classes within a subject and then to subjects, irrespective of pair count.

B.2.3 Source-only selection, calibration, and final refit

For outer test patient $s_\star$, the candidate set is

$$\mathcal{A}_{s_\star} = \{(q, \rho) : q \in \{6, 8\},\ \rho \in \mathcal{R}_{\mathrm{task}}\}, \quad (24)$$

subject to $q \leq C$. The task-specific representation sets are

$$\mathcal{R}_{\mathrm{MI}} = \{(\mathrm{CSP}, \mathrm{plain}), (\mathrm{PCA}, \mathrm{plain})\}, \tag{25}$$

$$\mathcal{R}_{\mathrm{P300}} = \{(\mathrm{xDAWN}, \mathrm{plain}), (\mathrm{xDAWN}, \mathrm{ERP\text{-}residual})\}. \tag{26}$$

One deterministically chosen source patient and budgets $12/12/6$ epochs produce a proxy rank over every declared candidate. Candidates within 0.01 of the best proxy Macro-F1 are ordered by a geometric joint score, minimum class F1, Macro-F1, and secondary measures; the top two are shortlisted. A proxy failure invalidates the outer fold rather than silently reducing the grid.

Both shortlisted candidates are then retrained at the full budget. We use three fixed source-validation patients in each non-EEGET fold and all five available source patients in each EEGET fold. Validation patients have equal weight. For a metric vector z containing Macro-F1, accuracy, balanced accuracy, AUROC, transformed kappa $\widetilde{\kappa} = \mathrm{clip}((\kappa + 1)/2, 0, 1)$, and P300 AUPRC when defined, the joint score is

$$\mathrm{GM}(z) = \exp\left(\frac{1}{|z|}\sum_{a \in z} \log \mathrm{clip}(a, 10^{-6}, 1)\right). \tag{27}$$

We first retain candidates whose mean raw Macro-F1 is within 0.01 of the best. We then prioritize, in order, worst-patient GM, mean GM, worst-patient Macro-F1, mean minimum-class F1, mean Macro-F1, secondary score, lower Macro-F1 variability, and smaller q. All predeclared validation patients must complete for a candidate to remain eligible.

Only after fixing the structure and raw-probability checkpoint do we calibrate decisions from pooled out-of-fold source-patient predictions. P300 uses one source-patient-balanced threshold. MI applies

$$\widehat{p}_{nk}^{(\tau)} \propto \exp(\log \widehat{p}_{nk} - \tau \log \pi_k^{\mathrm{src}}), \tag{28}$$

where π_k^{src} averages the class proportion over source patients and $\tau \in \{-0.5, -0.375, \ldots, 0.5\}$ is selected from source validation. Threshold/prior candidates must remain within the declared tolerances of the uncalibrated decision on accuracy, balanced accuracy, and kappa. The chosen stage lengths are the medians across source-validation patients, with minimum stage floors of 5, 6, and 3 epochs. Final training replays the corresponding prefix of the same cosine schedule on all source patients.

For every dataset, model, and held-out patient, we average seeds 11, 22, and 33. Dataset summaries then report the mean and sample standard deviation over held-out patients. Seeds are optimization replicates, not independent patients, and no best-seed selection is performed.

Table 5: Controls used to interpret the node–edge interaction.

Control	Intervention relative to QRG-EEG
Node-only VQC	Keeps the learned node encoder and local rotations; removes every two-qubit gate and every ZZ readout coordinate.
Fixed-entanglement VQC	Replaces the source-derived graph by a ring and feeds constant edge values, yielding trainable but trial-independent entangling angles.
Shuffled-edge QRG	Keeps the learned graph and per-trial edge multiset but applies a fixed one-position cyclic derangement before the edge gates.
Edge-only VQC	Never executes the node encoder; supplies constant-one node values while retaining trial-specific Riemannian edges.
Riemannian-VQC	Vectorizes the tangent matrix, compresses it to q values, and applies a generic VQC instead of pair-aligned edge operations.
Same-Graph GAT	Uses the same node encoder, edge values, topology, loss, source-validation budget, and approximately the same learned parameter count, but replaces the circuit by graph attention.
Matched classical interaction	Uses the same aligned node/edge inputs and budget in a parameter-matched classical interaction layer.

B.2.4 Exact controlled interventions

Tab. 5 operationally defines the mechanism controls used across RQ1 and RQ2, specifying which node, edge, topology, interaction, and readout pathways are retained, removed, replaced, or misaligned relative to QRG-EEG. For Same-Graph GAT and Matched Classical Interaction, the hidden

width is chosen to minimize the absolute relative gap from QRG-EEG's total learned parameter count under the same fold configuration. A gap above 10% invalidates the match rather than being reported as capacity controlled.

B.3 Formal assumptions and claim boundary

The proofs use the following assumptions.

Assumption B.1 (Implemented unitary convention)**.** Gates execute from left to right in code but multiply state vectors from the left. Executing $R_x(\omega_i^x)$ and then $R_y(\omega_i^y)$ therefore gives $\mathscr{R}_i = R_y(\omega_i^y)R_x(\omega_i^x)$.

Assumption B.2 (Local feature coordinates)**.** For mixed-derivative statements, (v_i, v_j, e_{ij}) are treated as separately intervenable coordinates. Source-fitted normalization parameters and all other coordinates are held fixed. This does not assert statistical independence of features computed from the same EEG trial.

Assumption B.3 (Nonsaturated interior)**.** Raw-feature derivatives are evaluated away from robust-clipping boundaries, with nonzero feature scales and trainable encoding scales. The encoded-angle identities themselves hold for every real angle.

Assumption B.4 (Controlled edge blindness)**.** The Node-only and Fixed-entanglement classes receive no trial-specific e_{ij} through preprocessing, an auxiliary branch, calibration, or the readout. Their trainable parameters are fixed across trials at evaluation.

Assumption B.5 (Alignment intervention)**.** Edge shuffling changes edge values assigned to generators while keeping node features, qubit identities, generator locations, and readout order fixed. A simultaneous permutation of the complete graph and circuit is a different intervention.

The theory does not assume that entanglement is present for every input, that training finds a particular parameter region, or that a circuit property implies a predictive gain. It gives exact observables and a restricted function-family distinction. Classical learners with access to $(\mathbf{v}, \mathbf{e})$ may represent or approximate the same functions.

B.4 Full proof of edge observability

Write $\chi_i = \cos\alpha_i$ and $\varsigma_i = \sin\alpha_i$. For $\mathscr{R}_i = R_y(\omega_i^y)R_x(\omega_i^x)$, direct Pauli conjugation gives

$$\mathscr{R}_i^\dagger Z_i \mathscr{R}_i = R_x^\dagger(\omega_i^x)\left[Z_i\cos\omega_i^y - \sigma_i^x\sin\omega_i^y\right]R_x(\omega_i^x) \tag{29}$$

$$= \cos\omega_i^x\cos\omega_i^y Z_i - \sin\omega_i^y\sigma_i^x + \sin\omega_i^x\cos\omega_i^y\sigma_i^y \tag{30}$$

$$= \lambda_i^z Z_i + \lambda_i^x\sigma_i^x + \lambda_i^y\sigma_i^y. \tag{31}$$

For $R_{ZZ}(\beta) = \exp(-\mathrm{i}\beta Z_1 Z_2/2)$,

$$R_{ZZ}^\dagger\sigma_1^x R_{ZZ} = \sigma_1^x\cos\beta - \sigma_1^y Z_2\sin\beta, \tag{32}$$

$$R_{ZZ}^\dagger\sigma_1^y R_{ZZ} = \sigma_1^y\cos\beta + \sigma_1^x Z_2\sin\beta, \tag{33}$$

with symmetric identities for qubit 2. The node-encoded state

$$|\mathrm{node}\rangle = [R_y(\alpha_1)\otimes R_y(\alpha_2)]|00\rangle \tag{34}$$

satisfies

$$\langle\sigma_i^x\rangle_{\mathrm{node}} = \varsigma_i, \qquad \langle\sigma_i^y\rangle_{\mathrm{node}} = 0, \qquad \langle Z_i\rangle_{\mathrm{node}} = \chi_i. \tag{35}$$

Substitution into Eqs. (31)– (33) yields

$$\langle Z_1\rangle = \lambda_1^z\chi_1 + \lambda_1^x\varsigma_1\cos\beta + \lambda_1^y\varsigma_1\chi_2\sin\beta, \tag{36}$$

$$\langle Z_2\rangle = \lambda_2^z\chi_2 + \lambda_2^x\varsigma_2\cos\beta + \lambda_2^y\chi_1\varsigma_2\sin\beta. \tag{37}$$

Applying the chain rule to the first line proves Eq. (14).

The pair observable exposes a different combination. Expanding $(\mathscr{R}_1^\dagger Z_1\mathscr{R}_1)(\mathscr{R}_2^\dagger Z_2\mathscr{R}_2)$, conjugating each Pauli product by R_{ZZ}, and applying Eq. (35) gives

$$\langle Z_1Z_2\rangle = \lambda_1^z\lambda_2^z\chi_1\chi_2 + \lambda_1^x\lambda_2^x\varsigma_1\varsigma_2 \tag{38}$$

$$+ (\lambda_1^z\lambda_2^y\varsigma_2 + \lambda_2^z\lambda_1^y\varsigma_1)\sin\beta \tag{39}$$

$$+ (\lambda_1^z\lambda_2^x\chi_1\varsigma_2 + \lambda_2^z\lambda_1^x\chi_2\varsigma_1)\cos\beta. \tag{40}$$

Thus local and pair readouts expose different functions of the same conditioned interaction; Eq. (40) does not imply that pair readout is uniformly more informative.

If $\mathscr{R}_1 = I$, then $\lambda_1^z = 1$ and $\lambda_1^x = \lambda_1^y = 0$, so $\langle Z_1 \rangle = \chi_1$. The reason is structural: $[Z_1, Z_1 Z_2] = 0$. More generally, an R_y-only mixer has $\lambda_1^y = 0$ and $\lambda_1^x = -\sin\omega_1^y$, so it can expose edge dependence through $\lambda_1^x \varsigma_1 \cos\beta$ but cannot produce the particular three-way term. A nonzero R_x component with $\cos\omega_1^y \neq 0$ supplies $\lambda_1^y \neq 0$ and the three-way interaction.

B.5 Full proof of the mixed interaction

The final term of the local-readout expression in Eq. (13) depends on all three encoded angles. Differentiating it gives

$$\frac{\partial^3}{\partial\alpha_1\partial\alpha_2\partial\beta}\left[\lambda_1^y \sin\alpha_1 \cos\alpha_2 \sin\beta\right] = -\lambda_1^y \cos\alpha_1 \sin\alpha_2 \cos\beta, \tag{41}$$

which is nonzero except on a union of lower-dimensional zero sets when $\lambda_1^y \neq 0$.

To connect encoded angles to raw normalized features, write, within an unclipped region,

$$\widetilde{v}_i = (v_i - \mu_i^v)/\operatorname{scale}_i^v, \tag{42}$$

$$\alpha_i'(v_i) = \frac{\alpha_{\max}\theta_i^N}{\operatorname{scale}_i^v} \operatorname{sech}^2(\theta_i^N \widetilde{v}_i), \tag{43}$$

$$\widetilde{e}_{12} = (e_{12} - \mu_{12}^e)/\operatorname{scale}_{12}^e, \tag{44}$$

$$\beta'(e_{12}) = \frac{\beta_{\max}\theta_{12}^E}{\operatorname{scale}_{12}^e} \operatorname{sech}^2(\theta_{12}^E \widetilde{e}_{12}). \tag{45}$$

The multivariate chain rule therefore gives

$$\frac{\partial^3\langle Z_1\rangle}{\partial v_1\partial v_2\partial e_{12}} = -\lambda_1^y \alpha_1'(v_1)\alpha_2'(v_2)\beta'(e_{12}) \cos\alpha_1 \sin\alpha_2 \cos\beta. \tag{46}$$

Outside the clipping interval the corresponding scaling-map derivative is zero; at the clipping boundary it is not classically differentiable. Strong $\tanh$ saturation can also make the derivative numerically small. These are degenerate cases, not contradictions of the encoded-angle identity.

For a theory-only linearization $\alpha_1 = a_1 v_1$, $\alpha_2 = a_2 v_2$, and $\beta = be$, the mixed factor has the exact expansion

$$\sin(a_1 v_1)\cos(a_2 v_2)\sin(be) = -\frac{1}{8}\sum_{\boldsymbol{\sigma}\in\{-1,1\}^3} \sigma_1\sigma_3 \exp\Big(\mathrm{i}[\sigma_1 a_1 v_1 + \sigma_2 a_2 v_2 + \sigma_3 be]\Big). \tag{47}$$

Every displayed frequency vector has nonzero coordinates for v_1, v_2, and e. This use of Fourier coordinates relies on existing PQC theory (Schuld et al., 2021); the QRG-specific result is the identification of the joint component–component–Riemannian-edge term for the implemented circuit. Because $\tanh$ is nonlinear, Eq. (47) is not a finite Fourier series in the raw EEG-derived values.

B.6 Full proof of restricted relational separation

For an edge-blind controlled circuit, write

$$\psi_0(\mathbf{v}, \mathbf{e}) = \langle 0|U_0^\dagger(\mathbf{v};\Theta)\mathsf{Obs}\,U_0(\mathbf{v};\Theta)|0\rangle. \tag{48}$$

Since neither U_0 nor Obs nor the fixed parameters depend on e_{ij}, $\psi_0(\mathbf{v}, \mathbf{e})$ is constant along that coordinate and

$$\frac{\partial\psi_0(\mathbf{v}, \mathbf{e})}{\partial e_{ij}} = 0. \tag{49}$$

This covers Node-only VQC and Fixed-entanglement VQC under the edge-blindness condition in Appendix B.3: a trainable fixed angle may change during training but does not change across evaluation trials as e_{ij} changes.

For QRG, select $\mathsf{Obs} = Z_1$, $\omega_1^x = \pi/2$, and $\omega_1^y = 0$. Then $\lambda_1^z = \lambda_1^x = 0$ and $\lambda_1^y = 1$, so

$$\psi_{\mathrm{QRG}}(\mathbf{v}, \mathbf{e}) = \sin \alpha_1(v_1) \cos \alpha_2(v_2) \sin \beta(e_{12}), \tag{50}$$

with

$$\frac{\partial \psi_{\mathrm{QRG}}}{\partial e_{12}} = \sin \alpha_1 \cos \alpha_2 \cos \beta \, \beta'(e_{12}). \tag{51}$$

The derivative is nonzero on an open set if β' is not identically zero. Hence at least one edge-varying function belongs to the QRG family and not to the defined edge-blind families.

This argument proves $\mathcal{F}_{\mathrm{QRG}} \setminus (\mathcal{F}_{\mathrm{node}} \cup \mathcal{F}_{\mathrm{fixed}}) \neq \varnothing$; it does not prove either edge-blind family is a subset of $\mathcal{F}_{\mathrm{QRG}}$. In particular, the implemented bias-free map $\beta(e) = \beta_{\max} \tanh(\theta^E e)$ does not necessarily reproduce every nonzero constant entangling angle for all e. Nor does the result separate QRG from an arbitrary classical model that receives e explicitly.

B.7 Full proof of the alignment corollary

Fix $\mathbf{v}$ and define $h_{\mathbf{v}}(\mathbf{e}) = \psi(\mathbf{v}, \Pi_{\mathcal{E}} \mathbf{e})$. If $\psi(\mathbf{v}, \mathbf{e}) = \psi(\mathbf{v}, \Pi_{\mathcal{E}} \mathbf{e})$ for every $\mathbf{e}$ in an open neighborhood, differentiation and the chain rule give

$$\nabla_{\mathbf{e}} \psi(\mathbf{v}, \mathbf{e}) = \nabla_{\mathbf{e}} h_{\mathbf{v}}(\mathbf{e}) = \Pi_{\mathcal{E}}^{\top} \nabla_{\mathbf{e}} \psi(\mathbf{v}, \Pi_{\mathcal{E}} \mathbf{e}), \tag{52}$$

which proves the necessary condition in Eq. (17). Its contrapositive proves that violation of the gradient relation rules out functional equality throughout the neighborhood. It does not necessarily imply unequal values at the single point $\mathbf{e}_0$ because two differentiable functions can intersect while having different gradients.

At a permutation-fixed point $\mathbf{e}_0 = \Pi_{\mathcal{E}} \mathbf{e}_0$, the local effect is especially clear. For a small perturbation δ,

$$\psi(\mathbf{v}, \mathbf{e}_0 + \delta) - \psi(\mathbf{v}, \mathbf{e}_0 + \Pi_{\mathcal{E}} \delta) = \nabla_{\mathbf{e}} \psi(\mathbf{v}, \mathbf{e}_0)^{\top} (I - \Pi_{\mathcal{E}}) \delta + o(\|\delta\|). \tag{53}$$

If the gradient is not permutation invariant, some δ makes the leading term nonzero. Shuffling can nevertheless leave a representation unchanged under a graph automorphism accompanied by symmetric encodings, parameters, and readout, or when exchanged edge values/sensitivities coincide.

B.8 Extension to a sparse multiqubit graph

For one block,

$$U_E = \prod_{(i,j) \in \mathcal{E}} e^{-\mathrm{i} \beta_{ij} Z_i Z_j / 2}, \qquad [Z_i Z_j, Z_k Z_l] = 0. \tag{54}$$

Consider local readout at node i. Edges not incident to i commute with σ_i^x and σ_i^y and therefore disappear from their conjugation. Repeatedly applying Eq. (33) over the incident neighborhood $\mathcal{N}(i)$ gives, by induction, Pauli strings of the form

$$\tau_i \prod_{j \in \mathcal{U}} Z_j, \qquad \mathcal{U} \subseteq \mathcal{N}(i), \quad \tau_i \in \{\sigma_i^x, \sigma_i^y\}, \tag{55}$$

with coefficients

$$\mathrm{sgn}(\mathcal{U}) \prod_{j \in \mathcal{U}} \sin \beta_{ij} \prod_{k \in \mathcal{N}(i) \setminus \mathcal{U}} \cos \beta_{ik}, \qquad \mathrm{sgn}(\mathcal{U}) \in \{-1, 1\}. \tag{56}$$

The induction holds because conjugating by one additional incident ZZ gate either retains the current Pauli through its cosine branch or toggles $\sigma_i^x \leftrightarrow \sigma_i^y$, appends Z_j, and takes its sine branch. Taking expectations in the node product state converts surviving strings into factors containing $\sin \alpha_i$ and products of $\cos \alpha_j$. Hence the sparse graph creates structured multivariate interactions over local graph neighborhoods. Some terms vanish for particular mixer settings or input angles; the construction does not guarantee that every algebraically available term survives training.

B.9 MULTIPLE BLOCKS AND THE PRECISE ROLE OF THE MIXER

The main mechanism already appears for $L = 1$. For $L > 1$, alternating node, edge, and mixer layers can expand an observable's Pauli light cone. The resulting expectation is an analytic trigonometric function of the encoded angles that enter this finite-depth circuit light cone; for a sparse graph and finite L, dependence on every encoded angle is neither required nor guaranteed. With linearized angle maps, repeated encodings can enlarge the accessible frequency set (Pérez-Salinas et al., 2020; Schuld et al., 2021). This statement does not imply monotonic growth in useful expressivity, gradient quality, generalization, or classification accuracy.

Lemma 4.1 concerns the current edge layer followed by immediate diagonal readout. In a multi-block circuit, a later R_y node layer is itself noncommuting and can expose phases produced by an earlier R_{ZZ} layer even if the explicit mixer is removed. Therefore the mathematically correct architectural statement is: At least one suitable noncommuting operation must lie after an edge-conditioned ZZ phase on its causal path to a Z/ZZ measurement. QRG-EEG places a trainable R_x–then–R_y mixer after every edge layer to provide this path within each block.

This is narrower than claiming that an entire $L = 2$ no-mixer circuit must be edge invariant.

B.10 DEGENERATE CASES AND EMPIRICAL SCOPE

Lemma 4.1, Theorem 4.2, Proposition 4.3, and Corollary 4.4 characterize properties available to the QRG-EEG circuit family under their stated conditions. They do not assert that every trained parameter configuration, input trial, or output logit exhibits a nonzero edge response. Tab. 6 identifies the encoding, circuit, readout, and comparison conditions that can suppress or qualify the derived interaction.

Table 6: Cases that suppress or qualify the derived interaction.

Case	Consequence
$\theta^E_{ij} = 0$ or clipped/saturated edge map	$\beta'(e_{ij})$ vanishes or becomes numerically negligible.
$\sin \alpha_i = 0$	Local node i supplies no transverse component to the derived term.
$\sin \omega^x_i = 0$ or $\cos \omega^y_i = 0$	$\lambda^y_i = 0$; the displayed three-way term vanishes.
Special β or α_j	The derivative can vanish at isolated angle surfaces.
Zero linear-head weight	A measurement may be edge sensitive while a particular output logit ignores it.
Edge-blind baseline receives $\mathbf{e}$ elsewhere	Proposition no longer applies.
Graph/circuit permutation symmetry	Shuffling may leave the representation unchanged.
Statistical dependence of $\mathbf{v}$ and $\mathbf{e}$	Does not alter coordinate-level derivatives, but prevents interpreting them as statistical independence.

The first four conditions suppress the relevant path inside the encoding or circuit. The linear-head condition acts after measurement, while the final three conditions delimit conclusions from edge-blind and edge-shuffling controls. The experiments evaluate consequences of the mechanism under training rather than re-proving the algebra: Tab. 5 defines the interventions, Tab. 2 reports node/edge and alignment controls, Tab. 1 reports matched classical comparisons, and Tab. 12 reports the post-edge mixer diagnostic. Accordingly, the theoretical contribution is architecture-specific: non-commuting post-edge dynamics can convert a trial-conditioned ZZ phase into measurable terms involving two component-aligned node variables and their Riemannian relation, yielding a restricted separation from the defined edge-blind controls, not from arbitrary classical models.

C EXPERIMENTAL PROTOCOL AND REPRODUCIBILITY DETAILS

This section fixes the cohorts, fold construction, comparator implementations, run matrix, software environment, and reporting estimands. Appendix B.2 separately specifies the fold-local geometric construction, candidate-ranking rule, optimization constants, and exact RQ2 interventions.

C.1 COHORTS AND FIXED SIGNAL RULES

Tab. 7 summarizes the task, class structure, and patient-level LOSO folds of the four primary cohorts; we refer to the ALS subset of bigP3BCI Study N as bigP3BCI-ALS throughout the paper. bigP3BCI aggregates studies with different populations and acquisition conditions; we therefore use only diagnosis-verified, acquisition-compatible Study N rather than pooling records by label name alone (Mainsah et al., 2025). BNCI2014-008 contributes its eight ALS participants in the P300 speller task (Riccio et al., 2013). The EEGET primary cohort contains only its six ALS participants in the three-class MI setting; healthy participants and the four-class setting are excluded (Ngo et al., 2024). UCL contributes all eight ALS patients, with every longitudinal session assigned to that patient's fold (Patel et al., 2025a). We analyze each dataset separately and never pool scores across P300 and MI. All datasets are resampled to 128 Hz. P300 uses a 0.5–20 Hz band and the 0–0.8 s post-stimulus interval. MI uses 4–40 Hz, with fixed 4 s UCL and 2 s EEGET windows. A $120\,\mu$V peak-amplitude gate is applied only when the stored signal has a validated voltage interpretation; EEGET follows the predeclared, label-blind bypass because its stored units do not support that gate. The node and covariance branches receive independently source-scaled copies of the same projected components. The covariance copy is not normalized to unit variance trial by trial, which would remove task-relevant amplitude and band-power variation.

Table 7: Primary ALS-EEG cohorts. Each patient defines one outer LOSO fold.

Dataset	Paradigm	Classes	Patients/ folds
bigP3BCI-ALS (Study N)	P300	Target, non-target	8
BNCI2014-008	P300	Target, non-target	8
EEGET-ALS-MI	MI	Left, right, rest	6
UCL-ALS-MI	MI	Left, right, rest	8
Total	P300 and MI	–	30

C.2 NESTED SOURCE-ONLY SELECTION AND REFITTING

For outer test patient $s_\star$, every fitted operation is reconstructed from the remaining patients. MI compares one-versus-rest CSP with PCA; P300 fixes xDAWN and compares plain with ERP-residual covariance. The candidate set is the Cartesian product of the two admissible task representations and $q \in \{6, 8\}$, subject to $q \le C$. In the reported final matrix, the QRG-EEG circuit, $L = 2$, and graph multiplier 1.5 are fixed across outer folds.

A fixed $12/12/6$-epoch proxy evaluates all candidates on one deterministic source validator and advances the top two. Both shortlisted candidates then complete full source-patient validation: three deterministic validators in non-EEGET folds and all five available source validators in EEGET folds. Raw probabilities select the structure and checkpoint before decision adjustment. Pooled out-of-fold source predictions subsequently select one patient-balanced threshold for P300 or one source-prior correction coefficient for MI. The selected stage lengths are replayed with the same cosine-schedule prefix on all source patients. Missing candidates or failed validators invalidate the fold rather than silently changing its search space. Eqs. (27)–(28) give the ranking and tie-breaking rules. This source-only procedure is distinct from the fold-wise reporting rule in Appendix C.5.

C.3 OPTIMIZATION, SIMULATOR, AND SOFTWARE ENVIRONMENT

QRG-EEG and its matched controls use maximum stage budgets of $60/60/30$ epochs, AdamW, weight decay 10^{-4}, gradient-norm limit 5, deterministic cosine schedules, and EMA decay 0.99. Tab. 4 lists the stage learning rates, minibatch size, angle bounds, initialization, and objective weights; Tab. 8 records the software environment. Generic learned baselines use subject-balanced classification loss and replay the source-selected prefix of their deterministic schedule during all-source refitting. The custom

Table 8: Recorded model software environment.

Component	Recorded version
GPU	NVIDIA RTX A6000 (48 GB)
NVIDIA driver	535.309.01
Python	3.11.16
PyTorch	2.5.1+cu121
MNE	1.12.1
MOABB	1.6.1
pyRiemann	0.12
scikit-learn	1.9.0

PyTorch statevector backend computes exact analytic expectations; it uses neither finite-shot sampling nor a hardware-noise model.

The supplementary fold logs record `wall_seconds` for final all-source fitting and held-out prediction. This field excludes preprocessing and source-only planning and is therefore an implementation-specific runtime measure, not hardware or end-to-end latency.

Table 9: Code-level definition of the prediction models. "Adapted" denotes a fold-compatible implementation of the cited mechanism, not an execution of the authors' official repository.

ID	Model	Implemented comparison
B1	TS–LR (Barachant et al., 2012)	Source-referenced tangent vector with class-balanced logistic regression.
B2	EEGNet (Lawhern et al., 2018)	Compact temporal, depthwise-spatial, and separable-convolution encoder with a linear head.
B3	Task-specific strong	P300: flattened source-fitted xDAWN components with shrinkage LDA (Rivet et al., 2009). MI: adapted six-band FBCNet-style grouped spatial filtering and log-variance readout (Mane et al., 2021).
B4	QEEGNet-style (Chen et al., 2024)	Adapted EEGNet embedding, linear compression to q coordinates, R_y encoding, ring CNOT layers, trainable R_y rotations, and Z readout.
B5	Task-specific geometric/statistical	P300: standardized Bayesian ridge discriminant on flattened source-fitted xDAWN components, reported as xDAWN–BLDA. MI: fold-compatible Graph-CSPNet-style reimplementation using six filter-band shrinkage SPD covariances, a shared orthonormal BiMap, LogEig-based symmetric vectorization, cyclic neighboring-band aggregation, mean pooling, and a linear classifier (Ju & Guan, 2024).
B6	Same-Graph GAT	Internal control using the identical component-aligned node encoder, Riemannian edge values, and sparse topology in two edge-aware graph-attention layers.
B7	Matched Classical Interaction	Internal control applying a shallow nonlinear head to node values and aligned products $e_{ij}v_iv_j$ under QRG-matched inputs, objective, validation budget, and learned parameter count.
–	QRG-EEG	Proposed component-aligned VQC with data-conditioned node rotations, pair-aligned R_{ZZ} interactions, noncommuting mixers, and analytic Z/ZZ readout.
S1	Riemannian-VQC (Andreev et al., 2023; Cattan & Quemy, 2023)	Tangent-vector compression to q coordinates followed by the generic VQC used for B4; removes pair alignment.
S2	CRGNet-style (Gao et al., 2022)	Adapted fusion of a 16-dimensional EEGNet feature with the full tangent vector through a two-layer nonlinear head.
S3	PLV-GAT-style (Patel et al., 2025b)	Adapted complete-channel graph with log-variance node attributes, PLV edge weights, and graph attention; evaluated only on the compatible UCL-ALS-MI cohort.

C.4 Implemented comparison matrix

Baseline-specific settings were prespecified as follows: B1 used class-balanced logistic regression with $C_{\mathrm{LR}} = 1$; B2/B4 used EEGNet $(F_1, D, F_2) = (8, 2, 16)$ and dropout 0.25, with $L = 2$ for B4; B3 used LSQR shrinkage LDA for P300 or four spatial filters per band for MI; and B5 used standardized Bayesian ridge for P300 or BiMap dimension $\min(8, C)$ and hidden width 16 for MI, with both MI baselines using $[4, 8]$, $[8, 12]$, $[12, 16]$, $[16, 20]$, $[20, 28]$, and $[28, 40]$ Hz bands. These prespecified values were singleton candidate sets; source validation selected $q \in \{6, 8\}$ where applicable and the epoch prefixes (1–60 for B2/B3-MI/B4/B5-MI; 1–60/1–60/1–30 for B6/B7), while B6/B7 searched $h \in \{1, \ldots, 64\}$ under the 10% parameter-matching tolerance. For MI B5, Tab. 9 presents a custom, fold-compatible reimplementation of the central SPD-graph mechanism described in Graph-CSPNet (Ju & Guan, 2024); it is not an execution of the authors' official repository. Each trial is represented by six filter-band shrinkage covariance matrices. A shared orthonormal BiMap preserves positive definiteness, a LogEig operation maps each matrix to a symmetric vector, and one graph layer aggregates the two cyclic neighboring bands before mean pooling and linear classi-

fication. We adapted the input dimensions, graph module, and prediction head to the varying ALS channel sets and the patient-level LOSO protocol. We therefore use *Graph-CSPNet-style* for this experimental comparator; *Graph-CSPNet* without the suffix refers only to the original published method.

C.5 Fold-Wise Reporting and Statistical Estimands

Each held-out patient is one independent reporting unit. Let $z^{(\nu)}_{dmsr}$ denote metric ν for dataset d, model m, held-out patient s, and seed $r \in \{11, 22, 33\}$. We first average all three seeds within the patient fold,

$$\bar{z}^{(\nu)}_{dms} = \frac{1}{3} \sum_{r \in \{11,22,33\}} z^{(\nu)}_{dmsr}. \tag{57}$$

For a dataset containing N_d patients, the reported location and dispersion are

$$\bar{z}^{(\nu)}_{dm} = \frac{1}{N_d} \sum_{s=1}^{N_d} \bar{z}^{(\nu)}_{dms}, \qquad s^{(\nu)}_{dm} = \sqrt{\frac{1}{N_d - 1} \sum_{s=1}^{N_d} \left(\bar{z}^{(\nu)}_{dms} - \bar{z}^{(\nu)}_{dm} \right)^2}. \tag{58}$$

Thus, the displayed standard deviation measures between-patient variability, not trial- or seed-level variability; patients with more trials or sessions receive no additional weight. No held-out score selects a seed.

Macro-F1 is primary. P300 co-reports binary AUROC, while MI co-reports balanced accuracy. Tab. 10 additionally reports accuracy and MI macro one-vs-rest AUROC; Cohen's κ is used only in the designated diagnostic comparisons. Trials and seeded reruns are never treated as independent patients. Because the "strongest comparator" in Tab. 1 is identified from the same dataset summary, that comparison is descriptive: we report wins/ties/losses and paired effect sizes without attaching a post-selection confirmatory claim. Full patient-fold distributions accompany the dataset summaries because six or eight patients cannot support fine-grained asymptotic inference.

Table 10: Additional accuracy and AUROC results.

Dataset	Metric	B1	B2	B3	B4	B5	B6	B7	QRG-EEG
bigP3BCI-ALS	Accuracy	0.515 ± 0.117	0.881 ± 0.013	0.884 ± 0.007	0.865 ± 0.027	0.865 ± 0.029	0.881 ± 0.021	0.888 ± 0.002	**0.891 ± 0.004**
	AUROC	0.520 ± 0.026	0.614 ± 0.064	0.585 ± 0.060	0.610 ± 0.073	0.589 ± 0.066	0.531 ± 0.029	0.593 ± 0.057	**0.653 ± 0.084**
BNCI2014-008	Accuracy	0.580 ± 0.172	0.829 ± 0.015	0.836 ± 0.015	0.838 ± 0.008	0.829 ± 0.018	0.833 ± 0.001	0.836 ± 0.011	**0.850 ± 0.009**
	AUROC	0.553 ± 0.039	0.791 ± 0.032	0.776 ± 0.040	0.791 ± 0.032	0.779 ± 0.042	0.607 ± 0.074	0.789 ± 0.034	**0.807 ± 0.030**
EEGET-ALS-MI	Accuracy	0.500 ± 0.056	0.511 ± 0.054	0.504 ± 0.083	0.495 ± 0.072	0.499 ± 0.003	0.505 ± 0.023	0.531 ± 0.048	**0.608 ± 0.066**
	AUROC	0.684 ± 0.074	0.680 ± 0.086	0.582 ± 0.144	0.676 ± 0.068	0.515 ± 0.036	0.628 ± 0.085	0.659 ± 0.080	**0.764 ± 0.077**
UCL-ALS-MI	Accuracy	0.416 ± 0.165	0.447 ± 0.052	0.455 ± 0.033	0.462 ± 0.074	0.453 ± 0.030	0.452 ± 0.019	0.484 ± 0.060	**0.544 ± 0.059**
	AUROC	0.665 ± 0.075	0.638 ± 0.064	0.627 ± 0.048	0.636 ± 0.080	0.621± 0.070	0.612 ± 0.049	0.646 ± 0.076	**0.713 ± 0.081**

Each cell reports mean ± sample standard deviation across held-out patients after averaging the three fixed seeds within each patient fold, as in Eqs. (57)–(58). Bold and underlining mark the highest and second-highest displayed means within each dataset–metric block. B5 is xDAWN–BLDA for P300 and the adapted Graph-CSPNet-style comparator for MI.

D Additional results

D.1 Accuracy and score-ranking metrics

Tab. 10 reports accuracy and AUROC as complementary metrics, computed from the same final seed-averaged patient-fold outputs used for Tab. 1; all displayed values, including the UCL Graph-CSPNet-style AUROC, are obtained directly from these outputs, and no value is imputed. Accuracy reflects decisions at the operating point but can be dominated by the non-target class on the imbalanced P300 datasets, so it is included for completeness and should be read alongside balanced accuracy and Macro-F1. AUROC instead evaluates score ranking independently of the operating threshold, using binary AUROC for P300 and macro one-vs-rest AUROC for the three-class MI tasks. Comparator identifiers B1–B7 follow the definitions in Tab. 9.

D.2 FOLD-WISE PAIRED ANALYSIS

Consistent with Tab. 1, the paired analysis first averages the three fixed-seed results within each patient–model fold, producing one Macro-F1 estimate per held-out patient and method. This aggregation prevents stochastic training runs from being treated as independent clinical observations and preserves the patient as the statistical unit. Within each dataset, the seed-averaged QRG-EEG score is paired with the comparator score obtained for the same held-out patient. The resulting patient-wise difference is defined as $\Delta = \text{Macro-F1}_{\text{QRG-EEG}} - \text{Macro-F1}_{\text{comparator}}$, so positive values favor QRG-EEG. Pairing the scores in this manner accounts for variation in patient-specific decoding difficulty and determines whether the dataset-level difference is consistent across patients rather than being driven by a small subset of folds.

Table 11: Fold-wise paired Macro-F1 comparison between QRG-EEG and the strongest non-QRG comparator under seed-averaged patient-fold scores.

Dataset	Comparator	W/T/L	Mean Δ	p_{Holm}	d_z
bigP3BCI-ALS	B4	8/0/0	0.043	0.0156	2.70
BNCI2014-008	B4	8/0/0	0.060	0.0156	1.44
EEGET-ALS-MI	B7	6/0/0	0.111	0.0156	4.15
UCL-ALS-MI	B1	8/0/0	0.116	0.0156	0.96

Δ is patient-wise $\text{Macro-F1}_{\text{QRG-EEG}}$ minus $\text{Macro-F1}_{\text{comparator}}$; W/T/L counts positive, zero, and negative differences. The reported p_{Holm} values are exact one-sided Wilcoxon signed-rank probabilities adjusted across the four datasets, and d_z is the paired standardized mean difference. Because each comparator is identified from the same dataset-level results, these statistics are descriptive rather than preregistered confirmatory tests. Comparator IDs are defined in Tab. 9.

For each dataset, the comparator is the non-QRG model with the highest dataset-level mean Macro-F1 under seed-averaged scores, which provides a stringent, although data-dependent, reference; averaging seeds before the comparison also prevents optimization replicates from being treated as independent patients. Tab. 11 summarizes each dataset by its win/tie/loss count, mean paired difference, Holm-adjusted exact Wilcoxon result, and paired standardized mean difference, which respectively capture directional consistency, absolute performance change, rank-based evidence, and effect magnitude relative to between-patient variation. Because the comparator is identified from the same dataset-level results used in the analysis, these statistics are interpreted descriptively rather than as independently confirmatory tests. Complementing these summaries, Fig. 6 retains every patient-level difference, making visible whether the dataset mean reflects a consistent shift or is driven by a small number of patients.

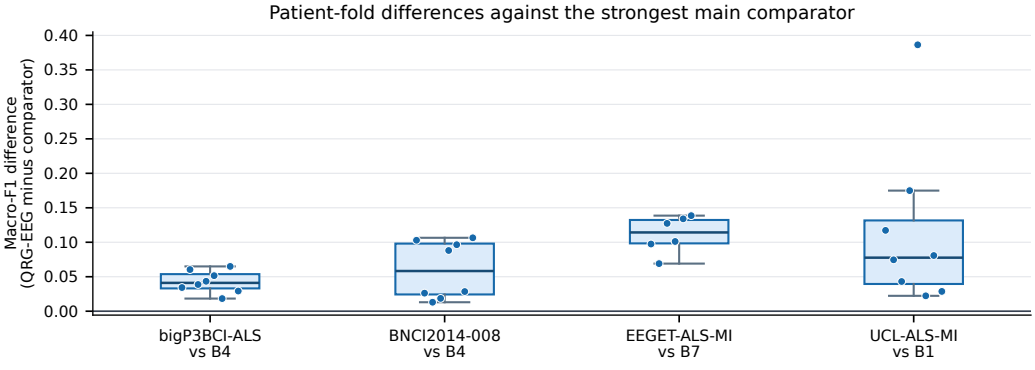

Figure 6: Patient-wise Macro-F1 difference between QRG-EEG and the strongest non-QRG comparator selected from seed-averaged dataset results. Each point represents one held-out patient after averaging the three fixed seeds; positive values favor QRG-EEG.

E MIXER SENSITIVITY

We test the corresponding practical design choice on two source-validation folds per dataset, keeping the graph, node encoder, objective, and source-only protocol fixed. The no-mixer condition removes both post-edge mixer rotations. The R_y-only condition retains the $\cos\beta$ edge path but sets $a_i^y = 0$, removing the displayed $\sin\beta$ mixed term. The full condition uses the implemented R_x-then-R_y sequence.

Table 12: Mixer sensitivity on a two-fold-per-dataset diagnostic subset. Values are means across the two source-validation folds.

	bigP3BCI-ALS			BNCI2014-008			EEGET-ALS-MI			UCL-ALS-MI		
Mixer	F1	BAcc	κ	F1	BAcc	κ	F1	BAcc	κ	F1	BAcc	κ
No mixer	.471	.500	.000	.598	.582	.211	.274	.361	.060	.328	.366	.079
R_y only	.489	.508	.032	.641	.628	.294	.341	.410	.118	.382	.411	.134
Full R_x–R_y	**.549**	**.543**	**.120**	**.699**	**.689**	**.400**	**.519**	**.498**	**.333**	**.474**	**.521**	**.203**

In Tab. 12, the descriptive ordering is full R_x–R_y > R_y-only > no mixer on all four datasets and all three reported metrics. This supports retaining the full post-edge mixer in the implemented circuit. It does not turn the empirical table into a proof of the lemma: in a two-block circuit, the second node encoding can expose a phase created in the first block even when the explicit mixers are removed. The experiment therefore tests the value of the implemented causal path, while the lemma characterizes the immediate one-block path.

F ARCHITECTURE SENSITIVITY: QUBIT COUNT AND DEPTH

We vary $q \in \{6, 8\}$ and $L \in \{2, 4\}$ on the same two-fold-per-dataset, diagnostic subset. Each axis changes alone while retaining the graph-conditioned node/edge construction and source-patient validation protocol.

Table 13: Qubit-count and depth sensitivity on the diagnostic subset. Values are means across two source-validation folds. Bold compares values only within the same axis and dataset.

		bigP3BCI-ALS			BNCI2014-008			EEGET-ALS-MI			UCL-ALS-MI		
Axis	Value	F1	BAcc	κ	F1	BAcc	κ	F1	BAcc	κ	F1	BAcc	κ
Qubits	$q = 6$	**.548**	**.553**	**.111**	.668	.672	.350	.513	.536	.330	**.493**	**.528**	**.219**
	$q = 8$	.524	.523	.079	**.689**	**.691**	**.384**	**.514**	**.538**	**.329**	.446	.487	.161
Depth	$L = 2$	**.549**	**.543**	**.120**	**.699**	**.689**	**.400**	**.519**	**.498**	**.333**	**.474**	**.521**	**.203**
	$L = 4$	.490	.493	.043	.631	.622	.290	.449	.481	.264	.421	.453	.152

In Tab. 13, the preferred qubit count is task dependent: $q = 6$ leads on bigP3BCI-ALS and UCL-ALS-MI, whereas $q = 8$ leads on BNCI2014-008 and narrowly on EEGET-ALS-MI. This pattern supports source-only selection between compact qubit counts rather than a universal value. In contrast, $L = 2$ leads $L = 4$ on every dataset and metric in this subset. The depth result supports the shallow two-block primary design, but the subset is a sensitivity analysis.

G CLASSICAL CONTROLS ON THE FROZEN NODE–EDGE REPRESENTATION

Classical approximation is a central concern for small simulated VQCs (Landman et al., 2023; Schreiber et al., 2023). We therefore compare three controls receiving the same source-fitted, frozen node values $\{\widetilde{v}_{b,i}\}$ and edge values $\{\widetilde{e}_{b,ij}\}$. Matched Classical Interaction (B7) uses aligned products $e_{ij}v_iv_j$ under the parameter-matching rule. The finite-RFF control applies $D = 256$ random

Fourier features with bandwidth $\sigma = \pi/2$ to the concatenated node–edge vector, followed by a linear classifier. This is a finite shift-invariant nonlinear control; it is not asserted to reconstruct the trained quantum kernel. The SVM-RBF control uses a class-balanced RBF kernel on the same vector ($C = 10$, $\gamma = \texttt{scale}$).

Table 14: Classical controls on the same frozen node–edge features as QRG-EEG. Values are mean Macro-F1 on the same two diagnostic outer folds per dataset. Δ is QRG-EEG minus each control on those folds.

	bigP3BCI-ALS		BNCI2014-008		EEGET-ALS-MI		UCL-ALS-MI	
Model	F1	Δ	F1	Δ	F1	Δ	F1	Δ
B7 Matched Classical	.469	+.088	.613	+.080	.394	+.119	.331	+.140
Finite RFF	.475	+.082	.622	+.071	.411	+.102	.342	+.129
SVM–RBF	.482	+.075	.638	+.055	.423	+.090	.358	+.113
QRG-EEG	**.557**	–	**.693**	–	**.513**	–	**.471**	–

QRG-EEG leads all three controls on each diagnostic dataset (Tab. 14). Because the controls receive the same frozen representation, the differences are not explained by access to additional node or edge variables. They show that the tested aligned-product, finite-RFF, and RBF decision functions do not reproduce the observed performance under their evaluated settings.

H Objective and Structural Sensitivity

We evaluate the contribution of each objective term and the local sensitivity of the frozen loss coefficients, ranking margin, and graph density on EEGET-ALS-MI. The loss-component ablation and the hyperparameter sweeps retain the same seeds, preprocessing, source-only fitting, and optimization budget. The coefficient, margin, and graph-density sensitivity results are computed only from source-patient validation folds.

Loss-component ablation. The complete objective uses $(w_{\mathrm{nat}}, w_{\mathrm{pc}}, w_{\mathrm{rank}}) = (0.65, 0.35, 0.08)$ with $m = 0.25$. When either classification term is removed, the remaining classification coefficient is set to one to preserve the total cross-entropy scale. Removing $\mathcal{J}_{\mathrm{rank}}$ leaves $(w_{\mathrm{nat}}, w_{\mathrm{pc}}) = (0.65, 0.35)$ unchanged. The full-objective row reuses the corresponding primary EEGET-ALS-MI patient folds.

Table 15: Loss-component ablation on EEGET-ALS-MI. Values are Macro-F1 mean $\pm$ sample standard deviation across six held-out patients after averaging the three fixed seeds within each patient. The final column reports $\Delta = \text{Macro-F1}_{\mathrm{condition}} - \text{Macro-F1}_{\mathrm{full}}$.

Objective	Macro-F1	Δ
Full $\mathcal{J}$	$\mathbf{.516 \pm .077}$	.000
Without $\mathcal{J}_{\mathrm{nat}}$	$.482 \pm .071$	$-.034$
Without $\mathcal{J}_{\mathrm{pc}}$	$.473 \pm .063$	$-.043$
Without $\mathcal{J}_{\mathrm{rank}}$	$.492 \pm .079$	$-.024$

Tab. 15 shows reductions of .034, .043, and .024 without $\mathcal{J}_{\mathrm{nat}}$, $\mathcal{J}_{\mathrm{pc}}$, and $\mathcal{J}_{\mathrm{rank}}$, respectively. These component-wise changes show complementary contributions under the evaluated setting; they do not by themselves identify a causal mechanism for cross-subject drift.

Coefficient, margin, and graph-density sensitivity. For the coefficient surface, we vary $w_{\mathrm{pc}} \in \{0.20, 0.35, 0.50\}$ and $w_{\mathrm{rank}} \in \{0.04, 0.08, 0.12\}$ while setting $w_{\mathrm{nat}} = 1 - w_{\mathrm{pc}}$. We then vary the ranking margin $m \in \{0.10, 0.25, 0.50\}$ and graph multiplier $g \in \{1.0, 1.5, 2.0\}$ separately around the frozen setting, where $|\mathcal{E}| = \mathrm{round}(gq)$. All other settings remain fixed, and all values reported in this sensitivity analysis are computed from source-patient validation predictions.

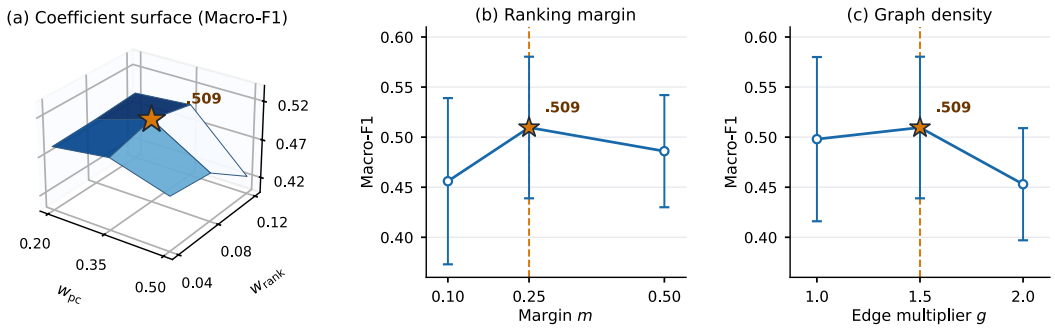

Figure 7: **Objective and structural sensitivity on EEGET-ALS-MI.** (a) Mean validation Macro-F1 over the joint patient-class and ranking-loss coefficient grid, with $w_{\text{nat}} = 1 - w_{\text{pc}}$; (b) ranking-margin sensitivity; and (c) graph-density sensitivity. All three panels report source-patient validation results. In panels (b) and (c), points show validation means and error bars show the corresponding standard deviations. Stars and dashed lines denote the frozen configuration $(w_{\text{pc}}, w_{\text{rank}}, m, g) = (0.35, 0.08, 0.25, 1.5)$, whose source-validation Macro-F1 is $.509 \pm .070$.

The frozen coefficient pair attains the highest observed mean in the declared grid. Increasing w_{pc} to .50 lowers Macro-F1 for every tested ranking weight, with the largest reduction at $w_{\text{rank}} = .12$, indicating that simultaneously strengthening both auxiliary contributions can displace natural-risk classification. For the ranking loss, $m = .10$ provides only weak separation pressure, whereas $m = .50$ imposes a stricter ordering constraint; the intermediate margin $m = .25$ performs best. Graph density shows a similar interior optimum: $g = 1.0$ restricts the number of component-pair interactions, whereas $g = 2.0$ adds weaker relations and produces the largest decrease. Thus $g = 1.5$ provides the best observed balance within the tested range. Because the error bars measure between-patient dispersion rather than confidence intervals, these diagnostics support the frozen configuration without implying statistical significance from error-bar overlap.

I Modern EEG Architecture and Cross-Subject Alignment

We extend the EEGET-ALS-MI comparison with two complementary baselines using the same six patient-level LOSO folds, fixed seeds, preprocessing, and patient-wise aggregation. EEG Conformer-style is our fold-compatible reimplementation of the convolution–attention mechanism in Song et al. (2023), with input and output dimensions adapted to EEGET-ALS-MI; the suffix "-style" makes clear that this is not an execution of the authors' official repository. EA+EEGNet applies Euclidean alignment before the B2 classifier: each patient's trials are left-multiplied by the inverse square root of that patient's mean covariance. The held-out patient's reference therefore uses unlabeled target trials, so this comparator is transductive and is reported separately from the source-only main matrix. QRG-EEG has the highest observed mean Macro-F1, balanced accuracy, and Cohen's κ in this supplementary comparison. EEG Conformer-style has the second-highest Macro-F1 under the same source-only boundary, whereas EA+EEGNet has the second-highest balanced accuracy and κ while using unlabeled held-out-patient covariates. These are descriptive mean comparisons over six patients; they show competitiveness under the evaluated settings rather than a general superiority claim.

J Full Related-Work Discussion

ALS-EEG decoding and cross-patient transfer. ALS BCI performance depends on more than motor progression: cognitive status, behavior, age, and signal quality can change both P300 and MI control (Riccio et al., 2013; Geronimo et al., 2016). Compact architectures therefore remain attractive in this low-sample regime. EEGNet controls capacity through temporal, depthwise, and separable convolutions; FBCNet adds filter-bank views and variance aggregation; and EEG Conformer and ATCNet combine convolutional priors with attention or temporal modeling (Lawhern et al., 2018; Mane et al., 2021; Song et al., 2023; Altaheri et al., 2023). LaBraM and CBraMod instead

Table 16: Supplementary comparison with a modern EEG architecture and cross-subject alignment on EEGET-ALS-MI. Values are mean ± sample standard deviation across six held-out patients after averaging the three fixed seeds within each patient. Bold indicates the highest mean and underlining the second-highest mean.

Model	Macro-F1	BAcc	κ
EEGNet (Lawhern et al., 2018)	$.344 \pm .106$	$.406 \pm .060$	$.126 \pm .096$
EEG Conformer-style (Song et al., 2023)	$\underline{.501 \pm .033}$	$.520 \pm .088$	$.293 \pm .102$
EA+EEGNet[†] (He & Wu, 2020; Lawhern et al., 2018)	$.492 \pm .063$	$\underline{.521 \pm .023}$	$\underline{.320 \pm .083}$
QRG-EEG	$\mathbf{.516 \pm .077}$	$\mathbf{.542 \pm .072}$	$\mathbf{.342 \pm .122}$

[†] EA+EEGNet estimates the held-out patient's Euclidean-alignment reference from unlabeled target trials. All other rows use no target-patient data during fitting, selection, or calibration.

transfer representations pretrained on broad EEG corpora (Jiang et al., 2024; Wang et al., 2025). These foundation models address data scarcity through external pretraining, whereas we study a compact source-cohort model and fit every transformation within each LOSO fold. Euclidean alignment and Riemannian recentering directly address subject shift and motivate our source-only fitting boundary (He & Wu, 2020; Junqueira et al., 2024; Zanini et al., 2018).

Riemannian and graph representations for EEG. Covariance matrices represent multichannel EEG on the SPD manifold, where geometry-aware means, distances, and tangent maps support classification and transfer (Barachant et al., 2012; Yger et al., 2017; Congedo et al., 2017). For P300 decoding, ERP-augmented covariance constructions retain cross-covariance with a target template and provide a stronger geometric reference than plain trial covariance (Barachant & Congedo, 2014). Geometric deep models extend this principle: CRGNet combines convolutional features with Riemannian processing; Tensor-CSPNet and Graph-CSPNet preserve structured SPD information; MAtt attends over manifold-valued representations; and TSMNet uses SPD domain-specific normalization for cross-domain adaptation (Gao et al., 2022; Ju & Guan, 2022; 2024; Pan et al., 2022; Kobler et al., 2022). EEG graph networks encode complementary neurophysiological or functional priors through regularized, local–global, or phase-synchrony graphs (Zhong et al., 2022; Ding et al., 2023; Patel et al., 2025b). Although deep graph convolution can over-smooth node representations (Li et al., 2018), we do not treat this as a universal limitation of graph learning. The defining difference is instead operational: QRG-EEG keeps node evidence and pairwise geometry separately addressable, and applies edge (i, j) directly to the interaction between the identically indexed qubits.

Quantum EEG learning and graph-conditioned circuits. QEEGNet processes a compressed EEGNet representation with a quantum layer (Chen et al., 2024). pyRiemann-qiskit and its ERP case study map covariance or tangent-space features into generic quantum classifiers (Andreev et al., 2023; Cattan & Quemy, 2023); these are the closest precedents at the intersection of EEG, Riemannian geometry, and quantum classification. They do not maintain a component-indexed graph in which each tangent relation controls its corresponding qubit pair. Beyond EEG, QGNNs (Verdon et al., 2019) encode adjacency through graph-structured quantum dynamics, equivariant quantum graph circuits (Mernyei et al., 2022) formalize permutation-aware architectures, and weighted-graph circuits condition interactions on continuous edge values (Skolik et al., 2023). QRG-EEG builds on this family rather than claiming the generic node-to-qubit or edge-to-interaction idea. Repeated node/edge encoding across the two fixed blocks is related to data re-uploading, but is not isolated as a separate empirical contribution (Pérez-Salinas et al., 2020).

Evaluation of variational quantum models. Parameter-shift rules enable analytic circuit gradients (Schuld et al., 2019), but trainability does not by itself demonstrate a useful or uniquely quantum model: barren plateaus, efficient simulation regimes, and classical approximations remain relevant concerns (Larocca et al., 2025; Cerezo et al., 2025). Random-Fourier and model-derived surrogates further show why a quantum model should be compared with controls that receive equivalent information and capacity (Landman et al., 2023; Schreiber et al., 2023). We therefore compare QRG-EEG with embedding-only quantum, vectorized Riemannian-quantum, same-graph, and parameter-matched classical alternatives, and restrict conclusions to predictive performance under exact simulation.