MULTI-VIEW HYBRID QUANTUM LEARNING WITH CLASS-CONDITIONAL SOURCE ALIGNMENT FOR CROSS-SUBJECT EEG

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-subject electroencephalography (EEG) learning combines few independent participants, heterogeneous task information, and distribution shift between people. Existing hybrid quantum EEG models compress all features before a single variational quantum circuit (VQC), which discards view-specific structure and leaves subject shift unaddressed. We introduce MVQAN-CORAL, a source-only architecture that preserves temporal, spatial, spectral, statistical, covariance, and task-dependent geometric views. Each view is processed by an independent shallow VQC and a lightweight residual encoder, a VQC-conditioned distribution weights both branches, and the fused quantum representation is regularized by class-conditional source-moment alignment and class-center separation. Under subject- or patient-disjoint evaluation on BNCI-P300, UCL ALS-EEG, and EEGET-ALS, MVQAN-CORAL obtains Macro-F1 scores of $86.5\pm0.5\%$, $66.8\pm0.5\%$, and $86.6\pm0.4\%$, exceeding the strongest of seven reimplemented architectures by 5.8, 2.0, and 11.0 points. Replacing each VQC by a LayerNorm–MLP matched to within 1% of the parameter count lowers Macro-F1 by 4.7, 3.1, and 3.0 points. Objective ablations identify conditional alignment as the largest individual contributor on BNCI and UCL, and the matched substitution isolates an additional encoder-specific increment within the aligned system. The circuits are analytically simulated bounded trigonometric feature maps; we claim no computational or hardware quantum advantage, and reported deviations quantify three-seed optimization variability under fixed outer folds.

1 INTRODUCTION

Subject-independent EEG models must generalize from repeated measurements of a few people to a person absent from fitting and selection. The number of trials can be large while the number of independent domains remains small, allowing a model to exploit participant-specific structure. We consider three binary settings with different prediction units: P300 target detection, left-versus-right motor-imagery decoding in people with amyotrophic lateral sclerosis (ALS), and window-level ALS-status prediction under subject-disjoint evaluation. The third setting is not a BCI control task or a clinical diagnostic endpoint. Task-aware pipelines remain competitive in this regime because xDAWN, common spatial patterns (CSP), filter banks, and Riemannian covariance geometry impose useful structure before learning (Rivet et al., 2009; Muller-Gerking et al., 1999; Ang et al., 2012; Barachant et al., 2012). Compact neural networks can learn complementary representations, while domain-generalization objectives reduce discrepancies among source subjects. These observations motivate a model that preserves distinct EEG views and treats each source participant as a domain. Near-term hybrid quantum models use classical preprocessing followed by trainable parameterized circuits (Preskill, 2018; Benedetti et al., 2019; Cerezo et al., 2021). Existing EEG variants typically apply one VQC after global compression (Olvera et al., 2024; Chen et al., 2024). Such a circuit cannot recover distinctions already removed between ERP timing, regional activity, band power, trial statistics, and covariance geometry. A VQC also does not, by itself, correct class-conditional subject shift.

We propose MVQAN-CORAL (Figure 1) as an integration of view-preserving representation learning and source-only conditional alignment. The VQC is treated as a structured encoder inside

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.

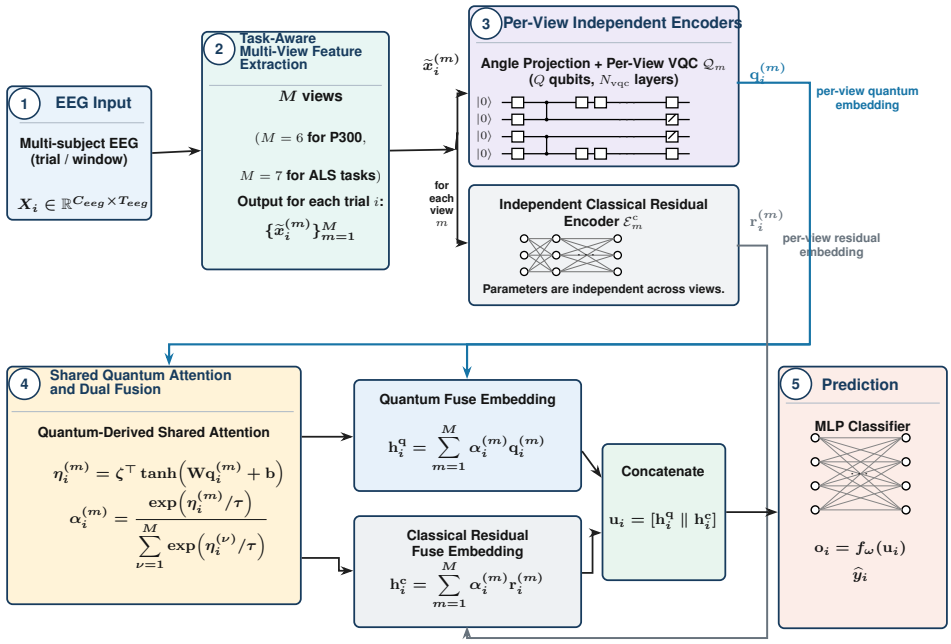


Figure 1: Overview of MVQAN-CORAL. Source-fitted transformations produce complementary EEG views. Each view enters an independent variational circuit and residual encoder. A distribution computed from the VQC embeddings weights both branches. Classification uses their concatenation; class-conditional alignment and separation act only on the fused quantum representation during source training. No held-out subject contributes to preprocessing, fitting, checkpointing, threshold selection, or architecture selection. The block labeled as shared attention denotes the VQC-conditioned classical softmax weighting of Eq. 3.

this system rather than as the sole source of performance. Our objective ablation identifies alignment as the largest individual contributor on two datasets, whereas a matched encoder substitution measures the additional VQC-associated increment under otherwise identical conditions.

The contribution is threefold. First, we formulate a task-aware hybrid architecture that preserves view identity through independent shallow VQCs and a residual path, with one VQC-conditioned view distribution shared across both branches. Second, we align sample-weighted class moments across source subjects after fusion and combine this with class-center regularization. Third, we separate complete-system comparison from component attribution through subject-disjoint evaluation, an objective decomposition, leave-one-view-out analysis, and a below-1% parameter-gap encoder control. The evidence supports the integrated design and an architecture-specific VQC increment; it does not establish quantum advantage or the necessity of a VQC over every structured classical map.

2 Related Work

Structured and cross-subject EEG learning. CSP, xDAWN, filter-bank methods, and tangent-space classifiers encode task-specific spatial or covariance structure (Muller-Gerking et al., 1999; Ramoser et al., 2000; Ang et al., 2012; Rivet et al., 2009; Barachant et al., 2012). EEGNet, ATCNet, LMDA-Net, and EEG Conformer learn compact or attention-based spatiotemporal representations (Lawhern et al., 2018; Altaheri et al., 2023; Miao et al., 2023; Song et al., 2023). Source-only domain generalization instead learns from several labeled source domains and evaluates an unseen domain (Ben-David et al., 2010; Gulrajani & Lopez-Paz, 2021). EEG-DG aligns marginal and conditional source distributions (Zhong et al., 2023); DMMR uses denoising mutual reconstruction for cross-subject emotion recognition (Wang et al., 2024); and State Mamba aligns spatiotemporal EEG states and transitions (Weng et al., 2026). The latter methods use different tasks and inputs, so we position them conceptually rather than treating their published numbers as directly comparable.

Table 1: Task-aware views. The first five are common; P300 adds xDAWN, whereas UCL and EEGET add both tangent and CSP views.

View	Construction	Intended information
Temporal	Statistics over four time windows	ERP or trial evolution
Spatial	Region–window channel summaries	Coarse spatial organization
Spectral	Log power in task-specific FFT bands	Rhythmic activity
Task-statistical	ERP amplitudes or channel log variance	Paradigm-specific summaries
Covariance	Regularized, trace-normalized covariance	Cross-channel dependence
xDAWN tangent	Source-fitted xDAWN covariance geometry	P300 enhancement
Riemannian tangent	Source-fitted tangent coordinates	Covariance geometry
CSP log variance	Source-fitted CSP filters	Discriminative variance

CORAL and conditional source alignment. CORrelation ALignment (CORAL) reduces domain discrepancy by matching second-order feature statistics (Sun & Saenko, 2016). Standard CORAL is usually marginal: it does not require same-class observations from different domains to share moments. In our setting, each source subject is a domain and labels are available for source training. We therefore match both the mean and covariance of each valid subject–class group to a pooled reference for that class. This avoids aligning different classes merely because their marginal subject distributions differ.

Hybrid quantum representation learning. Parameterized circuits yield trainable expectation-value maps and implicit feature spaces (Benedetti et al., 2019; Cerezo et al., 2021; Schuld & Killoran, 2019). QEEGNet and the Olvera hybrid model apply VQCs to globally compressed EEG inputs (Chen et al., 2024; Olvera et al., 2024). MVQAN-CORAL instead applies separate circuits to semantically distinct views and evaluates the encoder through a controlled replacement. Because finite rotation circuits induce Fourier-like functions (Schuld et al., 2021; Perez-Salinas et al., 2020), a generic MLP is not an exhaustive classical analogue (Section 6).

3 Method

3.1 Problem Setup and Task-Aware Views

Let $\mathcal{D} = \{(\mathbf{X}_i, y_i, s_i)\}_{i=1}^{J}$, where $\mathbf{X}_i \in \mathbb{R}^{C_{\mathrm{eeg}} \times T_{\mathrm{eeg}}}$, $y_i \in \{0, 1\}$, and s_i identifies a participant domain. In each outer fold, source-training, source-validation, and held-out test domains $\mathcal{S}_{\mathrm{tr}}, \mathcal{S}_{\mathrm{va}}, \mathcal{S}_{\mathrm{te}}$ are disjoint. Every fitted transform, including standardization, PCA, xDAWN, CSP, and tangent-space mapping, uses $\mathcal{S}_{\mathrm{tr}}$ only.

For view $m \in \{1, \ldots, M\}$,

$$\mathbf{x}_i^{(m)} = \phi_m(\mathbf{X}_i), \qquad \widetilde{\mathbf{x}}_i^{(m)} = \Psi_m(\mathbf{x}_i^{(m)}), \tag{1}$$

where ϕ_m constructs the view and Ψ_m is its source-fitted scaling/PCA map. Table 1 makes the central inputs explicit; complete bands, windows, covariance regularization, and dimensionality rules appear in Appendix D.

3.2 View-Specific Encoders and Shared Weighting

A learned projection maps each processed view to $2Q$ bounded rotation angles,

$$\mathbf{a}_i^{(m)} = \pi \tanh\Big(g_m(\widetilde{\mathbf{x}}_i^{(m)})\Big) \in [-\pi, \pi]^{2Q}. \tag{2}$$

The two halves parameterize RY and RZ data rotations. The view-specific circuit $\mathcal{Q}_m(\mathbf{a}_i^{(m)}; \boldsymbol{\theta}_m)$ applies trainable RY–RZ layers and ring-CNOT entanglement, followed by all-qubit Pauli-Z measurements and finite classical post-processing, to produce $\mathbf{q}_i^{(m)} \in \mathbb{R}^Q$. An independent residual encoder $\mathcal{E}_m^{\mathrm{c}}$ produces $\mathbf{r}_i^{(m)} \in \mathbb{R}^{D_{\mathrm{c}}}$.

The residual path is intended to preserve information compressed by the low-qubit encoder, not to learn a second unrelated view selector. We therefore compute one distribution from the VQC

embeddings and reuse it for both branches:

$$\eta_i^{(m)} = \boldsymbol{\zeta}^\top \tanh(\mathbf{W}\mathbf{q}_i^{(m)} + \mathbf{b}), \qquad \alpha_i^{(m)} = \frac{\exp(\eta_i^{(m)}/\tau)}{\sum_{\nu=1}^{M} \exp(\eta_i^{(\nu)}/\tau)}, \tag{3}$$

$$\mathbf{h}_i^{\mathrm{q}} = \sum_m \alpha_i^{(m)} \mathbf{q}_i^{(m)}, \qquad \mathbf{h}_i^{\mathrm{c}} = \sum_m \alpha_i^{(m)} \mathbf{r}_i^{(m)}, \tag{4}$$

$$\mathbf{u}_i = [\mathbf{h}_i^{\mathrm{q}} \| \mathbf{h}_i^{\mathrm{c}}], \qquad \mathbf{o}_i = f_\omega(\mathbf{u}_i) \in \mathbb{R}^2. \tag{5}$$

We call this mechanism *VQC-conditioned shared view weighting*; it is a classical softmax scorer whose inputs are quantum expectation embeddings, not a quantum attention primitive.

3.3 Class-Conditional Moment Alignment

Classification alone leaves the fused representation free to encode subject identity. To reduce this source-domain variation without observing the held-out participant, we align same-class source moments. For mini-batch $\mathcal{B}$, define $\mathcal{B}_{s,k} = \{i \in \mathcal{B} : s_i = s, y_i = k\}$, $\mathcal{B}_k = \{i \in \mathcal{B} : y_i = k\}$, and retain $\mathcal{V}_\mathcal{B} = \{(s,k) : |\mathcal{B}_{s,k}| \geq 2\}$. The pooled moments $\bar{\boldsymbol{\mu}}_k, \bar{\boldsymbol{\Sigma}}_k$ are computed from all samples in $\mathcal{B}_k$; hence they are sample-weighted, not equal-weight averages of subject moments. With subject–class moments $\boldsymbol{\mu}_{s,k}, \boldsymbol{\Sigma}_{s,k}$ measured on $\mathbf{h}_i^{\mathrm{q}}$,

$$\mathcal{L}_{\mathrm{cor}} = \frac{1}{|\mathcal{V}_\mathcal{B}|} \sum_{(s,k)\in\mathcal{V}_\mathcal{B}} \left[\frac{\|\boldsymbol{\mu}_{s,k} - \bar{\boldsymbol{\mu}}_k\|_2^2}{Q} + \frac{\|\boldsymbol{\Sigma}_{s,k} - \bar{\boldsymbol{\Sigma}}_k\|_F^2}{Q^2} \right]. \tag{6}$$

Unlike marginal CORAL, Eq. 6 does not force different labels toward one shared distribution. Empty valid-group sets contribute zero. The pooled reference is sample-weighted, whereas the outer sum assigns equal weight to every valid subject–class group.

3.4 Class Geometry and Training

Let $\widehat{\mathbf{h}}_i^{\mathrm{q}} = \mathbf{h}_i^{\mathrm{q}} / \max(\|\mathbf{h}_i^{\mathrm{q}}\|_2, \epsilon_{\mathrm{n}})$ with a small constant $\epsilon_{\mathrm{n}} > 0$; let $\mathcal{K}_\mathcal{B}$ be the classes with at least two samples in $\mathcal{B}$, $\boldsymbol{\vartheta}_{k,\mathcal{B}}$ the mean normalized embedding of class $k \in \mathcal{K}_\mathcal{B}$, and $\mathcal{P}_\mathcal{B}$ the unordered pairs of distinct classes in $\mathcal{K}_\mathcal{B}$. The separation loss combines within-class compactness with a squared hinge between valid class centers:

$$\mathcal{L}_{\mathrm{sep}} = \frac{1}{|\mathcal{K}_\mathcal{B}|} \sum_{k\in\mathcal{K}_\mathcal{B}} \frac{1}{|\mathcal{B}_k|} \sum_{i\in\mathcal{B}_k} \|\widehat{\mathbf{h}}_i^{\mathrm{q}} - \boldsymbol{\vartheta}_{k,\mathcal{B}}\|_2^2 + \frac{1}{|\mathcal{P}_\mathcal{B}|} \sum_{(k,k')\in\mathcal{P}_\mathcal{B}} \left[\max(0, \delta - \|\boldsymbol{\vartheta}_{k,\mathcal{B}} - \boldsymbol{\vartheta}_{k',\mathcal{B}}\|_2)\right]^2. \tag{7}$$

With $\varpi_{i,k} = \mathrm{softmax}(\mathbf{o}_i)_k$, the classification term is the class-weighted focal loss (Lin et al., 2017)

$$\varepsilon_i = -\rho_{y_i} \log \varpi_{i,y_i}, \qquad \mathcal{L}_{\mathrm{cls}} = \frac{1}{|\mathcal{B}|} \sum_{i\in\mathcal{B}} (1 - e^{-\varepsilon_i})^\gamma \varepsilon_i, \tag{8}$$

and the complete objective is

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{cls}} + \lambda_{\mathrm{cor}}\mathcal{L}_{\mathrm{cor}} + \lambda_{\mathrm{sep}}\mathcal{L}_{\mathrm{sep}}. \tag{9}$$

Training follows one fixed sequence: construct the outer domains, fit every transform on source training, optimize Eq. 9, choose the checkpoint and probability threshold on source validation Macro-F1, and evaluate the held-out domain once. Appendix E specifies corner cases, widths, coefficients, and optimization. The executed model uses $\tau = 0.75$, a 12-unit view scorer, a $22 \to 24 \to 2$ classifier, $\lambda_{\mathrm{cor}} = 0.01$, $\lambda_{\mathrm{sep}} = 0.015$, and $\delta = 0.75$; these values are fixed across the three tasks.

Selection of fixed coefficients. We selected the scalar coefficients once using only source-training and source-validation domains and then froze them for all three datasets; no held-out domain entered this choice. Candidate configurations were ranked by mean source-validation Macro-F1 under the same domain-disjoint protocol used for checkpointing. The temperature τ controls the entropy of the

Table 2: Prediction tasks and independent outer units. T/NT denotes target/non-target and L/R denotes left/right. Dataset counts and fixed pair construction are detailed in Appendix C.

Dataset	Eligible cohort	Folds	Held-out unit	Prediction and fold aggregation
BNCI	10 participants; T/NT P300	10	One participant	Epoch metrics; equal mean over LOSO folds
UCL	8 ALS patients; L/R imagery	8	One patient	Trial metrics; equal mean over LOSO folds
EEGET	6 ALS + 170 controls; status	6	One ALS + one fixed control	2-s-window metrics; equal mean over pair-out folds

shared view distribution, while λ_{cor} and λ_{sep} scale moment alignment and class geometry relative to classification. Because separation operates on normalized embeddings, whose pairwise distance is bounded by 2, $\delta = 0.75$ specifies a moderate center margin. Base inverse-frequency weights are $\rho_k^{(0)} = |\mathcal{D}_{\mathrm{tr}}|/(2|\mathcal{D}_{\mathrm{tr},k}|)$; we use $\rho_0 = \rho_0^{(0)}$, $\rho_1 = 1.25\rho_1^{(0)}$, and $\gamma = 2$ to emphasize difficult positive examples after fold-specific class balancing.

Mini-batch validity. A subject–class group enters Eq. 6 only with at least two batch samples, so its covariance is defined, and class centers follow the same rule; subject-aware stratification increases the number of valid groups without using validation or test data. An auxiliary term is zero when its valid set is empty, so classification remains well defined.

3.5 Representation Analysis

We characterize the implemented encoder and the alignment objective; proofs are in Appendix G.

Proposition 1 (Structured VQC view map)**.** *For fixed trainable parameters $\boldsymbol{\theta}_m$, every Pauli-Z expectation of $\mathcal{Q}_m$ is a finite multivariate trigonometric polynomial in $\mathbf{a}_i^{(m)}$ with values in $[-1, 1]$, and the ring-CNOT entanglement can create products of angles assigned to different qubits.*

Because the weights in Eq. 3 are nonnegative and sum to one, each fused embedding lies in the convex hull of its per-view embeddings. Consequently, $\|\mathbf{h}_i^{\mathrm{q}}\|_2 \le \max_m \|\mathbf{q}_i^{(m)}\|_2$; the number of views alone cannot inflate this norm bound (Lemma 1), although an added view can change the convex hull.

Assumption 1 (Local margin regularity)**.** *With the residual branch fixed, the binary logit margin is locally $\chi(\mathbf{z}) = \boldsymbol{\beta}^\top \mathbf{z} + \mathbf{z}^\top \mathbf{G}_\chi \mathbf{z} + \kappa$ in the fused quantum representation $\mathbf{z} = \mathbf{h}^{\mathrm{q}}$, where $\mathbf{G}_\chi = \mathbf{G}_\chi^\top$ and the class-mean norms are bounded by Ω_μ.*

Proposition 2 (Moment-controlled margin drift)**.** *Under Assumption 1, for a source subject–class distribution $P_{s,k}$ and the pooled class reference $\bar{P}_k$,*

$$\left|\mathbb{E}_{P_{s,k}}\chi - \mathbb{E}_{\bar{P}_k}\chi\right| \le \left(\|\boldsymbol{\beta}\|_2 + 2\Omega_\mu\|\mathbf{G}_\chi\|_2\right)\|\boldsymbol{\mu}_{s,k} - \bar{\boldsymbol{\mu}}_k\|_2 + \|\mathbf{G}_\chi\|_F\,\|\boldsymbol{\Sigma}_{s,k} - \bar{\boldsymbol{\Sigma}}_k\|_F.$$

Here $\boldsymbol{\mu}_{s,k}, \boldsymbol{\Sigma}_{s,k}$ and $\bar{\boldsymbol{\mu}}_k, \bar{\boldsymbol{\Sigma}}_k$ are the moments of $P_{s,k}$ and $\bar{P}_k$, whose mini-batch estimates enter Eq. 6. The right-hand side contains the mean and covariance discrepancies whose Q-normalized squared values are penalized by Eq. 6. Holding the local margin coefficients fixed, reducing these squared discrepancies tightens the displayed source-domain drift bound. The result does not control risk on an unseen participant: it characterizes the optimization target and explains why class-conditional alignment is more specific than marginal CORAL. Neither proposition implies computational or statistical quantum advantage.

4 Experiments

4.1 Tasks, Outer Evaluation, and Aggregation

Aggregation and scope. Table 2 lists the prediction task, held-out unit, and fold aggregation of each dataset. BNCI and UCL enumerate all eligible people. EEGET uses six fixed ALS-indexed pair-out folds because one held-out person would yield a single-class test set; every ALS participant appears once with a fixed control slot. Metrics are computed per fold and averaged with equal fold

weight, so repeated epochs or windows cannot increase a person's weight. Each seed yields one full-registry aggregate, and the displayed deviation is then computed across three seeds. It measures optimization variation under fixed people, not population uncertainty. EEGET remains window classification, not a participant-level diagnosis.

BNCI2014-009 is loaded through MOABB (BNCI Horizon 2020, 2014; Jayaram & Barachant, 2018); UCL retains left/right imagery from eight ALS patients (Patel et al., 2025); and EEGET uses EEG-only, non-overlapping 2-s windows capped at 160 per participant (EEGET-ALS Dataset, n.d.; Ngo et al., 2024). Validation domains are selected only from the remaining sources. Appendix C gives preprocessing, counts, pair construction, and leakage checks.

Metric computation and thresholding. For each seed and outer fold, a single decision threshold $t^\star$ is selected by source-validation Macro-F1. The same thresholded predictions are used to compute Macro-F1, balanced accuracy, and accuracy; no metric-specific threshold is fitted. AUROC and AUPRC use the corresponding unthresholded class-1 probabilities. Each metric is first computed within a held-out fold, averaged with equal fold weight to form one seed-level aggregate, and finally summarized across the three seeds.

4.2 Comparators, Controls, and Optimization

We reimplement seven published architectures under the same folds, seeds, metric code, stopping rule, and source-validation threshold selection: EEGNet, ATCNet, EEG-DG, QEEGNet, the Olvera hybrid VQC, LMDA-Net, and EEG Conformer (Lawhern et al., 2018; Altaheri et al., 2023; Zhong et al., 2023; Chen et al., 2024; Olvera et al., 2024; Miao et al., 2023; Song et al., 2023). Their numbers are task-adapted reproduction results, not values copied from the original papers. QEEGNet and Olvera retain much smaller, globally compressed architectures; comparisons with them assess complete systems and cannot isolate encoder type. Appendix H documents every adaptation and tuning budget.

The attribution control, PM-Classical, changes only the principal per-view encoder: each VQC is replaced by a LayerNorm–MLP–LayerNorm map while views, residual encoders, shared weighting, objectives, classifier, folds, and optimization remain fixed. Deterministic width matching leaves parameter gaps of 0.72%, 0.63%, and 0.79%; Appendix J gives the complete five-metric comparison.

The full model uses $Q = 6$, two variational layers, residual width $D_{\mathrm{c}} = 16$, AdamW, and seeds $\{0, 1, 2\}$. Circuits are simulated with analytic Pauli expectations on PennyLane 0.45.1 `default.qubit` through the PyTorch interface, using the default simulator differentiation (`diff_method="best"`) without finite-shot estimation. The 4×3 qubit–depth grid is a separate sensitivity study (Appendix M), and Appendix F gives the full optimization and selection policy.

5 Results

Complete-system comparison. In Table 3(a), MVQAN-CORAL exceeds the strongest reimplemented comparator by 5.8 Macro-F1 points on BNCI, 2.0 on UCL, and 11.0 on EEGET, and Appendix I reports the same ordering for balanced accuracy, accuracy, AUROC, and AUPRC. These are comparisons between complete systems that differ in preprocessing, capacity, and objectives, so the gaps to QEEGNet and Olvera cannot be attributed to the VQC. Figure 2 relates performance to model size: MVQAN-CORAL uses fewer parameters than ATCNet, EEG Conformer, and EEG-DG, but more than EEGNet, QEEGNet, Olvera, and LMDA-Net.

Controlled encoder comparison. Table 3(b) replaces only the principal per-view encoder. The VQC version improves Macro-F1 over its matched MLP by 4.7, 3.1, and 3.0 points and is higher on every displayed metric, at parameter gaps below 1%. The comparison is specific to the implemented MLP and does not cover periodic or trigonometric classical maps.

Interaction between alignment and representation. The gain from conditional alignment depends on the representation it regularizes (Appendix K). Adding the same alignment to a classical multi-view attention model changes Macro-F1 by 0.2 points on BNCI (66.2 to 66.4) and by 2.2 on UCL (54.8 to 57.0), whereas adding $\mathcal{L}_{\mathrm{cor}}$ to the classification-only MVQAN-CORAL objective adds 17.5 and

Table 3: Primary predictive comparisons (%). Values are mean ± standard deviation over three seed-level aggregates under fixed outer folds. Panel (a) compares complete reimplemented systems; panel (b) changes only the principal encoder.

(a) Reimplemented-architecture Macro-F1 and parameter count

Model	BNCI	UCL	EEGET	Params (B/U/E)
EEGNet	77.5±0.6	64.8±0.7	63.9±0.7	1.5k/1.6k/1.2k
ATCNet	77.7±0.5	62.0±0.7	56.0±0.8	36.9k/37.1k/36.1k
EEG-DG	74.0±0.7	63.0±0.7	68.6±0.6	933.8k/934.0k/933.3k
QEEGNet	68.9±0.6	60.7±0.7	65.2±0.6	5.4k/5.5k/5.2k
Olvera Hybrid VQC	66.6±0.7	50.5±1.1	33.3±0.7	2.0k/2.2k/1.9k
LMDA-Net	79.9±0.6	63.5±0.8	73.1±0.7	∼6.8k/∼6.9k/∼5.9k
EEG Conformer	80.7±0.6	64.3±0.7	75.6±0.6	∼54k/∼55k/∼48k
MVQAN-CORAL	**86.5±0.5**	**66.8±0.5**	**86.6±0.4**	18.8k/21.0k/16.7k

(b) Parameter-matched encoder substitution

Dataset	Encoder	Macro-F1	BAcc	AUROC	AUPRC	Params
BNCI	MLP	81.8±0.7	82.5±0.7	82.0±0.6	71.4±0.8	18,640
	VQC	**86.5±0.5**	**87.0±0.6**	**84.2±0.5**	**74.4±0.8**	18,775
UCL	MLP	63.7±0.8	64.3±0.8	71.9±0.8	74.1±0.8	20,820
	VQC	**66.8±0.5**	**67.4±0.8**	**74.8±0.7**	**75.9±0.7**	20,951
EEGET	MLP	83.6±0.6	84.1±0.6	89.8±0.5	91.5±0.5	16,520
	VQC	**86.6±0.4**	**86.9±0.6**	**92.0±0.4**	**94.2±0.3**	16,651

11.4 points. A single VQC applied after global PCA compression reaches 54.9, 48.4, and 73.3, which is 31.6, 18.4, and 13.3 points below the full model; with 486 parameters, this baseline also differs in capacity, so it bounds rather than isolates the value of per-view encoding. Proposition 2 concerns moment discrepancies within a given representation; these results show that the empirical benefit of reducing them depends strongly on that representation.

Objective ablation. Conditional alignment is the largest single addition on BNCI and UCL (Table 4): relative to classification alone, it adds 17.5 and 11.4 points, compared with 6.7 and 5.8 from separation. On EEGET the gains are 2.3 and 1.6. The full objective adds a further 2.6, 2.9, and 1.6 points beyond alignment alone, so separation complements rather than replaces alignment. The classification-only model is 11.1 and 12.3 points below EEGNet on BNCI and UCL. Together with the encoder control, this indicates that alignment is needed for the strongest BNCI and UCL results and that, within the aligned system, the VQC outperforms the matched MLP.

Table 4: **Objective ablation** (Macro-F1, %). Architecture, views, encoder, and evaluation protocol are fixed; only the active loss terms change.

Objective	BNCI	UCL	EEGET
$\mathcal{L}_{\text{cls}}$	66.4±0.8	52.5±1.0	82.7±0.6
$+\mathcal{L}_{\text{cor}}$	83.9±0.6	63.9±0.7	85.0±0.5
$+\mathcal{L}_{\text{sep}}$	73.1±0.7	58.3±0.8	84.3±0.5
Full	**86.5±0.5**	**66.8±0.5**	**86.6±0.4**

Views and circuit capacity. Removing any view lowers Macro-F1 (Appendix L). The largest drops are 6.5 points for xDAWN on BNCI, 6.5 for CSP on UCL, and 5.6 for Riemannian tangent coordinates on EEGET, and even the least influential view costs at least 2.4 points. In the qubit–depth grid (Figure 3), six qubits and two layers give the highest mean on every dataset; eight qubits lower Macro-F1 by 0.6, 0.8, and 0.5 points, and a third layer lowers it by 0.9, 1.1, and 0.6 points. Each grid cell is a separate retraining run, so its six-qubit, depth-two means (85.6, 66.1, 85.7) differ slightly from the primary estimates; we treat the grid as descriptive sensitivity evidence (Appendix M).

Dataset-specific observations. On BNCI, non-targets dominate, so the Macro-F1 gain exceeds the accuracy gain, and removing the xDAWN or task-statistical view costs 6.5 or 5.4 points, tying the result to ERP structure. UCL gives the smallest margin (2.0 points over EEGNet) on eight patients; with three seeds, this margin does not establish a population-level ranking. On EEGET, the model predicts the inherited status label of 2-s windows in six held-out control–ALS pairs, so the 11.0-point margin concerns window classification rather than diagnosis. Without alignment, removing the residual branch collapses EEGET Macro-F1 to 34.6 (Table 16), which shows that the residual path is essential when the quantum representation is not regularized.

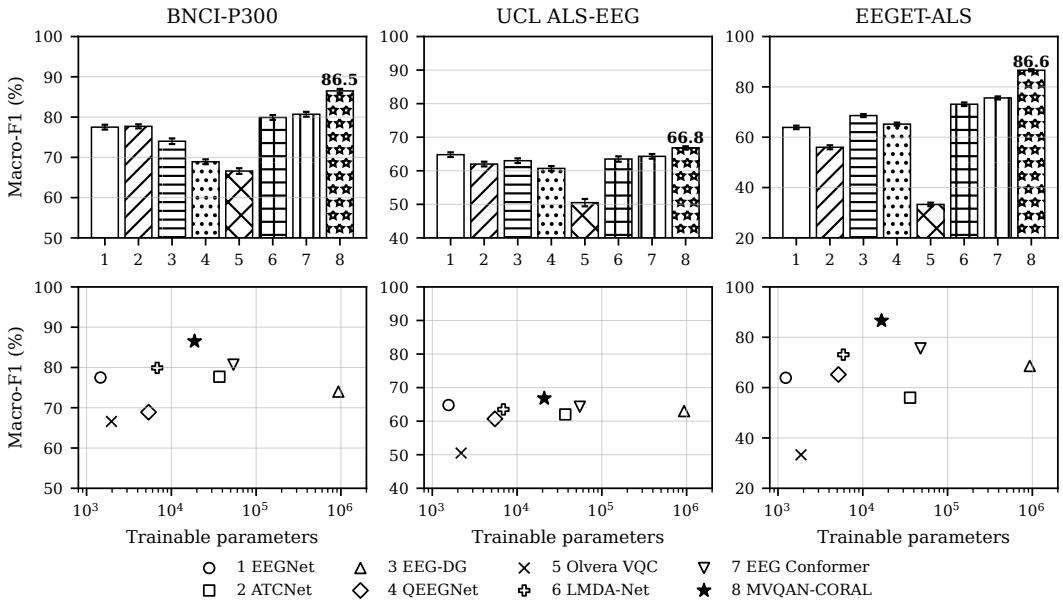


Figure 2: Complete-system context. Top: Macro-F1 (mean ± SD over three seeds) of the reimplemented architectures. Bottom: Macro-F1 versus trainable parameters on a logarithmic axis. Both rows compare complete pipelines; encoder attribution comes only from Table 3(b).

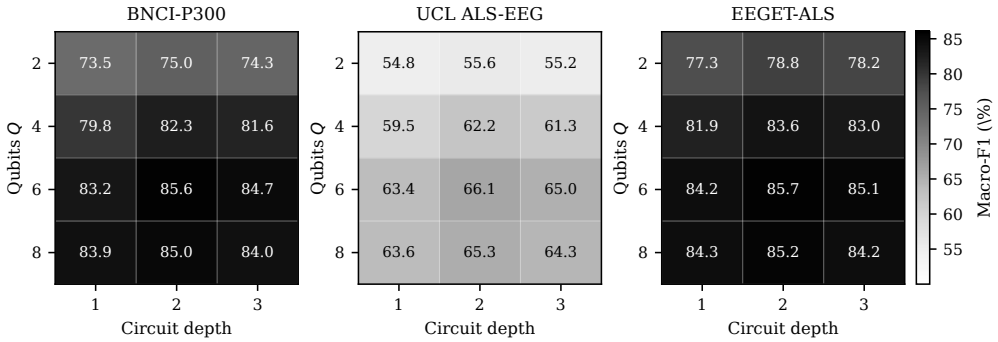


Figure 3: Controlled Macro-F1 sensitivity to qubit count Q and variational depth N_{vqc}, with all other model components fixed. Within the tested grid, six qubits and two variational layers give the highest mean on each dataset. Each cell is a separate retraining run; the grid is descriptive and does not replace the frozen primary estimates.

Embedding geometry. Figure 4 compares held-out fused quantum embeddings with and without alignment. BNCI and UCL show more coherent local organization in the displayed aligned folds, whereas EEGET shows heterogeneous pair-specific effects; t-SNE panels are qualitative (van der Maaten & Hinton, 2008), and the concatenated representation is shown in Appendix N. In the displayed BNCI fold, alignment raises balanced accuracy from 67.05% to 87.03% and accuracy from 80.96% to 86.79%: the target epochs that overlap the non-target mass without alignment form a separate region, and the remaining errors concentrate along its boundary, whereas in the displayed UCL fold accuracy rises from 55.31% to 67.35% but the two imagery classes still overlap substantially, consistent with the smallest aggregate margin in Table 3(a). In the displayed EEGET pair, both representations already separate the two classes and alignment slightly lowers fold accuracy (88.75% without alignment), so the 1.9-point aggregate EEGET gain from alignment (84.7 to 86.6 Macro-F1; Table 16) is not uniform across held-out pairs.

Threshold-dependent and ranking metrics. On BNCI, where targets form one sixth of the epochs, MVQAN-CORAL exceeds the best reimplemented accuracy by 1.0 point (86.8 versus 85.8) but the best balanced accuracy by 5.7 points (87.0 versus 81.3), so its Macro-F1 gain reflects improved minority-class recall at the source-selected threshold; the threshold-free AUROC and AUPRC gains

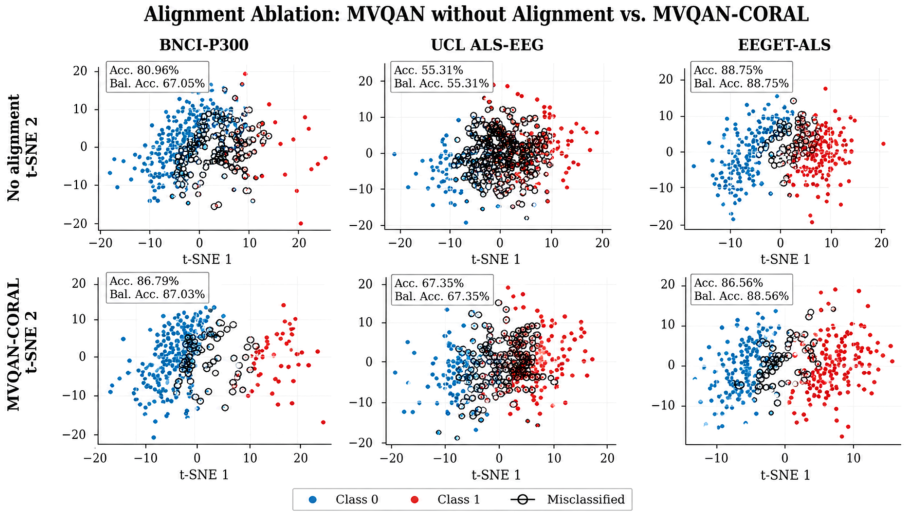


Figure 4: Held-out t-SNE diagnostics of the fused quantum representation, comparing MVQAN without alignment (top) with MVQAN-CORAL (bottom). Annotations are fold-specific. The plots are qualitative; aggregate results use all participant-disjoint outer folds.

are 1.9 points each (Appendix I). On EEGET, the AUROC and AUPRC gains over the strongest comparator are 6.7 and 10.0 points, so the margin is not a thresholding effect. The Olvera model illustrates the opposite case: its EEGET Macro-F1 of 33.3 and balanced accuracy of 50.0 correspond to near-single-class predictions at the selected threshold, although its AUROC of 66.9 indicates above-chance ranking.

6 Limitations

Our evidence is architecture-specific: within the aligned multi-view system, per-view VQCs outperform the parameter-matched LayerNorm–MLP control. This control does not cover trigonometric or random-Fourier maps, and the complete-system comparison omits same-fold MDM, tangent-space LDA/SVM, and FBCSP pipelines. QEEGNet and Olvera remain smaller, globally compressed systems, while DMMR and State Mamba address other tasks; none isolates encoder type. We therefore restrict encoder attribution to the implemented VQC–MLP substitution. The registry includes ten BNCI participants, eight UCL patients, and six EEGET ALS participants, so three-seed deviations measure optimization variability over fixed folds rather than population uncertainty (Appendix P). BNCI results are epoch-level without character accuracy or information-transfer rate; EEGET results are 2-s-window-level without participant aggregation or diagnostic interpretation. Finally, analytic simulation and heterogeneous compute conditions support no runtime claim (Appendix O); finite-shot sampling, device noise, and hardware execution remain untested.

7 Conclusion

MVQAN-CORAL combines task-aware view preservation, independent shallow VQCs, VQC-conditioned view weighting shared with a lightweight residual pathway, and source-only class-conditional regularization. Across the three evaluated subject-disjoint EEG tasks, it attains the highest mean Macro-F1 among seven reimplemented architectures and exceeds PM-Classical on every reported metric. The objective, leave-one-view-out, and qubit–depth analyses indicate that the gains arise from the interaction of view-wise encoding, conditional source alignment, class separation, and residual information preservation rather than from parameter count alone, and Proposition 2 links the alignment objective to the source-domain margin drift it controls. Validation on larger clinical cohorts, participant-level endpoints, structured classical controls, and finite-shot quantum execution are the next steps.

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.

Reproducibility statement. Appendix A provides the supplementary roadmap and Appendix B consolidates notation. Appendices C–R specify preprocessing, source-only fitting, dimensions, objectives, coefficient selection, optimization, differentiation, folds, parameter matching, metrics, ablations, compute conditions, and reporting. The code and fixed fold registry reproduce the reported comparisons.

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.

Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in Neuroscience*, 6:39, 2012. doi: 10.3389/fnins.2012.00039.

Alexandre Barachant, Stephane 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.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79(1–2):151–175, 2010. doi: 10.1007/s10994-009-5152-4.

Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, 2019. doi: 10.1088/2058-9565/ab4eb5.

Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, M. Sohaib Alam, Shahnawaz Ahmed, Juan Miguel Arrazola, Carsten Blank, Alain Delgado, Soran Jahangiri, Keri McKiernan, Johannes Jakob Meyer, Zeyue Niu, Alain Plourde, and Nathan Killoran. PennyLane: Automatic differentiation of hybrid quantum–classical computations. 2018.

BNCI Horizon 2020. BNCI horizon 2020 database: Dataset 009-2014. BNCI Horizon 2020 public data repository, 2014. Dataset 009-2014, covert and overt ERP-based BCI; CC BY-NC-ND 4.0; accessed 2026-07-14.

M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3:625–644, 2021. doi: 10.1038/s42254-021-00348-9.

Chi-Sheng Chen, Samuel Yen-Chi Chen, Aidan Hung-Wen Tsai, and Chun-Shu Wei. QEEGNet: Quantum machine learning for enhanced electroencephalography encoding. In *Proceedings of the IEEE Workshop on Signal Processing Systems*, pp. 153–158, 2024. Also available as arXiv:2407.19214.

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.

EEGET-ALS Dataset. EEGET-ALS dataset. Kaggle dataset, n.d. Kaggle ALS/control subject archive; accessed 2026-07-14.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti Hamalainen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7:267, 2013. doi: 10.3389/fnins.2013.00267.

Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. In *International Conference on Learning Representations*, 2021.

Vinay Jayaram and Alexandre Barachant. MOABB: Trustworthy algorithm benchmarking for BCIs. *Journal of Neural Engineering*, 15(6):066011, 2018. doi: 10.1088/1741-2552/aadea0.

Ian T. Jolliffe. *Principal Component Analysis*. Springer, 2 edition, 2002.

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.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017.

Zhengqing Miao, Meirong Zhao, Xin Zhang, and Dong Ming. Lmda-net: A lightweight multi-dimensional attention network for general eeg-based brain–computer interfaces and interpretability. *NeuroImage*, 276:120209, 2023. doi: 10.1016/j.neuroimage.2023.120209.

Johannes Muller-Gerking, Gert Pfurtscheller, and Henrik Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology*, 110(5):787–798, 1999. doi: 10.1016/S1388-2457(98)00038-8.

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

Cynthia Olvera, Oscar Montiel Ross, and Yoshio Rubio. EEG-based motor imagery classification with quantum algorithms. *Expert Systems with Applications*, 247:123354, 2024. doi: 10.1016/j.eswa.2024.123354.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Rishan Patel, Dai Jiang, Barney Bryson, Tom Carlson, Andreas Demosthenous, and Andrew Geronimo. Longitudinal ALS EEG dataset for motor imagery studies. University College London Research Data Repository, 2025. CC BY 4.0.

Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Mathieu Perrot, and Edouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Adrià Perez-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.

John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018. doi: 10.22331/q-2018-08-06-79.

Herbert Ramoser, Johannes Muller-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.

Bertrand Rivet, Antoine Souloumiac, Virginie Attina, and Guillaume Gibert. xDAWN algorithm to enhance evoked potentials: Application to brain–computer interface. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2061–2064, 2009. doi: 10.1109/ICASSP.2009.4959629.

Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122(4):040504, 2019. doi: 10.1103/PhysRevLett.122.040504.

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.

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.

Baochen Sun and Kate Saenko. Deep CORAL: Correlation alignment for deep domain adaptation. In *European Conference on Computer Vision Workshops*, pp. 443–450, 2016. doi: 10.1007/978-3-319-49409-8_35.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

Yiming Wang, Bin Zhang, and Yujiao Tang. DMMR: Cross-subject domain generalization for EEG-based emotion recognition via denoising mixed mutual reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 628–636, 2024. doi: 10.1609/aaai.v38i1.27819.

Weining Weng, Yang Gu, Yuan Ma, Yuchen Liu, Yingwei Zhang, and Yiqiang Chen. State mamba: Spatiotemporal EEG state-space model with dynamic brain alignment for cross-subject representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 17850–17858, 2026. doi: 10.1609/aaai.v40i21.38843.

Xiao-Cong Zhong, Qisong Wang, Dan Liu, Zhihuang Chen, Jing-Xiao Liao, Jinwei Sun, Yudong Zhang, and Feng-Lei Fan. EEG-DG: A multi-source domain generalization framework for motor imagery EEG classification, 2023.

Table 5: Notation shared with the main paper. D_{c} denotes the residual-embedding dimension, and $\mathcal{E}_m^{\mathrm{c}}$ denotes the view-specific classical residual encoder.

Symbol	Meaning
$J, S, M, Q, N_{\mathrm{vqc}}, D_{\mathrm{c}}$	Samples, subject domains, views, qubits, circuit depth, and residual width
i, m, ν, s, k	Sample, view, dummy-view, subject-domain, and class indices
$\mathcal{S}_{\mathrm{tr}}, \mathcal{S}_{\mathrm{va}}, \mathcal{S}_{\mathrm{te}}$	Source-training, source-validation, and held-out test-domain sets
$\mathbf{X}_i$	EEG trial in $\mathbb{R}^{C_{\mathrm{eeg}} \times T_{\mathrm{eeg}}}$
$\mathbf{x}_i^{(m)}, \widetilde{\mathbf{x}}_i^{(m)}$	Raw and training-fitted processed vectors for view m
Ψ_m	Training-fitted standardization/PCA operator
$\mathbf{a}_i^{(m)}$	Bounded RY–RZ angle vector
$\mathcal{Q}_m, \mathcal{E}_m^{\mathrm{c}}$	View-specific VQC and classical residual encoder
$\mathbf{q}_i^{(m)}, \mathbf{r}_i^{(m)}$	Per-view quantum and residual embeddings
$\eta_i^{(m)}, \alpha_i^{(m)}$	Unnormalized score and normalized view weight
$\mathbf{h}_i^{\mathrm{q}}, \mathbf{h}_i^{\mathrm{c}}, \mathbf{u}_i$	Quantum fusion, residual fusion, and concatenated representation
$\mathbf{o}_i, \varpi_{i,k}$	Two-class logits and predicted probability
ρ_k, γ	Source-training class weight and focal exponent
$\mathcal{B}_{s,k}, \mathcal{V}_{\mathcal{B}}$	Mini-batch subject–class subset and valid moment groups
$\mathcal{B}_k, \mathcal{K}_{\mathcal{B}}, \mathcal{P}_{\mathcal{B}}$	Class subset, valid classes, and valid center pairs
$\boldsymbol{\mu}_{s,k}, \bar{\boldsymbol{\mu}}_k$	Subject–class and pooled class means
$\boldsymbol{\Sigma}_{s,k}, \bar{\boldsymbol{\Sigma}}_k$	Subject–class and pooled class covariances
$\boldsymbol{\vartheta}_{k,\mathcal{B}}$	Normalized mini-batch center of class k
$\chi, \boldsymbol{\beta}, \mathbf{G}_\chi, \kappa$	Local margin probe and its affine–quadratic coefficients
$\mathcal{L}_{\mathrm{cls}}, \mathcal{L}_{\mathrm{cor}}, \mathcal{L}_{\mathrm{sep}}$	Classification, moment-alignment, and class-separation losses
$\mathcal{L}_{\mathrm{total}}$	Complete training objective

A Supplementary Material

This supplement supplies the details needed to interpret and reproduce the main paper. It proceeds from notation and subject-disjoint protocols to view construction, the exact model and objectives, optimization, representation characterization, comparator implementation, and complete experimental evidence. In particular, it distinguishes three questions that must not be conflated: whether the complete system is competitive with reimplemented EEG architectures, whether the VQC outperforms the specific parameter-matched MLP substitution inside the proposed system, and which training objective supplies the largest individual gain. The final sections state the compute conditions, aggregation convention, evidence boundaries, and reproducibility checks. Figures already shown at readable size in the main paper are not duplicated; exact numerical tables are retained here.

B Notation

Table 5 groups symbols by role. The first rows define dataset cardinalities, indices, and subject-domain partitions; the middle rows describe view preprocessing and the two encoder branches; and the final rows define fusion, mini-batch statistics, and objectives. In particular, D_{c} is reserved for residual dimensionality, $\mathcal{E}_m^{\mathrm{c}}$ for the corresponding encoder, $\mathcal{V}_{\mathcal{B}}$ for valid subject–class moment groups, and $\mathbf{G}_\chi$ for the quadratic term in the local margin analysis. The pooled class moments $\bar{\boldsymbol{\mu}}_k$ and $\bar{\boldsymbol{\Sigma}}_k$ are computed from samples in $\mathcal{B}_k$; they are not equal-weight averages of subject-specific moments.

C Datasets and Subject-Disjoint Evaluation

For every outer fold, $\mathcal{S}_{\mathrm{tr}}$, $\mathcal{S}_{\mathrm{va}}$, and $\mathcal{S}_{\mathrm{te}}$ are pairwise disjoint. All fitted transforms use $\mathcal{S}_{\mathrm{tr}}$ only; $\mathcal{S}_{\mathrm{va}}$ controls model selection, stopping, and thresholding; $\mathcal{S}_{\mathrm{te}}$ is evaluated once. Fold identities are shared by every model and seed.

C.1 BNCI-P300

BNCI2014-009 is loaded through MOABB (BNCI Horizon 2020, 2014; Jayaram & Barachant, 2018). Target is encoded as 1 and non-target as 0. The pipeline applies 0.1–30 Hz filtering, extracts 0–0.8 s

Table 6: Dataset and split statistics. BNCI counts follow the public grid-speller metadata; the UCL volume includes all recorded conditions before rest exclusion; and EEGET counts precede the deterministic per-subject window cap. Post-filtering class counts are generated directly from the fixed fold registry.

Dataset	Subj.	Subject classes	Recorded task volume	Ch.	Outer protocol	Test unit
BNCI-P300	10	Trial-level target/non-target	17,280 epochs: 14,400 NT / 2,880 T	16	LOSO (10 folds)	1 subject
UCL ALS-EEG	8	All patients have ALS	160 trials/patient across recorded conditions; L/R retained	19	Patient-out	1 patient
EEGET-ALS	176	170 control / 6 ALS	9 sessions/subject; 2-s windows capped at 160/subject	32	6 ALS-indexed pair-out folds	1 control + 1 ALS

stimulus-locked epochs, and resamples to 128 Hz. Each of the ten outer folds holds out one subject; one remaining subject is validation and eight are training.

C.2 UCL ALS-EEG

The UCL archive contains eight patient-specific MATLAB records with left, right, and rest trials from 19 electrodes (Patel et al., 2025). Rest is excluded; left/right are encoded 0/1. All trials from a patient file inherit that patient's identifier. Each outer fold holds out one patient; one source patient is validation and the remaining patients are training.

C.3 EEGET-ALS

The EEGET-ALS archive contains six ALS and 170 control participants with EEG and eye tracking (Ngo et al., 2024; EEGET-ALS Dataset, n.d.). We use EEG only, encode control/ALS as 0/1, split continuous recordings into non-overlapping 2-s windows, and retain at most 160 windows per subject. A one-subject test fold would contain only one label, so the fixed registry contains six ALS-indexed outer folds. Each fold holds out one ALS participant and one control participant, and its validation domains are drawn from the remaining sources with both labels represented. All six ALS participants appear once; the paired control identity is read from the same seed-invariant fold registry for every method rather than resampled during training. Thus the protocol evaluates six control slots rather than exhaustively covering the 170-control pool. The model predicts the inherited status label of each 2-s window. Metrics are computed within each held-out pair and averaged equally across the six folds; the reported result is therefore window classification under subject-disjoint pair-out evaluation, not a single diagnostic decision per participant.

C.4 Leakage Checks

A fold is retained only if: (i) subject identifiers are disjoint; (ii) validation/test contain both classes; (iii) scalers, PCA, xDAWN, tangent mappings, CSP, and covariance parameters are fitted on training sources; (iv) class weights and balancing use training sources; and (v) stopping and thresholds use validation sources only. These checks instantiate the source-only evaluation described in the main paper. Repeated trials from a held-out participant never appear in training or validation, and no statistic estimated from the held-out fold enters feature construction, loss weighting, early stopping, threshold selection, or hyperparameter choice. Consequently, the large epoch/window counts in Table 6 do not inflate the number of independent evaluation units.

Table 6 also clarifies the distinction between the outer unit and the prediction unit. BNCI and UCL predict trial or epoch labels for one held-out participant. EEGET predicts window labels for a held-out control–ALS pair. In all cases, generalization is evaluated only on participants absent from fitting and selection, but the EEGET metric must not be described as participant-level diagnostic accuracy.

D TASK-AWARE VIEWS AND PREPROCESSING

Every trial produces five base views. The temporal view concatenates mean, standard deviation, maximum, and minimum over four windows. P300 uses 0–0.2, 0.2–0.4, 0.4–0.6, and 0.6–0.8 s; other tasks divide the available trial into four equal windows. The spatial view partitions channels into up to four ordered groups and summarizes region–window mean, standard deviation, and maximum. This is a coarse ordered summary, not an anatomical-region claim when montages are heterogeneous.

The spectral view uses one-sided FFT power and $\log(1 + \cdot)$. P300 bands are 0.5–4, 4–8, 8–13, and 13–30 Hz; ALS-related bands are 4–8, 8–12, 12–16, 16–24, and 24–35 Hz. The task-statistical view uses ERP-window amplitude statistics for P300 and channel-wise log variance, mean, and standard deviation for ALS tasks.

For covariance, let $\mathbf{X}_i^{\circ}$ denote the channel-centered trial. We compute

$$\mathbf{\Sigma}_i^{\mathrm{emp}} = \frac{\mathbf{X}_i^{\circ}(\mathbf{X}_i^{\circ})^{\top}}{T_{\mathrm{eeg}} - 1}, \tag{10}$$

$$\mathbf{\Sigma}_i^{\mathrm{reg}} = \mathbf{\Sigma}_i^{\mathrm{emp}} + 10^{-3}\frac{\mathrm{tr}(\mathbf{\Sigma}_i^{\mathrm{emp}})}{C_{\mathrm{eeg}}}\mathbf{I}. \tag{11}$$

The trace-normalized upper triangle is vectorized. BNCI adds an xDAWN-enhanced covariance tangent view; UCL and EEGET add regularized tangent-space and CSP log-variance views (Rivet et al., 2009; Barachant et al., 2012; Muller-Gerking et al., 1999; Chen et al., 2010). All supervised transforms are rebuilt per outer fold.

A StandardScaler is fitted per view on source training. PCA is applied only above 32 dimensions; supervised views are capped at 16 dimensions (Jolliffe, 2002; Pedregosa et al., 2011). Validation and test use the fitted transforms unchanged.

Role of the view construction. The five base views provide complementary summaries for every task, while the supervised geometry views encode paradigm-specific structure. xDAWN tangent features emphasize stimulus-locked covariance changes for P300. Riemannian tangent and CSP features summarize covariance geometry and discriminative sensorimotor variance for the ALS-related tasks. Importantly, supervised does not mean test-informed: every xDAWN, tangent-space, and CSP object is rebuilt from source-training observations inside its outer fold. Appendix L quantifies each branch contribution without changing this rule.

E EXACT MODEL AND OBJECTIVES

E.1 PER-VIEW ENCODERS

The full model uses $Q = 6$, $N_{\mathrm{vqc}} = 2$, $D_{\mathrm{c}} = 16$, and dropout 0.20. The angle projection is

$$\boldsymbol{\xi}_i^{(m)} = \mathrm{Dropout}(\mathrm{GELU}(\mathbf{\Pi}_{1,m}\,\mathrm{LN}(\widetilde{\mathbf{x}}_i^{(m)}))), \tag{12}$$

$$\mathbf{a}_i^{(m)} = \pi\tanh(\mathbf{\Pi}_{2,m}\boldsymbol{\xi}_i^{(m)}) \in [-\pi, \pi]^{2Q}. \tag{13}$$

The VQC applies RY/RZ data encoding, two layers of trainable RY/RZ gates, ring CNOTs, and all-qubit Pauli-Z readout. LayerNorm–Linear–GELU–LayerNorm produces $\mathbf{q}_i^{(m)} \in \mathbb{R}^Q$. The classical path applies LayerNorm–Linear–GELU–Dropout–Linear–LayerNorm to yield $\mathbf{r}_i^{(m)} \in \mathbb{R}^{D_{\mathrm{c}}}$. Parameters are independent across views.

E.2 VIEW WEIGHTING, FUSION, AND CLASSIFIER

The residual branch is intended to preserve information that the low-qubit map may compress. To prevent it from learning an independent selector that could bypass the principal encoder, one view distribution is computed from the VQC embeddings and reused for both branches. This is a

VQC-conditioned classical weighting mechanism, not a quantum attention primitive.

$$\eta_i^{(m)} = \boldsymbol{\zeta}^\top \tanh(\mathbf{W}\mathbf{q}_i^{(m)} + \mathbf{b}), \tag{14}$$

$$\alpha_i^{(m)} = \frac{\exp(\eta_i^{(m)}/\tau)}{\sum_{\nu=1}^{M} \exp(\eta_i^{(\nu)}/\tau)}, \quad \tau = 0.75, \tag{15}$$

$$\mathbf{h}_i^{\mathrm{q}} = \sum_m \alpha_i^{(m)} \mathbf{q}_i^{(m)}, \quad \mathbf{h}_i^{\mathrm{c}} = \sum_m \alpha_i^{(m)} \mathbf{r}_i^{(m)}, \tag{16}$$

$$\mathbf{u}_i = [\mathbf{h}_i^{\mathrm{q}} \| \mathbf{h}_i^{\mathrm{c}}], \quad \varpi_{i,k} = \mathrm{softmax}(f_\omega(\mathbf{u}_i))_k. \tag{17}$$

The view-scoring hidden width is 12; the classifier input/hidden widths are 22/24. Compact MVQAN-CORAL uses $Q = 4$, depth 1, $D_{\mathrm{c}} = 4$, and dropout 0.25.

E.3 Exact Classification, Alignment, and Separation

Balanced training weights are $\rho_k^{(0)} = |\mathcal{D}_{\mathrm{tr}}|/(2|\mathcal{D}_{\mathrm{tr},k}|)$, followed by $\rho_0 = \rho_0^{(0)}$ and $\rho_1 = 1.25\rho_1^{(0)}$. The implemented focal-style objective is

$$\varepsilon_i = -\rho_{y_i} \log \varpi_{i,y_i}, \tag{18}$$

$$\mathcal{L}_{\mathrm{cls}} = |\mathcal{B}|^{-1} \sum_{i \in \mathcal{B}} (1 - e^{-\varepsilon_i})^\gamma \varepsilon_i, \quad \gamma = 2. \tag{19}$$

Within mini-batch $\mathcal{B}$, define $\mathcal{B}_k = \{i \in \mathcal{B} : y_i = k\}$ and $\mathcal{B}_{s,k} = \{i \in \mathcal{B} : s_i = s, y_i = k\}$. The pooled class moments use every mini-batch sample with label k:

$$\bar{\boldsymbol{\mu}}_k = \frac{1}{|\mathcal{B}_k|} \sum_{i \in \mathcal{B}_k} \mathbf{h}_i^{\mathrm{q}}, \tag{20}$$

$$\bar{\boldsymbol{\Sigma}}_k = \frac{1}{|\mathcal{B}_k| - 1} \sum_{i \in \mathcal{B}_k} (\mathbf{h}_i^{\mathrm{q}} - \bar{\boldsymbol{\mu}}_k)(\mathbf{h}_i^{\mathrm{q}} - \bar{\boldsymbol{\mu}}_k)^\top. \tag{21}$$

Thus the reference is sample-weighted; subjects with more valid samples in the current mini-batch contribute proportionally. Subject–class moments $\boldsymbol{\mu}_{s,k}, \boldsymbol{\Sigma}_{s,k}$ are defined analogously on $\mathcal{B}_{s,k}$, and $\mathcal{V}_{\mathcal{B}}$ contains the groups with at least two examples:

$$\mathcal{L}_{\mathrm{cor}} = \frac{1}{|\mathcal{V}_{\mathcal{B}}|} \sum_{(s,k) \in \mathcal{V}_{\mathcal{B}}} \left[\frac{\|\boldsymbol{\mu}_{s,k} - \bar{\boldsymbol{\mu}}_k\|_2^2}{Q} + \frac{\|\boldsymbol{\Sigma}_{s,k} - \bar{\boldsymbol{\Sigma}}_k\|_F^2}{Q^2} \right]. \tag{22}$$

The outer average assigns equal weight to each valid subject–class group. If $\mathcal{V}_{\mathcal{B}}$ is empty, this term is zero.

Normalized embeddings form valid class centers only when $|\mathcal{B}_k| \geq 2$. The separation loss is exactly Eq. 7 of the main paper. If no class is valid, $\mathcal{L}_{\mathrm{sep}} = 0$; if only one center is valid, the compactness term is retained and the center-margin term is zero. The final coefficients are $\lambda_{\mathrm{cor}} = 0.01$, $\lambda_{\mathrm{sep}} = 0.015$, and margin $\delta = 0.75$.

Interaction of the two branches and three objectives. Only the concatenation $\mathbf{u}_i = [\mathbf{h}_i^{\mathrm{q}} \| \mathbf{h}_i^{\mathrm{c}}]$ enters the classifier. The VQC-conditioned weights determine which views are emphasized in both branches, while the residual pathway can retain information compressed by the low-qubit representation. Alignment and separation act only on $\mathbf{h}_i^{\mathrm{q}}$ during source training; they add no inference-time inputs and never use a held-out subject. Appendix K separates the objectives and reports the broader architecture interactions.

F Optimization and Model Selection

In Table 7 the optimization settings define a single source-validation policy rather than a collection of dataset-specific test-tuned choices. Macro-F1 controls checkpointing, scheduling, and probability-threshold selection because it treats both classes symmetrically under imbalance. The epoch cap

Table 7: Final optimization settings. Every choice and checkpoint uses source training/validation only.

Item	Setting
Optimizer / learning rate	AdamW / 10^{-3}
Weight decay	10^{-4}
Maximum epochs / patience	120 / 22
Scheduler	ReduceLROnPlateau, factor 0.5
Gradient clipping	5.0
Quantum device / shots	PennyLane `default.qubit` / analytic (`shots=None`)
Differentiation	PyTorch interface; PennyLane 0.45.1 default `best` simulator differentiation
Quantum / classical batch	32 / 128 on GPU (32 CPU)
Threshold grid	0.05–0.95, step 0.01
Selection metric	source-validation Macro-F1
Seeds	$\{0, 1, 2\}$

is an upper bound: patience-based early stopping selects the actual training duration using source-validation subjects. The smaller quantum-model batch is imposed by analytic circuit simulation and should not be read as an efficiency benefit. Conversely, the larger GPU batch for pure PyTorch models is an execution detail rather than additional access to data.

For each dataset, seed, and fold: construct held-out and validation domains; balance only source training where required; fit views and transforms; train with $\mathcal{L}_{\text{total}}$; select the threshold and checkpoint by validation Macro-F1; and evaluate the untouched test fold once. The same fold registry is reused by every method. This fixed sequence is the operational definition of the leakage-free protocol used by every result table.

Quantum differentiation. The executed QNodes use PennyLane 0.45.1 `default.qubit`, analytic expectations (`shots=None`), and the PyTorch interface. The implementation does not override PennyLane's differentiation setting and therefore uses the default `diff_method="best"` simulator differentiation. No parameter-shift override, finite-shot estimator, or hardware gradient is used. We report the recorded configuration rather than retroactively assigning a more specific backend label not stored in the run artifacts.

Primary configuration versus sensitivity grid. The six-qubit, depth-two primary estimates and the reported 4×3 grid come from separate complete runs. The grid does not reuse the primary checkpoints, and its held-out cells are not substituted for the primary table. The archived records do not establish that the configuration was prespecified independently of all earlier exploratory work. We therefore describe the grid as exploratory sensitivity evidence, not as an independent confirmatory selection experiment. Appendix M reports the exact grid and the numerical discrepancies.

G Representation Characterization and Proofs

This section proves the statements of Sec. 3.5 and states the fusion bound used there. The results characterize the implemented representation and loss; they are not generalization guarantees and do not establish that a VQC is preferable to a classical Fourier or trigonometric map.

Proof of Proposition 1. Every RY or RZ matrix entry is a linear combination of $\sin(a_j/2)$ and $\cos(a_j/2)$. Trainable rotations and CNOTs contribute matrices independent of the data angles. A finite product of these matrices therefore has entries that are finite polynomials in the sine/cosine factors. The expectation $\langle 0|U^\dagger Z_\ell U|0\rangle$ is formed by finite sums and products of such entries and can be rewritten as a finite multivariate trigonometric polynomial. Since Z_ℓ has operator norm one, the expectation lies in $[-1, 1]$. CNOTs couple tensor factors, so the polynomial can contain products involving angles assigned to different connected qubits. □

The proposition motivates a structured bounded hypothesis class; it does not establish that this class is universally better than a classical neural network (Schuld et al., 2021).

Table 8: Executed tuning budget. We did not perform broad post-hoc architecture sweeps for the reimplemented comparators; the held-out test domain was never used for selection.

Method	Architecture/tuning space	Source-validation procedure	Budget
MVQAN-CORAL	Primary $Q = 6$, depth 2; separate exploratory $Q \in \{2, 4, 6, 8\}$ and depth $\in \{1, 2, 3\}$ sensitivity reruns	Early stopping and threshold by source-validation Macro-F1; grid not presented as independent prespecification	Primary run + 12 diagnostic settings
PM-Classical	Integer hidden-width search until parameter gap $< 1\%$	Same folds, losses, stopping, threshold	Deterministic width match
EEGNet / ATCNet / EEG-DG	Released architecture adapted only to input/output shape	Common early stopping and threshold; no test tuning	1 released architecture each
QEEGNet / Olvera VQC	Reported circuit design adapted to binary task	Common early stopping and threshold; no test tuning	1 reported architecture each
LMDA-Net / EEG Conformer	Released/default architecture adapted to channels/classes	Common early stopping and threshold; no test tuning	1 released architecture each

Lemma 1 (Bounded dual fusion)**.** *If* $\|\mathbf{q}_i^{(m)}\|_2 \le \Omega_{\mathrm{q}}$ *and* $\|\mathbf{r}_i^{(m)}\|_2 \le \Omega_{\mathrm{c}}$ *for all* m*, then* $\|\mathbf{h}_i^{\mathrm{q}}\|_2 \le \Omega_{\mathrm{q}}$, $\|\mathbf{h}_i^{\mathrm{c}}\|_2 \le \Omega_{\mathrm{c}}$*, and* $\|\mathbf{u}_i\|_2 \le \sqrt{\Omega_{\mathrm{q}}^2 + \Omega_{\mathrm{c}}^2}$.

Proof. Softmax gives nonnegative weights summing to one. The triangle inequality bounds each weighted sum; concatenation adds squared norms. □

Proof of Proposition 2. Use $\mathbb{E}[\mathbf{z}^\top \mathbf{G}_\chi \mathbf{z}] = \mathrm{tr}(\mathbf{G}_\chi \boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top \mathbf{G}_\chi \boldsymbol{\mu}$. Cauchy–Schwarz bounds the linear term, and Frobenius duality bounds the trace term. Adding and subtracting $\bar{\boldsymbol{\mu}}_k^\top \mathbf{G}_\chi \boldsymbol{\mu}_{s,k}$ bounds the quadratic mean difference by $2\Omega_\mu \|\mathbf{G}_\chi\|_2 \|\boldsymbol{\mu}_{s,k} - \bar{\boldsymbol{\mu}}_k\|_2$. □

The statement concerns source-domain local drift; it does not inspect the held-out domain or bound the complete classifier.

H Reimplemented Architectures, Parameter Matching, and Tuning

Seven published architectures are reimplemented and adapted to the three binary tasks. EEGNet retains temporal, depthwise, and separable convolutions; ATCNet retains temporal windows, attention, and TCN blocks; and EEG-DG retains its multi-source invariance mechanism (Lawhern et al., 2018; Altaheri et al., 2023; Zhong et al., 2023). QEEGNet and Olvera retain their reported globally compressed VQC pipelines, while LMDA-Net and EEG Conformer retain multidimensional attention and convolutional Transformer encoding, respectively (Chen et al., 2024; Olvera et al., 2024; Miao et al., 2023; Song et al., 2023). The numerical entries are results of these task-adapted reimplementations under our protocol; they are not values copied from the original publications.

QEEGNet and Olvera are substantially smaller than MVQAN-CORAL and do not share its views, residual branch, fusion, or auxiliary objectives. Their rows therefore compare complete pipelines and cannot attribute a performance gap to quantum encoding alone. Encoder attribution is reserved for PM-Classical, which holds those other components fixed. Even that control compares the implemented VQC with a generic MLP, not with every structured classical basis.

PM-Classical changes only the principal per-view encoder. Hidden widths are selected by deterministic integer search to minimize the parameter-count difference, producing gaps of 135, 131, and 131 parameters (0.72%, 0.63%, and 0.79%).

Table 8 separates three activities. First, each reimplemented comparator retains one released or reported task-adapted architecture and shares the same source-validation stopping and threshold rule. Second, the parameter-matched control performs only the deterministic width search required to match capacity; it does not search on test performance. Third, the 4×3 circuit grid is a separate exploratory sensitivity analysis and is not used to replace the separately executed primary estimates. The comparison is therefore controlled at the evaluation-protocol level, but it is not an exhaustive

Table 9: BNCI-P300 full baseline comparison (%).

Model	Macro-F1	BAcc	Accuracy	AUROC	AUPRC	Params
ATCNet	77.7±0.5	78.3±0.7	85.4±0.4	80.4±0.5	70.9±0.8	36,858
EEGNet	77.5±0.6	79.4±0.4	85.8±0.6	80.1±0.6	70.9±0.7	1,458
EEG-DG	74.0±0.7	71.6±0.6	84.0±0.7	82.1±0.5	62.5±0.9	933,768
QEEGNet	68.9±0.6	67.7±0.5	83.7±0.5	81.5±0.7	50.6±0.9	5,382
Olvera Hybrid VQC	66.6±0.7	79.6±0.6	72.7±0.5	81.6±0.7	68.3±0.5	1,966
LMDA-Net	79.9±0.6	80.6±0.6	85.2±0.4	81.9±0.5	71.9±0.7	∼6,800
EEG Conformer	80.7±0.6	81.3±0.5	85.5±0.5	82.3±0.5	72.5±0.7	∼54,000
MVQAN-CORAL	**86.5±0.5**	**87.0±0.6**	**86.8±0.6**	**84.2±0.5**	**74.4±0.8**	18,775

Table 10: UCL ALS-EEG full baseline comparison (%).

Model	Macro-F1	BAcc	Accuracy	AUROC	AUPRC	Params
EEGNet	64.8±0.7	66.3±0.7	65.4±0.5	73.8±0.8	75.0±0.8	1,554
EEG-DG	63.0±0.7	64.4±0.7	64.4±0.6	72.5±0.8	75.0±0.7	933,960
ATCNet	62.0±0.7	63.1±0.9	63.1±0.6	70.4±0.5	71.3±0.8	37,050
QEEGNet	60.7±0.7	61.6±0.8	61.6±0.8	70.0±0.7	71.6±0.7	5,478
Olvera Hybrid VQC	50.5±1.1	51.6±0.7	51.6±0.8	52.3±1.1	54.3±0.8	2,190
LMDA-Net	63.5±0.8	64.2±0.8	64.2±0.7	72.8±0.8	73.9±0.8	∼6,900
EEG Conformer	64.3±0.7	64.8±0.7	64.8±0.6	73.3±0.7	74.3±0.7	∼55,000
MVQAN-CORAL	**66.8±0.5**	**67.4±0.8**	**67.4±0.7**	**74.8±0.7**	**75.9±0.7**	20,951

architecture search for every method. Rounded parameter totals for LMDA-Net and EEG Conformer are marked as approximate.

I Full Reimplemented-Architecture Results

The left panel of the expanded-comparison figure in the main paper provides the visual Macro-F1 ranking. This section supplies the complete numerical evidence behind that panel. Tables 9–11 report Macro-F1, balanced accuracy, accuracy, AUROC, AUPRC, and trainable parameters under exactly the same subject-level folds, seed aggregation, labels, and source-validation thresholding as MVQAN-CORAL. Approximate parameter totals are marked explicitly for LMDA-Net and EEG Conformer.

I.1 Cross-Architecture Results

Cross-dataset pattern. The left panel of the main-paper expanded-comparison figure shows that MVQAN-CORAL has the highest Macro-F1 mean on all three tasks. The margin is nevertheless task dependent: it is moderate on BNCI-P300, narrowest on the eight-patient UCL cohort, and largest on the EEGET window-classification task under subject-disjoint pair-out evaluation. The reported standard deviations describe limited seed-to-seed variation under fixed folds; they are not confidence intervals over new subjects. Because Macro-F1 alone cannot distinguish threshold-dependent class balance from ranking quality, Tables 9– 11 provide both thresholded and ranking metrics.

BNCI-P300. Table 9 shows that the proposed model improves Macro-F1 from the strongest reproduced value of 80.7 (EEG Conformer) to 86.5, a 5.8-point gain. The accuracy margin is smaller—86.8 versus 85.8 for EEGNet—because the majority non-target class dominates raw accuracy. The larger gains in balanced accuracy and Macro-F1 therefore indicate better treatment of both classes, while the AUROC and AUPRC improvements over the strongest corresponding reproduced values are 1.9 points each. This metric pattern is consistent with the reason Macro-F1 was selected as primary for the imbalanced P300 task.

UCL ALS-EEG. The UCL results in Table 10 are interpreted more cautiously. The proposed model reaches 66.8 Macro-F1, 2.0 points above EEGNet, and leads the remaining metrics by 0.9–2.0 points over their strongest reproduced means. The gains are consistent but smaller than on the other datasets, which is unsurprising for eight heterogeneous patients and a difficult cross-patient imagery

Table 11: EEGET-ALS full baseline comparison (%).

Model	Macro-F1	BAcc	Accuracy	AUROC	AUPRC	Params
EEG-DG	68.6±0.6	68.8±0.8	68.8±0.6	70.8±0.5	66.7±0.5	933,320
QEEGNet	65.2±0.6	66.6±0.7	66.6±0.8	82.9±0.6	81.3±0.6	5,158
EEGNet	63.9±0.7	67.7±0.6	67.7±0.7	75.4±0.7	76.5±0.4	1,234
ATCNet	56.0±0.8	59.2±0.6	59.2±0.7	61.1±0.9	62.8±0.9	36,138
Olvera Hybrid VQC	33.3±0.7	50.0±0.7	50.0±0.7	66.9±0.5	59.7±0.6	1,870
LMDA-Net	73.1±0.7	73.8±0.7	73.8±0.6	83.2±0.5	82.0±0.6	∼5,900
EEG Conformer	75.6±0.6	76.4±0.6	76.4±0.5	85.3±0.5	84.2±0.5	∼48,000
MVQAN-CORAL	**86.6±0.4**	**86.9±0.6**	**86.8±0.5**	**92.0±0.4**	**94.2±0.3**	16,651

Table 12: Controlled encoder comparison. Every non-encoder component and training condition is identical. Parameter gaps are below 1%.

Dataset	Model	Macro-F1	BAcc	Accuracy	AUROC	AUPRC	Params
BNCI	PM-Classical	81.8±0.7	82.5±0.7	84.9±0.5	82.0±0.6	71.4±0.8	18,640
	MVQAN-CORAL	**86.5±0.5**	**87.0±0.6**	**86.8±0.6**	**84.2±0.5**	**74.4±0.8**	18,775
UCL	PM-Classical	63.7±0.8	64.3±0.8	64.3±0.7	71.9±0.8	74.1±0.8	20,820
	MVQAN-CORAL	**66.8±0.5**	**67.4±0.8**	**67.4±0.7**	**74.8±0.7**	**75.9±0.7**	20,951
EEGET	PM-Classical	83.6±0.6	84.1±0.6	84.1±0.5	89.8±0.5	91.5±0.5	16,520
	MVQAN-CORAL	**86.6±0.4**	**86.9±0.6**	**86.8±0.5**	**92.0±0.4**	**94.2±0.3**	16,651

task. The result supports predictive utility under the stated folds; three seeds do not justify a formal significance claim.

EEGET-ALS. Table 11 exhibits the largest separation. Relative to EEG Conformer, the strongest reproduced Macro-F1 comparator, MVQAN-CORAL gains 11.0 points; the corresponding improvements are 10.5 in balanced accuracy, 10.4 in accuracy, 6.7 in AUROC, and 10.0 in AUPRC. Agreement between thresholded and ranking metrics reduces the likelihood that the result is explained only by the selected decision threshold. Nevertheless, the evaluation concerns window predictions in balanced held-out control–ALS pairs drawn from a highly imbalanced archive and must not be interpreted as a clinical diagnostic estimate.

The proposed model leads every displayed metric on every dataset, while the size of the gain varies with task and cohort. This establishes competitiveness against the seven executed reimplementations. It does not isolate whether the improvement comes from encoder type, model size, multi-view construction, or the auxiliary losses; those questions are addressed by Appendices J–M.

J Parameter-Matched Classical Replacement

Table 12 holds the task-aware views, residual encoders, shared weighting, fusion, classifier, objectives, folds, and optimization fixed and changes only the principal per-view encoder. Parameter differences are 135 on BNCI, 131 on UCL, and 131 on EEGET, corresponding to 0.72%, 0.63%, and 0.79%. Under these budgets, MVQAN-CORAL improves Macro-F1 by 4.7, 3.1, and 3.0 points and improves every other reported metric. Parameter count alone is therefore an inadequate explanation of this substitution result. The conclusion remains architecture-specific: the experiment supports the implemented VQC relative to the matched MLP inside MVQAN-CORAL, not universal superiority over classical networks, Fourier features, or geometry-specific classifiers.

K Objective and Architecture Ablations

Table 13 and Figure 5 vary only the active loss terms of the full architecture. Tables 14–16 report broader structural ablations, some of which alter several components and are therefore read as interaction tests rather than single-factor attribution.

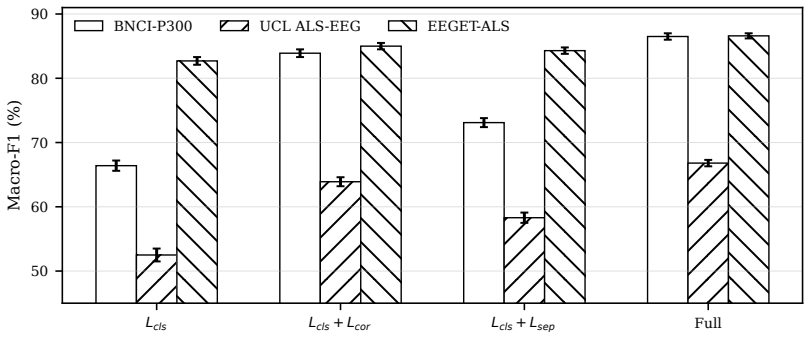


Figure 5: Objective-component Macro-F1 ablation.

Table 13: Objective Macro-F1 ablation (%). Each row uses the full architecture and changes only active loss terms.

Objective	BNCI	UCL	EEGET
$\mathcal{L}_{\text{cls}}$	66.4±0.8	52.5±1.0	82.7±0.6
$+\lambda_{\text{cor}}\mathcal{L}_{\text{cor}}$	83.9±0.6	63.9±0.7	85.0±0.5
$+\lambda_{\text{sep}}\mathcal{L}_{\text{sep}}$	73.1±0.7	58.3±0.8	84.3±0.5
Full	**86.5±0.5**	**66.8±0.5**	**86.6±0.4**

K.1 Objective-Component Interpretation

Table 13 shows that class-conditional moment alignment is the larger individual contributor on BNCI and UCL: adding $\mathcal{L}_{\text{cor}}$ to classification raises Macro-F1 by 17.5 and 11.4 points, compared with gains of 6.7 and 5.8 from $\mathcal{L}_{\text{sep}}$ alone. On EEGET the individual gains are smaller but remain positive (2.3 and 1.6 points). The complete objective is best in every case, adding 2.6, 2.9, and 1.6 points beyond the CORAL-only variant. Thus separation does not replace alignment; it supplies complementary class-geometry regularization after source-class moments have been brought closer.

K.2 Architecture-Interaction Interpretation

BNCI-P300. Table 14 shows that the full model exceeds the PCA single-VQC pipeline by 31.6 Macro-F1 points, so one globally compressed circuit is inadequate for this task. Within the no-alignment setting, mean pooling and learned view weighting are nearly tied (66.5 versus 66.6), and weighting alone therefore cannot explain the final score. Restoring alignment raises the complete model from 66.6 to 86.5, whereas adding CORAL to the much smaller classical MVAttention model changes Macro-F1 by only 0.2 points. The effectiveness of alignment depends on the representation on which it acts.

UCL ALS-EEG. Table 15 gives a similar but less pronounced pattern. The full model is 18.4 points above PCA plus one VQC and 14.5 points above MVQAN without alignment. Classical MVAttention benefits by 2.2 points from CORAL, but it remains 9.8 points below the full model. Mean pooling and learned view weighting again differ by only 0.2 points when alignment is absent, reinforcing an interaction among view-specific encoding, residual preservation, fusion, and the training objectives rather than weighting in isolation.

EEGET-ALS. The EEGET pattern in Table 16 is different and therefore informative. The full model is only 1.9 Macro-F1 points above the no-alignment model, but removing both alignment and the residual pathway lowers Macro-F1 from 84.7 to 34.6. The residual branch is consequently critical for this subject-window task when the quantum representation is not regularized. Mean pooling slightly exceeds learned view weighting without alignment, and classical MVAttention plus CORAL is already strong. These findings caution against treating any component as uniformly dominant across datasets.

Table 14: BNCI-P300 architecture ablations (%).

BNCI variant	Macro-F1	BAcc	Accuracy	AUROC	AUPRC	Params
PCA single VQC	54.9±1.1	54.4±0.7	80.1±0.6	62.7±0.7	25.0±0.8	486
Classical MVAttention	66.2±0.6	66.1±0.7	81.2±0.7	78.1±0.4	41.6±0.7	4,357
Classical MVAttention + CORAL	66.4±0.5	66.3±0.6	81.4±0.7	77.1±0.6	40.1±1.1	4,357
Compact MVQAN-CORAL	61.6±0.7	60.5±0.6	81.1±0.5	77.4±0.7	40.1±0.9	9,623
No alignment, no residual	63.0±0.8	62.0±0.6	81.2±0.5	76.3±0.7	35.5±0.8	5,365
Mean pooling, no alignment	66.5±0.6	67.0±0.6	80.8±0.6	77.7±0.5	41.4±0.8	18,575
MVQAN, no alignment	66.6±0.6	67.0±0.5	81.0±0.4	77.6±0.4	38.7±0.8	18,575
MVQAN-CORAL	**86.5±0.5**	**87.0±0.6**	**86.8±0.6**	**84.2±0.5**	**74.4±0.8**	18,775

Table 15: UCL ALS-EEG architecture ablations (%).

UCL variant	Macro-F1	BAcc	Accuracy	AUROC	AUPRC	Params
PCA single VQC	48.4±0.9	48.4±1.1	48.4±0.9	48.8±1.0	49.3±1.1	486
Classical MVAttention	54.8±1.0	55.0±0.6	55.0±0.8	58.0±0.8	59.8±0.6	4,835
Classical MVAttention + CORAL	57.0±0.7	57.2±0.8	57.2±0.7	59.2±0.7	62.9±0.7	4,835
Compact MVQAN-CORAL	57.8±0.7	57.8±0.8	57.8±0.7	60.0±0.8	59.7±0.9	10,747
No alignment, no residual	54.2±1.1	54.4±0.8	54.4±0.8	55.6±0.7	56.8±0.6	6,011
Mean pooling, no alignment	52.1±1.1	55.3±0.9	55.3±0.9	60.6±0.8	61.3±0.9	20,951
MVQAN, no alignment	52.3±1.0	55.3±0.6	55.3±0.8	59.8±0.8	60.7±0.9	20,951
MVQAN-CORAL	**66.8±0.5**	**67.4±0.8**	**67.4±0.7**	**74.8±0.7**	**75.9±0.7**	20,951

L Leave-One-View-Out Analysis

Each variant removes exactly one input branch while retaining the remaining encoders, shared VQC-conditioned weighting, residual path, objectives, folds, and optimization settings. Figure 6 visualizes the drop from the corresponding full model, while Tables 17–19 provide the absolute Macro-F1 values and seed dispersion.

L.1 View Contributions

The heatmap reports decreases rather than absolute scores, allowing relative dependence on each view to be compared across tasks with different difficulty levels. Every cell is positive, indicating that no branch is redundant under the fixed full-model protocol. Darker cells identify views whose removal is most damaging; they should be interpreted as predictive contribution within this architecture, not as causal neurophysiological importance.

BNCI-P300. Table 17 identifies xDAWN tangent geometry as the most influential branch: removing it decreases Macro-F1 by 6.5 points. Task-statistical and temporal views follow with losses of 5.4 and 4.6 points, consistent with the importance of stimulus-locked ERP timing and amplitude structure. Covariance and spectral summaries are also nontrivial, while the spatial view has the smallest but still material 3.2-point contribution. The result supports preserving multiple complementary P300 representations instead of relying on xDAWN alone.

UCL ALS-EEG. In Table 18, CSP log variance produces the largest decrease (6.5 points), followed by Riemannian tangent geometry (5.7) and the spectral view (4.6). This ordering is compatible with an imagery task in which discriminative variance, covariance structure, and mu/beta-range power are central. Temporal, task-statistical, covariance, and spatial summaries still contribute 2.5–4.1 points, so the supervised geometry views do not render the common base views unnecessary.

EEGET-ALS. Table 19 gives the largest role to Riemannian tangent geometry (5.6 points), followed by covariance (4.5) and CSP log variance (3.9). Unlike UCL, EEGET predicts subject status from segmented recordings rather than within-patient imagery labels, so covariance-level subject structure is especially informative. The temporal, spectral, statistical, and spatial branches still yield positive gains of 2.4–3.6 points. Across all three datasets, even the least influential view costs at least 2.4 points when removed, supporting the paper's view-preservation premise.

Table 16: EEGET-ALS architecture ablations (%).

EEGET variant	Macro-F1	BAcc	Accuracy	AUROC	AUPRC	Params
PCA single VQC	73.3±0.8	74.1±0.6	74.1±0.6	76.7±0.4	75.1±0.5	486
Classical MVAttention	84.6±0.6	81.6±0.5	83.6±0.4	91.0±0.3	91.3±0.2	3,373
Classical MVAttention + CORAL	85.9±0.4	82.9±0.3	84.9±0.6	91.2±0.3	91.6±0.5	3,373
Compact MVQAN-CORAL	82.2±0.4	84.2±0.2	83.2±0.2	86.9±0.2	87.5±0.2	7,823
No alignment, no residual	34.6±0.7	50.3±0.7	50.3±1.1	77.0±0.4	84.6±0.7	4,549
Mean pooling, no alignment	85.9±0.6	83.9±0.6	83.9±0.5	90.7±0.2	92.7±0.5	16,651
MVQAN, no alignment	84.7±0.3	85.8±0.4	85.8±0.3	91.6±0.4	93.3±0.5	16,651
MVQAN-CORAL	**86.6±0.4**	**86.9±0.6**	**86.8±0.5**	**92.0±0.4**	**94.2±0.3**	16,651

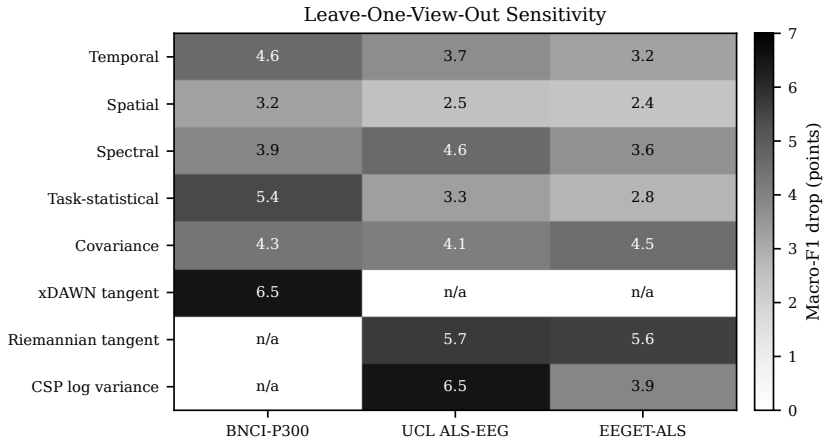


Figure 6: Macro-F1 drop after removing one task-aware view.

M VQC Qubit–Depth Sensitivity

This section provides the exact values underlying the sensitivity heatmap in Figure 3. The grid varies only qubit count and variational depth while holding views, residual width, shared weighting, objectives, folds, and model-selection rules fixed. Each cell is a separate retraining run; these values are descriptive and are not substituted for the frozen primary estimates.

M.1 Sensitivity Grid

Table 20 provides the exact 4×3 diagnostic grid for every dataset. Moving from two to four and then six qubits improves the best depth on every task; the gains flatten at eight qubits, and depth three is consistently below depth two at a fixed qubit count. The $Q = 6$, depth-two cells are 85.6, 66.1, and 85.7, which are 0.9, 0.7, and 0.9 points below the frozen primary estimates. The difference occurs because the grid retrained every configuration independently rather than reusing the primary checkpoints. At depth two, increasing to eight qubits lowers performance by 0.6, 0.8, and 0.5 points; increasing depth from two to three at six qubits lowers it by 0.9, 1.1, and 0.6 points. Because the archived records do not establish prospective independence from every earlier exploratory choice, the grid is interpreted as a descriptive sensitivity diagnostic rather than a confirmatory selection experiment.

N Parameter Context and Embedding Diagnostics

This section interprets the parameter–performance panel retained in the main paper, refers to the held-out fused-quantum diagnostic in Figure 4, and presents the final concatenated representation in Figure 7. None of these visualizations is used for model selection or replaces the fold-aggregated metrics.

Table 17: BNCI-P300 leave-one-view-out results.

BNCI removed view	Macro-F1	Drop
None	86.5±0.5	0.0
Temporal	81.9±0.7	4.6
Spatial	83.3±0.6	3.2
Spectral	82.6±0.7	3.9
Task-statistical	81.1±0.8	5.4
Covariance	82.2±0.7	4.3
xDAWN tangent	80.0±0.9	6.5

Table 18: UCL ALS-EEG leave-one-view-out results.

UCL removed view	Macro-F1	Drop
None	66.8±0.5	0.0
Temporal	63.1±0.8	3.7
Spatial	64.3±0.8	2.5
Spectral	62.2±0.9	4.6
Task-statistical	63.5±0.8	3.3
Covariance	62.7±0.9	4.1
Riemannian tangent	61.1±1.0	5.7
CSP log variance	60.3±1.0	6.5

N.1 Parameter–Performance Plot

The right panel of the main-paper expanded-comparison figure uses a logarithmic parameter axis so that the nearly one-million-parameter EEG-DG model does not visually compress the smaller architectures. MVQAN-CORAL uses far fewer parameters than EEG-DG and fewer than ATCNet and EEG Conformer, while using more than EEGNet, QEEGNet, Olvera, and LMDA-Net. Its position therefore supports a moderate-parameter performance–capacity trade-off rather than a minimum-parameter claim. The panel does not establish wall-clock efficiency because analytic quantum simulation and GPU-accelerated classical models use different execution paths. Encoder-specific capacity is evaluated more directly through the below-1% parameter differences in Table 12.

N.2 Interpretation of the Embedding Diagnostics

Main-paper Figure 4 visualizes $\mathbf{h}^{\mathrm{q}}$, on which $\mathcal{L}_{\mathrm{cor}}$ and $\mathcal{L}_{\mathrm{sep}}$ act; Figure 7 visualizes the concatenated classifier representation $\mathbf{u}$. The displayed folds show different local neighborhood changes across datasets, and UCL retains substantial class overlap. Black rings mark errors, so visually separated neighborhoods are not equivalent to correct classification. Because t-SNE can distort global distances, densities, and cluster sizes, both plots are qualitative fold-level diagnostics (van der Maaten & Hinton, 2008); the fold-aggregated tables remain the primary evidence.

O Compute Environment and Cost Reporting

Table 21 reports the execution conditions that can be recovered reliably from the final runs and explicitly identifies the missing timing logs. The quantum models were evaluated through Penny-Lane's analytic CPU simulator, whereas pure PyTorch baselines could use an NVIDIA T4. Average epoch time, total training time, and inference latency would therefore reflect both model design and different hardware/software backends. Because consistent timers were not retained across the heterogeneous Colab sessions, reconstructing those values after the fact would be misleading. The paper consequently reports parameter counts and predictive metrics, discloses the device and batch differences, and makes no runtime or hardware-efficiency claim.

The stack uses Python, PyTorch, PennyLane, MOABB, MNE, scikit-learn, PyRiemann, NumPy, and SciPy. PyTorch and PennyLane provide model execution (Paszke et al., 2019; Bergholm et al., 2018). Dataset and signal tooling follows MOABB and MNE (Jayaram & Barachant, 2018; Gramfort et al., 2013); classical transforms use scikit-learn (Pedregosa et al., 2011). Parameter counts include tensors with gradients enabled. Final scripts include timing instrumentation for future reruns; the present evidence supports predictive and parameter comparisons only.

Table 19: EEGET-ALS leave-one-view-out results.

EEGET removed view	Macro-F1	Drop
None	86.6±0.4	0.0
Temporal	83.4±0.6	3.2
Spatial	84.2±0.5	2.4
Spectral	83.0±0.6	3.6
Task-statistical	83.8±0.5	2.8
Covariance	82.1±0.7	4.5
Riemannian tangent	81.0±0.7	5.6
CSP log variance	82.7±0.6	3.9

Table 20: Controlled qubit–depth Macro-F1 sensitivity.

	BNCI-P300				UCL ALS-EEG				EEGET-ALS		
Q	D1	D2	D3	Q	D1	D2	D3	Q	D1	D2	D3
2	73.5	75.0	74.3	2	54.8	55.6	55.2	2	77.3	78.8	78.2
4	79.8	82.3	81.6	4	59.5	62.2	61.3	4	81.9	83.6	83.0
6	83.2	**85.6**	84.7	6	63.4	**66.1**	65.0	6	84.2	**85.7**	85.1
8	83.9	85.0	84.0	8	63.6	65.3	64.3	8	84.3	85.2	84.2

P Statistical and Metric Reporting

Reported uncertainties are standard deviations over seeds 0, 1, and 2 under fixed subject-level outer folds. Every seed evaluates the complete fold registry through the same aggregation routine. Macro-F1 averages the class F1 values; balanced accuracy averages class recall; AUROC and AUPRC use class-1 probabilities. Positive classes are P300 target, right imagery, and ALS. Three seeds describe run-to-run dispersion and do not support a formal significance claim.

Why multiple metrics are retained. Macro-F1 and balanced accuracy are threshold-dependent summaries that weight both classes symmetrically; accuracy can remain high when the majority class dominates, as illustrated by BNCI-P300. AUROC evaluates ranking across decision thresholds, while AUPRC is more sensitive to positive-class retrieval under imbalance. Agreement across these measures strengthens an empirical comparison, but the displayed standard deviations quantify only seed-to-seed variation under the fixed folds. They are not trial-level standard errors, subject-population confidence intervals, or evidence of statistical significance.

Aggregation convention. For each seed, the complete outer-fold registry is evaluated with the same preprocessing and metric code, reduced to one seed-level aggregate, and then summarized over the three seeds. This prevents a dataset with more repeated windows from contributing more nominal independent observations to the reported uncertainty. The full-model entries are copied from the same canonical aggregate files wherever they recur in baseline, controlled-comparison, and ablation tables.

Q Scope of Inference and Unperformed Controls

The reported evidence has four boundaries that are important for interpreting the contribution.

Participant uncertainty. The displayed deviations are standard deviations across three optimizer seeds after each seed has been aggregated over the fixed folds. They do not quantify variation across new participants. A population-level claim would require per-fold scores, paired participant-level differences, and a participant- or hierarchy-aware interval. We therefore make no formal significance claim for the 2.0-point UCL margin.

Structured classical encoders. PM-Classical controls parameter count and changes only the principal encoder, but its MLP is not a Fourier or trigonometric network. The experiment cannot distinguish a specifically quantum effect from the benefit of a periodic structured basis. No such result is inferred or imputed.

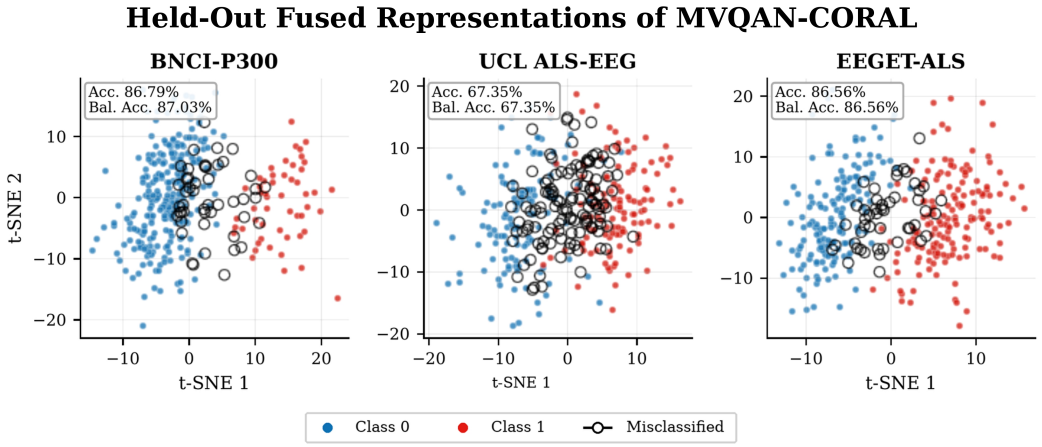


Figure 7: Held-out concatenated-representation t-SNE diagnostics. The figure is qualitative and is not used for model selection or statistical claims.

Table 21: Execution conditions. The runs used changing Colab sessions, and consistent per-epoch, total-training, and inference timers were not retained for every final method. We therefore do not report post-hoc timing estimates or claim runtime advantage.

Model family	Device	Batch size	Epoch cap	Timing log
MVQAN / VQC models	Analytic PennyLane simulator on CPU	32	120	Not retained consistently
Pure PyTorch baselines	NVIDIA T4 when available; CPU fallback	128 GPU; 32 CPU	120	Not retained consistently

Geometry-specific pipelines. The study supplies xDAWN, CSP, covariance, and Riemannian tangent views to the proposed model but does not report same-fold MDM, tangent-space LDA/SVM, or FBCSP pipelines. The complete-system table therefore establishes competitiveness only against the seven executed reimplementations. It does not establish that the hybrid model is necessary relative to every classical geometry-based pipeline.

Application endpoints. BNCI is evaluated at the target/non-target epoch level; no character-selection curve or information-transfer rate was retained. EEGET is evaluated at the 2-s-window level within held-out subject pairs; no participant-level voting or clinical diagnostic analysis was performed. These application-level outcomes require additional experiments.

R Reproducibility Audit

The accompanying reproducibility materials record fixed fold identities and class counts, train-only fitted transformations, the modality registry, shared metric code, saved seeds and configurations, and scripts for parameter matching, ablations, sensitivity analyses, tables, and figures. Repeated full-model entries are generated from the same canonical aggregate results. LMDA-Net and EEG Conformer parameter counts remain approximate because only rounded totals for the executed task-adapted configurations were retained.