

Unified EEG Classification Across Subjects and BCI Paradigms Using Complementary Geometric and Spectral Representations

Wentuo Liang, Roneel V. Sharan, and Junhua Li*, *Senior Member, IEEE*

Abstract— Deep learning is widely used for electroencephalography (EEG) decoding, but most models remain paradigm-specific. Unified decoding requires a single model to learn information across paradigms while retaining paradigm-dependent discriminative information. We propose the Complementary Dual-Expert Network (DuExNet), which is capable of 11-class cross-subject classification including motor imagery (MI), steady-state visual evoked potential (SSVEP), and mental workload (MW) paradigms. DuExNet consists of two complementary pathways. The first pathway learns a common global second-order representation enriched with spectral-statistical information. The second pathway captures sample-adaptive multiscale covariance structure to accommodate paradigm-dependent temporal variation. The calibrated probabilities from both pathways are combined by weighted fusion. We evaluated DuExNet using nine-fold cross-validation on three public datasets and compared it with DeepConvNet, EEGNet, CRGNet, LMDA-Net, and EEG-Deformer. DuExNet achieved an overall accuracy of 65.69% and a macro-averaged F1 score (Macro-F1) of 68.53%, exceeding the best counterpart by 4.25 and 4.55 percentage points, respectively. The statistical analysis showed that the performance improvement of the proposed model is significant. In addition, the ablation study demonstrated the effectiveness of combining the two pathways. Our study provides an effective model for cross-subject and cross-BCI-paradigm classification, aiming for universal models.

Clinical Relevance— DuExNet provides a single cross-subject decoder for motor imagery, SSVEP, and mental workload EEG. By reducing reliance on paradigm-specific models, this unified framework simplifies the development of multi-paradigm BCIs for neurorehabilitation, assistive communication, and cognitive state monitoring.

I. INTRODUCTION

Deep learning enables end-to-end learning of task-relevant representations from raw or minimally preprocessed EEG signals [1]. Existing EEG decoders employ diverse architectures and representation strategies, including convolutional networks, attention mechanisms, Transformers, and symmetric positive-definite (SPD) manifold modeling [2]–[6]. However, most are designed and optimized for individual EEG paradigms, requiring separate models for different brain-computer interface (BCI) tasks. Unified classification therefore requires a single model that can capture common structure across paradigms

while preserving paradigm-dependent discriminative information.

This objective is challenging because EEG paradigms differ in their neural responses and informative signal structures. MI primarily involves event-related changes in sensorimotor rhythms, whereas SSVEP exhibits stimulus-locked periodic responses. MW assessment instead depends on activity distributed across multiple frequency bands and cognitive-state-related statistics. These differences create distinct requirements for temporal receptive fields, spectral information, and statistical features. Global second-order representations can model dependencies across latent channels, but a single temporal scale may not accommodate sample- and paradigm-dependent variation. Unified decoding therefore requires a model that can jointly capture cross-channel dependencies, spectral and statistical characteristics, and temporal patterns at adaptive scales.

To address this challenge, we introduce DuExNet for a unified 11-class, cross-subject classification problem spanning MI, SSVEP, and MW. DuExNet contains two complementary prediction pathways within a joined latent feature space. The Global Log-Cholesky Backbone (GLCB) and Spectral-Statistical Residual Expert (SSRE) form a joint pathway that models global second-order geometry and supplements it with spectral and low-order temporal-statistical information. The Scale-Adaptive Fréchet Expert (SAFE) captures sample-adaptive multiscale covariance structure to accommodate temporal-scale variation across samples and paradigms. Finally, DuExNet calibrates and combines the probabilities from both pathways through weighted fusion. The main contributions are as follows:

- (1) We introduce DuExNet for unified cross-subject, cross-paradigm EEG classification, designed to capture common structure across paradigms while preserving paradigm-dependent discriminative information.
- (2) We develop SAFE to aggregate multiscale covariance representations adaptively, thereby accommodating temporal-scale variation across samples and EEG paradigms.
- (3) We construct a joint GLCB-SSRE pathway that combines global second-order geometry with spectral and low-order temporal-statistical information.

II. METHODOLOGY

A. Overall Framework

DuExNet consists of a joint GLCB-SSRE pathway and an independent SAFE pathway, as illustrated in Fig. 1. In the joint pathway, GLCB extracts global Log-Cholesky geometric features, while SSRE applies multiscale spectral and low-order temporal-statistical residual corrections to the GLCB logits, producing $\mathbf{z}_J$. In parallel, SAFE adaptively aggregates Log-Cholesky representations across temporal

W. Liang is with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Wuyi University, Jiangmen 529020, China.

R. V. Sharan is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK.

J. Li is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK, and also with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Wuyi University, Jiangmen 529020, China.

* indicates the corresponding author (Junhua Li, junhua.li@essex.ac.uk).

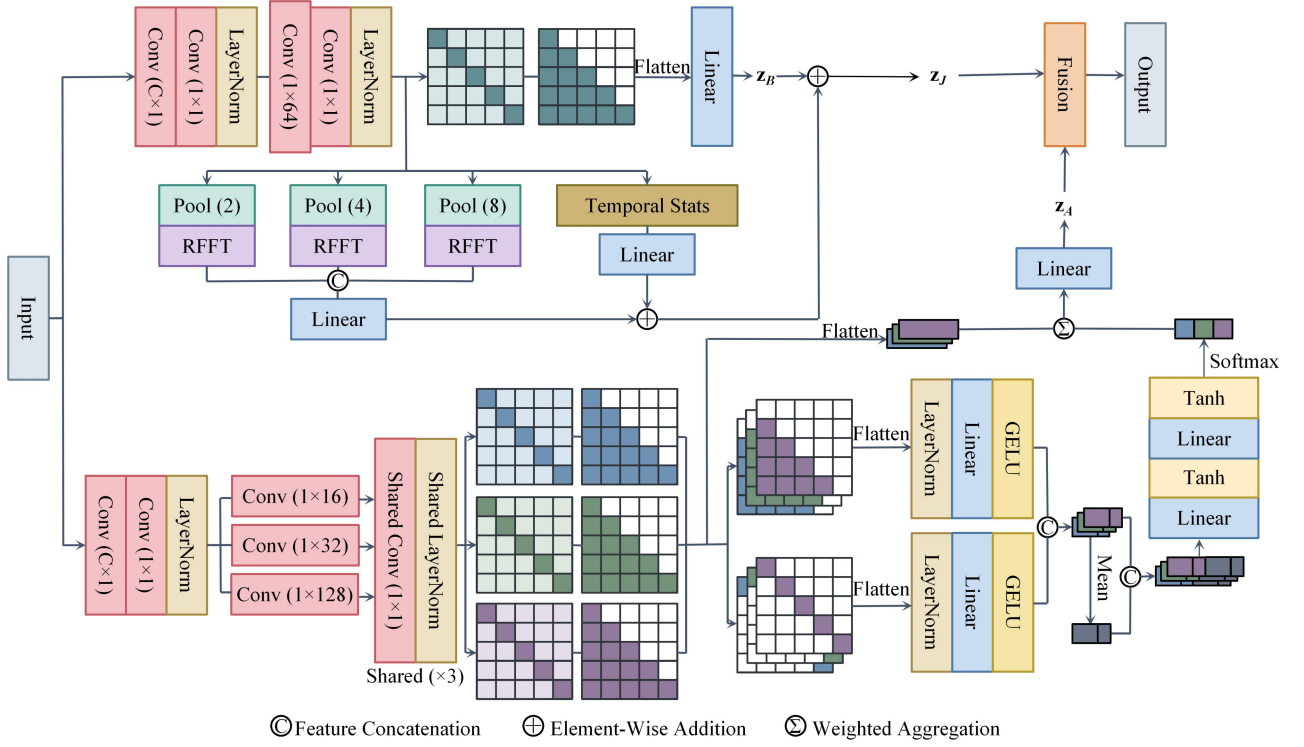


Fig. 1. Overall architecture of DuExNet for unified cross-subject, cross-paradigm EEG classification.

scales to produce $\mathbf{z}_A$. Both $\mathbf{z}_J$ and $\mathbf{z}_A$ belong to $\mathbb{R}^{N_{\text{cls}}}$, where N_{cls} denotes the number of classes. The joint pathway and SAFE are trained separately. After training, both pathways are frozen, and their output probabilities are calibrated using validation data. The calibrated probabilities are then combined through weighted fusion to obtain the final class probabilities.

B. GLCB-SSRE Joint Branch: Global Geometry and Residual Correction

An input EEG sample is denoted by $\mathbf{X} \in \mathbb{R}^{1 \times C \times T}$, where C and T denote the numbers of EEG channels and time points, respectively. GLCB first applies a spatial convolution along the electrode dimension, mapping the input to D latent channels, where $D=100$. The resulting features undergo pointwise convolution for channel mixing, followed by layer normalization along the temporal dimension. The normalized features are then processed by a temporal depthwise convolution with kernel length 64, stride 1, and same padding. A final pointwise convolution and temporal layer normalization yield the latent feature matrix $\mathbf{F} \in \mathbb{R}^{D \times T}$.

The temporal kernel length was selected to provide a receptive field that spans the rhythmic time scales relevant to the different tasks. At a sampling rate of 250 Hz, a 64-sample kernel corresponds to a temporal receptive field of 0.256 s. This receptive field covers at least one complete cycle of signals at frequencies of 4 Hz or higher. For MI, the relevant frequencies include the μ and β rhythms associated with event-related desynchronization and synchronization. The relevant frequency ranges also include the lowest SSVEP stimulation frequency used here (5.45 Hz) and the θ and α bands associated with MW. The selected kernel length therefore provides GLCB with a common local temporal scale that covers the principal rhythmic ranges considered across the three tasks.

Given $\mathbf{F} \in \mathbb{R}^{D \times T}$, GLCB estimates global second-order dependencies between latent channels using all T time points. GLCB temporally centers $\mathbf{F}$ and computes the unbiased sample covariance $\mathbf{S}$ using the denominator $T-1$. Shrinkage and diagonal loading are then applied as

$$\Sigma_\rho = (1-\rho)\mathbf{S} + \left(\rho \frac{\text{tr}(\mathbf{S})}{D} + \epsilon\right)\mathbf{I}_D. \quad (1)$$

Here, $\mathbf{I}_D$ is the D -dimensional identity matrix. We set $\rho=10^{-3}$ and $\epsilon=10^{-6}$ in all experiments. The shrinkage term partially contracts the sample covariance toward a scale-matched isotropic matrix, reducing covariance-estimation instability under finite sampling [7]. The diagonal-loading term $\epsilon\mathbf{I}_D$ ensures that Σ_ρ is strictly positive definite, thereby supporting stable Cholesky decomposition.

GLCB then performs Cholesky decomposition of the regularized covariance matrix and defines its Log-Cholesky coordinates as

$$\Sigma_\rho = \mathbf{L}\mathbf{L}^\top, \Phi(\Sigma_\rho) = [\text{vecl}_{\text{strict}}(\mathbf{L}), \log(\text{diag}(\mathbf{L}))]. \quad (2)$$

Here, $\Phi(\cdot)$ denotes the Log-Cholesky map from an SPD matrix to a Euclidean coordinate vector [8]. The matrix $\mathbf{L}$ is lower triangular with strictly positive diagonal entries. The operator $\text{vecl}_{\text{strict}}(\cdot)$ extracts the strictly lower-triangular entries in a fixed order. Consequently, $\Phi(\Sigma_\rho)$ has dimensionality $D(D+1)/2$. When $D=100$, the strictly lower-triangular and log-diagonal components contain 4,950 and 100 elements, respectively, yielding a 5,050-dimensional Log-Cholesky representation. This representation is passed through a linear classifier to produce the GLCB classification logits $\mathbf{z}_B$.

Given the 4-s input windows sampled at 250 Hz, and because the preceding convolutions preserve temporal

length, $\mathbf{F}$ contains 1,000 time points. For each pooling factor $w \in \{2, 4, 8\}$, SSRE applies temporal average pooling with both window length and stride equal to w , followed by a real fast Fourier transform (RFFT) with orthonormal normalization. The averaged spectral magnitude of each latent channel is then computed within predefined left-closed, right-open frequency bands constrained by the post-pooling Nyquist frequency. For $w=2$, the bands are [1,4), [4,8), [8,13), [13,20), [20,30), [30,45), and [45,60) Hz. For $w=4$, they are [1,4), [4,8), [8,13), [13,20), and [20,31) Hz. For $w=8$, they are [1,4), [4,8), [8,12), and [12,15.625) Hz. Concatenating the resulting feature vectors of dimensions $7D$, $5D$, and $4D$ yields the $16D$ -dimensional spectral feature vector $\mathbf{c}_{\text{spec}}(\mathbf{F})$.

SSRE also computes the temporal mean, population standard deviation, mean absolute value, and root-mean-square value of each latent channel from the unpooled $\mathbf{F}$. These statistics form the $4D$ -dimensional temporal-statistical feature vector $\mathbf{c}_{\text{stat}}(\mathbf{F})$. Separate linear residual heads map $\mathbf{c}_{\text{spec}}(\mathbf{F})$ and $\mathbf{c}_{\text{stat}}(\mathbf{F})$ to the residual-logit vectors $\mathbf{r}_{\text{spec}}$ and $\mathbf{r}_{\text{stat}}$, respectively. The two residual-logit vectors are equally weighted and added to the GLCB logits as $\mathbf{z}_J = \mathbf{z}_B + (\mathbf{r}_{\text{spec}} + \mathbf{r}_{\text{stat}})/2$.

C. SAFE: Scale-Adaptive Fréchet Expert

SAFE constructs covariance representations at three temporal scales and aggregates them through a sample-adaptive weighted Fréchet mean. It first applies a $C \times 1$ spatial convolution to map the input to $D=100$ latent channels. A pointwise convolution and temporal layer normalization then produce a shared latent feature basis. This basis is processed by three parallel temporal depthwise convolutions with kernel lengths $K \in \{16, 32, 128\}$, where K denotes the temporal kernel length. Each convolution uses a stride of 1 and same padding, thereby preserving the temporal length T . Unlike the single length-64 kernel used by GLCB, these kernels capture complementary local temporal scales.

The three branch outputs are passed in turn through a single shared 1×1 convolution followed by a single shared temporal layer-normalization module. These two modules reuse identical parameters for all three branches, whereas the temporal depthwise convolutions have branch-specific parameters. This process produces a latent feature matrix $\mathbf{F}_m \in \mathbb{R}^{D \times T}$ for each scale $m \in \{1, 2, 3\}$. The shared post-processing modules maintain a consistent interpretation of corresponding latent channels across scales. Consequently, the scale-specific covariance matrices can be represented in a common Log-Cholesky coordinate system.

For each scale, SAFE temporally centers $\mathbf{F}_m$, computes the shrinkage covariance matrix $\mathbf{\Sigma}_m \in \text{SPD}(100)$, and applies the Log-Cholesky mapping defined in Section II-B. The resulting 5,050-dimensional coordinates are partitioned into the strictly lower-triangular component $\mathbf{o}_m$ and the log-diagonal component $\mathbf{d}_m$ of the Cholesky factor.

Two coordinate encoders shared across scales process $\mathbf{o}_m$ and $\mathbf{d}_m$ separately using layer normalization, linear projection, and GELU activation. The resulting 64- and 16-dimensional vectors are concatenated into an 80-dimensional scale embedding used only to estimate the routing weights. The averaged embedding across the three scales is concatenated with each scale-specific embedding to form a 160-dimensional scoring vector. A shared scoring

network consisting of a $160 \rightarrow 32$ linear layer, a tanh activation, and a $32 \rightarrow 1$ linear layer produces an unconstrained scalar score r_m . The scores are subsequently bounded by a second tanh operation before temperature-scaled softmax normalization. The output layer of the scoring network is initialized to zero, assigning equal weights to the three scales at the start of training. Let $\mathbf{r} = (r_1, r_2, r_3)$. The final scale weights are

$$\alpha_m = 0.85 \text{ softmax}_m \left(\frac{2 \tanh(r/2)}{1.5} \right) + 0.05, \quad (3)$$

which ensures that $\alpha_m \geq 0.05$ and $\sum_{m=1}^3 \alpha_m = 1$.

SAFE then computes the weighted Fréchet mean by forming a convex combination of the uncompressed scale-specific coordinates under the Log-Cholesky metric [8]:

$$\Phi(\mathbf{\Sigma}) = \sum_{m=1}^3 \alpha_m \Phi(\mathbf{\Sigma}_m). \quad (4)$$

Finally, a linear classifier maps $\Phi(\mathbf{\Sigma})$ to the SAFE logits $\mathbf{z}_A$.

D. Training and Prediction Fusion in DuExNet

The joint GLCB-SSRE branch and SAFE are trained independently. Training of the joint branch proceeds in two stages. GLCB is first optimized using classification cross-entropy and is subsequently frozen. Its classification logits, spectral features, and temporal statistics are then extracted from the training and validation sets. The spectral and temporal-statistical features are standardized separately using their corresponding training-set means and standard deviations, with the standard deviation bounded below by 10^{-4} . The standardized features are passed through their respective zero-initialized linear residual heads, which are trained independently and sequentially. For $k \in \{\text{spec}, \text{stat}\}$, each head is optimized using

$$L_k = \text{CE}(\mathbf{z}_B + \mathbf{r}_k, y) + 0.1 \text{ mean}(\mathbf{r}_k^2). \quad (5)$$

Here, $\text{CE}(\cdot, y)$ denotes the multiclass cross-entropy loss between the predicted logits and ground-truth labels, and $\text{mean}(\cdot)$ denotes the arithmetic mean over all entries of its argument, namely, across the mini-batch and class dimensions. The second term penalizes excessively large residual logits. After training and separate checkpoint selection for each head based on validation accuracy, the complementary logits are computed as $\mathbf{z}_C = (\mathbf{r}_{\text{spec}} + \mathbf{r}_{\text{stat}})/2$, and the joint-branch logits are given by $\mathbf{z}_J = \mathbf{z}_B + \mathbf{z}_C$.

SAFE is trained using a combination of classification cross-entropy and scale-weight regularization. The regularizer contains a batch-level load-balancing term and a sample-level weight-smoothing term. The first discourages systematic routing collapse toward a single scale, whereas the weaker second term discourages overly concentrated routing at the sample level. Let B denote the batch size and $\mathbf{a}_i$ the scale-weight vector for sample i . The batch-average weight vector is $\bar{\mathbf{a}} = B^{-1} \sum_{i=1}^B \mathbf{a}_i$, and $\mathbf{u} = (1/3, 1/3, 1/3)$ denotes the uniform distribution. The SAFE objective is

$$L_A = \text{CE}(\mathbf{z}_A, y) + 0.01 \left[D_{\text{KL}}(\bar{\mathbf{a}} \parallel \mathbf{u}) + \frac{0.1}{B} \sum_{i=1}^B D_{\text{KL}}(\mathbf{a}_i \parallel \mathbf{u}) \right]. \quad (6)$$

Here, $D_{KL}(\mathbf{s}||\mathbf{q})=\sum_{m=1}^3 s_m \log(s_m/q_m)$ denotes the Kullback-Leibler divergence. The coefficient 0.01 controls the overall regularization strength, whereas 0.1 controls the sample-level contribution relative to the batch-level term.

After both pathways have been trained, all model parameters are frozen. Following the temperature-scaling approach in [9], pathway-specific temperatures are introduced before probability-level fusion. The temperature parameters and fusion weight are jointly selected using the validation-set logits. The resulting calibrated probabilities are then fused as

$$\mathbf{p}=\lambda \text{softmax}\left(\frac{\mathbf{z}_J}{T_J}\right)+(1-\lambda) \text{softmax}\left(\frac{\mathbf{z}_A}{T_A}\right). \quad (7)$$

Here, $T_J>0$ and $T_A>0$ are the pathway-specific temperatures, and $\lambda\in[0,1]$ is the probability weight assigned to the joint branch. The parameter triplet (λ, T_J, T_A) is selected jointly using the validation-set logits. We evaluate $\lambda\in\{0,0.025,\dots,1\}$, with both temperatures selected from $\{0.5,0.67,0.8,1.0,1.25,1.5,2.0,2.5,3.0\}$. Candidate combinations are ranked sequentially by higher validation accuracy, lower negative log-likelihood, and proximity to the neutral setting (0.5,1,1), where (0.5,1,1) corresponds to equal pathway weighting without temperature rescaling. The final prediction is obtained from the fused probability vector $\mathbf{p}$.

The test set is excluded from training and all model-selection procedures, including standardization, checkpoint, temperature, and fusion-weight selection, and is used only for final evaluation.

III. EXPERIMENTAL SETUP

A. Datasets and Data Partitioning

We used three public EEG datasets to evaluate unified cross-paradigm classification under a cross-subject setting. Table I summarizes the principal characteristics of the three datasets. BCI Competition IV 2a was used for MI with four classes: left hand, right hand, both feet and tongue [10]. OpenBMI was used for SSVEP with four stimulation frequencies: 5.45, 6.67, 8.57 and 12 Hz [11]. The MATB-II subset of COG-BCI was used for MW with low, medium and high mental workload levels [12].

To mitigate the potential influence of unequal sample sizes across paradigms, we included data from nine subjects in each dataset. This number was determined by BCI Competition IV 2a, which contained the fewest subjects. For OpenBMI and COG-BCI, the first nine subjects in ascending order of their original identifiers were selected. This selection was prespecified before model training and was independent of signal quality or classification performance. Subjects were organized into nine cross-dataset groups according to their original within-dataset identifiers. Each group comprised one subject from each dataset. The matching identifiers were

TABLE I
CHARACTERISTICS OF THE EEG DATASETS USED FOR UNIFIED
CROSS-PARADIGM CLASSIFICATION

Paradigm	Classes	Channels	f_s (Hz)	Sessions/Trials
MW	3	62	500	3/222
MI	4	22	250	2/288
SSVEP	4	62	1000	2/200

Note: f_s denotes sampling frequency.

used only for data organization and did not indicate that the three subjects were the same individual.

We then performed nine-fold cross-validation. In each fold, one group was held out for testing, one of the remaining eight groups was selected for validation using a fixed random seed, and the other seven groups were used for training. Each group served as the test set exactly once, and all models were evaluated using identical fold assignments.

B. Data Preprocessing

OpenBMI and COG-BCI were notch-filtered at 60 and 50 Hz, respectively. All three datasets were subsequently bandpass-filtered between 0.5 and 75 Hz. Using the 250 Hz sampling rate of BCI Competition IV 2a as the common rate, we downsampled OpenBMI and COG-BCI to 250 Hz. The continuous recordings from all three datasets were then segmented into non-overlapping 4-s windows. Spherical spline interpolation [13] was applied to each EEG window to interpolate the scalp potentials onto 19 standard channels of the international 10-20 system. Finally, each channel was z-score normalized across the temporal dimension within each window.

C. Compared Models, DuExNet Training and Evaluation

The compared models were DeepConvNet [2], EEGNet [3], CRGNet [4], LMDA-Net [5], and EEG-Deformer [6]. All models received the same preprocessed inputs and used the same fixed data partitions. Only the input and classification layers were adapted to accommodate 19-channel inputs and 11-class outputs. Except for the common training-duration and model-selection rules, each compared model retained the loss function, optimizer, and remaining hyperparameters reported in the original publication. All compared models were trained for at least 20 and at most 200 epochs, with early stopping enabled after the minimum training period and triggered by 15 consecutive epochs without improvement in validation accuracy.

DuExNet was trained using the procedure described in Section II-D. GLCB and SAFE used AdamW with an initial learning rate of 10^{-3} , a weight decay of 10^{-2} , and a batch size of 64. They followed the same training-duration and model-selection rules as the compared models.

The two SSRE residual heads were trained independently using AdamW with a learning rate of 10^{-2} , a weight decay of 10^{-3} , and a batch size of 1,024. Each head was trained for a minimum of 10 and a maximum of 60 epochs. Training stopped if validation accuracy did not improve for eight consecutive epochs, and the best checkpoint was retained separately for each head.

Model checkpoints, temperature parameters, and fusion weights were determined exclusively from the validation data. The test set was used only for final evaluation after all model and selection parameters had been fixed. We calculated accuracy and Macro-F1 separately for each test fold and report the mean $\pm$ standard deviation across the nine folds.

IV. RESULTS AND DISCUSSION

A. Performance Comparison

Samples from all three paradigms were jointly classified within the unified 11-class task, while paradigm-specific accuracies were calculated separately

TABLE II
CROSS-SUBJECT CLASSIFICATION PERFORMANCE OF DUEXNET AND COMPARED MODELS ACROSS THREE EEG PARADIGMS

Method	MW (%)	MI (%)	SSVEP (%)	Acc (%)	Macro-F1 (%)
DeepConvNet	53.49±10.14	45.35±8.27	91.53±6.32	59.90±5.18*	62.35±5.44
EEGNet	45.65±14.67	46.93±8.06	91.53±6.30	57.27±5.95*	59.53±6.11
CRGNet	53.87±9.18	48.50±9.68	92.36±6.66	61.36±5.68*	63.04±5.51
LMDA-Net	53.59±7.48	44.71±7.35	92.69±4.71	60.00±4.85*	62.55±5.09
EEG-Deformer	54.45±9.22	46.86±8.78	94.06±4.99	61.44±5.11**	63.98±4.03
Ours	56.79±8.42	54.71±14.68	96.33±3.47	65.69±5.75	68.53±5.80

Note: Values are percentages reported as mean ± standard deviation over nine folds. Boldface indicates the best averaged performance. Asterisks denote statistical significance (* $p < 0.05$, ** $p < 0.01$) using two-sided paired t-tests.

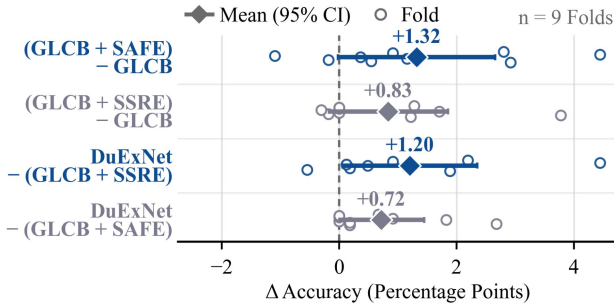


Fig. 2. Foldwise accuracy gains from adding SAFE and SSRE in the DuExNet ablation analysis.

only for performance analysis. As shown in Table II, DuExNet achieved the highest accuracies for MW, MI, and SSVEP under the common preprocessing pipeline and nine-fold cross-subject evaluation. Compared with the best counterpart for each paradigm, DuExNet improved accuracy by 2.34, 6.21, and 2.27 percentage points, respectively. It also achieved the highest overall accuracy and Macro-F1 score, exceeding the best counterpart by 4.25 and 4.55 percentage points, respectively. Paired t-tests further confirmed that DuExNet achieved significantly higher overall accuracy than each of the five comparison models.

B. Ablation Analysis

We evaluated GLCB, SAFE, GLCB + SAFE, GLCB + SSRE, and the full DuExNet using the same data partitions and preprocessing. Components shared across configurations were trained using the same optimization settings, epoch limits, early-stopping rules, and checkpoint-selection criteria.

Table III shows that adding either SAFE or SSRE to GLCB improved both accuracy and Macro-F1. The complete DuExNet also outperformed the two partial variants, GLCB + SAFE and GLCB + SSRE, on both metrics. Fig. 2 further shows that adding SAFE to GLCB + SSRE yielded positive foldwise accuracy gains in most folds. Adding SSRE to GLCB + SAFE likewise yielded positive accuracy gains in most folds.

TABLE III
ABLATION RESULTS FOR GLCB, SSRE, AND SAFE

Method	Acc (%)	Macro-F1 (%)
GLCB	63.66±5.81	66.28±5.56
SAFE	63.40±5.99	66.05±6.91
GLCB + SAFE	64.97±6.13	67.77±6.24
GLCB + SSRE	64.49±5.81	67.20±5.38
DuExNet (Full)	65.69±5.75	68.53±5.80

Note: Boldface indicates the best averaged performance.

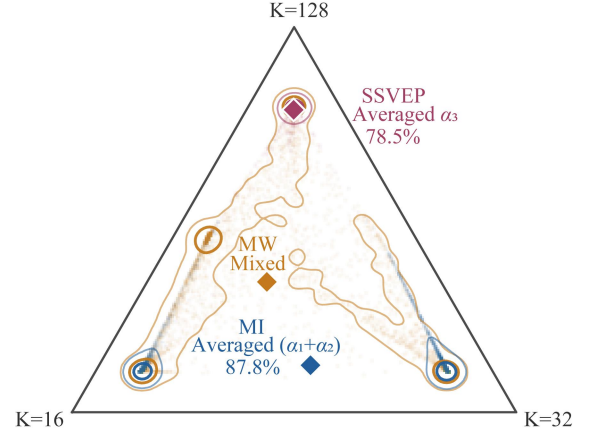


Fig. 3. Distribution of SAFE scale-routing weights across EEG paradigms. Percentages indicate the averaged weight for $K = 128$ in SSVEP and the averaged combined weight for $K = 16$ and $K = 32$ in MI.

GLCB models single-scale global geometry, SSRE supplies spectral and low-order temporal-statistical information, and SAFE captures multiscale geometric variation. The complete DuExNet outperformed the partial variants, indicating that integrating the GLCB-SSRE and SAFE pathways improves classification performance.

C. SAFE Scale Routing Across EEG Paradigms

SAFE exhibited distinct scale-routing patterns across paradigms (see Fig. 3). Among the three SAFE kernels, only the 128-sample kernel spanned at least one complete cycle of every SSVEP stimulation frequency at 250 Hz. For SSVEP samples, the averaged routing weight of the 128-sample scale was 78.5%. This pattern indicates a preference for a longer local filtering range when representing stable periodic responses.

For MI samples, the averaged combined routing weight of the 16- and 32-sample scales was 87.8%. SAFE applies each temporal kernel across the complete trial and estimates covariance using all time points. Thus, K controls the local filtering range without changing the overall analysis duration. This pattern suggests a preference for covariance representations derived from shorter local filtering ranges in MI.

MW samples showed no single dominant routing scale and exhibited sample-to-sample variation in scale weights. This result suggests a less pronounced overall temporal-scale preference for MW samples than for SSVEP and MI samples. Together with the ablation results, these routing patterns support the conclusion that SAFE adapts its scale aggregation across samples and paradigms.

V. CONCLUSION

DuExNet combines global Log-Cholesky geometry, spectral-statistical residuals, and multiscale covariance aggregation for unified cross-subject, cross-paradigm EEG classification. It achieved the highest accuracy for all three paradigms and outperformed all compared models in overall accuracy and Macro-F1. The ablation results further support the effectiveness of integrating the GLCB-SSRE and SAFE pathways. The routing patterns also revealed distinct scale-routing behaviors across paradigms.

Evaluation was limited to three datasets with nine subjects per paradigm, each paradigm drawn from a different dataset, and computational cost was not assessed. Future work should test larger cohorts, broader paradigms, consistent acquisition conditions, and deployment efficiency.

REFERENCES

- [1] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, Art. no. 051001, 2019.
- [2] R. T. Schirrmester et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [3] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Art. no. 056013, 2018.
- [4] C. Gao, W. Liu, and X. Yang, "Convolutional neural network and Riemannian geometry hybrid approach for motor imagery classification," *Neurocomputing*, vol. 507, pp. 180–190, 2022.
- [5] Z. Miao, M. Zhao, X. Zhang, and D. Ming, "LMMA-Net: A lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability," *NeuroImage*, vol. 276, Art. no. 120209, 2023.
- [6] Y. Ding et al., "EEG-Deformer: A dense convolutional transformer for brain-computer interfaces," *IEEE J. Biomed. Health Inform.*, vol. 29, no. 3, pp. 1909–1918, 2025.
- [7] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivar. Anal.*, vol. 88, no. 2, pp. 365–411, 2004.
- [8] Z. Lin, "Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 40, no. 4, pp. 1353–1370, 2019.
- [9] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, Sydney, NSW, Australia, 2017, pp. 1321–1330.
- [10] C. Brunner, R. Leeb, G. R. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI Competition 2008—Graz data set A," Graz Univ. Technol., Graz, Austria, 2008. [Online]. Available: https://www.bbci.de/competition/IV/desc_2a.pdf
- [11] M.-H. Lee et al., "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, Art. no. giz002, 2019.
- [12] M. F. Hinss, E. S. Jahanpour, B. Somon, L. Pluchon, F. Dehais, and R. N. Roy, "Open multi-session and multi-task EEG cognitive dataset for passive brain-computer interface applications," *Sci. Data*, vol. 10, Art. no. 85, 2023.
- [13] F. Perrin, J. Pernier, O. Bertrand, and J.-F. Echallier, "Spherical splines for scalp potential and current density mapping," *Electroencephalogr. Clin. Neurophysiol.*, vol. 72, no. 2, pp. 184–187, 1989.