Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Learned Population-Level Electrode Reduction for Motor Imagery BCI Decoding

Seema Devi

seema.devi@vips.edu

Vivekananda Institute of Professional Studies

Research Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11745609/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Learned Population-Level Electrode Reduction for Motor Imagery BCI Decoding

Seema Devi[1*]

[1] School of Engineering and Technology, Vivekananda Institute of Professional Studies-Technical Campus (VIPS-TC), Delhi, 110034, Delhi, India.

Corresponding author(s). E-mail(s): seema.devi@vips.edu;

Abstract

Motor imagery (MI) brain–computer interfaces (BCIs) can help people with motor disabilities, but EEG systems with many electrodes are difficult to set up and use. Many channel-selection methods choose electrodes separately for each user and are tested on only one dataset. In this study, we learn one common set of electrodes that can be used for all users without reselection. We train a squeeze-and-excitation (SE) channel-gating module on BCI Competition IV-2a and average its channel scores across subjects and cross-validation folds. This reduces the 22-channel montage to 14 channels. Using EEGNet, DeepConvNet, EEGTCNet, and ShallowConvNet, the 14-channel montage showed no significant difference in accuracy from the full 22 channels ($\boldsymbol{p > 0.05}$) and reduced computation (FLOPs) by 7–36%. We then tested the same 14 channels, without any re-selection, on the 54-subject OpenBMI dataset. They gave higher accuracy than the 20-channel motor-cortex baseline for all four models (+1.91 to +2.61 percentage points, $\boldsymbol{p < 0.05}$). The two channel sets are not identical in coverage, so part of this gain may come from the different electrodes included. Most selected channels lie over central and centro-parietal areas (C3, C4, C5, Cz, CP3, CP4), and 10 of the top 12 channels match a montage based on known motor-cortex regions. These results show that one learned set of electrodes can be reused across models and datasets, which can make MI-BCI systems simpler and cheaper.

Keywords: Brain–Computer Interface (BCI), EEG Electrode Selection, Squeeze-and-Excitation Attention, Motor Imagery, Deep Learning

1 Introduction

Brain–Computer Interface (BCI) systems translate brain activity into commands for external devices [1]. Among different BCI paradigms, motor imagery (MI) has been widely investigated for neurorehabilitation, assistive technologies, robotic control, gaming, and smart environments. Electroencephalography (EEG) is commonly used to record brain activity because it is non-invasive, relatively inexpensive, portable, and safe. However, the use of a large number of electrodes increases acquisition complexity and may introduce redundant or noisy information, while also increasing computational and practical requirements. Therefore, identifying a compact set of informative EEG channels is important for developing practical MI-BCI systems.

EEG channel-selection methods can generally be categorized into filter, wrapper, and embedded approaches. Filter methods rank channels using statistical, spatial, or feature-based criteria independently of the classifier [2–5]. Although computationally efficient, they may not capture channel relevance for the final classification task. Wrapper methods select channel subsets according to classifier performance [6], but their repeated evaluation of different channel combinations can be computationally expensive. In contrast, embedded approaches learn channel relevance jointly with the classification model, providing a task-dependent alternative to filter-based selection. Attention mechanisms have already been used for embedded channel selection in MI-EEG. Zhang et al. [12] used a sparse Squeeze-and-Excitation (SE) module, and Tong et al. [14] used efficient channel attention, with channel weights learned while training on each subject's data. Such subject-specific selection requires calibration data from every new user before a montage can be fixed, and it remains unclear whether a selected set is stable across subjects or effective on an independent dataset with a different recording system and electrode layout.

In this work, we use SE attention [7] to score channels for each subject and aggregate the scores across subjects and cross-validation folds into a single population-level ranking, from which a reduced 14-channel montage is selected from the 22-channel BCI Competition IV-2a montage. We then fix this montage and evaluate it, without re-selection, on an independent dataset (OpenBMI) and across different dl architectures.

The proposed montage is evaluated using EEGNet, DeepConvNet, EEGTCNet, and ShallowConvNet and compared with the complete 22-channel configuration using statistical testing. Furthermore, the learned 14-channel configuration is independently evaluated on the OpenBMI dataset to assess its cross-subject and cross-dataset applicability. To assess whether the SE-attention mechanism captures neurophysiologically meaningful information, we examined the top-12 selected channels and compared them against a predefined montage based on established motor imagery-related brain regions. The main contributions of this study are as follows:

- We propose an SE-attention-based channel-scoring method that learns per-channel importance end-to-end and aggregates it across subjects and folds into a global channel ranking.
- The resulting 14-channel montage is evaluated against the full 22-channel montage across four DL architectures (EEGNet, DeepConvNet, EEGTCNet, ShallowConvNet) using paired statistical testing.
- A channel-count ablation (8–22 channels) characterizes the accuracy–channel-count trade-off across all architectures.
- The top-12 attention-based channels are compared with a fixed, neurophysiologically motivated 12-channel montage to assess alignment with expert knowledge.
- The generalized channel set is validated on the independent OpenBMI dataset.

2 Methodology

2.1 Dataset

Two publicly available MI-EEG datasets were used: BCI Competition IV-2a (BCIC-IV-2a) [11] and OpenBMI [13]. BCIC-IV-2a contains EEG recordings from 9 subjects performing four-class MI tasks (left hand, right hand, feet, and tongue), recorded using 22 Ag/AgCl electrodes at 250 Hz over two sessions, with 288 trials per session. This dataset was used to learn the SE-Attention channel weights and generate the global channel ranking.

OpenBMI was used for independent validation of the selected montage. It contains left- and right-hand MI recordings from 54 subjects, acquired using 62 electrodes at 1000 Hz over two sessions. A predefined 20-channel motor-cortex subset was used as the baseline (All_20ch). The 14 channels selected from BCIC-IV-2a were mapped to OpenBMI by electrode labels and evaluated directly without re-ranking or dataset-specific channel selection. Since some selected channels (e.g., Fz, POz, P1, and P2) are outside the 20-channel baseline, the two montages are not nested. This evaluation assesses the cross-subject and cross-dataset transferability of the selected 14-channel montage. The complete details of the datasets are provided in Table 1.

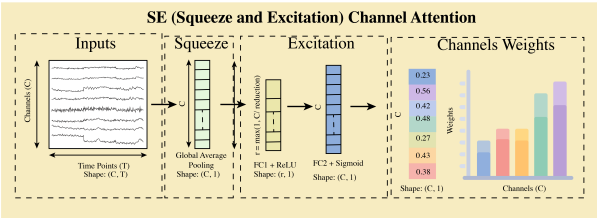

Fig. 1: Illustration of the Squeeze-and-Excitation (SE) channel attention mechanism. Global average pooling first summarizes each EEG channel, followed by two fully connected layers to learn a channel-wise attention weight. The learned weights are then broadcast and multiplied with the input feature map to emphasize informative channels and suppress less informative channels.

2.2 SE-Attention Channel Scoring

MI-EEG trial is represented as $X \in \mathbb{R}^{C\times T}$, where $C = 22$ is the number of MI-EEG channels and $T = 1000$ (2-6 sec) is the number of time samples, an SE-attention module is applied directly to the EEG channel dimension before convolutional feature extraction. The module learns an importance weight for each EEG channel during training as shown in Fig. 1.

First, a squeeze operation summarizes each channel by averaging its signal over time:

$$z_c = \frac{1}{T}\sum_{t=1}^{T} x_c(t), \qquad c = 1, \ldots, C, \tag{1}$$

The resulting channel descriptor $\mathbf{z}$ is passed through a two-layer bottleneck network to generate channel attention weights:

$$\mathbf{s} = \sigma(W_2\, \delta(W_1 \mathbf{z}))\,, \tag{2}$$

where $W_1 \in \mathbb{R}^{(C/r)\times C}$ and $W_2 \in \mathbb{R}^{C\times(C/r)}$ are learned weight matrices, r is the reduction ratio controlling the bottleneck dimension C/r, $\delta(\cdot)$ is a ReLU non-linearity, and $\sigma(\cdot)$ is a sigmoid.

The learned attention weight s_c is then used to rescale the corresponding EEG channel,

$$\mathbf{x}'_c = s_c\, \mathbf{x}_c, \tag{3}$$

before the reweighted EEG signals are passed to the convolutional backbone for feature extraction and classification. The network is trained end-to-end for each subject using 5-fold cross-validation. After training, the attention weights are extracted

Table 1: Details of the dataset used.

Dataset	Subjects	Sessions	Classes	Trials/ Session	Trials/ Class	Electrodes	Sampling Rate
BCIC-2A	9	2	4 (L, R, F, T)	288	72	22 Ag/AgCl	250 Hz
OpenBMI	54	2	2(L, R)	200	100	20* Ag/AgCl	250† Hz

Note: * Predefined 20-channel motor-cortex subset of the 62-channel OpenBMI montage (FC-5/3/1/2/4/6, C-5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6), as used in the original OpenBMI study. The recordings contain 62 electrodes in total. † Downsampled from 1000 Hz to 250 Hz for cross-dataset consistency.

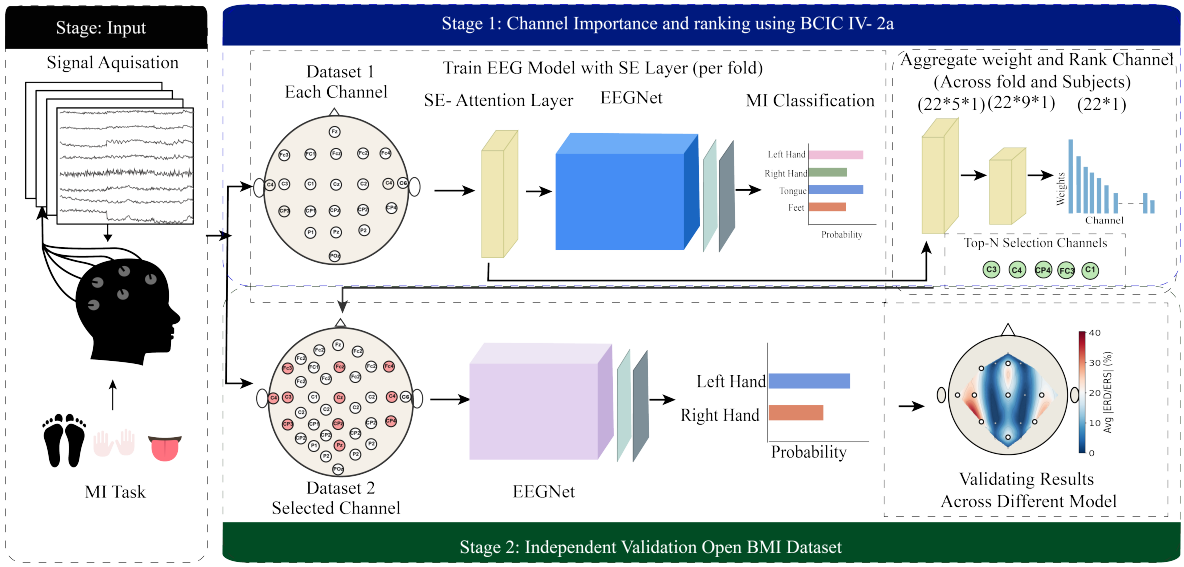

Fig. 2: Illustration of the proposed minimized EEG channel selection method. Global SE-Attention channel ranking based on the mean attention weight across subjects and cross-validation folds, followed by validation of the selected channels on a cross-dataset to assess their generalizability.

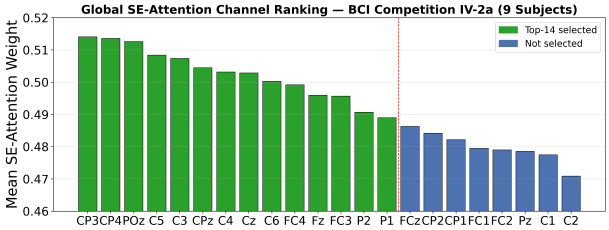

Fig. 3: Global SE-Attention channel ranking (mean attention weight across subjects and folds). The selected top-k channels are highlighted in green (here, $k = 14$).

for every trial and averaged across trials, folds, and subjects to obtain a global importance score for each EEG channel.

2.3 SE-Attention-Based Channel Selection

To identify the most informative EEG electrodes, the channel-wise attention weights produced by the Squeeze-and-Excitation (SE) block are analyzed as shown in Fig. 2. For each subject, the attention weights are averaged across all trials and cross-validation folds to obtain a stable importance score for each channel. These scores are subsequently aggregated to derive a channel-importance ranking.

Table 2: Subject-wise Top-5 SE-Attention channels and their gating weights.

Sub	Top-5 Channels (Weight)				
S1	CP4 (.53)	CP3 (.52)	C5 (.52)	C3 (.51)	CPz (.51)
S2	Cz (.55)	Fz (.54)	FC4 (.53)	C5 (.52)	POz (.52)
S3	C3 (.55)	CPz (.54)	C4 (.52)	CP4 (.51)	POz (.50)
S4	C5 (.52)	POz (.52)	CP3 (.51)	C6 (.51)	P2 (.51)
S5	POz (.52)	Fz (.52)	C5 (.51)	C3 (.51)	P2 (.51)
S6	FCz (.52)	FC3 (.51)	C3 (.51)	CP3 (.51)	CP4 (.50)
S7	C3 (.58)	C4 (.54)	CP3 (.53)	POz (.52)	CPz (.50)
S8	CP3 (.57)	POz (.56)	CP4 (.55)	C5 (.52)	CPz (.52)
S9	C4 (.52)	CP4 (.52)	CPz (.52)	C3 (.51)	POz (.51)

Given the channel importance scores $\{s_1, s_2, \ldots, s_{22}\}$, the set of selected channels is obtained as

$$\mathcal{C}_{\text{attn}}^{(k)} = \text{TopK}\left(\{s_1, \ldots, s_{22}\}, k\right), \qquad (4)$$

where k denotes the desired number of channels and $\text{TopK}(\cdot)$ returns the k channels with the highest attention scores.

Reduced EEG montages are constructed by retaining only the channels in $\mathcal{C}_{\text{attn}}^{(k)}$. Multiple values of k are evaluated to investigate the trade-off between the number of electrodes and classification performance. The resulting channel rankings

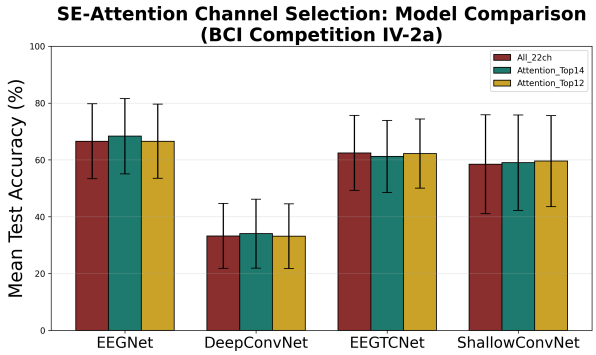

Fig. 4: Mean accuracy (with standard deviation) across the three evaluated channel sets, for EEGNet, DeepConvNet, EEGTCNet, and ShallowConvNet.

and the performance of different reduced-channel montages are presented in the Results section.

2.4 Deep Learning Models

The proposed approach is evaluated using different deep learning (DL) architectures for MI EEG classification: EEGNet [8], DeepConvNet [9], EEGTCNet [10], and ShallowConvNet [9]. These models were selected to represent lightweight, deep convolutional, temporal convolutional, and shallow convolutional architectures, respectively, providing a comprehensive assessment across diverse network designs. SE-attention-based channel importance is learned only once using the EEGNet model. The resulting top-k-channel subset is then fixed and directly applied to all other evaluated architectures, without further channel-selection training or recalculation of channel weights.

3 Results and Discussion

This section evaluates the effectiveness of the proposed SE-attention-based channel selection across several state-of-the-art deep learning architectures. First, the experimentally determined Top-k electrode montage is analyzed to identify an efficient subset of EEG channels. The performance of the selected channel configurations is then evaluated and compared with that of the original 22-channel montage across all architectures. Finally, the learned Top-12 montage is compared with a neurophysiology-motivated 12-channel electrode set to determine whether the proposed data-driven channel selection can reduce the number of electrodes while maintaining or improving classification performance.

3.1 Analysis of Channel Ranking and Electrode Selection

The SE-Attention assigns an importance weight to each EEG channel, enabling the identification of electrodes that contribute most to MI classification. To obtain a robust estimate, the attention weights were averaged across all subjects and cross-validation folds. Fig. 3 provides a visual representation of the resulting global ranking. The top-k channels, highlighted in the figure, were selected as the reduced electrode montage based on the global attention scores.

To further examine subject-specific variability, Table 2 lists the top five channels and their corresponding attention weights for each subject. Several channels, including CP3, CP4, C3, and POz, consistently appear among the highest-ranked electrodes across multiple subjects, indicating their stable contribution to MI decoding. At the same time, some variability is observed, with a few subjects assigning higher importance to channels such as Fz or FCz, reflecting the inherent inter-subject differences in EEG patterns.

Overall, the global and subject-wise analyses provide complementary evidence for channel importance. The consistency of the highly ranked electrodes supports the selection of a common 14-channel montage, while the subject-level analysis highlights the presence of individual differences. The selected reduced-channel montage is subsequently used to evaluate the impact of electrode reduction on different classification models

3.2 Impact of SE-Attention Channel Selection Across Network Architectures

Tables 3–6 present the subject-wise performance of EEGNet, DeepConvNet, EEGTCNet, and ShallowConvNet using the full 22-channel montage and the SE-Attention-based reduced montages (SE_Top14 and SE_Top12). Performance is reported in terms of accuracy, macro F1-score, and Cohen's κ.

EEGNet with the SE_Top14 configuration achieved the highest mean accuracy ($68.35\% \pm 13.24\%$), compared with $66.54\% \pm 13.19\%$ using all 22 channels, while SE_Top12 performs similarly to the baseline. DeepConvNet exhibits nearly identical performance across all channel configurations, indicating limited sensitivity to channel selection. For EEGTCNet, the full 22-channel montage and SE_Top12 produce comparable results, whereas SE_Top14 shows a slight decrease in performance. In contrast, ShallowConvNet shows a gradual improvement with channel reduction, achieving its comparable performance with SE_Top12 ($59.58\% \pm 16.03\%$).

Overall, the proposed SE-Attention-based channel selection maintains or slightly improves classification performance while reducing the number of electrodes. Although EEGNet and ShallowConvNet show modest gains with reduced montages, the statistical analysis in Section 3.3 indicates that these improvements are not statistically significant ($p > 0.05$). Figure 4 summarizes the mean accuracies for all models and channel configurations.

Table 3: Impact of SE-Attention channel selection on EEGNet: per-subject Accuracy (%), F1, and Cohen's κ.

Subject	All_22ch			SE_Top14			SE_Top12		
	Acc(%)	F1	κ	Acc(%)	F1	κ	Acc(%)	F1	κ
S1	77.08	76.73	0.6942	72.74	72.34	0.6363	72.03	71.76	0.6270
S2	53.71	53.46	0.3832	56.75	55.73	0.4241	52.65	52.21	0.3692
S3	84.47	84.44	0.7927	90.55	90.50	0.8739	85.93	85.84	0.8123
S4	50.44	50.32	0.3395	50.88	50.59	0.3454	54.21	53.69	0.3896
S5	60.65	60.70	0.4755	59.71	59.30	0.4623	59.57	59.64	0.4610
S6	46.23	46.54	0.2831	51.91	52.07	0.3588	46.98	47.05	0.2931
S7	77.83	77.68	0.7053	81.66	81.68	0.7559	81.30	81.23	0.7510
S8	77.12	77.05	0.6949	76.31	76.28	0.6840	77.79	77.83	0.7037
S9	71.36	71.14	0.6180	74.62	74.12	0.6615	68.48	67.83	0.5795
Mean ± std	66.54±13.19	66.45	0.5540	68.35±13.24	68.07	0.5780	66.55±13.04	66.34	0.5540

Table 4: Impact of SE-Attention channel selection on DeepConvNet: per-subject Accuracy (%), F1, and Cohen's κ.

Subject	All_22ch			SE_Top14			SE_Top12		
	Acc(%)	F1	κ	Acc(%)	F1	κ	Acc(%)	F1	κ
S1	23.49	12.55	-0.0201	23.49	11.13	-0.0213	25.84	14.21	0.0098
S2	36.11	34.82	0.1489	34.13	31.28	0.1236	38.02	36.31	0.1746
S3	24.54	10.01	-0.0030	24.54	9.86	-0.0000	24.69	10.62	0.0017
S4	28.42	18.79	0.0442	25.88	14.50	0.0023	26.49	17.47	0.0110
S5	50.07	48.87	0.3356	52.10	50.45	0.3626	49.13	47.21	0.3233
S6	26.70	14.51	0.0209	25.12	13.30	0.0020	24.74	11.85	-0.0055
S7	27.08	16.73	0.0241	35.88	27.14	0.1426	25.85	11.79	0.0029
S8	56.83	55.04	0.4241	58.45	56.23	0.4460	56.75	54.15	0.4232
S9	26.06	13.71	0.0187	26.89	13.90	0.0299	26.74	13.63	0.0280
Mean ± std	33.26±11.42	25.00	0.1104	34.05±12.13	25.31	0.1209	33.14±11.40	24.14	0.1077

3.3 Statistical Significance

Table 7 presents the results of the paired t-test and Wilcoxon signed-rank test comparing the reduced 14-channel configurations (SE_Top14) with the full 22-channel montage (All_22ch). Across all DL models, none of the comparisons were statistically significant ($p > 0.05$). Although EEGNet achieved the largest improvement with the SE_Top14 montage, this increase was not statistically significant.

These results indicate that reducing the electrode set from 22 to 14 or 12 channels using the proposed SE-Attention-based ranking preserves classification performance without a statistically significant loss. Therefore, the selected reduced montages provide a more compact EEG configuration while maintaining performance comparable to the full 22-channel montage on the BCI Competition IV-2a dataset.

3.4 Ablation: Accuracy vs. Number of Selected Channels

To characterize the accuracy/channel-count trade-off more finely, each model was retrained on larger SE-Attention-ranked subsets from 8 to 22 channels (Table 8).

For EEGNet, accuracy rises from 62.86% at 8 channels to a peak of 68.35% at 14 channels, then gently declines toward 66.54% at the full 22-channel montage, as shown in Table 3. DeepConvNet is comparatively insensitive to channel count, fluctuating within a narrow 30.04–34.05% band across the full ablation range, again reflecting a persistent architecture-level ceiling rather than a channel-selection effect. EEGTCNet's reduced-montage accuracy peaks at 12 channels (62.21%) among subsets below 22, but is exceeded by the full 22-channel montage (62.47%), a non-monotonic pattern suggesting that, for this architecture, channel count interacts with the specific subset chosen rather than following a smooth trend. ShallowConvNet peaks at 16 channels (59.71%) rather than at 12 or 14, indicating that the optimal montage size is itself somewhat backbone-dependent: EEGNet's peak is at 14 channels, EEGTCNet's at 12, and ShallowConvNet's at 16.

In other words, none of the 22-vs-reduced comparisons reach statistical significance regardless of whether 10, 12, or 14 channels are used (Table 7) covers 12 and 14; the same non-significant pattern is expected, so channel count in this 12–14 range is best chosen on efficiency grounds for a given backbone rather than on accuracy alone: EEGNet and DeepConvNet can plausibly drop to 10 channels at no cost and a further hardware saving, while

Table 5: Impact of SE-Attention channel selection on EEGTCNet: per-subject Accuracy (%), F1, and Cohen's κ.

Subject	All_22ch			SE_Top14			SE_Top12		
	Acc(%)	F1	κ	Acc(%)	F1	κ	Acc(%)	F1	κ
S1	68.97	68.45	0.5862	67.90	67.52	0.5718	64.98	64.50	0.5329
S2	48.90	48.62	0.3195	47.42	47.00	0.2997	48.13	47.32	0.3095
S3	83.81	83.75	0.7840	78.10	78.17	0.7078	79.19	79.03	0.7223
S4	47.46	46.92	0.3002	45.88	45.07	0.2783	49.56	49.00	0.3277
S5	58.26	58.26	0.4449	59.13	58.91	0.4557	60.80	59.83	0.4783
S6	41.86	42.33	0.2245	41.40	41.63	0.2183	42.70	40.87	0.2353
S7	70.69	70.20	0.6103	72.49	72.31	0.6341	75.23	74.91	0.6704
S8	72.77	73.19	0.6367	73.73	73.95	0.6497	72.62	72.88	0.6349
S9	69.47	68.93	0.5925	64.92	64.80	0.5321	66.67	66.16	0.5555
Mean ± std	62.47±13.22	62.29	0.4999	61.22±12.69	61.04	0.4831	62.21±12.18	61.61	0.4963

Table 6: Impact of SE-Attention channel selection on ShallowConvNet: per-subject Accuracy (%), F1, and Cohen's κ.

Subject	All_22ch			SE_Top14			SE_Top12		
	Acc(%)	F1	κ	Acc(%)	F1	κ	Acc(%)	F1	κ
S1	65.20	64.62	0.5357	69.11	68.32	0.5879	67.62	66.66	0.5679
S2	44.45	42.68	0.2600	47.00	44.82	0.2943	46.29	44.09	0.2848
S3	82.86	82.29	0.7713	82.49	81.90	0.7664	81.39	80.80	0.7517
S4	44.39	44.17	0.2592	45.88	45.73	0.2786	46.49	46.40	0.2868
S5	36.45	34.54	0.1562	35.36	33.79	0.1410	44.93	44.02	0.2669
S6	35.07	34.28	0.1345	37.12	36.75	0.1615	33.21	32.03	0.1101
S7	78.56	78.27	0.7148	77.69	77.38	0.7035	77.18	76.55	0.6967
S8	71.51	71.81	0.6199	71.14	71.42	0.6149	71.37	71.63	0.6179
S9	67.73	66.69	0.5694	65.30	64.51	0.5371	67.73	66.95	0.5696
Mean ± std	58.47±17.41	57.71	0.4468	59.01±16.81	58.29	0.4539	59.58±16.03	58.79	0.4614

EEGTCNet and ShallowConvNet are better served by keeping 12 channels.

3.5 Comparison with a Neurophysiology-Based Montage

To assess how closely the learned SE-Attention ranking agrees with domain knowledge, the SE_Top12 montage was compared against a fixed 12-channel montage built from neurophysiological priors (electrodes over the contralateral/midline motor cortex regions associated with each MI class): FC3, FCz, FC4, C5, C3, Cz, C4, C6, CP3, CPz, CP4, Pz. Ten of the twelve channels are shared between the two montages (C3, C4, C5, C6, Cz, FC3, FC4, CP3, CP4, CPz), indicating strong agreement between the purely learned SE-Attention ranking and hand-crafted neurophysiological priors over the sensorimotor strip. The two montages disagree only at the periphery: SE-Attention additionally selects Fz and POz (frontal midline and parieto-occipital electrodes, plausibly capturing residual attentional or visual-cue-related activity), whereas the neurophysiology-based montage instead includes Pz, FCz. This agreement supports the interpretation that the SE block's learned gating weights are recovering genuine, neurophysiologically plausible channel importance rather than an arbitrary artifact of training.

3.6 Comparison with Published Methods

Table 9 compares the proposed SE-Attention method with three recent EEG channel-selection approaches evaluated on the BCIC-IV-2a dataset. The previous methods—ECA-CNN, Signal Power CS, and SNR-based CS—reduce the number of channels but report accuracy losses of 6.24, 8.59, and 2.73 percentage points, respectively.

In contrast, the proposed SE-Attention reduces the 22-channel montage to 14 channels while achieving a small accuracy increase of 1.80 percentage points. However, the improvement is not statistically significant ($p = 0.155$). The selected channels also show strong correspondence with motor-related cortical regions.

Overall, the results indicate that SE-Attention can substantially reduce the EEG channel count while maintaining classification performance. The comparison shows that the proposed approach achieves better performance compared with existing methods, despite differences in data splits, classifiers, and evaluation protocols across studies.

Table 7: Paired statistical tests: each reduced channel set vs. its All_22ch baseline, per model.

Model	Channel Set	ΔAcc(%)	t	p_t	W	p_W
EEGNet	SE_Top14	+1.80	1.569	0.1553 (ns)	12.0	0.2500
EEGNet	SE_Top12	+0.00	0.004	0.9971 (ns)	22.0	1.0000
DeepConvNet	SE_Top14	+0.80	0.706	0.5002 (ns)	12.0	0.7353
DeepConvNet	SE_Top12	-0.12	-0.223	0.8295 (ns)	20.0	0.8203
EEGTCNet	SE_Top14	-1.25	-1.492	0.1741 (ns)	12.0	0.2500
EEGTCNet	SE_Top12	-0.26	-0.247	0.8113 (ns)	20.0	0.8203
ShallowConvNet	SE_Top14	+0.54	0.791	0.4518 (ns)	17.0	0.5508
ShallowConvNet	SE_Top12	+1.11	1.040	0.3287 (ns)	11.0	0.3828

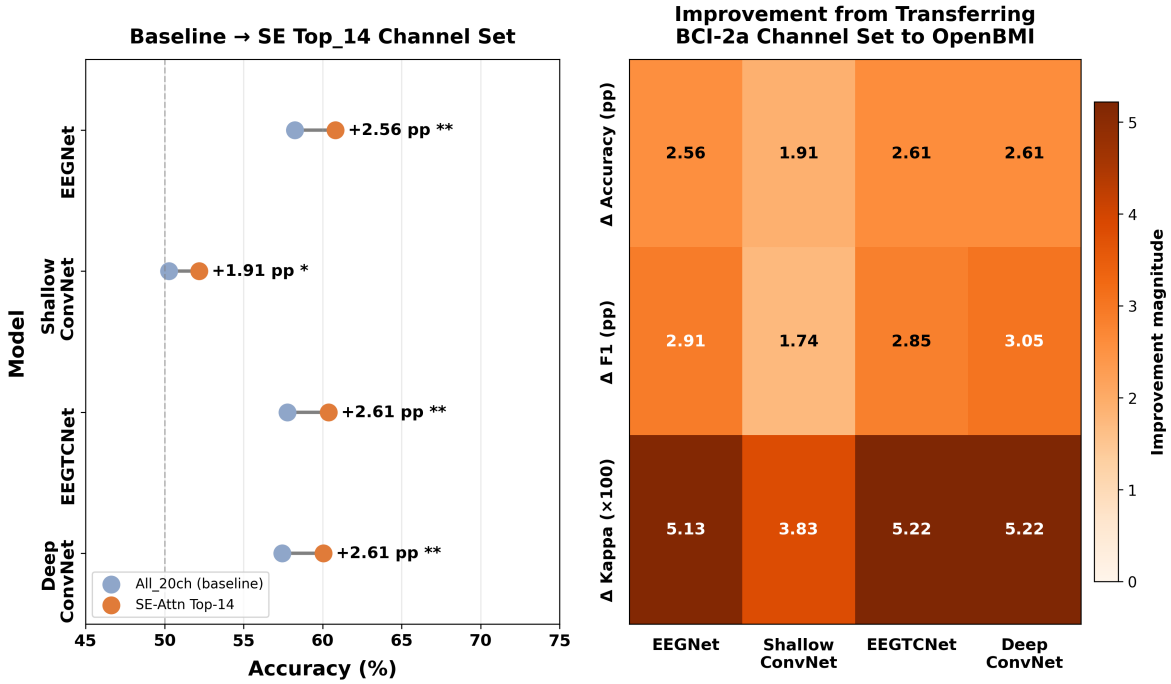

Fig. 5: Accuracy gains and cross-metric improvement from transferring the BCI-IV 2a-derived SE-Attention 14-channel subset to OpenBMI, across four MI decoding models (54 subjects, session 1 → session 2). **(Left)** Dumbbell plot of mean accuracy, baseline vs. transferred channel set, with significance ($^*p < 0.05$, $^{**}p < 0.01$). **(Right)** Heatmap of improvement magnitude across accuracy, F1, and kappa.

Table 8: Ablation study: mean accuracy (%) vs. number of SE-Attention selected channels.

Model	8ch	10ch	12ch	14ch	16ch	18ch	20ch	22ch
EEGNet	62.86	64.09	66.55	68.35	67.40	65.40	66.06	66.54
DeepNet	30.04	30.73	33.14	34.05	33.66	32.27	31.42	33.26
EEGTCNet	58.05	59.16	62.21	61.22	59.75	61.11	61.07	62.47
ShallowNet	54.78	56.80	59.58	59.01	59.71	58.32	57.64	58.47

3.7 Cross-Dataset Generalization: OpenBMI Validation

To evaluate the generalizability of the proposed channel-reduction approach beyond the BCIC-IV-2a dataset, we further validated the selected 14-channel configuration on the OpenBMI dataset [13]. OpenBMI contains recordings from 54 subjects performing two-class left- and right-hand motor imagery, using a 62-channel EEG montage based on the international 10-10 system (a 20-channel motor-cortex subset is also commonly used in downstream analyses). The same 14 channels selected from BCIC-IV-2a were evaluated on OpenBMI, and compared against the original 20-channel configuration across architectures (EEGNet, ShallowConvNet, EEGTCNet, and DeepConvNet) using subject-dependent 5-fold cross-validationw, where the cross-validation folds were used for validation and model-checkpoint selection. The selected checkpoint was then evaluated on the held-out session 2, which was never used during training or checkpoint selection. Rather than merely matching baseline performance, the transferred 14-channel configuration significantly outperformed the 20-channel baseline for all compared models, with accuracy gains of 1.91–2.61 percentage points, each confirmed as statistically significant by paired t-tests ($p < 0.05$ for all models, $p < 0.01$ for three of four) and corroborated by Wilcoxon signed-rank tests (Table 10). Figure 5 summarizes these gains:

Table 9: Comparison of EEG channel selection methods on the BCI Competition IV-2a dataset (4-class motor imagery, 9 subjects, 22 EEG channels sampled at 250 Hz). Results are reported under the experimental conditions used in the respective state-of-the-art studies, including their original evaluation protocols, classifiers, and data splits.

Method	Selection Basis	Neurophysio. Matched	Total Channels	Reduced Channels	ΔAcc (pp)	Classifier/ Model
ECA-CNN [14]	Attention weight	Partial	22	8	−6.24	EEGNet
Signal Power CS [15]	Signal power	Partial	22	11	−8.59	EEGNet
SNR-based CS + PRD [16]	Signal-to-noise ratio	Partial	22	11	−2.73	EEGNet
SE-Attention (Proposed)	**SE-Attention weight (learned)**	**Full**	**22**	**14**	**+1.80^{ns}**	**EEGNet**

$\Delta\text{Acc} = \text{Acc}_{\text{reduced}} - \text{Acc}_{\text{total}}$ (negative = accuracy loss); ^{ns} = not statistically significant, $p = 0.155$ (paired t-test, $n = 9$); Neurophysio. match — *Yes*: ERD/ERS or motor-cortex grounded, *Partial*: motor-region overlap only, *No*: purely data-driven; results reproduced directly from original publications without re-implementation.

Table 10: Cross-dataset transfer of the SE-Attention 14-channel subset (selected on BCI-IV 2a) to OpenBMI. For each of the 54 subjects, the decoder was trained with 5-fold cross-validation on session 1 trials and evaluated, using the checkpoint from the final training fold, on session 2 trials. Results are mean ± SD accuracy across subjects, 2-class MI (LH/RH). Significance is from a paired t-test against the All_20ch baseline (Wilcoxon signed-rank test agrees for all four models).

Model	Channel Set	Accuracy (%)	F1 (%)	Kappa	Δ Acc (pp)	Sig.
EEGNet	All_20ch	58.24 ± 9.77	56.06	0.1648	–	–
	SE_Top14	60.80 ± 12.84	58.97	0.2161	+2.56	$p = 0.0046^{**}$
ShallowConvNet	All_20ch	50.28 ± 6.54	49.10	0.0056	–	–
	SE_Top14	52.19 ± 8.10	50.84	0.0439	+1.91	$p = 0.0249^{*}$
EEGTCNet	All_20ch	57.76 ± 9.59	55.71	0.1553	–	–
	SE_Top14	60.37 ± 12.76	58.56	0.2075	+2.61	$p = 0.0018^{**}$
DeepConvNet	All_20ch	57.44 ± 8.57	55.45	0.1488	–	–
	SE_Top14	60.05 ± 12.58	58.50	0.2010	+2.61	$p = 0.0046^{**}$

* $p < 0.05$, ** $p < 0.01$ (paired t-test vs. baseline).

Table 11: Impact of Channel Reduction on Computational Efficiency Across Models

Model	Ch	Trainable Params	T_train (s)	T_pred) (s/sample)	FLOPs (M)
EEGNet	22	3,700	3.9764	0.0018	26.45
	14	3572	3.6586	0.0015	16.95
DeepNet	22	553,654	6.0391	0.0079	226.79
	14	548,654	5.4685	0.0019	210.94
ShallowNet	22	47,364	4.4812	0.0096	127.20
	14	34,567	3.1789	0.0015	81.05
EEGTCNet	22	4096	6.7236	0.0049	13.97
	14	3968	4.5663	0.0033	9.07

the dumbbell plot shows the consistent, model-wise shift in accuracy from baseline to the transferred channel set, while the heatmap highlights the relative improvement magnitude across accuracy, F1, and kappa, with ShallowConvNet showing the smallest, though still significant, effect. These results demonstrate that the 14-channel configuration selected on BCIC-IV-2a generalizes effectively to an independent dataset and montage, substantially reducing the number of required EEG electrodes while simultaneously improving, rather than merely preserving, decoding performance across models.

3.8 Training-Time Benchmark

Table 11 shows the computational benefits of reducing the EEG montage from 22 to 14 channels across four EEG decoding models. The channel reduction decreased the number of trainable parameters, training time, inference latency, and FLOPs for all models. In particular, inference time was reduced by up to 84% for ShallowConvNet and 76% for DeepConvNet, while FLOPs decreased by up to 36%. These results show that reducing the number

of EEG channels can lower computational requirements and support the development of compact, efficient, and real-time BCI systems.

4 Conclusion

This work proposes an SE-Attention-based channel-selection method for MI EEG decoding, in which the importance of each channel is learned end-to-end through a Squeeze-and-Excitation (SE) gating mechanism. The proposed 14-channel montage, derived by aggregating SE-Attention channel importance across subjects from the 22-channel BCIC-IV-2a montage, was evaluated on the independent OpenBMI dataset and compared with a predefined 20-channel motor-cortex subset of its 62-channel montage. The proposed montage was evaluated across multiple state-of-the-art models, including EEGNet, DeepConvNet, EEGTCNet, and ShallowConvNet, demonstrating that substantial channel reduction can be achieved while maintaining comparable classification performance.

The results show the practical effectiveness of SE-Attention for substantial reduction of the EEG channel count while preserving the classification performance. In BCIC-IV-2a, reducing the full 22-channel montage to 14 selected channels did not result in statistically significant accuracy degradation in the evaluated architectures, confirming that a large proportion of the original montage can be removed without compromising decoding performance. When the same 14-channel configuration, selected on BCIC-IV-2a, was transferred without re-selection to the independent 54-subject OpenBMI dataset, it significantly outperformed the full 20-channel baseline across all four architectures (accuracy gains of 1.91–2.61 percentage points, $p < 0.05$ for all models, $p < 0.01$ for three of four, confirmed by paired Wilcoxon signed-rank tests), with consistent gains in Cohen's kappa as well. This cross-dataset result indicates that the learned channel subset is not merely dataset-specific but captures a generalizable, task-relevant spatial pattern that transfers across montages and subject populations.

Furthermore, the learned SE-Attention ranking showed strong agreement with an independently constructed, neurophysiologically motivated montage, with 10 of the 12 selected channels overlapping and the selected channels concentrated over central and centro-parietal regions (C3, C4, C5, Cz, CP3, CP4, POz) known to be active during MI. This agreement indicates that the learned attention mechanism identifies spatially meaningful and task-relevant EEG channels rather than producing arbitrary channel selections.

Overall, these findings highlight SE-Attention as an effective, data-driven approach for developing compact EEG montages with reduced channel requirements, while matching or exceeding the performance of the full-channel configuration. The resulting reduction in electrode requirements can facilitate more compact, cost-effective, and practical BCI systems, while the demonstrated cross-dataset and cross-montage generalization, including the significant transfer gains observed on OpenBMI, further supports their potential for assistive applications for individuals with paralysis and neurological disabilities.

Funding

No funds, grants, or other support were received for the preparation of this research work.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain–computer interfaces for communication and control. *Clin. Neurophysiol.* 113(6), 767–791 (2002).

[2] Jin, J., Miao, Y., Daly, I., Zuo, C., Hu, D., Cichocki, A.: Correlation-based channel selection and regularized feature optimization for MI-based BCI. *Neural Netw.* 118, 262–270 (2019).

[3] Das, A., Suresh, S.: An effect-size based channel selection algorithm for mental task classification in BCI. In: *2015 IEEE SMC*, pp. 3140–3145. IEEE (2015).

[4] Luo, Y., Mu, W., Wang, L., Wang, J., Wang, P., Gan, Z., Zhang, L., Kang, X.: An EEG channel selection method for motor imagery based on Fisher score and local optimization. *J. Neural Eng.* 21(3), 036030 (2024).

[5] Varsehi, H., Firoozabadi, S.M.P.: An EEG channel selection method for motor imagery based BCI and neurofeedback using Granger causality. *Neural Netw.* 133, 193–206 (2021).

[6] Sun, H., Jin, J., Xu, R., Cichocki, A.: Feature selection combining filter and wrapper methods for motor-imagery based BCIs. *Int. J. Neural Syst.* 31(09), 2150040 (2021).

[7] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proc. IEEE CVPR*, pp. 7132–7141 (2018).

[8] Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGNet: a compact CNN for EEG-based BCIs. *J. Neural Eng.* 15(5), 056013 (2018).

[9] Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T.: Deep learning with CNNs for EEG decoding and visualization. *Hum. Brain Mapp.* 38(11), 5391–5420 (2017).

[10] Ingolfsson, T.M., Hersche, M., Wang, X., Kobayashi, N., Cavigelli, L., Benini, L.: EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. In: *2020 IEEE SMC*, pp. 2958–2965 (2020).

[11] Brunner, C., Leeb, R., Müller-Putz, G., Schlögl, A., Pfurtscheller, G.: BCI Competition 2008 – Graz data set A. Institute for Knowledge Discovery, Graz University of Technology, 16, 1–6 (2008).

[12] Zhang, Y., Nam, C.S., Zhou, G., Jin, J., Wang, X., Cichocki, A.: Temporally constrained sparse group spatial patterns for motor imagery BCI. *J. Neural Eng.* 18(1), 016004 (2021).

[13] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, and S.-W. Lee, "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, p. giz002, 2019.

[14] Tong, L., Qian, Y., Peng, L., Wang, C., Hou, Z.-G.: A learnable EEG channel selection method for MI-BCI using efficient channel attention. *Front. Neurosci.* 17, 1276067 (2023).

[15] K. Singh, N. Singha, A. K. Sharma, S. Bhalaik, and C. Kumar, "Optimized EEG sensor electrode configuration for motor imagery decoding with minimal accuracy loss and reduced cost," *IEEE Sensors Letters*, vol. 9, no. 12, pp. 1–4, 2025, doi: 10.1109/LSENS.2025.3622924.

[16] K. Singh, N. Singha, A. K. Sharma, S. Bhalaik, and C. Kumar, "A novel approach to optimized electrode selection and data augmentation for MI-EEG data," *IEEE Signal Processing Letters*, vol. 32, pp. 2174–2178, 2025, doi: 10.1109/LSP.2025.3569212.