What Causes Performance Degradation in Cross-Subject EEG Classification?

Yihe Wang*, Taida Li*, Yujun Yan, Wenzhan Song, **Xiang Zhang**

University of North Carolina at Charlotte, USA

xiang.zhang@charlotte.edu

Code: `https://github.com/DL4mHealth/EEG-Cross-Subject`

Motivation

- Cross-subject EEG classification is essential for deployment on **unseen subjects**.
- However, subject-independent evaluation often performs much worse than subject-dependent evaluation.
- This performance degradation is widely observed, but its underlying causes remain unclear.

Core Question

Why does EEG classification performance drop when moving from subject-dependent to subject-independent evaluation?

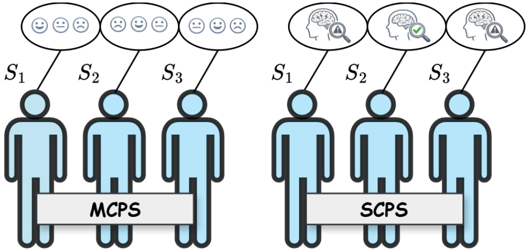
Two Types of EEG Classification Tasks

MCPS: Multi-Class-Per-Subject

- Labels vary within each subject.
- Examples: motor imagery, emotion recognition, ERP stimulus classification.
- Subject identity cannot determine the task label.

SCPS: Single-Class-Per-Subject

- Each subject has one fixed label.
- Examples: AD, PD, depression detection.
- Subject identity can become a label shortcut.



MCPS vs. SCPS EEG tasks.

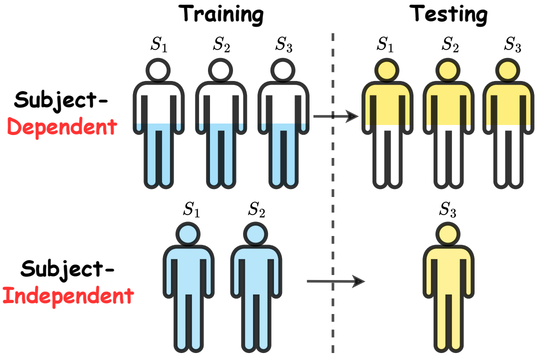
Evaluation Setups

Subject-Dependent

- Split samples randomly.
- Samples from the same subject may appear in both train and test sets.
- $S_{train} \cap S_{test} \neq \emptyset$

Subject-Independent

- Split by subjects.
- Test subjects are unseen during training.
- $S_{train} \cap S_{test} = \emptyset$



Subject-dependent vs. subject-independent evaluation.

Hypothesis

Inter-Subject Variability

- EEG distributions differ across subjects.
- Causes domain shift between training and test subjects.
- Affects both MCPS and SCPS tasks.

Shortcut Learning

- Models may learn subject-specific features instead of task-relevant features.
- Especially severe when subject identity is correlated with the class label.
- Mainly affects SCPS tasks.

Main Finding

MCPS degradation is mainly caused by inter-subject variability; SCPS degradation is largely caused by shortcut learning based on subject identity.

Experimental Design

- Model: EEGConformer for all experiments to avoid architecture-induced variance.
- Split ratio: 60% / 20% / 20% for train / validation / test.
- Repeated with five random seeds: 41–45.
- Metrics: F1 score and AUROC.

Dataset	**Task Type**	**Label**	**Subjects**	**Samples**	**Channels**	**Sampling Rate**
mTBI-ODD	MCPS (3-class)	Standard / Target / Novel	96	24,885	61	200Hz
MMIDB	MCPS (2-class)	Both Fists / Both Feet	109	4,897	64	160Hz
ADFTD	SCPS (3-class)	HC / AD / FTD	88	69,752	19	200Hz
PD-RS	SCPS (2-class)	HC / PD	149	23,839	60	200Hz
AOPD	MCPS & SCPS	Standard / Deviant; HC / PD	50	9,830	59	200Hz

Experiment 1: Performance Drop

Dataset	Setup	F1 Score	AUROC
mTBI-ODD (MCPS)	Sub-Dep	67.63±1.08	87.52±0.63
	Sub-Indep	65.77±2.11↓1.86	85.90±1.23↓1.62
MMIDB (MCPS)	Sub-Dep	65.98±1.87	72.53±1.83
	Sub-Indep	65.38±2.21↓0.60	71.26±2.59↓1.27
ADFTD (SCPS)	Sub-Dep	97.66±0.46	99.85±0.05
	Sub-Indep	47.89±3.93↓49.77	67.46±4.07↓32.39
PD-RS (SCPS)	Sub-Dep	98.55±0.27	99.89±0.05
	Sub-Indep	60.15±5.91↓38.40	66.51±8.01↓33.38

Observation

MCPS tasks show small F1 drops, while SCPS tasks show severe F1 drops.

Experiment 2: Controlled AOPD Comparison

Dataset	Setup	F1 Score	AUROC
AOPD (MCPS)	Sub-Dep	65.25±1.24	72.02±1.64
	Sub-Indep	64.35±1.66↓0.90	69.98±1.88↓2.04
AOPD (SCPS)	Sub-Dep	98.35±0.59	99.86±0.05
	Sub-Indep	67.69±9.23↓30.66	74.15±11.95↓25.71

Why AOPD?

- Same dataset.
- Same subjects and EEG recordings.
- Same train/test split.
- Contains both MCPS and SCPS labels.

Conclusion

The larger SCPS drop is caused by task formulation, not dataset-specific factors.

Experiment 3: Subject Discrimination

Dataset	Setup	F1 Score	AUROC
AOPD (MCPS)	Sub-Dep Sub-Indep	65.25±1.24 64.35±1.66↓0.90	72.02±1.64 69.98±1.88↓2.04
AOPD (SCPS)	Sub-Dep Sub-Indep	98.35±0.59 67.69±9.23↓30.66	99.86±0.05 74.15±11.95↓25.71
AOPD	Sub-Disc	**98.59**±0.88	**99.98**±0.01

Key Evidence

EEG signals contain strong subject-specific information.

- Subject discrimination is a 50-class classification problem.
- The model still achieves 98.59% F1 score.
- These features can be exploited as shortcuts in SCPS tasks.

Experiment 4: Label Permutation Strategy

- Goal: eliminate genuine task-related information while preserving label distribution.
- MCPS: permute labels at the **sample level**.
- SCPS: permute labels at the **subject level** so each subject still has a fixed arbitrary label.

MCPS Label Permutation		SCPS Label Permutation	
Original	Permuted	Original	Permuted
(0, S1)	(1, S1)	(0, S1)	(1, S1)
(1, S1)	(0, S1)	(0, S1)	(1, S1)
(0, S1)	(0, S1)	(0, S1)	(1, S1)
(1, S2)	(1, S2)	(1, S2)	(0, S2)
(0, S2)	(1, S2)	(1, S2)	(0, S2)
(1, S2)	(0, S2)	(1, S2)	(0, S2)

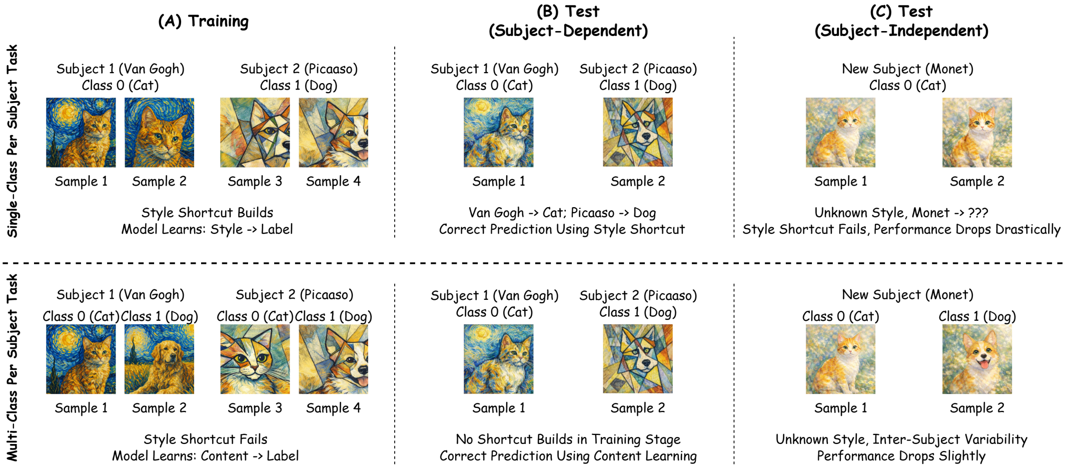
Experiment 4: Label Permutation Results

Dataset	Setup	F1 Score	AUROC
AOPD (MCPS)	Sub-Dep	65.25±1.24	72.02±1.64
	Sub-Dep-Perm	49.77±1.85	50.47±2.27
	Sub-Indep	64.35±1.66	69.98±1.88
	Sub-Indep-Perm	48.92±0.34	49.73±0.65
AOPD (SCPS)	Sub-Dep	98.35±0.59	99.86±0.05
	Sub-Dep-Perm	**98.15±0.29**	**99.79±0.10**
	Sub-Indep	67.69±9.23	74.15±11.95
	Sub-Indep-Perm	49.60±8.84	50.04±13.82

Key Result

After SCPS labels are permuted, subject-dependent performance still remains 98.15% F1.

Shortcut Learning: Image Analogy



Global Analysis

MCPS Tasks

- Labels vary within each subject.
- Subject identity is not a reliable predictor.
- Main challenge: inter-subject variability.

SCPS Tasks

- Each subject has a fixed label.
- Subject identity can predict labels under subject-dependent splits.
- Main challenge: shortcut learning plus inter-subject variability.

Implication

Subject-dependent evaluation can substantially overestimate performance on SCPS EEG tasks.

Takeaways and Future Directions

- Cross-subject performance degradation has different causes across EEG task formulations.
- MCPS tasks are mainly affected by domain shift from inter-subject variability.
- SCPS tasks are highly vulnerable to shortcut learning based on subject-specific features.
- Future work should disentangle subject identity from task-relevant neural features.
- Larger and more diverse pretraining datasets may help learn more generalizable EEG representations.

Thank You

Questions?