AutoTransfer: Automated Subject Transfer Learning with Censored Representations on Biosignals Data

Niklas Smedemark-Margulies[1], Ye Wang[2], Toshiaki Koike-Akino[2], Deniz Erdogmus[3]

[1]Khoury College of Computer Science, Northeastern University, Boston, MA 02115, USA. [2]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA. [3]College of Engineering, Northeastern University, Boston, MA 02115, USA.

Abstract

We frame the problem of **subject transfer learning** as a constrained optimization problem in which we seek to learn an encoder model that minimizes classification loss, subject to a constraint on independence between the latent representation and the subject label.

- We propose a new framework called "**AutoTransfer**" for automatically performing transfer learning on new datasets.
- AutoTransfer achieved **1st place** in subject-transfer task at BEETL AI challenge [1].
- We introduce three notions of independence which we call "**censoring modes**" to derive subject-invariant objective functions: (1) Marginal independence: $z \perp s$; (2) Class-conditional independence: $z \perp s \mid y$; and (3) Complementary independence: $z_1 \perp s$ and $\max I(z_2; s)$.
- For each censoring mode, we enforce these independence constraints using two penalties: mutual information or divergence (See Tab. 1).
- We provide a total of 15 censoring algorithms in the form of neural critic functions as well as analytic function approximations (See Tab. 2).
- We perform extensive experimentation, hyperparameter tuning, and model ensembling, showing superior performance in subject transfer learning on a variety of EEG, EMG, and ECoG datasets.

Censoring Objectives

- **Subject-Invariant Inference**: Consider a classification problem with data x, task labels y, and subject labels s. We train an encoder model $z = f_\theta(x)$ and a classifier model $\hat{y} = g_\phi(z)$ by adding a regularization term alongside the standard cross entropy loss:

$$(\theta^*, \phi^*) = \arg\min_{\theta,\phi} \mathcal{L}_{\text{task}} + \lambda \mathcal{L}_{\text{censor}} \quad (1)$$

- **Censoring Modes**: Here $\mathcal{L}_{\text{task}}$ is the main task loss and $\mathcal{L}_{\text{censor}}$ is a regularization term in the form of a **mutual information** penalty or a **divergence** penalty. The regularization term enforces **marginal** independence ($z \perp s$), **conditional** independence ($z \perp s \mid y$), or **complementary** independence ($z_1 \perp s$ and $\max I(z_2; s)$).

Table 1: High-level censoring penalties considered

Censoring Mode	Mutual Information	Divergence
Marginal	$I(z; s)$	$\mathcal{D}(q_\theta(z) \| q_\theta(z\|s))$
Conditional	$I(z; s\|y)$	$\mathcal{D}(q_\theta(z\|y) \| q_\theta(z\|s, y))$
Complementary	$I(z_1; s) - I(z_2; s)$	$\mathcal{D}(q_\theta(z_1) \| q_\theta(z_1\|s)) - \mathcal{D}(q_\theta(z_2) \| q_\theta(z_2\|s))$

- **Censoring Methods**: We consider various estimation methods for each censoring penalty:

Table 2: Censoring penalties and estimation methods

Penalty	Estimation Methods
Mutual Information	MIGE [2], Adversary [3]
Divergence	MMD/Pairwise MMD [4], BEGAN Disc [5]

- **Problem**: This framework results in a large set of combinatorial possibilities to apply in regularization terms. Because of **no free-lunch theorem**, there is no single algorithm performing best across all datasets.
- **Solution**: Our proposed AutoTransfer methods explores these censoring algorithms without manual trial-and-error, and selects the best settings according to performance on an unseen validation subject.

References

[1] *NeurIPS 2021 BEETL Competition: Benchmarks for EEG Transfer Learning*. https://beetl.ai.

[2] Liangjian Wen et al. "Mutual information gradient estimation for representation learning". In: *arXiv preprint arXiv:2005.01123* (2020).

[3] Ozan Özdenizci et al. "Learning invariant representations from EEG via adversarial inference". In: *IEEE access* 8 (2020), pp. 27074–27085. DOI: 10.1109/ACCESS.2020.2971600.

[4] Arthur Gretton et al. "A kernel two-sample test". In: *The Journal of Machine Learning Research* 13.1 (2012), pp. 723–773.

[5] David Berthelot, Thomas Schumm, and Luke Metz. "BEGAN: Boundary equilibrium generative adversarial networks". In: *arXiv preprint arXiv:1703.10717* (2017).

[6] Gregory Lee et al. "PyWavelets: A Python package for wavelet analysis". In: *Journal of Open Source Software* 4.36 (2019), p. 1237.

[7] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[8] *RSVP EEG Dataset*. https://repository.library.northeastern.edu/collections/neu:gm80jm78x.

[9] Perrin Margaux et al. "Objective and subjective evaluation of online error correction during P300-based spelling". In: *Advances in Human-Computer Interaction* 2012 (2012).

[10] *American Sign Language EMG Dataset*. Non-public data, taken with permission from Northeastern University Movement Neuroscience Lab.

[11] Kai J Miller. "A library of human electrocorticographic data and analyses". In: *Nature human behaviour* 3.11 (2019), pp. 1225–1235.

AutoTransfer Pipeline

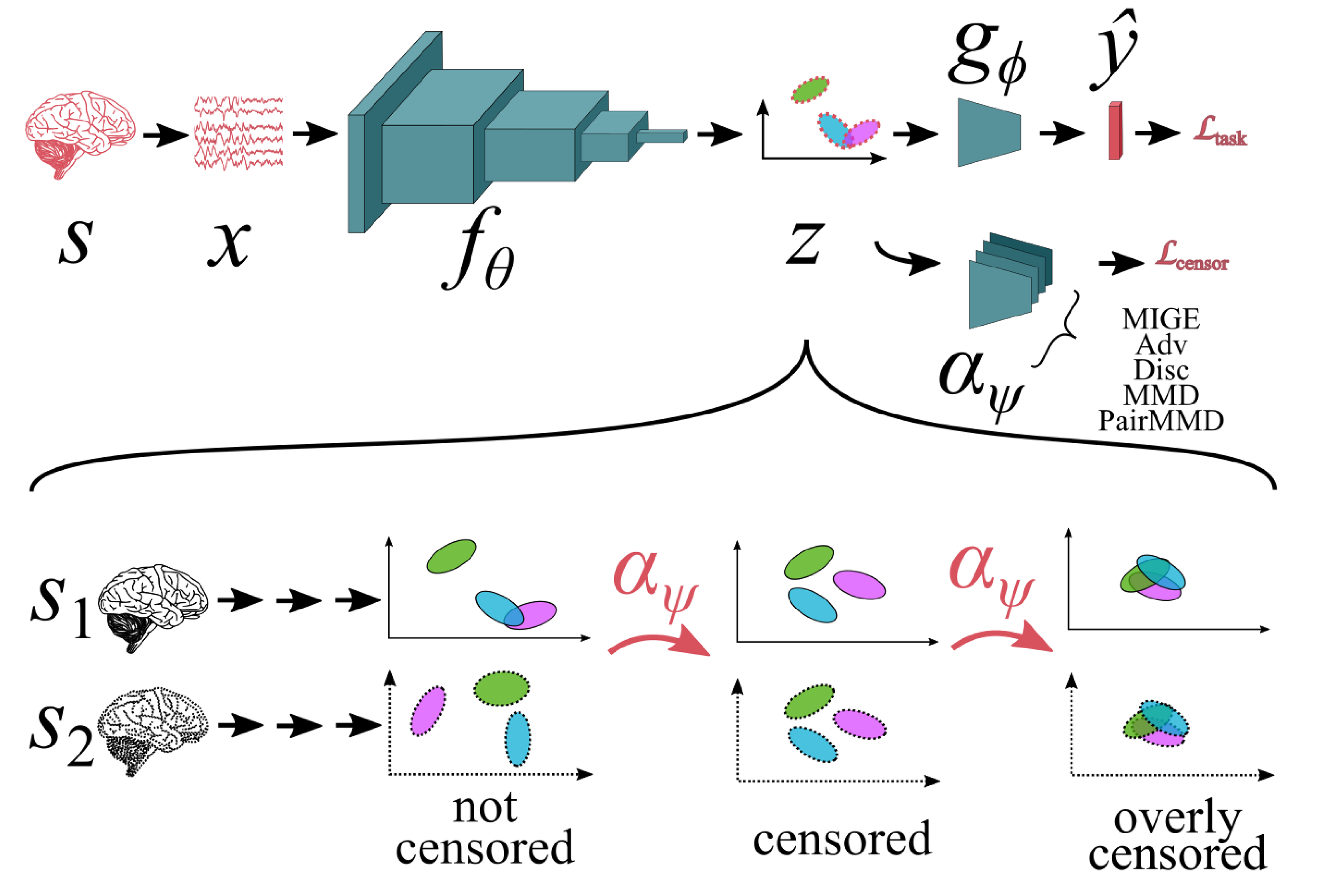

Figure 1: AutoTransfer pipeline for subject-invariant feature censoring in pre-shot transfer learning.

- Subject s produces data x, which is mapped by encoder f_θ to latent code z, and fed into g_ϕ to produce estimated class probabilities $\hat{y}$, giving task loss $\mathcal{L}_{\text{task}}$.
- Various censoring models α_ψ compute regularization penalty $\mathcal{L}_{\text{censor}}$ to enforce independence.
- Multiple subjects s_1, s_2 are encoded, and the penalty from α_ψ gradually changes the latent feature distribution during training.
- Different censoring algorithms having different strength of regularization are automatically explored to provide sufficient regularization without causing collapse.

Example Censoring Algorithms

Algorithm 1: Marginal MIGE Censoring

Input: Batch $\{(x_i, y_i, s_i)\}_{i=1}^N$, Encoder f_θ, No. nuisance values M, Score estimator F_{score}

Output: Gradient of MI

Subroutine $Est_{\nabla H}$ (*vectors* $\{z_i\}_{i=1}^N$)**:**

- $\nabla_\theta H \leftarrow 0$; fit F_{score} to $\{z_i\}$
- **for** i *in* $1 \dots N$ **do**
  - $r \leftarrow F_{score}(z_i)$ `// Eval Score`
  - add $r \cdot \nabla_\theta z_i$ to $\nabla_\theta H$
- **return** $\nabla_\theta H$

for i *in* $1 \dots N$ **do**
- $z_i \leftarrow f_\theta(x_i)$

$\nabla_\theta H(z) \leftarrow Est_{\nabla H}(\{z_i\})$

for m *in* $1 \dots M$ **do**
- $\mathcal{S}_m \leftarrow \{z_i : s_i = m\}$
- add $\frac{1}{M \cdot |\mathcal{S}_m|} Est_{\nabla H}(\mathcal{S}_m)$ to $\nabla_\theta H(z|s)$

return $\nabla_\theta H(z) - \nabla_\theta H(z|s)$

Algorithm 2: Conditional Censoring using BEGAN Discriminator

Input: Batch $\{(x_i, y_i, s_i)\}_{i=1}^N$, Encoder f_θ, No. nuisance values M, No. classes C, Prev. control trade-off $k_{prev} \in [0, 1]$, Control LR β

Output: Encoder's divergence penalty, Discriminator's objective, Next control trade-off value

for i *in* $1 \dots N$ **do**
- $z_i \leftarrow f_\theta(x_i)$

$\mathcal{L}_{p(z|y)} \leftarrow 0$; $\mathcal{L}_{p(z|s,y)} \leftarrow 0$

for c *in* $1 \dots C$ **do**
- add $\mathcal{L}^D(z_i : y_i = c)$ to $\mathcal{L}_{p(z|y)}$
- **for** r *in* $1 \dots M$ **do**
  - add $\mathcal{L}^D(z_i : s_i = r, y_i = c)/M$ to $\mathcal{L}_{p(z|s,y)}$

$\mathcal{L}_{\text{Disc}} \leftarrow \mathcal{L}_{p(z|y)} - k_{prev} \cdot \mathcal{L}_{p(z|s,y)}$

$\mathcal{L}_{\text{Enc}} \leftarrow \mathcal{L}_{p(z|s,y)}$

$k_{next} \leftarrow k_{prev} + \beta \cdot (0.5 \cdot \mathcal{L}_{p(z|y)} - \mathcal{L}_{p(z|s,y)})$

return $\mathcal{L}_{\text{Enc}}$, $\mathcal{L}_{\text{Disc}}$, clip($k_{next}$, 0, 1)

Algorithm 3: Complementary Adversarial Censoring

Input: Batch $\{(x_i, y_i, s_i)\}_{i=1}^N$, Encoder f_θ, Adversarial Classifier α_ψ

Output: Mutual Information penalty

$\mathcal{L}_{\text{total}} \leftarrow 0$

for i *in* $1 \dots N$ **do**
- `// Split latent representation`
- $(z_i^1, z_i^2) \leftarrow f_\theta(x_i)$
- `// Predict subj from each half`
- $q_\psi(s_i|z_i^1, y_i) \leftarrow \alpha_\psi(z_i^1, y_i)$
- $q_\psi(s_i|z_i^2, y_i) \leftarrow \alpha_\psi(z_i^2, y_i)$
- add $\mathcal{L}_{\text{CE}}(q_\psi(s_i|z_i^1), s_i)$ to $\mathcal{L}_{\text{total}}$
- subtract $\mathcal{L}_{\text{CE}}(q_\psi(s_i|z_i^2), s_i)$ from $\mathcal{L}_{\text{total}}$

return $\mathcal{L}_{\text{total}}$

Experimental Setting

Our censoring objectives can be combined with other standard deep learning techniques.

- We experiment with various Continuous Wavelet Transforms (CWT) [6] for preprocessing.
- We use ResNet18 [7] encoder model pretrained on image datasets.
- For new datasets, AutoTransfer tunes hyperparams over balanced accuracy on held-out subject.
- For the top 3 settings of each censoring method, we run leave-subject-out cross-validation (CV).
- In each fold, we reserve 1 validation and 1 test subject.

Results

- We evaluate our approach on diverse neurophysiological datasets: EEG Rapid Serial Visual Presentation (RSVP) event-related potentials [8], Error Potentials (ErrP) [9]; EMG American Sign Language (ASL) [10]; and ECoG facial recognition task [11].
- We verified that censoring can improve subject transfer performance across varied datasets.
- The ideal censoring mode and method is dataset dependent.
- Improvements are especially pronounced for subjects whose naive transfer performance is lower.
- AutoTransfer ranked **1st place** in cross-subject transfer task 1 of NeurIPS BEETL AI challenge.

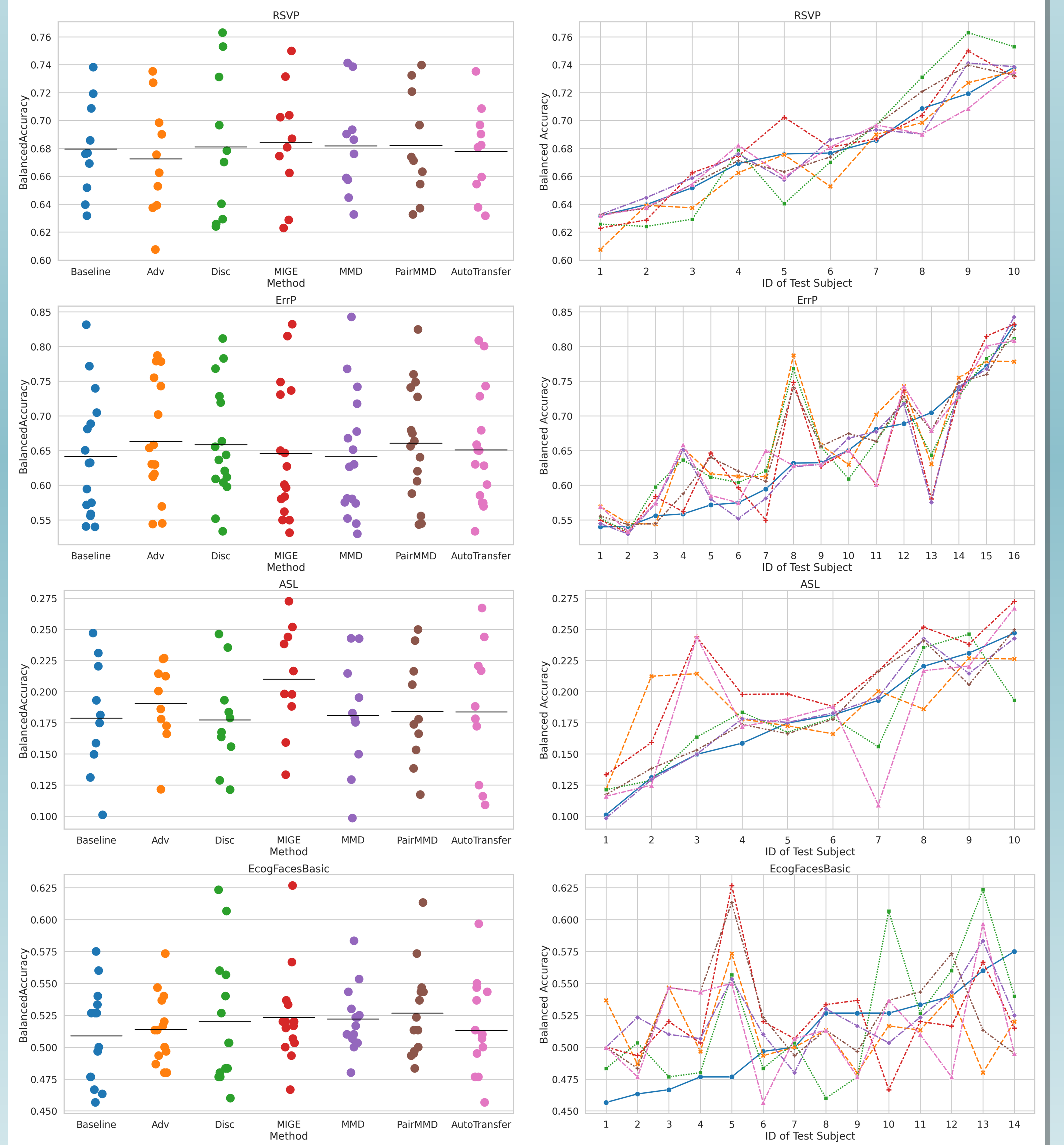

Figure 2: Subject transfer balanced accuracy. Left: Test score from each CV fold, black line indicates mean. Right: Accuracy vs test subject, sorted by baseline performance. Color coding matches for left and right.

Table 3: BEETL Task 1 Results: Sleep Stage Classification

Competition Stage	Censoring Method	Score (gap to competitor)
Leaderboard Testing	Baseline	68.22 (-3.92)
Leaderboard Testing	Marginal Adv	67.65 (-4.49)
Leaderboard Testing	Marginal PairMMD	65.68 (-6.46)
Leaderboard Testing	Marginal MIGE	66.81 (-5.33)
Final Testing	Baseline	68.69 $(+0.03)$
Final Testing	Conditional MIGE	67.23 (-1.43)
Final Testing	Complementary BEGAN Disc	68.41 (-0.25)
Final Testing	Conditional MMD	$\mathbf{69.23}$ $(+0.57)$

smedemark-margulie.n@northeastern.edu, yewang@merl.com, koike@merl.com