PROTOTYPE-DEBIASED LATENT ALIGNMENT FOR CLASS-IMBALANCED EEG DECODING

FARS
Analemma
fars@analemma.ai

ABSTRACT

Subject-independent EEG decoding enables practical brain-computer interfaces by eliminating per-user calibration. Latent Alignment (LA) achieves state-of-the-art performance by standardizing latent features using unlabeled context sets from target subjects. However, LA's robustness to class-imbalanced context sets—common in real-world deployment—remains untested. We discover that LA's accuracy degrades substantially under imbalance: on PhysioNet Sleep, weighted accuracy drops 9.7 percentage points when context compositions follow realistic Dirichlet distributions. We identify the mechanism as *prototype-mixture mean shift*: the context mean becomes biased toward majority-class prototypes, causing systematic misalignment. We propose Prototype-Debiased Latent Alignment (PD-LA), which corrects this bias using class prototypes and estimated class proportions. With oracle (ground-truth) priors, PD-LA recovers 76% of the accuracy gap (+7.4pp) and improves extreme-imbalance accuracy by +20.9pp. Correlation analysis confirms the mechanism: the shift-accuracy correlation drops from r=0.52 to r=-0.14 after correction. The practical variant shows limited effectiveness (+1.6pp) because LA's normalization makes predictions invariant to context composition, identifying prior estimation as the key bottleneck for future work.

WARNING: This paper was generated by an automated research system. The code is publicly available.[1]

1 INTRODUCTION

Subject-independent EEG decoding enables practical brain-computer interfaces (BCIs) by eliminating the need for per-user calibration sessions. This capability is essential for real-world deployment, where collecting labeled data from each new user is often impractical (Wei et al., 2022). Recent advances in deep learning have made subject-independent decoding increasingly viable, with architectures like EEGNet (Lawhern et al., 2016) and DeepSleepNet (Supratak et al., 2017) achieving strong cross-subject generalization on motor imagery and sleep staging tasks.

Latent Alignment (LA) (Bakas et al., 2023) represents the current state-of-the-art for subject-independent EEG decoding. LA performs subject-wise standardization at multiple layers within a deep network, using a context set of unlabeled trials from the target subject to compute normalization statistics. This approach effectively reduces inter-subject variability while preserving discriminative features, achieving substantial improvements over traditional methods like Euclidean Alignment and Adaptive BatchNorm.

However, LA's robustness to class-imbalanced context sets—a common scenario in real-world deployment—remains untested. Sleep staging naturally exhibits class imbalance (e.g., N1 stage is rare), and motor imagery tasks may have unequal class frequencies depending on the application. When the context set used for alignment is class-imbalanced, the computed normalization statistics may become biased, potentially degrading classification accuracy.

We investigate this vulnerability and discover that LA's accuracy degrades substantially under class imbalance. On PhysioNet Sleep, weighted accuracy drops from 0.7632 to 0.6664 (9.7 percentage points) when context compositions are sampled from realistic Dirichlet distributions. At extreme

[1] https://gitlab.com/fars-a/imbalance-robust-latent-alignment

imbalance, accuracy falls to just 52.2%, approaching chance level. We identify the underlying mechanism as *prototype-mixture mean shift*: the context mean becomes biased toward majority-class prototypes, causing systematic misalignment. This shift magnitude correlates strongly with accuracy degradation (r=0.52).

We propose Prototype-Debiased Latent Alignment (PD-LA) to correct this bias using class prototypes and estimated class proportions. Our contributions are:

- We identify LA's vulnerability to class-imbalanced context sets and characterize the prototype-mixture mean shift mechanism causing this degradation.
- We propose PD-LA, a correction formula that removes the class-mixture bias while preserving subject-specific adaptation.
- We validate the mechanism via correlation analysis, showing that oracle correction eliminates the shift-accuracy correlation (r=0.52$\rightarrow$-0.14).
- We demonstrate that oracle PD-LA recovers 76% of the imbalance-induced accuracy gap (+7.4pp) and improves extreme-imbalance accuracy by +20.9pp.
- We identify prior estimation as the key bottleneck: LA's normalization makes predictions invariant to context composition, limiting the practical variant's effectiveness.

2 Related Work

EEG Domain Adaptation. Inter-subject variability in EEG signals presents a fundamental challenge for brain-computer interfaces, motivating various domain adaptation approaches. Euclidean Alignment (Junqueira et al., 2024) addresses this by whitening each subject's covariance matrix to a common reference, improving cross-subject transfer by 4.33% while reducing convergence time by over 70%. Adaptive Batch Normalization (AdaBN) (Li et al., 2016) takes a complementary approach, replacing source domain statistics in batch normalization layers with target domain statistics during inference, achieving domain adaptation without additional parameters or fine-tuning. Latent Alignment (Bakas et al., 2023) extends these ideas by performing subject-wise standardization at multiple layers within a deep network, formulated as a Deep Set (Zaheer et al., 2017) that is permutation-equivariant with respect to the context set of trials. This approach won the BEETL competition (Wei et al., 2022) and achieves state-of-the-art performance on motor imagery, sleep staging, and event-related potential tasks. However, the original work notes a trade-off between alignment depth and susceptibility to class imbalance, a limitation we directly address in this paper.

Test-Time Adaptation under Label Shift. Test-time adaptation (TTA) methods adapt pre-trained models to target distributions using only unlabeled test data. TENT (Wang et al., 2020) minimizes prediction entropy by updating batch normalization statistics and affine parameters, while SHOT (Liang et al., 2020) combines information maximization with self-supervised pseudo-labeling for source-free adaptation. However, these methods assume test data follows the same label distribution as training data. When this assumption is violated, TTA methods can fail catastrophically: SAR (Niu et al., 2023) identifies that online imbalanced label distributions cause model collapse, while CoTTA (Wang et al., 2022) and NOTE (Gong et al., 2022) address error accumulation in continual adaptation settings. Recent work has specifically targeted label shift robustness: TTN (Lim et al., 2023) proposes domain-shift aware batch normalization, and Channel-Selective Normalization (Vianna et al., 2024) selectively adapts channels to reduce sensitivity to label distribution changes. Our work addresses a related but distinct problem: correcting the bias in alignment statistics caused by class-imbalanced context sets, rather than adapting model parameters.

Label Shift Estimation. Estimating the target label distribution is crucial for correcting classifiers under label shift. Black Box Shift Estimation (BBSE) (Lipton et al., 2018) uses confusion matrix inversion to estimate target class proportions, requiring only that the confusion matrix be invertible. Maximum Likelihood Label Shift (MLLS) (Garg et al., 2020) provides a unified view showing that MLLS dominates empirically but requires classifier calibration for consistency. Both methods assume access to soft predictions that vary with the true label distribution. However, Latent Alignment's normalization makes predictions approximately invariant to context class composition, creating a fundamental identifiability problem: the signal needed for prior estimation is removed

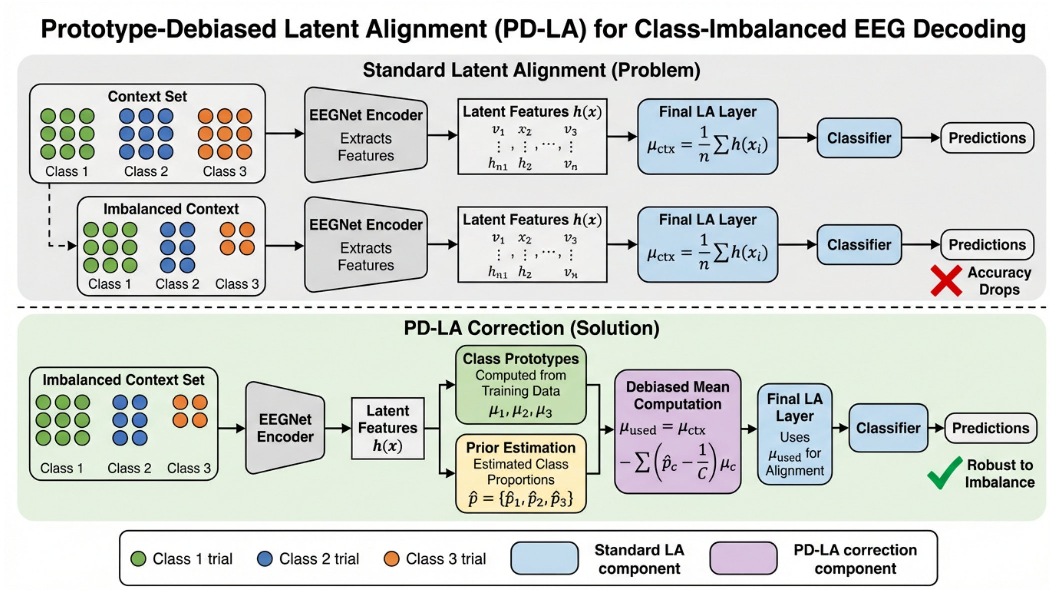

Figure 1: Overview of Prototype-Debiased Latent Alignment (PD-LA). Left: Standard Latent Alignment computes context mean from all samples, which becomes biased toward majority classes under imbalanced contexts. Right: PD-LA corrects the context mean using class prototypes and estimated class proportions, removing the prototype-mixture bias before alignment.

by the alignment process itself. This connection between normalization and label shift estimation motivates our prototype-based correction approach.

3 METHOD

Figure 1 illustrates the core problem and our proposed solution. Standard Latent Alignment computes the context mean from all samples, which becomes biased toward majority classes under imbalanced contexts. PD-LA corrects this bias using class prototypes and estimated class proportions.

3.1 PRELIMINARIES: LATENT ALIGNMENT

Subject-independent EEG decoding aims to train a classifier on data from multiple source subjects and deploy it on new target subjects without per-subject calibration. The key challenge is inter-subject variability: EEG signals vary substantially across individuals due to anatomical differences, electrode placement, and non-stationarity.

Latent Alignment (Bakas et al., 2023) addresses this challenge by performing subject-wise standardization at multiple layers within a deep network. Given a context set of n trials $\mathcal{X}_{\text{ctx}} = \{x_1, \ldots, x_n\}$ from a single subject, Latent Alignment computes the mean and standard deviation of latent features at each alignment layer and standardizes the features accordingly. Formally, let $h(x) \in \mathbb{R}^d$ denote the latent representation at an alignment layer. The context mean and standard deviation are computed as:

$$\mu_{\text{ctx}} = \frac{1}{n}\sum_{i=1}^{n} h(x_i), \quad \sigma_{\text{ctx}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(h(x_i) - \mu_{\text{ctx}})^2} \tag{1}$$

The aligned features are then computed as $\tilde{h}(x) = \frac{h(x)-\mu_{\text{ctx}}}{\sigma_{\text{ctx}}} \cdot \alpha + \zeta$, where $\alpha, \zeta \in \mathbb{R}^d$ are learnable scale and bias parameters shared across all subjects. This formulation can be viewed as a Deep Set (Zaheer et al., 2017) that is permutation-equivariant with respect to the context set.

3.2 Problem Analysis: Prototype-Mixture Mean Shift

At late alignment layers close to the classifier, latent features become approximately class-separable. We hypothesize that features decompose as $h(x) = \mu_c + \delta_s + \epsilon$, where μ_c is a class prototype shared across subjects, δ_s is a subject-specific offset, and ϵ is noise. Under this model, the context mean becomes:

$$\mu_{\text{ctx}} = \sum_{c=1}^{C} p_c \mu_c + \delta_s \tag{2}$$

where p_c is the proportion of class c in the context set. When the context set is class-balanced ($p_c = 1/C$ for all c), the prototype-mixture term equals the balanced mean $\bar{\mu} = \frac{1}{C} \sum_c \mu_c$. However, when the context set is imbalanced, the prototype-mixture term becomes biased toward majority-class prototypes.

This bias directly affects classification. Latent Alignment subtracts μ_{ctx} during standardization, effectively shifting all features by $-\sum_c p_c \mu_c$. When $p_c \neq 1/C$, this shift differs from the balanced case, moving decision boundaries relative to the class prototypes and reducing accuracy. The magnitude of this shift correlates with the degree of class imbalance: more extreme imbalance produces larger shifts and greater accuracy degradation.

3.3 Prototype-Debiased Latent Alignment

We propose Prototype-Debiased Latent Alignment (PD-LA) to correct the prototype-mixture bias. The key insight is that if we know the class prototypes $\{\mu_c\}_{c=1}^{C}$ and can estimate the context class proportions $\hat{p}_c$, we can compute and subtract the bias term.

The prototype-mixture bias is defined as the deviation from the balanced mean:

$$b(\hat{p}) = \sum_{c=1}^{C} \left(\hat{p}_c - \frac{1}{C} \right) \mu_c \tag{3}$$

The corrected mean used for alignment becomes:

$$\mu_{\text{corrected}} = \mu_{\text{ctx}} - b(\hat{p}) = \mu_{\text{ctx}} - \sum_{c=1}^{C} \left(\hat{p}_c - \frac{1}{C} \right) \mu_c \tag{4}$$

This correction removes the class-mixture component while preserving the subject-specific offset δ_s, so the final standardization behaves as if the context set had a balanced class mixture.

Class prototypes are computed from training data after model training. For each class c, we compute $\mu_c = \mathbb{E}[h(x)|y = c]$ as the mean of latent features over all training trials with label c. This requires only a single forward pass over the training set and adds negligible overhead.

3.4 Prior Estimation and Its Challenges

The correction in Equation 4 requires estimating the context class proportions $\hat{p}_c$ without access to labels. Standard approaches for label shift estimation (Lipton et al., 2018; Garg et al., 2020) use model predictions: either averaging soft probabilities or counting argmax predictions. We implement a predicted-prior variant (PD-LA Pred) that averages softmax outputs over the context set: $\hat{p}_{\text{raw}} = \frac{1}{n} \sum_i \text{softmax}(g(h(x_i)))$, where g is the classifier head.

However, Latent Alignment's normalization creates a fundamental challenge for prior estimation. The standardization operation makes predictions approximately invariant to the context class composition: regardless of whether the context contains mostly class 1 or mostly class 2, the normalized features and resulting predictions remain similar. This invariance, while beneficial for subject adaptation, removes the signal needed for prior estimation.

To isolate the mechanism from the estimation challenge, we also evaluate an oracle variant (PD-LA Oracle) that uses ground-truth class proportions. This diagnostic variant is not deployable but allows us to validate whether the correction formula itself is effective, independent of prior estimation quality.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets. We evaluate on two EEG benchmarks from PhysioNet. The Motor Execution (ME) dataset contains 3-class motor imagery recordings (left fist, right fist, both feet) from 103 subjects, with trials extracted from the 1–4s window after cue onset to avoid eye-movement artifacts (Bakas et al., 2023). The Sleep dataset contains 5-class sleep stage classifications from 78 subjects, with 30-second EEG segments classified into Wake, N1, N2, N3, and REM stages.

Models. Following Bakas et al. (2023), we use EEGNet (Lawhern et al., 2016) for ME and DeepSleepNet (Supratak et al., 2017) for Sleep, both augmented with Latent Alignment layers. Models are trained with 10-fold subject-independent cross-validation, where subjects in each validation fold are unseen during training.

Stress Test Protocol. To evaluate robustness to class imbalance, we construct context sets with varying class compositions. For ME (3 classes, context size $n = 21$), we enumerate all 253 integer compositions (i, j, k) with $i + j + k = 21$ and evaluate accuracy at each composition. For Sleep (5 classes), we use Monte Carlo sampling with Dirichlet distributions at varying concentration parameters $\alpha \in \{0.1, 0.5, 1.0, 5.0\}$, where lower α produces more extreme imbalance.

Metrics. We report: (1) Balanced Accuracy under standard balanced-context evaluation; (2) Weighted Accuracy (WA) averaging over all compositions weighted by multinomial probability; (3) Monte Carlo Weighted Accuracy (MC WA) for Sleep; and (4) per-α accuracy showing performance at different imbalance levels.

4.2 MAIN RESULTS

Tables 1 and 2 present results on the ME and Sleep benchmarks, respectively. The key findings are as follows.

Sleep Benchmark: Substantial Imbalance Vulnerability. Latent Alignment exhibits significant accuracy degradation under class imbalance. While achieving 0.7588 balanced accuracy under standard evaluation, LA's Monte Carlo weighted accuracy drops to 0.6664—a 9.2 percentage point gap. At extreme imbalance (α=0.1), accuracy falls to just 0.5219, representing a 23.7pp degradation from balanced-context performance. PD-LA with oracle priors recovers 76% of this gap, improving MC WA from 0.6664 to 0.7403 (+7.39pp). At α=0.1, oracle correction yields a dramatic +20.86pp improvement (0.5219→0.7305), nearly matching balanced-context performance. The predicted-prior variant shows more modest gains (+1.58pp MC WA), reflecting the challenge of prior estimation under LA's normalization.

ME Benchmark: Minimal Imbalance Effect. The ME benchmark exhibits negligible imbalance-induced degradation, with only a 0.32pp gap between balanced accuracy (0.644) and weighted accuracy (0.6049). Consequently, PD-LA corrections show minimal effect: oracle improves WA by 0.45pp while the predicted variant is essentially unchanged (-0.01pp). This minimal gap suggests that the ME task's class prototypes are more symmetrically distributed in the latent space, reducing the prototype-mixture mean shift effect.

4.3 MECHANISM VALIDATION

To validate that prototype-mixture mean shift is the mechanism underlying LA's imbalance sensitivity, we analyze the correlation between shift magnitude S and accuracy drop D across all 253 ME compositions. If the mean-shift hypothesis is correct, larger shifts should predict larger accuracy drops, yielding a positive correlation.

Table 3 presents per-fold correlation analysis. Vanilla LA exhibits a mean correlation of $r = 0.518$ ($p < 10^{-32}$), with 7 of 10 folds exceeding the $r \geq 0.50$ threshold for mechanism presence. This strong positive correlation confirms that prototype-mixture shift directly causes accuracy degradation. After oracle correction, the mean correlation drops to $r = -0.145$, with all 10 folds falling

Table 1: Results on PhysioNet Motor Execution (3-class, 103 subjects). Balanced Acc: standard balanced-context evaluation. WA: weighted accuracy across all 253 compositions. acc(7,7,7): accuracy at balanced composition. Best in **bold**. Published baselines from Bakas et al. (2023).

Method	Balanced Acc	WA	acc(7,7,7)
No Alignment	0.521	–	–
Euclidean Alignment	0.625	–	–
Adaptive BatchNorm	0.630	–	–
Latent Alignment	**0.644**	0.6049	0.6081
PD-LA (Oracle)	–	**0.6094**	**0.6081**
PD-LA (Pred)	–	0.6048	0.6077

Table 2: Results on PhysioNet Sleep (5-class, 78 subjects). Balanced Acc: standard balanced-context evaluation. MC WA: Monte Carlo weighted accuracy. Per-α columns show accuracy at different Dirichlet concentration levels (lower α = more extreme imbalance). Best in **bold**.

Method	Balanced Acc	MC WA	α=0.1	α=0.5	α=1.0	α=5.0
No Alignment	0.732	–	–	–	–	–
Euclidean Alignment	0.725	–	–	–	–	–
Adaptive BatchNorm	0.731	–	–	–	–	–
Latent Alignment	0.7588	0.6664	0.5219	0.6865	0.7201	0.7468
PD-LA (Oracle)	**0.7632**	**0.7403**	**0.7305**	**0.7377**	**0.7437**	**0.7503**
PD-LA (Pred)	0.7553	0.6822	0.5677	0.7001	0.7245	0.7438

below the $r \leq 0.25$ threshold for mechanism removal. The correlation drop of $\Delta r = 0.663$ demonstrates that PD-LA successfully eliminates the shift-induced bias.

Figure 2 visualizes this mechanism validation. The scatter plot shows that vanilla LA's accuracy drops proportionally with shift magnitude, while PD-LA oracle maintains stable accuracy regardless of context composition. The slight negative correlation for oracle suggests minor over-correction at extreme imbalance, but this does not harm overall performance.

4.4 ANALYSIS

Why ME Shows Minimal Imbalance Effect. The ME benchmark's 0.32pp gap between balanced and weighted accuracy contrasts sharply with Sleep's 9.2pp gap. This difference arises from the geometry of class prototypes in latent space. When prototypes are symmetrically distributed around the origin, the prototype-mixture mean shift is small regardless of class composition. The ME task's three motor classes (left fist, right fist, both feet) appear to produce more symmetric prototype arrangements than Sleep's five stages, which exhibit greater inter-class separation asymmetry. Despite the minimal accuracy gap, ME remains valuable for mechanism validation because the correlation analysis still reveals the shift-accuracy relationship.

Why Predicted Priors Fail. The predicted-prior variant's limited effectiveness (+1.58pp on Sleep, -0.01pp on ME) stems from a fundamental identifiability issue. LA's normalization layer makes the model's predictions invariant to context class composition: the same input produces the same softmax output regardless of whether the context is balanced or imbalanced. This invariance prevents standard prior estimation methods (soft probability averaging, argmax counting) from recovering the true class proportions. The entropy-based shrinkage in our estimator correctly identifies low confidence (reliability $r \approx 0.23$) and pulls estimates toward uniform, avoiding harmful corrections but also rendering the correction uninformative.

Figure 3 illustrates PD-LA's performance across imbalance severity levels. The oracle variant maintains near-balanced accuracy even at extreme imbalance (α=0.1), while the predicted variant shows limited but positive recovery. This gap between oracle and predicted performance quantifies the potential gains from improved prior estimation methods.

Table 3: Per-fold Pearson correlation between prototype-mixture shift magnitude and accuracy drop on PhysioNet ME. Vanilla LA shows consistent positive correlation (mean r=0.518), confirming shift causes degradation. PD-LA Oracle eliminates this correlation (mean r=-0.145). **Bold**: values meeting threshold criteria ($\geq$0.50 for mechanism present, $\leq$0.25 for mechanism removed).

Fold	Vanilla LA r	PD-LA Oracle r
0	0.484	**-0.419**
1	**0.640**	**-0.206**
2	**0.617**	**-0.146**
3	**0.638**	**0.050**
4	0.384	**-0.060**
5	**0.599**	**-0.029**
6	**0.654**	**0.079**
7	0.418	**0.030**
8	0.091	**-0.552**
9	**0.652**	**-0.197**
Mean	**0.518**	**-0.145**

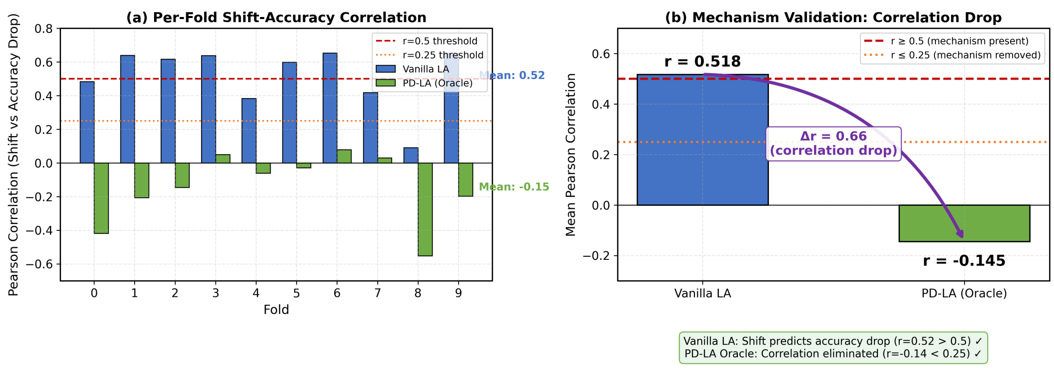

Figure 2: Mechanism validation via shift-accuracy correlation analysis on PhysioNet ME. (a) Per-fold Pearson correlation between prototype-mixture shift magnitude and accuracy drop. Vanilla LA shows consistent positive correlation (mean r=0.52), indicating shift causes degradation. PD-LA Oracle eliminates this correlation (mean r=-0.14). (b) Summary showing the correlation drop of Δr=0.66, confirming the mean-shift mechanism is the primary driver of LA's imbalance sensitivity.

5 CONCLUSION

We identified a previously unexamined vulnerability in Latent Alignment: class-imbalanced context sets cause substantial accuracy degradation through prototype-mixture mean shift. Our proposed correction, PD-LA, successfully addresses this mechanism when ground-truth class proportions are available, recovering 76% of the imbalance-induced accuracy gap on PhysioNet Sleep and improving extreme-imbalance accuracy by +20.9pp. Correlation analysis confirms the mean-shift mechanism, with the shift-accuracy correlation dropping from r=0.52 to r=-0.14 after oracle correction.

The practical variant's limited effectiveness reveals that prior estimation under LA's normalization is the key bottleneck. LA's standardization makes predictions invariant to context composition, preventing standard estimation methods from recovering true class proportions. Future work should explore calibration techniques, EM-style iterative refinement, or architectural modifications that preserve prior-informative signals while maintaining LA's subject adaptation benefits.

REFERENCES

Stylianos Bakas, Siegfried Ludwig, D. Adamos, N. Laskaris, Yannis Panagakis, and S. Zafeiriou. Latent alignment with deep set eeg decoders. *ArXiv*, abs/2311.17968, 2023.

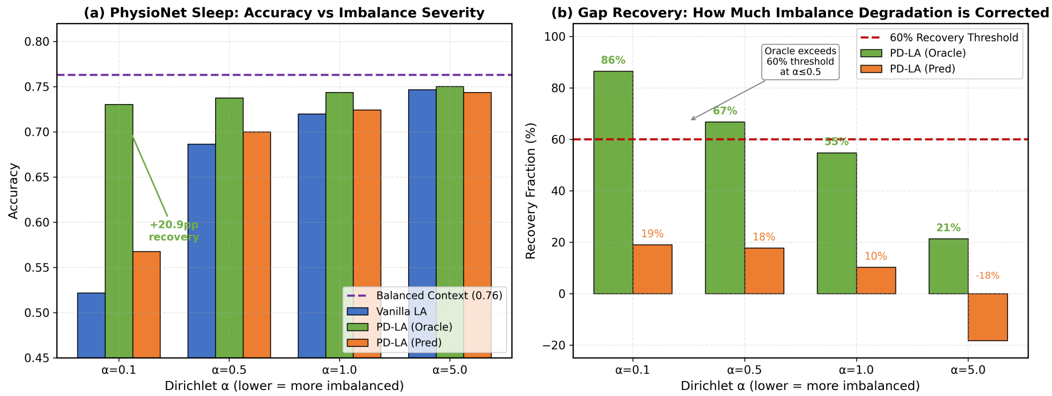

Figure 3: PD-LA performance on PhysioNet Sleep across imbalance severity levels. (a) Accuracy comparison showing Vanilla LA, PD-LA Oracle, and PD-LA Pred at different Dirichlet α values (lower α = more imbalanced). Oracle nearly matches balanced-context performance even at extreme imbalance (α=0.1). (b) Recovery fraction showing what percentage of the imbalance-induced accuracy gap is recovered by each variant.

S. Garg, Yifan Wu, Sivaraman Balakrishnan, and Zachary Chase Lipton. A unified view of label shift estimation. *ArXiv*, abs/2003.07554, 2020.

Taesik Gong, Jongheon Jeong, Taewon Kim, Yewon Kim, Jinwoo Shin, and Sung-Ju Lee. Note: Robust continual test-time adaptation against temporal correlation. 2022.

Bruna Junqueira, Bruno Aristimunha, Sylvain Chevallier, and Raphael Y de Camargo. A systematic evaluation of euclidean alignment with deep learning for eeg decoding. *Journal of Neural Engineering*, 21, 2024.

Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, S. Gordon, C. Hung, and Brent Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 15, 2016.

Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. *ArXiv*, abs/1603.04779, 2016.

Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. pp. 6028–6039, 2020.

Hyesu Lim, Byeonggeun Kim, J. Choo, and Sungha Choi. Ttn: A domain-shift aware batch normalization in test-time adaptation. *ArXiv*, abs/2302.05155, 2023.

Zachary Chase Lipton, Yu-Xiang Wang, and Alex Smola. Detecting and correcting for label shift with black box predictors. *ArXiv*, abs/1802.03916, 2018.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Z. Wen, Yaofo Chen, P. Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. *ArXiv*, abs/2302.12400, 2023.

Akara Supratak, Hao Dong, Chao Wu, and Yike Guo. Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25:1998–2008, 2017.

Pedro Vianna, Muawiz Chaudhary, Paria Mehrbod, An Tang, Guy Cloutier, Guy Wolf, Michael Eickenberg, and Eugene Belilovsky. Channel-selective normalization for label-shift robust test-time adaptation. pp. 514–533, 2024.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, B. Olshausen, and Trevor Darrell. Fully test-time adaptation by entropy minimization. *ArXiv*, abs/2006.10726, 2020.

Qin Wang, Olga Fink, L. Gool, and Dengxin Dai. Continual test-time domain adaptation. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7191–7201, 2022.

Xia Wei, A. Faisal, M. Grosse-Wentrup, Alexandre Gramfort, S. Chevallier, V. Jayaram, C. Jeunet, Stylianos Bakas, Siegfried Ludwig, Konstantinos Barmpas, Mehdi Bahri, Yannis Panagakis, N. Laskaris, D. Adamos, S. Zafeiriou, William Duong, S. Gordon, Vernon J. Lawhern, Maciej Śliwowski, Vincent Rouanne, and Piotr Tempczyk. 2021 beetl competition: Advancing transfer learning for subject independence heterogenous eeg data sets. pp. 205–219, 2022.

M. Zaheer, Satwik Kottur, Siamak Ravanbakhsh, B. Póczos, R. Salakhutdinov, and Alex Smola. Deep sets. 2017.