Does Time Hurt Like People?

Temporal Shift and Adaptation in Mobile Stress Sensing

Panyu Zhang · **Uichin Lee**
Graduate School of Data Science · Department of AI Computing
Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
Contact: panyu@kaist.ac.kr · uclee@kaist.ac.kr

Code

Paper

−5.25% AUROC drop from the random reference under cross-user evaluation

−8.05% AUROC drop from the random reference under temporal (chronological) evaluation

< 0.60 best temporal AUROC even with UDA (DANN-M3BAT: 0.592, +1.20%)

1 Motivation

- **Cross-user shift is the usual robustness test:** train on some users, evaluate on unseen users.
- **However, deployment is also longitudinal:** a model trained on earlier data is used later, after routines, contexts, and stressors change.
- **Features and the behavior–stress link can both drift,** so cross-user results need not imply reliable long-term performance.

RQ1 Does temporal shift degrade performance like cross-user shift does?

RQ2 Do unsupervised domain adaptation (UDA) methods help under temporal shift?

2 Dataset & Task

160 college students (users)

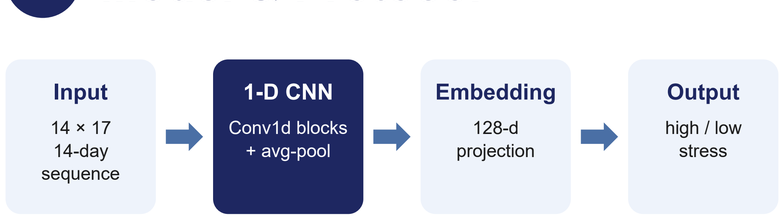
31,626 14-day prediction instances

14 × 17 days × daily sensing features

2017–2022 Sep 10, 2017 – Jun 25, 2022

- **College Experience Study (CES):** 17 daily passive smartphone-sensing features paired with self-reported stress EMAs.
- **Label:** high stress if the self-report is strictly above the user's median, computed from training/source rows only (leakage-free).
- **Metric:** pooled target AUROC, threshold-independent under label imbalance.

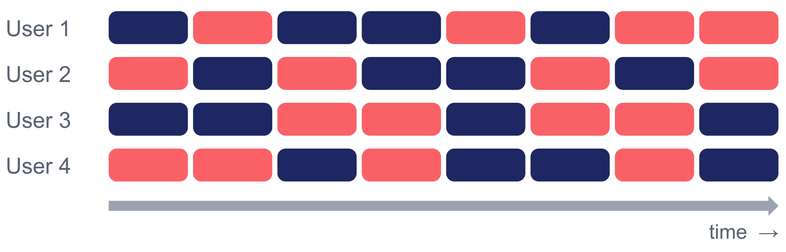
3 Model & Protocol



- **Ten UDA methods:** DANN, CDAN, DeepCORAL, ADDA, MCC, MCD, SHOT, CBST, DANN-M3BAT, CDAN-M3BAT.
- **M3BAT (Meegahapola et al., IMWUT 2024)** is the original multi-branch adversarial method: high/mid/low-shift feature branches weighted by Cohen's d, with a DANN discriminator. We extend it to CDAN (class-conditioned alignment), and all domain-based methods discriminate user identity instead of earlier vs. later.
- **Source-only model selection:** 30 Optuna trials and early stopping on source-validation AUROC; 5 paired repetitions (seeds 42–46); target labels only for the final score.

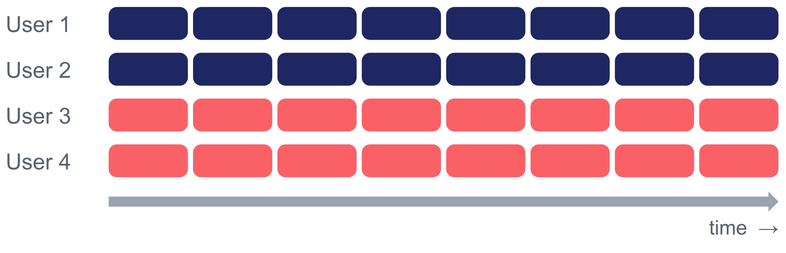
4 Three Evaluation Settings

Random 2-fold split



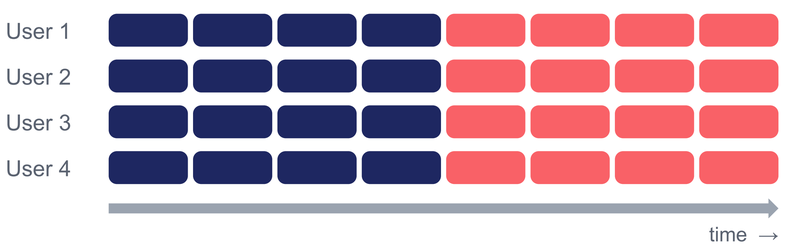
Users and time are mixed across folds: an optimistic i.i.d. reference, not a deployment setting.

Cross-user 2-fold



User-disjoint folds: zero-shot generalization to unseen users.

Temporal 50/50 (chronological)

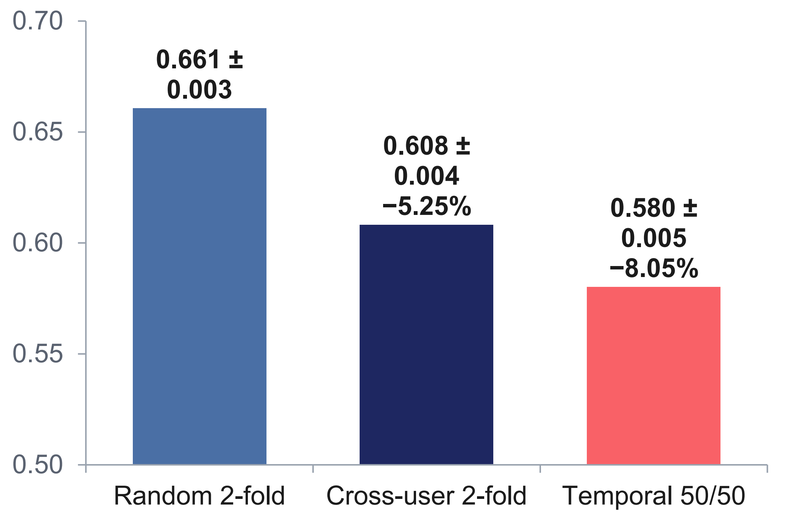


Each user's first half trains or adapts, the latter half is evaluated: generalization to future time.

Train / source Test / target

5 RQ1: Time hurts more than people

Baseline 1-D CNN AUROC. Axis starts at 0.5 (chance).

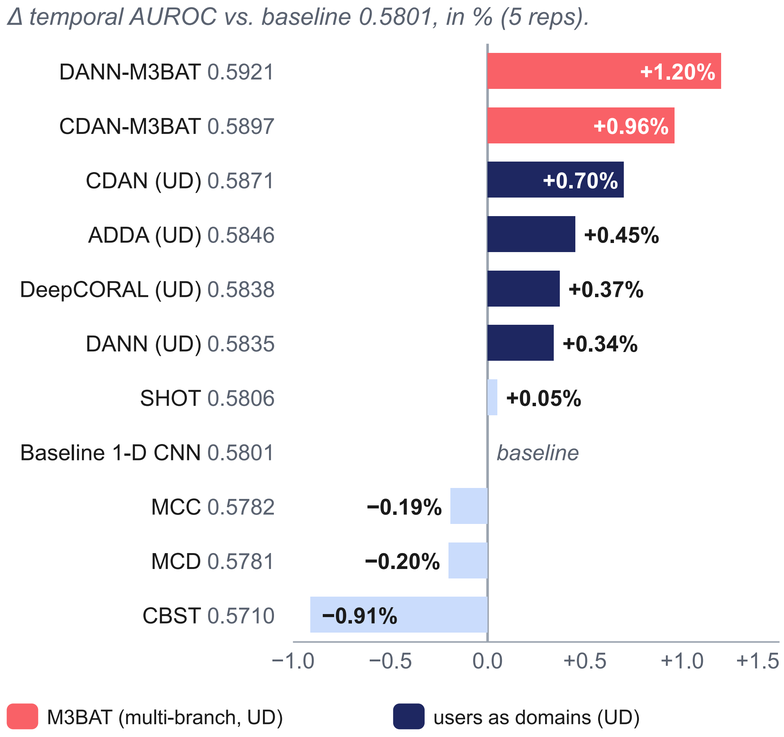


Labels drift too. High-stress prevalence rises from **30.67%** in the source period to **38.45%** in the target period (+7.78%).

- **The temporal gap exceeds the cross-user gap:** longitudinal drift reveals a deployment challenge that cross-user evaluation alone can miss (the random split is only an optimistic reference).

6 RQ2: UDA gains are small

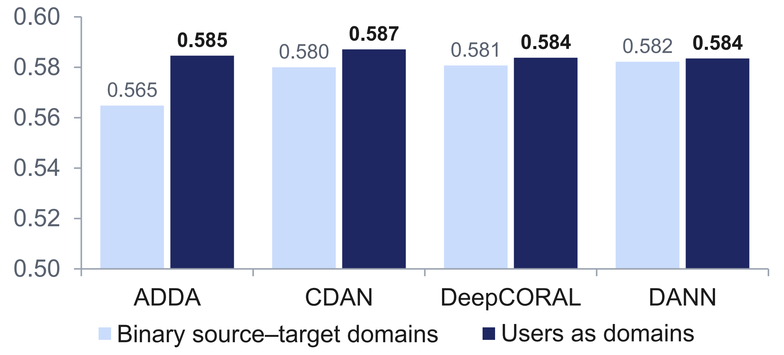
Δ temporal AUROC vs. baseline 0.5801, in % (5 reps).



- **All six user-as-domain methods beat the baseline;** methods without a domain term (SHOT, MCC, MCD, CBST) do not. The best AUROC, 0.592, is still below 0.60.

7 Users as domains help

Temporal AUROC. Axis starts at 0.5 = chance level.



- **Every convertible method improves over its binary earlier-vs-later form:** ADDA +1.98%, CDAN +0.71%, DeepCORAL +0.31%, DANN +0.13%. Multi-branch encoders add +0.26% (CDAN) and +0.86% (DANN) on top.

8 Takeaways & Future Work

- **Temporal shift deserves first-class evaluation,** not a secondary check after cross-user tests.
- **Inter-user variability is a useful adaptation signal** even under temporal drift; pooling users into two temporal domains wastes it.
- **Next: deployment-time update policies:** fixed-interval refresh, OOD-triggered test-time adaptation (TENT, SHOT), or a hybrid.
- **Limitations:** one cohort, one binary label, 2-fold dispersion only; drift sources are not disentangled.

Acknowledgments. Supported by the IITP grant (RS-2026-25552484) and the NRF grant (RS-2022-NR068758) funded by the Korean government (MSIT).

Data: College Experience Study (Nepal et al., IMWUT 2024), Kaggle: kaggle.com/datasets/subigyanepal/college-experience-dataset

Code: github.com/Kaist-ICLab/MobileStressTemporalAdaptation

Panyu Zhang and Uichin Lee. 2026. Does Time Hurt Like People? Temporal Shift and Adaptation in Mobile Stress Sensing.
UbiComp Companion '26, October 11–15, 2026, Shanghai, China.
https://doi.org/10.1145/3798063.3837183 · CC BY 4.0