Moodring: A Machine Learning Framework for Depression Prediction

Kunaal Agarwal[1,2], Matthew Clark[1], Afsaneh Doryab[1,3]

University of Virginia Department of: [1]Computer Science, [2]Biology, [3]Systems and Information Engineering

Motivation

Depression impacts **1 in 6 adolescents globally**, yet **treatment rates remain stagnant**; clinicians stress the importance of **objective**, tools to enable **earlier intervention** [1].

Global estimates show a **significant rise** in both **mild-to-moderate** and **moderate-to-severe depression** among adolescents between **1989** and **2022** ($P = 0.002$, $P = 0.034$), based on recent meta-analyses [1].

Study Design

Goal: predict depression severity from behavioral signals

- **63 Adolescents** over **24 weeks**
- Collected passive data from phones via AWARE framework
- Weekly **clinical assessments** using **patient surveys** (PHQ-8)
- **Feature categories** include **screen time, phone calls, battery usage, location, WiFi**

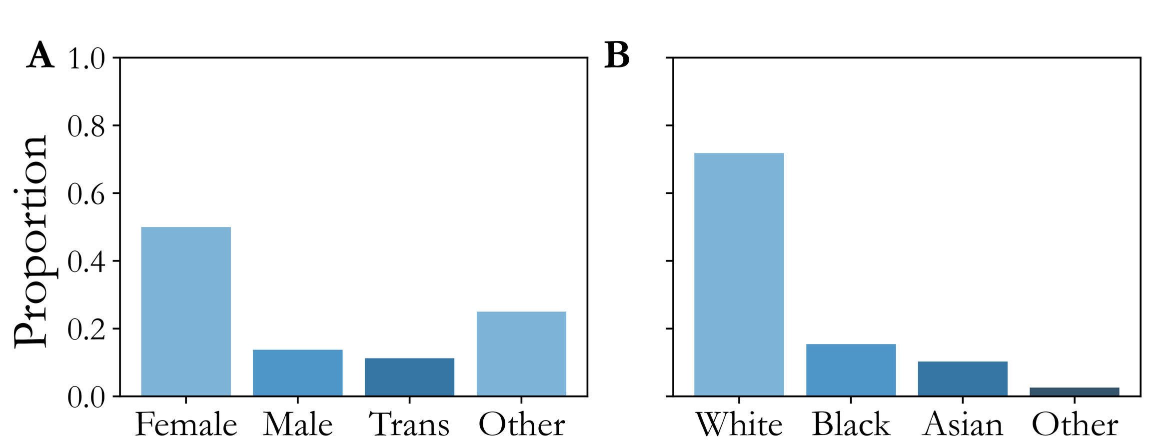


Fig 2. Demographics of study participants. A) most participants identified as female; B) majority of participants identified as white.

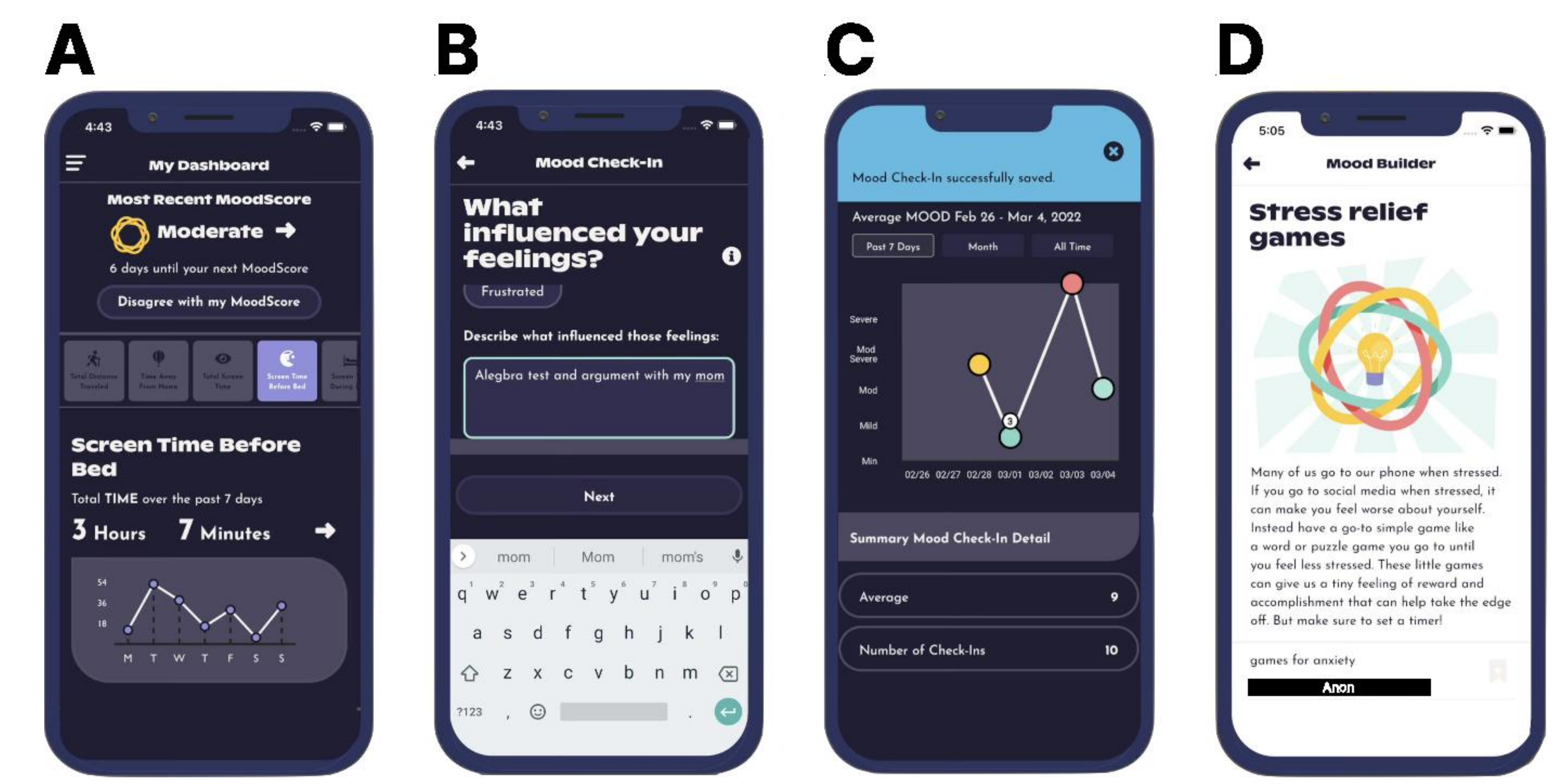
References

[1] [2] [3] [4]

Methods

Question	Abbreviation
Little interest or pleasure in doing things	PLEASURE
Feeling down, depressed, irritable or hopeless	HOPELESS
Trouble falling or staying asleep, or sleeping too much	SLEEP
Feeling tired or having little energy	TIRED
Poor appetite or overeating	APPETITE
Feeling bad about yourself - or that you are a failure or have let yourself or your family down	FAILURE
Trouble concentrating on things, such as school work, reading or watching television	CONCENTRATION
Moving or speaking so slowly that other people could have noticed? Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual	RESTLESS

Table 1. PHQ-8 items used as individual prediction targets.



- Used **ensemble regression models** to predict **PHQ-8 individual question and composite scores**.
- Built **personalized and population** models, both included cross-validation.
- Evaluated performance with **mean absolute error (MAE)** based on weekly clinical surveys.
- Applied cleaning and **multivariate imputation** to sensor data before training.
- Improved accuracy with **feature engineering** and **hyperparameter tuning**

Figure 3. Screenshots from the *Moodring* mobile app interface.
(A) Dashboard.
(B) Mood Check-In.
(C) Mood patterns over time.
(D) Mood Builder.

Results

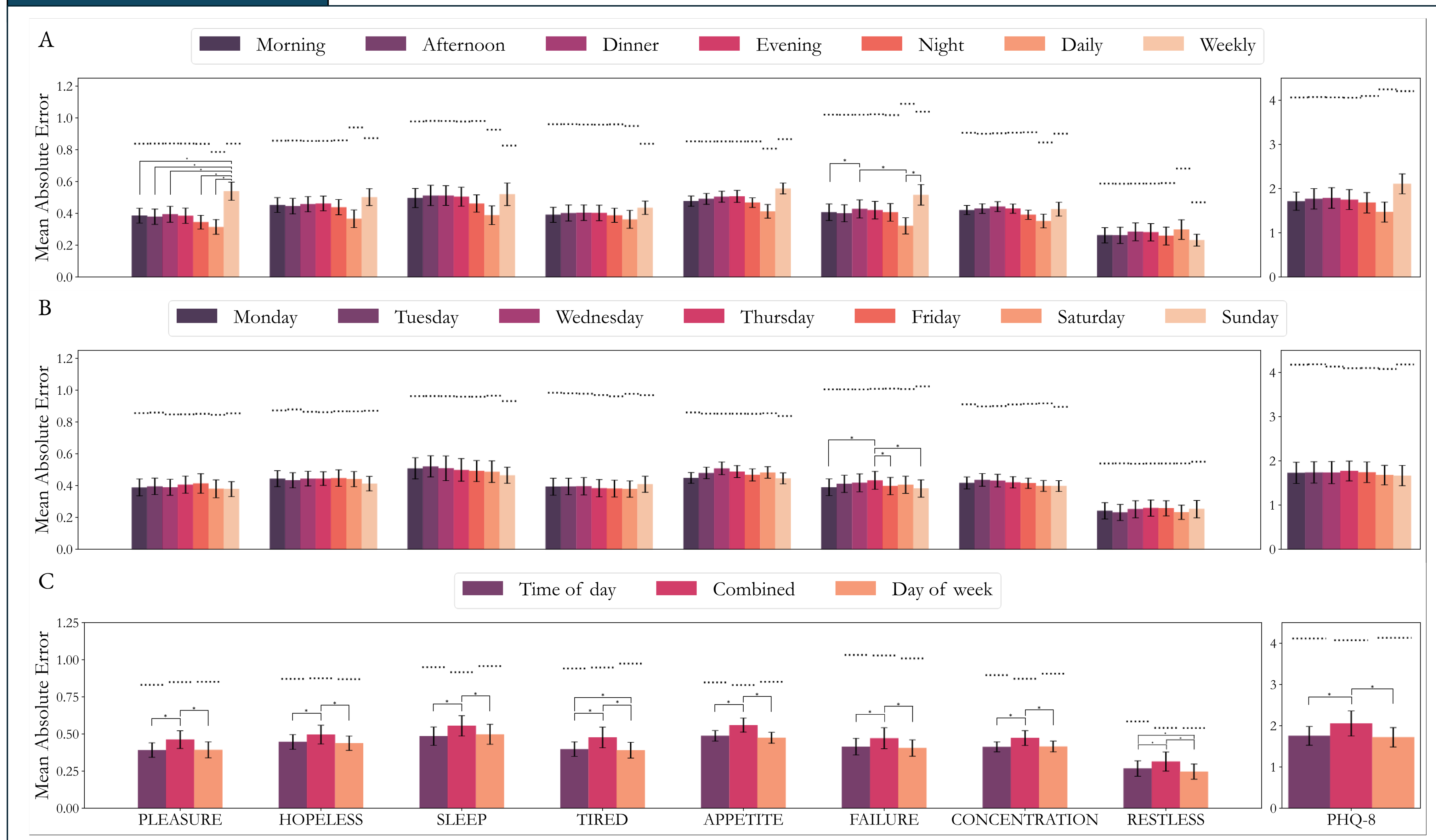


Figure 4. Model performance in predicting PHQ-8 labels across temporal groupings.
Only the best-performing algorithm (AdaBoost or Random Forest) is shown for each condition.
Dotted lines indicate expected error from a naïve baseline.
Asterisks (*) denote statistically significant differences in **MAE** ($p < .05$, Wilcoxon rank-sum test).

- Models showed **strong performance** in predicting **all PHQ-8 labels.**
- Achieved **1.47 MAE** for PHQ-8 (*range: 0-24*), **highlighting clinical potential**
- **Personalized models** significantly outperformed population-models
- **Temporal subsets** (e.g. morning or weekday data) **performed as well or better** than full training sets
- All time-of-day models outperformed weekly models
- **Findings support context-aware modeling as a strategy for improving depression prediction and minimizing invasiveness**