Predicting Craving-Related Emotions among Opioid Use Disorder Patients: Preliminary Results

Zachary King[1], Zoe Setiadi[2], Liana Hamdan[3], Hajar Ahmed[3], Bishal Lamichhane[1], Ashutosh Sabharwal[1], Ramiro Salas[3,4,5], Nidal Moukaddam*[3], & Akane Sano*[1]

[1]Department of Electrical and Computer Engineering, Rice University, Houston, TX
[2]Departments of Statistics and Psychological Sciences, Rice University, Houston, TX
[3]Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX
[4]Center for Translational Research on Inflammatory Diseases, Michael DeBakey VA Medical Center, Houston, TX
[5] The Menninger Clinic, Baylor College of Medicine, Houston, TX

***Abstract*—Individuals with Opioid Use Disorder (OUD) often struggle to maintain sobriety, with many experiencing relapse within the first year. While medication-assisted treatment (MAT) is among the most effective approaches, access to intensive care is often limited by financial barriers. Mobile health (mHealth) technologies offer a promising, cost-effective alternative by enabling continuous monitoring and timely intervention through tools such as ecological momentary assessments (EMAs), wearable sensors, and smartphone data. In this study, we explore the feasibility of using mHealth data to predict emotions that align with cravings in OUD patients undergoing MAT. Using data collected from EMAs, wearables, smartphone tracking, and surveys, we demonstrate that machine learning models can accurately predict emotional states associated with cravings. These findings highlight the potential of mHealth systems to support individuals with OUD through timely and scalable interventions.**

I. Introduction

Substance use disorder (SUD) is a major public health concern, affecting over 20 million Americans [1], and costing hospitals nearly 13.2 billion dollars a year [2]. Managing and intervening in SUD is crucial for reducing negative health outcomes and improving well-being. However, there are many barriers preventing healthcare providers from adequately treating SUD. Madras et al. defined five barriers: provider, institution, regulatory, financial, and other [3]. Many of these barriers stem from the substantial financial costs and the time investment required by providers to deliver effective treatment for substance abuse.

The addiction cycle starts with the initiation of sobriety, followed by withdrawal and stress, then cravings and adverse emotions, culminating ultimately in relapse [4]. Approximately 40-60% of individuals with SUD relapse within a year of achieving sobriety [5], and on average, it takes as many as 5.3 attempts to maintain long-term sobriety [6]. These statistics highlight the pervasiveness of the addiction cycle. However, we may be able to interrupt this cycle before a relapse occurs through Just-In-Time Adaptive Interventions (JITAI), which provides support at specific times based on an individual's needs [7].

This work was supported by the National Institutes of Health 1R01DA059925, Rice University Institute of Biosciences and Bioengineering Hamill Innovation Grants program, Sony Research Award Program.* are co-senior authors.

Researchers have explored mobile health (mHealth) technologies, including ecological momentary assessments (EMA) and wearable sensors, to monitor and intervene in SUD [8], [9], [10], [11]. While EMA remains a common tool for capturing momentary states, machine learning models leveraging data from wearable devices have shown promise in predicting stress and cravings with up to 86% accuracy [12]. GPS-based sensing and HRV-triggered JITAIs have also demonstrated predictive and therapeutic potential [13]. However, reliance on user-tagged events and surveys presents limitations, highlighting the need for minimally burdensome, context-aware sensing systems to support personalized treatment.

We address current limitations using multimodal analyses and modeling for patients with Opioid Use Disorders (OUD) undergoing Medication-assisted treatment (MAT) for one month [14]. The project employs a combination of lab and field studies, including functional magnetic resonance imaging (fMRI), questionnaires, wearables, a smartphone tracking application, and EMA to develop novel methods for modeling multimodal data. In this paper, we aim to explore preliminary craving and affect analyses and mHealth-based affect detection that can eventually guide adaptive interventions. The contributions of this paper can be summarized as follows:

- We present a multimodal longitudinal study of 35 OUD patients receiving MAT, capturing contextual, physiological, and EMA data.
- We provide a preliminary analysis of craving and emotions.
- We develop machine learning models to predict positive vs. negative emotions, which we identified as being linked to cravings through past literature and our analyses.

II. Study

A. Study Design

The study enrolled 35 participants (23 female and 12 male) receiving Buprenorphine/Naloxone from a clinic located at Ben Taub Hospital in Houston, Texas. Participants were between 21 and 64 years of age (mean age = 43, SD = 10.7). Most participants identified as White (26), followed by African American (8), and Other (1); 8 identified as Hispanic.

Eligible patients were referred by clinic psychiatrists and contacted by phone for consent and onboarding. At the initial visit, participants completed surveys and set up EMA, phone sensing, and Fitbit tracking. Final visits were scheduled 1–2 months later.

B. Wearable

Participants were provided with the Fitbit Inspire 2 and guided through setup using a prepared installation guide and pre-registered accounts for data access via the Fitbit API. The device collected minute-by-minute intraday data, including timestamps, step count, average heart rate, and sleep stages (awake, asleep, or restless).

C. Phone applications

Participants installed two smartphone apps alongside the Fitbit app (iOS $\geq$ 15, Android $\geq$ 9). The AWARE app [15] collected call and message logs, location, and screen usage, with all tracked data disclosed for privacy transparency. The CraveSense app prompted four daily EMAs, balancing data quality with participant burden. EMAs included sleep, stress, mood, and the Geneva Emotion Wheel (GEW) [16] (Fig. 1). The GEW is divided into four quadrants. The x-axis separates responses by emotion type, with negative emotions on the left and positive emotions on the right. The y-axis differentiates responses by control level, with low control emotions in the bottom and high control emotions in the top half. After choosing up to three emotions, the participants were then asked to provide the intensity of that emotion on a scale between 1 and 5. Participants were also asked if they were currently experiencing cravings for opioids or any other drug; if they indicated that they were experiencing a craving, they would then be asked to fill out the Brief Substance Craving Scale [17].

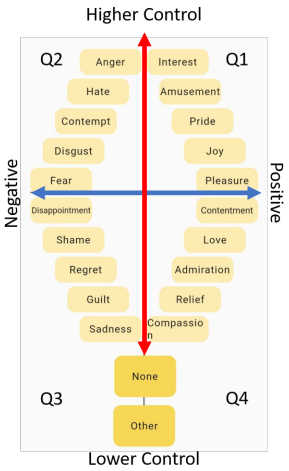

Fig. 1. Screenshot from the Cravesense APP of the Geneva Emotion Wheel (GEW). The Quadrants of the GEW are also displayed. The GEW splits emotions based on emotional state (x-axis) and level of control (y-axis).

TABLE I
FEATURES EXTRACTED FROM FITBIT AND AWARE DATA.

Modality	**Features**
Steps	Mean, Standard Deviation, Max
Heart Rate	Mean, Standard Deviation, Max, Min
Sleep	Time Asleep, Time Restless, Time Awake, Sleep Regularity Index
Phone Calls	# of Incoming and Outgoing Calls, Call Duration, Common Callers
SMS	Messages Sent, Messages Received, Common Texters
Location	Cluster Location, Location Count
Screen	Screen Usage

III. ANALYSES AND MODELING

A. Data Processing

We segment the data into 30-minute intervals and extract statistical features from each data source within these segments. A list of all the features extracted is found in Table I. For heart rate and step count, we calculate statistical features. For sleep, we extract four features that summarize the participant's sleep from the previous night: Time Awake, Time Restless, Time Asleep, and the Sleep Regularity Index (SRI). SRI represents the consistency of a person's sleep pattern by assessing the likelihood of sleep occurring 24 hours apart [18].

For phone data, we record the number and duration of phone calls, the number of text messages sent and received, and the duration of screen time in each interval. We rank callers and texters in ascending order by how often the participant interacts with them, and for each segment, we identify frequently contacted persons for calls and text messages. For GPS features, we cluster the locations by applying a dynamic k-Means clustering algorithm to stationary GPS data, determining the optimal number of clusters based on a maximum distance threshold. We assign the most frequently visited cluster location to each interval and record the number of unique locations visited within that interval.

B. Analysis

We analyze the reported emotional responses obtained through EMAs, particularly how they differed when participants indicated cravings. Many of our participants report few, if any, cravings (20% of responses). Therefore, we focus on participants who reported cravings and analyze whether their emotional states differ significantly between craving and non-craving moments. We evaluate these relationships using Cohen's Kappa to measure agreement between binary emotion labels (e.g., Q1–Q4, Positive/Negative, High/Low Control) and binary craving responses.

C. Machine Learning Models

We train models to predict emotion labels using data collected 90 to 120 minutes before each EMA response. This approach enables the delivery of interventions before craving episodes occur. In addition, given potential delays in wearable sensor data, particularly with devices such as Fitbit, predicting future emotional states allows sufficient time to collect the necessary data and stay ahead of participant cravings.

We explore several machine learning algorithms for both binary classification and regression tasks, drawing from previous research. We utilize algorithms such as XGBoost (XGB), Random Forest (RF), and Support Vector Machine (SVM) for binary classification. [8], [13], [19]. We opt not to employ more complex models, such as neural networks, due to the limited size of the labeled data. We also include a baseline model that labels the data points in the test set based on a probabilistic random choice, reflecting the distribution of class labels in the training set.

1) Class Labels: Based on positive and negative GEW emotions, we define a binary label, Positive vs. Negative (PVN), where responses are labeled as positive if only positive emotions are reported, negative if only negative emotions are reported, and excluded if both are reported. Given the high rate of negative emotions during craving instances compared to non-craving instances, and the opposite pattern observed for positive emotions, we hypothesize that the PVN label may serve as an effective proxy for predicting drug cravings.

2) Cross-Validation: We split the data into training and test sets with two kinds of strategies: Leave-One-Subject-Out (LOSO) and random. The random strategy is to split all users' data points randomly; therefore, the training and test sets contain data points from all users. We refrain from using a purely personalized approach due to insufficient labeled data points per participant. Although a personalized model might be feasible for some individuals, the data collected from Fitbit, AWARE, and EMAs varied significantly among participants. The LOSO strategy splits the data into observed and unobserved users. The training set contains data points from observed users, and the test set contains unobserved users. This strategy demonstrates the model's generalizability when predicting new participants' emotional states. Hyperparameters were tuned via exhaustive grid search within each cross-validation fold. For XGBoost and Random Forest we varied n_estimators (20-200), max_depth (2-5), and min_samples_split (1-3); for SVM we compared linear and Radial Basis Function (RBF) kernels with grids over C (0.1-10) and γ (0.0001-0.1).

3) Model Evaluation: Our model evaluation encompasses accuracy, precision, recall, and F1 score for each of the four machine learning algorithms, including the baseline model. Furthermore, we provide an interpretation of these models by highlighting the importance of features across all iterations of model prediction.

IV. Results & Discussion

A. Study Statistics

We collected 1,674 EMA responses, with participants reporting cravings in 171 instances. Response counts varied widely, averaging 60 per participant (SD = 50.1), with some submitting only one. Data availability also varied, averaging 25 days (SD = 20.9) of Fitbit data and 9.6 days (SD = 7.4) of AWARE data.

TABLE II
Distribution of classification labels. Q1 (Positive & High Control), Q2 (Negative & High Control), Q3 (Negative & Low Control), and Q4 (Positive & Low Control)

Class Label	Craving(responses=171)	Non-Craving(responses=1503)
Q1	36 (21%)	687 (46%)
Q2	90 (53%)	217 (14%)
Q3	100 (58%)	210 (14%)
Q4	72 (42%)	754 (50%)
Positive	82 (48%)	1029 (68%)
Negative	117 (68%)	353 (23%)
High Control	119 (70%)	846 (56%)
Low Control	131 (78%)	911 (61%)

TABLE III
Prediction results for Positive vs. Negative (PVN) 90 minutes ahead using Fitbit and AWARE data. Results average 20 runs (standard deviations in parentheses). RF: Random Forest, SVM: Support Vector Machine, XGB: XGBoost, LOSO: Leave-One-Subject-Out, CV: cross-validation. Random-CV ANOVA found significant metric differences; Tukey tests confirmed Random Forest differed from all other algorithms.

Algorithm	CV	Accuracy	Recall	Precision	F1
Random Baseline	Random	0.63 (0.13)	0.57 (0.23)	0.63 (0.13)	0.52 (0.18)
	LOSO	0.49 (0.21)	0.12 (0.24)	0.49 (0.21)	0.11 (14)
RF	Random	0.73* (0.06)	0.76* (0.12)	0.73* (0.06)	0.67* (0.09)
	LOSO	0.59 (0.25)	0.05 (0.08)	0.59 (0.25)	0.05 (0.09)
SVM	Random	0.64 (0.11)	0.45 (0.25)	0.64 (0.11)	0.45 (0.18)
	LOSO	0.52 (0.24)	0.11 (0.21)	0.52 (0.23)	0.07 (0.12)
XGB	Random	0.67 (0.12)	0.51 (0.24)	0.67 (0.12)	0.51 (0.19)
	LOSO	0.55 (0.24)	0.11 (0.21)	0.55 (0.24)	0.09 (0.15)

B. EMA Analysis

Table II presents the distribution of GEW values reported during craving and non-craving episodes. Although many participants reported few or no cravings, negative emotions were reported far more frequently during cravings (68%) than non-cravings (23%). Similar results were found for Stress and Mood. Average Stress during cravings was 3.6 (SD=1.4), but decreased to 2.4 (SD=1.6) during non-cravings. Average Mood was 3.0 (SD=1.2) during cravings, and increased to 3.8 (SD=1.0) during non-cravings.

Cohen's kappa scores, used to assess agreement between binary GEW labels and craving instances, showed modest alignment for labels linked to negative emotions: 0.28 for Q2, 0.33 for Q3, and 0.25 for Negative Emotions (Q2 and 3). This supports the emotion distribution patterns in Table II. The low frequency of reported cravings, especially among those early in recovery, contrasts with existing literature [20]. This may stem from participants' reluctance to disclose cravings to avoid jeopardizing their access to Buprenorphine/Naloxone or straining their relationship with psychiatric providers. This may be due to social desirability bias, where participants present themselves to be seen in a more favorable way to others, in this case, the research team [21].

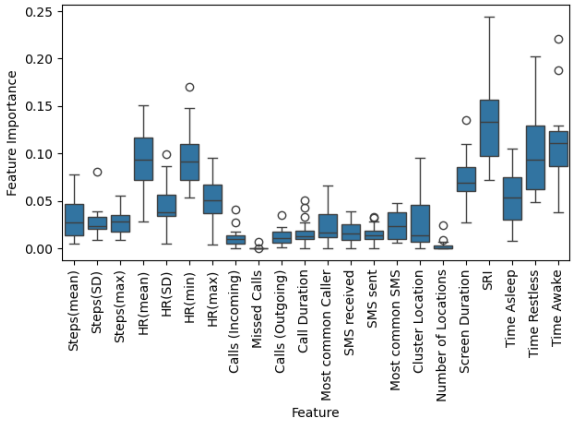

Fig. 2. Average Feature Importance across multiple runs using Random Forest, when predicting Positive vs. Negative (PVN).

C. Model Evaluation

The results in Table III demonstrate the model's ability to predict the PVN class label 90 minutes in advance. Notably, the RF classifier outperformed the random baseline, while the performance gains from SVM and XGB were marginal, if any. These differences in performance were confirmed using a one-way ANOVA, which showed that the random cross-validation results showed significant performance differences across algorithms. A Post-Hoc Tukey test revealed that RF was statistically different from the XGB, SVM, and baseline models across all four metrics. The disparity in performance between the LOSO and random cross-validation highlights the importance of incorporating personalization when deploying these models in real-world settings. Figure 2 presents a box-plot of feature importance for predicting PVN across all 20 iterations of our random cross-validation using the RF model, with Fitbit and AWARE data as input. The results highlight the significant role of sleep- and heart rate-related features.

V. Conclusion and Future Work

Our findings represent a significant step forward, demonstrating the feasibility of working with this population using mHealth tools to predict momentary affect and cravings. Despite the limited number of craving reports, our ability to predict related emotional states is promising, highlighting the potential of mHealth for monitoring and supporting recovery through timely interventions.

However, several limitations must be addressed in future work. The study's notable challenges related to participant engagement and data collection are noteworthy. To improve the robustness and reliability of our findings, we have updated our study protocol by incorporating engagement-based incentives and adopting more user-friendly third-party applications to increase participant engagement in our ongoing data collection. Analyses and modeling with a larger and more engaged cohort including other modality of data and parameters (e.g., temporal segmentation of features and labels) will be crucial for enhance performance and generalizability of our machine learning algorithms.

References

[1] L. R. McKnight-Eily, "Racial and ethnic disparities in the prevalence of stress and worry, mental health conditions, and increased substance use among adults during the covid-19 pandemic—united states, april and may 2020," *Morbidity and mortality weekly report*, vol. 70, 2021.

[2] C. Peterson, M. Li *et al.*, "Assessment of annual cost of substance use disorder in us hospitals," *JAMA network open*, vol. 4, no. 3, pp. e210 242–e210 242, 2021.

[3] B. K. Madras, N. J. Ahmad *et al.*, "Improving access to evidence-based medical treatment for opioid use disorder: strategies to address key barriers within the treatment system," *NAM perspectives*, vol. 2020, 2020.

[4] R. Zheng, L. Hao *et al.*, "Commentary: craving in opioid use disorder: from neurobiology to clinical practice," *Frontiers in psychiatry*, vol. 12, p. 615921, 2021.

[5] T. H. Brandon, J. I. Vidrine *et al.*, "Relapse and relapse prevention," *Annu. Rev. Clin. Psychol.*, vol. 3, pp. 257–284, 2007.

[6] J. F. Kelly, M. C. Greene *et al.*, "How many recovery attempts does it take to successfully resolve an alcohol or drug problem? estimates and correlates from a national study of recovering us adults," *Alcoholism: Clinical and Experimental Research*, vol. 43, no. 7, pp. 1533–1544, 2019.

[7] I. Nahum-Shani, S. N. Smith *et al.*, "Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support," *Annals of Behavioral Medicine*, pp. 1–17, 2018.

[8] S. Carreiro, K. K. Chintha *et al.*, "Wearable sensor-based detection of stress and craving in patients during treatment for substance use disorder: A mixed methods pilot study," *Drug and alcohol dependence*, vol. 209, p. 107929, 2020.

[9] B. T. Gullapalli, S. Carreiro *et al.*, "Opitrack: a wearable-based clinical opioid use tracker with temporal convolutional attention networks," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 5, no. 3, pp. 1–29, 2021.

[10] B. P. Chapman, B. T. Gullapalli *et al.*, "Impact of individual and treatment characteristics on wearable sensor-based digital biomarkers of opioid use," *NPJ Digital Medicine*, vol. 5, no. 1, p. 123, 2022.

[11] B. T. Gullapalli, S. Carreiro *et al.*, "Pharmacokinetics-informed neural network for predicting opioid administration moments with wearable sensors," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 21, 2024, pp. 22 892–22 898.

[12] R. K. Sah, M. McDonell *et al.*, "Adarp: A multi modal dataset for stress and alcohol relapse quantification in real life setting," in *2022 IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2022, pp. 1–4.

[13] D. Epstein, M. Tyburski *et al.*, "Prediction of stress and drug craving ninety minutes in the future with passively collected gps data. npj digit med. 20203, 26."

[14] N. Moukaddam, A. Sano *et al.*, "Turning data into better mental health: Past, present, and future," *Frontiers in Digital Health*, vol. 4, p. 916810, 2022.

[15] D. Ferreira, V. Kostakos *et al.*, "Aware: mobile context instrumentation framework," *Frontiers in ICT*, vol. 2, p. 6, 2015.

[16] V. Sacharin, K. Schlegel *et al.*, "Geneva emotion wheel rating study," *Center for Person, Kommunikation, Aalborg University, NCCR Affective Sciences. Aalborg University, Aalborg*, 2012.

[17] E. Somoza, S. Dyrenforth *et al.*, "In search of a universal drug craving scale," in *Annual Meeting of the American Psychiatric Association, Miami Florida*, 1995, pp. 121–132.

[18] A. J. Phillips, W. M. Clerx *et al.*, "Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing," *Scientific reports*, vol. 7, no. 1, p. 3216, 2017.

[19] A. J. Burgess-Hull, C. Brooks *et al.*, "Using machine learning to predict treatment adherence in patients on medication for opioid use disorder," *Journal of Addiction Medicine*, vol. 17, no. 1, pp. 28–34, 2023.

[20] J. Hampton, R. Eugene *et al.*, "Digital detection of craving and stress for individuals in recovery from substance use disorder: A qualitative study," *Drug and Alcohol Dependence Reports*, p. 100336, 2025.

[21] N. Bergen and R. Labonté, ""everything is perfect, and we have no problems": detecting and limiting social desirability bias in qualitative research," *Qualitative health research*, vol. 30, no. 5, pp. 783–792, 2020.