# Towards Automated, Interpretable and Unobtrusive Detection of Acute Marijuana Intoxication in the Natural Environment: Harnessing Smartphones, Wearables, Machine Learning and Explainable AI to Empower Clinical Decision Support for Just-In-Time Adaptive Interventions

Sang Won Bae, Tammy Chung, Tongze Zhang, Melik Ozolcer, Anind K Dey, Mohammad Rahul Islam

Submitted to: JMIR AI
on: August 29, 2023

# *Table of Contents*

**Original Manuscript** 5
**Supplementary Files** 34

# Towards Automated, Interpretable and Unobtrusive Detection of Acute Marijuana Intoxication in the Natural Environment: Harnessing Smartphones, Wearables, Machine Learning and Explainable AI to Empower Clinical Decision Support for Just-In-Time Adaptive Interventions

Sang Won Bae[1] PhD; Tammy Chung[2] PhD; Tongze Zhang[1] MSc; Melik Ozolcer[1] BE; Anind K Dey[3] PhD; Mohammad Rahul Islam[4] BSc

[1]Stevens Institute of Technology Human-Computer Interaction and Human-Centered AI Systems Lab AI for Healthcare Lab Hoboken US
[2]Rutgers University Institute for Health, Healthcare Policy and Aging Research Newark US
[3]University of Washington Information School Seattle US
[4]Stevens Institute of Technology Human-Computer Interaction and Human-Centered AI Systems Lab AI for Healthcare Lab, Hoboken US

**Corresponding Author:**
Sang Won Bae PhD
Stevens Institute of Technology
Human-Computer Interaction and Human-Centered AI Systems Lab
AI for Healthcare Lab
1 Castle Point Terrace
Hoboken
US

## *Abstract*

**Background:** Acute marijuana intoxication can impair motor skills and cognitive functions (e.g., attention, information processing). However, existing tools (e.g., blood, urine, saliva tests) do not accurately reflect ‘real-time’ acute marijuana intoxication.

**Objective:** Considering the absence of screening tools to detect acute marijuana intoxication and impairment-related harms, our objective is to examine whether integration of smartphone-based sensors with a wearable activity tracker (Fitbit), as more accessible devices using passive sensing, can enhance detection of episodes of acute marijuana intoxication in real-world settings. No prior work has determined the potential of utilizing data from both phone sensors and a wearable device to improve the accuracy of algorithms in detecting acute marijuana intoxication in real-life scenarios (‘outside of lab settings’), nor focused on developing explainable AI (XAI) to provide insights into the algorithmic decision-making process, specifically in detecting episodes of moderate-intensive marijuana intoxication, leveraging passive sensing technologies captured in real-world contexts.

**Methods:** To address these aims, we collected daily data using the Experience Sampling Method (ESM) for up to 30 days from 33 young adults using personal smartphone sensors and a Fitbit, and self-reported marijuana use. Participants provided subjective ratings of marijuana intoxication within 15 min of starting to use marijuana and during semi-random prompts 3 times per day: “low-intoxication” (rating?=?1–3) vs “moderate-intensive intoxication” (rating?=?4–10) vs. “not-intoxicated” (rating?=?0).

**Results:** Using the EXtreme Gradient Boosting Machine classifier (XGBoost) to model this data, our results indicated that the best model (MobiFit-model), which combined data from off-the-shelf mobile phone and wearable technologies, achieved accuracy of 99% (AUC=0.99, F1-score =0.85) in detecting acute marijuana intoxication (i.e., subjective sense of intoxication) in the natural environment. F1-score, which balances sensitivity and specificity, showed a significant improvement of 13% and 11% for the combined model (MobiFit) compared to using Mobile and Fitbit individually, respectively. Explainable AI (XAI) presented algorithmic decisions which revealed that self-reported moderate-intensive marijuana intoxication was associated with smartphone sensors and Fitbit features, specifically: elevated minimum heart rate, increased micro-movements, but reduced macro-movement (i.e., a smaller radius of gyration via GPS), and increased noise energy level around the participants.

**Conclusions:** This study demonstrates the promise that mobile phone sensors and off-the-shelf wearable devices hold for

automated and continuous detection of acute marijuana intoxication in daily life. Advanced algorithmic decision-making processes could provide insight into behavioral, physiological and environmental features' contributions that may be most useful, for example, in triggering the delivery of just-in-time interventions to prevent marijuana-related harm; however, in order to make the algorithm applicable in real-world settings, the usefulness and effectiveness of such algorithms-driven decisions need to undergo robust evaluation in collaboration with clinical experts.

(JMIR Preprints 29/08/2023:52270)
DOI: https://doi.org/10.2196/preprints.52270

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

# Original Manuscript

**Original Paper**

# Towards Automated, Interpretable and Unobtrusive Detection of Acute Marijuana Intoxication in the Natural Environment: Harnessing Smartphones, Wearables, Machine Learning and Explainable AI to Empower Clinical Decision Support for Just-In-Time Adaptive Interventions


## Abstract

**Background**. Acute marijuana intoxication can impair motor skills and cognitive functions (e.g., attention, information processing). However, existing tools (e.g., blood, urine, saliva tests) do not accurately reflect 'real-time' acute marijuana intoxication.
**Objective**. Considering the absence of screening tools to detect acute marijuana intoxication and impairment-related harms, our objective is to examine whether integration of smartphone-based sensors with a wearable activity tracker (Fitbit), as more accessible devices using passive sensing, can enhance detection of episodes of acute marijuana intoxication in real-world settings. No prior work has determined the potential of utilizing data from both phone sensors and a wearable device to improve the accuracy of algorithms in detecting acute marijuana intoxication in real-life scenarios ('outside of lab settings'), nor focused on developing explainable AI (XAI) to provide insights into the algorithmic decision-making process, specifically in detecting episodes of moderate-intensive marijuana intoxication, leveraging passive sensing technologies captured in real-world contexts.
**Methods**. To address these aims, we collected daily data using the Experience Sampling Method (ESM) for up to 30 days from 33 young adults using personal smartphone sensors and a Fitbit, and self-reported marijuana use. Participants provided subjective ratings of marijuana intoxication within 15 min of starting to use marijuana and during semi-random prompts 3 times per day: "low-intoxication" (rating = 1–3) vs "moderate-intensive intoxication" (rating = 4–10) vs. "not-intoxicated" (rating = 0).
**Results.** Using the EXtreme Gradient Boosting Machine classifier (XGBoost) to model this data, our results indicated that the best model (MobiFit-model), which combined data from off-the-shelf mobile phone and wearable technologies, achieved accuracy of 99% (AUC=0.99, $F_1$-score =0.85) in detecting acute marijuana intoxication (i.e., subjective sense of intoxication) in the natural environment. $F_1$-score, which balances sensitivity and specificity, showed a significant improvement of 13% and 11% for the combined model (MobiFit) compared to using Mobile and Fitbit individually, respectively. Explainable AI (XAI) presented algorithmic decisions which revealed that self-reported moderate-intensive marijuana intoxication was associated with smartphone sensors and Fitbit features, specifically: elevated minimum heart rate, increased micro-movements, but reduced macro-movement (i.e., a smaller radius of gyration via GPS), and increased noise energy level around the participants.
**Conclusion.** This study demonstrates the promise that mobile phone sensors and off-the-shelf wearable devices hold for automated and continuous detection of acute marijuana intoxication in daily life. Advanced algorithmic decision-making processes could provide insight into behavioral, physiological and environmental features' contributions that may be most useful, for example, in triggering the delivery of just-in-time interventions to prevent marijuana-related harm; however, in order to make the algorithm applicable in real-world settings, the usefulness and effectiveness of such algorithms-driven decisions need to undergo robust evaluation in collaboration with clinical experts.
Word count: 431 out of 450 maximum

### Keywords

Marijuana intoxication; smartphone-based sensors; wearables; Fitbit; Machine learning; eXtreme Gradient Boosting Machine classifier (XGBoost); algorithmic decision-making process; explainable artificial intelligence (XAI); clinical decision support tools; Just-In-Time Adaptive Interventions

## 1. Introduction

Acute effects of marijuana use can result in impaired motor skills and cognitive functioning (e.g., attention, information processing) [12, 31, 44]. These acute marijuana-related effects have been associated with adverse consequences such as poor academic and work performance, and increased risk for motor vehicle crashes and fatal collisions [31, 35]. THC (delta-9 tetrahydrocannabinol) is the principal psychoactive constituent of marijuana. This chemical component binds to receptors in the brain, which can result in a feeling of "euphoria" or subjective report of feeling "high" [39]. Due to risks associated with acute marijuana-related impairment caused by THC, there is a critical need to detect episodes of marijuana-related intoxication in real-time in the natural environment.

A few studies have used phone sensors or wearable devices to detect acute marijuana consumption. A lab study (n=10 participants) using smartphone sensors (accelerometer, gyroscope) to detect acute marijuana use (3% or 7% THC vs placebo) found that gait analysis using support vector machine resulted in accuracy of 92% ($F_1$-score =0.93) [24]. However, it is unclear whether model performance pertains to validation or test results. Another study developed an electrochemical biosensor ring to simultaneously and rapidly (within 3 minutes) detect salivary THC (minimum of

0.5 M) and alcohol (minimum of 0.2 mM) [29]. The wearable ring to simultaneously detect THC and alcohol was preliminarily validated in the lab using chemical assays, and with 1 participant [29]. Importantly, these studies using smartphone and wearable device to detect acute marijuana use were both conducted in lab settings, highlighting the need for smartphone and wearable sensor-based research conducted in naturalistic settings to increase ecological validity.

The ability to detect episodes of marijuana use in daily life would support the delivery of Just-In-Time [47] harm reduction interventions (e.g., avoid driving when intoxicated) [34]. Challenges exist, however, in detecting acute marijuana-related intoxication [20]. Existing testing methods (e.g., blood, urine, saliva, breath) are not useful for detecting acute marijuana-related intoxication or impairment in real-time [4]. THC could be detected in an individual's blood or urine for several days after consumption depending on factors such as recency, frequency, and chronicity of use [4]. Thus, a person who tests positive for THC might not be intoxicated or impaired at the time of testing [4]. Instead, we propose that passive sensing using personal smartphones could provide a method for detecting episodes of marijuana use in the natural environment using one's own subjective report of marijuana intoxication as ground truth. Our recent paper [55] shows that acute marijuana intoxication can be detected by mobile phone sensor-based features integrated with time features (e.g., day of the week, time of day) with 90% accuracy. We advance our previous work [55] by investigating potential benefits of adding data from a wearable device (Fitbit), leveraging physiological signals to improve detection of marijuana intoxication in naturalistic environments.

With a wearable device, we examined heart rate as a physiological marker of acute cannabis intoxication because an acute increase in resting heart rate (HR) is a consistent effect of marijuana use in lab studies [19, 27, 50]. Specifically, lab research has found that within 2-3 minutes of smoking marijuana, there is an acute increase (20-60% dose-dependent) in resting heart rate [27], which might represent a "biomarker" of the onset of a marijuana smoking episode. HR peaks 10-15 minutes after maximum THC levels, followed by a rapid decline [19, 27, 50]. Tolerance to acute effects of marijuana on HR may develop (e.g., from a mean increase of 44.6 to 6.6 beats per minute after 18-20 days of use) [19, 27, 50]. While between-subject variance is high, study participants have shown a linear increase in HR with higher marijuana doses in lab studies [19, 27, 50]. The increase in HR during episodes of marijuana use has been validated in lab settings, but has not yet been explored in a real-world context. We hypothesized that an acute increase in HR detected by an off-the-shelf wearable device (Fitbit) might represent an objective biomarker that is associated with subjective reports of marijuana intoxication in daily life.

Our goal in this study was to determine whether data from personal smartphone (e.g., accelerometer, GPS) and off-the-shelf wearable (Fitbit) device (e.g., heart rate) can be used to detect subjective marijuana intoxication ("feeling high") in the natural environment using a machine learning approach. No prior work has determined whether data collected from the combination of smartphone sensors and wearable sensor features can be used to detect marijuana intoxication outside of lab settings. Since more people own smartphones than a Fitbit device, we tested whether the additional "burden" of wearing a Fitbit device is justified for the detection of self-reported marijuana use, and if so, which Fitbit-derived features provide unique information in improving the accuracy of the best performing smartphone-based detection model. Specifically, we tested the performance of different types of sensor-based models using (1) only smartphone-based sensors, (2) only Fitbit data, and (3) the combination of smartphone-based sensors and Fitbit data. We then used Explainable Artificial Intelligence (XAI) to enhance understanding of key features associated with reports of intoxication. Identification of smartphone-based sensor and Fitbit features that can be used to accurately detect episodes of self-reported marijuana intoxication in the natural environment could ultimately be used to trigger Just-In-Time interventions. We hypothesize that the combination of mobile phone and Fitbit data (MobiFit model) will have better performance in detecting reports of acute marijuana intoxication compared to Mobile only and Fitbit only models. We also hypothesize that HR and activity (e.g., step count) data from Fitbit will be identified as important features contributing to detection of subjective intoxication, supporting the potential value and additional burden of collecting data from wearable device.

## 2. Methods

### 2.1 Recruitment and Participants

We recruited 33 young adults (ages 18-24, M=19.63, SD=1.80; 60.6% female-identifying) from the local community using flyers, ads, and a participant research registry to participate in a study on the effects of marijuana on health. The sample included 23 individuals who self-identified as White, 4 as Black and 6 as other race/ethnicity (i.e., Asian, Asian Indian, Hispanic or bi-racial). Eligibility criteria for study participation were: current marijuana use at least two times per week, mobile phone ownership, not currently seeking treatment for substance use, no self-reported history of psychosis, and not taking any medication or using any medical device (e.g., pacemaker) that could affect heart rate. The average age when participants first used marijuana was 16.48 (SD=1.84; range=13-22) and the average age of regular marijuana use (i.e., using marijuana at least once per month for at least six months) was 17.03 years (SD=1.72). In the sample, 24.2% reported daily marijuana use, 9% reported use 5-6 times per week, 66.7% reported use 2-4 times per week. Participants had an Android (3%) or iOS (97%) smartphone as their primary device.

### 2.2 IRB Approval and Ethical Consideration

This naturalistic, observational follow-along study was approved by our university's Institutional Review Board (IRB). As in similar IRB-approved observational studies [8], all participants were provided with information regarding local resources (e.g., treatment providers for medical and mental health services). The study obtained a National Institutes of Health Certificate of Confidentiality. Prior to study participation, research staff obtained written informed consent from eligible individuals, and reviewed participant confidentiality and privacy, including risks to confidentiality (e.g., possible breach), and methods that the study uses to protect participant confidentiality (e.g., secure data transmission protocols, data storage). Research staff made clear to individuals during the discussion of informed consent, and throughout the project, that participation was voluntary (e.g., that participants could turn off sensors at any time, or withdraw from participation at any time). As in other mobile health studies, informed consent included a review of the types of data to be collected, for how long data would be collected, and the purpose of data collection. As discussed with participants, research data are kept confidential to the extent possible, to protect participants' privacy and confidentiality. Prior research has used GPS with individuals who engage in illicit substance use (e.g., [13]) with protections for participant confidentiality.

### 2.3 Study Design

Participants completed a baseline lab assessment (interview, questionnaires, cognitive testing), and downloaded study apps from the AppStore or Google Play Store to their personal smartphones. Research staff trained participants to use the apps on their smartphone and the study-provided Fitbit Charge 2 for the purposes of data collection. Our mobile app delivered Experience Sampling Method (ESM) questions on marijuana use. The Fitbit Charge 2 is a wrist-worn device that collects data on heart rate, physical activity (e.g., step count), and sleep (e.g., time, duration, quality) (see Table 2 in Appendix 1 of the Supplementary file for details on Fitbit variables). We collected these passive, continuous, and objective sensor streams (smartphone sensor data, Fitbit) and self-reported subjective data on marijuana intoxication from participants for up to 30 days. We determined that a 30-day period would provide sufficient data given that participants would use marijuana multiple times during the data collection period based on study inclusion criteria regarding a minimum frequency of marijuana use. At the end of the study, participants completed a debriefing interview. Participants were compensated for time and effort in line with other research projects. Participants were compensated with US $75 for completing the baseline assessment, and US $25 for completing the study exit interview during which they provided feedback on their experiences in the study (qualitative data collected in a semi-structured interview). For each day on which >75% of data collection (e.g., Fitbit, ESM) was completed, they earned US $10. Participants were not additionally compensated for providing self-initiated reports of marijuana use.

## 2.4 Mobile Sensing Framework and Applications for Data Collection

### 2.4.1 AWARE app

**Our AWARE application is a mobile sensing framework [16] that collects data passively and continuously from smartphone sensors. The sensor data can be used to infer human behavior patterns using different types of sensors: Location (e.g., travelled distance, circadian rhythm), physical movement (e.g., acceleration, activity), device usage (e.g., unlock, charge, keypress, app usage), social patterns (e.g., communication and conversations), and environmental (e.g., wifi, Bluetooth, sound/ambient noise and light) context. To track natural behaviors of young adults in a real-life context, we built a mobile sensing app based on the AWARE framework that ran in the background 24/7 and collected passive sensor and meta-data (e.g., time-stamp communication logs) while young adults used their smartphones in daily life. Our app transferred the collected sensor data to a MySQL database on our secure server each day.**

### 2.4.2 Experience Sampling Method (ESM)

Our mobile phone application also was designed to capture self-reports of marijuana use by young adults. Two types of surveys were used: self-reports initiated when a participant used marijuana, and self-reports that were delivered at three fixed times each day (morning: 10am, afternoon: 3pm, evening: 8pm) to capture general behavior patterns throughout the day [59]. To accommodate participants' schedules, survey response windows were open for 5 hours; no reminders were provided to complete surveys after the initial notification was delivered. For self-initiated reports of marijuana use, participants were asked to initiate a report of marijuana use within 15 minutes of starting use (since the most common form of use was expected to be smoking or vaping), and to rate subjective intoxication "How high are you feeling right now?" (0 [none] to 10 [a lot]). Prior work has used a similar item to rate subjective marijuana intoxication [60]. Reports of marijuana use also asked the amount of marijuana consumed (in grams or hits), mode of use (e.g., bong, pipe), where and with whom they were with. After participants reported the start time, a reminder was sent two hours later to complete the "end session" survey (i.e., report the "end time" of the marijuana use episode). Participants were asked to report the end time as the time when they no longer felt high, if applicable. The three fixed time daily surveys (morning, afternoon, evening) collected information on context (e.g., location, companions); time since last marijuana use, marijuana craving, current mood and feelings (e.g., relaxed, anxious, sad), and other recent substance use (e.g., alcohol, tobacco). The app transferred participant self-reports to a secure server along with the sensor data each day.

### 2.4.3 Fitbit Charge 2

Participants were provided a Fitbit Charge 2 wearable device, asked to wear it as much as possible, and to keep it charged (e.g., charge the device when showering, but wear when sleeping). The Fitbit Charge 2 collected physiological data (e.g., heart rate), activity data (e.g., step count) and sleep. We hypothesized that heart rate and activity (e.g., step count) data could signal episodes of acute marijuana intoxication. We collected Fitbit data from the Fitbit server at the end of the study, using the Fitbit API.

## 2.5 Preparing self-report and Fitbit data for analysis

An episode of self-reported subjective marijuana intoxication was defined based on the ESM item: “How high are you feeling right now?”, rated 0-10 (0= “not high” to 10= “a lot”) [59] [60]. For inclusion in the analyses, the start and end time of the marijuana use episode had to be reported, so that the episode duration could be computed to permit labeling of the sensor stream. From all participants, we received 641 self-reports (Mean=9.86, Median=7 and SD=8.49) and 1556 with no marijuana use reports (see Figure 1). Of these 641 reports, 168 had a subjective intoxication rating of 0 and 10, and 6 had no subjective intoxication rating. After removing 6 self-reports where the subjective intoxication rating was not reported and 108 duplicate self-reports, we had 527 samples remaining. We further excluded reports for which both start and end time were missing (n=6), or only start time (n=110) or end time (n=9) was reported, or when the end time was reported as occurring earlier than the start time (n=45). We only included marijuana episodes when the duration of the smoking session was reported to be less than 3 hours, based on lab research in which researchers found that the effect of marijuana lasted < 3 hours [48]. In total, to be conservative in our analyses, we excluded 136 self-reported marijuana episodes for which the duration of the smoking episodes was longer than 3 hours and 1556 with no marijuana use reports [48].

**After excluding these episodes, there were 178 self-reported episodes of marijuana use (subjective high rating of 1-10) and 43 periods of “no marijuana use” (subjective high rating = 0) remaining (n=221). For our model building, we also had to exclude episodes for which we had no mobile sensor data (n=72), leaving a total of 221 marijuana self-reports. We also had to exclude episodes for which we had no Fitbit sensor data (n=17), leaving a total of 50 people. These 50 participants provided 132 marijuana use self-reports and 909 “no marijuana use” reports. We analyzed all reports from each participant, excluding those who only reported not using marijuana or had a rating of 0 subjective intoxications when using marijuana, leaving a total of 642 with no marijuana use report or who reported 0 subjective intoxications when using marijuana and 34 people. To prevent participants from using Fitbit incorrectly, we excluded users without heart rate data, leaving a total of 33 people, who provided a total of 769 events: 640 “no marijuana use” reports and 129 marijuana use self-reports.**

To capture behaviors from young adults in the real-world context when they were not using marijuana to compare with times when they reported “feeling high”, we used the afternoon (n=1151), and evening (n=950) surveys in which the participant reported “no/yes” to marijuana use (n=2111). In a total of 1,556 ESM reports, participants reported “no” to the ESM item “Did you smoke marijuana since the last report?” and the response to “When was the last time you used marijuana?” corresponded to the last self-initiated ESM survey. The 1,556 time-stamped surveys were labeled as “0” for the subjective rating of marijuana “high” in the final dataset.

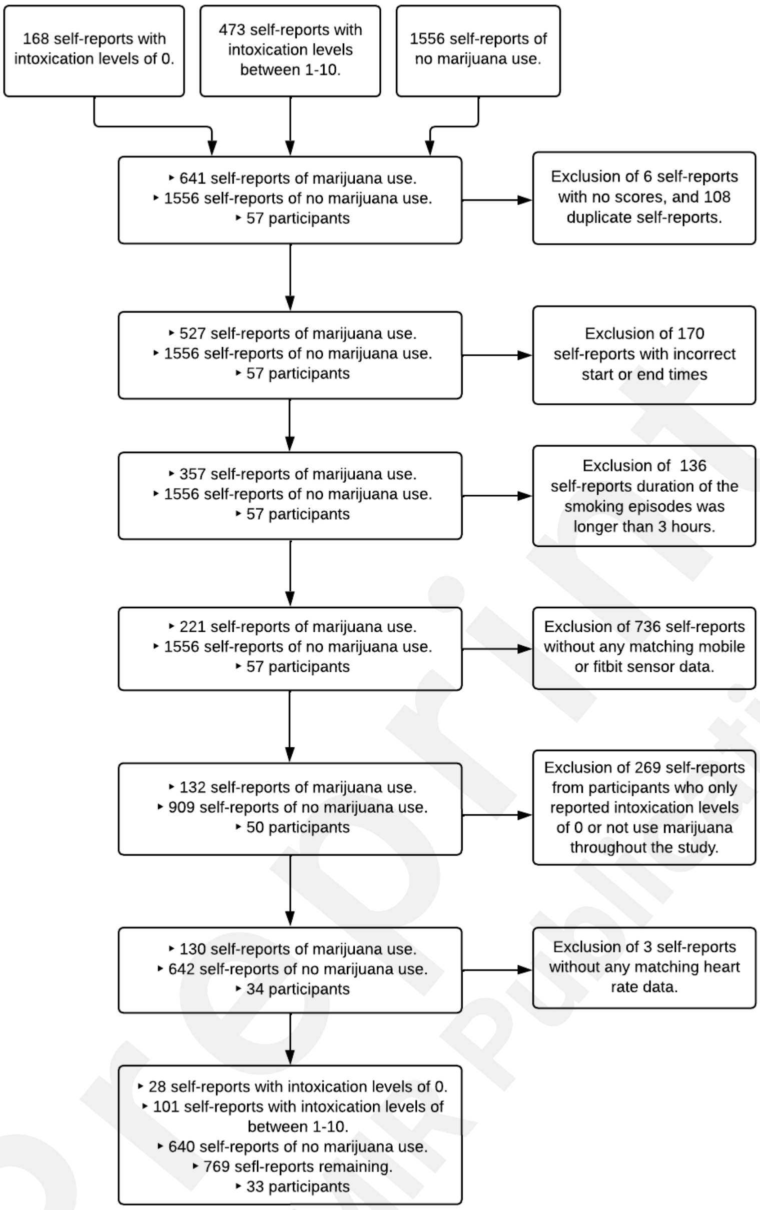


Figure 1. Flow chart of participants and data included in analyses

### 2.6 Extracting Smartphone and Fitbit Sensor Features

As in prior work, we extracted audio features to detect social interactions [30, 40], which might be associated with marijuana use. We computed device usage features such as smartphone unlock minutes and length of device interaction sessions. We extracted GPS features to examine movement patterns in daily life that might be associated with marijuana use [2, 3, 5, 9]: radius of gyration, time at a location cluster, total distances and number of clusters. Acceleration and phone angles were extracted. Features were extracted using the conversation plug-in, including attributes such as noise and voice. Finally, environmental features, including number of Bluetooth devices contacted, the most frequent wifi access point contacted, and light features (e.g., avg., and max. lux) were extracted. For almost all features, the minimum (min), maximum (max), average (avg), median (med), and standard deviation (std) were calculated. Additional information about the smartphone features is available in the Supplementary file.

All sensor feature statistics were extracted using a 5-minute time window. Specifically, we used a 5-minute segment because when marijuana is smoked in lab studies, there is an acute increase (20-60% dose-dependent acute increase) in resting heart rate within 2- 3 minutes (on average [19, 27], which might represent a "biomarker" of the onset of a marijuana smoking episode. In addition, participants reported marijuana use sessions that lasted an average of 75 minutes (SD=46.2). If we segmented data into larger time intervals (e.g., 30-minutes), we would likely include data that were not part of episodes of marijuana use, and might "average out" short-lived peaks in sensor data that signal important changes associated with marijuana use.

From the Fitbit, raw data on heart rate, sleep, and steps taken were extracted. To analyze Fitbit heart rate (HR) data, we first obtained heart rate and step count data using the Fitbit API, aggregated per-minute. We removed sensor streams when heart rate had a ‘0’ value. We extracted the following feature statistics: average, standard deviation, minimum, median, and maximum of heart rate within a 5-minute time window to explore relations between marijuana intoxication (“feeling of moderate-intensive intoxication” vs. “low-intoxication” vs. “no-intoxication”) and heart rate. We extracted resting heart rate by taking HR data when the participant was sedentary (i.e., no steps taken) for more than 5 minutes. To deepen our insights into HR patterns and their relevance for marijuana intoxication detection, we employed the extraction of heart rate feature characteristics, specifically the degree of peakedness (kurtosis) and asymmetry (skewness). These features were chosen based on their potential to offer distinct indicators of physiological changes associated with marijuana intoxication, as established in [62]. By analyzing kurtosis and skewness, we aim to capture nuanced variations in heart rate patterns that could serve as valuable markers for identifying marijuana intoxication within real-world settings that are not directly observable.

We suggest new features that can be used to identify subjective reports of marijuana intoxication. The Fitbit resting heart rate feature was calculated as follows. After aggregation within a 5-minute window, if a person did not move (i.e., no steps) within a window of 5 minutes, then the resting heart rate was computed during the window. The ‘pace’, ‘walk speed’ and ‘sedentary’ features were extracted as follows: Pace = time (minutes)/distance (meter), speed walk = 1/pace (meters/minute). For ‘moving’, at each minute, we checked whether the person was moving or not. If the number of steps per minute was greater than 0, then that particular minute was marked as moving (i.e., 1), otherwise not moving (i.e., 0). For the sedentary feature [58], each minute was checked for a sedentary bout. If the number of steps per minute was 0 then that particular minute was marked as a sedentary bout (i.e., 1) otherwise, it was not a sedentary bout (i.e., 0). Additionally, we extracted the number of minutes that someone is awake during the night when their sleep is disrupted the night before self-report of marijuana intoxication. (See details in Table 4 in the Supplementary file).

### 2.7 Ground Truth and Labeling Sensor Data

To achieve our goal of understanding behaviors in young adults during acute marijuana intoxication, we first needed to define a duration of marijuana use to label the mobile sensor data. Using reported start and end time of marijuana use, we identified episodes of marijuana use that were equal to or less than 3 hours in duration. We excluded the 3 hours of sensor data immediately after the “end time” report from analyses because we assumed that an individual could still be under the influence of marijuana during this period, which could affect the identification of periods when “no marijuana use” was reported (subjective intoxication rating=0). For example, if the duration of a marijuana use episode was less than 3 hours, the marijuana use session started at 6:00pm and continued until 6:30pm for example, then considering the remaining effect of marijuana use, we excluded the 3 hours from 6:30pm -9:30pm and labeled sensor streams as “no marijuana use” starting from 9:31 pm until the next episode of marijuana use occurred. We excluded the 30 minutes prior to the start time of reported episodes of marijuana use to be conservative based on findings in our pilot work indicating that self-initiated reports of marijuana use might be delayed between 5-15 minutes. For collecting non-marijuana candidates, we took sensor samples randomly throughout the day on which a participant did not use marijuana (i.e., non-marijuana days) to avoid any mixing of samples from marijuana use days. These non-marijuana candidates were then labeled using the morning/afternoon/evening surveys in which the participant reported “no” to the ESM item “Did you smoke marijuana since the last report?” and the response to “When was the last time you used marijuana?” was 5 hours prior to the ESM timestamp (see Fig. 2).

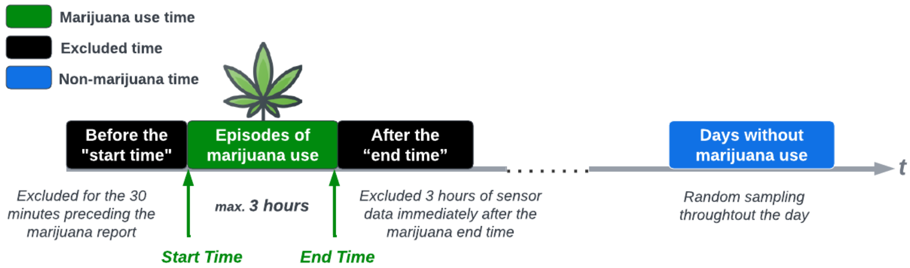


Figure 2. Marijuana use episodes and labeling principle

We intended to capture acute marijuana intoxication versus non-marijuana use. We labeled self-reported episodes of marijuana intoxication as a three-class classification problem: 0= “not intoxicated”, 1-3 rating of subjective marijuana intoxication (“low intoxication”), and 4-10 rating of subjective marijuana intoxication (“moderate-intensive intoxication”). In total, we labeled a total of 32,722 sensor stream samples (unit: 5 minute-window) as not intoxicated (154 windows from the self-initiated survey coded as 0 high, and 32,586 from the time-based self-reports), 423 sensor stream samples as “low intoxication” (ratings between 1-3; and 772 sensor stream samples as “moderate-intensive” (ratings between 4-10, where 10 = “a lot”).

Since we collected data from two devices (smartphone, Fitbit) the sample space for both devices were different. To make the results comparable we down sampled the mobile phone dataset by only keeping samples that were overlapping the available Fitbit data and combined the model sample space. As a result, we created three datasets (a) XGBoost-Mobile: mobile phone only samples, (b) XGBoost-Fitbit: Fitbit-only samples, and (c) XGBoost-MobiFit: mobile and Fitbit features combined.

### 2.8 Machine Learning Pipeline

We began by partitioning the labeled sensor data into training (80%) and test (20% holdout) datasets through a random split. Subsequently, we conducted leave-10%-samples-out cross-validation (CV) while employing Synthetic Minority Over-sampling Technique (SMOTE) for over-sampling. To report the ultimate model evaluation, we utilized the reserved test data (i.e., the 20% unseen data).

### Feature Selection, Hyper-Parameter Tuning and Cross-Validation

To avoid the effect of imbalanced data influencing model performance, in the training dataset, (1) we first removed features with a correlation coefficient higher than 0.9 and (2) used feature selection to choose the features with the Gini coefficient [57] importance greater than 0.005. Higher Gini values indicate greater feature importance. We then tried both over-sampling with SMOTE and random under-sampling of the majority class (i.e., “not- intoxicated”), so that the three classes (“moderate-intensive intoxication” (rating=4-10), “low-intoxication” (rating=1-3) and “not-intoxicated” (rating=0)) had the same number of training samples. Next, we leveraged Optuna [61] which uses Bayesian optimization to select the optimal combination of hyperparameters that maximizes the performance of the model. To ensure the accuracy of the hyperparameter results, we performed 10-fold CV on the models corresponding to the parameter combinations selected by Optuna. The 10-fold CV involves dividing the data into ten equal parts, training the model on nine parts, and evaluating the model's performance on the remaining part. This process is repeated ten times, with each of the ten parts used once for evaluation. By using this rigorous approach, we were able to select the optimal hyperparameters for our model, which ultimately led to higher accuracy and better performance on the test data. After the best set of hyperparameters was selected, a final model was trained on the 80% training data. Then, the model performance was evaluated on the predictions made on the 20% unseen test data. Finally, we conducted an XAI analysis to better understand the decision-making process of our final predictive model. We generated tree SHapley Additive exPlanations (SHAP) explanations on the unseen test data, which ensured that our findings are explainable for the data that the model has not seen. The entire study process is depicted in the diagram in Figure 3 below.

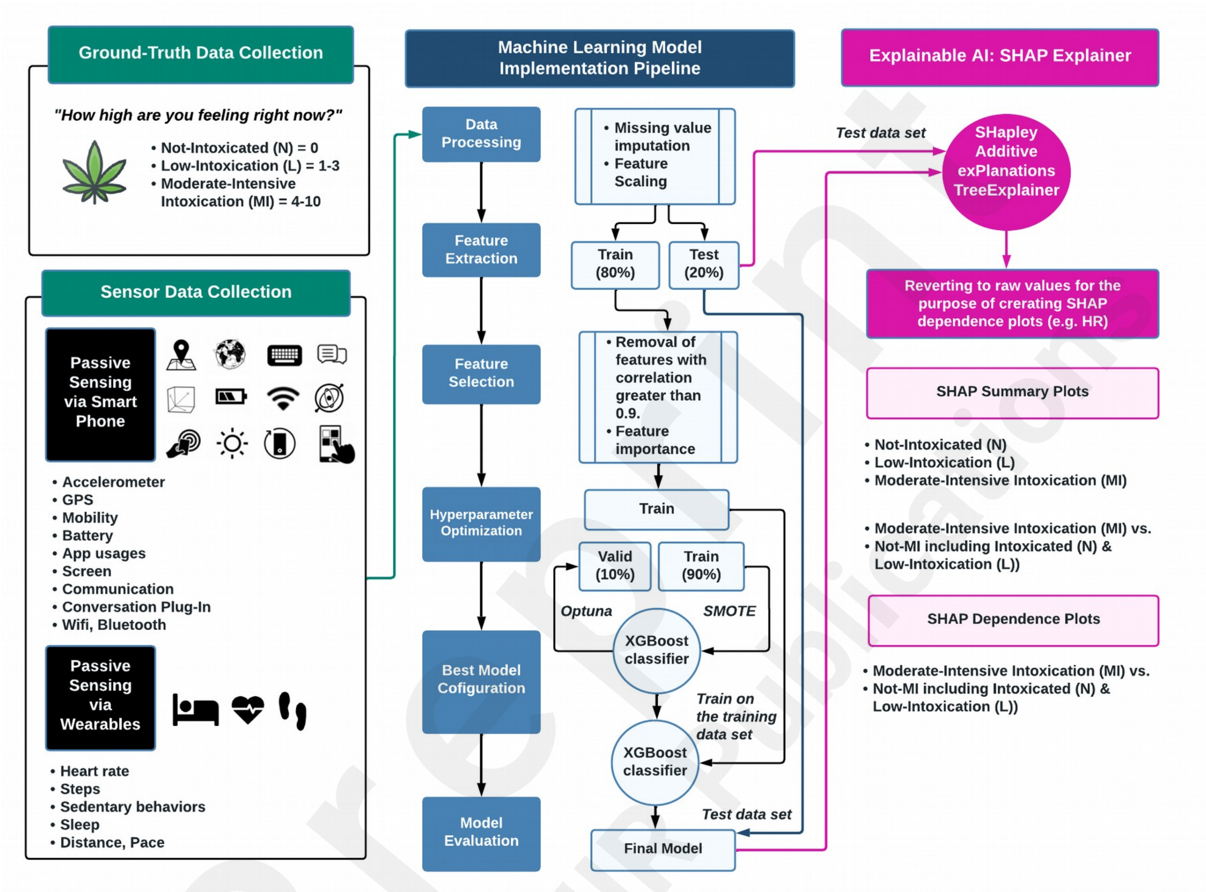


Figure 3. Study overview

## 2.9 Model Evaluation Metrics

We focus on three key metrics to evaluate model performance: $F_1$-score, recall and precision. To determine the best performing model, we select the model that maximizes the $F_1$-score. The $F_1$-score represents a balance between precision and recall [37], and for our desired use case, we need to maximize both precision and recall. Low precision means that we will have too many false positives (i.e., detecting marijuana intoxication when there is none) where we would mistakenly intervene or notify the participant. Too many false positives could erode trust in such an automated system. Low recall is also an issue, where we have too many false negatives (i.e., not detecting marijuana intoxication when the participant is intoxicated). An automated intervention system would not intervene when it should have and could result in participants unknowingly engaging in unsafe activities such as impaired driving under the influence of marijuana. Therefore, we focus on the $F_1$-score to evaluate our models, but also examine the resulting precision and recall. Specifically, given our imbalanced samples, we opted to use the AUC metric, which involves plotting the true positive rate (sensitivity or recall) against the false positive rate, and it offers a comprehensive performance assessment encompassing all conceivable classification thresholds. The AUC's resilience to class imbalance ensures a more comprehensive performance evaluation, providing a well-rounded perspective.

### 2.10 Explainable AI (XAI): Interpretation Approaches for Black-Box Machine Learning Models

To enhance the explainability of an algorithmic decision-making process, we utilized SHAP (SHapley Additive exPlanations), a widely used and influential interpretability method for machine learning models [53, 54]. Using SHAP to make predictions with explanations, our results provided valuable insights into the model's decision-making process. As such, we implemented a machine learning model to identify the top 30 most significant features associated with reports of marijuana intoxication. For these features, we generated feature importance, and SHAP summary graphs to better understand their contribution to the model's results (See Section 3.4). We chose to employ Gradient Boosted Trees due to their widespread use as state-of-the-art models. The utilization of tree SHAP in this context provides added advantages by diminishing the computation time for SHAP values from exponential to polynomial [53].

## 3. Results

### 3.1 Timing, Duration and Rating of Subjective Marijuana Intoxication

Over the 30-day study window, participants had an average of 14 days of active study participation and a median of 13 days; 129 ESM self-initiated reports of marijuana use meeting criteria for inclusion in the analysis were collected: 101 reports of subjective marijuana intoxication (“feeling high” 1-10 out of 10), 28 reports of feeling “not high” (0). We assign high = 0 to events that do not involve the use of marijuana.

Figures 4a and 4b show the distribution of self-reported subjective marijuana intoxication across participants. Most (n=75) episodes of subjective marijuana intoxication lasted between 30 min and 3 hours, with 54 episodes having a reported duration up to 30 minutes (Fig. 4a). Marijuana use was most often reported between 10 pm and 11 pm (n=24). Figure 3b shows the distribution of ESM responses throughout the day. The average response latency to an ESM prompt was 55 minutes (SD = 48 minutes), excluding cases when the scheduled prompt expired. Most self-initiated reports of marijuana use occurred in the evenings: 14.0% between 6-9pm, and 38.8% between and 9pm-midnight. On average, young adults rated their feeling of being high at 3.63 out of 10 (SD=2.72) when using marijuana (Fig. 4c).

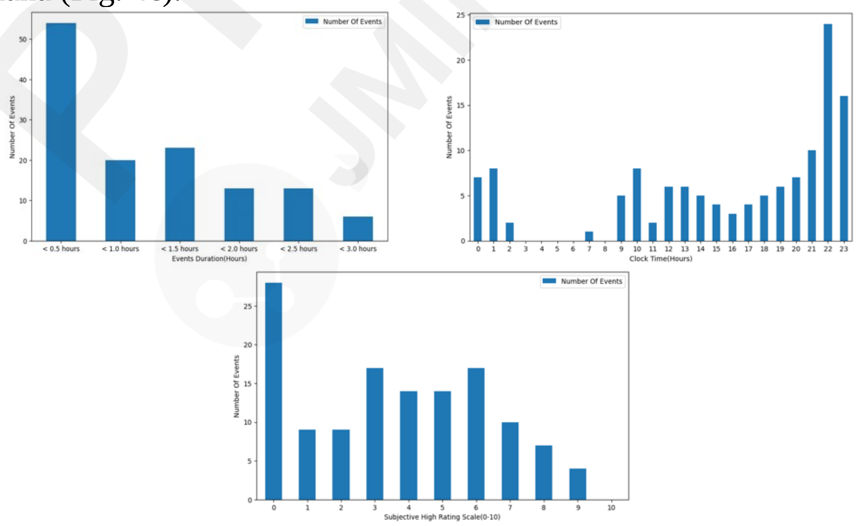


Figure 4. (a) Distribution of the duration of self-reported marijuana use episodes (n=129) across participants (left): x-axis refers to the window of smoking episodes. From left (30-minute) to right (3-hours). (b) Distribution of the start time of marijuana use episodes during the day (n=129) (middle) (c) Distribution of self-reported “feeling high” during marijuana use (x-axis= 0-10 scale representing an intensity of feeling high, 10= a lot) from the self-initiated reports of marijuana use (left). In our study, a value of 0 for the high report is labeled as “no-intoxication”

### 3.2 Model Comparison: Mobile Only, Fitbit Only and Mobile and Fitbit Integration

The objectives of the first part of our analysis are 1) to explore whether smartphone-sensor features only can be used for real-time detection to identify behaviors during reports of subjective marijuana intoxication (“feeling high”), and 2) to explore whether Fitbit data can "add value" to the model performance to detect subjective marijuana intoxication based on smartphone-sensor features to justify the added burden of Fitbit data collection. In order to explore whether the Fitbit device can "add value" to modeling marijuana intoxication behaviors, we conducted experiments to compare three machine learning models with the eXtreme Gradient Boosting (XGBoost) classifier 1) Smartphone-sensors only (XGBoost-Mobile), 2) Fitbit features only (XGBoost-Fitbit), 3) Smartphone and Fitbit features combined model (XGBoost-MobiFit), and present our best model.

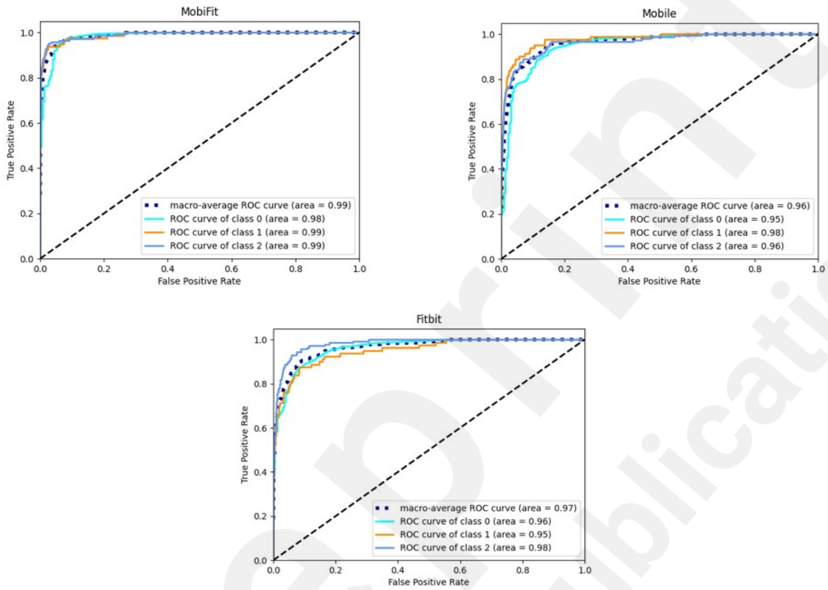


Figure 5. Model comparison to detect acute marijuana intoxication “low-intoxicated” (rating = 1–3) versus “moderate-intensive intoxicated” (rating = 4–10) versus “not-intoxicated” (rating=0). XGBoost-MobiFit: phone sensors and Fitbit (AUC = 0.99; Accuracy = 0.99) (left), XGBoost-Mobile: smartphone-based sensors (samples overlapping with Fitbit) (AUC = 0.96; Accuracy = 0.97) (center) and XGBoost-Fitbit: Fitbit only (AUC = 0.97; Accuracy = 0.98) (right)

Of the 3 models tested, the XGBoost-MobiFit model integrating the smartphone sensor with Fitbit data had the best performance, with 99% accuracy, 92% precision, 79% recall, 85% F1-score, and 99% AUC on the test dataset (See Fig. 5). These metrics show the XGBoost-MobiFit model's ability to accurately identify subjective moderate-intensive intoxication vs. low-intoxication vs. not-intoxicated. On the other hand, the Fitbit model (XGBoost-Fitbit) performed reasonably well, but not as well as the XGBoost-MobiFit model in detecting marijuana intoxication. XGBoost-Fitbit achieved an accuracy of 98%, 79% precision, 70% recall, 74% F1-score, and 97% AUC. These results suggest that utilizing only Fitbit data might not be as accurate in detecting subjective marijuana intoxication compared to the model integrated with smartphone sensor data. Based on these results, it can be concluded that the extra burden of wearing and charging the Fitbit wearable is likely justified in future deployments. The combined model (XGBoost-MobiFit), which utilizes both smartphone and Fitbit data, demonstrates improved performance in detecting subjective marijuana intoxication compared to using smartphone or Fitbit data alone.

Table 1. Comparison of three XGBoost Models using features selected in detecting moderate-intensive marijuana intoxication, low-intoxication, and not-intoxicated classes on the test dataset

| Machine Learning Model | AUC | F1-score | Recall | Precision | Accuracy |
|---|---|---|---|---|---|
| XGBoost-MobiFit | *0.99* | *0.85* | *0.79* | *0.92* | *0.99* |
| XGBoost-Mobile | 0.96 | 0.72 | 0.75 | 0.70 | 0.97 |

| | | | | | |
|---|---|---|---|---|---|
| XGBoost-Fitbit | 0.97 | 0.74 | 0.70 | 0.79 | 0.98 |

When Fitbit data are combined with Mobile data, we observed a significant improvement over the Fitbit-only model. The mobile only model achieved an AUC of 96%, F1-score of 72%, and recall of 75%, and precision of 70%. These results suggest that the inclusion of Fitbit data adds value beyond the utilization of smartphone-based sensor data alone (13% of improved F1-score). In summary, we suggest three key findings: The XGBoost-Mobile model had the lowest performance (F1-score =0.72, recall=0.75, precision=0.70); the XGBoost-Fitbit model (F1-score =0.74, recall=0.70, precision=0.79) also generally had lower performance than the combined model; and the XGBoost-MobiFit was the best performing model: F1-score (0.85), recall (0.79), and precision (0.92). In an earlier section, we highlighted the need for high precision and recall, and thus focused on the F1-score for identifying the best performing model.

### 3.3 Understanding Model Performance in Detecting the Risk State of "Moderate and Intensive Marijuana Intoxication"

To understand the predictability of the risk state of "moderate-intensive intoxication", our findings show that the MobiFit model using sensor features integrated based on mobile and Fitbit devices, outperformed both the mobile and the Fitbit only models in predicting specifically "moderate-intensive intoxication", with a substantial improvement of 20% and 18% in F1-score, respectively (Table 2). These results highlight the benefits of integrating two different devices, enhanced precision and recall for the moderate-intensive intoxication (MI) class compared to the not-intoxicated (N) and low-intoxicated (L) classes (Table 3).

Table 2. Performance comparison of three XGBoost models in detecting the subjective sense of moderate-Intensive marijuana intoxication (MI) class.

| ML Model | MI Precision | MI Recall | MI $F_1$–score | MI AUC |
|---|---|---|---|---|
| XGBoost-MobiFit | *0.89* | *0.76* | *0.82* | *0.99* |
| XGBoost-Mobile | 0.64 | 0.61 | 0.62 | 0.96 |
| XGBoost-Fitbit | 0.65 | 0.63 | 0.64 | 0.98 |

Table 3. Confusion matrix for (a) XGBoost-Mobifit (top), XGBoost-Mobile (middle), and XGBoost-Fitbit (bottom) model; for three classes not-intoxicated (N), low-intoxication (L), and moderate-intensive intoxication (MI) classes

| **XGBoost-MobiFit** | | ***Predicted*** | | |
|---|---|---|---|---|
| | | N | L | MI |
| ***Actual*** | N | *6541* | 7 | 13 |
| | L | 29 | *50* | 1 |
| | MI | 35 | 0 | *108* |

| **XGBoost-Mobile** | | ***Predicted*** | | |
|---|---|---|---|---|
| | | N | L | MI |
| ***Actual*** | N | *6452* | 59 | 50 |
| | L | 28 | *52* | 0 |
| | MI | 56 | 0 | *87* |

| **XGBoost-Fitbit** | | ***Predicted*** | | |
|---|---|---|---|---|
| | | N | L | MI |
| ***Actual*** | N | *6499* | 14 | 48 |
| | L | 41 | *39* | 0 |
| | MI | 52 | 1 | *90* |

The XGBoost-Mobile model exhibited a notably elevated false negative rate in the classification of instances labeled as “not-intoxicated”, often misclassifying them as “moderate-intensive intoxicated”. However, it exhibited a better discernment between “low-intoxicated” instances. In contrast, the XGBoost Mobifit model demonstrated a heightened true positive rate (TPR) in comparison to the other two machine learning models. This indicates that the XGBoost-Mobifit model accurately identified a significant proportion of moderate-intensive intoxication samples (the positive class) among the total samples belonging to that class. While the XGBoost-Mobile and Fitbit models achieved recall rates of 61% and 63%, respectively, in predicting individuals with moderate-intensive intoxication, it is noteworthy that they incorrectly predicted 56 and 53 out of 143 actual moderate-intensive intoxication samples as other classes, respectively. In contrast, the best-performing MobiFit model achieved 108 true positives out of the 143 actual moderate-intensive intoxication samples. This high precision suggests its strong performance. Despite this, it did miss 35 samples (as shown in Table 3).

### 3.4 Key Features Contributing to the Model Performance

To specifically explore the algorithms’ prediction of the “risk” state of “moderate-intensive intoxication”, we utilized SHAP (Shapley Additive exPlanations) summary visualizations [53, 54] to discern acute marijuana intoxication patterns. Through this analysis, we identified the main features that significantly contributed to the predictability of the machine learning model gauged by mean absolute SHAP values across all instances as well as specifically targeting “moderate-intensive intoxication” classes. In this analysis, eXplainable Artificial Intelligence (XAI) categorized the labels into two distinct groups: instances characterized by “moderate-intensive intoxication” and those in “other classes” (encompassing “low-intoxication” and “not-intoxicated”).

When interpreting the SHAP visualizations (Figs. 6 and 7), the length of each bar (graph on the left) reflects the extent of the corresponding feature's contribution. Longer bars indicate a stronger influence on prediction of the outcome. On the other hand, shorter bars have minimal impact on prediction. The SHAP summary plots (e.g., graph on the right), illustrate how features influence the moderate-intensive intoxication prediction class. These plots arrange features with the strongest influence at the top. The graph visually demonstrates how features impact the MI prediction across various values. The color shading indicates the direction in which the feature affects the prediction, such that blue refers to low values, purple corresponds to median values, and red indicates high values. Plots extending to the left make a negative contribution to the prediction, whereas plots extending to the right have a positive contribution to moderate-intensive intoxication prediction.

#### 3.4.1 Impact of Average Key Features on Model Output Magnitude

*The top five influential features in detecting the three classifications (Fig. 6, left) and affecting the moderate-intensive intoxication outputs (Fig. 6, right) encompassed the impact of time of day, radius of gyration, minimum heart rate, day of the week, and minutes awake during sleep.* Among physical activities and physiological signals, a diverse range of features extracted from various sensors, including those beyond time-based attributes from both mobile and Fitbit combined sensors, were chosen as the top 30 crucial elements for distinguishing the three distinct classes: not-intoxicated (N), low-intoxication (L), and moderate-intensive intoxication (MI). The SHAP value, signifying the average impact magnitude on the model's output, played a pivotal role in establishing this determination (Fig. 5, left).

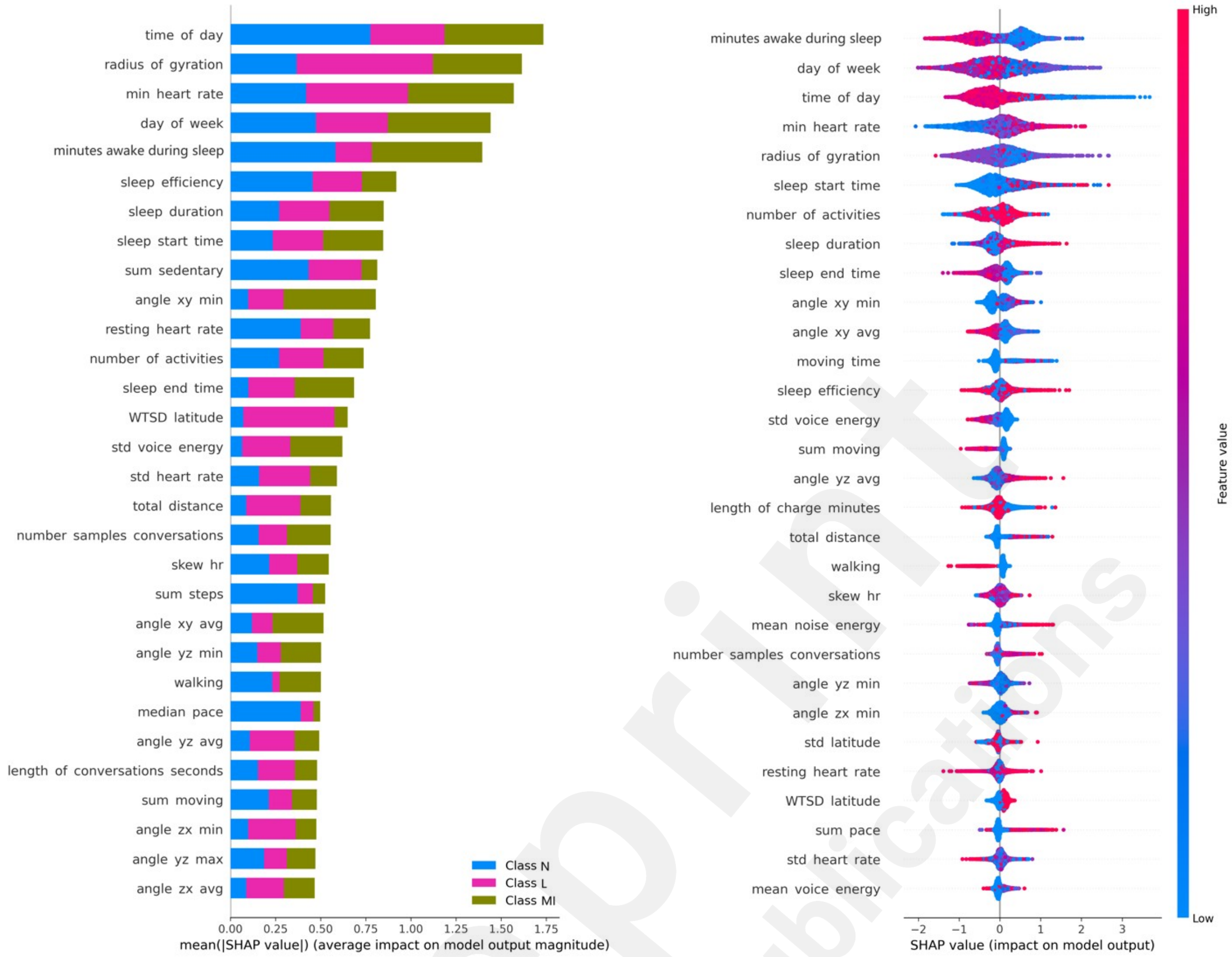


Figure 6. Explanations generated by SHAP summary plot. Impact of features on best performing XGBoost-MobiFit model (left) and binary model output identifying moderate-intensive intoxication (MI) (SHAP > 0) from non-moderate-intensive intoxication (N and L) classes (SHAP < 0) (right)

### 3.4.2 Impact of Unique Key Features on the Mobile and Fitbit Model Outputs

Similar to the MobiFit model (our best model), the Mobile model also highlighted key features that demonstrated overlapping impacts on the outcomes of the model. The only exception lies in the impact of specific movement and environmental context features, such as the number of Bluetooth samples, moving time, Wifi average, percent total noise, maximum magnitude of acceleration, and standard deviation of latitude (Fig. 6, top left and right). On the other hand, the Fitbit model exhibited an improved impact of four heart rate features, all of which ranked within the top 10 for all three classes (Fig. 6, bottom left) as well as for the MI classes compared to the non-MI classes (Fig. 7, bottom right).

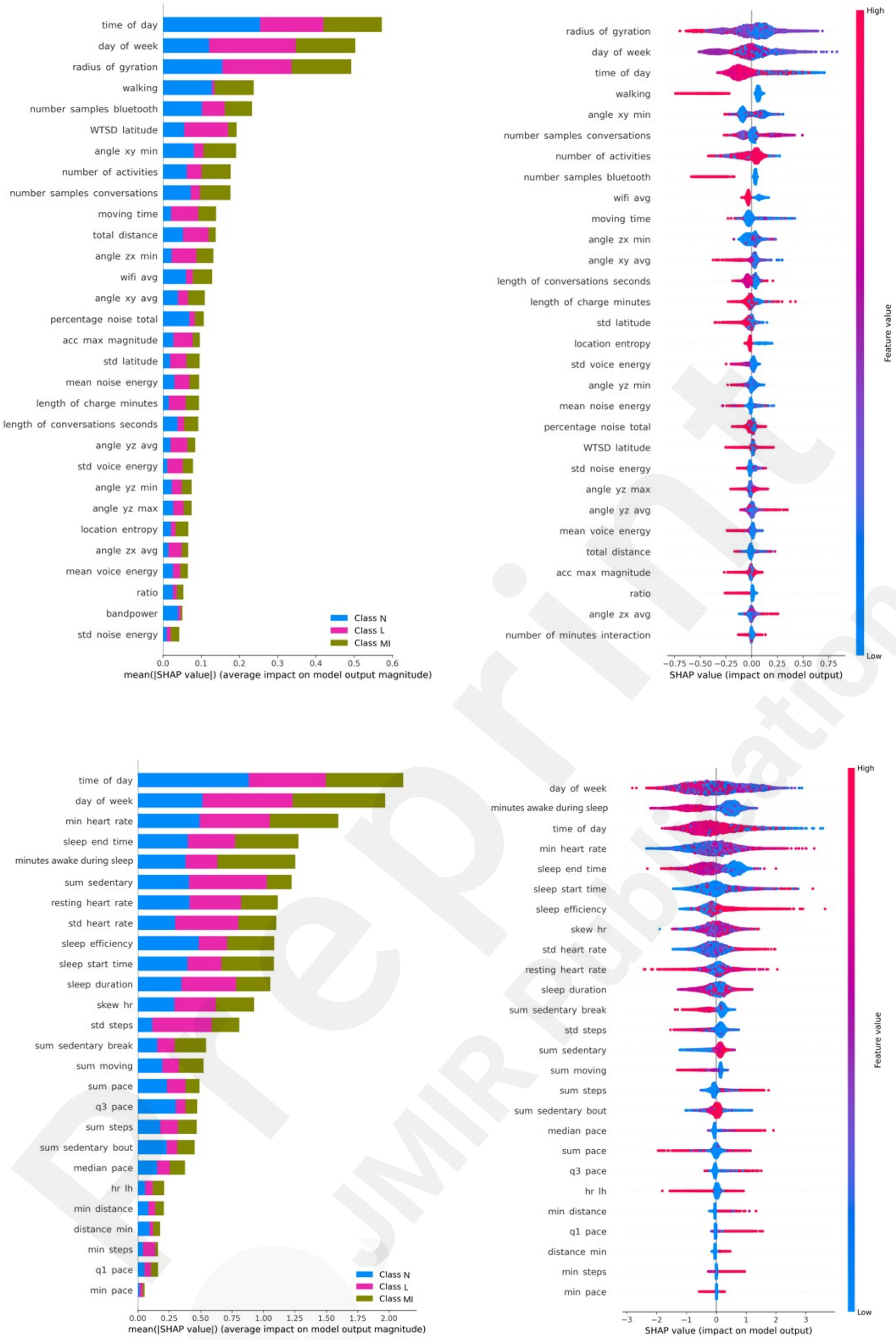


Figure 7. Explanations generated by SHAP summary plot. Impact of features on XGBoost-Mobile model (top left) and binary model output identifying moderate-intensive intoxication (MI) (SHAP > 0) from non-moderate-intensive intoxication (N and L) classes (SHAP < 0) (top right), impact of features on XGBoost-Fitbit model (bottom left) and binary model output identifying moderate-intensive intoxication (MI) (SHAP > 0) from non-moderate-intensive intoxication (N and L) classes (SHAP < 0) (bottom right)

### 3.5 Key Features Explaining Moderate-Intensive Intoxication

To specifically examine the influence of features on the “risk” state of moderate-intensive marijuana intoxication, we present comprehensive details regarding the prediction of each individual key feature within the model.

A partial dependence plot (PDP) (Fig. 8) provides information on the overall connection between a feature and the predicted outcome. The vertical axis represents SHAP values, signifying the effect of the chosen feature on predictions. The horizontal axis represents the real feature values across instances. Each plot point represents an instance's feature value and its corresponding SHAP value. An upward (rising) PDP slope indicates a positive impact of the feature on MI prediction, whereas a downward (decreasing) slope indicates a negative impact. To understand interactions between two features (min heart rate and sum of moving minutes presented in Fig. 8, top left), the surface on the PDP plot illustrates the combined impact of the two features on MI predictions. Greater values correspond to increased prediction values.

### 3.5.1 Elevated and Fluctuating Heart Rates

In the partial dependence plot, the SHAP values of minimum heart rates were significantly elevated from approximately 80 bpm (on average), peaking at 90 bpm and reaching up to 100 bpm (ranging from 60 bpm to 120 bpm, with a few data points exceeding 120 bpm), indicating moderate-intensive self-reported marijuana intoxication (SHAP value > 0) in young adults compared to other (not- and low-intoxicated) classes. The SHAP values clearly demonstrate *a positive increase in minimum heart rate associated with a higher likelihood of self-reported moderate to intensive marijuana intoxication*, irrespective of the impact of sum of moving minutes. The total moving minutes during self-report of moderate-intensive intoxication had an impact on elevations in the minimum heart rate, as shown in Fig. 8 (top left), where values in red refer to moving for max 5 minutes (our analysis employs 5-minute units). While heart rate can fluctuate due to various factors, as indicated by previous studies, including physical movements, consumption of substances like alcohol, caffeine, meal intake, and mental status (e.g., stress, anxiety), further investigation would be needed to explore impact of other factors.

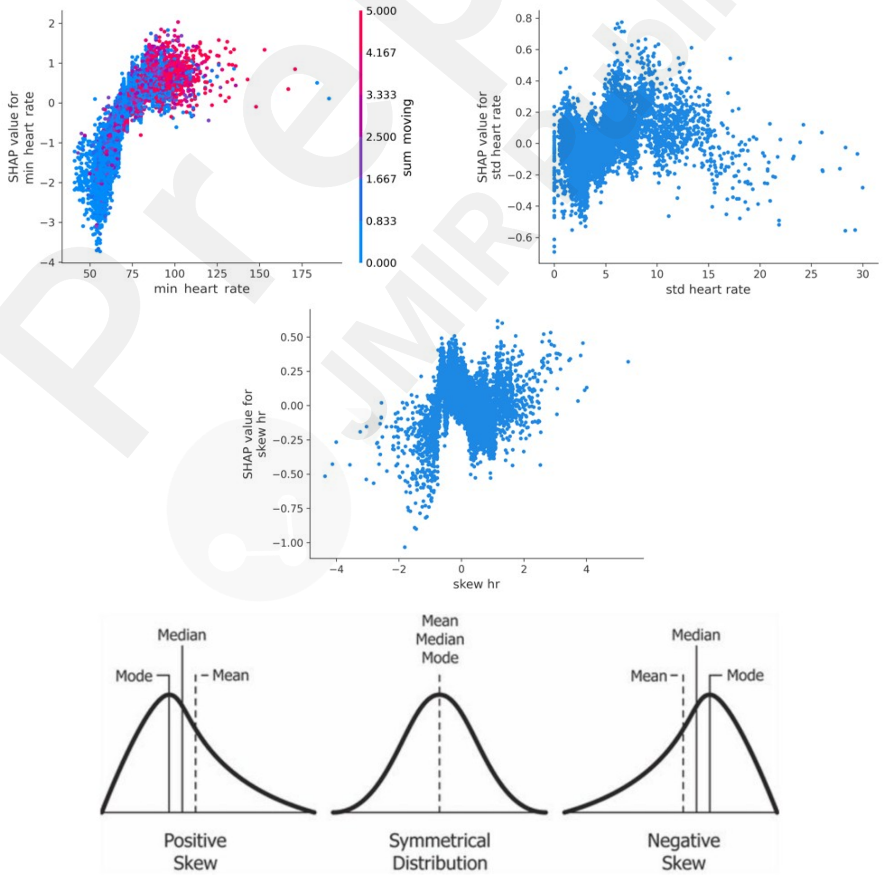


Figure 8. Interaction effects of total moving minutes on minimum heart rate values (top left), standard deviation (top middle), and skewness (top right) of heart rates, and an explanation of skewness [65] (bottom)

The data patterns for the standard deviation of heart rates exhibited fluctuations, but, in general, showed an increase when young adults reported moderate-intensive intoxication (Fig. 8, top middle). Negative skewness ("left-skewed" or "left-tailed") in heart rates was consistently linked with moderate-intensive intoxication. This indicates that there were more heart rate data points on the right side of the mean (referring to that the median is greater than the mean), resulting in a distribution stretched towards higher heart rate values (Fig. 8, top right).

### 3.5.2 Decreased Large-Scale Movements and Shifts in Micro-Movement Patterns

During states of moderate to intensive intoxication, individuals exhibited *a tendency to manifest relatively restricted large-scale movement***,** often limited to a radius of approximately 5 km. Notably, instances with radius of gyration data exceeding approximately 10 km were not associated with moderate to intensive intoxication. This observation implies that when in a state of self-reported moderate to intensive intoxication (rated 4-10), young adults exhibited a decreased inclination for extensive travel (Fig. 9, top left). Nonetheless, they still demonstrated movement within a radius of 5 km.

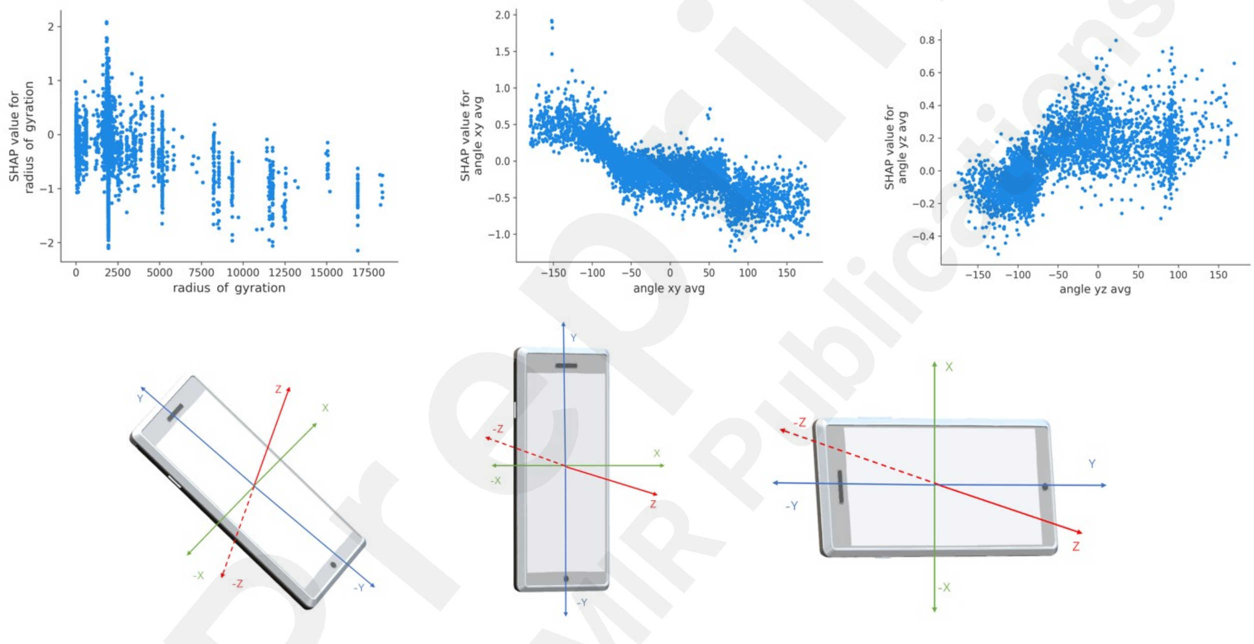


Figure 9. Influence of radius of gyration (unit: meters) (top left), average angles of XY (top middle) and YZ (top right), and reference: smartphone angles extracted from accelerometer sensors (bottom)

We presented the alterations in phone angles, as recorded by accelerometers, during episodes of moderate-intensive intoxication. Notably, the decrease in average XY-axis angles and the concurrent increase in YZ-axis angles exhibited a correlation with moderate-intensive intoxication levels (Fig. 9, top right). Interpreting the positive or negative values (-150 or +150) concerning how individuals who report being intoxicated utilize their smartphones may not be entirely straightforward. However, as "the axis directions and device side names remain consistent, regardless of the device's orientation" (Fig. 9, bottom right), the findings might indicate the occurrence of recurrent bodily movements when the phone is worn or hand movements when the phone is held. These movements may serve to *identify imbalanced states in which young adults could potentially transiently have less control of their body movements (i.e., less physical coordination)* when experiencing moderate to intensive intoxication (SHAP value > 0) under the influence of marijuana compared to their body movements when not- and low-intoxicated (SHAP value < 0).

### 3.5.3 Elevated Surrounding Noise Energy

Interestingly, while the variance of audio noise energy increased (with data points deviating further from the mean), *the mean noise energy demonstrated a decrease, yet it overall exhibited an upward trend* (Fig. 10 left). During phone conversations, young adults who reported moderate-intensive intoxication displayed discernible voice activity. However, their *speaking voice energy via our conversation plug-in exhibited no significant changes* that could serve as indicators of marijuana intoxication (Conversely, the number of conversation samples holds significance in the mobile model, as presented in Fig. 7, top). Intriguingly, instances involving individuals reporting moderate-intensive marijuana intoxication revealed *an increase in vocal variability, coupled with a subsequent reduction* (Fig. 10, middle).

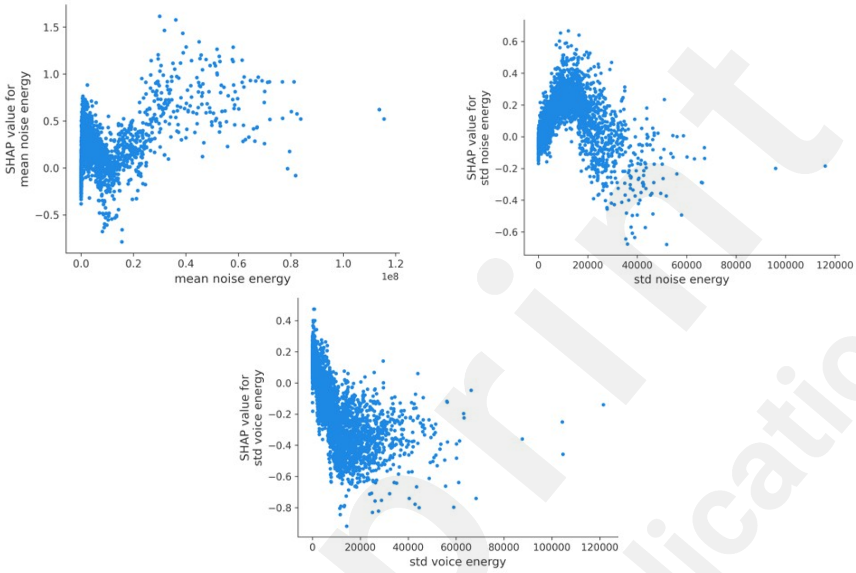


Figure 10. Influence of mean (left) and standard deviation (middle) noise energy, and standard deviation voice energy (right) (unit: Joule)

The analysis of surrounding sounds can provide valuable insights into the specific environments where individuals reporting moderate-intensive marijuana intoxication might be situated. *This could encompass moments of marijuana smoking, socializing with friends, or engaging with media such as television or music.* It's important to note that while GPS-generated features were the primary indicators, self-reported moderate-intensive marijuana intoxication might or might not directly link to places like shared social spaces (e.g., lounge), bars, pubs or clubs. Nevertheless, *it remains plausible that young adults who report moderate-intensive marijuana intoxication may choose to stay in noisy surroundings.*

### 3.5.4 Prolonged Sleep Patterns

We identified distinct sleep patterns linked to episodes of self-reported moderate-intensive intoxication. Notably, individuals who reported moderate-intensive marijuana intoxication (rated 4-10 out of 10) demonstrated *extended sleep durations, spanning approximately 8 to 11 hours* (Fig. 11, left) the day before self-reported intoxication. In contrast, instances with low or no reported intoxication tended to correspond to the healthy range of sleep durations, typically averaging around 6-7 hours of sleep, although some sleep patterns were as short as 2 hours.

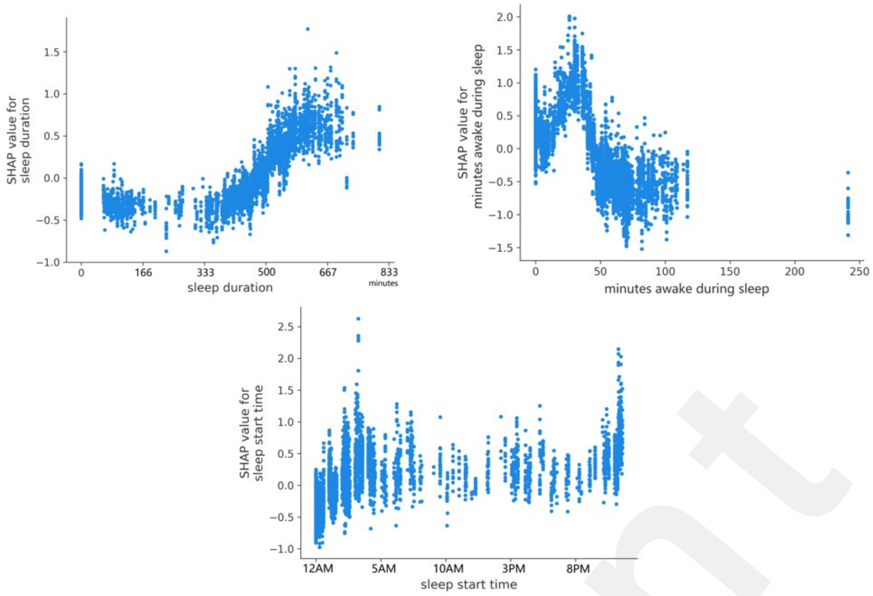


Figure 11. Total sleep duration (left), minutes awake during sleep (center), and sleep start time (right)

Interestingly, there was a *positive correlation between the duration of minutes awake after falling asleep and self-reported moderate-intensive marijuana intoxication, particularly within a specific timeframe (sleep min awake < 50 minutes).* However, an increase in extended minutes awake after falling asleep (if > 50 minutes, extending beyond approximately an hour) did not show any significant association with a likelihood of moderate-intensive marijuana intoxication (Fig. 11, center). Regarding sleep start times, the data indicated peaks at both 11 pm and early morning hours, with a rise in sleep start times continuing until around 4 am (Fig. 11, right).

To summarize our findings, elevated minimum heart rate values were clearly linked to a higher likelihood of self-reported moderate-intensive marijuana intoxication. However, we observed that travel patterns did not appear to increase. Phones angles displayed a likelihood of decrease or increase contingent on specific angles and young adults were positioned in discernibly noisy environments, but their speaking voice did not show any significant changes associated with moderate-intensive marijuana intoxication. Interestingly, extended sleep hours and minutes awake during sleep [64] the day before self-reported marijuana intoxication were associated with self-reported moderate-intensive marijuana intoxication.

### 3.6 Additional Analyses for Real-World Feasibility

To enhance the practicality of our machine learning model in real-world settings, we conducted supplementary analyses to evaluate our top-performing model, the XGBoost-MobiFit model, under different scenarios. These scenarios involved: (1) excluding location data, as some individuals might have privacy concerns about GPS collection or might have deactivated it during the study despite giving consent to researchers; (2) excluding sleep data in cases where users might not provide sleep information; and (3) excluding both location and sleep data. This approach aims to investigate the feasibility of offering more adaptable data collection options, potentially addressing privacy concerns.

In brief, the performance of the model excluding location features (XGBoost-MobiFit-GPS excluded) decreased 15% of F1-score compared to the best model. The sensitivity (recall) decreased 10%. Regarding exclusion of sleep data, analysis revealed a 24% (XGBoost-MobiFit-Sleep excluded) decrease in the F1-score compared to the performance of the best model. Upon excluding GPS and sleep features, the model (XGBoost-MobiFit-GPS-Sleep excluded) experienced a 16% reduction in F1-score (Table 7) and exhibited the lowest recall (Table 8) in identifying self-reported moderate-intensive marijuana intoxication classes, when compared to the best-performing model. Please refer to the Supplemental material (Appendix 4) for a detailed description of these additional analyses and results.

## 4. Discussion

### *Overview*

The ability to detect subjective report of acute marijuana intoxication (“feeling high”) in the natural environment using mobile sensors has the potential to enable Just-In-Time interventions [47] to reduce marijuana-related harms. To the best of our knowledge, this is the first study that demonstrates the impact of integrating smartphone-based and wearable sensor features on the enhancement of the predictability and interpretability of algorithms in detecting acute marijuana intoxication in naturalistic environments. As hypothesized, first, we found that the *XGB-Mobifit, smartphone sensor-based and Fitbit combined features with the eXtreme Gradient Boost Machine (XGBoost) outperformed (F1-score: 0.85)* the mobile and Fitbit only machine learning models. Combining Fitbit and smartphone data (XGBoost-MobiFit) significantly enhanced model performance by 13% compared to the smartphone-sensor only (XGBoost-Mobile) and by 11% compared to the Fitbit sensor only (XGBoost-Fitbit) model. Furthermore, XAI visualizations highlighted the significance of key sensor features, including elevated heart rates (ranked 3rd) as hypothesized as an indicator, reduced large-scale movements (ranked 2nd), alterations in micro-movement patterns, increased ambient noise energy, and disrupted sleep patterns (ranked in Top 30). These findings were observed beyond the influences of time of day and day of the week features (ranked 1st and 4th, respectively), as validated in [55], particularly during instances of self- reported subjective marijuana intoxication in naturalistic environments. Our findings demonstrate the promise that mobile phone sensors, tested for subjective cannabis intoxication in young adults [55] and alcohol intoxication in young adults [2, 3, 51], combined with wearable sensors, hold for automated, explainable and unobtrusive detection of acute subjective marijuana intoxication in the natural environment.

### 4.1 Interpretable Physiological Biomarkers of Marijuana Intoxication in Real-World Settings

To explain the results of the black-box machine learning models to detect marijuana intoxication in “everyday settings”, our study, which integrated sensors from smartphones and wearable devices, identified key sensor features and used XAI to facilitate interpretation of model results. *As our results are consistent with prior research conducted in controlled laboratory settings that consistently found an acute increase in resting heart rate following marijuana use [19, 27, 50], using heart rate - minimum heart rate within a 5-minute analysis window in our study - as an objective biomarker could be a valuable tool for detecting marijuana intoxication “outside of laboratory settings”*. Although it is not as clear as the literature identified resting heart rate, minimum resting heart rate plays a key role in contributing to the detection of self-reported marijuana intoxication. Other physiological signals, such as respiration, can be incorporated with heart rate to better capture marijuana intoxication [63]. While many factors can affect heart rate, our study yielded significant heart rate features and insights from the elevated heart rate patterns during self-reported acute marijuana intoxication. Subsequent research could delve into the associations between heart rate and additional physiological and behavioral indicators of marijuana use.

The use of explainable artificial intelligence (XAI) visualization could help increase transparency and accountability when conducted as part of a substance use detection system [51]. It is promising to use XAI because it can enable researchers and clinicians to learn about how the algorithms arrived at the decisions and identified key attributes, providing an opportunity to improve the accuracy of detection and reliability and increase trust over time. Ultimately, it is up to each individual to weigh the potential benefits of a detection and intervention system against privacy concerns and personal values.

### 4.2 Enhancing Participant Engagement for Improved Feasibility in Data Collection Deployment: Strategies to Ensure Privacy Preservation

To elucidate the advantages of employing combined sensor features from two devices while addressing potential privacy concerns, particularly related to location data, we aim to offer participants additional configuration choices rather than resorting to study withdrawal for GPS sensor deactivation. This is demonstrated by our testing of the best-performing model, XGBoost-MobiFit, wherein we excluded location features. The analysis revealed a 15% (XGBoost-MobiFit-GPS excluded) decrease in the F1-score compared to the performance of the best model. As proposed by Bae et al. [51], collecting GPS data and utilizing rounded GPS data extraction (i.e., less precise location data) could be a viable approach. This avoids the use of raw latitude and longitude, which may contain sensitive information about specific locations. Researchers and clinicians could consider providing alternative options instead of completely disabling GPS, as it contributes to the accuracy of the model.

Moreover, in order to assess the efficacy of our top-performing model in the context of real-time detection when young adults are consuming marijuana, we conducted tests after excluding sleep-related features (see Appendix 4 in the supplementary material). The analysis revealed a 24% (XGBoost-MobiFit-Sleep excluded) decrease in the F1-score compared to the performance of the best model. This scenario might be applicable when individuals remove the wristband during sleep or opt to take off the device due to discomfort while sleeping. Alternatively, the situation could arise if individuals forget to reattach the device after taking a shower at night. There is a trade-off between model performance and a privacy-preserving approach. While participants may find benefits in having the option to disable sensors when necessary, it is important to note that this could potentially lead to a decrease in the predictability of acute marijuana intoxication.

We believe that by building a system that prioritizes privacy and user autonomy, we can provide a valuable tool to reduce marijuana-related harm to both individuals and society as a whole. In the future, when the model is paired with an intervention, the intervention could be isolated to the one's own devices, but there are instances when sharing information (with the person's consent) about marijuana use episodes with clinicians or one's social support network could be valuable. This is all to say that while the technical aspects of our detection system can be applied in a way to minimize invasiveness from a privacy perspective, each user will have to decide for themselves whether the value that a detection and intervention system provides is worth the tradeoff to minimize marijuana-related harms to self and the broader community.

## 5. Limitations and Future Work

We now describe some limitations of our work. First, we used subjective self-reports reports as ground-truth for training our machine learning model to identify marijuana intoxication. This study extends prior ESM work which codes self-reported marijuana use as yes or no [43] by asking participants to rate marijuana intoxication from 0-10, which may be subject to recall or other biases in reporting. The classification scheme of low-intoxication (1-3), moderate-intensive marijuana intoxication (4-10), and not-high (0) could potentially lead to classification errors. The broad categorization might overlook nuanced differences within these categories, which could affect the accuracy of the classifiers. We plan future analyses to examine the performance of mobile and wearable sensors (smartphone-based and Fitbit) against different thresholds for a subjective marijuana intoxication outcome. Another limitation was the level of compliance (63%) in completing the morning, afternoon, and evening surveys, which could be improved, since no reminders were provided to facilitate completion. We do not know if all episodes of marijuana use were reported by participants, which could limit model performance. As discussed, there is not yet a real-time biological test that could be used as the gold standard against which to validate self-reported subjective marijuana intoxication or marijuana-related impairment. We collected up to 30 days of daily data, in 33 young adults, which provided a starting point for the analyses, but results warrant replication in a larger sample, and over longer periods of daily data collection. Our findings may not be generalizable beyond young adults, recruited from the community, who were not seeking treatment, living in an area where non-medical use of marijuana is illegal. Finally, our model performed the best when tested on the same people that the model was trained on (there was no overlap between training and testing data). While there is a valid use case for this, it assumes that we can always collect labeled training data for participants for whom we would like to apply the model. With a larger test population, exploration of more sophisticated (possibly multi-sensor) features, and improved tuning of models, we hope that our future refined models will generalize to new participants from whom no training data would be required. To obtain generalizability, continued development of the model to establish norms in a larger sample is needed. At the same time, our model needs to be improved to be applied to unseen new participants. When examining heart rate, it is important to also track activities, body movements, and environmental context (e.g., day, time, location) because an acute increase in heart rate by itself is non-specific and may not be associated with onset of smoking marijuana. False alarms triggered by the algorithm could erode trust in an automated system, whereas low sensitivity to actual marijuana use could result in marijuana-related harm. Therefore, it is important to investigate the interplay between human activities associated with marijuana intoxication and physiological signals in a larger population, and how these interactions can contribute to intervention delivery in real-world contexts.

Our models used a 5-minute window size for classifying self-reported marijuana intoxication: “moderate-intensive intoxication” vs. “low-intoxication” vs. not-intoxicated, using smartphone and Fitbit combined features. We suggest that as 5-minute windows are quite frequent in the context of a marijuana usage session (more than half of our participants’ sessions > 30 minutes), our best performing model could be used to detect self-reported marijuana intoxication in near real-time, and thus trigger interventions in near real-time. The ability to intervene in near real-time is important because participants could return to daily activities (e.g., bicycle or e-bike, drive a car) when they are no longer feeling high, but THC still remains in their body affecting their cognitive function and motor coordination [31]. Detecting marijuana episodes and intoxication can be an important first step toward intervening in a timely manner to assist people who are under the influence of marijuana.

Our best detection model is unlikely to misclassify a "high" state as not-high, which demonstrates the possibility of using our detection algorithm with unseen data in the real-world context. On the unseen test set, we obtained 85% precision (92% precision for three classes) in specifically identifying self-reported moderate-intensive marijuana intoxication. Passive sensing using smartphones-based sensors has been investigated in the context of alcohol intoxication [2, 3, 51], and here we extend this previous research to self-reported marijuana intoxication [55] beyond smartphone-based sensors, which can ultimately be useful for JIT interventions [47] having benefits of integrated mobile phone and wearable sensors to reduce marijuana-related harm. The value to society and to individuals of reducing marijuana-related harm is obvious. As a personal decision to support detection of acute marijuana use, individuals who use marijuana, for example, would keep their phone charged and with them when they are using marijuana, and wear a device (e.g., Fitbit) and keep it charged as well. Our detection model is simple enough that the collected data never needs to leave their phone, since feature extraction could be run on the phone, along with the model itself. This reduces privacy concerns somewhat.

## 6. Conclusions

Our study demonstrates that an integration of smartphone-based sensors and wearable features from Fitbit improves the detection of self-reported subjective acute marijuana intoxication in the natural environment among young adults. A combination of smartphone sensor and Fitbit features (XGBoost-MobiFit model) achieved the best $F_1$-score (0.85) balancing between sensitivity and specificity in detecting self-reported moderate-intensive marijuana intoxication compared to self-reported low-intoxication and not-intoxicated. Results suggest that Fitbit data improved XGBoost performance (additional 13% $F_1$-score) in detecting self-reported marijuana intoxication (vs not-intoxicated), over and above the smartphone sensor model only, which potentially justifies the added burden of wearing Fitbit for detection among young adults. Key smartphone and Fitbit sensor features associated with self-reported “moderate-intensive intoxication” included an extended set of statistical measures leveraging elevated minimum, standard deviation and skewness of heart rates, and changes of smartphone angles, increased average surrounding noise energy, smaller radius of gyration in the immediate environment, and prolonged sleep patterns the night before self-reported marijuana intoxication. Future work includes refining the smartphone and Fitbit sensor algorithm in larger samples and exploring the use of the algorithms generated by explainable AI to support the design of Just-In-Time interventions for clinicians to deliver context-adaptive personalized interventions to minimize potential marijuana-related harms (e.g., intoxicated driving). These harm reduction interventions could reduce the frequency and severity of acute marijuana-related harm in young adults.

## Acknowledgement

This study was supported by the National Institute on Drug Abuse (R21 DA043181/U01 DA056472), the Stevens Startup grant, and the Provost scholarship.

## References

1. Amin Ahsan Ali, Syed Monowar Hossain, Karen Hovsepian, Md Mahbubur Rahman, Kurt Plarre, and Santosh Kumar; 2012. mPuff: automated detection of cigarette smoking puffs from respiration measurements. In Proceedings of the 11th international conference on Information Processing in Sensor Networks: 269–280.
2. Sangwon Bae, Tammy Chung, Denzil Ferreira, Anind K Dey, and Brian Suffoletto; 2018. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. Addictive behaviors 83 (2018): 42–47.
3. Sangwon Bae, Denzil Ferreira, Brian Suffoletto, Juan C Puyana, Ryan Kurtz, Tammy Chung, and Anind K Dey; 2017. Detecting drinking episodes in young adults using smartphone-based sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 2 (2017): 1–36.
4. Michel Bédard, Sacha Dubois, and Bruce Weaver; 2007. The impact of cannabis on driving. Canadian Journal of Public Health 98, 1 (2007): 6–11.
5. Hilary F. Byrnes, Brenda A. Miller, Douglas J. Wiebe, Christopher N. Morrison, Lillian G. Remer, and Sarah E. Wiehe. [n.d.]. Tracking Adolescents With Global Positioning System-Enabled Cell Phones to Study Contextual Exposures and Alcohol and Marijuana Use: A Pilot Study. 57, 2 ([n. d.]): 245 – 247. https://doi.org/10.1016/j.jadohealth.2015.04.013
6. Stephanie Carreiro, Keerthi Kumar Chintha, Sloke Shrestha, Brittany Chapman, David Smelson, and Premananda Indic; 2020. 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 (2020), 107929.
7. Stephanie Carreiro, Hua Fang, Jianying Zhang, Kelley Wittbold, Shicheng Weng, Rachel Mullins, David Smelson, and Edward W Boyer; 2015. iMStrong: deployment of a biosensor system to detect cocaine use. Journal of medical systems 39, 12 (2015): 186.
8. Stephanie Carreiro, David Smelson, Megan Ranney, Keith J Horvath, Rosalind W Picard, Edwin D Boudreaux, Rashelle Hayes, and Edward W Boyer; 2015. Real-time mobile detection of drug use with wearable biosensors: a pilot study. Journal of Medical Toxicology 11, 1 (2015): 73–79.
9. Basile Chaix. 2018. Mobile Sensing in Environmental Health and Neighborhood Research. Annual Review of Public Health 39, 1 (2018):367–384. https://doi.org/10.1146/annurev-publhealth-040617-013731 arXiv:https://doi.org/10.1146/annurev-publhealth-040617-013731 PMID: 29608869.
10. Keerthi Kumar Chintha, Premananda Indic, Brittany Chapman, Edward W Boyer, and Stephanie Carreiro; 2018. Wearable biosensors to evaluate recurrent opioid toxicity after naloxone administration: a Hilbert transform approach. In Proceedings of the... Annual Hawaii International Conference on System Sciences. Annual Hawaii International Conference on System Sciences, Vol.

2018. NIH Public Access: 3247.
11. MA Cohn, AS Rao, M Broudy, S Birch, H Watson, N Atkins, B Davis, FD Stott, and MA Sackner; 1982. The respiratory inductive plethysmograph: a new non-invasive monitor of respiration. Bulletin europeen de physiopathologie respiratoire 18, 4 (1982): 643.
12. Deirdre A. Conroy, Megan E. Kurth, Kirk J. Brower, David R. Strong, and Michael D. Stein. [n.d.]. Impact of Marijuana Use on Self-Rated Cognition in Young Adult Men and Women. 24, 2 ([n. d.]): 160–165. https://doi.org/10.1111/ajad.12157
13. David H Epstein, Matthew Tyburski, William J Kowalczyk, Albert J Burgess-Hull, Karran A Phillips, Brenda L Curtis, and Kenzie L Preston; 2020. Prediction of stress and drug craving ninety minutes in the future with passively collected GPS data. NPJ Digital Medicine 3, 1 (2020): 1–12.
14. Catharine Fairbairn, Dahyeon Kang, and Nigel Bosch. [n.d.]. Using Machine Learning for Real-Time BAC Estimation from a New-Generation Transdermal Biosensor in the Laboratory. https://doi.org/10.31234/osf.io/849er
15. Catharine E. Fairbairn and Dahyeon Kang. [n.d.]. Temporal Dynamics of Transdermal Alcohol Concentration Measured via New-Generation Wrist-Worn Biosensor. 43, 10 ([n. d.]): 2060–2069. https://doi.org/10.1111/acer.14172
16. Denzil Ferreira, Vassilis Kostakos, and Anind K Dey; 2015. AWARE: mobile context instrumentation framework. Frontiers in ICT 2 (2015): 6.
17. Charlotte Goldfine, Jeffrey T Lai, Evan Lucey, Mark Newcomb, and Stephanie Carreiro; 2020. Wearable and Wireless mHealth Technologies for Substance Use Disorder. Current Addiction Reports (2020): 1–10.
18. August F Holtyn, Eugene Bosworth, Lisa A Marsch, Bethany McLeman, Andrea Meier, Elizabeth C Saunders, Emre Ertin, Md Azim Ullah, Shahin Alan Samiei, Monowar Hossain, et al; 2019. Towards detecting cocaine use using smartwatches in the NIDA clinical trials network: Design, rationale, and methodology. Contemporary clinical trials communications 15 (2019): 100392.
19. GL Huber, DL Griffith, and PM Langsjoen; 1988. The effects of marihuana on the respiratory and cardiovascular systems. Marijuana: An International Research Report. National Campaign Against Drug Abuse Monograph 7 (1988).
20. Marilyn A Huestis and Michael L Smith; 2018. Cannabinoid markers in biological fluids and tissues: revealing intake. Trends in molecular medicine 24, 2 (2018): 156–172.
21. Masudul Haider Imtiaz, Raul I Ramos-Garcia, Volkan Yusuf Senyurek, Stephen Tiffany, and Edward Sazonov; 2017. Development of a multisensory wearable system for monitoring cigarette smoking behavior in free-living conditions. Electronics 6, 4 (2017), 104.
22. Masudul H Imtiaz, Raul I Ramos-Garcia, Shashank Wattal, Stephen Tiffany, and Edward Sazonov; 2019. Wearable sensors for monitoring of cigarette smoking in free-living: A systematic review. Sensors 19, 21 (2019), 4678.
23. Bob Lansdorp, William Ramsay, Rashad Hamid, and Evan Strenk. [n.d.]. Wearable Enzymatic Alcohol Biosensor. 19, 10 ([n. d.]). https://doi.org/10. 3390/s19102380
24. Ruojun Li, Emmanuel Agu, Ganesh Balakrishnan, Debra Herman, Ana Abrantes, Michael Stein, and Jane Metrik; 2019. WeedGait: Unobtrusive Smartphone Sensing of Marijuana-Induced Gait impairment By Fusing Gait Cycle Segmentation and Neural Networks. In 2019 IEEE Healthcare Innovations and Point of Care Technologies,(HI-POCT). IEEE: 91–94.
25. Susan E. Luczak and Vijay A. Ramchandani. [n.d.]. Special issue on alcohol biosensors: Development, use, and state of the field: Summary, conclusions, and future directions. 81 ([n. d.]): 161–165. https://doi.org/10.1016/j.alcohol.2019.07.001
26. Kara L Lynch, Y Ruben Luo, Shirin Hooshfar, and Cassandra Yun; 2019. Correlation of breath and blood Δ9-tetrahydrocannabinol concentrations and release kinetics following controlled administration of smoked cannabis. Clinical chemistry 65, 9 (2019): 1171–1179.
27. Madelaine O Maykut; 1985. Health consequences of acute and chronic marihuana use. Progress in Neuro-Psychopharmacology and Biological Psychiatry 9, 3 (1985): 209–238.

28. Vladimir B Mikheev, Stephanie S Buehler, Marielle C Brinkman, Courtney A Granville, Timothy E Lane, Alexander Ivanov, Kandice M Cross, and Pamela I Clark; 2020. The application of commercially available mobile cigarette topography devices for e-cigarette vaping behavior measurements. Nicotine and Tobacco Research 22, 5 (2020): 681–688.
29. Rupesh K Mishra, Juliane R Sempionatto, Zhanhong Li, Christopher Brown, Nathalia M Galdino, Rushabh Shah, Shuyang Liu, Lee J Hubble, Kara Bagot, Susan Tapert, et al; 2020. Simultaneous detection of salivary Δ9-tetrahydrocannabinol and alcohol using a Wearable Electrochemical Ring Sensor. Talanta 211 (2020), 120757.
30. David C. Mohr, Mi Zhang, and Stephen M. Schueller. [n.d.]. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. 13 ([n. d.]): 23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949
31. Engineering National Academies of Sciences. [n.d.]. The Health Effects of Cannabis and Cannabinoids: The Current State of Evidence and Recommendations for Research. https://doi.org/10.17226/24625
32. Alessandro Ortis, Pasquale Caponnetto, Riccardo Polosa, Salvatore Urso, and Sebastiano Battiato; 2020. A Report on Smoking Detection and Quitting Technologies. International Journal of Environmental Research and Public Health 17, 7 (2020), 2614.
33. Abhinav Parate, Meng-Chieh Chiu, Chaniel Chadowitz, Deepak Ganesan, and Evangelos Kalogerakis; 2014. Risq: Recognizing smoking gestures with inertial sensors on a wristband. In Proceedings of the 12th annual international conference on Mobile systems, applications, and services. 149–161.
34. Eric R Pedersen, Justin F Hummer, Dipali Venkataraman Rinker, Zach K Traylor, and Clayton Neighbors; 2016. Measuring protective behavioral strategies for marijuana use among young adults. Journal of studies on alcohol and drugs 77, 3 (2016): 441–450.
35. Kristina T. Phillips, Michael M. Phillips, Trent L. Lalonde, and Kayla N. Tormohlen. [n.d.]. Marijuana use, craving, and academic motivation and performance among college students: An in-the-moment study. 47 ([n. d.]): 42–47. https://doi.org/10.1016/j.addbeh.2015.03.020
36. Thomas M Piasecki; 2019. Assessment of alcohol use in the natural environment. Alcoholism: clinical and experimental research 43, 4 (2019): 564–577.
37. D. Powers. [n.d.]. Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. /paper/Evaluation%3A-fromPrecision%2C-Recall-and-F-measure-to-Powers/6d03a38c5ddb7c7cd1ceb59b28907dc918c5d83a
38. Raul I Ramos-Garcia, Stephen Tiffany, and Edward Sazonov; 2016. Using respiratory signals for the recognition of human activities. In 2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE: 173–176.
39. Pertwee RG. [n.d.]. Handbook of Cannabis. Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199662685. 001.0001/acprof-9780199662685 Publication Title: Handbook of Cannabis.
40. Sohrab Saeb, Emily G Lattie, Stephen M Schueller, Konrad P Kording, and David C Mohr; 2016. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 4 (2016), e2537.
41. Philipp Scholl, Nagihan Kücükyildiz, and Kristof Van Laerhoven. [n.d.]. When do you light a fire?: capturing tobacco use with situated, wearable sensors. In UbiComp '13 Adjunct (2013). https://doi.org/10.1145/2494091.2497284
42. Philipp M Scholl and Kristof Van Laerhoven; 2012. A feasibility study of wrist-worn accelerometer based detection of smoking habits. In 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. IEEE: 886–891.
43. Randi Melissa Schuster, Robin J. Mermelstein, and Donald Hedeker. [n.d.]. Ecological Momentary Assessment of Working Memory Under Conditions of Simultaneous Marijuana and Tobacco Use. 111, 8 ([n. d.]): 1466–1476. https://doi.org/10.1111/add.13342

44. J Cobb Scott, Samantha T Slomiak, Jason D Jones, Adon FG Rosen, Tyler M Moore, and Ruben C Gur;2018. Association of cannabis with cognitive functioning in adolescents and young adults: a systematic review and meta-analysis. JAMA psychiatry 75, 6 (2018): 585–595.
45. Volkan Y Senyurek, Masudul H Imtiaz, Prajakta Belsare, Stephen Tiffany, and Edward Sazonov; 2019. A comparison of SVM and CNN-LSTM based approach for detecting smoke inhalations from respiratory signal. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE: 3262–3265.
46. Volkan Y Senyurek, Masudul H Imtiaz, Prajakta Belsare, Stephen Tiffany, and Edward Sazonov; 2020. A CNN-LSTM neural network for recognition of puffing in smoking episodes using wearable sensors. Biomedical Engineering Letters (2020): 1–9.
47. Joshua M Smyth and Kristin E Heron; 2016. Is providing mobile interventions" just-in-time" helpful? An experimental proof of concept study of just-in-time intervention for stress management. In 2016 IEEE Wireless Health (WH). IEEE: 1–7.
48. Tory R Spindle, Edward J Cone, Nicolas J Schlienz, John M Mitchell, George E Bigelow, Ronald Flegel, Eugene Hayes, and Ryan Vandrey; 2018. Acute effects of smoked and vaporized cannabis in healthy adults who infrequently use cannabis: a crossover trial. JAMA network open 1, 7 (2018):e184841–e184841.
49. Yan Wang, Daniel J. Fridberg, Robert F. Leeman, Robert L. Cook, and Eric C. Porges. [n.d.]. Wrist-worn alcohol biosensors: Strengths, limitations, and future directions. 81 ([n. d.]): 83–92. https://doi.org/10.1016/j.alcohol.2018.08.013
50. Lineke Zuurman, Annelies E Ippel, Eduard Moin, and Joop MA Van Gerven; 2009. Biomarkers for the effects of cannabis and THC in healthy volunteers. British journal of clinical pharmacology 67, 1 (2009): 5–21.
51. Bae S, Suffoletto B, Zhang T, Chung T, Ozolcer M, Islam M, Dey A. Leveraging Mobile Phone Sensors, Machine Learning, and Explainable Artificial Intelligence to Predict Imminent Same-Day Binge-drinking Events to Support Just-in-time Adaptive Interventions: Algorithm Development and Validation Study JMIR Form Res 2023;7:e39862 URL: https://formative.jmir.org/2023/1/e39862 DOI: 10.2196/39862
52. Lakmal Meegahapola, Florian Labhart, Thanh-Trung Phan, Daniel Gatica-Perez , Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous TechnologiesVolume 5Issue 3Article No.: 121pp 1–26https://doi.org/10.1145/3478126
53. Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. "From local explanations to global understanding with explainable AI for trees." Nature machine intelligence 2, no. 1 (2020): 56-67.
54. Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." Advances in neural information processing systems 30 (2017).
55. Sang Won Bae, Tammy Chung, Rahul Islam, Brian Suffoletto, Jiameng Du, Serim Jang, Yuuki Nishiyama, Raghu Mulukutla, Anind Dey. Mobile phone sensor-based detection of subjective cannabis intoxication in young adults: A feasibility study in real-world settings, Drug and Alcohol Dependence, Volume 228; 2021, 108972, ISSN 0376-8716, .
56. Steigerwald, Stacey, Peggy O. Wong, Beth E. Cohen, Julie H. Ishida, Marzieh Vali, Erin Madden, and Salomeh Keyhani. Smoking, vaping, and use of edibles and other forms of marijuana among US adults. Annals of internal medicine 169, no. 12 (2018): 890-892.
57. Yitzhaki, S., & Schechtman, E;2013. The Gini methodology: a primer on a statistical methodology (pp. 11-31). New York: Springer.
58. Bae, Sangwon, Anind K. Dey, and Carissa A. Low. "Using passively collected sedentary behavior to predict hospital readmission." In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 616-621; 2016.
59. Chung T, Bae SW, Mun EY, Suffoletto B, Nishiyama Y, Jang S, Dey AK. Mobile Assessment of Acute Effects of Marijuana on Cognitive Functioning in Young Adults: Observational Study. JMIR

Mhealth Uhealth; 2020 Mar 10;8(3):e16240. doi: 10.2196/16240. PMID: 32154789; PMCID: PMC7093776.
60. Mokrysz, C., Freeman, T., Korkki, S. et al. Are adolescents more vulnerable to the harmful effects of cannabis than adults? A placebo-controlled study in human males. Transl Psychiatry 6, e961 (2016). https://doi.org/10.1038/tp.2016.225
61. Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework.
62. K. Tara, A. K. Sarkar, M. A. G. Khan and J. R. Mou, "Detection of cardiac disorder using MATLAB based graphical user interface (GUI)," 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh; 2017, pp. 440-443, doi: 10.1109/R10-HTC.2017.8288994.
63. NIDA. "What are marijuana's effects on other aspects of physical health?." National Institute on Drug Abuse;13 Apr. 2021, https://nida.nih.gov/publications/research-reports/marijuana/what-are-marijuanas-effects-on-other-aspects-of-physical-health Accessed ;10 Aug. 2023.
64. Shrivastava D, Jung S, Saadat M, Sirohi R, Crewson K. How to interpret the results of a sleep study. J Community Hosp Intern Med Perspect; 2014 Nov 25.4(5):24983
65. https://en.wikipedia.org/wiki/Skewness ; 2023 Aug 28

## Abbreviations

| | |
|---|---|
| N | not-intoxicated |
| L | low-intoxication |
| MI | moderate-intensive intoxication |
| THC | delta-9 tetrahydrocannabinol |
| HR | heart rate |
| ESM | experience sampling method |
| API | application programming interface |
| MobiFit | Mobile and Fitbit sensors combined |
| XGBoost | eXtreme Gradient Boosting Machine classifier |
| SMOTE | synthetic minority over-sampling technique |
| Optuna | hyperparameter optimization framework |
| CV | cross-validation |
| AUC | Area Under the ROC Curve |
| TPR | true positive rate |
| XAI | explainable artificial intelligence |
| SHAP | SHapley Additive exPlanations |
| PDP | partial dependence plot |
| JITAI | just-in-time adaptive intervention |

# Supplementary Files