

mCrave: Continuous Estimation of Craving During Smoking Cessation

Soujanya Chatterjee*, Karen Hovsepian[∇], Hillol Sarker*, Nazir Saleheen*
Mustafa al’Absi[†], Gowtham Atluri[†], Emre Ertin[⊥], Cho Lam[∇]
Andrine Lemieux[†], Motohiro Nakajima[†], Bonnie Spring^ψ, David W. Wetter[∇], Santosh Kumar*
University of Memphis* Troy University[∇] University of Minnesota[†] Ohio State University[⊥]
Northwestern University^ψ Rice University[∇]
{schtrrj1, hsarker, nsleheen, skumar4}@memphis.edu* khovsepian@troy.edu[∇]
{malabsi, lemie008, mnakajim}@d.umn.edu[†] gowtham@cs.umn.edu[†] ertin.1@osu.edu[⊥]
{cho.lam, david.wetter}@rice.edu[∇] bspring@northwestern.edu^ψ

ABSTRACT

Craving usually precedes a lapse for impulsive behaviors such as overeating, drinking, smoking, and drug use. Passive estimation of craving from sensor data in the natural environment can be used to assist users in coping with craving. In this paper, we take the first steps towards developing a computational model to estimate cigarette craving (during smoking abstinence) at the minute-level using mobile sensor data. We use 2,012 hours of sensor data and 1,812 craving self-reports from 61 participants in a smoking cessation study. To estimate craving, we first obtain a continuous measure of stress from sensor data. We find that during hours of day when craving is high, stress associated with self-reported high craving is greater than stress associated with low craving. We use this and other insights to develop feature functions, and encode them as pattern detectors in a Conditional Random Field (CRF) based model to infer craving probabilities.

Author Keywords

Mobile Health; Smoking Cessation; Craving; Stress

ACM Classification Keywords

H.1.2. Models and Principles: User/Machine Systems

INTRODUCTION

Tobacco smoking is known to cause serious health issues such as cancer, respiratory diseases, cardiovascular diseases and metabolic diseases [1]. Smoking is responsible for 480,000 deaths per year in US alone, according to estimates from the Centers for Disease Control and Prevention (CDC) [1]. World-wide, first and second hand smoke causes over 6 million deaths per year, according to estimates from the World Health Organization (WHO) [3]. Encouragingly, nearly 7 out of 10 (68.8%) adult smokers report a desire to quit and almost

half of all smokers attempt to quit each year [1], however only 6.2% of those attempts are successful [2]. The majority relapses [14] in the first few days after a quit attempt.

Decades of prior research on smoking cessation with self-report has found that major predictors of a smoking lapse include stress (or negative affect) [42], smoking cues (e.g., seeing a cigarette) [40, 46], and craving [9, 18, 38]. Similar to other impulsive behaviors, craving is prevalent during the first few postquit days. Passive estimation of craving from sensor data in the natural environment [20] can enable the development of novel mobile tools to address the adverse consequences of craving during smoking abstinence.

Recent works on continuously estimating similar mental processes in the natural environment such as stress [13] from mobile physiological sensors, and [30] electrodermal activity (EDA) and detection of generalized tonic-clonic (GTC) seizures automatically are encouraging developments. But, to the best of our knowledge, there has not been any work to estimate craving from sensor data in the natural environment.

Estimating craving automatically from physiological response in the field environment is more challenging than estimation of stress. There are several reasons for this. First, the pathways linking perception of stress and its manifestation in physiology are now well established. In the case of identifying stress-related arousal, various features from physiological signals, such as heart rate variability (HRV) from Electrocardiogram (ECG) have been shown to be effective. To the best of our knowledge, no such work exists for inferring craving from physiological response.

Second, there have been works [21] showing activation of specific parts of the brain in response to craving but activation of these components of the brain are not known to have a specific and identifiable manifestation in physiological arousal. Third, no gold standard exists that can be used as labels for training and testing a craving model, especially in field settings. Therefore, we are limited to using self-reports as labels. Given the inherent variability of self-reports in capturing mental states, even for well-researched phenomenon such as stress, the best correspondence between physiological response and self-reports collected in the field is 0.71 [13].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
UbiComp '16, September 12–16, 2016, Heidelberg, Germany.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-4461-6/16/09...\$15.00.
<http://dx.doi.org/10.1145/2971648.2971672>

In this paper, we take the first step towards developing a computational model to estimate craving for each minute (during smoking abstinence) using sensor data. We use multiple key insights that are well-supported by craving research (conducted using self-reports). First, rather than targeting estimation of craving as a general phenomenon, we restrict ourselves to craving estimation during the post-quit period of smoking cessation, when craving estimation has the highest clinical utility [6].

The second insight is that although craving may not manifest itself clearly in a visible physiological response, high-craving moments may lead to elevation in the quitters' stress levels as they struggle to cope with craving [24]. However, stress may not always be due to craving, and in some cases, craving may not result in a stress response. Hence, stress can't be used as a direct surrogate of craving. The final insight we use is that time of day does have a noticeable effect on self-reported craving. This effect is supported by smoking abstinence research [26], and was observed in our data as well.

We analyze 2,012 hours of sensor data and 1,812 craving self-reports from 61 participants in a smoking cessation study in their post-quit period. We first verify each of the above insights in our dataset with appropriate statistical tests. We use a recent model of stress measure [13] for stress assessment.

In our analysis of craving self-reports during the post-quit period, we observe that craving is moderate in the morning, decreases in late morning, increases substantially in the afternoon (after lunch) and increases further in the evening-night. This observation is in line with a similar finding reported in an independent study [26].

In our analysis of association between craving and stress, we find that during the hours of day when craving is high, stress associated with self-reported high craving is significantly greater than stress associated with self-reported low craving. On the other hand, during hours of a day when craving is relatively low, stress associated with self-reported high craving is not significantly greater than that associated with self-reported low craving. In addition, we note that high craving in a minute is generally followed by high craving in the next minute, since craving dissipates gradually over time. Similarly, low craving in a minute is mostly followed by low craving in the next minute.

Prior to describing the *mCrave* model we present a statistical and exploratory analysis of craving, with an emphasis on gaining insights into the relationship between craving, time of day, and stress. In particular, we propose and statistically test two hypotheses describing this relationship.

Next, we use these insights to propose our *mCrave* model for craving estimation, which is based on a linear-chain Conditional Random Fields (CRF) model. The model is provided with observed input signals which contain salient patterns indicating heightened craving at the locations of the patterns. On the basis of these patterns, the model is capable of several types of inferences, such as inferring the probability of any given sequence of low/high craving labels, inferring the most probable sequence of labels over any period, or even infer-

ring the time series of marginal probabilities of high craving for each minute. As part of the training algorithm, the model learns the weights of the aforementioned patterns such that it can make accurate inferences. To do this, the model requires ground-truth craving labels for at least a subset of the data. A strength of the CRF model is that it can also make use of minute-to-minute transition patterns, such as craving to non-craving, and vice-versa, to improve the inference accuracy.

BACKGROUND

In this section, we illustrate the smoking lapse process, the craving process, and the utility of estimating craving.

Smoking Lapse Process

In the abstinence phase, deprivation of nicotine leads to withdrawal symptoms such as anxiety, sadness, anger, concentration impairment, increased hunger, and others [27, 36]. The withdrawal effect may be accentuated by environmental stimuli (e.g., visual exposure to smoking cues, alcohol) and social triggers (e.g., social gathering of friends who are smokers) [40]. Quitters who are better able to cope with these withdrawal effects (e.g., craving and stress) are relatively more successful in maintaining abstinence [5, 27]. For lapsers, a rapid increase in negative affect and smoking urges/cravings is associated with the first smoking lapse [4, 42]. Unfortunately, the first lapse usually leads to a full relapse [16].

Craving Process

Craving or urge to smoke is generally conceptualized as the motivational state of desire for nicotine [15, 38]. Although urge and craving may have different meaning [19], they can be used interchangeably as there is a correlation of 0.96 between self-reported urge and craving [38, 40, 47]. Hence, we consider urge and craving interchangeably hereafter.

During the post-quit abstinence phase, individuals experience high craving soon after quitting that reduces as time progresses [18, 38, 49]. Periodic episodes of prolonged craving are associated with stress (or negative affect) [24, 39, 42], time of day [26], smoking cues [38, 40], and alcohol consumption [40]. During abstinence, individuals often need to cope with prolonged, recurring, and intensified craving effects in order to maintain abstinence. Individuals experiencing higher craving soon after quitting are more likely to relapse [18].

Assessing Craving During Smoking Abstinence

Data collected in smoking cessation studies [9, 17, 26, 48] rely mostly on retrospective recall of craving and stress (or negative affect) that may be prone to recall biases and errors [12]. To reduce these problems, Ecological Momentary Assessments (EMA) [41, 44] collect repeated momentary self-reports in the natural environment, longitudinally.

The use of EMA in behavioral medicine is a significant methodological advancement, but it also has several limitations. First, EMA depends on the participants' volition to answer probes, making them prone to noncompliance due to not being available at the time of the prompt [34], or to reporting burden, which reduces data integrity. Second, EMA prompted at random times may miss the most opportune moment (e.g., intense stress and subsequent craving prior to a

lapse). Third, recall from the past (from minutes to hours) may still suffer from recall bias and error.

Smoking researchers have envisioned [37,43] the use of ubiquitous computing (e.g., smartphone and wearable sensors) in smoking cessation research. Recent advances in ubiquitous computing have now made it feasible to obtain critical measures (e.g., stress) from sensors [11,13,28].

Benefits of Minute-level Craving Estimation

In this work, we estimate craving at the minute-level during the abstinence phase in smoking cessation. Continuous estimation of craving during abstinence has several utilities.

First, the minute-level objective estimation of craving will not rely on participants' self-assessment, which eliminates recall-bias and noncompliance. Second, participant burden of frequent self-reporting can be reduced. Third, researchers can analyze craving data at a higher temporal resolution (e.g., prior to a lapse) and assess lapse risk. Fourth, this work accelerates the discovery of a model that can reliably predict first lapse or high risk situations during the abstinence period.

RELATED WORKS

Research on Craving from Self-reports

Craving has been studied extensively in smoking cessation research via self-reports. It has been found that escalation of stress and craving during a quit attempt may contribute to smoking lapse [6]. In addition, there is an increased risk for lapse following a sudden spike in craving [25]. Dynamic changes in craving during smoking cessation [38] show that it is episodic in nature.

Time of day effect on craving has also been widely reported [26]. During abstinence, craving is less in the morning, it gets elevated in the afternoon (i.e., after lunch) and increases further in the evening.

Associations of stress and craving have also been studied. For example, stress is associated with increased craving intensity and decreased self-control for rewarding substances (e.g., nicotine/cigarettes) in abstinent smokers [24].

Most of the research on craving, such as the ones discussed here, however, has been based on self-reports (e.g., EMA). In addition to suffering from reporting biases and response burden, low temporal resolution of self-reports has prevented a study of craving in the minutes preceding a high-risk or lapse situation. Our proposed model for continuous estimation of craving can now facilitate investigation of craving around significant clinical events, including the minutes prior to lapses.

Continuous Estimation of Mental States from Sensors

Advances in ubiquitous computing have resulted in several models for continuous estimation of the mental states of humans. For example, [45] demonstrates a method to dynamically infer multiple levels of user frustration caused by system response delays using physiological sensors. Computational models also exist to continuously infer physiological stress using electrocardiogram (ECG) and respiration data [13,29]. These contemporary works indicate that ubiquitous sensors

have progressed to the point that they can continuously measure behavior, physiology, and mental states from sensor data. Our work makes use of some of these models (e.g., stress) as inputs, but none of these existing models can be used directly to develop a model for estimating craving.

DATA COLLECTION

We describe the smoking cessation study [32] whose data was used in the development of the *mCrave* model. The study was approved by the Institutional Review Board (IRB).

Devices and Sensor Measurements

Wearable Sensor Suite: Participants in the study wore a wireless physiological sensor suite (AutoSense [11]) underneath their clothes. The wearable sensor suite consisted of two-lead electrocardiograph (ECG), 3-axis accelerometer, and respiration sensors. Participants also wore an inertial sensor (i.e., smartwatch), which includes a 3-axis accelerometer and a 3-axis gyroscope, on each wrist. Each sensor transmitted the sensor data continuously to a mobile phone. AutoSense respiration sensor has its own battery and it lasts for 10 days on a 750 mAh battery. It uses a low-powered ANT Radio to connect with the phone. The phone (which collects GPS data continuously and keeps its wireless radio on for data reception) lasts for 13 hours on a single charge. The smartwatch we use lasts 3 days on a 500 mAh battery. The sampling rate is 21.3 Hz for the respiration sensor, 64 Hz for the ECG sensor, and 16 Hz for each axis of the accelerometer on the smartwatch.

Mobile Phone: Participants were given a smartphone to carry. It receives and stores data from sensors on the body and on the phone. It also collects information via EMA, which captures the characteristics of situational factors associated with craving. These factors include stress and physical activity levels. In the smoking cessation study, participants used the phone to report (at random prompts) their *craving level* on a Likert scale of 1-6.

Smoking Cessation Study

Participants: Participants were 61 smokers (27 females) with mean age of 37 ± 12.54 and years-of-education of 14 ± 1.82 . Ethnically, there were 47 Caucasians, 10 African-Americans, 2 Native Hawaiian, and 2 from multiple races. All participants reported smoking 10 or more cigarettes per day for at least 2 years, and reported a high motivation to quit. To qualify, participants had to pass a screening session prior to being enrolled in the study. The screening included assessment of current medical and mental health status and history of any major medical and psychiatric illness. Screening also included assessment of smoking behavior, mood, and other behavioral health measures. Participants were excluded if they had ongoing major medical or psychiatric problems and if they had other comorbid psychiatric and substance use problems. Also, participants who were not entrained to the normal day/light diurnal cycle were excluded to control for variation in diurnal physiological activity and behaviors.

Protocol: Once enrolled, the participants picked a smoking quit date. Two weeks prior to their quit date, subjects wore

the sensor suite for 24 hours in their natural environment. After completion of the 24 hours of monitoring, which we call the pre-quit session, participants returned to the lab for their second visit. Smoking cessation counseling was provided starting at this second visit. Then the participants returned to the lab on the assigned quit date to attend a counseling session and to begin the 72 hours of monitoring in the field; this we refer to as the post-quit session. They came back to the lab each day to confirm smoking status via an expired breath sample in a carbon monoxide (CO) monitor. During each day of monitoring (24 hours pre-quit and 72 hours post-quit), the participants wore the sensor suite during awake hours, and completed on the mobile phone, 12 EMAs (i.e., self-reports) daily. All participants were compensated monetarily for their time and effort (\$430 after successful completion).

Total Data Collected: We collected data from 61 participants. The participants wore the sensor suite for a total of 2,766 hours (754 pre-quit and 2,012 post-quit).

Lapse Detection: For the *mCrave* model development, we use the data collected (both sensor and self-report) during the post-quit, but pre-lapse phase. This requires detection of first-lapse events. Although participants undergo carbon monoxide (CO) testing each day and are asked to self-report their lapse events, none of these are temporally precise enough to precisely mark lapse occurrence in the sensor data (that is collected at the rate of tens of hertz). Therefore, we use a recently developed model (*puffMarker*) for detecting first smoking lapse events from sensor data.

puffMarker [32] is a multi-sensor approach for pinpointing the timing of first lapse in smoking cessation. *puffMarker* uses data collected from two wearable sensors, breathing pattern captured from a RIP sensor and hand gestures captured using 6-axis inertial sensors (3-axis accelerometers and 3-axis gyroscopes) worn on wrists. It uses inertial sensor data to identify hand-to-mouth gestures and applies a machine learning model on the corresponding respiration data to detect deep inhalation and exhalation pattern expected during smoking. By using both of these sensing modalities, *puffMarker* achieves good accuracy. When applied to 3 days of post-quit data from 33 lapsers (from the participant pool of 61) [32], it correctly pinpoints the timing of first lapse in 28 participants; data from the other five was not available due to sensor-non-wearing or lost data at the time of first lapse.

Participant Selection for mCrave Modeling: After applying the *puffmarker* model on lapsers, a lapse time was established for the lapsers. For *mCrave*, to ensure uniformity and sufficiency of self-reported craving data (that is used as labels), we selected those participants (from both lapsers and abstainers) who had a minimum of 2 craving reports in each time of day (morning, afternoon and evening/night) during their abstinent (i.e., post quit, but pre-lapse) period. As a result, 16 participants were excluded either due to lack of reporting their craving level or because they lapsed before evening of the first post-quit day. We use the remaining 45 participants, who contributed 1,557 self-reports of craving. Further, if there was a missing stress inference at the minute adjacent to a craving report, due to activity or data loss, we excluded

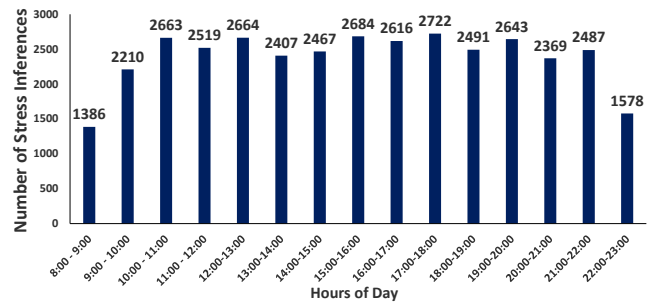


Figure 1. Distribution of Stress Inferences across hours of a day. Number of participants = 45

that craving report from our analysis. We are thus left with 1,109 craving self-reports, of which 109 are used to initialize the model and the remaining 1,000 for training and testing.

STRESS ASSESSMENT FROM SENSOR DATA

We use the *cStress* model for stress assessment. *cStress* uses Electrocardiogram (ECG) and respiration to infer stress. This model is applied to a set of features computed from each minute of sensor data, whereby consecutive minutes are non-overlapping. The model determines whether this minute’s sensor readings correspond to a physiological response to stressors. Features used in the model include 80th percentile of R-R intervals and Heart Rate Variability (HRV) from ECG, and mean inspiration-expiration ratio and median of stretch from respiration [13]. This model was shown to classify stress and non-stress minutes with 95% accuracy on independent subject validation (different from training set) in lab testing. It also showed that using HRV measure alone from ECG, as has been the case in several prior works [22, 23], leads to a significant drop in F1 score (0.56 vs. 0.78). Finally, it was evaluated against self-report from independent set of 20 participants who wore sensors for a week in the field and was found to have a median F1 score of 0.71 [13]. Subsequently, it was also shown to have a median F1 score of 0.717 with self-reports collected from a different set of 38 participants who wore sensors for four weeks in the field [35]. This stress model was recently validated on our smoking cessation dataset and was found to have a median F1 score of 0.68 [33].

The *cStress* model, briefly described above, provides a continuous inference of stress, scaled to be between 0 and 1, for every 1-minute of sensor data. This time series of 1-minute probability-like measures of stress, for a particular participant, is termed throughout the rest of the paper as “*stress likelihood*”. Figure 1 shows the hourly distribution of stress inferences of an awake day across all participants.

ASSOCIATION OF TIME OF DAY, STRESS, AND CRAVING DURING ABSTINENCE

Here we describe two hypotheses and support them with statistical tests on our data. The hypotheses provide key insights to the construction of the craving estimation model. Figure 2 provides an overview of the analysis structure.

Craving likelihood based on hour (time) of day

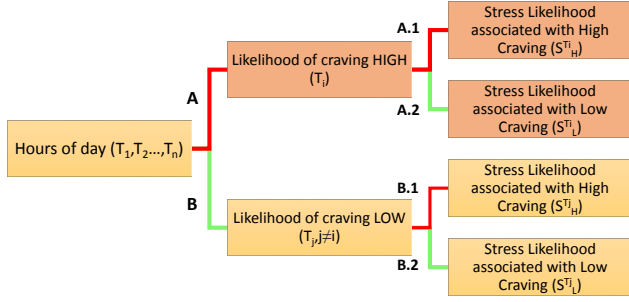


Figure 2. Analysis of Hour (time) of day, craving, and physiological stress.

| Name | Description |
|-------------|--|
| T_i | High Craving Likelihood Hour (High Vulnerable Hour) |
| T_j | Low Craving Likelihood Hour (Low Vulnerable Hour) |
| $S_H^{T_i}$ | Stress likelihood associated with self-reported high craving during high vulnerable hour |
| $S_L^{T_i}$ | Stress likelihood associated with self-reported low craving during high vulnerable hour |
| $S_H^{T_j}$ | Stress likelihood associated with self-reported high craving during low vulnerable hour |
| $S_L^{T_j}$ | Stress likelihood associated with self-reported low craving during low vulnerable hour |

Table 1. Variables and parameters

Self-reported craving data has been collected during the awake hours of the participants. Participants respond to assessments containing a craving item on a Likert scale of 1-6 prompted at random times in a day. We analyze the self-reported craving assessments (1,557 self-reports) of 45 participants. Individual differences in self-reported craving ratings affect the comparison of inter-individual ratings. To overcome this challenge, we compute a z -score transformation of the craving ratings for each participant p_i , separately. For participant p_i , we compute the mean, μ_i and standard deviation, σ_i of ratings of all the craving assessments reported by that participant. Now, for craving assessment c_{ij} of participant p_i , we compute the z -score \hat{z}_{ij} as,

$$\hat{z}_{ij} = \frac{c_{ij} - \mu_i}{\sigma_i}, i = 1, 2, \dots, n \quad \text{and} \quad j = 1, 2, \dots, k,$$

where k is the total number of craving assessments for participant p_i and $n = 45$ is the total number of participants.

Craving ratings with z -score greater than 0 are marked as **high craving episodes** and those with z -score less than 0 are marked as **low craving episodes**. We observe that during abstinence, craving is lower during morning, with mean craving z score of -0.132 ($p(C) = 0.451^1$), craving increases during the afternoon, with mean craving z score of 0.033 ($p(C) = 0.533$), and further increases during the evening, with mean craving z score of 0.106 ($p(C) = 0.545$). We find that craving

¹ $p(C)$ refers to probability of high craving episodes.

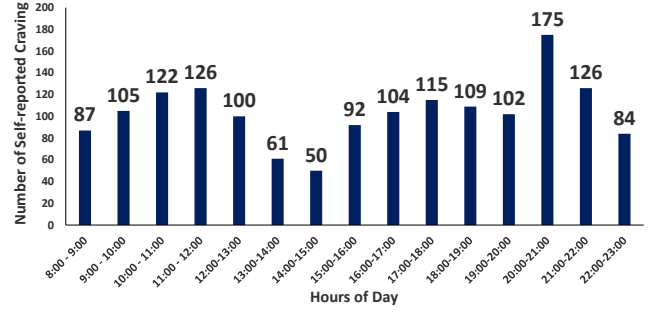


Figure 3. Total number of self-report assessments during hours of day across all participants. Number of participants, $n = 45$

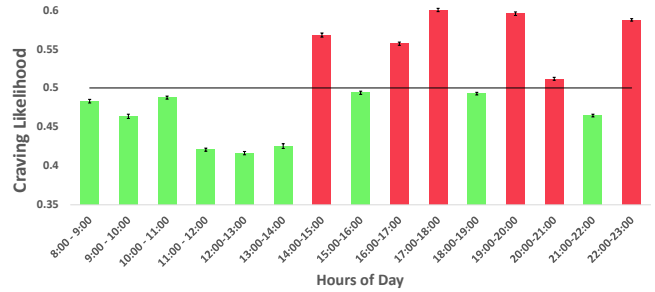


Figure 4. Craving distribution across hours of a day. Hours with *High* craving likelihood are marked with red. Hours with *Low* craving likelihood are marked with green

during the afternoon ($n = 178$) is significantly greater than in the morning ($n = 404$) ($p = 0.024$ obtained using one-tailed *Wilcoxon rank-sum test*). Also, craving during the evening ($n = 527$) is significantly greater than in the morning ($n = 404$) ($p < 0.0001$ obtained using one-tailed *Wilcoxon rank-sum test*), however, there is no significant difference in craving during evening ($n = 527$) and afternoon ($n = 178$) ($p = 0.489$ obtained using two-tailed pairwise *Wilcoxon rank-sum test*). Our finding agrees with that reported in [26] on different data.

We further estimated the distribution of normalized self-reported craving episodes during each awake hour (T_8, T_9, \dots, T_{22} , represent 15 hours of awake day, where T_9 represents the hour of day greater than equal to 9:00 and less than 10:00), across all the participants. Figure 3 illustrates the total number of self-report assessments during each hour of a day. We compute the craving likelihood, as the proportion of **high craving episodes** reported in each hour,

$$\hat{p}_c = \frac{\text{total number of assessments reported as high craving}}{\text{total number of self-reported assessments}}$$

Now, prior to classifying each hour of a day as high or low craving likelihood hour, we need to know the level of precision for the expected craving likelihood in each hour. We used the bootstrap method (random resampling with replacement) [10] in order to obtain mean and standard error for craving likelihood in each hour. We computed the 95% confidence interval (CI) of craving likelihood in each hour. We marked the set of hours of a day, as *high craving likelihood*

hours, T_i , (see Table 1) where both the upper and lower limits of 95% CI of craving likelihood, are greater than 0.5 (indicating majority voting). On the other hand, we marked the set of hours of a day as *low craving likelihood hours*, T_j ($j \neq i$), (see Table 1) where both the upper and lower limits of 95% CI of craving likelihood are less than or equal to 0.5. Figure 4 illustrates the distribution of reported craving across all participants during hours of a day. We observe that the hours of day when *craving likelihood is high* are,

$$T_i = [T_{14}, T_{16}, T_{17}, T_{19}, T_{20}, T_{22}]$$

Similarly, hours of day when *craving likelihood is low* are,

$$T_j = [T_8, T_9, T_{10}, T_{11}, T_{12}, T_{13}, T_{15}, T_{18}, T_{21}]$$

Shown in Figure 2, the hours of a day is divided into hours T_i when craving likelihood is high (red line marked as A), and T_j when craving likelihood is low (green line marked as B). The hours of day with high craving likelihood (T_i) are referred to as *High Vulnerable hours*; those with low craving likelihood (T_j) are referred to as *Low Vulnerable hours* (see Table 1).

Physiological Stress and Craving

Across all participants ($n = 45$), we compute the stress likelihood at the nearest minute adjacent to each craving self-report. As described previously, all the 1,109 craving self-reports selected for analysis have a stress inference at the minute adjacent to the craving report. Next, we group the craving reports and their associated stress likelihood according to hours, T_i (see Table 1) and hours T_j (see Table 1).

We hypothesize that physiological stress response is associated with craving during the smoking abstinence period, but, not always. We believe that during specific times of a day, high stress response may be associated with high craving (e.g., when participants become aware of the thought that they will not be able to smoke anymore, or under influence of smoking cues like alcohol), however, during other times stress response may be elevated due to other reasons (e.g., busy working in order to meet a deadline).

Following are the two alternate hypotheses,

(H01) During the hours (or times) of a day, T_i when participants are highly vulnerable, the stress likelihood associated with high craving (denoted as, $S_H^{T_i}$ in Figure 2, A.1) is significantly greater than the stress likelihood associated with low craving (denoted as, $S_L^{T_i}$ in Figure 2, A.2).

(H02) During the hours (or times) of a day, T_j when participants are not highly vulnerable, there is significant difference between the stress likelihood associated with high (denoted as, $S_H^{T_j}$ in Figure 2, B.1) and low craving (denoted as, $S_L^{T_j}$ in Figure 2, B.2).

We performed two statistical tests in order to provide a convincing rationale behind the above hypotheses.

First: To assess **H01**, we performed a two sample right tailed *Wilcoxon rank-sum test* for $S_H^{T_i}$ and $S_L^{T_i}$ (see Table 1) with $n = 269$ and $n = 202$, respectively. Interestingly, we found that $S_H^{T_i}$ is significantly greater than $S_L^{T_i}$ ($p = 0.012$).

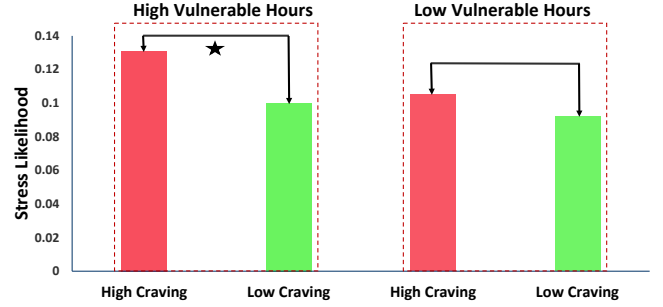


Figure 5. Median Stress likelihood associated with self-reported high craving is significantly greater than that associated with self-reported low craving during high vulnerable hours (marked with the star), however there is no significant difference between Median Stress likelihood associated with self-reported high craving and that associated with self-reported low craving during low vulnerable hours

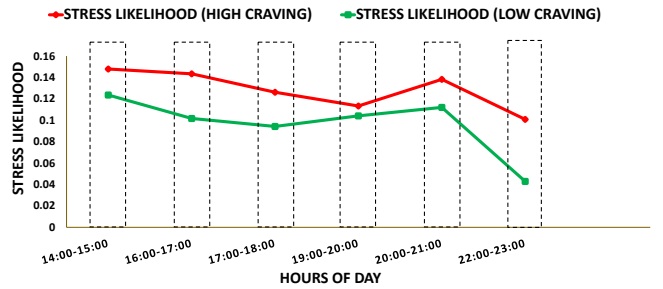


Figure 6. Median Stress Likelihood associated with self-reported high craving (red) and self-reported low craving (green) during high vulnerable hours of day. Median Stress Likelihood associated with high craving (red) significantly greater than that associated with low craving (green)

We observed that median of $S_H^{T_i}$ samples is 0.131 (mean = 0.196 ± 0.193) and median of $S_L^{T_i}$ samples is 0.1 (mean = 0.159 ± 0.176). Left half of Figure 5 shows the comparison between them. In order to assess **H02**, we performed a two sample two tailed *Wilcoxon rank-sum test* for $S_H^{T_j}$ and $S_L^{T_j}$ (see Table 1) with $n = 295$ and $n = 343$, respectively. We found that $S_H^{T_j}$ and $S_L^{T_j}$ are not significantly different ($p = 0.229$). We observed that the median of $S_H^{T_j}$ is 0.105 (mean = 0.167 ± 0.172) and the median of $S_L^{T_j}$ is 0.092 (mean = 0.152 ± 0.161). The right half of Figure 5 shows the comparison between them. Interestingly, in both cases, the mean stress likelihood is significantly greater than the median stress likelihood. This can be explained by the fact that stress likelihood follows a right-skewed beta distribution [35].

Second: The, number of self-reports in each hour is not the same (mean = 70.125 ± 27.959). Consequently, the result of *Wilcoxon rank-sum test* can be biased. Hence, we performed a second test, where we first computed the median stress likelihood associated with high and low self-reported craving in each hour of a day. Next, in order to assess **H01**, we performed a right tailed pairwise *Wilcoxon sign-rank test* for median of $S_H^{T_i}$ and median of $S_L^{T_i}$ in each high vulnerable hour, $n = 6$ pairs. We found that median of $S_H^{T_i}$ is significantly higher than the median of $S_L^{T_i}$ ($p = 0.015$). Figure 6 illustrates the samples used in this test. In order to assess **H02**, we

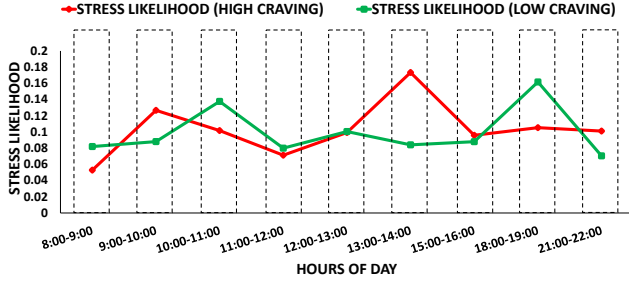


Figure 7. Median Stress Likelihood associated with High craving (red) and low craving (green) during low vulnerable hours. No significant difference between Median Stress Likelihood associated with high craving (red) and that associated with low craving (green)

performed a two sided pairwise *Wilcoxon sign-rank test* on the median of $S_H^{T_j}$ and the median of $S_L^{T_j}$ in each low vulnerable hour, $n = 9$ pairs. We found that the median of $S_H^{T_j}$ and the median of $S_L^{T_j}$ are not significantly different ($p = 0.945$). Figure 7 illustrates the samples used in this test.

THE MCRAVE MODEL

In this section, we describe our proposed Conditional Random Fields model, capable of inferring the probabilities of high craving continuously on a minute-by-minute basis, over a span of an entire day, provided we know the minute-level stress likelihoods over the course of the same day. Conditional Random Fields (CRFs) is a well-established and highly flexible class of graphical models for defining probability distributions over sequences of inter-dependent categorical random variables, conditioned on some observed evidential data.

To apply CRFs to our problem, we first define a sequence of random output variables $C = \{C(i) \in \{-1: \text{low craving}, +1: \text{high craving}\} | i = 1 \dots n\}$, which mark the binary craving levels for all n recorded minutes of the given day. The observed evidential data made available to the model consist of previously defined $\hat{p}_c(t) |_{t=8 \dots 23}$, which are a time series of non-day-specific hourly craving likelihoods spanning hours from 8:00 (8AM) to 23:00 (11PM), and $S_z(i) |_{i=1 \dots n}$, which are a time-series of standardized minute-level stress likelihoods, produced by standardizing, i.e., subtracting the daily mean and dividing by the daily standard deviation, the output of the *cStress* model. The standardization is important for the sake of producing a population-wide model, which is robust in the face of day-specific differences in baseline mean and standard deviation of stress likelihoods. The model defines the joint conditional distribution $P_\theta(C | \hat{p}_c, S_z)$. Once the parameters θ of this distribution are learned, we can use the model to infer the marginal probabilities of high craving for every minute of the day.

The key advantage of CRFs is their ability to use salient features that target the complex patterns and relationships that are expected to exist between the output variables $C(i)$ and the evidential data \hat{p}_c , and S_z . The ability to use these features can make CRFs a better choice in a supervised learning setting than such generative models as Bayesian Networks. Like Bayesian Networks, however, CRFs can also model structural

patterns and probabilistic dependencies between output variables, which gives them an edge in structural learning over other feature-rich models, such as Support Vector Machines and Logistic Regression.

The full conditional probability of a sequence of high/low labels, defined by our CRF model, is:

$$P_\theta(C | \hat{p}_c, S_z) = \frac{1}{Z(\hat{p}_c, S_z)} \Psi(C, \hat{p}_c, S_z) \quad (1)$$

$$\Psi(C, \hat{p}_c, S_z) = \exp \left\{ \sum_{i=1}^n \sum_j^k f_j(C, \hat{p}_c, S_z, i) \right\} \quad (2)$$

The function Z above is the so-called partition function, and it is used to normalize the numerator, and make sure that $P_\theta(C | \hat{p}_c, S_z)$ is scaled between 0 and 1 and adds up to 1. Computing Z presents one of the main challenges in CRF inference, because it is defined as the sum of $\Psi(C', \hat{p}_c, S_z)$ over all possible C' , of which there are exponentially many. Luckily, in a linear-chain CRF, this sum can be computed efficiently using a dynamic program called Sum-Product Message Passing (or Exact Belief Propagation).

The $f_j(C, \hat{p}_c, S_z, i)$ are the previously-mentioned feature functions, which are divided into so-called ‘local’ features, capturing the compatibility between observed evidential data and the craving label at a single minute i , and ‘pairwise’ features, which capture dependencies between successive craving labels. A positive output of f_j signals an agreement between the variables and the evidential data, whereas a negative value is a sign of a disagreement. A value close to 0 indicates an ambiguous/low signal.

The performance of the model depends on the discriminative power and generalizability of the feature functions encoded in the model. Below, we list the specific ‘local’ and ‘pairwise’ features used in our model, which are inspired by the analyses and hypotheses mentioned in the previous sections.

- $f_1(C, \hat{p}_c, S_z, i) = \alpha_1 C(i)(\hat{p}_c(t) - 0.5)$
Hour t contains minute i . This feature measures how well a craving label at a minute level agrees with the likelihood of craving at an hour level. If the output is positive, the sign of craving label $C(i)$ agrees with the sign of $\hat{p}_c(t) - 0.5$, whereas a negative output indicates a disagreement between them. The absolute value is the degree of agreement/disagreement. The constant α_1 weighs the role of this feature/pattern in deciding the local compatibility of $C(i)$, relative to all other features.
- $f_2(C, \hat{p}_c, S_z, i) = \alpha_2 C(i)(S_z(i) + \gamma \hat{p}_c(t) - \gamma 0.5)$
This feature expands on the previous feature and measures the three-way agreement among minute-level stress likelihood, hourly craving likelihood, and minute-level craving labels. Additionally, this feature is based on construction of a linear separating boundary between the **low craving** minutes ($C(i) = -1$) and **high craving** minutes ($C(i) = +1$) in the space $S_z(i) \times \hat{p}_c(t)$, with $\hat{p}_c(t)$ as the x -axis. The boundary has an x -intercept at $\hat{p}_c(t) = 0.5$, and is parametrized by the constant γ , denoting the slope of the boundary. The relationship it encodes is that if the craving

likelihood is low, it requires a much higher likelihood of stress to indicate high craving at that minute. On the other hand, if there is already a high likelihood of craving during that hour, it does not take as much stress to trigger high craving.

- $f_3(C, \hat{p}_c, S_z, i) = \alpha_3 \mathbf{1}_{\{\hat{p}_c(t) > 0.5\}} S_z(i) C(i)$
This feature is based on the hypothesis **H01**, discussed in the previous section, and detects instances of high stress likelihood associated with self-reported high craving during high vulnerable hours.

- $f_4(C, \hat{p}_c, S_z, i) = \begin{cases} \beta_1 & \text{if } C(i) = C(i+1) = -1 \\ \beta_2 & \text{if } C(i) = -1, C(i+1) = 1 \\ \beta_3 & \text{if } C(i) = 1, C(i+1) = -1 \\ \beta_4 & \text{if } C(i) = C(i+1) = 1 \end{cases}$

This *pairwise* feature measures the compatibility scores for all transitions of craving labels from minute to minute. For example, β_3 measures the compatibility of transitioning from ‘high craving’ at minute i to ‘low craving’ at minute $i + 1$.

The constants $\alpha_i, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \beta_4, \gamma$ comprise the parameter vector θ , and they can be learned using supervised learning via Maximum Likelihood Estimation, provided we have a sequence of ground-truth labels for C . Luckily, these weights can be learned even if we have ground-truth labels for only a portion of the minutes, as is the case with EMA-based sampling of self-reports throughout the day, by defining the partial likelihood of parameters given the sample of ground-truth labels. We use a L_2 regularized likelihood, to improve the generalization and convergence characteristics of the learning process. We use a regularization constant λ to tune the role of regularization.

Once the model is trained, it can be applied to provide several different inferences. The inference that we find to be most useful is the inference of $P(C_i = \text{high craving} | \hat{p}_c, S_z)_{i=1..n}$ — the marginal probabilities of high craving for all minutes of the day. They can be computed efficiently using an algorithm almost identical to the one used to compute Z . These marginal probabilities are used during validation of the model, enabling us to compute various classification performance metrics, such as F1, accuracy (hit-rate), Area-under-the-Curve (AUC), and others.

FINDINGS

In this section, we report the experimental results and findings obtained with our proposed *mCrave* model. The only hyperparameter used during learning is the choice of the regularization constant λ . To find the optimal value for this constant, as well as to validate the model’s generalization performance in a robust manner, we performed leave-one-participant-out (LOPO) cross validation.

The training data consists of daily self-reported ground-truth labels for 45 participants. In total, there are 1,000 ground-truth craving labels over all days, 505 of them from the *high craving* class, and 495 from *low craving* class. The LOPO results are summarized below.

Craving estimation during High Vulnerable Hours (T_i)

| | | Estimated by <i>mCrave</i> | | |
|--------|--------------|----------------------------|-------------|-------|
| | | High Craving | Low Craving | Total |
| Actual | High Craving | 222 (78.4%) | 61 (21.6%) | 283 |
| | Low Craving | 77 (34.7%) | 145 (65.3%) | 222 |
| Total | | 299 | 206 | 505 |

Table 2. Confusion Matrix for High Vulnerable hour. Overall Accuracy is 72.9% (against base accuracy 56.4%) with kappa 0.429.

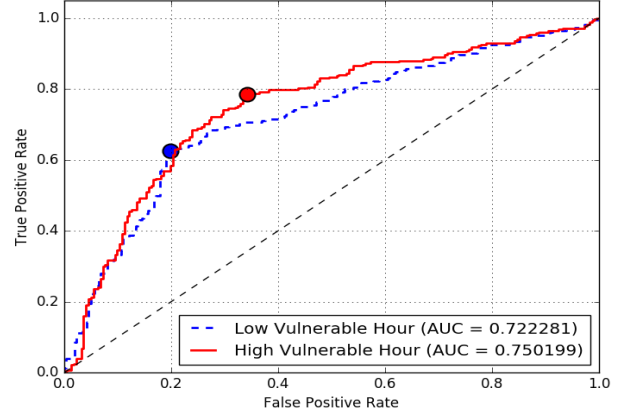


Figure 8. Receiver Operating Characteristic(ROC) curve

Out of the total number of craving ground-truth labels, 505 are during high vulnerable hours. Of these, 283 belong to *high craving* class and 222 to *low craving* class. As Table 2 shows, the overall accuracy of classifying these craving self-reports is 72.9%. Furthermore, the model attained a true positive rate of 78.4%, where positive class is *high craving* class, and a true negative rate of 65.3% cases. The precision, recall, F₁ score, and Area-Under-Curve (AUC) of the model are 0.742, 0.784, 0.764 and 0.750, respectively.

Craving estimation during Low Vulnerable Hours (T_j)

On the other side, 495 ground-truth craving labels are during the low vulnerable hours. Of these, 225 belong to the *high craving* class, and 270 to the *low craving* class. We obtained an overall accuracy of 72.2%. From the confusion matrix in Table 3, we found that the model attained a true positive rate of 62.7% cases, and a true negative rate of 79.6% cases. The model obtained a precision of 0.719, a recall of 0.627, an F₁ score of 0.673 and Area-Under-Curve (AUC) of 0.722.

For comparison, the median F1 score of the *cStress* model is 0.68 [33] for continuous estimation of stress and its agreement with self-reported stress in the field setting. As described previously, the *cStress* model is based on extensive prior works on stress and physiology, yet its accuracy is limited when comparing against self-report in the field setting, due to wide variability and occasional inconsistencies in self-reports. Therefore, the accuracy of craving estimation is quite good for a first model, as its comparison is also with self-reports collected in the field setting.

In addition to the confusion table, we plotted the ROC curves for the LOPO results, comparing the low vulnerable and high vulnerable curves side-by-side, as shown in Figure 8. We observed that the performance of the model during high vulner-

| | | Estimated by <i>mCrave</i> | | |
|--------|--------------|----------------------------|--------------------|-------|
| | | High Craving | Low Craving | Total |
| Actual | High Craving | 141 (62.7%) | 84 (37.3%) | 225 |
| | Low Craving | 55 (20.4%) | 215 (79.6%) | 270 |
| Total | | 196 | 299 | 495 |

Table 3. Confusion Matrix for Low Vulnerable hour. Overall Accuracy is 72.2% (against base accuracy 53.7%) with kappa 0.424.

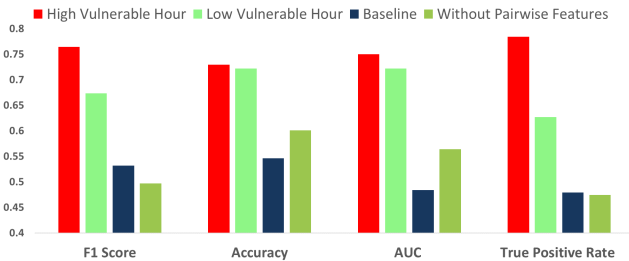


Figure 9. Model performance metrics for High Vulnerable, Low Vulnerable Hour, Baseline, Without *Pairwise* Feature

able hours is better than during low vulnerable hours (further illustrated in Figure 9), which agrees with our hypothesis that the high vulnerable hours are marked by a stronger connection between stress and craving, which in turn leads to better classification of craving self-reports conditioned to stress.

Utility of the *pairwise* feature

As alluded to earlier, the fact that we can incorporate ‘pairwise’ features, targeting probabilistic dependency patterns between successive craving labels, was a major factor in choosing CRFs as our model. The decision to include these pairwise features was based on the reasoning that high craving during a given minute is generally followed by high craving during the next minute, since craving dissipates gradually over time. Similarly, low craving during a minute is normally followed by low craving in the next minute. In order to test this idea and assess the discriminative power of the pairwise feature, we test our model after removing feature $f_A(C, \hat{p}_c, S_z, i)$. We observed that the model’s average performance over the high and low vulnerable hours drops significantly, as shown in Figure 9. This justifies our decision to include these pairwise probabilistic dependency patterns.

Comparison to Baseline

Finally, since there does not exist any prior model for craving estimation to which we can compare the performance of *mCrave*, we construct a likely candidate model. Since stress is a significant component of the *mCrave* model, we use *cStress* as a baseline. In other words, if the output of stress model was used directly as a surrogate of craving, how well will this model perform in comparison with our *mCrave* model.

Figure 9 shows the performance of this baseline model. We found that craving estimation directly from stress likelihoods attained an accuracy of 54.6% with a kappa score of 0.069, F_1 score of 0.532, recall of 0.479, and Area-Under-Curve (AUC) of 0.484. In comparison, the *mCrave* model performed significantly better (see Figure 9), which demonstrates the utility of *mCrave* modeling.

DISCUSSIONS, LIMITATIONS AND FUTURE WORK

This work makes several interesting observations. First, it showed that using only a few measures (i.e., stress and time of day), it is feasible to estimate craving. Second, it showed that stress, by itself, can’t be used as a surrogate of craving; it must be used in conjunction with time of day in an appropriate model to produce a good estimate of craving. Third, it showed that the accuracy of estimating craving does depend on time of day, i.e., higher accuracy may be obtained for high-vulnerable hours, when participants are more likely to lapse.

Limitations. Since this work is a first step towards estimating craving, it has several limitations that present exciting opportunities for future research in both the UbiComp and health research communities.

First, since craving is a psychological construct, estimating it from mobile sensor data is inherently difficult. Complicating it further is the challenge of obtaining reliable labels that can be used for training and testing of the model. Often, self-report is the only feasible label that can be obtained conveniently from the field setting.

Whenever sensor data is used to model the perception of a subjective phenomenon such as craving (especially at the minute-level granularity), there are limits to the level of accuracy that can be achieved from such models. This is due to inherent variabilities in the self-report. In contrast to objective phenomena such as physical activity, eating events, smoking events, etc., for which accuracies in the upper nineties can be expected, accuracy for subjective phenomenon modeling is usually limited to seventies. This is because the consistency among self-report items for the same construct are limited to lower eighties. Therefore, a consistency score of 0.7 or above among self-reported items are considered to be good [7]. In particular, [33] found that consistency among self-reported stress items was 0.76 in our dataset. The performance of our *mCrave* model should be viewed from this perspective.

Second, our model uses only stress and time of day. But, craving may be a result of other biological phenomena or due to cue exposure. Nicotine deprivation throughout the night may result in morning craving. Similarly, craving after lunch may be habitual, situational, or biological. In other cases, exposure to alcohol, seeing someone else smoking, smelling smoke, seeing a cigarette pack, or reading a social media message may trigger craving. Incorporating this information in the model can potentially improve the model’s accuracy. Hence, future smoking studies that incorporate other sensing modalities to collect geolocation (from GPS), digital exposure (from social media, calendars etc.), visual exposure (from smart eyeglasses), detection of eating from hand gestures (e.g., to know when lunch is over), etc. can assess the utility of these new data sources in improving the estimation of craving, especially in the low vulnerable hours.

Additionally, user demographics may be associated with craving and smoking behavior. For instance, female smokers show high craving reactivity to smoking-related cues relative to male smokers [31] while another EMA-based study

revealed that black smokers report greater levels of craving during the day than white smokers [8]. These suggest that more research is needed to carefully assess demographic influences on craving during abstinence.

Third, in this smoking cessation study, the duration of post-quit was chosen to be 3 days. This is because the first 3 days are the most critical days in smoking cessation, which captures the most intense withdrawal symptoms in abstinent smokers. Majority of participants lapse during this period; in our smoking cessation study, 53% lapsed in the first 3 days. Since this was the first smoking cessation study with continuous monitoring using physiological sensors (for stress assessment) and smartwatches (for lapse detection), the duration of the study was limited to ensure successful data capture in the most vulnerable abstinence period.

Now that the feasibility of capturing both stress and detection of smoking lapse from sensors in a real-life smoking cessation study has been established, longer studies can be pursued in future that can include these and other sensors to capture richer data sets. A longer study is likely to lead to new insights and improvements in the craving estimation modeling. It can be helpful in revealing several additional information such as trends in craving as withdrawal symptoms recede (for those who continue to abstain beyond the first 3 days).

Fourth, although we demonstrate feasibility of estimating craving, the model is not perfect and is unlikely to be perfect even in the future. It can both miss high-craving episodes as well as detect false ones. This may be due to factors such as errors/noise in the self-reported ground-truth craving labels, errors in the stress time series, as well as confounding variables, such as location, food/drink intake and others. Even with further improvements in the model, some inaccuracies may be inherent due to use of machine learning modeling approaches (for example, the fact that we have only a sparse sampling of ground-truth labels for all minutes of the day, we train the model by partial likelihood maximization, which may have played a role in lowering the model's generalization performance) and due to inherent between-person and between-situation variabilities. Therefore, any intervention delivery that are based on craving estimation models must deal with these inaccuracies.

Fifth, even if craving can be estimated perfectly, smoking lapse can sometimes occur without being preceded by craving, e.g., when offered cigarette by a friend, relative, or a colleague. Therefore, not all smoking lapse episodes can be prevented by intervening at all high-craving moments, which itself may be infeasible if they are too frequent in the day. What the model promises to do, instead, is to help prepare an individual to better tolerate craving in the abstinence period and increase the chances of remaining abstinent or delaying lapse. Once an individual acquires sufficient self-efficacy or tolerance to potent cues and biologically triggered craving episodes, they may become more likely to remain abstinent.

Future Research. In addition to addressing limitations and improving the model, this work presents other opportunities

for future research. A major research direction is to incorporate the *mCrave* model in smoking cessation interventions.

First, majority of interventions for smoking cessation are developed for delivery upon request, or at set times. These interventions may need to be revised or adapted for delivery in response to sensor-detected craving episodes.

Second, to become widely useful in the society, the clinical utility of *mCrave* model in the management of craving should be established by developing and evaluating sensor-triggered just-in-time mobile intervention via randomized clinical trials that can be triggered based on the estimation of craving. Issues to be considered include the interruption-like nature of these interventions, inaccuracy of sensor detections, opportunity for personalization by making use of sensor data to estimate user's availability [34], and others.

Third, methods need to be developed to deal with the high-frequency of sensor outputs (i.e., each minute) to find the most opportune moments to intervene. Some recent research in the context of stress intervention provides some initial foundations for this direction [35].

Fourth, effective visualizations could be developed to help the users visualize their craving patterns and gain useful insights into the associated contexts that might increase or decrease their craving during abstinence.

CONCLUSION

This work presented a computational model to estimate craving continuously from mobile physiological sensors. Use of an explainable model in *mCrave* helps retain the insights regarding the phenomena and get confirmation from domain experts. Doing so is critical to bridging the gap in such multi-disciplinary works. Although numerous improvements can be made to improve the sensitivity and specificity of the model (e.g., by incorporating geoexposure, visual exposure, and digital exposure data), establishing the feasibility of automatically estimating craving opens up numerous exciting research opportunities with significant potential to improve health and wellness. Development of appropriate interventions (potentially using assisting technologies such as smartphones, wearables, smart clothing, etc.) that are suitable for delivering at vulnerable moments can help build tolerance and improve smoking cessation success rates. Since craving plays an important role in several impulsive behaviors such as overeating, drinking, and drug use, research can be pursued to estimate craving for these behaviors as well.

ACKNOWLEDGMENTS

We thank Shahin Alan Samiei and Barbara Burch Kuhn from University of Memphis. We also thank study coordinators at Universities of Minnesota and Memphis. The authors acknowledge support by the National Science Foundation under award numbers CNS-1212901 and IIS-1231754 and by the National Institutes of Health under grants R01CA190329, R01MD010362, R01DA035502 (by NIDA) through funds provided by the trans-NIH OppNet initiative, and U54EB020404 (by NIBIB) through funds provided by the trans-NIH Big Data-to-Knowledge (BD2K) initiative.

REFERENCES

1. Centers for Disease Control and Prevention, Health Effects of Cigarette Smoking - Fact Sheet. http://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/, Accessed: March 2016.
2. Centers for Disease Control and Prevention, Health Effects of Cigarette Smoking - Fact Sheet. <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6044a2.htm/>, Accessed: June 2016.
3. World Health Organization, Tobacco - Fact Sheet. <http://www.who.int/mediacentre/factsheets/fs339/en/>, Accessed: March 2016.
4. al'Absi, M., Hatsukami, D., and Davis, G. L. Attenuated adrenocorticotrophic responses to psychological stress are associated with early smoking relapse. *Psychopharmacology* 181, 1 (2005), 107–117.
5. al'Absi, M., Hatsukami, D., Davis, G. L., and Wittmers, L. E. Prospective examination of effects of smoking abstinence on cortisol and withdrawal symptoms as predictors of early smoking relapse. *Drug and alcohol dependence* 73, 3 (2004), 267–278.
6. Allen, S. S., Bade, T., Hatsukami, D., and Center, B. Craving, withdrawal, and smoking urges on days immediately prior to smoking relapse. *Nicotine & Tobacco Research* 10, 1 (2008), 35–45.
7. Bland, J. M., and Altman, D. G. Statistics notes: Cronbach's alpha. *Bmj* 314, 7080 (1997), 572.
8. Carter, B. L., Paris, M. M., Lam, C. Y., Robinson, J. D., Traylor, A. C., Waters, A. J., Wetter, D. W., and Cinciripini, P. M. Real-time craving differences between black and white smokers. *The American Journal on Addictions* 19, 2 (2010), 136–140.
9. Doherty, K., Kinnunen, T., Militello, F. S., and Garvey, A. J. Urges to smoke during the first month of abstinence: relationship to relapse and predictors. *Psychopharmacology* 119, 2 (1995), 171–178.
10. Efron, B., and Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science* (1986), 54–75.
11. Ertin, E., Stohs, N., Kumar, S., Raji, A., al'Absi, M., and Shah, S. Autosense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field. In *ACM SenSys* (2011), 274–287.
12. Hammersley, R. A digest of memory phenomena for addiction research. *Addiction* 89, 3 (1994), 283–293.
13. Hovsepian, K., al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M., and Kumar, S. cstress: towards a gold standard for continuous stress assessment in the mobile environment. In *ACM UbiComp* (2015), 493–504.
14. Hughes, J. R., Keely, J., and Naud, S. Shape of the relapse curve and long-term abstinence among untreated smokers. *Addiction* 99, 1 (2004), 29–38.
15. Kassel, J. D., and Shiffman, S. What can hunger teach us about drug craving? a comparative analysis of the two constructs. *Advances in Behaviour Research and Therapy* 14, 3 (1992), 141–167.
16. Kenford, S. L., Fiore, M. C., Jorenby, D. E., Smith, S. S., Wetter, D., and Baker, T. B. Predicting smoking cessation: who will quit with and without the nicotine patch. *Jama* 271, 8 (1994), 589–594.
17. Killen, J. D., Fortmann, S. P., Kraemer, H. C., Varady, A., and Newman, B. Who will relapse? symptoms of nicotine dependence predict long-term relapse after smoking cessation. *Journal of consulting and clinical psychology* 60, 5 (1992), 797.
18. Killen, J. D., Fortmann, S. P., Newman, B., and Varady, A. Prospective study of factors influencing the development of craving associated with smoking cessation. *Psychopharmacology* 105, 2 (1991), 191–196.
19. Kozlowski, L. T., and Wilkinson, D. A. Use and misuse of the concept of craving by alcohol, tobacco, and drug researchers. *British journal of addiction* 82, 1 (1987), 31–36.
20. Kumar, S., Nilson, W., Pavel, M., and Srivastava, M. Mobile health: Revolutionizing healthcare through trans-disciplinary research. *IEEE Computer* 46, 1 (2013), 28–35.
21. McClernon, F. J., Hiott, F. B., Huettel, S. A., and Rose, J. E. Abstinence-induced changes in self-report craving correlate with event-related fmri responses to smoking cues. *Neuropsychopharmacology* 30, 10 (2005), 1940–1947.
22. McEwen, B. S. Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences* 840, 1 (1998), 33–44.
23. McEwen, B. S. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological reviews* 87, 3 (2007), 873–904.
24. McKee, S. A., Sinha, R., Weinberger, A. H., Sofuoglu, M., Harrison, E. L., Lavery, M., and Wanzel, J. Stress decreases the ability to resist smoking and potentiates smoking intensity and reward. *Journal of psychopharmacology* 25, 4 (2011), 490–502.
25. Moore, T. M., Seavey, A., Ritter, K., McNulty, J. K., Gordon, K. C., and Stuart, G. L. Ecological momentary assessment of the effects of craving and affect on risk for relapse during substance abuse treatment. *Psychology of Addictive Behaviors* 28, 2 (2014), 619.
26. Perkins, K. A., Briski, J., Fonte, C., Scott, J., and Lerman, C. Severity of tobacco abstinence symptoms varies by time of day. *Nicotine & Tobacco Research* 11, 1 (2009), 84–91.
27. Piasecki, T. M., Fiore, M. C., and Baker, T. B. Profiles in discouragement: two studies of variability in the time course of smoking withdrawal symptoms. *Journal of abnormal psychology* 107, 2 (1998), 238.

28. Plarre, K., Raji, A., Hossain, S., Ali, A., Nakajima, M., Al'absi, M., Ertin, E., Kamarck, T., Kumar, S., Scott, M., et al. Continuous inference of psychological stress from sensory measurements collected in the natural environment. In *IEEE/ACM IPSN* (2011), 97–108.
29. Plarre, K., Raji, A., Hossain, S., Ali, A., Nakajima, M., al'Absi, M., Ertin, E., Kamarck, T., Kumar, S., Scott, M., Siewiorek, D., Smailagic, A., and Wittmers, L. E. Continuous Inference of Psychological Stress From Sensory Measurements Collected in the Natural Environment. In *ACM/IEEE IPSN* (2011).
30. Poh, M.-Z., Loddenkemper, T., Swenson, N. C., Goyal, S., Madsen, J. R., and Picard, R. W. Continuous monitoring of electrodermal activity during epileptic seizures using a wearable sensor. In *IEEE EMBC* (2010), 4415–4418.
31. Saladin, M. E., Gray, K. M., Carpenter, M. J., LaRowe, S. D., DeSantis, S. M., and Upadhyaya, H. P. Gender differences in craving and cue reactivity to smoking and negative affect/stress cues. *The American Journal on Addictions* 21, 3 (2012), 210–220.
32. Saleheen, N., Ali, A. A., Hossain, S. M., Sarker, H., Chatterjee, S., Marlin, B., Ertin, E., al'Absi, M., and Kumar, S. puffmarker: a multi-sensor approach for pinpointing the timing of first lapse in smoking cessation. In *ACM UbiComp* (2015), 999–1010.
33. Sarker, H., Hovsepian, K., Nahum-Shani, I., Murphy, S. A., Spring, B., Ertin, E., al'Absi, M., Nakajima, M., and Kumar, S. From markers to interventions—the case of just-in-time stress intervention. In *Mobile Health: Sensors, Analytic Methods, and Applications*, J. Rehg, S. A. Murphy, and S. Kumar, Eds. Springer, 2016.
34. Sarker, H., Sharmin, M., Ali, A., Rahman, M., Bari, R., Hossain, S., and Kumar, S. Assessing the availability of users to engage in just-in-time intervention in the natural environment. In *ACM UbiComp* (2014), 909–920.
35. Sarker, H., Tyburski, M., Rahman, M., Hovsepian, K., Sharmin, M., Epstein, D. H., Preston, K. L., Furr-Holden, C. D., Milam, A., Nahum-Shani, I., al'Absi, M., and Kumar, S. Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data. In *ACM CHI* (2016).
36. Shiffman, S. Dynamic influences on smoking relapse process. *Journal of personality* 73, 6 (2005), 1715–1748.
37. Shiffman, S. Reflections on smoking relapse research. *Drug and alcohol review* 25, 1 (2006), 15–20.
38. Shiffman, S., Engberg, J. B., Paty, J. A., Perz, W. G., Gnys, M., Kassel, J. D., and Hickcox, M. A day at a time: predicting smoking lapse from daily urge. *Journal of Abnormal psychology* 106, 1 (1997), 104.
39. Shiffman, S., Gnys, M., Richards, T. J., Paty, J. A., Hickcox, M., and Kassel, J. D. Temptations to smoke after quitting: a comparison of lapsers and maintainers. *Health Psychology* 15, 6 (1996), 455.
40. Shiffman, S., Paty, J. A., Gnys, M., Kassel, J. A., and Hickcox, M. First lapses to smoking: within-subjects analysis of real-time reports. *Journal of consulting and clinical psychology* 64, 2 (1996), 366.
41. Shiffman, S., Stone, A. A., and Hufford, M. R. Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* 4 (2008), 1–32.
42. Shiffman, S., and Waters, A. J. Negative affect and smoking lapses: a prospective analysis. *Journal of consulting and clinical psychology* 72, 2 (2004), 192.
43. Shiffman, S., West, R. J., and Gilbert, D. G. Recommendation for the assessment of tobacco craving and withdrawal in smoking cessation trials. *Nicotine & Tobacco Research* 6, 4 (2004), 599–614.
44. Stone, A. A., and Shiffman, S. Ecological momentary assessment (ema) in behavioral medicine. *Annals of Behavioral Medicine* (1994).
45. Taylor, B., Dey, A., Siewiorek, D., and Smailagic, A. Using physiological sensors to detect levels of user frustration induced by system delays. In *ACM UbiComp* (2015), 517–528.
46. Waters, A. J., Shiffman, S., Sayette, M. A., Paty, J. A., Gwaltney, C. J., and Balabanis, M. H. Cue-provoked craving and nicotine replacement therapy in smoking cessation. *Journal of consulting and clinical psychology* 72, 6 (2004), 1136.
47. West, R. Use and misuse of craving.
48. West, R., Hajek, P., and Belcher, M. Severity of withdrawal symptoms as a predictor of outcome of an attempt to quit smoking. *Psychological medicine* 19, 04 (1989), 981–985.
49. West, R., Hajek, P., and Belcher, M. Time course of cigarette withdrawal symptoms while using nicotine gum. *Psychopharmacology* 99, 1 (1989), 143–145.