puffMarker: A Multi-Sensor Approach for Pinpointing the Timing of First Lapse in Smoking Cessation

Nazir Saleheen[⋆], Amin Ahsan Ali[▽], Syed Monowar Hossain[⋆], Hillol Sarker[⋆], Soujanya Chatterjee[⋆]
Benjamin Marlin[∨], Emre Ertin[◇], Mustafa al'Absi[†], Santosh Kumar[⋆]

University of Memphis[⋆] University of Dhaka[▽] University of Massachusetts, Amherst[∨]
The Ohio State University[◇] University of Minnesota Medical School[†]
{nsleheen, smhssain, hsarker, schttrj1, skumar4}@memphis.edu[⋆] aminali@du.ac.bd[▽]
marlin@cs.umass.edu[∨] ertin.1@osu.edu[◇] malabsi@d.umn.edu[†]

ABSTRACT

Recent researches have demonstrated the feasibility of detecting smoking from wearable sensors, but their performance on real-life smoking lapse detection is unknown. In this paper, we propose a new model and evaluate its performance on 61 newly abstinent smokers for detecting a first lapse. We use two wearable sensors — breathing pattern from respiration and arm movements from 6-axis inertial sensors worn on wrists. In 10-fold cross-validation on 40 hours of training data from 6 daily smokers, our model achieves a recall rate of 96.9%, for a false positive rate of 1.1%. When our model is applied to 3 days of post-quit data from 32 lapsers, it correctly pinpoints the timing of first lapse in 28 participants. Only 2 false episodes are detected on 20 abstinent days of these participants. When tested on 84 abstinent days from 28 abstainers, the false episode per day is limited to 1/6.

Author Keywords

Mobile health (mHealth); wearable sensors; smartwatch, smoking detection; smoking cessation

ACM Classification Keywords

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

INTRODUCTION

Smoking accounts for nearly one of every five deaths in the United States [1, 14]. Smoking cessation rates are improving with advances in treatments and interventions, but are still in single digits. A primary hurdle in achieving a higher success rate is a lack of methods that can intervene or deliver treatment at the right moment when an abstinent smoker is most vulnerable [25]. Advances in mobile technology have created an opportunity to deliver an intervention anytime and anywhere if a potential smoking lapse event can be predicted in advance. However, to find sensor-based predictors of a smoking lapse [4, 8, 5, 7, 10, 13, 21, 22, 26, 27] (e.g., rapid rise in stress [24] or proximity to a tobacco outlet), timing of a lapse event (especially the first lapse event, the most clinically relevant event as it usually leads to full relapse [22]) needs to be determined accurately. Then, data mining methods can be applied on the time series of mobile sensor data to identify the antecedents and precipitants of a smoking lapse (in a smoking cessation study). The traditional method of self-reporting a smoking lapse event [3, 9, 22, 23] lacks the temporal precision needed to identify predictors in a continuous stream of sensor data.

There have been several recent works on detecting a smoking episode from wearable sensors. They include tracking hand gestures during smoking by inertial sensors worn on the wrist [17, 28] and tracking deep inhalation and exhalation in the breathing pattern via Respiratory Inductive Plethysmography (RIP) sensors [2, 11]. But, their performance is reported mostly on training data collected in supervised settings, with the exception of RisQ [17], which was tested on 4 smokers who wore 9-axis inertial sensors for 4 hours a day over 3 days. They report detection of 27 smoking episodes (out of 30) and report a false positive rate of 2/3 per day (8 sessions out of 12 person days).

While RisQ is the most promising smoking detection method, it uses a 9-axis wrist sensor, whereas most modern smartwatches (e.g., Microsoft Band, Apple Watch) have only 6-axis inertial sensors, for better energy-efficiency. Since RisQ relies on quaternions that need all 9-axis, it is not clear how to adapt this method to 6 axis. Most importantly, none of the above described methods have been evaluated on data collected in a smoking cessation study and hence their performance for first lapse detection is not known.

In this paper, we propose a new method called *puffMarker* to detect smoking puffs that is sufficiently robust for use in smoking cessation studies. We adopt an explainable modeling approach so as to obtain better interpretability and generalizability. *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. Since a participant may use both hands to smoke, they are provided two wrist sensors to wear, one on each wrist. Both sensors nicely complement each other and hence provide a better detection accuracy. To provide an intuition of the benefit of using these two diverse sensors, we show signals captured during smoking, walking, and eating in Figure 1. We only

UbiComp '15, September 7–11, 2015, Osaka, Japan.

ACM 978-1-4503-3574-4/15/09...$15.00.
http://dx.doi.org/10.1145/2750858.2806897

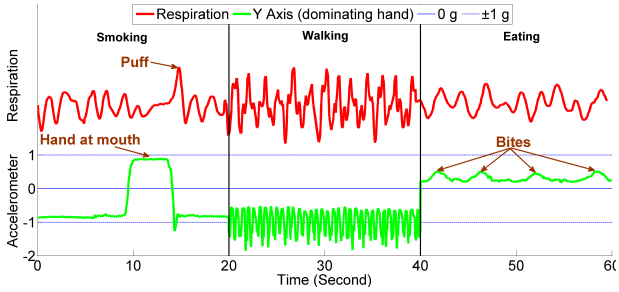

Figure 1. Comparison of respiration and wrist accelerometer (y-axis) signal between smoking, walking, and eating.

show y-axis for the inertial sensors since it has distinct pattern for hand gesture. We observe that during smoking, the hand comes to the mouth, and is immediately followed by a deep inhalation. During walking, the hand is downwards with a pendulum like movement and respiration is faster. During eating (cereal), the hand comes at the mouth, however deep inhalation, observed during smoking, is absent in such activities [16]. We later describe details of the *puffMarker* model (see Figure 4 and its accompanying description).

We train the model on 40 hours of data from 6 regular smokers, where each of the 470 puffs were carefully marked. In 10-fold cross-validation on the training data, the model achieves a recall rate of 96.9%, for a false positive rate of 1.1%. We applied the *puffMarker* model to a smoking cessation study with 61 participants, where each participant wore the sensors for one day while smoking ad lib and for 3 days since quitting. Among 61 participants, 33 lapsed within three days (verified by a CO monitor) — 17 lapsed on the first day, 12 on the second day, and 4 on the third day. We apply our model on these data and report 7 key findings.

1. **Recall:** Among 33 lapsers, one is eliminated due to high data loss; Of the remaining 32, first lapse is detected in 28.
2. **False Positives:** When tested on 20 abstinent days from 32 lapsers, only two false episodes are detected. When tested on 84 abstinent days (946 hours) of data from 28 abstainers, false episode per day is limited to 1/6.
3. **Lapse Progression:** The average number of smoking episodes is 1.1 on the lapse day, 2.75 on the day after lapse, and 3.56 on 2 days after lapse.
4. **Puff Count:** A regular smoking session contains an average of 15 puffs, but the first lapse episode contains an average of only 6.5 puffs. Number of puffs in a smoking episode increases to 9.5 puffs on the day after lapse and 11 puffs on 2 days after lapse.
5. **Temporal Inaccuracy of Self-report:** Out of 28 first lapse events detected by *puffMarker*, 9 were not self-reported, 15 were reported (an average of 41 minutes) after lapse, and 4 were reported (an average of 12.7 minutes) before lapse.
6. **Recall Inaccuracy:** Nine lapsers who did not self-report, recalled the lapse time upon CO verification in the lab next day. Temporal inaccuracy for these recalls was larger, ranging from 107 minutes before to 205 minutes after.
7. **Hand Pattern:** Among 61 smokers, 33 smoke using the right hand, 18 use the left hand, and 10 use both hands.

In summary, ours is the first work to show that precise moment of first lapse can indeed be detected in a real-life smoking cessation study. Given the critical nature of the first lapse in smoking cessation (as it marks the first event in cessation failure and usually leads to full relapse [22]), this work lays the groundwork for development of mobile-based just in-time intervention for smoking cessation.

RELATED WORK

We discuss related works that could be considered for detecting first lapse in a smoking cessation study. There have been several recent works on finding methods to detect smoking episodes from sensor data. The E-cigarette can record the timing of puffs and smart lighters can detect when it is lit [20]. Either one can detect the timing of the first lapse, but only if participants remember to use these devices at the time of their first lapse. Alternatively, if the smoking spot is under video surveillance, then the timing of the first lapse can be detected [29], but only if the participant is under surveillance at the time of their first lapse.

For detecting smoking of regular cigarettes without any instrumentation of lighters or being under surveillance, several wearable sensor-based methods have been proposed. They include tracking hand gestures during smoking by inertial sensors worn on the wrist [17, 28] and tracking deep inhalation and exhalation in the breathing pattern via respiratory inductive plethysmography (RIP) sensors [2, 11].

In [2], 161 puffs were collected from 10 participants while they wore a respiration sensor to capture breathing pattern. A machine learning model obtained a precision of 0.91 and recall of 0.81 in 10-fold cross validation. In [11], 20 participants wore a radio frequency sensor on the wrist and on the collar to track hand reaching mouth. They also wore a respiration sensor to capture breathing pattern. They performed 12 activities, including smoking in different postures. An average precision of 0.87 and a recall of 0.81 was reported.

In [28], 6 participants wore 4 accelerometers (wrist and upper arm of dominant hand, other wrist and ankle) and performed smoking and other activities for a total of 11.8 hours (consisting of 34 smoking episodes or 481 puffs). Recall and precision rates of upper seventies and lower eighties is obtained for a machine learning model.

The above works demonstrate a potential for detection of smoking events via experimentation in controlled setting. A recent work, RisQ [17], reported evaluation of a smoking detection model from wrist sensors in the field environment. In this work, data from 15 volunteers were collected for a total of 17 smoking episodes for training. The smoking episodes included smoking alone, in a group while having a conversation and smoking while walking around. The volunteers wore a 9-axis inertial measurement unit (IMU) on the wrist for an average of 2 hours each. Out of 369 puffs and 5,228 other gestures collected over 28 hours, their model achieves a precision of 0.91 and recall of 0.81. They applied their model on 4 users who wore 9-axis inertial sensors for 4 hours each on 3 days in the field. On this field dataset, they reported a recall rate of 90% (27 out of 30 sessions detected) and a false

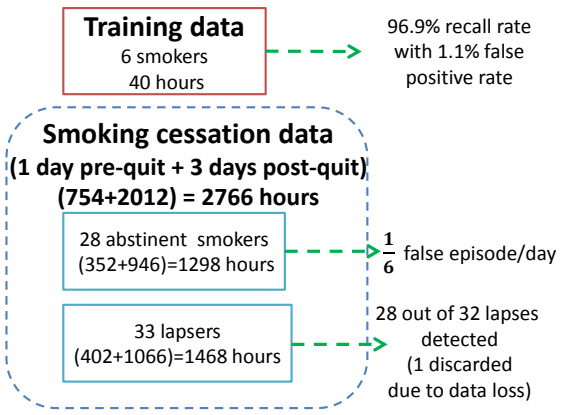

Figure 2. Data Summary of Training and Smoking Cessation Study

positive rate or 2/3 episodes per day (8 false sessions in 12 person days).

Although our work builds upon [17] and [2], it makes several novel contributions. First, RisQ [17] uses 9-axis wrist sensor (3-axis accelerometer, 3-axis gyroscope, 3-axis magnetometer), whereas our method uses 6 axis IMU (supported by many modern smartwatches such as Microsoft Band and Apple Watch). Since gesture recognition is invariant to the absolute orientation of the subject in the earths inertial frame, 6 axis IMU provides sufficient degrees of freedom. Second, prior work [17] relies on generic gesture recognition algorithms based on inertial tracking of the absolute orientation of the wrist, we developed a lightweight recognition algorithm tailored for the smoking gesture with much reduced computational complexity and sampling rate requirements. Third, mPuff [2] classified respiration cycles into puff and non-puff, but has a high false alarm rate; it falsely detects 150 out of 1,000 respiration cycles as puffs, making it ineffective for use in the natural environment. In contrast, our *puffMarker* model falsely detects only 1 out of 1,117 respiration cycles as puffs. Fourth, ours is the first work that was applied to data collected from a real-life smoking cessation study. All other prior works reported their results on only regular smoking training data. Fifth, our work is the first one to detect first lapses, which is most challenging due to significantly smaller number of puffs (45%). Sixth, RisQ and mPuff were both evaluated on only 4 users in the field environment while we evaluate on 61 users, making our work clearly the largest-ever study for sensor-based detection of smoking. Seventh, ours is the first work that combines respiration and wrist movement data and shows how inclusion of wrist movement detection can increase the performance of respiration based detector [2]. Finally, performance of our system (recall of 96.9% and false positive rate of 1.1%) is better than any previously reported work even on training data.

DATA COLLECTION

We describe details of the user study for collecting training data for the *puffMarker* model and the smoking cessation study where the *puffMarker* model was applied.

Wearable Sensor Suite: Participants in both studies wore a wireless physiological sensor suite (AutoSense [6]) 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 on each wrist that includes a 3-axis accelerometer and a 3-axis gyroscope. 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 for the respiration sensor is 21.3 Hz and that for the accelerometer and gyroscope on smartwatch are 16 Hz for each of the six axes.

Mobile Phone: Participants were given a smart phone to carry. It receives and stores data from sensors on the body and on the phone. It also collects self-reports in response to random prompts which capture characteristics of situational factors associated with smoking. These factors include stress, physical activity levels, posture, places visited, and commuting episodes. In the training study, an observer marks the timing of each puff on the phone. In the smoking cessation study, participants used the phone to report the beginning of smoking episodes by pressing a button.

Data Collection for Model Training

We collected data from 6 daily smokers. They wore the sensors for a total of 40.3 hours in field as they went about their daily lives. Each time they smoked, they were accompanied by an observer who who was instructed to mark (on study phone) a puff when the participant held cigarette between the lips and inhaled smoke. From the marking, we thus obtained the timing when the hand is at the mouth during smoking. This dataset contains 32 smoking episodes (that includes smoking while standing, sitting, walking, and being in a conversation) with 470 puff markings. In 179 instances out of the 470 puff markings, there were wireless data losses and noise due to physical movement or loosening of the respiration belt. We use the remaining 291 puff instances for which we have acceptable respiration and wrist sensor data.

Smoking Cessation Study

Participants: The participants were cigarette smokers who reported smoking 10 or more cigarettes per day for at least 2 years, and who reported high motivation to quit. To qualify, participants had to pass a screening session prior to being enrolled in the study. The screening includes assessment of current medical and mental health status and history of any major medical and psychiatric illness. Screening also includes 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 entertained into the normal day/light diurnal cycle were excluded to control for variation in diurnal physiological activity and behaviors.

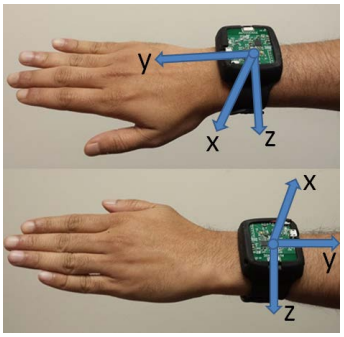

Figure 3. Depiction of mounting of inertial sensors on the wrist and the orientation of their axes.

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 hour monitoring, which we call the pre-quit session, subjects come back to the lab for their second visit. Smoking cessation counseling is provided starting at this second visit to the lab. Then the subjects come back to the lab on the assigned quit date to attend a counseling session and to begin the 72 hour of monitoring in the field; this we refer to as the post-quit session. They come back to the lab each day to confirm smoking status by capturing an expired breath sample in a carbon monoxide (CO) monitor. During each day of monitoring (24 hour pre-quit and 72 hour post-quit), the participants wear the sensor suite during awake hours, and on the mobile phone, complete 12 Ecological Momentary Assessments (EMAs) (i.e., self-reports) daily.

Data Collected: We collected data from 61 participants. The participants wore the sensor suite for a total of 2,766 hours. A summary of collected data is shown in Figure 2.

DATA PROCESSING AND MODEL DEVELOPMENT

Respiration, 3-axis accelerometers, and 3-axis gyroscopes provide us with 7 concurrent time series of data that are all time-stamped when they are received on the mobile phone[1]. The x, y and z axes of the accelerometers and gyroscopes on the wristband are aligned with each other. The directions of the axes of the wristband sensors are shown in Figure 3.

Overview of *puffMarker* Model

Figure 4 presents an overview of all the data analysis steps. First, we remove outliers and impute missing data. Second, we describe a method to detect hand-at-mouth gestures that segments (i.e., creates windows in) the time series of sensors data. These windows are assessed for representing a puff. Hand gesture during puffing a cigarette is typically composed

[1]We note that timestamping of sensor data upon receipt on the phone does not adversely affect time snchronization needed for our method. Data from sensors are transmitted to the phone tens of times each second and therefore the delay from sampling to reception on the phone is of the order of milliseconds. Our time granularity requirements for classification (discussed later) is of the order of 3-7 seconds, duration of hand staying at the mouth. Each respiration cycle is 3-6 seconds long. Hence, millisecond level errors in time synchronization between respiration and inertial sensor data does not adversely affect our model.

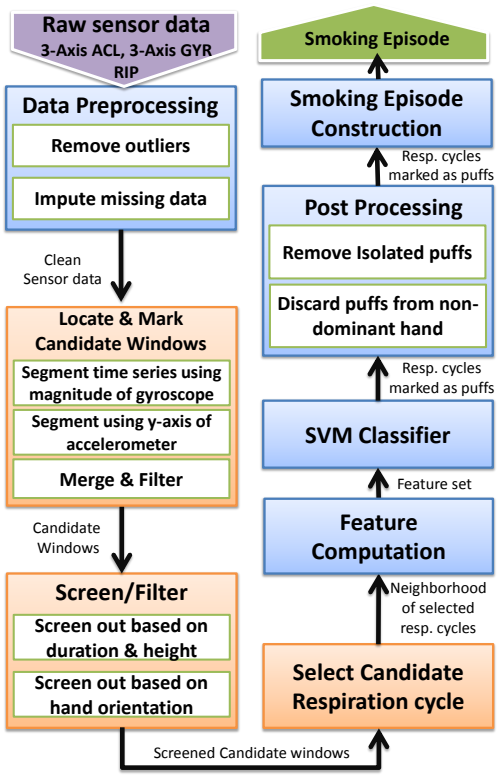

Figure 4. An overview of key modeling steps for detecting smoking puffs and constructing smoking episodes in the *puffMarker* model.

of three sub-gestures that usually occur in the following sequence — hand moving to the mouth, hand being at mouth while taking a puff, and hand moving away from the mouth.

puffMarker locates puff events by detecting the segments in the wrist sensor time series that contain hand-at-mouth gestures. When an accelerometer is stationary, any accelerometer axis aligned precisely with the earth's downward gravitational field will result in a measurement of $-1g$ in that axis. Due to the mounting of the sensor on the wrist (see Figure 3), we observe a positive value on y-axis when the right hand is held in an upward direction, whereas we get a negative value on the y-axis of the left wrist accelerometer.

However, when an accelerometer is moving, an axis will measure the combination of linear acceleration due to movement and the component of gravity in the direction of that axis. The gyroscope axes, on the other hand, measure the rate of change of rotation or angular velocity about the axes. We use the measurements obtained from both the accelerometers and gyroscope to detect hand-at-mouth gestures.

As a third step, we develop criteria to screen out segments (or windows) of data that do not correspond to smoking puffs. Using the detected hand-at-mouth segments, we identify the accompanying respiration cycles that potentially correspond to smoking puffs.

Fourth, we compute features from those candidates' respiration cycles and hand-at-mouth segments that pass the preceding criteria. For classification, we train a support vector ma-

chine model that uses the input features to classify the qualifying candidate windows of data into *puff* and *non-puff*.

Fifth, we apply two simple post-processing steps to further reduce false alarms. Isolated puffs that do not fit within the distribution of inter-puff duration in a regular smoking episode are filtered out. Also, the model is applied to data collected from each wrist, but if a participant is observed to use only one hand for smoking during the pre-quit phase, then smoking puffs detected from non-dominant hand are filtered out. Finally, a smoking episode is constructed if sufficient number of smoking puffs (in a cluster) are detected in close vicinity.

Data Preprocessing

Sensor data are collected in the field environment where they are subjected to various sources of noises, losses, and degradation in quality. We develop a series of screening methods to clean the sensor data.

Removing outliers: Windows of sensor data when a participant is not wearing the sensors (determined using methods presented in [19]) are removed from analysis.

Imputation of missing data: In a real-life environment, some data packets are lost in wireless transmission. We interpolate if sample loss is limited to one packet containing 5 samples as described in [19]. Any longer loss burst is not imputed to maintain data quality. After outlier removal and imputation, the data is ready for further processing.

Locating and Marking Windows of Interest

We observe that during cigarette smoke inhalation, the hand must remain stationary at the mouth for few seconds. We look for these stationary moments in the inertial sensor data stream. We argue that for detecting hand-at-mouth gestures, this method of locating puffs is more robust than tracking the hand trajectory, which can vary widely depending on the body posture (i.e., sitting, standing, walking) and hand position.

In order to detect hand-at-mouth gestures, we first segment the sensor data from both wrists to find relatively stationary segments and discard all non-stationary segments. To find the body location when the hand is relatively stationary, we use the magnitude of the gyroscope axes. Any movement of the hand will manifest as a rotation about one of the gyroscope axes and cause the magnitude value to increase independent of the direction of rotation. When there is very little movement of the hand, the gyroscope magnitude will be low since it is not affected by gravity. Therefore, a hand-at-mouth gesture can be detected by finding segments where the gyroscope magnitude time series attains low values and is preceded and followed by peaks. The first peak is due to the hand moving towards the mouth and the second one is due to the hand moving away from the mouth.

It should, however, be noted that simple thresholding on the magnitude values to locate hand-at-mouth segments may not work well in practice. The average magnitude of a hand-at-mouth segment during walking is usually higher than that of stationary segments during standing or sitting. This is because, when a person is walking, the whole body is moving and there will always be some movement of the hands. Even

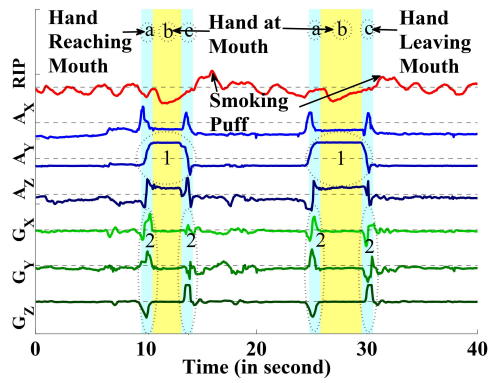

Figure 5. Hand at mouth segment detection. (A_X, A_Y, A_Z) and (G_X, G_Y, G_Z) present the signals of accelerometers and gyroscopes. The circled area 1 represents the effect of y-axis of accelerometer when hand is at mouth and the circled area 2 represents the changes in gyroscope when hand is reaching the mouth and hand is leaving the mouth.

when the hand is at mouth while taking a puff, there is some movement of the wrists due to taking steps.

Therefore, for the hand-at-mouth gestures during walking, we expect to find segments that attain low magnitude values compared to the average magnitude during walking. Moreover, in several instances, we observe that these relatively low amplitude values are sometimes higher than the amplitude of peaks corresponding to the hand movements before and after the hand-at-mouth stationary segments. The amplitude of these peaks depend on the rest position of the hand before and after the puff.

While standing or walking, the hand usually hangs beside the body and, therefore, we observe a larger peak amplitude. On the other hand, while sitting, the hands may be resting on the thighs, or on the armrest of a chair. In these cases, the hand is moving a shorter distance to the mouth and hence the peak amplitudes are lower. Also, sometimes, participants smoke with their hand hanging near the mouth. In such instances, the amount of movement is even lower, producing the lowest peak amplitudes. Therefore, setting the threshold value as high as the amplitude of hand-at-mouth segments during walking does not correctly identify hand-at-mouth segments in several other situations. This necessitates a segmentation method that is adaptive to the current level of movement so as to find segments that are relatively stationary.

We use a procedure that makes use of two moving averages to detect rise and fall in the gyroscope magnitude time series. Such methods are commonly used by investors in stock markets to identify price rise and fall [15]. More specifically, we perform the following steps.

1. We compute a fast (0.8 second window) and a slow (8 second window) moving average of the gyroscope magnitude. The fast moving average closely follows the dynamic nature of the signal, while the slow moving average represents the level of movement in the neighborhood (see Figure 5). The window size of the slow moving average corresponds to the length of a smoking hand gesture length, which usually lasts from 3-7 seconds [12]. Because of this

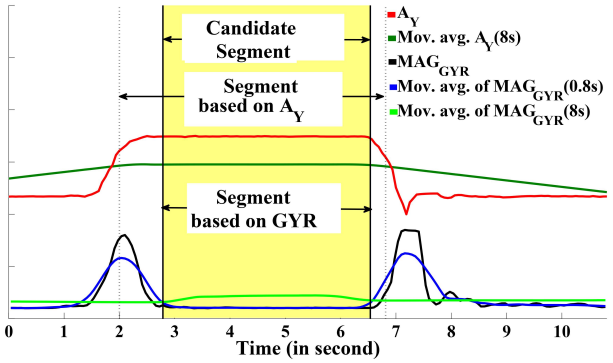

Figure 6. Locating a candidate segment and identifying its boundaries.

window size, during smoking hand-at-mouth gestures, the slow moving average is computed over windows that always contain the peaks due to the hand moving towards and away from the mouth. Therefore, during these segments, the slow moving average attains higher values compared to the fast moving average. The 0.8 second window size for the fast moving average is empirically chosen so that all segments that contain hand-at-mouth gestures during puffs are detected.

2. We select segments where the fast moving average lies below the slow moving average. These segments are demarcated by two consecutive crossing-over points of the two moving averages. The first of them corresponds to the location after which the fast moving average moves below the slower one and the second crossing-over point corresponds to the location after which the fast moving average rises above the slower one. The selected segments are, therefore, the ones where the magnitude values are relatively lower than the average magnitude of the neighborhood.

While segmenting data using the above method, we also check whether at that time y-axis of accelerometer is in an upward direction or not. Change in the magnitude of the gyroscope indicates that the hand is in motion. If, at that moment, the y-axis of the accelerometer changes from low to high (for the right hand sensor), it indicates the hand is moving in an upward direction. An opposite change in the y-axis of the accelerometer at the end of a gesture segment indicates the hand is moving in a downward direction. Figure 6 shows how segmentation is performed using both gyroscope and y-axis of accelerometer. This method of segment identification is designed to include all segments that are likely to contain a hand-at-mouth gesture, but may be over-inclusive.

Data Reduction via Non-candidate Segment Exclusion

We narrow down our search space by excluding hand-to-mouth segments in the inertial sensor data stream that are unlikely to represent a puff. We employ three methods, all of which are trained (i.e., determining parameters) using training data where each puff was carefully marked. In this dataset, we refer to inertial data segments that correspond to hand-at-mouth gestures when taking a puff as *puff segments*.

Appropriate Degree of Movement? We compute the mean difference between the fast and slow moving averages for each of the *puff segments*. Using the minimum of these distances as a threshold we discard all segments for which the mean distance is lower than this threshold (50 degree/second).

Appropriate Duration of Hand-to-mouth Gesture? We determine the duration of each *puff segments*. Inertial data segments that have a duration more that 3 standard deviations away from the mean duration (less than 0.8 seconds or greater than 5 seconds) are excluded.

Proper Hand Orientation When The Hand At Mouth? When the hand is at or near the mouth, it may not be for taking a puff, such as when touching the hair, fixing eyeglasses, yawning, etc. To determine whether the hand is properly oriented during a hand-to-mouth gesture (e.g., as it usually is when taking a puff), we determine orientation of the hand by computing the pitch and roll angles.

The pitch and roll angles at a particular orientation indicate the amount of rotation about the x and y axis respectively required to reach the particular orientation from an initial orientation. We assume that in the initial orientation, the hand is kept horizontal with the palm facing down (z-axis is aligned with the gravitational field). Roll and pitch angles can be computed from either accelerometer or gyroscope. However, in the presence of linear acceleration, the orientation angles computed from accelerometers are usually less accurate. On the other hand, roll and pitch angles can be computed by integrating the angular velocity measurements obtained from gyroscope. However, a small error in angular velocity measurement may lead to large integration errors.

Since we are interested in computing the orientation at times when the hand is relatively stationary, relying on only accelerometer measurements is sufficiently accurate for our purpose. For each segment, we compute the average of each axis forming the vector (a_x, a_y, a_z). Following the method proposed in [18], we compute pitch(θ) = $(-a_y)/(-a_z)$ and roll(ϕ) = $a_x/\sqrt{{a_y}^2 + {a_z}^2}$.

We note that pitch and roll may suffer from gimbal lock, which refers to the lack of bi-continuous map between the spherical coordinates (pitch/roll) and the torus surface of the rotations. As far as the classification of the static orientation is concerned, the lack of continuous map is not an issue, the pitch/roll still uniquely defines the orientation and therefore each static orientation can be uniquely classified. The problem appears when dynamic measurements are averaged to get an average orientation in a window of data. This could be remedied going to an intermediate redundant representation such as quaternions to implement the averaging. We have chosen a simpler approach to address this problem. We use a window of measurements for computing pitch and roll. From the window, we only take those values that do not suffer from Gimbal lock. Since our window consists of at least 0.8 seconds worth of samples, most of our windows have valid measurements for pitch and roll even if some values are momentarily affected by Gimbal lock due to orientation alignment.

We next describe our method for handling change in hands (left vs. right). Since the direction of the x-axis (and y-axis) of the accelerometer on the left and right wrists are opposite

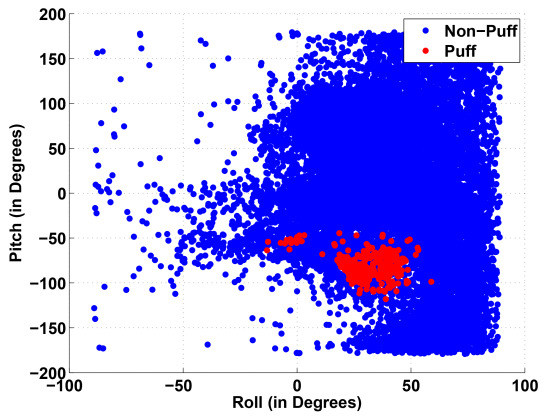

Figure 7. Scatter plot of roll and pitch angles for the *puff segments* and non-*puff segments*.

to each other, we first negate the x and y axis measurements of the left wrist sensor. In this way, the corresponding axes point in the same directions in both wrist sensors. By convention, the roll and pitch angles are positive when there is clockwise rotation about the y-axis and x-axis respectively. Therefore, when either arm is lifted from the horizontal position, pitch angles will have the same negative sign. However, inward (or outward) rotation of the left hand is in the opposite direction for an inward (or outward) rotation of the right hand. Therefore, in order to obtain the same sign for roll angles in both hands (that are)mirror images of each other, we negate the sign of the roll angles obtained for the left hand.

To determine proper values of pitch and roll, we compute these values for each *puff segments* and for non-*puff segments* (see Figure 7 for a scatter plot). We observe that the roll and pitch angles are slightly correlated with each other. For each segment, we compute the Mahalanabis distance[2] from the distribution of roll and pitch angles of *puff segments*. Mahalanabis distance d is computed by $d = (\mathbf{x} - \mu)S^{-1}(\mathbf{x} - \mu)'$, where $\mathbf{x} = (x_{roll}, x_{pitch})$ is the vector representing the roll and pitch angles of a segment and μ and S are the mean vector and co-variance matrix computed from roll and pitch angles of *puff segments*. Since for *puff segments*, the distance should be lower than other segments, we find a threshold t_d so that all *puff segments* are below it. We set the value of t_d to the largest value from all *puff segments* from training data and use it as a threshold. We discard all segments that have distance greater than $t_d = 10.15$ degree square.

Candidate Respiration Cycle Selection

After filtering out all non-candidate segments, we identify a respiration cycle that corresponds to each candidate hand-to-mouth gesture segment. We find the respiration cycles by computing the peak and valley locations in the respiration signal. Respiration signal reaches the peak once smoke is completely inhaled and exhalation of smoke usually occurs once cigarette is removed from the mouth. For each candidate hand-at-mouth gesture segment detected, we, therefore, select the first respiration cycle whose peak occurs after the end of the segment. This respiration cycle is a candidate for puff, if respiration signal is missing in that segment we use the hand-at-mouth segment as candidate.

A respiration cycle, however, can be associated with two different candidate segments, one from each hand, whose end times are close to each other. To avoid the situation where the training data contains a puff and a non-puff instance that are both associated with the same respiration cycle but different candidate segments, we only consider the non-smoking regions of the dataset as the source of non-puff instances.

Feature Computation

Respiration Features from [2]: From each respiration cycle, we compute the 17 respiration features presented in [2]. These features capture the characteristics of a respiration cycle and the relative changes in these characteristics. First, *Inhalation Duration*, *Exhalation Duration*, which correspond to the time required to breathe in and breathe out respectively are used as features. The next two features, *IE Ratio* and *Respiration Duration* are defined as the ratio of inhalation duration to the exhalation duration and their sum respectively. *Stretch* is defined to be the difference between the maximum (legitimate) amplitude and the minimum (legitimate) amplitude the signal attains within a respiration cycle.

Forward and backward first differences of a feature are defined as the difference between the value of this feature obtained from current respiration cycle and that from the next cycle and previous cycle, respectively. Since the smoking puff is different than neighboring respiration cycles, the forward and backward first differences of the values of the inhalation duration, exhalation duration, respiration duration, and stretch are also used as features to capture the relative changes in breathing pattern. Ratio of exhalation duration and stretch values to the average of the feature values of neighboring cycles are also used as features that capture the relative change in respiration. Finally, stretch of a respiration cycle is also divided into upper and lower parts with respect to the running mean of the valley amplitude of the respiration signal and these are used as features. The upper stretch magnitude is computed by taking the difference of peak amplitude and running mean value of the valley amplitudes of signal cycles (*ValleyAmplitudeMean*), while the lower stretch magnitude is computed by taking the absolute difference of minimum amplitude in a respiration cycle and *ValleyAmplitudeMean*.

New Respiration Features: In addition to the above features, we propose two new features. These features are computed from the rate of change signal obtained by taking the first derivative of the respiration signal. The maximum and minimum values that the rate of change signal attains within a respiration cycle are used as features.

Inertial Features: From the candidate hand-at-mouth gesture segments, we compute the mean, median, standard deviation, and quartile deviation of **magnitude of gyroscope**, **pitch**, and **roll**. This gives us a total of 12 features.

[2]An alternative measure could be Euclidean distance but the Euclidean distance is blind to correlated variables while the Mahalanabis distance takes the co-variances into account, which lead to elliptic decision boundaries in the 2D case, as opposed to the circular boundary in the Euclidean case.

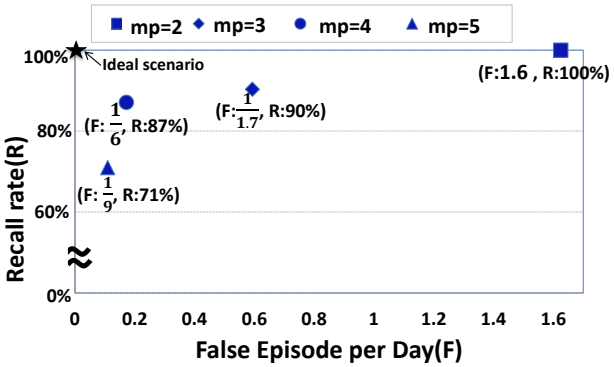

Figure 8. Recall versus false episode per day for different value of *mp*

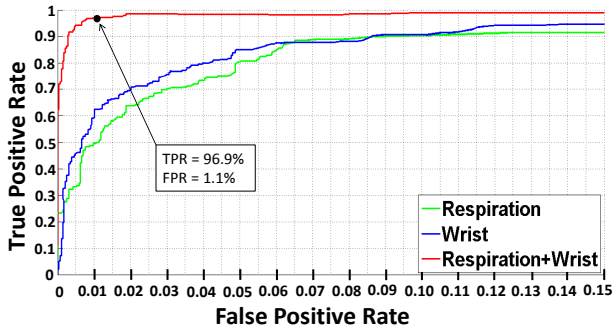

Figure 9. True Positive (recall) Rate vs. False Positive Rate of three classifiers for respiration-only model, wrist-only model, and for the combined model.

	Classified as puffs	Classified as non-puffs
puffs	282	9
non-puffs	40	3505

Table 1. Confusion Matrix for training data using 10-fold cross validation; Recall=96.9%, Precision=87.5%, Accuracy=98.7%, False Positive Rate=1.1%, Kappa=0.91

Model Development

We extract features from both the candidate segments and the corresponding respiration cycle from the training data. We train a two-class Support Vector Machine (SVM) classifier using this training data to detect puffs.

Post Processing

After obtaining a classification from the SVM model, we conduct two post-processing steps. Our post-processing steps are similar in purpose to the use of random forests and conditional random field used in prior works [17, 28] to construct a smoking episode from individual puffs. We opt for rule-based methods for better explainability of the resulting model.

Remove Isolated Puffs: We call a detected puff an *isolated puff* if no other puff is within two standard deviations of the mean inter-puff duration (i.e., 28 ($\pm$18.6) seconds). An isolated puff is unlikely to be part of a smoking episode.

Discard Puffs from the Non-dominant Hand: We observe that among 61 participants, 33 always smoke using their right hand, 18 smoke using left hand, and only 10 switch hands, sometimes switching hands even within a smoking episode. This points to the utility of using wrist sensors on both hands in a smoking cessation study. But, since majority of the participants smoke using only their dominant hand, puffs detected from their non-dominant hand can be discarded.

Constructing a smoking episode: After removing isolated puffs, we are left with clusters of (2 or more) puffs in the data stream. We use a simple rule-based method to declare a cluster of puffs as a smoking episode, i.e., if it contains at least *mp* (minimum puff count) puffs. To find an appropriate value for *mp*, we analyze the recall rate for detecting first lapses in lapsers and false episode detection rate in abstinent smokers in our smoking cessation study data. Figure 8 presents the recall and false episode per day rates for different values of *mp*. We observe that the best result is achieved when $mp = 4$.

EVALUATION AND APPLICATION

We now describe the performance of *puffMarker* on both training data and on the smoking cessation data.

Performance on Training Data

In addition to the *puffMarker* model that uses both respiration and wrist sensors, we also construct a wrist-only model to understand the performance expected if only wrist sensors are used due to its greater convenience and ease of wearing. To understand the improvement in accuracy due to each of the two sensor types, we also analyze the performance if only respiration sensor were used.

Training data for the puff classifier consists of 291 puffs and 44,696 respiration cycles that reduces to 3,545 non-puff cycles after applying our preprocessing steps. We build three different classifiers based on i) respiration-only features, ii) wrist-only features, and iii) features from both sensors.

Figure 9 presents the true positive rate of puffs detected versus true positive rate for the classifiers in 10-fold cross-validation. We observe that the performance of wrist-only model is better than that of respiration only. Combining both sensors results into significant improvement and makes it suitable for robust performance in the field setting. We pick an operating point on the ROC curve that is closest to the top left corner. It corresponds to a recall rate of 96.9% and false positive rate of 1.1%. A confusion matrix with various metrics for this operating point is presented in Table 1.

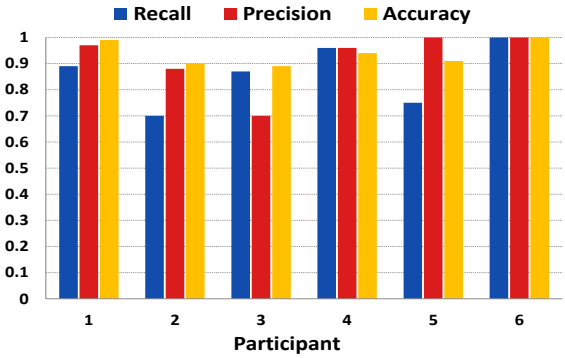

Figure 10. Leave one subject out cross validation on training data.

	puffMarker	Wrist-only
False episode per day	$\frac{1}{6}$	$\frac{1}{1.71}$
# detected lapses	$\frac{28}{32}(87.5\%)$	$\frac{24}{32}(75\%)$

Table 2. Comparison between *puffMarker* and Wrist-only model

We further investigate the generalizability of our combined model by performing leave-one-subject-out cross-validation. In this experiment, the model is learned from five participants' data and evaluated on the sixth. In each experiment, threshold on the score of the SVM classifier is set to a value so that it achieves at least 95% recall rate on the training data. From Figure 10, we observe that recall and precision are usually high with a minimum value of 0.7. It indicates that our method can generalize to new users. The difference in recall or precision can be attributed to difference in the proportion of confounding activities (e.g., conversation and physical activity) present in each participant's data.

Performance on Smoking Abstinence Data

We collected data on 61 participants in a smoking cessation study in which 33 lapsed within three days (verified by a CO monitor) — 17 lapsed on the first day, 12 on the second day, and 4 on the third day. We apply our model on these data and report several findings that include recall rate for detecting first lapse, false episode per day on abstinence data, lapse progression in lapsers, puff count in first lapse episodes, and temporal inaccuracy in self-report or recall of first lapses.

Detection of First Lapses

The CO report ascertained that among the 61 participants, 33 lapsed during their post quit session; 22 of them self-reported their lapse and 11 mentioned it in their next day interview. One participant is excluded from analysis because of high ($> 80\%$) sensor data loss around the neighborhood of lapse self-report. Of the 32 lapsers for whom data is available, *puffMarker* detects 28 lapse episodes. We can now derive various novel results on the nature of first lapse that has previously not been known. We report them in the following.

False Episode Detection

To analyze the false positive rate of *puffMarker*, we apply it to the data collected from 28 participants who did not lapse (confirmed by CO testing) during three days of post quit session. From each of these participants, we obtained an average of 11.2 hours of data per day for 3 days, for a total of 946 hours of data. Since the participants did not smoke on these days, all episodes detected by *puffMarker* are false positives. Out of these 28 participants, we get zero false positives for 22 participants, one episode for 2 participants, one episode each on two days for 2 participants, two episodes on two days for 1 participant, and four false episodes in one day for the final participant. We get a total of 14 falsely detected episodes in 84 days, for a false episode per day of 1 every six days.

We also analyze the false episode per day on abstinent days of lapsed smokers. Since 12 participants lapsed on the second day and 4 lapsed on the third day, we have a total of 12+8=20 abstinent days on these participants. We get false positive on only one participant who lapsed on the third day. *puffMarker* detects one episode on the first post-quit day and another episode on the second day.

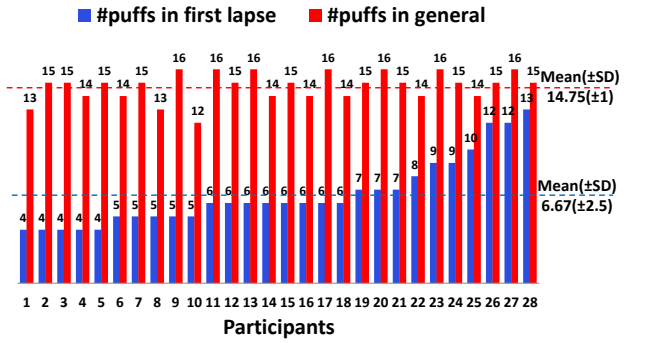

Figure 11. Puff variation in Lapse episode and regular smoking episode

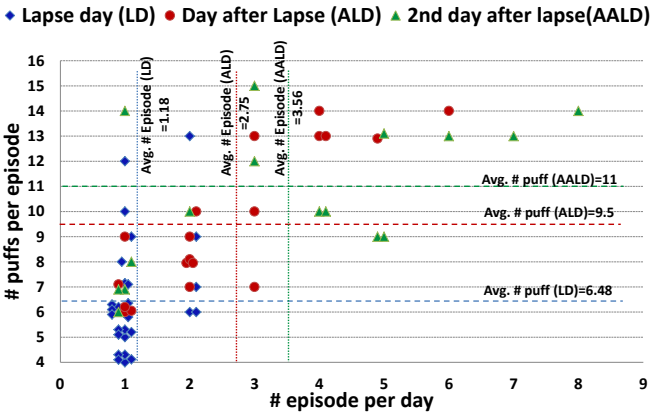

Figure 12. Progression of the lapse process. Number of smoking episodes per day and number of puffs per episode following a lapse are shown for the lapse day, day after lapse, and 2 days after lapse.

Performance of Wrist-only Model

In future, it may be desirable to use only wrist sensors in smoking cessation studies due to its convenience. We, therefore, also apply the wrist-only model on the smoking cessation data. Table 2 presents the performance of wrist-sensor-only model. We observe that the wrist-only model detects 24 of the 32 lapse events and the number of false episodes detected is 49 (or 1/1.71 per day).

Characterizing the Lapse process

Now that we have a model to detect the first lapse and subsequent lapses in a smoking cessation study, we can get some new insights into the lapse process that were not observable earlier. We report three novel findings. First, we analyze the number of puffs taken during the first lapse smoking event (see Figure 11). We find that the number of puffs in the first lapse event after quitting is significantly lower than that during regular smoking episodes. The average number of puffs in first lapse events is 6.67 (± 2.5), whereas the average number of puffs in regular smoking episodes is 14.75 (± 1). This has been suspected by smoking researchers as smokers are trying to resist smoking in the post-quit period, but our data now provides the first objective evidence.

Second, we analyze the number of smoking episodes per day on the lapse day, the day after lapse day, and 2 days after the lapse day. Third, we analyze the number of puffs per episode on these three days. Figure 12 shows data for both of these

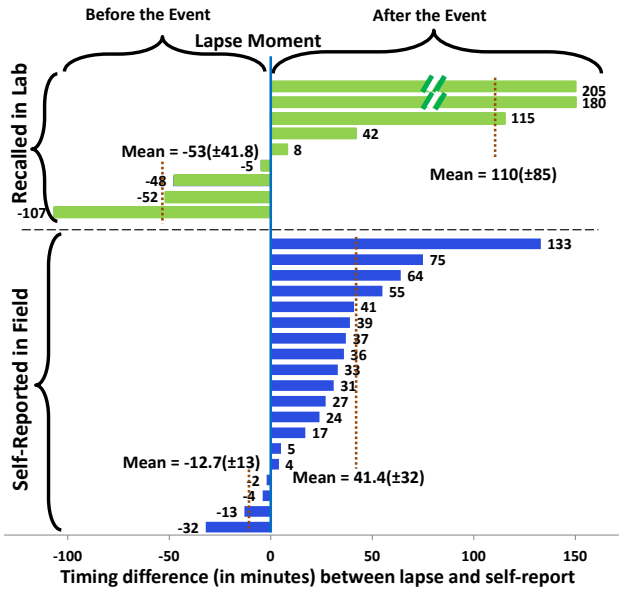

Figure 13. Temporal inaccuracy of self-report

metrics. Lapsers smoke an average of 1.18 times on the lapse day, an average of 2.75 times on the day after lapse day, and an average of 3.56 times on the second day after lapse day. For the number of puffs per smoking episode, we observe that the number of puffs per smoking episode is 6.5 on the lapse day, 9.5 on the day after lapse day, and 11 on 2 days after the lapse day.

The above objective data support prior observations (based on self-reports) that once a participant lapses, they gradually increase smoking frequency and eventually relapse fully. It is interesting to observe that not only the smoking frequency increases every day after lapse, but the number of puffs per episode also increases. Consequently, the total number of puffs per day increases rapidly; it is 7.7 on lapse day, 26.1 on day after lapse, and 39.2 on 2 days after lapse. This study only observed abstinent smokers for only 3 post-quit days; longer studies in future may reveal the entire progression process.

Temporal Precision in Self-report or Recall of First Lapse

Accurately locating the timing of the first lapse has been considered critical for the development of interventions. Smoking researchers have suspected that lapsers do not report their first lapse event promptly, partly due to being overwhelmed at the moment of this first failure in their cessation attempt. Now that we can pinpoint the timing of first lapse for 28 lapsers at the granularity of a respiration cycle (i.e., second-level accuracy), we can analyze the temporal imprecision in self-report or recall of the first lapse episodes (see Figure 13).

Of the 28, 9 did not self-report, 15 reported after the lapse, and 4 reported before they lapsed. The average delay for post-report was 41.4 minutes and that for a pre-report was 12.7 minutes. The temporal inaccuracy was even greater for those 9 who recalled the timing of lapse event next day in the lab. This is the first time that temporal inaccuracy in self-report of a smoking lapse has been reported.

CONCLUSION, LIMITATIONS, AND FUTURE WORK

This is the first work to show that timing of first lapse in smoking cessation can indeed be detected using wearable sensors in real-life environment. It is also the first to show the temporal inaccuracy in self-report or recall of first smoking lapse. From a computational modeling perspective, it presents an explainable model for gesture recognition from 6-axis inertial sensors worn on wrist. It also presents a reusable approach for combining the inertial sensor data with respiration data for better detection accuracy by leveraging the diversity of these two sensor types.

But, this work has several limitations that present numerous opportunities for future works. First, the model itself can be improved in multiple ways. For example, we use a simple rule for episode construction. More sophisticated models can potentially improve the detection accuracy for smoking episodes. Personalized models that use pre-quit data of each participant to calibrate the model may provide an even better accuracy. Second, detection of other related behaviors (e.g., eating, drinking, brushing, driving) using our modeling approach from wrist-mounted inertial sensors and respiration sensor is an interesting opportunity for future work. Third, we use wrist sensors on both wrists as the smoking activity can be performed with either hand. The same is true for other activities such as eating, typing, etc. But, wearing wrist sensors on both wrists may not be as prevalent outside of scientific studies. Obtaining similar accuracy of detection with only one sensor worn on the dominant (or non-dominant) hand is another interesting future work opportunity.

Fourth, our model was developed using data collected for cigarette smoking and hence may not directly work for cigars, e-cigarettes, hookah, etc. Fifth, since our model filters out isolated puffs and does not consider puffs to constitute a smoking episode unless there are 4 puffs in close vicinity. Hence, it may not work for detecting first lapses that consists of 3 or fewer puffs. Sixth, it also may not work when several people share a cigarette. In such a case, the time between successive puffs becomes longer than usual. Seventh, replication of our method in other populations can further improve its validity and utility. Finally, our work opens up a very rich area of research for discovering efficacious just-in-time interventions that can be triggered from predictors detected by sensors such as GPS, smart eyeglasses, electronic and social media, and physiological sensors.

ACKNOWLEDGEMENTS

We thank Rummana Bari, Mahbubur Rahman and Moushumi Sharmin from University of Memphis for their contributions to processing of respiration data and manuscript revisions. 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 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. Smoking-attributable mortality, years of potential life lost, and productivity losses - United States, 2000 - 2004. *Morbidity and Mortality Weekly Report 57*, 45 (2008), 1226–1228.
2. Ali, A., Hossain, S., Hovsepian, K., Rahman, M., Plarre, K., and Kumar, S. mPuff: Automated detection of cigarette smoking puffs from respiration measurements. In *Proc. ACM IPSN* (2012).
3. Ashton, H., Watson, D., Marsh, R., and Sadler, J. Puffing frequency and nicotine intake in cigarette smokers. *The British Medical Journal* (1970), 679–681.
4. Brendryen, H., Kraft, P., and Schaalma, H. Looking inside the black box: Using intervention mapping to describe the development of the automated smoking cessation intervention'happy ending'. *The Journal of Smoking Cessation 5*, 1 (2010), 29–56.
5. Doherty, K., Kinnunen, T., Militello, F., and Garvey, A. Urges to smoke during the first month of abstinence: relationship to relapse and predictors. *Psychopharmacology 119*, 2 (1995), 171–178.
6. Ertin, E., Stohs, N., Kumar, S., Raij, 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 *Proc. of ACM SenSys* (2011), 274–287.
7. Hughes, J., and Hatsukami, D. Signs and symptoms of tobacco withdrawal. *Archives of General Psychiatry 43*, 3 (1986), 289–294.
8. Hymowitz, N., Sexton, M., Ockene, J., and Grandits, G. Baseline factors associated with smoking cessation and relapse. *Preventive Medicine 20*, 5 (1991), 590–601.
9. Kalman, D. The subjective effects of nicotine: methodological issues, a review of experimental studies, and recommendations for future research. *Nicotine & Tobacco Research 4*, 1 (2002), 25–70.
10. Killen, J., and Fortmann, S. Craving is associated with smoking relapse: Findings from three prospective studies. *Experimental and Clinical Psychopharmacology 5*, 2 (1997), 137–142.
11. Lopez-Meyer, P., Tiffany, S., and Sazonov, E. Identification of cigarette smoke inhalations from wearable sensor data using a support vector machine classifier. In *Proc. IEEE EMBC* (2012).
12. Marian, C., O'Connor, R. J., Djordjevic, M. V., Rees, V. W., Hatsukami, D. K., and Shields, P. G. Reconciling human smoking behavior and machine smoking patterns: implications for understanding smoking behavior and the impact on laboratory studies. *Cancer Epidemiology Biomarkers & Prevention 18*, 12 (2009), 3305–3320.
13. Matheny, K., and Weatherman, K. Predictors of smoking cessation and maintenance. *Journal of Clinical Psychology 54*, 2 (1998), 223–235.
14. Mokdad, A., Marks, J., Stroup, D., and Gerberding, J. Actual causes of death in the united states, 2000. *The Journal of the Americal Medical Association 291*, 10 (2004), 1238–1245.
15. Murphy, J. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. New York Institute of Finance, 1999.
16. Palmer, J., and Hiiemae, K. Eating and breathing: interactions between respiration and feeding on solid food. *Dysphagia 18*, 3 (2003), 169–178.
17. Parate, A., Chiu, M.-C., Chadowitz, C., Ganesan, D., and Kalogerakis, E. RisQ: Recognizing smoking gestures with inertial sensors on a wristband. In *Proc. ACM MobiSys* (2014).
18. Pedley, M. Tilt sensing using a three-axis accelerometer, 2014.
19. Rahman, M., Bari, R., Ali, A., Sharmin, M., Raij, A., Hovsepian, K., and et. al.˙Are we there yet?: Feasibility of continuous stress assessment via wireless physiological sensors. In *Proc. of ACM BCB* (2014), 479–488.
20. Scholl, P. M., Kücükyildiz, N., and Laerhoven, K. V. When do you light a fire?: Capturing tobacco use with situated, wearable sensors. In *Proc. ACM UbiComp Workshop on Human Factors and Activity Recognition in Healthcare, Wellness, and Assistend Living* (2013).
21. Shiffman, S. Reflections on smoking relapse research. *Drug and Alcohol Review 25*, 1 (2006), 15–20.
22. Shiffman, S., Paty, J., Gnys, M., Kassel, J., and Hickcox, M. First lapses to smoking: Within-subjects analysis of real-time reports. *Journal of Consulting and Clinical Psychology 64*, 2 (1996), 366–379.
23. Shiffman, S., Scharf, D., Shadel, W., Gwaltney, C., Dang, Q., Paton, S., and Clark, D. Analyzing milestones in smoking cessation: illustration in a nicotine patch trial in adult smokers. *Journal of Consulting and Clinical Psychology 74*, 2 (2006), 276–285.
24. Shiffman, S., and Waters, A. Negative affect and smoking lapses: A prospective analysis. *Journal of Consulting and Clinical Psychology 72*, 2 (2004), 192–201.
25. Spohr, S. A., Nandy, R., Gandhiraj, D., Vemulapalli, A., Anne, S., and Walters, S. T. Efficacy of sms text message interventions for smoking cessation: A meta-analysis. *Journal of Substance Abuse Treatment* (2015).
26. Stitzer, M., and Gross, J. Smoking relapse: the role of pharmacological and behavioral factors. *Progress in Clinical and Biological Research 261* (1988), 163–184.
27. Swan, G., Ward, M., and Jack, L. Abstinence effects as predictors of 28-day relapse in smokers. *Addictive Behaviors 21*, 4 (1996), 481–490.

28. Tang, Q., Vidrine, D. J., Crowder, E., and Intille, S. S. Automated detection of puffing and smoking with wrist accelerometers. In *Proc. Pervasive Health)* (2014).

29. Wu, P., Hsieh, J., Cheng, J., Cheng, S., and Tseng, S. Human smoking event detection using visual interaction clues. In *Proc. IEEE Int'l Conf. Pattern Recognition* (2010).