Examining the Social Context of Alcohol Drinking in Young Adults with Smartphone Sensing

LAKMAL MEEGAHAPOLA, Idiap Research Institute & EPFL, Switzerland

FLORIAN LABHART, Idiap Research Institute, Switzerland

THANH-TRUNG PHAN, Idiap Research Institute, Switzerland

DANIEL GATICA-PEREZ, Idiap Research Institute & EPFL, Switzerland

According to prior work, the type of relationship between a person consuming alcohol and others in the surrounding (friends, family, spouse, etc.), and the number of those people (alone, with one person, with a group) are related to many aspects of alcohol consumption, such as the drinking amount, location, motives, and mood. Even though the social context is recognized as an important aspect that influences the drinking behavior of young adults in alcohol research, relatively little work has been conducted in smartphone sensing research on this topic. In this study, we analyze the weekend nightlife drinking behavior of 241 young adults in a European country, using a dataset consisting of self-reports and passive smartphone sensing data over a period of three months. Using multiple statistical analyses, we show that features from modalities such as accelerometer, location, application usage, bluetooth, and proximity could be informative about different social contexts of drinking. We define and evaluate seven social context inference tasks using smartphone sensing data, obtaining accuracies of the range 75%-86% in four two-class and three three-class inferences. Further, we discuss the possibility of identifying the sex composition of a group of friends using smartphone sensor data with accuracies over 70%. The results are encouraging towards supporting future interventions on alcohol consumption that incorporate users' social context more meaningfully and reducing the need for user self-reports when creating drink logs for self-tracking tools and public health studies.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in ubiquitous and mobile computing**; **Smartphones**; **Mobile phones**; **Mobile devices**; *Empirical studies in collaborative and social computing*; • **Computer systems organization** → Sensors and actuators; • **Applied computing** → *Consumer health*; *Health informatics*; *Sociology*; Psychology.

Additional Key Words and Phrases: passive sensing, smartphone sensing, nightlife, alcohol, drinking, social context, young adults, mobile sensing, self-reports, interaction sensing, continuous sensing

1 INTRODUCTION

In western countries, alcohol consumption is a leading risk factor for mortality and morbidity [46]. The consumption of several drinks in a row, commonly referred as binge drinking or heavy drinking, can lead to many

Authors' addresses: Lakmal Meegahapola, lmeegahapola@idiap.ch, Idiap Research Institute & EPFL, Switzerland; Florian Labhart, florian@idiap.ch, Idiap Research Institute, Switzerland; Thanh-Trung Phan, trungphan@idiap.ch, Idiap Research Institute, Switzerland; Daniel Gatica-Perez, gatica@idiap.ch, Idiap Research Institute & EPFL, Switzerland.

short-term adverse consequences not only for the person drinking (e.g., unprotected sex, injury, accidents or blackouts [52]) but also at the family and community levels (e.g. violence, drunk driving [49, 58]). On a larger time frame, heavy alcohol consumption can also lead to long-term consequences, such as poor academic achievement, diminished work capacity, alcohol dependence and premature death [68]. Adolescence and early adulthood appear as a particularly critical period of life for the development of risky alcohol-related behaviors since heavy alcohol consumption in late adolescence appears to persist into adulthood [113]. In order to limit excessive drinking among adolescents and young adults, it is essential to understand the etiology and antecedents of drinking occasions [47]. Prior work in social and epidemiological research on alcohol has emphasized the importance of the social context in shaping people's alcohol use and motives [28, 47, 70] in the sense that the consumption of alcohol or not, and the amounts consumed, vary depending on the presence or absence of family members [85, 88, 109, 110], of friends or colleagues [26, 76, 82, 88], and of the spouse or partner [59, 61, 66]. Additionally, a recent literature review showed that although the type of company is generally not a significant direct predictor of alcohol-related harm, young adults tend to experience more harm, independent of increased consumption, when they drink in larger groups [104].

Recent developments in ambulatory assessment methods (i.e., the collection of data in almost real time, e.g. every hour, and in the participant's natural environment [97, 100]) using smartphones made it possible to assess the type and the number of people present over the course of real-life drinking occasions [50, 53]. Compared to cross-sectional retrospective surveys traditionally used in alcohol epidemiological and psychological research, this type of approach allows to capture the interplay between drinking behaviors and contextual characteristics at the drinking event level in more detail [51]. For instance, evidence shows that larger numbers of drinking companions are associated with increased drinking amounts over the course of an evening or night [99, 106], and that this relationship is mediated by the companions' gender [107]. By repetitively collecting information from the same individuals over multiple occasions, ambulatory assessment methods are able to capture a large diversity of social contexts of real-life drinking occasions (e.g. romantic date with a partner, large party with many friends, family dinner) with the advantages of being free of recall bias and of participants serving as their own controls.

In addition to the possibility of capturing in-situ self-reports, smartphone-based apps have the potential to provide just-in-time adaptive interventions (JITAI) and feedback [71, 78]. Feedback systems primarily rely on identifying users' internal state or the context that they are in, to offer interventions or support (feedback) [45, 103]. Leveraging these ideas, recent studies in alcohol research have used mobile apps to provide interventions to reduce alcohol consumption using questionnaires, self-monitoring, and location-based interventions [4, 36, 116]. Furthermore, mobile sensing research has used passive sensing data from wearables and smartphones to infer aspects that could be useful in feedback systems, such as inferring drinking nights [92], inferring non-drinking, drinking, and heavy drinking episodes [6], identifying walking under alcohol influence [40], and detecting drunk driving [23]. Hence, given that the characteristics of the social context have been identified as central elements of any drinking event, it appears as a central target for inferring drinking occasions. However, to the best of our knowledge, mobile sensing has not been widely used to automatically infer the social context of alcohol drinking events. To further understand the importance of identifying social context using mobile sensing, consider the following example. If an app could infer a heavy-drinking episode (as shown by [6]), it could provide an intervention. However, there is a significant difference between drinking heavily alone or with a group of friends [32, 98]. Drinking several drinks in a row alone might indicate that the person is in emotional pain or stressed (also known as "coping" drinking motive) [47, 98]. However, drinking several drinks is common when young adults go for a night-out with friends [32]. In a realistic setting, for a mobile health app to provide useful interventions or feedback, the knowledge of the social context, in addition to knowing that the user is in a heavy-drinking episode, could be vital. Hence, understanding fine-grained contextual aspects related to alcohol consumption

using passive sensing is important, and could also open new doors in mobile interventions and feedback systems for alcohol research.

Further, there are a plethora of alcohol tracking, food tracking, and self-tracking applications in app stores, that primarily rely on user self-reports [72, 75, 92]. Even though gaining a holistic understanding regarding eating or drinking behavior is impossible without capturing contextual aspects regarding such behaviors, prior work has shown that people tend to reduce the usage of apps that require a large number of self-reports, and tend to use health and well-being applications that function passively [71]. Mobile sensing offers the opportunity to infer attributes that otherwise require user self-reports, hence reducing user burden [71, 72, 74]. In addition, mobile sensing could infer attributes to facilitate search acceleration in food/drink logging apps [39]. The social context of drinking alcohol is a variable that could benefit from smartphone sensing in an alcohol tracking application. As a whole, the idea of using smartphone sensing, in addition to capturing self-reports, is to gain a holistic understanding regarding the user context passively, that could otherwise take a long time-span if collected using self-reports. Considering all these aspects, We address the following research questions:

RQ1: What social contexts around drinking events can be observed by analyzing self-reports and smartphone sensing data corresponding to weekend drinking episodes of a group of young adults?
RQ2: Can young adults' social context of drinking be inferred using sensing data? What are the features that are useful in making such inferences?
RQ3: Are social context inference models robust to different group sizes? Can mobile sensing features infer the sex composition (same-sex, mixed-sex, opposite-sex), when drinking is done in a group of friends or colleagues?

By addressing the above research questions, our work makes the following contributions:

Contribution 1: Using a fine-grained mobile sensing dataset that captures drinking event level data from 241 young adults in a European country, we first show that there are differences in self-reporting behavior among men and women, regarding drinking events done with family members and with groups of friend/colleagues. Next, using various statistical techniques, we show that features coming from modalities such as accelerometer, location, bluetooth, proximity, and application usage are informative regarding different social contexts around which alcohol is consumed.
Contribution 2: We first define seven social context types, based on the number of people in groups (e.g., alone, with another person, with one or more people, with two or more people) and the relationship between the participant and others in the group (e.g., family or relatives, friends or colleagues, spouse or partner). Then, based on the above context types, we evaluate four two-class and three three-class inference tasks regarding the social context of drinking, using different machine learning models, obtaining accuracies between 75% and 86%, with all passive smartphone sensing data. In addition, we show that models that only take inputs from single sensor modalities such as accelerometer and application usage, could still perform reasonably well across all seven social context inferences, providing accuracies over 70%.
Contribution 3: For the specific case of drinking with friends or colleagues, we show that mobile sensor data could infer the sex composition of groups (i.e. same-sex, mixed-sex, or opposite-sex) in a three-class inference task, obtaining an accuracy of 75.8%.

The paper is organized as follows. In Section 2, we describe the background and related work. In Section 3, we describe the study design, data collection procedure, and feature extraction techniques. In Section 4 and Section 5, we present a descriptive analysis and a statistical analysis of dataset features. We define and evaluate inference tasks in Section 6. Finally, we discuss the main findings in Section 6, and conclude the paper in Section 7.

2 BACKGROUND AND RELATED WORK

2.1 The Social Context of Drinking Alcohol

While there are numerous definitions for the term social context in different disciplines, in this paper, we borrow the concept commonly used in alcohol research [20, 47, 55, 70], which refers to either one or both of the following aspects: (1) *type of relationship*: the relationship between an individual and the people in the individual's environment with whom she or he is engaging, and (2) *number of people*: the number of people belonging to each type of relationship, with whom the individual is engaging. By combining the two aspects, a holistic understanding of the social context of drinking of an individual can be attained.

The consumption of alcohol is associated with different contextual characteristics. These characteristics include the type of setting (e.g., drinking location), its physical attributes (e.g., light, temperature, furniture), its social attributes (e.g., type, size, and sex-composition of the drinking group, on-going activities), and the user's attitudes and cognition [70]. Applied to real-life situations, this conception underlines the changing nature of the drinking context, in the sense that the variety of situations during which alcohol might be consumed is rather large. For instance, across three consecutive days, the same person might drink in a restaurant during a date with a romantic partner, join a large party at a night club with many attendees, and finally, join a quiet family dinner at home.

Among all contextual characteristics, the composition of the social context is a central element of any drinking occasion, since the consumption of alcohol is predominantly a social activity for non-problematic drinkers [98]. Among adolescents and young adults, previous literature has shown that amounts of alcohol consumed on any specific drinking occasion vary depending on the type and number of people present [20]. The type of relationship that received the most attention so far is the presence of friends, in terms of number and of sex composition. Converging evidence shows that the likelihood of drinking [11] and drinking amounts are positively associated with the size of the drinking group [30, 99, 106]. Unfortunately, the group size is generally used as a continuous variable, preventing the identification of a threshold at which the odds of drinking in general or drinking heavily increase. Evidence regarding the sex composition of the group, however, provided mixed results, with some studies indicating that more alcohol is consumed in mixed-sex groups [65, 107] while others suggesting that this might rather be the case in same-sex groups [2]. The influence of the presence of the partner (e.g. boyfriend or girlfriend) within a larger drinking group has not been investigated, but evidence suggests that alcohol is less likely to be consumed and in lower amounts in a couple situation (i.e., the presence of the partner only) [38, 107]. It should be noted that these studies only suggest correlational links between the contextual characteristics and drinking behaviors and should not be interpreted as causal relationships.

The presence or absence of members of the family also play an important role in shaping adolescents and young adults' drinking behaviors. In particular, the presence of parents and their attitude towards drinking are often described as being either limiting or facilitating factors, but evidence in this respect is inconclusive. For instance, the absence of parental supervision was found to be associated to an increased risk for drinking at outdoor locations and young adults' home [108] suggesting that their presence might decrease this risk. However, another study shows that parents' knowledge about the happening of a party is negatively associated with the presence of alcohol, but there was no relationship between whether a parent was present at the time of the party and the presence of alcohol [29]. Lastly, parents might also facilitate the use of alcohol by supplying it, especially to underage drinkers [31, 38]. To sum up, evidence on the impact of the presence or absence of parents on young people's drinking appears mixed as this might be related to their attitude towards drinking, with some parents being more tolerant or strict than others [83]. Lastly, it should be noted that the presence of siblings has rarely been investigated, but unless they have a supervision role in the absence of parents, their role within the drinking group might be similar to one of friends.

2.2 Alcohol Consumption and Mobile Phones

2.2.1 Mobile Apps for Interventions in Alcohol Research. Many mobile apps in alcohol research focus on providing interventions or feedback to users to reduce alcohol consumption [21, 25, 36]. Crane et al. [21] conducted a randomized controlled trial using the app called "Drink Less", to provide interventions. This app relied on user self-reports, and they concluded that the app helped reduce alcohol consumption. Moreover, Davies et al. [25] conducted a randomized controlled trial with an app called "Drinks Meter", that provided personalized feedback regarding drinking. This app also used self-reports to provide feedback. Similarly, many mobile health applications in alcohol research that provide users with interventions or feedback, primarily used self-reports [15, 37, 81]. Regarding sensing, Gustafson et al. [36] deployed an intervention app called ACHESS, that provided computer-based cognitive behavioral therapy and additional links to useful websites, and this app provided interventions to users when they entered pre-defined high-risk zones, primarily relying on location sensing capabilities of the smartphone. LBMI-A [27] by Dulin et al. is another study that is similar to ACHESS. As a summary, alcohol epidemiology research that used mobile apps primarily targeted interventions based on self-reports or simple sensing mechanisms. Even though many studies have identified that self-reports are reasonably accurate to capture alcohol consumption amounts [64], studies have also stated that heavy-drinking episodes are often under-reported when self-reporting [80]. In addition, unless there is a strong reason for users to self-report, there is always the risk of users losing motivation to use the app over time.

2.2.2 Smartphone Sensing for Health and Well-Being. Smartphones allow sensing health and well-being aspects via continuous and interaction sensing techniques, both of which are generally called passive sensing [71]. This capability has been used in areas such as stress [13, 67], mood [62, 102], depression [14, 112], well-being [57, 63], and eating behavior [12, 72, 74]. If we consider drinking related research in mobile sensing, Bae et al. [6] conducted an experiment with 30 young adults for 28 days, and used smartphone sensor data to infer non-drinking, drinking, and heavy-drinking episodes with an accuracy of 96.6%. They highlighted the possibility of using such inferences to provide timely interventions. Santani et al. [92] deployed a mobile sensing application among 241 young adults for a period of 3 months, to collect sensor data around weekend nightlife events. They showed that sensor features could infer drinking and non-drinking nights with an accuracy of 76.6%. Kao et al. [40] proposed a phone-based system to detect feature anomalies of walking under the influence of alcohol. Further, Arnold et al. [3] deployed a mobile application called Alco Gait, to classify the number of drinks consumed by a user into sober (0-2 drinks), tipsy (3-6 drinks) or drunk (greater than 6 drinks) using gait data, obtaining reasonable accuracies. While most of these studies focused on detecting drinking events/episodes/nights, we focus on inferring the social contexts of drinking events.

2.2.3 Event Detection and Event Characterization in Mobile Sensing. Smartphone sensing deployments can be classified into two based on the study goal [71]: (a) Event Detection (e.g. drinking alcohol, eating food, smoking, etc.) and (b) Event Characterization (characteristics of the context that helps understand the event better – e.g. social context, concurrent activities, ambiance, location, etc.). For domains such as eating behavior, there are studies regarding both event detection (identifying eating events [9, 77], inferring meal or snack episodes [12], inferring food categories [73]) and event characterization (inferring the social context around eating events [72]). Inferring mood [62, 96] as well as identifying contexts around specific moods [24] has been attempted in ubicomp. However, even though alcohol epidemiology researchers have attempted to characterize alcohol consumption to gain a more fine-grained understanding about drinking, mobile sensing research has not been focused on the social context aspect thus far, even though some studies have looked into event detection [3, 6, 40, 92]. Hence, we aim to address this research gap by focusing on the social context of drinking alcohol using smartphone sensing.

3 DATA, FEATURES, AND TASKS

3.1 Mobile Application, Self-Reports, and Passive Sensing

We obtained a dataset regarding young adults' nightlife drinking behavior, from our previous work [92]. This dataset contains smartphone sensor data and self-reports regarding the drinking behavior of a 241 young adults (53% men) in Switzerland, during weekend nights, throughout a period of three months, and was collected as a collaboration between alcohol researchers, behavioral scientists, and computer scientists. In this section, we briefly describe the study design, the data collection procedure, and feature extraction technique. A full description regarding the ethical approval, deployment, and the data collection procedure can be found in [1, 92, 93].

3.1.1 Mobile App Deployment. To collect data from study participants, an android mobile application was deployed, and this app had two main components: **(a) Drink Logger:** used to collect in-situ self-reports during weekend nights (Friday and Saturday nights, from 8.00pm to 4.00am next day). The app sent notifications hourly, asking whether users wanted to report a new drink; and **(b) Sensor Logger:** used many passive sensing modalities to collect data, including both continuous (accelerometer, battery, bluetooth, location, wifi) and interaction (applications, screen usage) sensing. The application was deployed from September to December 2014. The study participants were young adults with ages ranging from 16 to 25 years old (mean=19.4 years old, SD=2.5). More details regarding the deployment can be found in [92].

3.1.2 Self-Reports. Whenever they were about to drink an alcoholic or non-alcoholic drink, participants were requested to take a picture of it and to describe its characteristics and the drinking context using a series of self-reported questionnaires [54]. Participants labeled the drink type using a list of 6 alcoholic drinks (e.g. beer, wine, spirits, etc.) and 6 non-alcoholic drinks (e.g. water, soda, coffee, etc.). Then, in accordance with the definition of social context we adopted in Section 2.1, participants reported the type and number of people present for each of the following categories: (a) partner or spouse; (b) family or relatives; (c) male friends or colleagues; (d) female friends or colleagues; and (e) other people (called *type of relationship* in the remainder of the paper). These five categories were adopted from prior work in alcohol research [55]. Next, for each type of relationship, participants reported the *number of people* using a 12-point scale with 1-point increments from 0 to 10, plus 'more than 10'; with the exception of partner or spouse which could either be absent (coded as 0) or present (1). This scale was designed to measure variations in the social context, following the assumption that the presence of each person counts within small groups, but that the additional value of each extra person is less important within larger groups (e.g. 10 or more people). Further, information about participants including age, sex, occupation, education level, and accommodation were collected in a baseline questionnaire. Overall, by selecting self-reports of situations when participants reported the consumption of an alcoholic drink, we were left with 1254 self-reports for the analysis.

3.1.3 Passive Smartphone Sensing. To gain a fine-grained understanding about users' drinking behavior, passive sensing data were collected during the same time period when participants self-reported alcohol consumption events. The chosen sensing modalities were Accelerometer (ACC), Applications (APP), Location (LOC), Screen (SCR), Battery (BAT), Bluetooth (BLU), Wifi (WIF), and Proximity (PRO). A dataset summary is given in Table 1 and an extensive description is given in [86, 92].

3.2 Aggregation and Matching of Self-Reports and Passive Sensing Data

Prior studies that used this dataset primarily considered *user-night* as the point of analysis (e.g., inferring nights of alcohol consumption vs. no alcohol consumption [92], inferring heavy-drinking nights [86], etc.). However, in this study, we consider drink-level data, that is more fine-grained. We prepared the self-report dataset such

Table 1. Summary of features extracted from mobile sensors (134). Sensor data are aggregated for every 10-minute time slot from 8pm to 4am. For all the given features, average, minimum, and maximum were calculated during the matching phase, hence resulting in a total of 402 sensing features for each alcohol consumption event.

Sensor – Feature Type (# of features)	Sensor Description Feature Description
Location	Location data were continuously collected for a time period of 1-minute during each 2-minute time slot. Collected data included data source, longitude, latitude, signal strength, and accuracy.
– Attributes (10)	{min., max., med., avg., std.} of avg. of speed and sensor accuracy
– Signal (3)	3 signal strengths (GPS, network, unknown)
Accelerometer	Values from all three axes of the sensor were collected, 10 seconds continuously, at a frequency of 50Hz, during every minute. we calculated (a) basic statistics from raw sensor data from the X, Y, and Z-axes [92]; (b) aggregated statistics related to acceleration (m, mNew, dm) and signal magnitude area (mSMA) by combining data from three axes [41, 69, 86]; and (c) angle between acceleration and the gravity vector [86, 92].
– Raw (15)	{min., max., med., avg., std.} of avg. of xAxis, yAxis, zAxis of accelerometer
– Angle (15)	{min., max., med., avg., std.} of angle of xAxis, yAxis, zAxis with g vector
– Dynamic (20)	{min., max., med., avg., std.} of mSMA, dm, m, mNew values
Bluetooth	The list of available devices was captured as Bluetooth logs, once every 5 minutes. Features such as the number of devices around, signal strengths, and empty scan counts were captured.
– Count (4)	number of bluetooth IDs surrounding devices, records, bluetooth scan count, empty scan count
– Strength (5)	{min., max., med., avg., std.} of bluetooth strength signal of surrounding devices
Wifi	The list of available devices was captured as WiFi logs, once every 5 minutes. Features such as the number of hotspots around, signal strengths, and the empty scan counts were captured.
– Count (2)	wifi record , wifi id set
– Attributes (10)	{min., max., med., avg., std.} of level, frequency of wifi hotspot
Application	Applications were categorized into 33 groups (e.g., art & design, food & drink, social, games, etc.) based on the categorization provided in google play store [33]. Using the categories and running apps, statistics such as the total number of running apps and the number of running apps based on categories were calculated [62, 86, 92].
– Count (2)	app count, app record
– Category (33)	normalized 33-bin histogram of 33 application categories
Proximity	Several statistical featured were derived using raw values of the proximity sensor.
– Count (1)	proximity records
– Distance (5)	{min., max., med., avg., std.} of distance from phone to objects
Battery	Battery levels were captured once every five minutes, and status changes were captured whenever a change occured. Hence, several features including battery full, discharging, charging, battery level, and whether the phone is plugged-in or not were derived.
– Status (5)	5 battery statuses
– Level (5)	{min., max., med., avg., std.} of battery level
– Count (2)	count of battery records and plugged times
Screen	Screen data were recorded whenever the screen status changed. Using the captured data, we derived the percentage of screen-on time.
– Usage (1)	percentage of screen on time

that each entry corresponds to a drinking event. Then, to combine sensor data and self-reports in a meaningful manner, we used the following two-phase technique, that was adopted from prior ubicomp research [5, 12, 72, 96]:

Phase 1 (Aggregation): We aggregate raw sensor data for every ten-minute window throughout the night. Different techniques were used for the aggregation for sensors. Hence, for a user-night, we have 48 ten-minute windows, from 8.00pm to 4.00am next day. For each feature derived from each sensor, we have 48 values (6 ten-minute windows per hour X eight hours per night) for a user-night. For instance, if there is a feature F_1 derived from sensor S_1, for each for each user and for each night, F_1 would have 48 values, that represent time windows from 8.00pm-8.09pm, 8.10pm-8.19pm, 8.20pm-8.29pm, until 03.50am-03.59am of next day.

Phase 2 (Matching): During this phase, features are matched to alcohol consumption self-reports using a one-hour window (approximately, from 30 minutes before the alcohol consumption self-report, to 30 minutes after the drinking self-report). For instance, if the drinking was reported at 10.08pm, we calculate the average (_avg), maximum (_max), and minimum (_min) values for each feature using values corresponding to six ten-minute
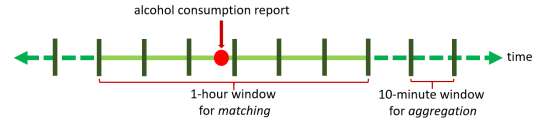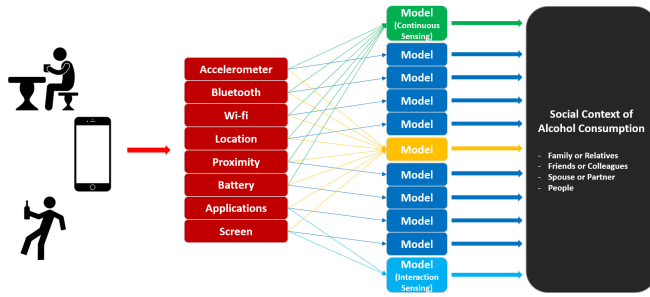
Fig. 1. A schematic diagram representing the summary of the study

Fig. 2. Diagram summarizing the two-phase technique for combining self-reports and sensing data

windows (obtained in Phase 1) from 9.40pm to 10.39pm, and match those value to the self-report. The idea behind this aggregation is to capture sensor data around drinking events, expecting that contextual cues could be informative of different social contexts. The summary of passive sensing features is provided in Table 1. These features were derived for every ten-minute time slot throughout the night, as discussed in Phase 1 (Aggregation), resulting in a total of 134 features. Then, in accordance with the procedure in Phase 2 (Matching), a one-hour time window around each drinking self-report was considered. By considering the average, minimum, and maximum of the six values corresponding to the one-hour time window, 402 passive sensing features were included for each alcohol drinking self-report in the final dataset. This two-phase technique is summarized in Figure 2. After following this technique for all self-reports and sensor features, we obtained a dataset. We removed data points with incomplete sensor data (152 records with unavailable sensor data, mainly location, wifi, or bluetooth data), not enough data for the matching phase (102 records, for drinking events that were done between 8.00pm-8.30pm and 3.30am-4.00am), and self-reports that were produced while visiting other countries (59 records, when the participant traveled while being in the study). The final dataset contained 941 complete drinking reports with sensor features.

3.3 Deriving Two-Class and Three-Class Social Context Features

In Section 2, we described how social contexts such as being with/without family members, friends/colleagues, and spouse/partner could be associated with drinking behavior. In addition, under Section 3.1.2, we described the type of social contexts reported by participants. Among them, features such as *with male friends/colleagues*, *with female friends/colleagues*, and *with family members* had twelve-point scales, and *with partner/spouse* had a two-point scale. However, for the purpose of this analysis, we reduced the twelve-point scale to low-dimensional scales (two-point and three-point), with the objective of capturing social context group dynamics, that are meaningful in terms of drinking events such as: being alone, with another person, or in a group of two or more. We followed the following steps.

First, except for the feature *with partner or spouse* which is already two-class, we minimized the scale of other features to two-classes and three-classes. For two-class features, the values could be either zero or one, whereas – **zero:** the participant is not with anyone belonging to the specific social context; and **one:** the participant is with one or more others belonging to the specific social context (hence, in a group). For three-class features, the values could be either zero, one, or two as follows – **zero:** the participant is not with anyone belonging to the specific social context; **one:** the participant is with one other person belonging to the specific social context (hence, in a group of two people); **two:** the participant is with two or more people belonging to the specific social context (hence, in a larger group).

Then, we derived several new features using the existing features:

Table 2. Summary of social contexts in the final dataset.

| Social Context	Classes	Interpretation
$family_{two}$	2	without vs. with one/more family members/relatives
$partner_{two}$	2	without vs. with the partner/spouse
$friends_{two}$	2	without vs. with one/more friends/colleagues
$people_{two}$	2	without vs. with one/more people
$family_{three}$	3	without vs. with one vs. with two/more family members/relatives
$friends_{three}$	3	without vs. with one vs. with two/more friends/colleagues
$people_{three}$	3	without vs. with one vs. with two/more people
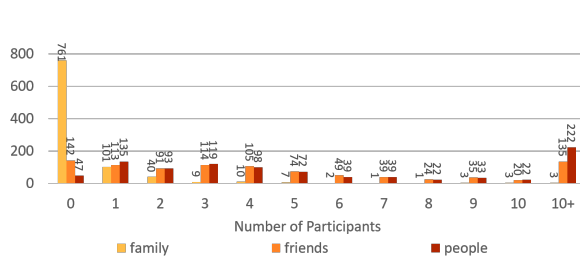
Fig. 3. Distribution of original self-report features (family, friends) and a derived feature (people)
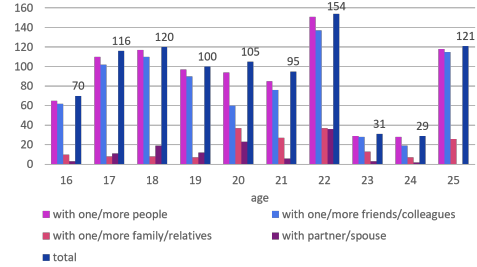

Fig. 4. Self-Reports in terms of Age

- *without friends/colleagues vs. with friends/colleagues* (two-class): this aggregated features about men and women friends/colleagues into a single two-class variable by discarding the sex demographic attribute of friends/colleagues.
- *without friends/colleagues vs. with another friend/colleague vs. with two/more friends/colleagues* (three-class): this aggregated features about the men and women friends/colleagues into a single three-class feature.
- *without people vs. with people* (two-class): this feature combines all the two-class social contexts to estimate the overall two-class social context of the user.
- *without people vs. with another person vs. with two/more people* (three-class): this feature combines all the other three-class social contexts and the two-class feature *with partner/spouse*, to estimate the overall three-class social context of the user.

The final set of social context features used for this study are summarized in Table 2. In accordance with the definition of social context proposed in Section 3.3, these features capture two aspects. First, they capture the relationships between the study participant and people engaging with the participant during alcohol consumption. Second, they capture group dynamics for each relationship (e.g., alone, with another person – small group of two people, with two/more people – comparatively large group, etc.). According to prior work in alcohol research, both perspectives are important to obtain a fine-grained understanding about drinking behavior [8, 70]. The summary of our analytical setting is presented in Figure 1.

4 DESCRIPTIVE ANALYSIS (RQ1)

In this section, we provide a descriptive analysis regarding self-reports using demographic information, to understand the nature of the aggregate drinking behavior of participants.

Self-Report Distribution Breakdown Based on Social Contexts: Figure 3 provides a distribution of self-reports. Self-reports for partner is not shown here because it is only a binary response, and is included in Figure 5.
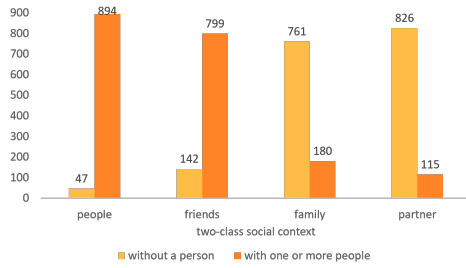
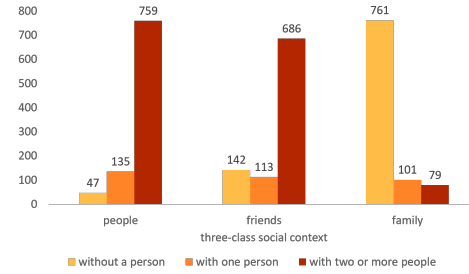Fig. 5. Distribution of two-class social contexts

Fig. 6. Distribution of three-class social contexts
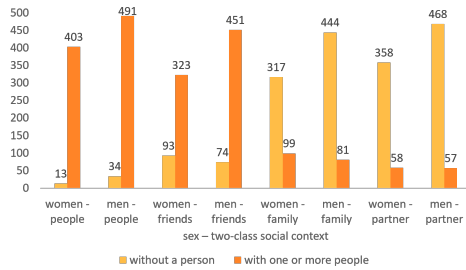

Fig. 7. Self-Reports in terms of Sex and Two-Class Social Context
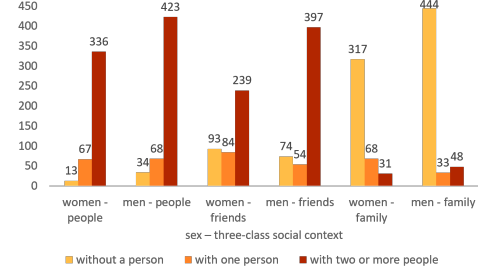

Fig. 8. Self-Reports in terms of Sex and Three-Class Social Context

Figure 5 and Figure 6 provide a distribution of self-reports for four two-class social contexts and three three-class social contexts, respectively. Results in Figure 5 show that only 47 (5.0%) drinking occasions were done alone as compared to 894 (95.0%) occasions that were done with one/more people. Out of these 894 reports, 799 (89.4%) were reported to have happened with two/more friends/colleagues. According to Figure 6, these 799 reports consist of 113 (14.1%) reports that were done with one friend/colleague and 686 (85.9%) reports that were done with two/more friends/colleagues, hence in a larger group. As a summary, participants consumed alcohol while being alone only on a small portion of occasions. This result is comparable to prior alcohol research, which shows that solitary drinking episodes are less frequent as compared to other social contexts [8]. Moreover, the presence of two or more friends/colleagues was reported well over more than half of all drinking occasions (686/941 = 72.9%). This result too is in line with prior work that state that young adults tend to drink alcohol for social facilitation and peer acceptance [8].

Self-Report Distribution Breakdown Based on Sex: In Figure 7 and Figure 8, we present distributions of self-reports, based on sex and social context pairs. Results indicate that social contexts 'people' and 'friends' reported more drinking occasions with one/more people, for both men and women, whereas social contexts 'partner' and 'family' have significantly high number of drinking drinking events that were reported to be done alone. In addition, for the social context 'friends', Figure 8 shows that the proportion of self-reports in groups of two/more (239) is just over half for women (239/416 = 57.5%), whereas for men, drinking events with two/more friends/colleagues is 75.6% (397/525), which is almost a 20% difference for two sexes. This suggests that men reported a higher proportion of drinking occasions in groups of two/more people. This result is consistent with prior literature that state that men tend to drink in larger social contexts (specially with friends/colleagues) whereas women are less likely to do so [108]. Further, women participants have reported drinking with family members 99 times (99/416 = 23.8%), whereas men only reported to have done so 81 times (81/525 = 15.4%), that is about 9% less than women.

Table 3. Pearson Correlation Co-efficient (PCC) and Point-Biserial Correlation Co-efficient (PBCC) for Sensor Features and Social Contexts (two-class and three-class). With the top 5 features for each Social Context are included in the table. p-values are denoted with the following notation: p-value$\leq 10^{-4}$:****; p-value$\leq 10^{-3}$:***

		family		partner		friends		alone	
	Feature (Sensor)	PBCC	Feature (Sensor)	PBCC	Feature (Sensor)	PBCC	Feature (Sensor)	PBCC
two-class	angleXMax_min (ACC)	0.19 (+) ****	yAxisAvgMax_avg (ACC)	0.14 (-) ****	mMean_avg (ACC)	0.22 (+) ****	mMed_max (ACC)	0.19 (+) ****
	angleXMed_min (ACC)	0.19 (+) ****	food_and_drink_avg (APP)	0.13 (+) ***	mSMAMean_avg (ACC)	0.22 (+) ****	mSMAMed_max (ACC)	0.19 (+) ****
	angleYAvg_min (ACC)	0.18 (+) ****	food_and_drink_min (APP)	0.12 (+) ***	mMed_max (ACC)	0.22 (+) ****	mMean_max (ACC)	0.19 (+) ****
	angleXAvg_min (ACC)	0.18 (+) ****	food_and_drink_min (APP)	0.12 (+) ***	mSMAMed_max (ACC)	0.22 (+) ****	mSMAMean_max (ACC)	0.19 (+) ****
	angleYMax_min (ACC)	0.18 (+) ****	zAxisAvgMin_avg (ACC)	0.12 (+) ***	mMax_avg (ACC)	0.21 (+) ****	dmListStd_max (ACC)	0.18 (+) ****
	Feature (Sensor)	PCC	-	-	Feature (Sensor)	PCC	Feature (Sensor)	PCC
three-class	angleXMed_min (ACC)	0.18 (+) ****	-	-	mMean_avg (ACC)	0.24 (+) ****	mMean_avg (ACC)	0.23 (+) ****
	angleXMax_min (ACC)	0.18 (+) ****	-	-	mSMAMean_avg (ACC)	0.24 (+) ****	mSMAMean_avg (ACC)	0.23 (+) ****
	angleYAvg_min (ACC)	0.18 (+) ****	-	-	mMed_avg (ACC)	0.23 (+) ****	yAxisAvgMin_min (ACC)	0.23 (-) ****
	angleYMin_min (ACC)	0.17 (+) ****	-	-	mSMAMed_avg (ACC)	0.23 (+) ****	mMean_max (ACC)	0.23 (+) ****
	angleXMean_min (ACC)	0.17 (+) ****	-	-	mSMAMax_avg (ACC)	0.23 (+) ****	mSMAMean_max (ACC)	0.23 (+) ****

Self-Report Distribution Breakdown Based on Age: As shown in Figure 4, participants' age ranged from 16 to 25. Except for ages 23 and 24 (31 and 29 self-reports, respectively), all other ages had over 70 self-reports. Moreover, the highest proportion of situations with one/more friends/colleagues (115/121 = 95.0%) was reported by participants aged 25. The lowest proportion of situations with partner/spouse (0%) was reported by the same age group.

5 STATISTICAL ANALYSIS (RQ1)

5.1 Pearson and Point-Biserial Correlation for Social Contexts and Passive Sensing Features

We conducted Pearson (PCC) [105] and Point-biserial (PBCC) [10] correlation analyses to measure the strength and the direction of the relationships between each of the three-class (takes values 0,1, and 2) and two-class (takes values 0 and 1) features and passive sensing features. The results of the top five features with the highest PCC or PBCC with each social context are summarized in Table 3. For a majority of social contexts (family$_{two}$, family$_{three}$, friends$_{two}$, friends$_{three}$, people$_{two}$, and people$_{three}$), multiple accelerometer-related features were among the top five features based on the correlation co-efficient values. The exception is the social context partner$_{two}$, where application features (e.g. food and drink) were among the top five. However, most of the values suggested, at best, weak positive or negative relationships.

5.2 Statistical Analysis of Dataset Features

Table 4 shows statistics such as t-statistic [43], p-value [34], and Cohen's-d (effect size) with 95% confidence interval (CI) [56] for the top five features in the dataset for the seven different social contexts. For two-class social contexts, the objective is to identify passive sensing features that help discriminate between: without people (alone) and with one/more people (group). Here, the term group is used because it could either be a small group of two to three people, or a large group of more than ten people. Further, for three-class social contexts, the objective is to identify passive sensing features that help discriminate between: (a) without people (alone) vs. with one person (sgroup); (b) with one person (sgroup) vs. with two/more people (lgroup); and (c) without people (alone) vs. with two/more people (lgroup), where sgroup and lgroup stands for small group and large group, respectively. The features are ordered by the descending order of t-statistics and Cohen's-d values. In addition, prior work stated the lack of sufficient informativeness in p-values [60, 115]. For this reason, we calculated the Cohen's-d [89] to measure the statistical significance of features. We adopted the following rule-of-thumb, commonly used to interpret Cohen's-d values: 0.2 = small effect size; 0.5 = medium effect size; and 0.8 = large effect size. According to this notion, the higher the value of Cohen's-d, the higher the possibility of discerning the two groups using the

Table 4. t-statistic (T) (p-value$\leq 10^{-4}$:****; p-value$\leq 10^{-3}$:***; p-value$\leq 10^{-2}$:**), and Cohen's-d (C) with 95% confidence intervals (* if confidence interval include zero). Top five features are shown in decreasing order.

family$_{three}$

	Feature	T	Feature	C
alone vs. sgroup	blueStrengthMax_avg (BLU)	4.41****	blueStrengthMed_avg (BLU)	0.42
	blueStrengthAvg_avg (BLU)	4.21****	zAxisAvgStd_min (ACC)	0.41
	blueStrengthMed_avg (BLU)	4.21****	mMin_min (ACC)	0.41
	blueStrengthMin_avg (BLU)	3.87***	dmListStd_min (ACC)	0.40
	blueStrengthMax_min (BLU)	2.71**	zAxisAvgMean_min (ACC)	0.40
sgroup vs. lgroup	zAxisAvgMean_min (ACC)	3.22**	dmListMean_min (ACC)	0.38
	zAxisAvgMedian_min (ACC)	3.08**	zAxisAvgStd_min (ACC)	0.36
	video_players_min (APP)	2.84**	anglezMax_min (ACC)	0.34
	zAxisAvgMax_avg (ACC)	2.71**	system_avg (APP)	0.32
	video_players_avg (APP)	2.70**	zAxisAvgMax_avg (ACC)	0.32
alone vs. lgroup	system_min (APP)	5.83****	mMin_min (ACC)	0.49
	anglexMax_min (ACC)	5.64****	zAxisAvgStd_min (ACC)	0.49
	anglexMedian_min (ACC)	5.64****	anglezMedian_min (ACC)	0.48
	angleyMean_min (ACC)	5.55****	anglexMax_min (ACC)	0.47
	angleyMin_min (ACC)	5.54****	zAxisAvgMean_min (ACC)	0.46

friends$_{three}$

	Feature	T	Feature	C
alone vs. sgroup	yAxisAvgMin_max (ACC)	4.05****	dmListMean_max (ACC)	0.37
	xAxisAvgStd_avg (ACC)	3.76***	xAxisAvgStd_avg (ACC)	0.34
	anglexStd_avg (ACC)	3.72***	anglexStd_avg (ACC)	0.33
	xAxisAvgStd_max (ACC)	3.65***	dmListMedian_max (ACC)	0.33
	yAxisAvgMax_avg (ACC)	3.55***	angleyMin_max (ACC)	0.32
sgroup vs. lgroup	mMedian_avg (ACC)	3.44***	yAxisAvgMin_avg (ACC)	0.38
	mMean_avg (ACC)	3.43***	yAxisAvgMean_avg (ACC)	0.35
	mSMAMedian_avg (ACC)	3.41***	yAxisAvgMedian_avg (ACC)	0.35
	mSMAMean_avg (ACC)	3.41***	mMax_min (ACC)	0.33
	mNewStd_avg (ACC)	3.15**	mMedian_avg (ACC)	0.33
alone vs. lgroup	mMean_avg (ACC)	8.61****	mMean_avg (ACC)	0.56
	mSMAMean_avg (ACC)	8.60****	mSMAMean_avg (ACC)	0.56
	mMedian_max (ACC)	8.44****	mMedian_max (ACC)	0.55
	mSMAMedian_max (ACC)	8.40****	mSMAMedian_avg (ACC)	0.55
	mMax_avg (ACC)	8.40****	mMax_avg (ACC)	0.54

people$_{three}$

	Feature	T	Feature	C
alone vs. sgroup	mSMAMedian_max (ACC)	3.81***	zAxisAvgMedian_max (ACC)	0.41
	mMedian_max (ACC)	3.81***	xAxisAvgStd_max (ACC)	0.41
	anglexStd_max (ACC)	3.79***	xAxisAvgMedian_min (ACC)	0.40
	xAxisAvgStd_max (ACC)	3.63***	xAxisAvgMean_max (ACC)	0.40
	mMean_max (ACC)	3.60***	angleyMax_max (ACC)	0.38
sgroup vs. lgroup	yAxisAvgMin_min (ACC)	5.30****	mNewStd_avg (ACC)	0.43
	yAxisAvgMin_avg (ACC)	5.11****	mMean_avg (ACC)	0.42
	yAxisAvgMean_min (ACC)	5.10****	mSMAMean_avg (ACC)	0.42
	yAxisAvgMedian_min (ACC)	5.07****	mSMAMax_avg (ACC)	0.42
	yAxisAvgMax_min (ACC)	4.18****	mMax_avg (ACC)	0.41
alone vs. lgroup	yAxisAvgMin_min (ACC)	7.44****	xAxisAvgStd_max (ACC)	0.77
	zAxisAvgMin_min (ACC)	6.76****	zAxisAvgMedian_max (ACC)	0.77
	xAxisAvgMin_min (ACC)	6.63****	angleyMax_max (ACC)	0.75
	yAxisAvgMedian_min (ACC)	6.54****	anglezMin_max (ACC)	0.75
	yAxisAvgMean_min (ACC)	6.31****	mMedian_avg (ACC)	0.73

people$_{two}$

	Feature	T	Feature	C
alone vs. group	yAxisAvgMin_max (ACC)	4.05****	dmListMean_max (ACC)	0.37
	xAxisAvgStd_avg (ACC)	3.76***	xAxisAvgStd_avg (ACC)	0.34
	anglexStd_avg (ACC)	3.72***	anglexStd_avg (ACC)	0.33
	xAxisAvgStd_max (ACC)	3.65***	dmListMedian_max (ACC)	0.33
	yAxisAvgMax_avg (ACC)	3.55***	angleyMin_max (ACC)	0.32

family$_{two}$

	Feature	T	Feature	C
alone vs. group	anglexMax_min (ACC)	6.82****	mMin_min (ACC)	0.46
	anglexMedian_min (ACC)	6.82****	zAxisAvgStd_min (ACC)	0.46
	angleyMean_min (ACC)	6.64****	anglezMedian_min (ACC)	0.44
	anglexMin_min (ACC)	6.49****	dmListStd_min (ACC)	0.43
	angleyMax_min (ACC)	6.42****	zAxisAvgMean_min (ACC)	0.43

partner$_{two}$

	Feature	T	Feature	C
alone vs. group	food_and_drink (APP)	4.55****	yAxisAvgMax_avg (ACC)	0.43
	food_and_drink (APP)	4.54****	zAxisAvgMin_avg (ACC)	0.33
	food_and_drink (APP)	4.55****	proximityRecord_avg (PRO)	0.29
	zAxisAvgMin_avg (ACC)	4.17****	yAxisAvgStd_avg (ACC)	0.27
	zAxisAvgMean_avg (ACC)	3.35****	angleyStd_avg (ACC)	0.27

friends$_{two}$

	Feature	T	Feature	C
alone vs. group	mMean_avg (ACC)	8.10****	mMean_avg (ACC)	0.52
	mSMAMean_avg (ACC)	8.08***	mSMAMean_avg (ACC)	0.52
	mMedian_max (ACC)	7.97****	mMedian_avg (ACC)	0.50
	mSMAMedian_max (ACC)	7.94****	mSMAMedian_avg (ACC)	0.50
	mMax_avg (ACC)	7.89****	yAxisAvgStd_max (ACC)	0.50

feature. In addition, 95% confidence intervals for Cohen's-d were calculated, and if the interval does not overlap with zero, the difference can be considered as significant [60].

For the social context family$_{three}$, features from the bluetooth sensor were among the top five in terms of t-statistic and Cohen's-d, for the combination alone vs. sgroup. In addition, all the top five features had Cohen's-d values closer to medium effect size. Further, a total of 122 features had Cohen's-d values above small effect size and confidence intervals not including zero. For the combinations sgroup vs. lgroup and alone vs. lgroup, the majority of features were from the accelerometer and two features (video_player and system) were from application usage. In addition, if the hierarchy of the social contexts alone, sgroup, and lgroup is considered, sgroup is in the middle, sandwiched by alone and lgroup, that are further apart, hence, it would be easier to discern between these two groups. This is indicated in the results for the combination alone vs. lgroup, that have higher t-statistics

and Cohen's-d values (some around medium effect size) compared to the other two combinations (alone vs. sgroup and sgroup vs. lgroup). Furthermore, for the social contexts $friends_{three}$ and $people_{three}$, for all three combinations, all features in the top five in terms of both t-statistic and Cohen's-d are from the accelerometer. In addition, for $friends_{three}$, features in the combination alone vs. lgroup had high t-statistics and Cohen's-d values above medium effect size. In fact, 14 features, all of which are from the accelerometer had Cohen's-d values above medium effect size. In addition, for $people_{three}$, 44 features had Cohen's-d values above medium effect size, and the highest ones were closer to large effect size, meaning that these accelerometer feature could discriminate between alone and lgroup social contexts.

For two-class social contexts $people_{two}$, $family_{two}$, and $friends_{two}$, all features in the top five for both t-statistic and Cohen's-d were from the accelerometer. Further, only $friends_{two}$ had features with Cohen's-d above medium effect size among all four two-class social contexts. However, for $partner_{two}$, several features from application usage (food and drink app usage) were among the top five for t-statistics. In addition, a feature from the proximity sensor had a Cohen's-d of 0.29, which is above small effect size. As a summary, results from the statistical analysis suggest that for all the social contexts, accelerometer features could be informative of the group dynamic. In addition, for social contexts related to partner/spouse, app usage behavior and proximity sensors could be informative. Moreover, bluetooth sensor had high statistical significance in discriminating social contexts related to family members.

6 SOCIAL CONTEXT INFERENCE

6.1 Two-Class and Three-Class Social Context Inference (RQ2)

In this section, we use all the available smartphone sensing features and implement seven social context tasks, using features defined in Section 3.3 as target variables. The tasks include four two-class inference tasks and three three-class inference tasks: (1) $family_{two}$, (2) $partner_{two}$, (3) $friends_{two}$, (4) $people_{two}$, (5) $family_{three}$, (6) $friends_{three}$, and (7) $people_{three}$. In this phase, we used scikitlearn [84] and keras [18] frameworks together with python, and conducted experiments with several model types: (1) Random Forest Classifier [22], (2) Naive Bayes [90], (3) Gradient Boosting [79], (4) XGBoost [17], and (5) AdaBoost [94]. These models were chosen by considering the tabular nature of the dataset, interpretability of results, and small size of the dataset. In addition, we used the leave k-participants out strategy (k = 20) when conducting experiments, where testing and training splits did not have data from the same user, hence avoiding possible biases in experiments. Further, similar to recent ubicomp studies [5, 44, 74], we used the Synthetic Minority Over-sampling Technique (SMOTE) [16] to obtain training sets for each inference task. As recommended by Chawla et al. [16], when and where necessary, we under-sampled the majority class/classes to match over-sampled minority class/classes to create balanced datasets, hence not over-sampling unnecessarily beyond doubling the minority class size. In addition, we also calculated the area under the curve (AUC) (for three-class inferences, one vs. the rest technique, using macro averaging) using the receiver operator characteristics (ROC) curves. All experiments were repeated for ten iterations. We report mean and standard deviation of accuracies, and mean of AUC using results from the ten iterations.

Table 5 summarizes the results of the experiments. All the two-class inference tasks achieved accuracies over 80%. Moreover, all the three-class inferences achieved accuracies over 75%. When considering model types, Random Forest classifiers performed the best across five out of the seven inference tasks ($family_{two}$, $partner_{two}$, $friends_{two}$, $family_{three}$, and $people_{three}$) and Gradient Boosting had higher accuracies for two inference tasks ($people_{two}$ and $friends_{three}$). Generally, all models included in the study, except for Naive Bayes, performed reasonably well. Further, low standard deviation values suggest that regardless of the samples used for training and testing, the models generalized reasonably well. AUC scores followed a similar trend as the accuracy. These

Table 5. Mean (Ā) and Standard Deviation (A$_\sigma$) of inference accuracies and the mean area under the curve of the receiver operator characteristic curve (AUC), calculated from 10 iterations, using five different models, for two-class and three-class tasks, with attributes such as family, friends/colleagues, spouse/partner, and alone. Results are presented as: Ā (A$_\sigma$), AUC

	Target Variable	Random Forest	XG Boost	Ada Boost	Gradient Boost	Naive Bayes
two-class	baseline	50.0 (0.0), 50.0	50.0 (0.0), 50.0	50.0 (0.0), 50.0	50.0 (0.0), 50.0	50.0 (0.0), 50.0
	family$_{two}$	**86.1** (3.1), 84.8	82.6 (3.4), 73.2	82.5 (4.2), 71.7	83.2 (3.7), 74.2	65.6 (6.9), 67.6
	partner$_{two}$	**87.4** (2.6), 82.6	84.6 (4.6), 74.7	83.5 (3.8), 69.3	84.7 (5.1), 78.8	68.6 (8.2), 64.2
	friends$_{two}$	**80.1** (2.9), 81.3	78.3 (4.5), 75.2	78.0 (4.1), 70.4	78.7 (3.7), 73.7	64.0 (4.2), 61.4
	people$_{two}$	83.3 (3.2), 79.2	84.1 (3.1), 75.2	82.7 (4.3), 79.5	**84.2** (2.8), 76.9	72.3 (4.7), 67.3
three-class	baseline	33.3 (0.0), 50.0	33.3 (0.0), 50.0	33.3 (0.0), 50.0	33.3 (0.0), 50.0	33.3 (0.0), 50.0
	family$_{three}$	**85.9** (2.2), 80.5	81.4 (3.1), 73.2	73.9 (3.7), 63.2	83.0 (2.6), 70.2	63.1 (4.2), 61.8
	friends$_{three}$	76.7 (2.3), 78.2	77.1 (3.6), 70.1	71.0 (3.1), 68.8	**77.6** (2.8), 72.3	62.2 (5.1), 63.5
	people$_{three}$	**78.3** (2.1), 75.3	71.2 (2.6), 69.8	67.1 (3.8), 67.8	73.8 (3.2), 73.1	57.9 (6.7), 67.2

results suggest that passive mobile sensing features could be used to infer both two-class and three-class social contexts related to alcohol consumption, with reasonable performance.

6.2 Social Context Inference for Different Sensors (RQ2)

Prior work in mobile sensing has argued for multiple inference models for the same inference task, in the case of sensor failure [71, 92, 114]. For instance, during a weekend night, young adults could be concerned for the battery life of their phone, and could turn-off bluetooth, wifi, and location sensors that drain the battery faster. In such cases, having separate inference models that use different data sources to infer the same target attribute could be beneficial. In addition, prior work has segregated passive sensing modalities into Continuous Sensing (using embedded sensors in the smartphone) and Interaction Sensing (sensing the users' phone usage and interaction behavior) [71]. Considering these aspects, we conducted experiments for different feature groups based on the sensing modality (accelerometer, applications, battery, bluetooth, proximity, location, screen, and wifi) and the following feature group combinations that are meaningful in the context of drinking and young adults:

- **Continuous Sensing (ConSen):** These sensing modalities use embedded sensors to capture context. Examples are accelerometer, battery, bluetooth, proximity, location, and wifi. ConSen contains features from all these sensing modalities, and this feature group combination can measure the capability of the smartphone in inferring the social context of drinking, even if the user does not necessarily use the smartphone, because the considered sensing modalities sample data regardless of the phone usage behavior.
- **Interaction Sensing (IntSen):** These sensing modalities capture the phone usage and interaction behavior. Examples include screen events and application usage. In addition, these sensing modalities do not fail often because there is no straightforward way for users to turn-off interaction sensing modalities. Furthermore, these sensing modalities consume far less power compared to continuous sensing. In this context, this feature group combination could measure the capability of a smartphone to infer the social context of drinking, based on the way young adults use and interact with the smartphone.

For the two above mentioned feature groups, we conducted experiments using the same procedure as given in Section 6.2. Even though we got results for all models, we only present results for random forest classifiers in Table 6 because they output feature importance values which are useful to interpret results in Section 6.3, and they provide the results with highest accuracy and AUC values for a majority of inference tasks. Even though the accuracies were well above baselines for both two-class and three-class inference tasks, the lowest accuracies were recorded for SCR. This could either be because of the far too small dimensionality (only three features) or because the features were less informative. For the inference of social context partner$_{two}$, APP provided the highest accuracy of 82.92% followed by ACC that provided an accuracy of 81.21%. This suggests that the

Table 6. Social Context Inference accuracy breakdown for sensor type based feature groups and feature group combinations using Random Forest classifiers. Both the mean (Ā) and standard deviation (A_σ) of accuracies from cross validation are reported in addition to the mean area under the curve (AUC) from reciever operator characteristics graph (ROC)

Feature Group	two-class				three-class		
(# of features)	$family_{two}$ $\bar{A}$ (A_σ), AUC	$partner_{two}$ $\bar{A}$ (A_σ), AUC	$friends_{two}$ $\bar{A}$ (A_σ), AUC	$people_{two}$ $\bar{A}$ (A_σ), AUC	$family_{three}$ $\bar{A}$ (A_σ), AUC	$friends_{three}$ $\bar{A}$ (A_σ), AUC	$people_{three}$ $\bar{A}$ (A_σ), AUC
Baseline	50.0 (0.0), 50.0	50.0 (0.0), 50.0	50.0 (0.0), 50.0	50.0 (0.0), 50.0	33.3 (0.0), 50.0	33.3 (0.0), 50.0	33.3 (0.0), 50.0
ACC (150)	83.3 (2.4), 80.2	81.2 (3.1), 79.6	74.9 (3.0), 72.5	81.1 (3.1), 81.4	82.6 (1.9), 78.5	71.5 (2.5), 70.1	72.4 (2.7), 72.0
APP (105)	82.9 (3.5), 80.7	82.9 (3.0), 79.2	74.3 (2.4), 76.7	80.5 (2.5), 81.1	81.4 (2.4), 81.5	69.5 (3.1), 69.1	71.9 (2.3), 70.2
BAT (36)	78.7 (2.8), 76.7	77.5 (3.6), 77.3	71.5 (3.0), 73.6	78.1 (3.3), 72.1	77.5 (2.7), 75.6	66.1 (3.1), 67.8	68.1 (2.6), 68.4
BLU (27)	74.6 (2.8), 72.1	75.5 (3.6), 73.8	69.0 (2.8), 70.3	74.0 (3.1), 73.1	69.3 (2.9), 68.8	59.4 (2.7), 61.4	60.5 (2.8), 66.4
PRO (18)	74.1 (3.1), 71.9	75.6 (2.3), 74.5	69.9 (2.8), 68.2	75.8 (2.7), 76.8	70.4 (2.2), 73.2	59.1 (3.8), 61.9	60.7 (2.4), 62.9
LOC (39)	79.2 (3.0), 77.1	78.9 (2.7), 78.2	74.1 (3.2), 76.1	77.6 (2.6), 76.9	77.2 (2.6), 76.4	67.0 (3.3), 69.1	69.5 (2.9), 68.7
SCR (3)	68.3 (4.5), 61.1	69.7 (4.7), 60.3	64.5 (4.6), 62.8	71.9 (3.1), 67.2	62.2 (5.6), 60.7	54.1 (4.3), 55.2	54.8 (5.5), 56.1
WIF (36)	77.5 (2.9), 78.1	77.1 (2.9), 76.9	68.8 (3.7), 70.2	75.3 (3.9), 76.1	73.1 (1.9), 73.0	61.6 (2.9), 63.6	64.5 (2.8), 67.1
ConSen (306)	85.7 (2.6), 82.1	86.8 (2.9), 80.8	79.5 (3.2), 80.2	82.9 (2.1), 81.4	85.3 (1.9), 76.8	76.7 (2.7), 76.9	77.9 (3.0), 76.1
IntSen (96)	83.3 (2.0), 80.1	83.1 (2.5), 81.7	76.5 (2.9), 76.3	81.6 (2.5), 81.5	82.3 (2.7), 78.2	71.4 (2.7), 71.2	73.2 (2.7), 72.7
ALL (402)	86.1 (3.1), 84.8	87.4 (2.6), 82.6	80.1 (2.9), 81.3	83.3 (3.2), 79.2	85.9 (2.2), 80.5	76.7 (2.3), 78.2	78.3 (2.1), 75.3

app usage behavior during drinking events is informative of whether participants are with a partner/spouse or not. This could also be related to prior work regarding *partner phubbing* [19, 91] that could lead to relationship dissatisfaction and disappointment. People might try to avoid phubbing (hence use the phone less/differently than normal) when they are with their partner/spouse. Furthermore, except for this inference, for all other social context inferences, the highest accuracies were obtained using ACC (in the range of 71.52% to 83.33%). This suggests that physical activity levels and movement dynamics around drinking events could be used to infer social contexts such as family ($family_{two}$ and $family_{three}$), friends ($family_{two}$ and $family_{three}$), and people ($people_{two}$ and $people_{three}$). In addition, results from the AUC followed a similar trend to accuracies. For two-class inferences, except for SCR, all other modalities reported AUC scores above 70%. However, for three-class inferences, except for ACC, all other modalities reported AUC scores below 70%. Further, except for SCR, for all the other inferences, standard deviation scores were reasonably low, suggesting that inference results hold regardless of the training and testing splits. High standard deviations for SCR could be because of the low number of features, which was also reflected with low accuracies and AUC scores.

Feature group combinations ConSen and IntSen provided similar accuracies for all inference tasks even though ConSen achieved slightly better than IntSen for each inference. While ConSen outperformed ACC and APP for all the inferences, IntSen had slightly lower accuracies for social contexts $family_{three}$ (82.36%) and $friend_{three}$ (71.44%) as compared to ACC, that had accuracies 82.60% and 71.52% for the respective inferences. Standard deviation scores for both ConSen and IntSen were low. In addition, AUC scores too were above 70% for all cases, which is a reasonable result. Finally, the results suggest that IntSen could provide reasonably high accuracies as compared to ConSen in case of sensor failure, and in the worst case scenario, ACC provides fair accuracies for all the inference tasks, which is satisfactory given that it is just one sensing modality.

6.3 Feature Importance for Social Context Inferences (RQ2)

In Figure 9, we show the top twenty feature importance values for each inference presented in Section 6.2. These values were captured from the output of trained random forest models, when using all features. We obtained feature importance values for all features in each iteration, and report the mean value for each feature. The sensor modality that was present throughout all seven inferences was the accelerometer (ACC). This is congruent with the results presented in the statistical analysis (Section 5). This suggests that physical activity levels and phone motion dynamics could help infer different types of social contexts of drinking occasions. This makes sense for

(a) family$_{three}$

(b) friends$_{three}$

(c) people$_{three}$

(d) family$_{two}$

(e) partner$_{two}$

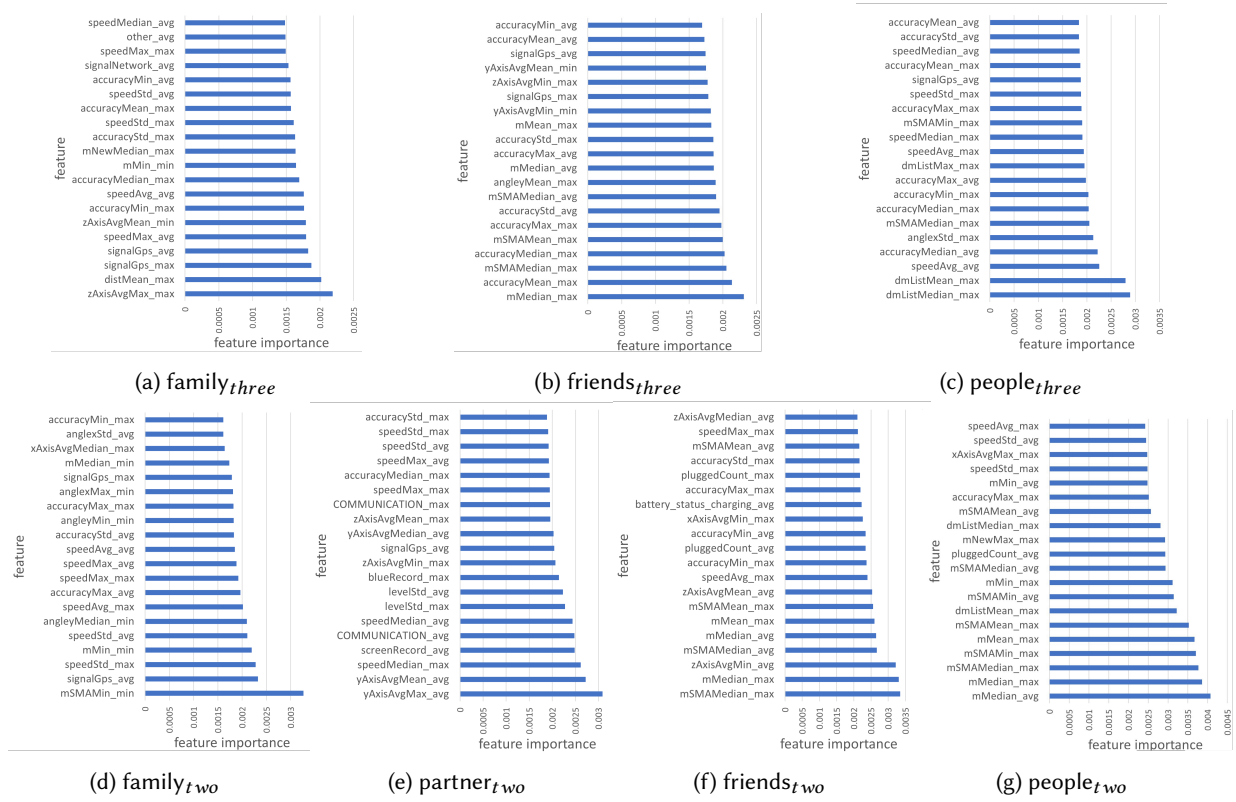(f) friends$_{two}$

(g) people$_{two}$

Fig. 9. Feature importance values from random forest classifiers with all features, for different social contexts.
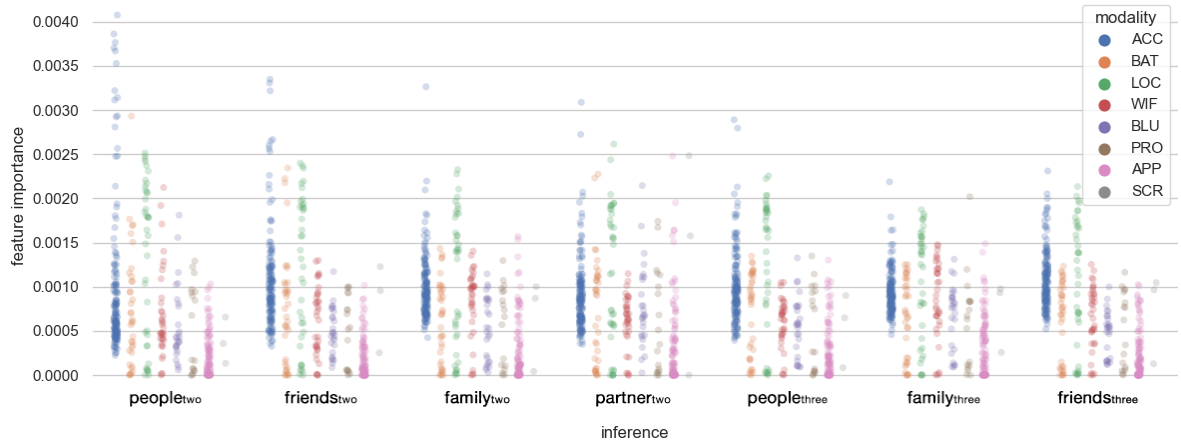

Fig. 10. Feature importance value distributions for different sensing modalities in different inferences when using all features

certain situations because it is highly unlikely that a person would drink and dance alone on a weekend night, while this might happen when people are in larger groups (with both family and friends).

The second most common modality across all inferences is location (LOC). Features that capture the speed of the phone (speedMedian_avg, speedMax_avg, etc.), accuracy of the signal (accuracyMin_max, accuracyMean_avg, etc.), and signal type and strength (signalGps_max, signalNetwork_avg, etc.) are present across all inferences. Specially, for both two-class and three-class inferences regarding family, location features regarding GPS signal strength and speed filled up a majority of top five features. This suggests that location-related features have captured certain differences with regards to group dynamics in the social context family. Even though interaction sensing modalities (APP, SCR) were not present among all the social contexts, $partner_{two}$ had several features (COMMUNICATION_avg, COMMUNICATION_max, etc.) regarding communication app usage (e.g. viber, whatsapp, messenger, etc.) and also screen usage (screenRecord_avg). Given interaction sensing modalities capture the phone usage behavior, this suggests that people use their phone differently when they are drinking alcohol with their partner as opposed to not being with him or her.

In Figure 10, we plot a distribution of feature importance values for all social context inferences, grouped by different sensing modalities. This provides an overview of the informativeness of sensing modalities in making inferences. The most sparse distribution across all inferences came from the ACC, for the social context $people_{two}$. Overall, accelerometer produced the most informative feature, for all seven social contexts. Location features had comparatively high values for all seven social contexts. Even though location features were not among the highest for any inference, mean feature importance for location modalities was even higher than for accelerometer features (because the location feature distribution is negatively skewed). In addition, except for WIF, all other modalities had comparatively sparse and wider distributions for the context $partner_{two}$. To sum up, the takeaways from this analysis are: accelerometer features (ACC) were informative for all inferences, location features (LOC) were generally informative across all inferences too, application usage (APP) and screen usage (SCR) features (interaction sensing) were informative for $partner_{two}$ while not being comparatively informative for other inferences, and except for Wifi features (WIF), all other features had wider distributions for $partner_{two}$.

6.4 Effect of Varying Group Sizes (RQ3)

In the previous analyses, we considered group dynamics as follows: (a) two-class social contexts - *without* vs. *with one/more people* and (b) three-class social contexts - *without* vs. *with one person* vs. *with two/more people*. Hence, while the two-class inference mostly relate to the absence of a particular type of people, the three classes effectively tried to infer the presence of groups of varying sizes (with one person, with two/more people). If we consider the three-class inferences, *with one person* means that it is a group of two people including the participant, and *with two/more people* means that the group has a minimum size of three people including the participant, hence both classes capture different group sizes, with the former being a small group, and the latter being a larger group in comparison. Given that there is no gold standard regarding the definition of the size of the drinking group (as highlighted in Section 2), in this section, we aim to change the size of these two groups by changing the threshold called *grouping threshold*, which was always equal to *one* in previous sections (e.g., without vs. with *one*/less people vs. with two/more people), for three-class inferences. To this aim, we increase the value of the grouping threshold from one to ten, to investigate how it affects the inference accuracy. One and ten were chosen as highest and lowest thresholds because those were the highest and lowest values available in self-reports to define three-classes.

We conducted the evaluation with the three three-class inferences using the same approach mentioned in Section 6.1, and the results are summarized in Figure 11. For $friend_{three}$ and $people_{three}$, inference accuracies decreased when increasing the grouping threshold, meaning that the model was not good at discerning the three classes when the threshold was around three (alone vs. with three/less people vs. with four/more people) and

four (alone vs. with four/less people vs. with five/more people). However, when increasing the threshold further, the accuracies increased back to the same level as when the threshold was equal to one. What this means is that random forest classifier is not performing well when the small and large groups are defined by thresholds of the range three to five. This result is not surprising because any kind of nightlife-related activities available for a small group of people (they might find a table to fit altogether in a pub or a restaurant, they might easily travel with a cab, they might all gather in a living room) would result in a large heterogeneity of sensor data as compared to a larger group (e.g. ten or more people). This is because of the differences in behavior when people are in large groups, as opposed to small groups. Consequently, this would result in a lower inference accuracy when social contexts with three, four, or five people are in both the small group and the large group classes of the three-class inference. Consider an example where the grouping threshold is three, where samples with a group of three people would fall into the small group and samples with a group of four or more people would be included in the large group of the three-class inference.

According to the distribution given in Figure 3, for the variable $friends_{three}$, when 114 samples (group of 3) and 105 samples (group of 4) fall into small and large groups in the inference, both classes have homogeneous sensor data, hence making it difficult for the model to discriminate between the classes. Conversely, the range of activities is smaller for larger groups due to its size, resulting in a lower heterogeneity of sensor data within groups, and consequently, higher inference accuracy for higher grouping thresholds. For example, consider the grouping threshold of ten where the small group would have ten or less people and the large group would have eleven or more people. According to Figure 3, for the variable $people_{three}$, there are 221 samples of eleven or more (clearly a large group), and over 300 samples of groups with three to ten people, with a majority of data

Fig. 11. Three-Class Social Context Inference Accuracies for Different Grouping Thresholds

coming from small group sizes such as three (135 samples), four (93 samples), and five (119 samples). This leads to heterogeneous data between small and large groups, because small group consists of data predominantly from groups of three, four, or five, and the large group is predominantly containing groups of 10+ people. This makes it easier for the model to discriminate between the three classes, hence leading to higher accuracies. On the other hand, for $family_{three}$, increasing threshold had an opposite effect, and increased the performance of the models. Again, this might be explained by the lower diversity of choices of activities and contexts to be sensed in family contexts, that tend to be highly routinized. Finally, results suggest that, regardless of the grouping threshold, models performed reasonably well for all $family_{three}$ inferences. In addition, except for grouping thresholds from three to five, for all other threshold, $friends_{three}$ and $people_{three}$ showed reasonable performance with accuracies over 70%. Hence, according to this analysis, having different grouping thresholds seems a valid design choice depending on the application and the use-case.
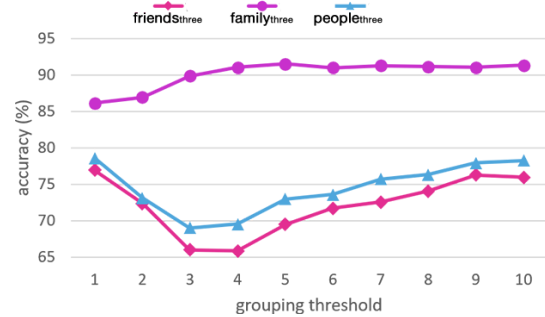
6.5 Sex Composition of Groups of Friends (RQ3)

In the related work section, we described the importance of identifying the gender composition of people in groups when consuming alcohol. For example, we described how prior work discussed about men or women feeling more comfortable when drinking with groups of same sex friends [2, 107]. In this section, we define and evaluate a three-class inference task for drinking episodes that are done with friends/colleagues (N = 799), with the classes: same-sex (389), opposite-sex (97), and mixed-sex (313). This feature was derived using the demographic sex

Table 7. Sex-Composition Inference results for alcohol consumption episodes done with friends, in the order: Ā (A$_\sigma$), AUC

Target Variable	Random Forest	XG Boost	Ada Boost	Gradient Boost	Naive Bayes
baseline	33.3 (0.0), 50.0	33.3 (0.0), 50.0	33.3 (0.0), 50.0	33.3 (0.0), 50.0	33.3 (0.0), 50.0
sex_composition$_{three}$	75.8 (2.5), 74.9	69.6 (4.1), 70.7	66.1 (6.1), 66.2	71.5 (3.7), 70.9	58.7 (7.3), 59.9

attribute of the participant and the men and women friends/colleagues present in the drinking occasion, as reported by the participants. We followed the same approach as in Section 6.1 to conduct the evaluation, and the results are presented in Table 7. According to results, the random forest classifier performed the best with an accuracy of 75.86%, followed by gradient boosting that had an accuracy of 71.57%. The ten highest feature importance values for the inference that were obtained using the random forest classifiers included six features from LOC (related to GPS signal strength and the accuracy of the signal, e.g. accuracyMedian_max, accuracyMax_max, etc.) and four from the ACC (reading from the z axis, e.g. zAxisAvgMedian_avg, zAxisAvgMin_avg; and aggregated m statistic - mMedian_max). This result suggests that mobile sensing features can to some degree classify the sex composition of drinking groups, but more in-depth work would be needed to understand this phenomenon.

7 DISCUSSION

Features: It is worth noting that for modalities such as ACC and LOC, we generated simple statistical features that do not need extensive processing of the dataset. If we consider the ACC, while features proved to be informative in inferring different social contexts, the only set of features we used are statistical features from the three axes, angles between the gravity vector and axes, and aggregate features that combine the values of three-axes (Section 3.1.3). It is also worth noting that these feature are less interpretable in the context of alcohol consumption. For example, the feature mMedian_max had the highest feature importance value for friends$_{three}$, as shown in Figure 9a. While this feature represents the overall acceleration of the phone at a time period closer to the drinking event, it is difficult to interpret it compared to more interpretable features such as step count or activity type. If such features were derived using the accelerometer data, the interpretation could have been much simpler. However, we were not able to derive them due to limitations in the original dataset (sampling frequency, lack of gyroscope data, etc.). Future work could consider using low-power consuming libraries such as Google Activity Recognition API to obtain activity types and native step counters available in modern smartphones to obtain step counts, hence obtaining more interpretable features. In addition, researchers could also look into using other sensing modalities such as ambient light sensor, typing and touch events, and notification clicking behaviors.

Ethical Considerations: The goal of this study is to support public health research. Hence, it is essential to be aware of ethical implications. For public health, the inferences done in this work are anonymous in the sense that no identities of individuals are inferred when inferring social contexts. However, certain social contexts such as 'being with a partner' could be more sensitive, because identifying the presence of such people could potentially reveal sensitive information about them, even though they might not have agreed to have their location indirectly reported. Given that social context is relational, it is critical that during data collection, social companions (friends, family, etc.) agree that their presence is reported (even as an aggregate). Future studies should consider these aspects. Furthermore, for future interactive health systems that would be used by individuals and their health providers, it is fundamental to have clarity on who could access inferred data regarding social contexts, given their sensitive nature. Further, running social context inferences on-device, rather than on servers, would help preserve the privacy of users and others interacting with them. More generally, participants' respect of privacy and well-being should be the guiding lights of any future design of mobile health systems regarding alcohol consumption.

Limitations and Future Work: We prepared the drinking event level dataset (in Section 3.2) without assuming any relationship between two drinking events that occur consecutively, hence, we considered alcohol drinking events to be independent of each other. However, in reality, there could be a relationship between drinking events of the same person during the same night. Understanding such relationships is a complex problem, and it needs further examination. Another limitation of our work is that it does not capture complex relationships within family members. For example, young adults might prefer drinking with their brother or same-age members of the family, where as they might not feel comfortable drinking with their parents. In addition, the perception of parents and other family members could differ significantly, and it could affect the drinking behavior in the vicinity of family members. Furthermore, the partner's/spouse's perception towards alcohol consumption is another variable that was not captured during this study. These aspects need further investigation. In addition, it is worthwhile to note that, inferring the social context of drinking does not directly help overcome health problems. This is not the intention of this work, as it would oversimplify the problem. However, inferring the social context of drinking would assist or complement other inferences such as drinking occasions, drinking nights, drink vs. drunk in ubicomp and alcohol research [5, 6, 27, 36, 86, 87, 92]. In this respect, the inference of social context might help to provide meaningful and context-aware interventions that might decrease the amounts consumed, and, as a consequence, less adverse alcohol-related consequences. The design of such interventions is beyond the scope of the paper.

Another important aspect is the choice of time windows for aggregation and matching phases. Even though we presented results for the dataset obtained with ten-minute time window for aggregation and one-hour time window for matching, we conducted evaluations with different time windows. We obtained the best results using these time windows, and hence, considering space limitations, we only presented results for these windows. It is worth noting that the time window would affect the number of self-reports included in the study. For instance, if the matching time window is two hours, we need to discard all self-reports from 8.00pm to 9.00pm because we would not have enough sensor data for reports done between those time windows. The same applies to drinking events done between 3.00am and 4.00am. Further, it is worth noting that, regardless of the time window and the resulting dataset size, we obtained inference results comparable to the ones we presented in Table 5, with differences of the range 0.4% (best-case scenario) to 12% (worst-case scenario).

An important topic for future work is regarding the drinking motives of young adults. As we discussed in Section 2, drinking motives could be the primary driving factor why young adults choose specific social contexts to drink [47, 48]. Hence, examining the associations between such drinking motives and smartphone sensing data could further advocate the idea of building holistic mobile health systems that consider not only the alcohol consumption, but also other factors associated to the event. Furthermore, even though there were multiple comparisons in the statistical analysis, we did not use Bonferroni correction for p-values [111]. Hence, the results with p-values should be interpreted with caution.

Importance of Diversity-Awareness: The drinking behavior of people differ significantly depending on age, sex, drinking culture, beverage preferences, as well as how people perceive drinking alcohol [6, 35, 92]. For example, in some Asian countries, drinking alcohol might not be socially accepted while it is a societal norm in Europe and North America [7, 101]. Hence, it is worth pointing out that this study regarding the drinking behavior of young adults in Switzerland is exploratory, and the results cannot be directly assumed as being representative of the drinking behavior in other countries. Recent work has highlighted the importance of considering diversity-awareness when building social platforms using machine learning models and mobile sensing data [42, 95].

8 CONCLUSION

In this study, we examined the weekend drinking behavior of 241 young adults in Switzerland using self-reports and passive smartphone sensing data. Our work emphasized the importance of understanding the social context of drinking, to obtain a holistic view regarding the alcohol consumption behavior. With multiple statistical analyses, we show that features from modalities such as accelerometer, location, bluetooth, and application usage could be informative about social contexts of drinking. In addition, we define and evaluate seven inference tasks obtaining accuracies of the range 75%-86% in two-class and three-class tasks, showing the feasibility of using smartphone sensing to detect social contexts of drinking occasions. We believe these findings could be useful for ubicomp and alcohol epidemiology researchers towards in implementing future mobile health systems with interventions and feedback mechanisms.

ACKNOWLEDGMENTS

REFERENCES

[1] 2014. *Youth@Night.* Retrieved February 8, 2021 from http://www.youth-night.ch/

[2] Birgitta Ander, Agneta Abrahamsson, and Disa Bergnehr. 2017. 'It is ok to be drunk, but not too drunk': party socialising, drinking ideals, and learning trajectories in Swedish adolescent discourse on alcohol use. *Journal of Youth Studies* 20, 7 (Aug. 2017), 841–854.

[3] Z. Arnold, D. Larose, and E. Agu. 2015. Smartphone Inference of Alcohol Consumption Levels from Gait. In *2015 International Conference on Healthcare Informatics*. 417–426.

[4] Sophie Attwood, Hannah Parke, John Larsen, and Katie Morton. 2017. Using a mobile health application to reduce alcohol consumption: a mixed-methods evaluation of the drinkaware track & calculate units application. *BMC Public Health* 17 (05 2017).

[5] Sangwon Bae, Tammy Chung, Denzil Ferreira, Anind K. Dey, and Brian Suffoletto. 2018. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive Behaviors* 83 (2018), 42–47. Ambulatory Assessment of Addictive Disorders.

[6] Sangwon Bae, Denzil Ferreira, Brian Suffoletto, Juan C. Puyana, Ryan Kurtz, Tammy Chung, and Anind K. Dey. 2017. Detecting Drinking Episodes in Young Adults Using Smartphone-Based Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 5 (June 2017), 36 pages.

[7] Yatan Balhara and Shachi Mathur. 2012. Alcohol: A Major Public Health Problem – South Asian Perspective. *Addictive Disorders & Their Treatment* 11 (06 2012), 101–120.

[8] Kenneth Beck, Amelia Arria, Kimberly Caldeira, Kathryn Vincent, Kevin O'Grady, and Eric Wish. 2008. Social Context of Drinking and Alcohol Problems Among College Students. *American journal of health behavior* 32 (07 2008), 420–30.

[9] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 37 (Sept. 2017), 20 pages.

[10] Edward J. Bedrick. 2005. *Biserial Correlation.* American Cancer Society.

[11] Melina Bersamin, Sharon Lipperman-Kreda, Christina Mair, Joel Grube, and Paul Gruenewald. 2016. Identifying Strategies to Limit Youth Drinking in the Home. *Journal of Studies on Alcohol and Drugs* 77, 6 (Nov. 2016), 943–949.

[12] Joan-Isaac Biel, Nathalie Martin, David Labbe, and Daniel Gatica-Perez. 2018. Bites'N'Bits: Inferring Eating Behavior from Contextual Mobile Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 125 (Jan. 2018), 33 pages.

[13] Yekta Said Can, Bert Arnrich, and Cem Ersoy. 2019. Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *Journal of Biomedical Informatics* 92 (2019), 103139.

[14] Luca Canzian and Mirco Musolesi. 2015. Trajectories of Depression: Unobtrusive Monitoring of Depressive States by Means of Smartphone Mobility Traces Analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 1293–1304.

[15] Giuseppe Carra, Cristina Crocamo, Francesco Bartoli, Daniele Carretta, Alessandro Schivalocchi, Paul Bebbington, and Massimo Clerici. 2016. Impact of a Mobile E-Health Intervention on Binge Drinking in Young People: The Digital-Alcohol Risk Alertness Notifying Network for Adolescents and Young Adults Project. *Journal of Adolescent Health* 58 (05 2016), 520–6.

[16] Nitesh Chawla, Kevin Bowyer, Lawrence Hall, and W. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res. (JAIR)* 16 (06 2002), 321–357.

[17] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. *CoRR* abs/1603.02754 (2016).

[18] François Chollet. 2015. keras. https://github.com/fchollet/keras.

[19] Varoth Chotpitayasunondh and Karen M. Douglas. 2016. How phubbing becomes the norm: The antecedents and consequences of snubbing via smartphone. *Computers in Human Behavior* 63 (2016), 9 – 18.

[20] M. J. Cox, K. Sewell, K. L. Egan, S. Baird, C. Eby, K. Ellis, and J. Kuteh. 2019. A systematic review of high-risk environmental circumstances for adolescent drinking. *Journal of Substance Use* 24, 5 (Sept. 2019), 465–474.

[21] David Crane, Claire Garnett, Susan Michie, Robert West, and Jamie Brown. 2018. A smartphone app to reduce excessive alcohol consumption: Identifying the effectiveness of intervention components in a factorial randomised control trial. *Scientific Reports* 8 (03 2018).

[22] Adele Cutler, David Cutler, and John Stevens. 2011. *Random Forests*. Vol. 45. 157–176.

[23] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan. 2010. Mobile phone based drunk driving detection. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. 1–8.

[24] Victor-Alexandru Darvariu, Laura Convertino, Abhinav Mehrotra, and Mirco Musolesi. 2020. Quantifying the Relationships between Everyday Objects and Emotional States through Deep Learning Based Image Analysis Using Smartphones. 4, 1, Article 7 (March 2020), 21 pages.

[25] Emma L Davies, Adam J Lonsdale, Sarah E Hennelly, Adam R Winstock, and David R Foxcroft. 2017. Personalized Digital Interventions Showed no Impact on Risky Drinking in Young Adults: A Pilot Randomized Controlled Trial. *Alcohol and Alcoholism* 52, 6 (08 2017), 671–676.

[26] Danielle M. Dick, Jason L. Pagan, Candice Holliday, Richard Viken, Lea Pulkkinen, Jaakko Kaprio, and Richard J. Rose. 2007. Gender Differences in Friends Influences on Adolescent Drinking: A Genetic Epidemiological Study. *Alcoholism: Clinical and Experimental Research* 31, 12 (2007), 2012–2019.

[27] Patrick Dulin, Vivian Gonzalez, Diane King, Danielle Giroux, and Samantha Bacon. 2013. Smartphone-Based, Self-Administered Intervention System for Alcohol Use Disorders: Theory and Empirical Evidence Basis. *Alcoholism treatment quarterly* 31 (07 2013).

[28] Bridget Freisthler, Sharon Lipperman-Kreda, Melina Bersamin, and Paul Gruenewald. 2014. Tracking the when, where, and with whom of alcohol use: Integrating ecological momentary assessment and geospatial data to examine risk for alcohol-related problems. *Alcohol Research: Current Reviews* 36, 1 (2014).

[29] Bettina Friese and Joel W. Grube. 2014. Teen Parties: Who Has Parties, What Predicts Whether There Is Alcohol and Who Supplies the Alcohol? *The Journal of Primary Prevention* 35, 6 (Dec. 2014), 391–396.

[30] Owen Gallupe and Martin Bouchard. 2013. Adolescent parties and substance use: A situational approach to peer influence. *Journal of Criminal Justice* 41, 3 (May 2013), 162–171.

[31] Conor Gilligan, Kypros Kypri, Natalie Johnson, Marita Lynagh, and Stephanie Love. 2012. Parental supply of alcohol and adolescent risky drinking: Parental supply of alcohol and adolescent drinking. *Drug and Alcohol Review* 31, 6 (Sept. 2012), 754–762.

[32] Gerhard Gmel, Jacques Gaume, Mohamed Faouzi, Jean-Pierre Kulling, and Jean-Bernard Daeppen. 2008. Who Drinks Most of the Total Alcohol in Young Men—Risky Single Occasion Drinking as Normative Behaviour. *Alcohol and Alcoholism* 43, 6 (2008), 692–697.

[33] Google. 2021. *Choose a category and tags for your app or game.* Retrieved January 29, 2021 from https://bit.ly/39r8zPp

[34] Greenland Sander, Senn Stephen J., Rothman Kenneth J., Carlin John B., Poole Charles, Goodman Steven N., and Altman Douglas G. 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology* 31, 4 (2016), 337–350.

[35] Ulrike Grittner, Sharon Wilsnack, Sandra Kuntsche, Thomas Greenfield, Richard Wilsnack, Arlinda Kristjanson, and Kim Bloomfield. 2019. A Multilevel Analysis of Regional and Gender Differences in the Drinking Behavior of 23 Countries. *Substance Use & Misuse* 55 (12 2019), 1–15.

[36] David Gustafson, Fiona Mctavish, Ming-Yuan Chih, Amy Atwood, Roberta Johnson, Michael Boyle, Michael Levy, Hilary Driscoll, Steven Chisholm, Lisa Dillenburg, Andrew Isham, and Dhavan Shah. 2014. A Smartphone Application to Support Recovery From Alcoholism A Randomized Clinical Trial. *JAMA psychiatry* 71 (03 2014).

[37] Leanne Hides, Catherine Quinn, Wendell Cockshaw, Stoyan Stoyanov, Oksana Zelenko, Daniel Johnson, Dian Tjondronegoro, Lake-Hui Quek, and David J. Kavanagh. 2018. Efficacy and outcomes of a mobile app targeting alcohol use in young people. *Addictive Behaviors* 77 (2018), 89–95.

[38] Kristina M. Jackson, Jennifer E. Merrill, Nancy P. Barnett, Suzanne M. Colby, Caitlin C. Abar, Michelle L. Rogers, and Kerri L. Hayes. 2016. Contextual influences on early drinking: Characteristics of drinking and nondrinking days. *Psychology of Addictive Behaviors* 30, 5 (2016), 566–577.

[39] Jisu Jung, Lyndal Wellard-Cole, Colin Cai, Irena Koprinska, Kalina Yacef, Margaret Allman-Farinelli, and Judy Kay. 2020. Foundations for Systematic Evaluation and Benchmarking of a Mobile Food Logger in a Large-Scale Nutrition Study. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 47 (June 2020), 25 pages.

[40] Hsin-Liu (Cindy) Kao, Bo-Jhang Ho, Allan C. Lin, and Hao-Hua Chu. 2012. Phone-Based Gait Analysis to Detect Alcohol Usage. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing

Machinery, New York, NY, USA, 661–662.

[41] Dean Karantonis, Michael Narayanan, Merryn Mathie, Nigel Lovell, and B.G. Celler. 2006. Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring. *Information Technology in Biomedicine, IEEE Transactions on* 10 (02 2006), 156 – 167.

[42] Mohammed Khwaja, Sumer S. Vaid, Sara Zannone, Gabriella M. Harari, A. Aldo Faisal, and Aleksandar Matic. 2019. Modeling Personality vs. Modeling Personalidad: In-the-Wild Mobile Data Analysis in Five Countries Suggests Cultural Impact on Personality Models. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 88 (Sept. 2019), 24 pages.

[43] Tae Kim. 2015. T test as a parametric statistic. *Korean Journal of Anesthesiology* 68 (11 2015), 540.

[44] Takumi Kondo, Haruka Kamachi, Shun Ishii, Anna Yokokubo, and Guillaume Lopez. 2019. Robust Classification of Eating Sound Collected in Natural Meal Environment. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) *(UbiComp/ISWC '19 Adjunct).* Association for Computing Machinery, New York, NY, USA, 105–108.

[45] Santosh Kumar, Wendy J. Nilsen, Amy Abernethy, Audie Atienza, Kevin Patrick, Misha Pavel, William T. Riley, Albert Shar, Bonnie Spring, Donna Spruijt-Metz, Donald Hedeker, Vasant Honavar, Richard Kravitz, R. Craig Lefebvre, David C. Mohr, Susan A. Murphy, Charlene Quinn, Vladimir Shusterman, and Dallas Swendeman. 2013. Mobile Health Technology Evaluation: The mHealth Evidence Workshop. *American Journal of Preventive Medicine* 45, 2 (2013), 228 – 236.

[46] Emmanuel Kuntsche and Gerhard Gmel. 2013. Alcohol consumption in late adolescence and early adulthood - where is the problem? *Swiss medical weekly* 143 (07 2013), 0.

[47] Emmanuel Kuntsche, Ronald Knibbe, Gerhard Gmel, and Rutger Engels. 2005. Why do young people drink? A review of drinking motives. *Clinical Psychology Review* 25, 7 (2005), 841 – 861.

[48] Emmanuel Kuntsche and Sandra Kuntsche. 2009. Development and Validation of the Drinking Motive Questionnaire Revised Short Form (DMQ–R SF). *Journal of clinical child and adolescent psychology : the official journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53* 38 (11 2009), 899–908.

[49] Emmanuel Kuntsche, Sandra Kuntsche, Johannes Thrul, and Gerhard Gmel. 2017. Binge drinking: Health impact, prevalence, correlates and interventions. *Psychology & health* 32 (05 2017), 1–42.

[50] Emmanuel Kuntsche and Florian Labhart. 2014. The future is now-using personal cellphones to gather data on substance use and related factors: Editorial. *Addiction* 109, 7 (July 2014), 1052–1053.

[51] Emmanuel Kuntsche and Florian Labhart. 2014. The future is now–using personal cellphones to gather data on substance use and related factors. *Addiction* 109, 7 (2014), 1052–1053.

[52] Florian Labhart, Michael Livingston, Rutger Engels, and Emmanuel Kuntsche. 2018. After how many drinks does someone experience acute consequences–determining thresholds for binge drinking based on two event-level studies. *Addiction* 113, 12 (2018), 2235–2244.

[53] Florian Labhart, Skanda Muralidhar, Benoit Massé, Lakmal Meegahapola, Emmanuel Kuntsche, and Daniel Gatica-Perez. 2021. Ten seconds of my nights: Exploring methods to measure brightness, loudness and attendance and their associations with alcohol use from video clips. *PLOS ONE* 16, 4 (04 2021), 1–21.

[54] Florian Labhart, Flavio Tarsetti, Olivier Bornet, Darshan Santani, Jasmine Truong, Sara Landolt, Daniel Gatica-Perez, and Emmanuel Kuntsche. 2020. Capturing drinking and nightlife behaviours and their social and physical context with a smartphone application – investigation of users' experience and reactivity. *Addiction Research & Theory* 28, 1 (2020), 62–75.

[55] Florian Labhart, Samantha Wells, Kathryn Graham, and Emmanuel Kuntsche. 2014. Do Individual and Situational Factors Explain the Link Between Predrinking and Heavier Alcohol Consumption? An Event-Level Study of Types of Beverage Consumed and Social Context. *Alcohol and Alcoholism* 49, 3 (May 2014), 327–335.

[56] Daniël Lakens. 2013. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 4: 863. *Frontiers in psychology* 4 (11 2013), 863.

[57] Nicholas D. Lane, Mu Lin, Mashfiqui Mohammod, Xiaochao Yang, Hong Lu, Giuseppe Cardone, Shahid Ali, Afsaneh Doryab, Ethan Berke, Andrew T. Campbell, and Tanzeem Choudhury. 2014. BeWell: Sensing Sleep, Physical Activities and Social Interactions to Promote Wellbeing. *Mob. Netw. Appl.* 19, 3 (June 2014), 345–359.

[58] Anne-Marie Laslett, Robin Room, Jason Ferris, Claire Wilkinson, Michael Livingston, and Janette Mugavin. 2011. Surveying the range and magnitude of alcohol's harm to others in Australia. *Addiction* 106, 9 (2011), 1603–1611.

[59] Katrin Leadley, Catherine L Clark, and Raul Caetano. 2000. Couples' Drinking Patterns, Intimate Partner Violence, and Alcohol-Related Partnership Problems. *Journal of Substance Abuse* 11, 3 (2000), 253 – 263.

[60] Dong Kyu Lee. 2016. Alternatives to P value: confidence interval and effect size. In *Korean journal of anesthesiology.*

[61] Kenneth Leonard and Pamela Mudar. 2003. Peer and partner drinking and the transition to marriage: A longitudinal examination of selection and influence processes. *Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors* 17 (07 2003), 115–25.

[62] Robert LiKamWa, Yunxin Liu, Nicholas D. Lane, and Lin Zhong. 2013. MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services* (Taipei, Taiwan)

(MobiSys '13). Association for Computing Machinery, New York, NY, USA, 389–402.

[63] Mu Lin, Nicholas D. Lane, Mashfiqui Mohammod, Xiaochao Yang, Hong Lu, Giuseppe Cardone, Shahid Ali, Afsaneh Doryab, Ethan Berke, Andrew T. Campbell, and Tanzeem Choudhury. 2012. BeWell+: Multi-dimensional Wellbeing Monitoring with Community-guided User Feedback and Energy Optimization. In *Proceedings of the Conference on Wireless Health* (San Diego, California) *(WH '12).* ACM, New York, NY, USA, Article 10, 8 pages.

[64] Tomi Lintonen, Salme Ahlstrom, and Leena Metso. 2004. The Reliability of Self-Reported Drinking in Adolescence. *Alcohol and Alcoholism* 39, 4 (07 2004), 362–368.

[65] Sharon Lipperman-Kreda, Laura J. Finan, and Joel W. Grube. 2018. Social and situational characteristics associated with adolescents' drinking at party and non-party events. *Addictive Behaviors* 83 (Aug. 2018), 148–153.

[66] Claire G. Lisco, Dominic J. Parrott, and Andra Teten Tharp. 2012. The role of heavy episodic drinking and hostile sexism in men's sexual aggression toward female intimate partners. *Addictive Behaviors* 37, 11 (2012), 1264 – 1270.

[67] Hong Lu, Denise Frauendorfer, Mashfiqui Rabbi, Marianne Schmid Mast, Gokul T. Chittaranjan, Andrew T. Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. 2012. StressSense: Detecting Stress in Unconstrained Acoustic Environments Using Smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12).* ACM, New York, NY, USA, 351–360.

[68] E. Jane Marshall. 2014. Adolescent Alcohol Use: Risks and Consequences. *Alcohol and Alcoholism* 49, 2 (01 2014), 160–164.

[69] M.J. Mathie, B.G. Celler, Nigel Lovell, and Adelle Coster. 2004. Classification of basic daily movements using a triaxial accelerometer. *Medical & biological engineering & computing* 42 (10 2004), 679–87.

[70] Dennis McCarty. 1985. *Environmental factors in substance abuse.* Springer, 247–281.

[71] Lakmal Meegahapola and Daniel Gatica-Perez. 2021. Smartphone Sensing for the Well-Being of Young Adults: A Review. *IEEE Access* 9 (2021), 3374–3399.

[72] Lakmal Meegahapola, Salvador Ruiz-Correa, and Daniel Gatica-Perez. 2020. Alone or With Others? Understanding Eating Episodes of College Students with Mobile Sensing. In *19th International Conference on Mobile and Ubiquitous Multimedia* (Essen, Germany) *(MUM 2020).* Association for Computing Machinery, New York, NY, USA, 162–166.

[73] Lakmal Meegahapola, Salvador Ruiz-Correa, and Daniel Gatica-Perez. 2020. Protecting Mobile Food Diaries from Getting Too Personal. In *19th International Conference on Mobile and Ubiquitous Multimedia* (Essen, Germany) *(MUM 2020).* Association for Computing Machinery, New York, NY, USA, 212–222.

[74] Lakmal Meegahapola, Salvador Ruiz-Correa, Viridiana del Carmen Robledo-Valero, Emilio Ernesto Hernandez-Huerfano, Leonardo Alvarez-Rivera, Ronald Chenu-Abente, and Daniel Gatica-Perez. 2021. One More Bite? Inferring Food Consumption Level of College Students using Smartphone Sensing and Self-Reports. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 26 (2021), 28 pages.

[75] Jochen Meyer, Judy Kay, Daniel A. Epstein, Parisa Eslambolchilar, and Lie Ming Tang. 2020. A Life of Data: Characteristics and Challenges of Very Long Term Self-Tracking for Health and Wellness. *ACM Trans. Comput. Healthcare* 1, 2, Article 11 (March 2020), 4 pages.

[76] C D Mohr, S Averna, D A Kenny, and F K Del Boca. 2001. "Getting by (or getting high) with a little help from my friends": an examination of adult alcoholics' friendships. *Journal of Studies on Alcohol* 62, 5 (2001), 637–645.

[77] Mehrab Bin Morshed, Samruddhi Shreeram Kulkarni, Richard Li, Koustuv Saha, Leah Galante Roper, Lama Nachman, Hong Lu, Lucia Mirabella, Sanjeev Srivastava, Munmun De Choudhury, Kaya de Barbaro, Thomas Ploetz, and Gregory D Abowd. 2020. A Real-Time Eating Detection System for Capturing Eating Moments and Triggering Ecological Momentary Assessments to Obtain Further Context: System Development and Validation Study. *JMIR Mhealth Uhealth* 4, 3 (18 dec 2020), e20625.

[78] Inbal Nahum-Shani, Shawna Smith, Bonnie Spring, Linda Collins, Katie Witkiewitz, Ambuj Tewari, and Susan Murphy. 2016. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine* 52 (09 2016).

[79] Alexey Natekin and Alois Knoll. 2013. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics* 7 (2013), 21.

[80] Jeremy Northcote and Michael Livingston. 2011. Accuracy of Self-Reported Drinking: Observational Verification of 'Last Occasion' Drink Estimates of Young Adults. *Alcohol and Alcoholism* 46, 6 (09 2011), 709–713.

[81] Renee O'Donnell, Ben Richardson, Matthew Fuller-Tyszkiewicz, and Petra Staiger. 2019. Delivering Personalized Protective Behavioral Drinking Strategies via a Smartphone Intervention: a Pilot Study. *International Journal of Behavioral Medicine* 26 (06 2019).

[82] D. Wayne Osgood, Daniel T. Ragan, Lacey Wallace, Scott D. Gest, Mark E. Feinberg, and James Moody. 2013. Peers and the Emergence of Alcohol Use: Influence and Selection Processes in Adolescent Friendship Networks. *Journal of Research on Adolescence* 23, 3 (2013), 500–512.

[83] Hilde Pape and Elin K. Bye. 2017. Drinking with parents: Different measures, different associations with underage heavy drinking? *Nordisk alkohol- & narkotikatidskrift: NAT* 34, 6 (Dec. 2017), 445–455.

[84] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine*

Learning Research 12 (2011), 2825–2830.

[85] Peggy L. Peterson, J. David Hawkins, Robert D. Abbott, and Richard F. Catalano. 1994. Disentangling the Effects of Parental Drinking, Family Management, and Parental Alcohol Norms on Current Drinking by Black and White Adolescents. *Journal of Research on Adolescence* 4, 2 (1994), 203–227.

[86] Thanh-Trung Phan, Florian Labhart, Skanda Muralidhar, and Daniel Gatica-Perez. 2020. Understanding Heavy Drinking at Night through Smartphone Sensing and Active Human Engagement. In *Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Atlanta, GA, USA) *(PervasiveHealth '20)*. Association for Computing Machinery, New York, NY, USA, 211–222.

[87] Thanh-Trung Phan, Skanda Muralidhar, and Daniel Gatica-Perez. 2019. Drinks & Crowds: Characterizing Alcohol Consumption through Crowdsensing and Social Media. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 2, Article 59 (June 2019), 30 pages.

[88] Evelien A. P. Poelen, Ron H. J. Scholte, Gonneke Willemsen, Dorret I. Boomsma, and Rutger C. M. E. Engels. 2007. Drinking by parents, siblings, and friends as predictors of regular alcohol use in adolescents and young adults: a longitudinal twin-family study. *Alcohol and Alcoholism* 42, 4 (05 2007), 362–369.

[89] Marnie E. Rice and Grant T. Harris. 2005. Comparing Effect Sizes in Follow-Up Studies: ROC Area, Cohen's d, and r. *Law and Human Behavior* 29, 5 (01 Oct 2005), 615–620.

[90] Irina Rish. 2001. An Empirical Study of the NaÃ¯ve Bayes Classifier. *IJCAI 2001 Work Empir Methods Artif Intell* 3 (01 2001).

[91] James A. Roberts and Meredith E. David. 2016. My life has become a major distraction from my cell phone: Partner phubbing and relationship satisfaction among romantic partners. *Computers in Human Behavior* 54 (2016), 134 – 141.

[92] D. Santani, T. Do, F. Labhart, S. Landolt, E. Kuntsche, and D. Gatica-Perez. 2018. DrinkSense: Characterizing Youth Drinking Behavior Using Smartphones. *IEEE Transactions on Mobile Computing* 17, 10 (2018), 2279–2292.

[93] Darshan Santani and Daniel Gatica-Perez. 2015. Loud and Trendy: Crowdsourcing Impressions of Social Ambiance in Popular Indoor Urban Places. In *Proceedings of the 23rd ACM International Conference on Multimedia* (Brisbane, Australia) *(MM '15)*. ACM, 211–220.

[94] Robert E Schapire. 2013. Explaining adaboost. In *Empirical inference*. Springer, 37–52.

[95] Laura Schelenz, Ivano Bison, Matteo Busso, Amalia de Götzen, Daniel Gatica-Perez, Fausto Giunchiglia, Lakmal Meegahapola, and Salvador Ruiz-Correa. 2021. The Theory, Practice, and Ethical Challenges of Designing a Diversity-Aware Platform for Social Relations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. ACM, 11.

[96] Sandra Servia-Rodríguez, Kiran K. Rachuri, Cecilia Mascolo, Peter J. Rentfrow, Neal Lathia, and Gillian M. Sandstrom. 2017. Mobile Sensing at the Service of Mental Well-Being: A Large-Scale Longitudinal Study. In *Proceedings of the 26th International Conference on World Wide Web* (Perth, Australia) *(WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 103–112.

[97] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. 2008. Ecological Momentary Assessment. *Annual Review of Clinical Psychology* 4, 1 (2008), 1–32.

[98] Carillon J. Skrzynski and Kasey G. Creswell. 2020. A systematic review and meta-analysis on the association between solitary drinking and alcohol problems in adults. *Addiction* (Dec. 2020), add.15355.

[99] Koen Smit, Martine Groefsema, Maartje Luijten, Rutger Engels, and Emmanuel Kuntsche. 2015. Drinking Motives Moderate the Effect of the Social Environment on Alcohol Use: An Event-Level Study Among Young Adults. *Journal of Studies on Alcohol and Drugs* 76, 6 (Nov. 2015), 971–980.

[100] Joshua Smyth and Arthur Stone. 2003. Ecological Momentary Assessment Research in Behavioral medicine. *Journal of Happiness Studies* 4 (02 2003), 35–52.

[101] Bundit Sornpaisarn, Kevin Shield, Jakob Manthey, Yuriko Limmade, Wah Yun Low, Vo Van Thang, and Jurgen Rehm. 2020. Alcohol consumption and attributable harm in middle-income South-East Asian countries: Epidemiology and policy options. *International Journal of Drug Policy* 83 (2020), 102856.

[102] Dimitris Spathis, Sandra Servia-Rodriguez, Katayoun Farrahi, Cecilia Mascolo, and Jason Rentfrow. 2019. Passive Mobile Sensing and Psychological Traits for Large Scale Mood Prediction. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Trento, Italy) *(PervasiveHealth'19)*. Association for Computing Machinery, New York, NY, USA, 272–281.

[103] D. Spruijt-Metz and W. Nilsen. 2014. Dynamic Models of Behavior for Just-in-Time Adaptive Interventions. *IEEE Pervasive Computing* 13, 3 (2014), 13–17.

[104] Abigail K. Stevely, John Holmes, Simon McNamara, and Petra S. Meier. 2020. Drinking contexts and their association with acute alcohol-related harm: A systematic review of event-level studies on adults' drinking occasions. *Drug and Alcohol Review* 39, 4 (May 2020), 309–320.

[105] Richard Taylor. 1990. Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography* 6, 1 (1990), 35–39.

[106] Johannes Thrul and Emmanuel Kuntsche. 2015. The impact of friends on young adults' drinking over the course of the evening-an event-level analysis: Impact of friends on young adults' drinking. *Addiction* 110, 4 (April 2015), 619–626.

[107] Johannes Thrul, Florian Labhart, and Emmanuel Kuntsche. 2017. Drinking with mixed-gender groups is associated with heavy weekend drinking among young adults. *Alcohol Research: Current Reviews* 112, 3 (2017), 432–439.

[108] Johannes Thrul, Sharon Lipperman-Kreda, and Joel W. Grube. 2018. Do Associations Between Drinking Event Characteristics and Underage Drinking Differ by Drinking Location? *Journal of Studies on Alcohol and Drugs* 79, 3 (May 2018), 417–422.

[109] Haske Van Der Vorst, Rutger C.M.E. Engels, Wim Meeus, and Maja Dekovic. 2006. The impact of alcohol-specific rules, parental norms about early drinking and parental alcohol use on adolescentsdrinking behavior. *Journal of Child Psychology and Psychiatry* 47, 12 (2006), 1299–1306.

[110] Haske Van Der Vorst, Rutger C. M. E. Engels, Wim Meeus, Maja Dekovic, and Jan Van Leeuwe. 2005. The role of alcohol-specific socialization in adolescents' drinking behaviour. *Addiction* 100, 10 (2005), 1464–1476.

[111] Victoria Vickerstaff, Rumana Omar, and Gareth Ambler. 2019. Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes. *BMC Medical Research Methodology* 19 (06 2019).

[112] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 43 (March 2018), 26 pages.

[113] Aaron White and Ralph Hingson. 2013. The Burden of Alcohol Use Excessive Alcohol Consumption and Related Consequences Among College Students. *Alcohol research : current reviews* 35 (01 2013), 201–18.

[114] W. Yao, Y. Liu, D. Zhou, Z. Pan, M. J. Till, J. Zhao, L. Zhu, L. Zhan, Q. Tang, and Y. Liu. 2018. Impact of GPS Signal Loss and Its Mitigation in Power System Synchronized Measurement Devices. *IEEE Transactions on Smart Grid* 9, 2 (2018), 1141–1149.

[115] Koji Yatani. 2016. *Effect Sizes and Power Analysis in HCI.* Springer International Publishing, Cham, 87–110.

[116] Chuang-wen You, Kuo-Cheng Wang, Ming-Chyi Huang, Yen-Chang Chen, Cheng-Lin Lin, Po-Shiun Ho, Hao-Chuan Wang, Polly Huang, and Hao-Hua Chu. 2015. *SoberDiary: A Phone-Based Support System for Assisting Recovery from Alcohol Dependence.* Association for Computing Machinery, New York, NY, USA, 3839–3848.