One More Bite? Inferring Food Consumption Level of College Students Using Smartphone Sensing and Self-Reports

LAKMAL MEEGAHAPOLA, Idiap Research Institute & EPFL, Switzerland
SALVADOR RUIZ-CORREA, IPICYT, Mexico
VIRIDIANA DEL CARMEN ROBLEDO-VALERO, IPICYT, Mexico
EMILIO ERNESTO HERNANDEZ-HUERFANO, IPICYT, Mexico
LEONARDO ALVAREZ-RIVERA, IPICYT, Mexico
RONALD CHENU-ABENTE, University of Trento, Italy
DANIEL GATICA-PEREZ, Idiap Research Institute & EPFL, Switzerland

While the characterization of food consumption level has been extensively studied in nutrition and psychology research, advancements in passive smartphone sensing have not been fully utilized to complement mobile food diaries in characterizing food consumption levels. In this study, a new dataset regarding the holistic food consumption behavior of 84 college students in Mexico was collected using a mobile application combining passive smartphone sensing and self-reports. We show that factors such as sociability and activity types and levels have an association to food consumption levels. Finally, we define and assess a novel ubicomp task, by using machine learning techniques to infer self-perceived food consumption level (eating as usual, overeating, undereating) with an accuracy of 87.81% in a 3-class classification task by using passive smartphone sensing and self-report data. Furthermore, we show that an accuracy of 83.49% can be achieved for the same classification task by using only smartphone sensing data and time of eating, which is an encouraging step towards building context-aware mobile food diaries and making food diary based apps less tedious for users.

CCS Concepts: • **Human-centered computing** → **Mobile computing**; **Smartphones**; **Mobile phones**; **Empirical studies in ubiquitous and mobile computing**; • **Applied computing** → *Consumer health*; *Health informatics*; *Sociology*.

Additional Key Words and Phrases: passive mobile sensing, health, well-being, eating behavior, food consumption, overeating

Authors' addresses: Lakmal Meegahapola, lmeegahapola@idiap.ch, Idiap Research Institute & EPFL, Switzerland; Salvador Ruiz-Correa, salvador.ruiz@ipicyt.edu.mx, IPICYT, Mexico; Viridiana del Carmen Robledo-Valero, viridiana.robledo@ipicyt.edu.mx, IPICYT, Mexico; Emilio Ernesto Hernandez-Huerfano, emilio.hernandez@ipicyt.edu.mx, IPICYT, Mexico; Leonardo Alvarez-Rivera, leonardo.alvarez@ipicyt.edu.mx, IPICYT, Mexico; Ronald Chenu-Abente, chenu@disi.unitn.it, University of Trento, Italy; Daniel Gatica-Perez, gatica@idiap.ch, Idiap Research Institute & EPFL, Switzerland.

1 INTRODUCTION

Many young adults show a tendency to adopt unhealthy eating practices during college years, when they undergo significant lifestyle changes such as leaving home, meeting new friends, starting a career, and developing relationships [82, 89]. Even though young adults are relatively healthy compared to other older populations, unhealthy eating habits at this age could lead to adverse health outcomes such as cardiovascular diseases, overweight conditions and obesity in the long term [11, 43, 82]. Due to these reasons, researchers in nutrition, behavioral science, and psychology are extensively studying causes and contexts of food consumption, specially among college students [53, 89, 136, 141]. Moreover, prior research in these domains have linked factors such as social context [52], eating location [41], availability and types of food [117], and psychological aspects [51] to food consumption behavior. With increasing smartphone coverage among young adults and the availability of a plethora of mobile health (mHealth) applications [78], smartphones have become a ubiquitous tool that can help young adults adhere to healthier food consumption practices [73].

The most common use case in mHealth apps is fitness tracking, where such apps keep track of daily activity levels *passively* by providing insights to users in terms of step count, activity levels, and activity types [6]. "Food and Nutrition" is another major category of mHealth applications [78], and many widely used commercial apps such as MyFitnessPal [3], Lose It! [2], Apple Health [7], and Samsung Health [8] allow users to keep their food intake as a mobile food diary, providing basic statistical insights so people can adhere to healthier eating patterns. However, in their current state, these applications do not yet make full use of smartphone sensing capabilities to provide additional insights (with the notable exception of the camera, which is used to photograph food), even though food consumption level is directly related to many aspects that can be sensed passively such as stress [70], mood [67, 90], activities [38, 124], and sociability [26, 80]. Even though attention has been given to visually identifying food types via mobile apps [100], characterizing meal and snack eating behavior using smartphone sensing and self-reports [24], and identifying eating and overeating events using wearable sensing [143], relatively less research has been conducted to leverage smartphone sensing capabilities to obtain data from people to analyze food consumption levels [75].

To our knowledge, while food intake recognition has been studied in ubicomp research [24], the specific overeating phenomenon has not been studied using passive smartphone sensing and self-report datasets. Using such a rich combination of data sources allows to analyze eating behavior of college students using knowledge from nutrition and mobile sensing research by associating food consumption levels to aspects such as mobile app usage, location, activity levels, sociability, and food types. This approach allows for comparisons with findings about self-perceived food consumption levels in prior nutrition and behavioral science research (which validates some of the observed trends), and also to provide novel insights regarding techniques to build mobile food journaling systems that leverage passive sensing to identify behaviors of college students associated to overeating, and to provide them with valuable insights and interventions regarding their food consumption. In this study, we address the following research questions.

RQ1: What behaviors and contextual patterns around food consumption levels can be observed by analyzing everyday eating episodes of a group of college students obtained via passive smartphone sensing and self-reports?
RQ2: Can self-percieved food consumption level be inferred using contextual data obtained through a mobile application?

By addressing the above research questions, this paper makes the following contributions.

(1) We conducted a new mobile data collection campaign in Mexico with 84 college students to capture their eating behavior including food consumption levels (eating as usual, overeating, undereating). Studies of populations in Latin America are not common in ubicomp research. During our study, we collected 3278 self-reports including 1911 meal reports and 1367 snack reports for two participant cohorts spanning 37 and 23 days, respectively. In

addition to self-reported data, we collected passive sensing data related to eating context, physical activity levels, and phone usage to characterize eating episodes in a holistic perspective.

(2) We defined and evaluated a novel ubicomp task by inferring eating more than usual ("overeating"), less than usual ("undereating"), and "as usual" episodes using passive smartphone sensing and self-report data, with an accuracy of 87.81% in a 3-class classification task. We also show that an inference accuracy of 83.61% can be achieved for the same 3-class classification task even if food category-related features are not used in model training. Moreover, we show that the same inference task can be done with an accuracy of 83.49% by only using passive smartphone sensing features and time of eating. These results show that in the set of participants, food consumption level was not only driven by food type or category as assumed in traditional food journals and calorie based calculations, but also by other contextual factors associated to eating episodes. Moreover, results from our study illustrates the potential of using passive smartphone sensing alone or together with mobile food journals towards building context-aware food-related mobile systems.

2 DEFINING FOOD CONSUMPTION LEVEL

In prior research, food consumption level has had both objective interpretations (nutrition science-based) [28, 113, 137] and subjective ones (nutrition and psychology-based) [40, 98, 123, 125, 131, 137], and there is no unique way to define it [49, 69]. The *objective* food consumption level attempts to capture the exact calorie consumption during eating episodes from a purely nutritional standpoint. In lab studies, calorie intake is pre-calculated before offering food to participants [28]. Under this objective interpretation, a person should eat only as much as is necessary to offset her/his caloric demands, and overeating occurs if the food intake exceeds this amount [49]. Many currently available mobile food diaries such as MyFitnessPal [3], Samsung Health [8], and other research studies [37] attempt to capture this attribute using self-reports by requesting the users to enter each food type and the amount they eat. Even though the target here is to capture the objective calorie intake, there is by design a subjective element because users self-report it, and it is known that people often fail to report volume/weight of a dish accurately [137]. However, even if caloric intake is correctly reported and calculated, defining food consumption level as overeating and undereating according to this approach is complicated according to Herman et al. [49, 125], because it depends on a plethora of factors: (a) individual factors such as metabolic rates, activity levels, age, gender, height, weight; and (b) measurement factors, i.e., the unit of calculation for overeating is usually caloric deficit per day (nutritionists often do it at the day, week, or meal/snack episode level). This implies that for the same person, eating the exact same amount of food on a more active day could be overeating on a slow day. Hence, the process gets more complex as factors add on, and it could get particularly difficult and inaccurate if overeating and undereating episodes are determined based on self-reports that reflect food types and volumes. In addition, a recent study by Jung et al. [58] emphasized how currently available mobile food logging systems can be troublesome to users because of the tedious manual data entry process, hence leading to low adoption rates.

Contrary to the objective view of food consumption level, nutrition researchers have also widely used *subjective* measures to capture food consumption levels of people by considering the psychology of food consumption [59, 69, 97, 98, 117, 122, 123, 125, 126], also known as self-perceived food consumption level. This view is primarily based on the idea that, when you ask people whether they overate or not, more often than not, the answer would be based on an eating episode level, and the self-perceived amount of food they have eaten [49]. This measure has often been used as a proxy to the actual amount of food people have eaten. Field et al. [40] showed that self-perceived food consumption level can be similar to real food consumption levels, and these self-reports are valid to determine bulimic episodes in adolescents. Moreover, Williamson et al. [137] examined the relation between self-reported caloric intake (similar to self-reports regarding caloric intake in MyFitnessPal and Samsung Health) and self-perceived overeating, concluding that there is a positive relationship between the two variables

for all four groups of people they considered: (1) suffering from bulimia nervosa, (2) compulsive binge eaters, (3) obese, and (4) not having any of the three previous conditions. Due to the above mentioned factors, many prior studies have used self-perceived food consumption level as a proxy to the objective food consumption level, although we are not aware of any study that establishes detailed guidelines of when self-reported subjective overeating and objective overeating coincide or not. Furthermore, prior work in nutrition research suggests that adverse behavioral and emotional effects of overeating arise not only after people eat an objective large amount of food, but even when people simply think that they have overeaten (self-perceived overeating) compared to their prior beliefs or current social context [69, 87, 88]. This is why many studies regarding psychology and eating behavior consider self-perceived food consumption level to be an important attribute, specially when considering eating as a holistic process to understand eating behavior [25, 56, 120].

Williamson et al. [137] captured overeating episodes by asking participants to report their perception on whether they overate or not (a binary choice). In a study by Ruddock and Hardman [98], self-perceived food consumption level was examined using a three-level coding system (eating more-than/less-than/as usual). Moreover, Vartanian et al. [125] used a five-point likert scale (1-5) in their study regarding food consumption levels where 1, 3, and 5 corresponded to "ate much less than I normally eat", "ate similar to the amount I normally eat", and "ate much more than I normally eat" respectively. By *normal* or *as usual*, what these studies meant is in comparison to their past behavior, and how they perceive societal norms regarding normal food intake. Following this literature, in this paper we define self-perceived food consumption level as *"eating more than (overeating condition), less than (undereating condition), or roughly the same (as usual condition) amount of food during an eating episode, in relation to the person's own estimated consumption, beliefs, and norms"*. Hence, from here onwards in this paper, we use the terms "food consumption level", "overeating", "eating as usual" or "undereating" to denote the self-perceived and self-reported attributes.

3 RELATED WORK

3.1 Internal and Contextual Factors Affecting Overeating

If we consider food consumption level, prior research has shown several factors that affect it such as psychological aspects, sociability, activity levels and types, and food types [49, 50, 59, 83, 111, 116, 135]. A recent review study discussed the complex relationship between psychology and food intake, describing that stress and mood affect food consumption behavior, and it can lead to both overeating or undereating depending on external or psychological stressors [111]. A recent article explained that stress can lead to lack of appetite in the short term (due to the high levels of hormone epinephrine secreted by adrenaline glands) that would lead to undereating [1]. On the flip side, long term exposure to stress can cause people to overeat with the initial expectation of overcoming stress and bad mood that is driven by hormonal activity (due to another hormone released from the adrenaline gland called Cortisol which increases appetite). Many other studies confirm these findings; some show that positive and calm affect, influenced by contextual factors (social gatherings, celebrations, partying) can lead to overeating in the short term [28, 31, 83], while other studies show that longer term exposure to stress can lead to eating disorders [33, 110]. Further explaining the relationship between human psychology and overeating, Vartanian et. al. [126] described that many people perceive the eating amount to be highly driven by internal factors such as hunger, feeling, and emotions rather than external factors such as social context. In another study, Bongers et. al. [27] showed that overeating is not only driven by mood and negative affect, but also external contextual factors. Studies also suggest that people who overeat do not self-perceive it at the moment, and usually feel sometime after consuming the food [49]. This could be due to the rapid nature of food consumption, and the time the body takes to generate the sensation of being full [15, 49, 101]. Considering this fact, we designed our mobile surveys retrospectively rather than in-situ [104] as described in Section 5. In this study, we focus on using a mobile food diary to identify short term food consumption levels with eating episode-level data.

3.2 Mobile Health Apps to Analyze Eating Behavior

Nutrition researchers have carried out mobile food diary based studies to understand compliance among young adults and children [10, 102], claiming that there is potential for young people to independently record their food intake using photos. Most of the mobile food diary based studies rely on users taking and labeling pictures, and manually entering calorie levels [75]. This is a tedious process that can disengage people due to boredom [58]. Hence, studies say that the focus of modern food diaries has been to reduce the effort by users to enter details, such that benefits of having such diaries outweigh the effort [12]. Following this, a direction pursued by mobile sensing and computer vision researchers is to identify food categories and estimated calorie intake with photos [29, 57, 66, 76]. However, most of these applications still lack in accuracy to be applicable in any unrestricted setting. As Min et. al. studied in a survey [75], these studies have focused on food recognition and use heterogeneous data sources such as social media and recipe datasets. In contrast, inferring food consumption levels using contextual passive sensing data and self-reports is not a widely attempted task in food computing research due to its challenging nature starting from obtaining a relevant data with accurate ground truth.

Even though mobile food diaries have been studied extensively as stand-alone applications, food intake is not necessarily a stand-alone activity, and is associated to factors such as psychology, activity levels, social context, and the general behavior of people as described in Section 3.1. Moreover, prior research highlights the importance of building diaries that are not only aware of the food intake, but also the contextual, psychological, and social modeling of the eating episode [32, 34, 79]. Hence, our focus is to find associations between perceived food consumption levels and features from passive sensing and self-reports, using the intuition that representing an eating episode by merely the food intake and type is not enough, but all the contextual and individual cues captured via passive sensing features should be considered.

3.3 Mobile Sensing to Analyze Eating Behavior

The use of mobile sensing in food-related studies can be categorized into two main types: (1) detecting and identifying eating events and time of eating using contextual sensing; and (2) after eating events are detected, identifying food types and characterizing eating events by identifying associations with user context, food consumption levels, and other attributes.

Detecting and identifying eating events has been primarily approached using wearable sensing and smartphone sensing. Rahman et. al. [91] used wearable sensing to predict about-to-eat moments with a recall of 77%. Bedri et. al. [20] studied a wearable system called EarBit to detect chewing moments, with 90.1% and 93% accuracies in lab-settings and outside-lab-settings, respectively. Thomaz et. al.[118] used a wrist wearable to detect eating episodes, reporting F1-scores of the range 70%-76% in two lab based experiments. In summary, these previous works attempt to identify the time of eating.

When the time of eating is detected, the next step is to characterize eating events [77]. Madan et. al. [71] used mobile phone sensing to understand food consumption of US university students for a period of nine months, and concluded that healthy food consumption patterns are related to the health behavior of other people that they associate with (sociability). Seto et. al.[107] used smartphone sensing to identify eating behavior using self-reports and passive sensing, and concluded that the eating environment affects the eating patterns. Biel et. al. [24] deployed a mobile sensing application to track eating behavior of 122 Swiss university students. They gathered over 4440 eating events, and performed an eating occasion inference (meal vs. snack) task with an accuracy of 84% using data such as location, time of the day, and time since the last food intake. Vu et. al. [129] laid out a vision for wearable food intake monitoring systems, and emphasized the importance of identifying overeating moments. Zhang et.al. [143] used a wrist wearable to detect overeating episodes based on the number of feeding gestures. They worked with the assumption that the higher the number of feeding gestures, the higher the calorie intake and objective food consumption level, and showed correlations in objective calorie consumption
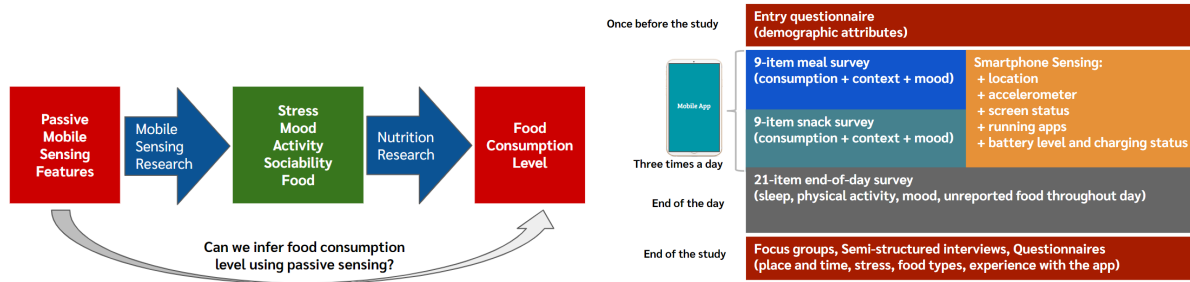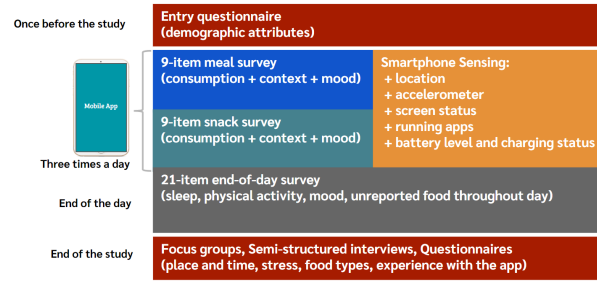
Fig. 1. Objective of the Study

Fig. 2. Block Diagram of Data Collection

and the number of feeding gestures. In our study, we consider eating events as holistic [24, 25, 120], and attempt to understand self-perceived food consumption level using passive smartphone sensing and self-reports. We show that food consumption level is related not only to the type of food, but also the sociability, mood, stress, and other behaviors, some of which can be captured or estimated using passive sensing features. As a summary, our study differs from previous work in five aspects: (1) we use smartphone sensing combined with mobile food diaries; (2) we collect a new mobile sensing and self-report dataset to study this phenomenon; (3) we consider eating as a holistic event instead of just considering the type and amount of food; (4) we show that the self-perceived food consumption level can be inferred using smartphone sensing and self-reports; and (5) we provide preliminary evidence of using just *passive smartphone sensing* features and time of eating, as informative features for the inference of food consumption level of college students.

4 STUDY DESIGN

4.1 Study Objective and Hypothesis

The primary objectives of this study are to investigate links between food consumption level and features derived using passive sensing, and to leverage such links to automatically infer food consumption level, as summarized in Figure 1. Prior literature has shown that passive sensing features can be used to infer psychological and contextual aspects such as stress [70, 106], mood [67], activity types [6, 8, 24], sociability [13, 23, 46, 47], and food types [24, 77, 107]. In addition, a plethora of prior nutrition and behavioral science studies have linked the above aspects to food consumption levels [49, 50, 74, 126] as discussed in Section 3. Knowing that smartphone sensing features have shown correlations to certain attributes, which have also been connected to food consumption levels as shown in Figure 1, our objective is to leverage these relationships studied in prior literature to use passive sensing for inference of food consumption levels. In other words, as passive sensing features have been linked to aspects such as stress, mood, activity, sociability, and food; and as these aspects have been linked in nutrition literature to food consumption levels like overeating; our hypothesis is that mobile sensing features could be used to infer self-percieved food consumption levels.

4.2 Mobile Application

We used a native android mobile application called i-Log to collect data from volunteers [142]. The app was developed at the University of Trento with Java, and data were initially stored in a SQLite database in the smartphone. Moreover, the system uses Google Firebase as a notification broker to send push notifications. When the phone is connected to a WiFi network and the phone has sufficient battery capacity, anonymized data were uploaded to Cassandra DB database in secure servers, hence freeing up the internal storage. The app has three main components: (a) push notification system to prompt users to complete questionnaires; (b) mobile surveys to record self-reports; and (c) passive smartphone sensing component to log sensor data.

4.3 Push Notification System

Given the nature of our study, we decided to use retrospective questionnaires to obtain self-reports from users. A notification strategy is important to get ground truth at the correct time, and we did not want users to report food intake in-situ, as the feeling of overeating is often experienced by users after a certain time from eating as described in Section 3.1. Hence, with the expectation of users retrospectively reporting their last food intake, we sent notifications to users 3 times per day, prompting them to fill in the food intake questionnaire. Push notifications were sent during the time slots 10.00 to 11.00 am, 3.00 to 4.00 pm, and 9.00 to 10.00 pm, that were chosen after considering the eating behavior of participants (note that lunch and dinner times in Mexico are later than those in the USA or most of Europe). Further, our expectation was not to collect data regarding all food intakes, but to collect as accurate data as possible regarding the 3 eating periods.

4.4 Food Intake Questionnaires

When users click a notification, the app opens and they are asked to indicate whether they wish to report a meal or a snack. The 9-item questionnaire was prepared with user experience in mind such that it can be filled within 1-2 minutes. Initially, we asked people about the number of eating episodes they had within the last 4 hours. Then we asked users to focus on their last food intake and answer 8 more questions that collected data including how long ago the eating episode had occurred (1-30 minutes ago, 30-60 minutes ago, etc.), food categories (meat, fish, bread, cereals, dairy, etc.), social context of eating (alone, with a date, with a groups of friends, etc.), semantic eating location (home, university restaurant, cafe, etc.), and concurrent activities (reading, socializing, watching TV, etc.). Then, two more questions were asked regarding their state of mind including general mood/affect at the time of eating (5-point scale from very negative to very positive) and stress level at the time of eating (5-point scale from very stressed to very calm). These questions were adapted from existing literature used to determine mood and stress level [67, 90]. Moreover, similar to Vartanian et al. [125], the app asked users to indicate their food consumption level as: (1) significantly less food than usual, (2) slightly less food than usual, (3) about the same as usual, (4) slightly more food than usual, and (5) significantly more food than usual. For our analysis, we define options 1 and 2 as "undereating", option 3 as "as usual", and options 4 and 5 as "overeating" in accordance with the definition we laid out in Section 2. Finally, even though participants answered end-of-day questionnaires, we did not include it in the analysis presented here.

4.5 Passive Smartphone Sensing

The average amount of time users take to provide self-reports is 6-8 minutes per day. To have a fine grained understanding about behavioral routines connected to eating, the app had passive smartphone sensing capabilities to collect data regarding users throughout the day. (a) **App Usage** - We collected data regarding app usage once every 5 seconds. Prior research has indicated that app usage behavior could be associated to user context [109, 138] and internal states [39, 144]; (b) **Accelerometer** - Prior research has shown that accelerometer data can be used to derive daily activity levels of people [65, 119]. Some studies have also associated these activity levels to well-being-related aspects [90, 106]. The app collected data from accelerometer with a frequency of 20Hz; (c) **Battery Events** - Whether the phone is plugged into a charging port, and to what kind of a port it is connected to (usb, alternative current, unknown) provide cues regarding the context of users including the availability of charging facilities nearby and also the general phone usage behavior [16]. Moreover we collected battery level throughout the day at every 1% change in battery level, and each time charging status changed; (d) **Screen Events** - The app logged details regarding the status of the screen including when the screen is turned on or off [9, 16]. These details can be used to estimate the amount of time the phone screen is on around eating events. This is an important attribute when examining eating behavior because prior work has discussed the adverse consequences of excessive phone usage around eating episodes [4, 112, 115]; (e) **Location** - This has

Table 1. Summary of Phases Including Workshop Participation and Number of Recruited Volunteers

Phase	# of Days	# of Participants	# of Recruited Volunteers	Population
I	37	32	29 (90.6%)	University 1
II	23	90	55 (61.1%)	University 2
Total	60	122	84 (68.9%)	-

been used in many studies to quantify daily movement [17, 140]. The app logged location of users with both GPS and phone network based on the availability of sensors once every 10 minutes.

5 DATA COLLECTION AND PRE-PROCESSING

5.1 Participant Recruitment, Study Approach, and Ethical Considerations

This study considered participants living in San Luis Potosi City in Mexico, which has 1.2 million inhabitants, and is home of diverse student populations. We considered college students from two of the main (private and public) universities in the city. People of this age group are tech savvy and the smartphone coverage among them is high [78]. In addition, young adults are open to change their behavior, specifically in relation to their eating and physical activity habits [94–96]. Considering all the above factors, a campaign was launched in June 2019 to announce a mobile sensing data challenge. Two workshops were held in August and October 2019, where the goals of the study were introduced to potential participants, describing how data collection would be carried out, and how data would be used for research. After a basic screening process, interested participants voluntarily filled out a consent form and entered the study filling an entry questionnaire (see Figure 2). Then, they installed the app on their smartphones. The average age of study participants was 23.4 years (SD: 3.51), mean BMI was 24.14 (SD: 4.68), and the cohort had 44% males and 56% females. As an appreciation for participating in the study, all of them were rewarded with a gift pack and a t-shirt. Participants had to meet certain criteria: (a) own an android smartphone and (b) not having eating disorders like bulimia or anorexia.

The initial phase of the study was run from September 3 to October 9, 2019. The second phase ran between November 19 and December 12, 2019. Hence we collected data for a total of 60 days from 84 university students. We made sure that the data collection was done during a time period with no major examinations or university level events that would alter the usual behavior of students. The data summary is described in Table 1. Phase I involved students who study subjects not relevant to nutrition from University 1 (Universidad Autónoma de San Luis Potosi), and Phase II was deployed in University 2 (Universidad Tangamanga) where participants were nutrition science students. During the data collection campaign, we sent a total of 7898 notifications to users, got 3895 responses (49.3% response rate), and 3278 of these responses corresponded to fully complete reports including 1911 meal reports and 1367 snack reports. After the mobile data collection phase, we conducted 10 semi-structured interviews, 1 eight-person focus group, and 64 short questionnaires with open-ended questions. The goal was to understand possible links between food consumption and the variables captured in the app, including places, times, stress, and food types. All material was originally in Spanish, then translated to English for this publication. A detailed description of the experimental procedures of the project was reviewed and approved by the local institutional authorities at IPICYT. The institutional review included approval of a Data Protection Impact Assessment document, a Privacy statement and Participant Consent forms, a Data Processing agreement among project partners, and a Declaration of Commitment agreement.

5.2 Pre-Processing the Dataset for Analysis

The goal of our analysis was to investigate eating episode level data. Hence, we chose each food intake self-report as a data point in our dataset. To integrate sensor and survey data, we followed an approach suggested in prior mobile sensing literature [24, 103, 106], where for each event of focus, in this case for each eating episode, passive

sensing data would be aggregated using a defined *time window*. We selected a time window of one hour which would mean that for each food intake event, we aggregate passive sensing data half an hour before and after the event starting time. We chose this time window considering prior research regarding characterizing eating events [24] and from a preliminary analysis regarding food consumption level (explained in Section 8). We started the procedure by finding the adjusted eating time because self-reports were done retrospectively. As mentioned in the previous section, we asked users "how long before they had the last meal". Using the answer for this question, we adjusted the timestamp of each food intake report to estimate the actual time of the eating episode. As an example, if the time of the self-report is 2pm, the answer for the question is 30-60 minutes ago (on average 30+60/2 = 45 minutes ago), the adjusted time of eating is estimated as 1.15pm (2pm - 45 minutes). Hence, using the one-hour time window, each eating event would be considered as a one hour eating episode. If the adjusted eating time is denoted by T, the time window would be the one hour from $T - 30$ minutes to $T + 30$ minutes. Next, we describe how each data modality was processed to associate it with eating episodes.

Accelerometer: following an approach similar to [24], for each 10-minute slot of the day, we generated features (aggregated sum of all values and sum of absolute values) using accelerometer value for axes x, y, and z. Then, depending on the adjusted time of an eating event (T), we considered three 10-minute bins before that eating episode (T-30 to T-20, T-20 to T-10, and T-10 to T), and three 10-minute bins after the start of the eating episode (T to T+10, T+10 to T+20, and T+20 to T+30). This way of pre-processing led to creating 18 features using accelerometer values. We use abbreviations to name the features generated using this methodology: (a) abs - calculated using absolute values of the accelerometer data; (b) bef - feature is calculated considering data before T, from T-30 to T; and (c) aft - feature is calculated considering data after T, from T to T+30.

Apps: we selected the ten most frequently used apps in the dataset. Then, during the hour associated with the eating episode, we determined whether each of those apps were used or not, hence resulting binary values for features in feature group *App*.

Location: using location traces, we calculated radius of gyration (a commonly used metric in mobile sensing [17, 140]) within the hour of consideration associated to the eating episode. Moreover, for each user, we generated stay regions throughout the whole day. Hence, using self-report labels (home, university, etc.), we generated labels for passively sensed stay regions of users, and we call that feature as *location* in our analysis. Moreover, for the location feature, we only used location degraded in precision for location privacy reasons (keeping only 4 decimal points).

Screen: using screen-on/off events in the dataset, we calculated the number of times the screen was turned on during the time slot, similar to prior literature [9, 16].

Battery: Similar to [16], we calculated the average battery level and also whether any charging events were detected during the time of eating episode. Battery and Screen events are used as proxies to smartphone usage behavior [9, 16].

6 DESCRIPTIVE DATA ANALYSIS (RQ1)

6.1 Food Types and Food Consumption Level

Figure 3 shows how different types of foods (meals in the top row, and snacks in the bottom row) were reported. When the number of meals is 1 (top-left), the figure shows three clear peaks at around 10am, 2pm, and 8pm that correspond to breakfast, lunch, and dinner times, respectively. The peak for lunch is visible when the number of meal events is 2, the dinner time peak is visible in the top-right figure. Moreover, when considering plots for snack intakes (bottom row), the figures for "how_many" values of 2 and more than 2 are sparse, while in the plot for "how_many" having 1, there are two peaks including one around noon, which is also a time period between breakfast meal at 10am and lunch meal peak at 2pm.

Table 2. Summary of the 56 features used in the analysis. Feature Group describes the type of features and 'Examples' are some feature names. The number in parenthesis next to categorical indicates the number of categories in categorical features.

| Feature Group - # of Features	Description	Type	Example Features
Activity (ACC) - 18	Features derived using accelerometer	numerical	acc_x, acc_xabs, acc_zbef
App Usage (APP) - 10	Features derived using app usage	categorical(2)	facebook, instagram, etc.
Food Category (FOOD) - 15	Features derived using category of food	categorical(2)	fats & oils, meat, eggs, fish
Battery Events (BAT) - 2	Charging status	categorical(2)	charging status
	Battery level of the phone	numerical	battery level
Location (LOC) - 2	Radius of gyration calculated using location	numerical(1)	radius of gyration
	Location of the person	categorical(12)	location
Temporal (TIME) - 2	Hour of the day and minute of the day	numerical	hours, minutes
Screen (SCR) - 1	Number of screen on/off events	numerical	screen events
Psychological (PSY) - 2	Derived using mood and stress self-reports	categorical(5)	mood, stress
Context (CON) - 3	Self-reported contextual details about eating	categorical	activity, social, location
Overeating (OVER) - 1	Self-reported food consumption level	categorical(5)	overeating
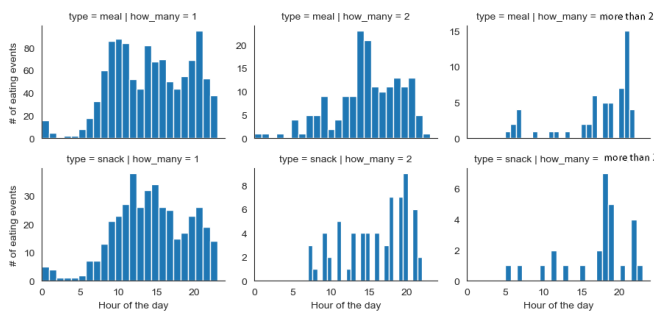
Fig. 3. Bar charts for food intake reports with hour of the day in the x-axis and number of eating events in the y-axis. "how_many" is the self-reported value from food intake questionnaire regarding how many eating episodes have occured in the past four hours.
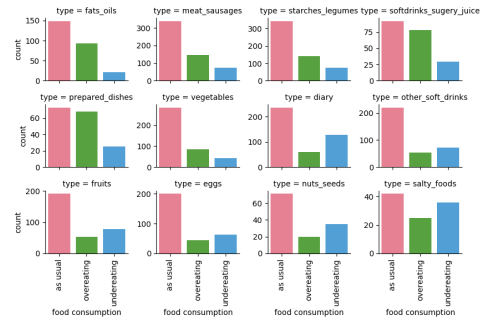
Fig. 4. Bar charts for ten different food categories. In each subplot, the x axis indicates the three classes regarding food consumption level, and the y-axis indicates the total numbers of reported cases.

In the dataset, the total number of meal episodes are distributed as 746 (55%) as usual , 319 (24%) undereating, and 285 (21%) overeating. Hence, the number of undereating and overeating episodes are comparable. For snacks, the numbers are distributed as 247 (49%) as usual, 190 (38%) undereating, and 63 (13%) overeating. Hence, the percentage of overeating episodes during snacks is relatively low compared to undereating, which is reasonable given that usual snacks are smaller in size. In addition, the percentage of undereating snack episodes is significantly higher (1.5 times the percentage) than for undereating meal episodes. Moreover, Figure 6 shows that a higher number of overeating events are found closer to lunch and dinner meal peaks.

As shown in Figure 4, when considering food categories and consumption levels, overeating was associated 93 times (35.4%) when "fats and oils" were present in the food as compared to undereating, which was reported only 21 times (7%). Food type "meat and sausages" was also associated to overeating almost twice the number of times as compared to undereating. For the food type "soft drinks or sugary juices", overeating episode reports (46%) were almost equal to the number of "as usual" episodes (39%). Moreover, for prepared dishes, the relation among "as usual", "overeating", and "undereating" was 3:3:1, which suggests that participants were highly likely to have reported "overeating" or "as usual" when eating prepared dishes, that are typically found in restaurants, bars, or cafes. These results match prior literature that links overeating to eating out in restaurants where food is often
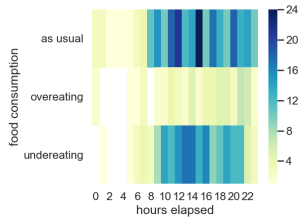
Fig. 5. Heatmap for hourly food consumption episodes considering **snack** intake. "hours elapsed" denotes the # of hours from the start of the day.
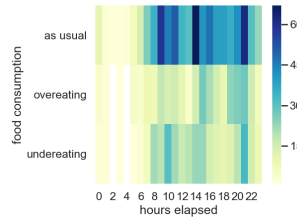
Fig. 6. Heatmap for hourly food consumption episodes considering **meal** intake. "hours elapsed" denotes the # of hours from the start of the day.
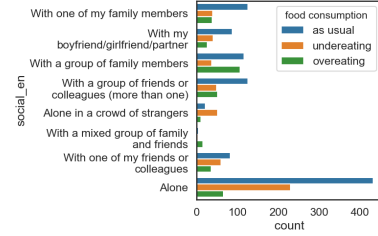
Fig. 7. Bar chart representing the number of eating events related to different social contexts and associated food consumption levels.

more palatable [72]. Moreover, our results are consistent with prior research that also suggests people reported overeating in the presence of fats and oils, sugary drinks, and meat [55, 132].

6.2 Sociability and Food Consumption Level

We considered 8 contexts of sociability in our analysis. 3 social contexts correspond to groups (with family and friends, family members, friends or colleagues), 3 contexts correspond to eating with another person (with one family member, boyfriend/girlfriend/partner, one friend/colleague), and the remaining 2 contexts (alone, alone in a crowd) correspond to eating alone. Figure 7 shows the distribution of food consumption episodes across different social contexts, and it indicates that all group eating episodes had a higher number of overeating cases (substantially for family; slightly for the other two) compared to undereating episodes. On the other hand, when eating alone, 230 reports indicated undereating, and in the social context "alone in a crowd of strangers", the percentage of undereating episodes (61.4%) is higher than the "as usual" percentage (25.3%). In summary, these results suggest that when participants were in groups, they had more overeating episodes as compared to undereating episodes. In eating behavior research, this phenomenon has been called *social facilitation* [49, 50]. These results match prior research that discusses eating as a holistic social event and how people tend to eat more food than they usually eat when they are in more social contexts [28, 49, 50, 74]. On the other hand, when alone, the number of undereating episodes rose significantly. Further, when participants were with another person (partner, family member, friend), more undereating episodes were reported compared to overeating in all cases. This could be because of the effect called *impression management*, that deals with lowering food intake to please one's companions [49, 50]. Moreover, apart from face-to-face sociability, widespread mobile adoption have made people sociable on the online sphere as well. Some studies have examined the differences and similarities of online and real world sociability of people in different countries [92]. In our dataset, apps such as facebook, whatsapp, and instagram showed different usage levels around the three food consumption levels (e.g. usage of social apps such as facebook and instagram have shown a higher frequency closer to overeating reports as opposed to undereating reports). These observations conform with prior literature that have indicated associations between social media use and food consumption [86, 121, 130].

6.3 Mood, Stress, and Food Consumption Level

According to Figure 9, when the mood was negative or very negative, people tended to report undereating. This result is puzzling, although some literature in nutrition and psychology has suggested that loss of appetite has been indeed reported when people go through short-term episodes of negative moods [1, 105]. The number of reported overeating episodes increases when going from neutral to positive mood. Our analysis shows that out of all the overeating reports with a very positive emotion (124 reports), the majority (47.58%) has reported home
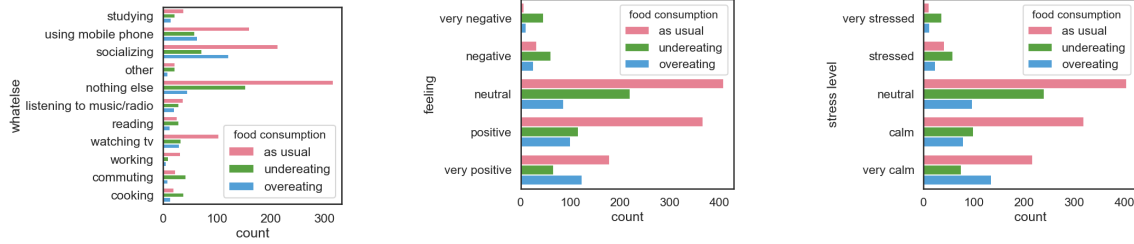
Fig. 8. Bar chart representing reported concurrent activities done while eating.

Fig. 9. Mood while eating and count of reported instances.

Fig. 10. Stress while eating and count of reported instances.

as the eating location. As a whole, the number of reports with negative or slightly negative mood (183 reports) is low as compared to positive and very positive mood (952 reports). As shown in Figure 10, a similar overall trend can be seen for the reported stress level too. This similarity is reflected in the high and positive Pearson correlation coefficient (0.77, p < 10^{-5}) between mood and stress [22].

6.4 Concurrent Activities and Food Consumption Level

Figure 8 shows the types of concurrent activities done while eating as reported by the participants. "Nothing Else" category was reported 514 times, and it had the lowest percentage of overeating reports (8.2%) out of all the different concurrent activities. As literature suggests, this might be because when participants are fully focused on the eating event, they are mostly aware of their food consumption level [95]. Moreover, results suggest that participants reported more overeating episodes as compared to undereating, when they socialized while eating, which is consistent with prior research that suggested social eating as an instance when people overeat [28, 83]. Furthermore, using mobile phone while eating also corresponds to more overeating episodes compared to undereating. Watching TV had about the same numbers for overeating and undereating. This result is consistent with a recent study that explored the relationship between screen time (mobile or TV) and unhealthy diets [112]. Moreover, another recent study suggested that distracted eating can lead to overeating [95], which indeed happens when eating while using the phone or watching TV. This is also why studying screen time and app usage are important when analyzing overeating episodes with a holistic approach. Moreover, activities such as commuting, cooking, and reading had higher percentages of undereating as compared to both overeating and as usual. Recent reports regarding commuting and overeating suggest that long commutes could significantly influence overeating behavior [5, 54].

7 STATISTICAL ANALYSIS (RQ1)

7.1 Pearson and Point-Biserial Correlation for Self-Report Features

We conducted Pearson [114] (for mood, stress, and social context) and Point-biserial [21] correlation analysis (for food types and food categories, because they are binary values) to understand feature relationships between self-report features and food consumption level, and some results are summarized in Table 3. Mood, Stress, and Sociability showed correlation values above +0.29 indicating close to moderate relationships. This is once again conforming to prior nutrition studies such as [49, 50, 74, 126], that linked psychological aspects and sociability to food consumption level. On the other hand, we observed a correlation value of +0.22355, that is a weak positive linear relationship for food type and food consumption level. In our analysis, for feature food type, a feature value of 1 was used for meals (usually bigger in size, and had higher number of overeating episodes compared to undereating) and 0 was used for snacks (usually smaller in size, and had far fewer overeating events, higher number of undereating episodes). Hence, it means that meal intakes have correlated positively with food

Table 3. Pearson and Point-Biserial correlation analysis for some self-report features and food consumption level.

Features	Value	Features	Value
food consumption level	1 (+)	food meat sausages	.31827 (+)
mood	.35927 (+)	food fats oils	.32637 (+)
stress	.29719 (+)	food starches lugumes	.30412 (+)
food type	.22355 (+)	food softdrinks sugery juice	.25710 (+)
social context	.29165 (+)	food prepared dishes	.24144 (+)

Table 4. Comparative statistics of 20 features across classes "overeating" and "undereating" (OverUnder) and "overeating" and "as usual" (OverUsual): t-statistic, p-value (* if p>0.01, ** if p>0.1, *** if p>0.5), and Cohen's-d with 95% confidence intervals. Features are sorted based on the decreasing order of t-statistics.

OverUnder				OverUsual			
Feature	Group	t-statistic	Cohen's-d [95% CI]	Feature	Group	t-statistic	Cohen's-d [95% CI]
social context	CON	12.35076	0.85354, [0.71099, 0.99609]	social context	CON	5.76663	0.53366, [0.34955, 0.71777]
fats or oils	FOOD	10.02969	0.65577, [0.51561, 0.79594]	microsoft launcher	APP	5.01027	0.39534, [0.21242, 0.57827]
meat or sausages	FOOD	9.45595	0.63970, [0.49970, 0.77970]	mood	PSY	4.64106	0.41359, [0.23053, 0.59665]
mood	PSY	8.76104	0.61134, [0.47163, 0.75106]	prepared dishes	FOOD	4.36667	0.36789, [0.18515, 0.55062]
starches or legumes	FOOD	8.71631	0.59054, [0.45103, 0.73006]	fats or oils	FOOD	4.35349	0.38456, [0.20172, 0.56741]
stress	PSY	8.30656	0.57668, [0.43470, 0.71606]	softdrinks or sugery juices	FOOD	3.96672	0.33864, [0.15610, 0.52119]
softdrinks or sugery juices	FOOD	7.42293	0.49218, [0.35353, 0.63083]	stress	PSY	3.94471	0.35331, [0.17067, 0.53594]
prepared dishes	FOOD	6.87340	0.45509, [0.31673, 0.59345]	location	CON	3.84508	0.33838, [0.15584, 0.52092]
vegetables	FOOD	6.62784	0.44431, [0.30603, 0.58259]	minutes elapsed	TIME	3.57003	0.33825, [0.15572, 0.52079]
concurrent activity	CON	5.58480	0.39475, [0.25680, 0.53269]	hours elapsed	TIME	3.52190	0.33338, [0.15087, 0.51589]
microsoft launcher	APP	5.58480	0.34018, [0.20257, 0.47780]	screen on count	SCR	2.52200*	0.22040, [0.03847, 0.40232]
acc z abs bef	ACC	4.41168	0.30777, [0.17033, 0.44521]	meat or sausages	FOOD	2.09344*	0.19240, [0.01058, 0.37422]
acc z abs	ACC	4.37907	0.30347, [0.16605, 0.44089]	facebook	APP	1.99841*	0.17961, [-0.00216, 0.36139]
facebook	APP	3.20637	0.21770, [0.08065, 0.35476]	acc z bef	ACC	1.84707*	0.17348, [-0.00828, 0.35523]
acc z abs aft	ACC	3.14326	0.21902, [0.08196, 0.35607]	salty foods	FOOD	1.67583*	0.14556, [-0.03612, 0.32723]
location	CON	2.86509	0.19687, [0.05988, 0.33385]	starches or legumes	FOOD	1.63866**	0.15024, [-0.03144, 0.33193]
minutes elapsed	TIME	1.94971*	0.13457, [-0.00224, 0.27139]	acc z	ACC	1.61030**	0.15039, [-0.03130, 0.33208]
whatsapp	APP	1.88304*	0.12936, [-0.00744, 0.26616]	acc z abs bef	ACC	1.51921**	0.14886, [-0.03283, 0.33054]
hours elapsed	TIME	1.81924*	0.25618, [0.11897, 0.39338]	acc y	ACC	1.40647**	0.12889, [-0.05274, 0.31052]
instagram	APP	1.31946**	0.08977, [-0.04696, 0.22650]	acc y aft	ACC	1.39948**	0.12972, [-0.05191, 0.31135]

consumption level of the participants. Food categories such as meat and sausages, food starches and legumes, fats and oils have values above +0.3 indicating weak positive linear relationships with food consumption levels. Similar findings were previously reported in some nutrition studies [55, 132] where they reported overeating in the presence of categories of foods such as fats and oils and meat and sausages.

7.2 Statistical Analysis of Dataset Features

Table 4 (left) shows statistics such as t-statistic [60], p-value [44], and Cohen's-d (effect size) with 95% confidence interval (CI) [63] for all the features in the dataset for two groups: **Over**eating and **Under**ating (**OverUnder**). Hence, the objective is to identify features that would allow to discriminate between "overeating" (n = 348) and "undereating" (n = 509) episodes. Table 4 (right) shows the same set of statistics considering the two groups **Over**eating and As **Usual** (**OverUsual**). The objective here is to identify discriminating features for two groups "overeating" (n = 348) and "as usual" (n = 993) episodes, which are closer in range to each other. In both tables, the features are ordered by the descending order of t-statistics. Moreover, since p-values are not sufficiently informative [64, 139], we additionally calculated Cohen's-d [93] to help understand the statistical significance of the features. To interpret Cohen's-d, we used a commonly used rule-of-thumb: small effect size = 0.2, medium effect size = 0.5 and large effect size = 0.8. Moreover, we calculated 95% confidence interval for Cohen's-d. The higher the Cohen's-d value, the stronger the possibility of discriminating the two groups using the considered feature. If the confidence interval does not include zero, confidence interval of the effect size depicts the reliability

of the effect size. If the confidence interval is on positive side, the feature could be promising for discrimination among the groups. Moreover, the narrower the confidence interval, the higher the reliability of the calculated effect size.

In Table 4 (left), social context is the feature with the highest t-statistic, and it has a large effect size too. Hence, it shows that social context during an eating episode is the most discriminating feature with regard to OverUnder. In Table 4 (right), social is also the feature with the highest t-statistic, which suggests that it is the most discriminating feature for OverUsual. Moreover, the t-statistic and effect size for this feature have dropped from 12.35 and 0.85 to 5.76 and 0.53 from OverUnder to OverUsual. This suggests that the social feature has a higher discriminating capability for OverUnder as opposed to OverUsual. Another feature from the feature group CON that appeared in both tables is location. However, location had a higher t-statistic in OverUsual as compared to OverUnder. The effect size too increased from 0.19 (small effect) to 0.33 between OverUnder and OverUsual, hence making location a better feature to discriminate "overeating" and "as usual" episodes. The feature about concurrent activity has a t-statistic of 5.58 and a medium effect size for OverUnder. However, it is not among the top 20 discriminating features for OverUsual.

When considering food types, "fats or oils" was the feature from the FOOD feature group with the highest t-statistic and effect size for OverUnder. However a place down in OverUsual, whereas prepared dishes had a t-statistic of 4.36 and less than medium affect size. Even features such as "meat or sausages", "softdrinks or sugery juices", and "starches and legumes" switched positions in the two tables showing that features that affect the discrimination of the three food consumption level classes can differ in relevance. However, "starches or legumes" and "salty foods" contain zero in the corresponding confidence interval for OverUsual, suggesting that they might not be reliable features to distinguish between these two classes. Another feature "vegetables" appeared in OverUnder with a medium effect size, but disappeared from the Table for OverUsual. In both tables, mood and stress showed differences across groups. The effect sizes were 0.61 (resp. 0.57) in OverUnder, and 0.41 (resp. 0.35) in OverUsual. This result again tallies with previous findings in nutrition and behavioral psychology research as discussed in Section 3.

If we consider app usage, "facebook" and "microsoft launcher" are discriminant of the two classes in OverUnder with high t-statistics and small, medium effect sizes, respectively. Even though "whatsapp" and "instagram" appear in OverUnder, negative values in the confidence interval suggest that those are not reliable features for discrimination. When considering activity level derived from accelerometer, OverUnder contains three important features from z-axis, that have slightly high t-statistics and small-to-medium effect sizes. However, in OverUsual, these features have gotten smaller effects sizes with negative values in confidence intervals. Interestingly, minutes elapsed and hours elapsed features from the group TIME have moved from low t-statistics and unreliably low effect sizes in OverUnder to higher t-statistics and at least small effect sizes in OverUsual. This suggests that the time of the day at which food intake has been reported would be important in deciding between "overeating" and "as usual" episodes, but not between "overeating" and "undereating" episodes. Moreover, screen events feature showed medium effect size and high t-statistics in OverUsual. It also supports findings from prior literature that screen time is an important variable in determining "overeating" and unhealthy dieting habits [112]. Further, features related to battery events (BAT) and radius of gyration (derived using location data) did not seem to be significant enough features to discriminate food consumption levels of consideration in either table.

8 FOOD CONSUMPTION LEVEL INFERENCE (RQ2)

8.1 Three-class Food Consumption Level Inference

The three-class inference task uses different subsets of features in the training set, and calculates classification accuracy, precision, and recall. The target classes were *overeating*, *undereating*, and *as usual*. We used python with scikitlearn and keras in this phase, and we conducted experiments using several model types (in the decreasing

order of accuracies for inference task G5): random forest, naive bayes, gradient boosting, neural networks, XGboost, AdaBoost, and support vector classifiers. However, considering space limitations and aspects such as interpretability and model personalization, we present inference results for two models as follows:

(a) Random forest classifier (RF) with ntree values between 50 - 500: we got the highest accuracy values for inference tasks using RFs. More importantly, RFs models output the feature importance values used in inference, hence enabling us to understand and interpret the results.

(b) Neural network (NN) with 3-4 layers with relu activations, dropout for regularization, and binary cross entropy loss function (number of layers, number of nodes in intermediate layer/s changed depending on the feature group and the input dimensions; hyper-parameter tuning was done for each feature group with the goal of obtaining the best model for the task): NNs provided with reasonable accuracies, and allow transfer learning without much sophistication [81]. Hence, we present results for NNs here because we use them for transfer learning to personalize models in Section 8.3.

We followed leave k-participants out strategy (k = 15) for all the experiments when preparing the dataset, where training and testing splits did not have data from the same user, hence avoiding this possible source of bias in the evaluation procedure. Moreover, when preparing the dataset, we made sure that the classes are balanced by up-sampling the minority classes and down-sampling majority class to get a balanced dataset of 2400 records. The baseline for experiments is 33.3% since the classes were balanced in all inference tasks. We conducted experiments for individual feature groups and meaningful feature group combinations:

Self Reports without FOOD (G1): This corresponds to self-reports that would not be available in a traditional mobile food diary, such as reports regarding eating context (sociability, concurrent activities), psychological state (mood, stress) together with the time of eating. The objective is to show that even without capturing the types and amounts of food, it is still possible to infer food consumption level. An envisaged application scenario of this inference is where the mobile health app simply captures these few self-reports instead of all the food consumption details, hence making the user experience better in terms of lower burden of manual data input.

Self Reports (G2): This corresponds to all the self-reported features including food types and categories. In addition to features in G1, this also captures the types of food consumed by participants. This inference would reaffirm the relationship shown in Figure 1 with regard to the association between food consumption level and aspects such as mood, stress, sociability, activities, and food.

Passive Smartphone Sensing with TIME (G3): This feature group combination contained a single self-reported variable (time of eating), and a set of passively sensed features without any user input, such as accelerometer, app usage, location, screen usage, and battery events. Importantly, this group reflects an envisaged mobile health application usage scenario where participants only report that they ate, hence capturing the time of eating, without typing all the details about the food types and amounts, sociability, concurrent activities, hence making it less tedious. In addition, prior work has examined the use of passive mobile sensing features to determine the time of eating [20, 91, 118], which is a separate open research question. Hence, this feature group combination denotes a envisaged use case that depends on *near-passive* sensing.

All Feature Groups without FOOD (G4): This contained all feature groups except for FOOD. Hence, this set of feature would require the same set of user involvement/effort as in G1. As we are following a holistic approach regarding food consumption, the goal here is to evaluate whether only knowing about the contextual factors and sensing data without knowing the food types and amounts could characterize the food consumption levels.

All Features (G5): This used all the available features to demonstrate the potential of a future mobile food diary that is driven by passive smartphone sensing in addition to traditional self-reports. This feature group captures food related details, contextual and socio-psychological attributes, and passive sensing data.

All Features w/ PCA (G6): Out of the 3 commonly used multi-modal fusion techniques: (a) data, (b) feature, and (c) decision) [35, 108], all inference tasks in this study except for G6 used feature-level fusion, that feeds a

Table 5. Three-class food consumption inference (overeating, undereating, as usual) accuracy, precision, and recall obtained with a random forest classifier (RF) and a neural network (NN) using different feature group combinations.

Feature Group Name (# of Features)	RF			NN		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Baseline	33.33%	-	-	33.33%	-	-
SCR (1)	40.26%	42.65%	40.97%	31.35%	33.46%	15.91%
APP (10)	47.52%	51.29%	46.74%	45.21%	48.68%	44.05%
PSY (2)	50.82%	52.36%	50.59%	44.88%	51.81%	44.54%
LOC (2)	56.43%	56.31%	56.68%	32.34%	29.63%	33.94%
CON (3)	62.37%	62.27%	62.37%	45.87%	55.08%	45.70%
BAT (2)	63.03%	61.58%	62.69%	50.03%	40.36%	51.64%
FOOD (15)	65.34%	66.11%	65.38%	60.72%	60.83%	60.78%
TIME (2)	67.65%	67.14%	67.01%	56.30%	56.04%	56.14%
ACC (18)	76.89%	83.89%	69.76%	47.52%	47.01%	57.89%
G1: CON + PSY + TIME (7)	81.19%	81.45%	80.91%	62.19%	62.39%	62.11%
G2: CON + PSY + TIME + FOOD (22)	82.50%	82.61%	83.56%	66.68%	67.29%	67.25%
G3: ACC + APP + LOC + SCR + BAT + TIME (35)	83.49%	83.19%	82.84%	73.26%	73.33%	72.20%
G4: ACC + APP + LOC + SCR + BAT + CON + PSY + TIME (40)	83.61%	83.99%	83.57%	79.20%	79.26%	79.17%
G5: All Features (55)	87.81%	87.97%	88.37%	82.17%	82.19%	82.95%
G6: All Features w/ PCA (principle components=4) (4)	83.53%	83.46%	83.54%	76.67%	76.94%	76.53%

processed feature map into a classifier. In G6, we examined feature extraction and dimensionality reduction with principal component analysis (PCA), that fuses the features before feeding them into the classifier.

Results of the experiments (Table 5) show that RFs perform better than NNs across all feature groups and evaluation measures. Hence, in this section, only the results from RFs are discussed. Table 5 shows that individual feature groups such as ACC (accelerometer data), TIME (time of the day), FOOD (types of food consumed – similar to a traditional food diary), BAT (battery events), and CON (context when consuming food) have accuracies above 60%, while the highest accuracy of 76.89% corresponded to ACC feature group. This suggests that activity levels derived from the smartphone can be used to distinguish food consumption levels, to some degree. This is justifiable because prior work in nutrition and behavioral sciences have discussed the relation between food consumption levels and activity levels [49, 135]. G2 provides an idea regarding accuracies that can be obtained with currently available mobile food diaries that fully rely on participant self-reports. Accuracy, precision, and recall had values in the range 81% to 83% suggesting that self-reports are able to classify the three food consumption levels.

G3 shows that it has an even higher accuracy when compared to G1 or G2. Given that most prior research in nutrition has relied on self-reports regarding food categories and volumes to analyze food consumption behavior, this result shows that it is worth looking into passive smartphone sensing for cues regarding food consumption levels, given that there are many aspects such as app use, screen time, sociability that relate to the way people consume food in modern societies. Moreover, this result aligns with the hypothesis we presented in Section 4.1 and Figure 1 with regard to the possibility of inferring food consumption levels primarily using mobile sensing features.

G4, with an accuracy of 83.61% shows the benefit of considering eating as a holistic event as compared to just food categories and volumes. This accuracy, which is above the accuracy obtained with G2, again shows the importance of the holistic view of eating. Finally, by combining all the feature groups in G5, the model achieved an accuracy of 87.81%, precision of 87.97%, with a recall of 88.37%, all of which are encouraging. In addition, for G6, we got the best results with 4 principle components, an accuracy of 83.53%, which is higher than G1 and G2, although it is still lower than the corresponding feature level fusion (G5) that used the same set of features. These result suggests that passive smartphone sensing can be of great value when incorporated to mobile food
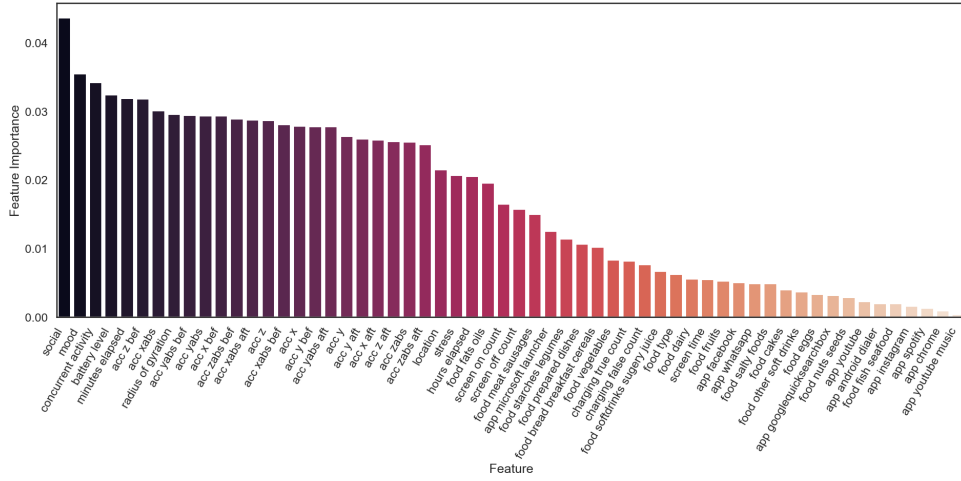
Fig. 11. Feature Importance generated using the RF for G5 Inference

diaries that are currently based only on self-reports. Further, these results also highlight the potential that passive smartphone sensing has as part of mobile applications for food monitoring with less intrusive usage scenarios.

In Figure 11, we plot the importance of features in classifying food consumption levels using RFs for G5 inference, hence containing all the features. Results indicate that social, mood, and concurrent activity are the self-reported features that were among the top 10 features. In addition, passive smartphone sensing features such as battery level, acc z bef, acc x abs, radius of gyration, acc y abs bef, and acc y abs were among the top 10 features in terms of feature importance. This indicates that activity levels and phone battery level around food intake events are useful as part of a rich model. It is also important to note that even though food category related features had high t-statistics and effect sizes in Table 4, Figure 11 shows that those features did not have comparatively high feature importance values in the RF.

8.2 Statistical Analysis Regarding Incorrect Inferences

Given that we have shown the capability of inferring food consumption levels using passive smartphone sensing and self-reports, in this section, we seek to understand which of the features have affected the most in getting incorrect inferences. To do that, we prepared three datasets using "All Features" (G5) with leave k-participants out strategy (k = 15) for Training and Testing splits: (1) Training - the dataset used for training models; (2) Testing-Correct - the dataset for which models provided correct inferences in testing; (3) Testing-Incorrect - the dataset for which models provided incorrect inferences in testing. We repeated the food consumption level inferences for five iterations, and for each iteration, we obtained different versions of the above mentioned three-fold datasets that are of different sizes. For each iteration, we calculated the t-statistic, p-value, and Cohen's-d for the following dataset pairs, and averaged the results: (**T1**) "Testing-Correct vs. Testing-Incorrect" - the objective is to identify features that are different across these two datasets using their feature distributions. If the difference is higher (captured with t-statistic and Cohen's-d), that feature could hint at the reasons for the model to incorrectly classify each of the dataset instance; (**T2**) "Training vs. Testing-Correct" - we expect the features from training set and testing-correct to be in the same distribution as testing instances were correctly classified using the RF model, hence potentially leading to low Cohen's-d values; and (**T3**) "Training vs. Testing-Incorrect" - The classifier inferred these instances incorrectly, possibly due to deviations between the testing and training feature distributions. Hence, the expectation is for these differences to have higher Cohen's-d values. The results are

Table 6. Comparative statistics of features across groups: t-statistic (T), p-value (P), and Cohen's-d (C) with 95% confidence intervals (* if confidence interval include zero). Features are sorted based on the decreasing order of t-statistics.

		T1: Testing-Correct vs. Testing-Incorrect			T2: Training vs. Testing-Correct			T3: Training vs. Testing-Incorrect		
	Feature	T	P	C	Feature	T	P	C	Feature	T	P	C
Passive Sensing	charging false	2.1036	0.0362	0.5762	screen on count	2.09259	0.0366	0.1621	google quick search	2.5010	0.0126	0.4885
	charging true	2.0236	0.0439	0.3574	acc x abs aft	2.07056	0.0386	0.1359*	charging true	1.3348	0.1807	0.2582*
	youtube	1.5775	0.1157	0.3192	app spotify	1.6195	0.1056	0.1271*	location	1.3265	0.1849	0.1947*
	acc x abs	1.5193	0.1297	0.3174	app z abs bef	1.4221	0.1552	0.1037*	acc x aft	1.3035	0.1928	0.1887*
	app x bef	1.4541	0.1470	0.3088	acc z	1.3731	0.1700	0.0988*	screen on count	1.1436	0.2530	0.1726*
Food Reports	food cakes	1.9318	0.0543	0.3431	food cakes	1.8611	0.0629	0.1265*	food fish seafood	1.3427	0.1796	0.2341*
	food nuts seeds	1.2331	0.2184	0.1726*	food prepared dish	1.7289	0.0840	0.1174*	food cakes	1.3338	0.1825	0.2223*
	food fish seafood	1.1611	0.2465	0.2052*	food other soft drinks	1.4794	0.1392	0.1081*	food nuts seeds	1.1879	0.2351	0.1530*
	food fats oils	0.8363	0.4036	0.1332*	food meat sausages	1.0385	0.2994	0.0733*	food fats oils	0.6915	0.4894	0.1034*
	food dairy	0.7685	0.4427	0.1147*	food dairy	0.8286	0.4074	0.0596*	food other soft drinks	0.6844	0.4938	0.1018*
Self-Reports	social context	0.5656	0.5717	0.0402*	concurrent activity	1.2816	0.2003	0.1932*	mood	0.7131	0.4762	0.1133*
	stress	0.5189	0.6039	0.0372*	mood	0.7942	0.4273	0.1317*	social context	1.8927	0.0594	0.3238*
	mood	0.4242	0.6714	0.0301*	stress	0.6212	0.5343	0.0894*	stress	0.2641	0.7918	0.0410*

presented in Figure 12, that shows a plot of Cohen's-d value of all features, in the descending order, for the three dataset combinations. Table 6 summarizes features with highest t-statistics and Cohen's-d values for the three dataset combinations from different feature categories such as passive sensing, food reports, and other self-reports.

T2 showed the lowest set of Cohen's-d values as given in Figure 12. This is intuitively expected as the features in Training and Testing-Correct might have similar distributions. In T3, the Cohen's-d values are much higher than T2. This could be because the distribution of training features and testing-incorrect features are slightly different. In Table 6, the passive sensing features that had the highest t-statistics and Cohen's-d for this dataset combination are google quick search (APP), charging true (BAT), location (LOC), acc x aft (ACC), and screen on count (SCR). Out of these features, the usage of google quick search app had a Cohen's-d closer to medium effect size. Moreover, all these features except for screen events were not included in the top 20 features in Table 4. Furthermore, in T1, the highest Cohen's-d values were obtained for charging false (BAT), charging true (BAT), youtube (APP), acc x abs (ACC), and acc x bef (ACC), none of which are in Table 4, suggesting that these features (when considered alone) were less informative in discerning food consumption levels, and similarly, these features had larger distribution differences between correct and incorrect inferences. This also suggests that classification models might not have learned to properly classify food consumption levels using these features. Moreover, for all the iterations, incorrect inferences were prevalent for all the users in all testing sets. Hence, if incorrect inferences were calculated based on per user level in the test set, the percentages ranged from 9% - 29% for feature combination G5. This provides an indication that the incorrect inferences were not because of users who were having highly deviated behaviors, and also that the model has generalized reasonably well to the data from all participants.

8.3 Model Personalization

In Table 5, we presented inference results without any personalization by only considering eating event-level features, and by not considering any temporal features. In this section, we attempt to personalize inferences by considering aspects regarding inference models (model-wise) and by including additional temporal features to the dataset (data-wise), with the expectation of increasing the accuracy of food consumption level inference. Finally, we show results for NN models because they allow transfer learning transfer learning in a relatively straightforward manner.
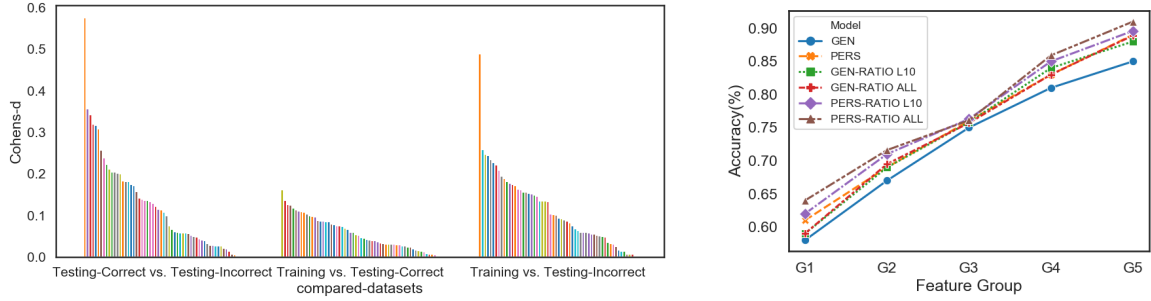
Fig. 12. Cohen's-d value for features across different dataset groups. Note that the set of features shown in Table 6 are a subset of features shown here, categorized based on feature types.

Fig. 13. Accuracies for different feature group combinations among GEN and PERS.

8.3.1 Transfer Learning for Model Personalization. In a prior study, LiKamwa et al. [67] have shown that general models are sufficient when users do not provide enough data for personalized model training. However, when the training dataset size increases, e.g., when collecting data for more days, the personalized models outperformed general models for inferences. In our dataset, we have limited data from individuals, hence restricting the possibility of training separate personal models for all users. Hence, we trained models on general data, and then follow a transfer learning approach to retrain the model with individual data. In our dataset, 11 users reported more than 50 food intake reports. For this task, we used the training procedures and feature group combinations presented in Section 8.1. The two model types we used were:

(1) GEN: the model is trained on data from all users. However, as we described in the training procedure in Section 8.1, we do not use data from the same individual in training and testing splits. Hence, this model has been trained on a dataset in which the specific user's data is not included. When testing, we used 30% split of user's data.

(2) PERS: We retrained the already trained model using a split (70%) from user's personal data, and test it on the unused 30% split of personal data.

We carried out experiments for the 11 users having more than 50 food reports, and included the average results in Figure 13 by repeating experiments for five iterations. Results show that for all feature group combinations, PERS outperformed the GEN with accuracies of the range 0.95% (G3) – 4.08% (G5).

8.3.2 Using Temporal Features for Model Personalization. The inferences done in previous sections did not consider food intake events to be linked to each other, as we considered all the food intake events to be mutually exclusive from even past food intake events of the same person. Adding temporal features could possibly increase accuracy. Hence, we introduce *four* new features: (1) previous food consumption level (prev_level) - capturing the previous food consumption level reported by the individual; and (2) three ratios to the dataset that capture the temporal evolution of reported food consumption levels of an individual participant. The ratios are: (2a) overeating ratio (oer): the total number of overeating episodes reported in the past, divided by the total number of episodes up to that time; (2b) undereating ratio (uer): the total number of undereating episodes reported in the past, divided by the total number of episodes by that time; (2c) as-usual ratio (aur): the total number of as-usual episodes reported in the past, divided by the total number of as-usual episodes by that time. Hence, these three ratios capture the participants' reporting history.

Moreover, when calculating the ratios, we examined two temporal windows: (1) Last ten food intakes (L10), which captures the recent food consumption levels; and (2) All previous food intakes (ALL), which captures the overall history of an individual. We assessed the efficacy of these features using the following two tasks:

Table 7. Summary of the Qualitative Analysis.

Topic	Insight & Quotes
Stress	Stress can play a key role with regard to eating behaviors. Stress has caused undereating or overeating in participants.
	"[...] we really eat according to how we feel at that moment." (Female 30 yo)
	"[..] when I'm stressed I tend to skip meals and end up eating later more food than I should including grease food." (Female 24 yo)
Food types	Food types and portions eaten by volunteers varied significantly. Variability depends on place, social context, stress level, and activity load. Food eaten at the school cafeteria and places around campus often have lower quality and variety than food prepared at home. Junk and fatty food are widely available to students, which sometimes fosters unhealthy eating behaviors.
	"Nutritious things are generally more expensive, and take longer to cook. Therefore, we choose to eat fast food." (Female 22 yo)
	"The time we dedicate to preparing our food, we want to live at a very fast pace, and we prioritize spending time on social networks. If we spent some of that time preparing our food, we would consume a much smaller amount of fast food." (Male 24 yo)
Places and Times	Participants mainly eat at home and the school cafeteria. Some prepare their food at home and take it to school, and others go to restaurants or cheap kitchens near school. Eating schedules are mainly determined by study/work load. However, some students only eat when they find the opportunity to do so, or have free time (they have no defined schedules). In general, reported times for breakfast, lunch and dinner are between 6:00 a.m. to 12:00 p.m., 1:00 p.m. to 6:00 p.m. and 9:00 p.m. 00 to 00:00, resp.
	"I eat when I have free time" (Female, 21 yo)
	"I mainly eat at the university cafeteria or places around. I sometimes eat at home" (Male 22 yo)
	"I prepare breakfast and dinner, and my mother cooks lunch for me" (Female 20 yo)

(1) GEN-RATIO: we re-trained all the models for feature groups G1 to G5 including the four features, for both temporal windows, using the approach presented in Section 8.1, and tested using the data from aforementioned 11 participants.

(2) PERS-RATIO: we re-trained all the personal models for feature groups G1-G5 for 11 participants, for both temporal windows, using the approach presented in Section 8.3.1 by including the four additional features that capture their temporal evolution.

As shown in Figure 13, results suggest that using these four features increased the accuracies for all four model-feature combinations (GEN-RATIO L10, GEN-RATIO ALL, PERS-RATIO L10, PERS-RATIO ALL) with accuracies in the range 0.1% to 3.9%. In addition, for the feature prev_level, for groups overeating and undereating, we obtained a Cohen's-d of 0.30, which suggests that this feature is informative of overeating vs. undereating episodes. Moreover, the three ratios had Cohen's-d values between 0.12 and 0.16, that are less than small effect size. In summary, the model performance can be increased by using features that consider the longitudinal aspects of data.

9 DISCUSSION

9.1 Feedback from Participants

We start this section by summarizing some of the feedback obtained from the study participants (semi-structured interviews, focus group, and questionnaire) in Table 7.

9.2 Passive Smartphone Sensing for Characterizing Food Consumption Levels

Results presented in prior sections confirm that our hypothesis regarding food consumption levels and passive sensing features (Section 4.1) is valid. Most smartphones are capable of both continuous sensing (feature groups APP and SCR) and interaction sensing (feature groups such as ACC, BAT, TIME, LOC) for behavioral modeling. Obviously, these sensing modalities do not directly capture the food type or internal aspects that nutrition and behavioral science researchers have linked to food consumption levels in the past. However, in Section 8.1, we showed the potential of using passive smartphone sensing to infer food consumption level. What these modalities sense is not the food type or psychological aspects, but the physical activity levels and smartphone usage behavior. Given that physical activity levels have been linked to stress and mood in prior mobile sensing literature [90, 106],

we believe these passive sensing modalities contain contextual information that could directly relate to food consumption behavior, and that is the reason why inferring food consumption levels with an accuracy of 83.49% using passive smartphone sensing and time related features was feasible. This is one of the first studies in this direction, and there are plenty of opportunities to explore *eating as a holistic event* as we did here. Given that computer vision researchers are focused on identifying food intake types and levels using images of the food portion (as discussed in Section 3), we could expect food consumption self-reports to get automated in future mobile food diaries. However, considering that eating is a holistic event driven by many factors, further research to determine food consumption behavior could enable advanced mobile food diaries that do no solely depend on user input to generate recommendations and interventions.

9.3 Further Informative Features Regarding Food Consumption Levels.

We acknowledge that the features we generated from passive modalities are simple and easier to interpret when associated with eating episodes. However, there is ample opportunity to build upon these findings, and develop novel features that could discriminate food consumption levels with higher accuracies. For example, in this study, all the accelerometer features are single dimensional and we did not use linear acceleration or 2D resultant acceleration features due to some limitations in the data collection process (not having data from gyroscope to match accelerometer traces so that gravity biases could be removed [19, 48]). Moreover, when considering app usage behavior, the features we used only determined whether a particular app was used or not during the time window of the eating episode. However, advanced research could be done to determine usage times of each app during eating episodes, hence obtaining a comprehensive understanding regarding app usage behavior related to eating. Moreover, using a low-power API such as Google Activity Recognition API to detect activity types could generate new features that might be beneficial in characterizing eating events.

9.4 Accounting for Diversity

The eating behavior of people in different countries vary depending on a plethora of factors such the culture, type of food they consume, concurrent activities while eating, and how they perceive events such as eating [36, 62, 127]. Hence, it is important to clarify that the results from the deployment of our application in Mexico are exploratory and not representative of the food consumption behavior of people from other regions. Moreover, if other aspects apart from food are considered, there are already known differences with regard to factors such as sociability [84, 99], activity levels [18, 45], and phone usage [68, 85] in different countries, and these aspects could get reflected in smartphone sensing datasets. For example, a study regarding sociability of university students in Mexico and USA showed that Mexican students perceived themselves to be less sociable compared to how Americans perceived themselves, although in reality Mexicans were more sociable [92]. Moreover, results show that Americans socialize more in private environments or by interacting through social media. On the other hand, Mexican students preferred to be more social in-person with people who are around them [92]. A similar trend was visible in our results as depicted in Section 7.2, where social was the feature with highest t-statistic and effect size among all features used in the analysis. Hence, we could expect differences in passive sensing data obtained from students in these two countries. It is fundamental to consider human diversity in smartphone sensing studies, and we believe more studies should integrate these diversity aspects in the future. Hence, future research could look into deploying mobile food diaries with sensing capabilities in diverse user groups based on ethnicity, culture, and geographic regions. In our opinion, the goal of such studies is to build models for mobile food diaries that generalize well enough to cater and adapt to diverse user populations. Even though our study is focused on college students of a Latin American country, we believe that this is a first step in this direction.

9.5 Limitations and Future Directions

Similar to other prior mobile sensing in-the-wild deployments [16, 24, 103, 106], we used feature level fusion techniques in the inference task except for G6. However, other decision level fusion techniques could be explored in future work for similar mobile sensing datasets. Further, this study used data from college students in Mexico. Even though the features we used did not contain country-specific details, studies similar to this could be carried out among diverse user groups within Mexico and also in other countries to determine the validity and applicability of our approach to user groups of diverse age and occupations. Moreover, we prepared the base dataset (in Section 5) assuming that individual eating episodes are mutually exclusive (similar to prior studies regarding mobile app deployments in-the-wild [24, 67, 90, 103, 106]). By making this assumption, we created a general dataset for the worst-case scenario where no temporal relationship between eating events for a given individual is assumed, hence avoiding any personalization effect due to considering longitudinal features from the same individual. However, in Section 8.3.2, we show how some basic longitudinal features can be used to personalize inferences, hence leading to higher accuracies for food consumption level inferences. We believe it is best to examine these longitudinal aspects in depth, in a separate study that explicitly focuses on these aspects.

Similar to other in the real-life app deployment, it becomes challenging to verify all the self-reports given by participants [24, 67, 103, 106]. For example, in our study, even if photos of food types were captured, it is difficult to verify other self-reports such as sociability, concurrent activities, mood, and stress. While we used different techniques to monitor the completion of food intake reports in line with prior work [24, 103], we believe that these are challenges that are common, specially during real-life field experiments where subjective reporting is used to collect ground truth labels. As future work, an interesting question relates to possible gender differences in mobile food diaries. Prior literature has suggested gender differences in food choice [134], activity levels [14], app usage [61], and different factors that we discovered in this study to be affecting food consumption level. Moreover, the number of participants in our study was 84, which is reasonable given the recent food studies in smartphone sensing by Seto et. al. [107] (12) and Biel et. al. [24] (122). We collected data from participants for a longer time span compared to Biel al. (10 days) and Seto et al. (6 days). Further, many other recent smartphone sensing studies with app deployments had participants of the range 50 - 100 [16, 30, 42, 133]. Based on this, we believe that the total number of events we captured through the study is enough to reasonably answer the research questions we addressed here. Future work could investigate larger sample sizes, which could allow to study issues of generality and personalization in more depth. Finally, even though our analysis contained multiple comparisons, we did not use Bonferroni correction for p-values [128]. Hence, these values should be interpreted with caution.

10 CONCLUSION

In this study, we examined the eating behavior of 84 college students using a dataset consisting of smartphone sensing data and self-reports. We demonstrated that behavioral mobile sensing and self-report features of college students around eating events can be used to infer self-reported food consumption levels. We achieved an accuracy of 87.81% for the 3-class food consumption level inference task, showing the potential of passive sensing together with self-reports in future mobile food diaries. Further, we show that an accuracy of 83.49% can be achieved for the same classification task by only using passive sensing data and time of eating events, in an envisaged simplified mobile food diary usage scenario. This performance suggests that smartphone sensing opens the possibility to detect self-perceived food consumption levels in future mobile health applications, which would need to be validated further with larger and more diverse user populations.

ACKNOWLEDGMENTS

REFERENCES

[1] 2018. Why stress causes people to overeat. Retrieved April 29, 2020 from https://www.health.harvard.edu/staying-healthy/why-stress-causes-people-to-overeat

[2] 2019. Lose It! Retrieved April 28, 2020 from https://www.loseit.com/

[3] 2019. MyFitnessPal. Retrieved April 28, 2020 from https://www.myfitnesspal.com/

[4] 2019. Put down that mobile phone - it's making you fat! Scientists warn using a smartphone at mealtimes can lead to an expanding waistline. Retrieved October 27, 2020 from https://www.dailymail.co.uk/news/article-6738357/Scientists-warn-using-smartphone-mealtimes-lead-expanding-waistline.html

[5] 2020. Commuter health. Retrieved May 1, 2020 from https://www.apple.com/ios/health/

[6] 2020. Google Fit - Coaching you to a healthier and more active life. Retrieved February 12, 2020 from https://www.google.com/fit/

[7] 2020. A more personal Health app. For a more informed you. Retrieved February 12, 2020 from https://www.apple.com/ios/health/

[8] 2020. Samsung Health App. Retrieved February 12, 2020 from https://www.samsung.com/us/support/owners/app/samsung-health

[9] Saeed Abdullah, Elizabeth L. Murnane, Mark Matthews, Matthew Kay, Julie A. Kientz, Geri Gay, and Tanzeem Choudhury. 2016. Cognitive Rhythms: Unobtrusive and Continuous Sensing of Alertness Using a Mobile Phone. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 178–189.

[10] Tanisha Aflague, Carol Boushey, Rachael Guerrero, Ziad Ahmad, Deborah Kerr, and Edward Delp. 2015. Feasibility and Use of the Mobile Food Record for Capturing Eating Occasions among Children Ages 3–10 Years in Guam. *Nutrients* 7, 6 (Jun 2015), 4403–4415.

[11] Nabil Alshurafa, Jayalakshmi Jain, Rawan Alharbi, Gleb Iakovlev, Bonnie Spring, and Angela Pfammatter. 2018. Is More Always Better? Discovering Incentivized MHealth Intervention Engagement Related to Health Behavior Trends. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 153 (Dec. 2018), 26 pages.

[12] A. H. Andrew, G. Borriello, and J. Fogarty. 2013. Simplifying mobile phone food diaries. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. 260–263.

[13] Martin Atzmueller and Katy Hilgenberg. 2013. Towards Capturing Social Interactions with SDCF: An Extensible Framework for Mobile Sensing and Ubiquitous Data Collection. In *Proceedings of the 4th International Workshop on Modeling Social Media (MSM '13)*. Association for Computing Machinery, New York, NY, USA, Article 6, 4 pages.

[14] Mario Azevedo, Cora Araújo, Felipe Reichert, Fernando Siqueira, Marcelo Silva, and Pedro Hallal. 2007. Gender differences in leisure-time physical activity. *International journal of public health* 52 (02 2007), 8–15.

[15] Nathan H. Azrin, Michael J. Kellen, Jeannie Brooks, Chris Ehle, and Veronica Vinas. 2008. Relationship Between Rate of Eating and Degree of Satiation. *Child & Family Behavior Therapy* 30, 4 (2008), 355–364.

[16] 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.

[17] Gianni Barlacchi, Christos Perentis, Abhinav Mehrotra, Mirco Musolesi, and Bruno Lepri. 2017. Are you getting sick? Predicting influenza-like symptoms using human mobility behaviors. *EPJ Data Science* 6 (12 2017), 27.

[18] Adrian Bauman, Guansheng Ma, Frances Cuevas, Zainal Omar, Temo Waqanivalu, Philayrath Phongsavan, Kieren Keke, and Anjana Bhushan. 2011. Cross-national comparisons of socioeconomic differences in the prevalence of leisure-time and occupational physical activity, and active commuting in six Asia-Pacific countries. *Journal of Epidemiology & Community Health* 65, 1 (2011), 35–43.

[19] Akram Bayat, Marc Pomplun, and Duc A. Tran. 2014. A Study on Human Activity Recognition Using Accelerometer Data from Smartphones. *Procedia Computer Science* 34 (2014), 450 – 457. The 9th International Conference on Future Networks and Communications (FNC'14)/The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC'14)/Affiliated Workshops.

[20] 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 Article 37 (Sept. 2017), 20 pages.

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

[22] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. *Pearson Correlation Coefficient*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1–4.

[23] Ethan M. Berke, Tanzeem Choudhury, Shahid Ali, and Mashfiqui Rabbi. 2011. Objective Measurement of Sociability and Activity: Mobile Sensing in the Community. *The Annals of Family Medicine* 9, 4 (2011), 344–350.

[24] 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.

[25] Carole A. Bisogni, Laura Winter Falk, Elizabeth Madore, Christine E. Blake, Margaret Jastran, Jeffery Sobal, and Carol M. Devine. 2007. Dimensions of everyday eating and drinking episodes. *Appetite* 48, 2 (2007), 218 – 231.

[26] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex (Sandy) Pentland. 2014. Daily Stress Recognition from Mobile Phone Data, Weather Conditions and Individual Traits. In *Proceedings of the 22Nd ACM International Conference on Multimedia (MM '14)*. ACM, New York, NY, USA, 477–486.

[27] Peggy Bongers, Anastacia de Graaff, and Anita Jansen. 2016. 'Emotional' does not even start to cover it: Generalization of overeating in emotional eaters. *Appetite* 96 (2016), 611 – 616.

[28] Peggy Bongers, Anita Jansen, Remco Havermans, Anne Roefs, and Chantal Nederkoorn. 2013. Happy eating. The underestimated role of overeating in a positive mood. *Appetite* 67 (2013), 74 – 80.

[29] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101 – Mining Discriminative Components with Random Forests. In *Computer Vision – ECCV 2014*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 446–461.

[30] Mehdi Boukhechba, Alexander R. Daros, Karl Fua, Philip I. Chow, Bethany A. Teachman, and Laura E. Barnes. 2018. DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones. *Smart Health* 9-10 (2018), 192 – 203. CHASE 2018 Special Issue.

[31] Hilde Bruch. 1964. Psychological Aspects of Overeating And Obesity. *Psychosomatics* 5, 5 (1964), 269 – 274.

[32] E. A. Carroll, M. Czerwinski, A. Roseway, A. Kapoor, P. Johns, K. Rowan, and M. C. Schraefel. 2013. Food and Mood: Just-in-Time Support for Emotional Eating. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 252–257.

[33] J Chua, Stephen Touyz, and AJ Hill. 2004. Negative mood-induced overeating in obese binge eaters: An experimental study. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity* 28 (05 2004), 606–10.

[34] Tegan Cruwys, Kirsten E. Bevelander, and Roel C.J. Hermans. 2015. Social modeling of eating: A review of when and why social influence affects food intake and choice. *Appetite* 86 (2015), 3 – 18. Social Influences on Eating.

[35] B. V. Dasarathy. 1997. Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proc. IEEE* 85, 1 (1997), 24–38.

[36] John B.F. de Wit, F. Marijn Stok, Derek J. Smolenski, Denise D.T. de Ridder, Emely de Vet, Tania Gaspar, Fiona Johnson, Lyliya Nureeva, and Aleksandra Luszczynska. 2015. Food Culture in the Home Environment: Family Meal Practices and Values Can Support Healthy Eating and Self-Regulation in Young People in Four European Countries. *Applied Psychology: Health and Well-Being* 7, 1 (2015), 22–40.

[37] Tamara Denning, Adrienne Andrew, Rohit Chaudhri, Carl Hartung, Jonathan Lester, Gaetano Borriello, and Glen Duncan. 2009. BALANCE: Towards a Usable Pervasive Wellness Application with Accurate Activity Inference. In *Proceedings of the 10th Workshop on Mobile Computing Systems and Applications (HotMobile '09)*. Association for Computing Machinery, New York, NY, USA, Article 5, 6 pages.

[38] W. Frank Epling and W. David Pierce. 1988. Activity-based anorexia: A biobehavioral perspective. *International Journal of Eating Disorders* 7, 4 (1988), 475–485.

[39] R. Ferdous, V. Osmani, and O. Mayora. 2015. Smartphone app usage as a predictor of perceived stress levels at workplace. In *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. 225–228.

[40] Alison E. Field, C. Barr Taylor, Angela Celio, and Graham A. Colditz. 2004. Comparison of self-report to interview assessment of bulimic behaviors among preadolescent and adolescent girls and boys. *International Journal of Eating Disorders* 35, 1 (2004), 86–92.

[41] Purificacian Garcia-Segovia, Robert J. Harrington, and Han-Seok Seo. 2015. Influences of table setting and eating location on food acceptance and intake. *Food Quality and Preference* 39 (2015), 1 – 7.

[42] Jiaqi Gong, Yu Huang, Philip I. Chow, Karl Fua, Matthew S. Gerber, Bethany A. Teachman, and Laura E. Barnes. 2019. Understanding behavioral dynamics of social anxiety among college students through smartphone sensors. *Information Fusion* 49 (2019), 57 – 68.

[43] Holly C. Gooding, Carly Milliren, Christina M. Shay, Tracy K. Richmond, Alison E. Field, and Matthew W. Gillman. 2016. Achieving Cardiovascular Health in Young Adulthoodâ€"Which Adolescent Factors Matter? *Journal of Adolescent Health* 58, 1 (2016), 119 – 121.

[44] 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.

[45] Regina Guthold, Melanie J. Cowan, Christine S. Autenrieth, Laura Kann, and Leanne M. Riley. 2010. Physical Activity and Sedentary Behavior Among Schoolchildren: AÂ 34-Country Comparison. *The Journal of Pediatrics* 157, 1 (2010), 43 – 49.e1.

[46] Gabriella Harari, Sandrine Mueller, Clemens Stachl, Rui Wang, Weichen Wang, Markus Buehner, Peter Rentfrow, Andrew Campbell, and Samuel Gosling. 2019. Sensing Sociability: Individual Differences in Young Adults' Conversation, Calling, Texting, and App Use Behaviors in Daily Life. *Journal of Personality and Social Psychology* 119 (05 2019).

[47] Gabriella M. Harari, Samuel D. Gosling, Rui Wang, Fanglin Chen, Zhenyu Chen, and Andrew T. Campbell. 2017. Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. *Computers in Human Behavior* 67 (2017), 129 – 138.

[48] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. 2014. Gravity and Linear Acceleration Estimation on Mobile Devices. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MOBIQUITOUS '14)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 50–59.

[49] C.P. Herman, J. Polivy, and T. Leone. 2005. 6 - The psychology of overeating. In *Food, Diet and Obesity*, David J. Mela (Ed.). Woodhead Publishing, 115 – 136.

[50] C. P. Herman and J. Polivy. 2003. Dieting as an exercise in behavioral economics. *Time and decision: Economic and psychological perspectives on intertemporal choice* (2003).

[51] Marion M. Hetherington. 2007. Cues to overeat: psychological factors influencing overconsumption. *Proceedings of the Nutrition Society* 66, 1 (2007), 113–123.

[52] Suzanne Higgs and Jason Thomas. 2016. Social influences on eating. *Current Opinion in Behavioral Sciences* 9 (2016), 1 – 6. Diet, behavior and brain function.

[53] Ya-Li Huang, Won O. Song, Rachel A. Schemmel, and Sharon M. Hoerr. 1994. What do college students eat? Food selection and meal pattern. *Nutrition Research* 14, 8 (1994), 1143 – 1153.

[54] Lee HM Hwang JH. 2007. A Study on Lifestyles, Dietary Habits, Nutrition Knowledge and Dietary behaviors of Male University Students According to Residence Type. *Korean Journal of Community Nutrition* (2007), 381–395.

[55] J.R. Ifland, H.G. Preuss, M.T. Marcus, K.M. Rourke, W.C. Taylor, K. Burau, W.S. Jacobs, W. Kadish, and G. Manso. 2009. Refined food addiction: A classic substance use disorder. *Medical Hypotheses* 72, 5 (2009), 518 – 526.

[56] Margaret M. Jastran, Carole A. Bisogni, Jeffery Sobal, Christine Blake, and Carol M. Devine. 2009. Eating routines. Embedded, value based, modifiable, and reflective. *Appetite* 52, 1 (2009), 127 – 136.

[57] S. Jiang, W. Min, L. Liu, and Z. Luo. 2020. Multi-Scale Multi-View Deep Feature Aggregation for Food Recognition. *IEEE Transactions on Image Processing* 29 (2020), 265–276.

[58] 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.

[59] Gregory S. Keenan, Louise Childs, Peter J. Rogers, Marion M. Hetherington, and Jeffrey M. Brunstrom. 2018. The portion size effect: Women demonstrate an awareness of eating more than intended when served larger than normal portions. *Appetite* 126 (2018), 54 – 60.

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

[61] Saskia Klenk, Doreen Reifegerste, and Rebecca Renatus. 2017. Gender differences in gratifications from fitness app use and implications for health interventions. *Mobile Media & Communication* 5, 2 (2017), 178–193.

[62] D Kromhout, A Keys, C Aravanis, R Buzina, F Fidanza, S Giampaoli, A Jansen, A Menotti, S Nedeljkovic, and M Pekkarinen. 1989. Food consumption patterns in the 1960s in seven countries. *The American Journal of Clinical Nutrition* 49, 5 (05 1989), 889–894.

[63] 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.

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

[65] I-Min Lee and Eric J Shiroma. 2014. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *British Journal of Sports Medicine* 48, 3 (2014), 197–201.

[66] Yanchao Liang and Jianhua Li. 2017. Computer vision-based food calorie estimation: dataset, method, and experiment. *ArXiv* abs/1705.07632 (2017).

[67] 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 (MobiSys '13)*. ACM, New York, NY, USA, 389–402.

[68] Soo Ling Lim, Peter Bentley, Natalie Kanakam, Fuyuki Ishikawa, and Shinichi Honiden. 2014. Investigating Country Differences in Mobile App User Behavior and Challenges for Software Engineering. *IEEE Transactions on Software Engineering* 41 (09 2014).

[69] Emma V. Long, Lenny R. Vartanian, C. Peter Herman, and Janet Polivy. 2020. What does it mean to overeat? *Eating Behaviors* 37 (2020), 101390.

[70] 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 (UbiComp '12)*. ACM, New York, NY, USA, 351–360.

[71] Anmol Madan, Sai T. Moturu, David Lazer, and Alex (Sandy) Pentland. 2010. Social Sensing: Obesity, Unhealthy Eating and Exercise in Face-to-Face Networks. In *Wireless Health 2010 (WH '10)*. Association for Computing Machinery, New York, NY, USA, 104–110.

[72] Megan A. McCrory, Paul J. Fuss, Nicholas P. Hays, Angela G. Vinken, Andrew S. Greenberg, and Susan B. Roberts. 1999. Overeating in America: Association between Restaurant Food Consumption and Body Fatness in Healthy Adult Men and Women Ages 19 to 80.

Obesity Research 7, 6 (1999), 564–571.

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

[74] David Mela. 2005. *Food, Diet, and Obesity.* Elsevier. https://www.elsevier.com/books/food-diet-and-obesity/mela/978-1-85573-958-1

[75] Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2019. A Survey on Food Computing. *ACM Comput. Surv.* 52, 5, Article Article 92 (Sept. 2019), 36 pages.

[76] Weiqing Min, Linhu Liu, Zhengdong Luo, and Shuqiang Jiang. 2019. Ingredient-Guided Cascaded Multi-Attention Network for Food Recognition. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*. Association for Computing Machinery, New York, NY, USA, 1331–1339.

[77] Mark Mirtchouk, Dana McGuire, Andrea Deierlein, and Samantha Kleinberg. 2019. Automated Estimation of Food Type from Body-worn Audio and Motion Sensors in Free-Living Environments. *Proceedings of machine learning research* 106 (08 2019), 641–662.

[78] Mobius. 2019. 11 surprising mobile health statistics. Retrieved April 28, 2020 from https://www.mobius.md/blog/2019/03/11-mobile-health-statistics/

[79] 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.

[80] Sai T Moturu, Inas Khayal, Nadav Aharony, Wei Pan, and Alex (Sandy) Pentland. 2011. Using Social Sensing to Understand the Links Between Sleep, Mood, and Sociability. In *Proceedings of IEEE International Conference on Social Computing*.

[81] S. J. Pan and Q. Yang. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 22, 10 (2010), 1345–1359.

[82] Guy S. Parcel, Lana D. Muraskin, and Carolina M. Endert. 1988. Community education: Study group report. *Journal of Adolescent Health Care* 9, 6, Supplement (1988), S41 – S45.

[83] K.A. Patel and D.G. Schlundt. 2001. Impact of moods and social context on eating behavior. *Appetite* 36, 2 (2001), 111 – 118.

[84] Pekka and Antti Kouvo. 2007. LINKED OR DIVIDED BY THE WEB?: Internet use and sociability in four European countries. *Information, Communication & Society* 10, 2 (2007), 219–241.

[85] Iryna Pentina, Lixuan Zhang, Hatem Bata, and Ying Chen. 2016. Exploring privacy paradox in information-sensitive mobile app adoption: A cross-cultural comparison. *Computers in Human Behavior* 65 (2016), 409 – 419.

[86] Thanh-Trung Phan and Daniel Gatica-Perez. 2017. Healthy fondue dinner: Analysis and Inference of Food and Drink Consumption Patterns on Instagram. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. Association for Computing Machinery, New York, NY, USA, 327–338.

[87] Janet Polivy and C. Peter Herman. 2020. Overeating in Restrained and Unrestrained Eaters. *Frontiers in Nutrition* 7 (2020), 30.

[88] Janet Polivy, C. Peter Herman, and Rajbir Deo. 2010. Getting a bigger slice of the pie. Effects on eating and emotion in restrained and unrestrained eaters. *Appetite* 55, 3 (2010), 426 – 430.

[89] V. M. Quick and C. Byrd-Bredbenner. 2013. Disturbed eating behaviours and associated psychographic characteristics of college students. *Journal of Human Nutrition and Dietetics* 26, s1 (2013), 53–63.

[90] Kiran K. Rachuri, Mirco Musolesi, Cecilia Mascolo, Peter J. Rentfrow, Chris Longworth, and Andrius Aucinas. 2010. EmotionSense: A Mobile Phones Based Adaptive Platform for Experimental Social Psychology Research. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp '10)*. ACM, New York, NY, USA, 281–290.

[91] Tauhidur Rahman, Mary Czerwinski, Ran Gilad-Bachrach, and Paul Johns. 2016. Predicting "About-to-Eat" Moments for Just-in-Time Eating Intervention. In *Proceedings of the 6th International Conference on Digital Health Conference (DH '16)*. Association for Computing Machinery, New York, NY, USA, 141–150.

[92] Nairan Ramirez-Esparza, Matthias R. Mehl, Javier Alvarez-Bermudez, and James W. Pennebaker. 2009. Are Mexicans more or less sociable than Americans? Insights from a naturalistic observation study. *Journal of Research in Personality* 43, 1 (2009), 1 – 7.

[93] 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.

[94] Amy Richards, Kendra K. Kattelmann, and Cuirong Ren. 2006. Motivating 18- to 24-Year-Olds to Increase Their Fruit and Vegetable Consumption. *Journal of the American Dietetic Association* 106, 9 (2006), 1405 – 1411.

[95] Eric Robinson, Paul Aveyard, Amanda Daley, Kate Jolly, Amanda Lewis, Deborah Lycett, and Suzanne Higgs. 2013. Eating attentively: a systematic review and meta-analysis of the effect of food intake memory and awareness on eating. *The American Journal of Clinical Nutrition* 97, 4 (02 2013), 728–742.

[96] Eric Robinson, Ellis Harris, Jason Thomas, Paul Aveyard, and Suzanne Higgs. 2013. Reducing high calorie snack food in young adults: A role for social norms and health based messages. *The international journal of behavioral nutrition and physical activity* 10 (06 2013), 73.

[97] Natti Ronel and Galit Libman. 2003. Eating Disorders and Recovery: Lessons from Overeaters Anonymous. *Clinical Social Work Journal* 31 (06 2003), 155–171.

[98] Helen K. Ruddock and Charlotte A. Hardman. 2018. Guilty pleasures: The effect of perceived overeating on food addiction attributions and snack choice. *Appetite* 121 (2018), 9 – 17.

[99] Alan Russell, Craig Hart, Clyde Robinson, and Susanne Olsen. 2003. Children's sociable and aggressive behaviour with peers: A comparison of the US and Australia, and contributions of temperament and parenting styles. *International Journal of Behavioral Development* 27, 1 (2003), 74–86.

[100] Doyen Sahoo, Wang Hao, Shu Ke, Wu Xiongwei, Hung Le, Palakorn Achananuparp, Ee-Peng Lim, and Steven C. H. Hoi. 2019. FoodAI: Food Image Recognition via Deep Learning for Smart Food Logging. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. ACM, New York, NY, USA, 2260–2268.

[101] B. Y. Salazar Vazquez, M. A. Salazar Vazquez, G. Lopez Gutierrez, K. Acosta Rosales, P. Cabrales, F. Vadillo-Ortega, M. Intaglietta, R. Perez Tamayo, and G. W. Schmid-Schoenbein. 2016. Control of overweight and obesity in childhood through education in meal time habits. The 'good manners for a healthy future' programme. *Pediatric Obesity* 11, 6 (2016), 484–490.

[102] Sarah-Jeanne Salvy, Kayla de la Haye, Julie C. Bowker, and Roel C.J. Hermans. 2012. Influence of peers and friends on children's and adolescents' eating and activity behaviors. *Physiology & Behavior* 106, 3 (2012), 369 – 378. Proceedings from the 2011 meeting of the Society for the Study of Ingestive Behavior (SSIB).

[103] 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 (Oct 2018), 2279–2292.

[104] Norbert Schwarz. 2012. *Retrospective and Concurrent Self-Reports: The Rationale for Real-Time Data Capture.*

[105] Hans Selye. 1976. *Stress without Distress.* Springer US, Boston, MA, 137–146.

[106] 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 (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 103–112.

[107] Edmund Seto, Jenna Hua, Lemuel Wu, Victor Shia, Sue Eom, May Wang, and Yan Li. 2016. Models of Individual Dietary Behavior Based on Smartphone Data: The Influence of Routine, Physical Activity, Emotion, and Food Environment. *PLOS ONE* 11, 4 (04 2016), 1–16.

[108] R. Sharma, V. I. Pavlovic, and T. S. Huang. 1998. Toward multimodal human-computer interface. *Proc. IEEE* 86, 5 (1998), 853–869.

[109] Alain Shema and Daniel E. Acuna. 2017. Show Me Your App Usage and I Will Tell Who Your Close Friends Are: Predicting User's Context from Simple Cellphone Activity. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 2929–2935.

[110] Christine Sheppard-Sawyer, Richard McNally, and Jennifer Fischer. 2000. Film-induced sadness as a trigger for disinhibited eating. *The International journal of eating disorders* 28 (10 2000), 215–20.

[111] Minati Singh. 2014. Mood, food, and obesity. *Frontiers in psychology* 5 (09 2014), 925.

[112] Neza Stiglic and Russell M Viner. 2019. Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews. *BMJ Open* 9, 1 (2019).

[113] Emma J Stinson, Susanne B. Votruba, Colleen A Venti, Marisol Perez, Jonathan Krakoff, and Marci E. Gluck. 2018. Food insecurity is associated with maladaptive eating behaviors and objectively measured overeating. *Obesity (Silver Spring, Md.)* 26 (2018), 1841 – 1848.

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

[115] Ethan Teo, Daniel Goh, Kamalakannan M. Vijayakumar, and Jean C. J. Liu. 2018. To Message or Browse? Exploring the Impact of Phone Use Patterns on Male Adolescents' Consumption of Palatable Snacks. *Frontiers in Psychology* 8 (2018), 2298.

[116] Heatherton TF and Baumeister RF. 1991. Binge eating as escape from self-awareness. *Psychol Bull.* (09 1991), 86–108.

[117] J Graham Thomas, Sapna Doshi, Ross D. Crosby, and Michael R Lowe. 2011. Ecological momentary assessment of obesogenic eating behavior: combining person-specific and environmental predictors. *Obesity* 19 8 (2011), 1574–9.

[118] Edison Thomaz, Irfan Essa, and Gregory D. Abowd. 2015. A Practical Approach for Recognizing Eating Moments with Wrist-Mounted Inertial Sensing. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 1029–1040.

[119] Stewart G. Trost. 2007. State of the Art Reviews: Measurement of Physical Activity in Children and Adolescents. *American Journal of Lifestyle Medicine* 1, 4 (2007), 299–314.

[120] Bengisu Tulu, Carolina Ruiz, Joshua Allard, Joseph Acheson, Andrew Busch, Andrew Roskusku, Gage Heeringa, Victor Jaskula, Jessica Oleski, and Sherry Pagoto. 2017. SlipBuddy: A Mobile Health Intervention to Prevent Overeating. (01 2017).

[121] Pixie Turner and Carmen Lefevre. 2017. Instagram use is linked to increased symptoms of orthorexia nervosa. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity* 22 (03 2017).

[122] Tatjana van Strien, C. Peter Herman, and Marieke W. Verheijden. 2009. Eating style, overeating, and overweight in a representative Dutch sample. Does external eating play a role? *Appetite* 52, 2 (2009), 380 – 387.

[123] Tatjana van Strien, C. Peter Herman, and Marieke W. Verheijden. 2012. Eating style, overeating and weight gain. A prospective 2-year follow-up study in a representative Dutch sample. *Appetite* 59, 3 (2012), 782 – 789.

[124] Elizabeth A Vandewater, Mi suk Shim, and Allison G Caplovitz. 2004. Linking obesity and activity level with children's television and video game use. *Journal of Adolescence* 27, 1 (2004), 71 – 85. Video Games and Public Health.

[125] Lenny Vartanian, Natalie Reily, Samantha Spanos, C Herman, and Janet Polivy. 2017. Self-reported overeating and attributions for food intake. *Psychology & health* 32 (01 2017), 1–10.

[126] Lenny R. Vartanian, Natalie M. Reily, Samantha Spanos, Lucy C. McGuirk, C. Peter Herman, and Janet Polivy. 2017. Hunger, taste, and normative cues in predictions about food intake. *Appetite* 116 (2017), 511 – 517.

[127] Carine A Vereecken, Joanna Todd, Chris Roberts, Caroline Mulvihill, and Lea Maes. 2006. Television viewing behaviour and associations with food habits in different countries. *Public Health Nutrition* 9, 2 (2006), 244–250.

[128] 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).

[129] Tri Vu, Feng Lin, Nabil Alshurafa, and Wenyao Xu. 2017. Wearable Food Intake Monitoring Technologies: A Comprehensive Review. *Computers* 6, 1 (Jan 2017), 4.

[130] Morgan Walker, Laura Thornton, Munmun [De Choudhury], Jaime Teevan, Cynthia M. Bulik, Cheri A. Levinson, and Stephanie Zerwas. 2015. Facebook Use and Disordered Eating in College-Aged Women. *Journal of Adolescent Health* 57, 2 (2015), 157 – 163.

[131] Birgitte Wammes, Boudewijn Breedveld, Stef Kremers, and Johannes Brug. 2006. The 'balance intervention' for promoting caloric compensatory behaviours in response to overeating: a formative evaluation. *Health Education Research* 21, 4 (04 2006), 527–537.

[132] B Wammes, S French, and J Brug. 2007. What young Dutch adults say they do to keep from gaining weight: self-reported prevalence of overeating, compensatory behaviours and specific weight control behaviours. *Public Health Nutrition* 10, 8 (2007), 790–798.

[133] 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.

[134] Jane Wardle, Anne Haase, Andrew Steptoe, Maream Nillapun, Kiriboon Jonwutiwes, and France Bellisle. 2004. Gender Differences in Food Choice: The Contribution of Health Beliefs and Dieting. *Annals of behavioral medicine : a publication of the Society of Behavioral Medicine* 27 (05 2004), 107–16.

[135] K.R. Westerterp. 2005. 4 - Physical activity and obesity. In *Food, Diet and Obesity*, David J. Mela (Ed.). Woodhead Publishing, 76 – 85.

[136] World Health Organization (WHO). 2020. Obesity and overweight. Retrieved April 28, 2020 from https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

[137] Donald A. Williamson, David H. Gleaves, and Olga J. Lawson. 1991. Biased perception of overeating in bulimia nervosa and compulsive binge eaters. *Journal of Psychopathology and Behavioral Assessment* 13 (1991), 257–268.

[138] Tingxin Yan, David Chu, Deepak Ganesan, Aman Kansal, and Jie Liu. 2012. Fast App Launching for Mobile Devices Using Predictive User Context. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys '12)*. Association for Computing Machinery, New York, NY, USA, 113–126.

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

[140] Yang Yue, Tian Lan, Anthony G.O. Yeh, and Qing-Quan Li. 2014. Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behaviour and Society* 1, 2 (2014), 69 – 78.

[141] Tok Chen Yun, Siti Rohaiza Ahmad, and David Koh Soo Quee. 2018. Dietary Habits and Lifestyle Practices among University Students in Universiti Brunei Darussalam. *The Malaysian Journal of Medical Sciences : MJMS* 25 (2018), 56 – 66.

[142] Mattia Zeni, Ilya Zaihrayeu, and Fausto Giunchiglia. 2014. Multi-Device Activity Logging. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 299–302.

[143] Shibo Zhang, William Stogin, and Nabil Alshurafa. 2018. I sense overeating: Motif-based machine learning framework to detect overeating using wrist-worn sensing. *Information Fusion* 41 (2018), 37 – 47.

[144] Konglin Zhu, Xiaoyi Zhang, Bin Xiang, and Lin Zhang. 2015. Exploiting User Context and Network Information for Mobile Application Usage Prediction. In *Proceedings of the 7th International Workshop on Hot Topics in Planet-Scale MObile Computing and Online Social NeTworking (HOTPOST '15)*. Association for Computing Machinery, New York, NY, USA, 25–30.