Protecting Mobile Food Diaries from Getting too Personal

Lakmal Meegahapola
Idiap Research Institute and EPFL
Switzerland
lmeegahapola@idiap.ch

Salvador Ruiz-Correa
IPICYT
Mexico
salvador.ruiz@ipicyt.edu.mx

Daniel Gatica-Perez
Idiap Research Institute and EPFL
Switzerland
gatica@idiap.ch

ABSTRACT

Smartphone applications that use passive sensing to support human health and well-being primarily rely on: (a) generating low-dimensional representations from high-dimensional data streams; (b) making inferences regarding user behavior; and (c) using those inferences to benefit application users. Meanwhile, sometimes these datasets are shared with third parties as well. Human-centered ubiquitous systems need to ensure that sensitive attributes of users are protected when applications provide utility to people based on such behavioral inferences. In this paper, we demonstrate that inferences of sensitive attributes of users (gender, body mass index category) are possible using low-dimensional and sparse data coming from mobile food diaries (a combination of sensor data and self-reports). After exposing this potential risk, we demonstrate how deep learning techniques can be used for feature transformation to preserve sensitive user information while achieving high accuracies for application-related inferences (e.g. inferring the type of consumed food). Our work is based on two datasets of daily eating behavior of 160 young adults from Switzerland (N_{CH}=122) and Mexico (N_{MX}=38). Results show that using the proposed approach, accuracies in the order of 75%-90% can be achieved for application related inferences, while reducing the sensitive inference to almost random performance.

CCS CONCEPTS

• **Human-centered computing** → **Mobile computing**; **Smartphones**; **Mobile phones**; **Empirical studies in ubiquitous and mobile computing**; • **Social and professional topics** → *Gender*; • **Applied computing** → *Consumer health*; *Health informatics*; *Sociology*.

KEYWORDS

mobile sensing, smartphone sensing, eating behavior, food journal, food diary, privacy, demographic attribute, sensitive attribute

ACM Reference Format:
Lakmal Meegahapola, Salvador Ruiz-Correa, and Daniel Gatica-Perez. 2020. Protecting Mobile Food Diaries from Getting too Personal. In *19th International Conference on Mobile and Ubiquitous Multimedia (MUM 2020), November 22–25, 2020, Essen, Germany*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3428361.3428468

MUM 2020, November 22–25, 2020, Essen, Germany

ACM ISBN 978-1-4503-8870-2/20/11…$15.00
https://doi.org/10.1145/3428361.3428468

1 INTRODUCTION

There is a booming industry around mobile sensing applications for human health and well-being. Research has attempted to infer health and well-being related attributes such as stress [37, 52, 75], emotions and mood [70, 83], well-being [47, 50, 91], alcohol consumption [15, 77], and other behavior [69] using smartphone sensor data and self-reports. In the specific case of food and nutrition, studies demonstrated the potential to identify eating occasions using wearable cameras and sensors [17, 78, 84, 85]. Moreover, automatically inferring attributes related to eating patterns (e.g. meal and snack patterns, social context of eating, food categories, etc.) using smartphone sensing has been emphasized as a key to design the next generation of food-related health and well-being applications, as these patterns would be invaluable for mobile interventions, dietary monitoring, and fitness applications [17, 21, 79, 84, 85].

Most commercial mobile food diary-based health and well-being applications such as Samsung Health [12], Google Fit [8], and Apple Health [10] passively sense activity information by transforming high-dimensional sensor data from accelerometer, location, gyroscope, and other sensors into low-dimensional features such as step count, semantic location, and activity type. Moreover, they collect data regarding food intake as food diaries [12]. Such applications usually provide an option for users to provide sensitive information such as gender, body mass index (BMI), and age, claiming that if they provide such data, personalized services could be provided with better quality of service [4, 6, 11]. While some users might be willing to provide such data, other users would prefer to use the application without providing sensitive information, thus setting trade-offs that involve personalization, privacy, and utility, when using applications and services [28, 86, 90]. How this conundrum plays a role in ubiquitous computing is described in [16], which emphasizes the need for privacy-preserving systems. Moreover, according to the terms of use of several mobile health apps [4, 6, 11], this is exactly why they use personalization for users who provide such sensitive information, and non-personalized algorithms for users who refuse to provide such data, but still opt to use the app.

Another concerning issue is that tech companies who own such low-dimensional data have often sold data to third parties (i.e. advertisers, insurance companies, etc.), and there is no full clarity as to how companies use data [16, 22, 45, 56, 64, 68, 71]. According to recent reports from consumer protection agencies [1, 7], this trend is not diminishing. Even though data might be anonymized before sharing it with third parties, it is not fully understood whether such low-dimensional data can be used to infer sensitive attributes without user consent using models developed by those third parties, specially for health-related information including food intake and activity levels. For example, a health insurance company can obtain anonymized food intake data through data brokers, and use a machine-learning model developed by them to infer sensitive attributes such as BMI (whether it is high or low, which is an indicator of the overall weight condition of a person), which could guide the decision to insure a person or not [2, 3]. Even though it is an unethical practice, these risks exist.

Given this context, in this paper, we examine two mobile sensing and self-reporting datasets about eating behavior. Using sparse, low-dimensional features regarding eating behavior, our paper addresses two research questions:

RQ1: Do everyday eating practices captured via low-dimensional, sparse smartphone sensing data pose a risk regarding the possible inference of sensitive attributes (gender and BMI as examples), and thus pose a risk about sharing such behavioral traces online?

RQ2: Can sensitive attributes embedded in low- dimensional and sparse datasets be preserved with deep learning techniques based on multi-task neural networks and autoencoders, such that high accuracies in essential application-oriented inferences can still be achieved?

By addressing the above research questions, our work contains the following contributions.

Contribution 1: We demonstrate that using low-dimensional features generated from sensors and self-reports regarding the eating behavior of two independent sets of university students (in Switzerland and Mexico), there is a risk of inferring sensitive attributes such as gender and BMI category with accuracies of the range 74% - 78%, while also using the same feature set for six application inferences at the episode level produces accuracies of the range 73%-86%. This shows a potential risk to users of mobile apps related to food diaries, as people might not be aware that sensitive attributes can be inferred in cases when they have not provided such information to the application, or when anonymized datasets are shared with third parties who might have trained models that could infer sensitive attributes:

Contribution 2: We show that by using a deep learning-based autoencoder architecture with a modified loss function, we are able to generate features that obscure the inference of sensitive attributes, while achieving high accuracies for application inferences, hence preserving their utility. Using this technique, gender and BMI-category inference accuracies dropped to around 50%, while application inference accuracies were maintained above 75% for most inferences in both the datasets. Moreover, we demonstrate the applicability of this approach across datasets from two different countries for different sensitive and application inference combinations.

This article is organized as follows. In Section 2, we present the key definitions and usage scenarios. In Section 3, we examine related work regarding ubiquitous computing, food, and demographic attribute inference using mobile sensing data. Next, we introduce the dataset used in this study in Section 4. In Section 5, we show the risks related to sensitive attribute inference, and in Section 6 we present our proposed deep learning-based technique to combat such issue in low-dimensional data. We discuss the implications of our work in Section 7 and conclude the paper in Section 8.

2 KEY DEFINITIONS AND USAGE SCENARIOS

Sensitive Inference: These tasks attempt to infer sensitive information regarding mobile health application users (e.g. gender, BMI, weight, height, etc.). The key concept we show in this paper is that there is a risk of inferring such sensitive attributes from datasets that at first would not seem amenable for such tasks (e.g. data from mobile food diaries) in cases when users do not explicitly provide such data. As a summary, the sensitive inferences we consider in this study are: (**S1**) Gender Inference - Men vs. Women; and (**S2**) BMI-Category Inference - Specifically, we used BMI values with a median split to define two categories: High BMI (higher than the median value), and Low BMI (lower than the median value) of each dataset. Moreover, using gender and BMI in the context of food diaries is relevant because literature in nutrition research suggests links between gender and BMI and eating habits, and the importance of altering eating habits to the consumption of certain foods like dairy products, meat, and oils, in order to control weight and other health conditions [32, 76, 93]. The choice of these two attributes was done considering the available attributes in the datasets we used. However, depending on the use-case or context, others attributes can be chosen when the technique is used on other datasets. In addition, in this study, we use the term *gender* to designate the demographic *sex* variable, aware of the fact that the two terms are neither identical nor binary [87, 95]. We use gender for sake of consistency with most previous literature in computing that refers to tasks such as gender inference [13, 25, 42, 44, 49, 80, 89] as described in Section 3, while acknowledging the limitations that this previous work has had about the conceptualization of gender [9].

Application Inference: These are inferences done on mobile food diaries and health applications to benefit users. In the context of this paper, we consider five useful inferences about food types (A1-A5) done using low-dimensional data such as: (**A1**) Meal vs. Snack, (**A2**) Sweet vs. Non-Sweet, (**A3**) Dairy vs. Non-Dairy, (**A4**) Fatty vs. Non-Fatty, and (**A5**) Meat vs. Non-Meat. We attempt to infer whether food contained fats and oils, sweets, dairy, or meat; all of which could be important in multiple ways considering food diaries because, for example, higher consumption of fruits is considered healthier [57], and high consumption of meat might provide high amounts of protein [39, 82] and at the same time might increase risks of several diseases including cardiovascular problems and cancer [35, 65]. Hence, according to prior literature, if most of these food intake related variables can be inferred using contextual and activity related data, such inferences would be valuable for users of mobile food diary apps [21, 23, 34, 43, 79, 85, 92].

Usage Scenarios: Recently, given the appearance of frameworks to regulate the collection and use of personal data like the European General Data Protection Framework (GDPR) [88], there is a push for explicitly not collecting personal information from app users without a clear purpose [29, 41, 67]. Currently, two problems in the current operation of mobile food diaries are:

(1) It is not known if mobile food diaries that collect low dimensional data can (or do) infer sensitive user attributes using machine learning models trained with other similar datasets, even when users do not provide such information. In principle, there can be apps that are trusted or not trusted by users, and depending on the trust level, users might decide what data and sensors the app is given access to [53, 54].

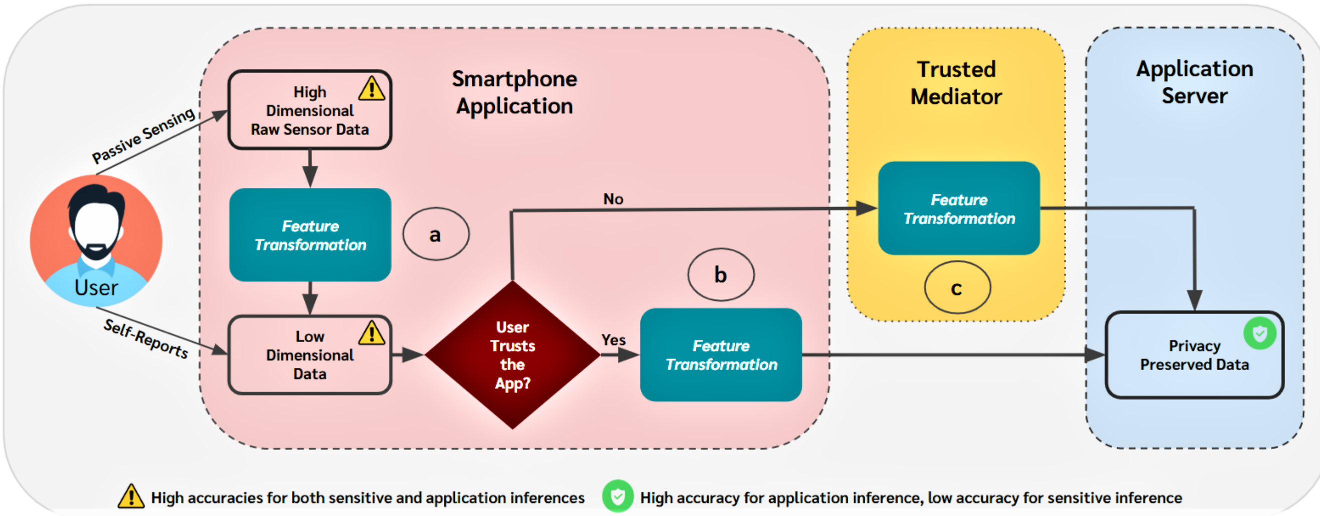

Figure 1: Envisaged Usage Scenarios for the Feature Transformation Technique

(2) It is not known if any third parties who get access to health and food related data (even anonymized data) can (or do) infer sensitive attributes, which represents a privacy risk to users [7, 30, 63, 68].

Considering trusted and non-trusted app usage scenarios suggested in prior work, we suggest the following envisaged usage scenarios for the technique we propose (Figure 1):

Trusted Apps (Figure 1–b): With trusted apps, user data will be directly sent to an app server from the smartphone. Users will be provided with insights, analytics, and interventions. The feature transformation (the approach we propose) should be carried out in the mobile application before data is sent to application servers. This way, application providers can assure users that their sensitive data would be protected to a greater extent (to a level where their data is unlikely to reveal sensitive attributes even if data are breached after a server hack). Currently, similar guarantees are provided by companies such as Apple and Google where they publicly declare to users that they use a plethora of techniques (e.g. differential privacy, federated learning, etc.) to preserve sensitive user data. In this case, since only the application server is aware of the feature transformation, even though the server can generate application inferences using the transformed data, for any other entity who gets to access the data, there is lower utility as the data is uninterpretable after the transformation. Hence, our technique enables applications to provide utility to users, while providing additional guarantees regarding user privacy.

Non-Trusted Apps (Figure 1–C): According to this idea, users grant applications only a limited range of inferences (instead of providing permission to sensors, users provide permission for inferences) via a trusted intermediate component/app in/outside the mobile phone. This idea regarding non-trusted apps and how data should be transformed using trusted mediators in the smartphone before sending data to the cloud was initially proposed by Malekzadeh et al. [53–55] for high dimensional data. The key idea behind this is that the non-trusted app server would send details to the trusted mediator regarding the inference that it needs to do in order to provide a service to the user. Knowing the exact inference, the trusted mediator would do a feature transformation to the data, such that inference required by the app server can be performed but any sensitive attribute inference is made difficult. However, what we show in this study is that feature transformation on high-dimensional sensor data is not sufficient, specially in mobile food diaries, if such techniques still produce low-dimensional data with a high accuracy. Hence, we need an additional step for feature transformation of low-dimensional data.

3 RELATED WORK

Food Diaries, Mobile Sensing, and Food Intake. While few studies have revealed adverse consequences of using food diaries for certain individuals [81], in many prior studies, food diaries have proven to be effective in eating behavioral change and interventions [60, 94]. There are many commercial applications [5, 6, 8, 12] that are intended for tracking food consumption by manual entry of food information in a mobile application, and research applications that automatically infer food type and nutritional content using an image of the food dish [43, 73]. Moreover, there are studies [14, 17, 84, 85] that emphasize the use of wearables to automate the generation of food diaries. Studies by Thomaz et al. [84, 85] focused on using wearables to identify eating moments. Going a step forward, Mirtchouk et al. [61] used wearables to identify food types as well.

If we specifically consider smartphone sensing applications in this domain, Biel et al. [21] used a smartphone sensing based application with self-reports (similar to a food diary) to track the eating behavior of university students. They demonstrated to be able to infer meal vs. snack occasions (food types) with an accuracy of more than 85% using a random forest classifier. They suggested that such automatic inferences are important in contexts such as mobile nutritional interventions and mobile recommendation engines. In another recent study, Seto et al. [79] emphasized the importance of smartphone sensing based applications to monitor eating behavior in order to promote healthier lifestyles by providing timely interventions. They provided preliminary evidence that ties routines, physical activity levels, and food environment. Apart from these studies, there have been studies that examined how meal and snack eating behavior can affect aspects such as diet control and weight control in nutrition research [18, 27, 31, 40, 48]; and how physical activity levels derived from wearable devices are associated with food consumption behavior and weight loss [38, 66, 72]. It should be noted that, while the main goal of these studies is to examine eating patterns, meal intake or snacking behavior, our study is focused on examining whether such food related behavioral traces can also reveal sensitive attributes of users, and to propose a methodology to address this potentially problematic issue.

Inference of Demographic Attributes using Mobile Sensing Data. Inference of demographic attributes has been a topic in the mobile computing community for some time. Early studies [24, 96] used high-dimensional features, and the main correlating feature for the specific attribute we study (gender) was the raw accelerometer trace. Moreover, in the recent past, many studies have been published regarding inferring demographic attributes such as gender and age [42, 62, 89] using raw accelerometer and gyroscope traces of wearables and mobile phones. In these papers, sensor data are again high-dimensional and raw, and need to be transformed in many ways to engineer features.

Compared to these existing studies, our work focuses on eating behavior, demonstrating that there is a privacy risk of inferring gender and BMI-category using low-dimensional, sparse mobile sensing and self-report data. According to Kotz et al. [45], if we specifically focus on mobile health applications that contain personal health records and food diaries in addition to all the passively sensed data, the privacy risk could be even higher.

Protecting Sensitive Attributes Embedded in High Dimensional Data Streams. While demographic attributes are important for various applications including mobile recommendation engines [24, 96], it should also be noted that application users might not have provided this information, or might not even know that their data is being used for such purposes. To tackle this issue, a plethora of adversarial learning techniques have been proposed [20, 33, 51]. In the context of high-resolution mobile sensor data, Malekzadeh et al. [53–55] proposed techniques based on autoencoders and generative adversarial networks (GAN) to replace sensor traces that reveal sensitive information with non-sensitive information [55], and augment high-resolution sensor data streams to preserve privacy of users when sharing data with servers via a trusted mediator [53], while making sure that activity inference accuracy is kept high (above 85%). Hence, they show that these privacy-preserved high-resolution information traces can be utilized to generate low-dimensional features such as step count and activities accurately. Our study focuses on reducing sensitive information leakages from low-dimensional sparse data traces, which is a challenging problem. With our results, we highlight that privacy preservation of high-dimensional or high-resolution data is not enough to ensure user privacy when low-dimensional information can be engineered accurately to infer sensitive information (see the illustration in Figure 1). Further, we argue that using autoencoder-based feature transformation techniques on low-dimensional mobile data regarding eating behavior can increase user privacy because generating additional features from these datasets is highly challenging.

4 DATASETS AND PRE-PROCESSING

Switzerland Dataset (CH-Dataset): We used a mobile sensing dataset called *Bites'n'Bits* from our previous research [21, 36]. It contains smartphone sensor data, self-reported data, and activity data of fitbit wearables from 122 students of a Swiss university. The smartphone application allowed users to self-report details regarding eating events in-situ (denoted by **C**: time of eating, social context of eating, food types and categories, concurrent activities, etc.). Further, their activity levels were captured using a fitbit wearable (denoted by **A**: step count). Moreover, passive sensing data regarding the context such as location of eating events were captured (also part of the **C** data category). Moreover, the demographic attributes (denoted by **D**) that were captured from participants were gender and BMI (participants self-reported their height and weight, using which BMI was calculated). In the final dataset, there are 4448 eating reports (3414 meals, 1034 snacks). All the users who took part in the study were between 18-26 in age, with a mean age of 20.5 years, and there were 65% men and 35% women.

Mexico Dataset (MX-Dataset): We collected another dataset using the same approach as in [21, 36], from 38 university student in San Luis Potosi, Mexico. The dataset had self-reported features similar to the CH-Dataset (**C**), and instead of the FitBit wearable, activity levels of participants were captured using accelerometer sensor in the smartphone (**A**: x, y, and z axis values of the accelerometer). Moreover, this dataset contained additional features that revealed details about the user context (**C:** app usage, radius of gyration, screen events, battery charging events). Gender and BMI were captured as self-reported demographic attributes (**D**). The dataset contained 1031 fully complete eating episode reports (642 meals, 389 snacks). The average age of study participants was 23.4 years, and the cohort had 44% men and 56% women.

The datasets we used for our analysis only contained high-level information (i.e. low-dimensional data) describing the eating behavior. This made sure that high-dimensional features that have been shown to reveal sensitive attributes (e.g. accelerometer traces) are not present in this data, while examining whether such inference can be based on low-dimensional data. The two datasets contained one entry per each eating event, and both datasets have been prepared using the same procedure as suggested by Biel et al. [21] where passive sensing data around eating events are aggregated from the time span **T**-β to **T**+β when a self-report about the eating event was captured at time **T**. Hence, these data contained all the self reported information regarding the eating event, and sensing data around (before and after) eating events in a 2β time window. The datasets processed has been prepared using β values: $\beta = 2$ hours in CH-Dataset and $\beta = 30$ minutes in MX-Dataset (we present results for these β values because best sensitive inference results were obtained for those values). Table 1 shows a summary of features and target variables. Group D shows the target variables for Sensitive inferences. Group F shows the target variables for Application inferences. The rest of the features were used as input features in the inference models.

Table 1: Feature groups are Demographic (D), Contextual (C), Food Category (F), and Activity (A). Type describes whether the feature is categorical (CA) or numerical (NU), and if it is categorical, how many categories are represented by the feature. The total number of features are 18 and 44 in the CH and MX datasets, respectively.

CH-Dataset				MX-Dataset			
Feature	**Description**	**Type**	**Group**	**Feature**	**Description**	**Type**	**Group**
gender	Man/Woman	CA(2)	D	gender	Man/Woman	CA(2)	D
bmi	Body Mass Index category (high/low)	CA(2)	D	bmi	Body Mass Index category (high/low)	CA(2)	D
meal_snack	Whether it is a meal or a snack	CA(2)	F	meal_snack	Whether it is a meal or a snack	CA(2)	F
sweet	Whether it is a sweet food or not	CA(2)	F	fatty	Whether food is fatty or non-fatty	CA(2)	F
dairy	Whether the food contains dairy or not	CA(2)	F	meat	Whether the food contains meat or not	CA(2)	F
time_since_meal	Time in minutes, since the last meal	NU	C	time_since_meal	Time in minutes, since the last meal	NU	C
time_in_min	Time of the day	NU	C	time_in_min	Time of the day	NU	C
where	Location of eating	CA(10)	C	where	Location of eating	CA(10)	C
withwhom	Social context of a eating (alone, friends, etc)	CA(4)	C	withwhom	Social context of a eating (alone, friends, etc)	CA(8)	C
whatelse	Concurrent activities while eating	CA(17)	C	whatelse	Concurrent activities while eating	CA(11)	C
steps_X_Y	Features derived using fitbit step counts	NU	A	charging or not	Whether the phone is charging when eating	CA(2)	C
	X = total, median, mean or std. deviation			battery_level	phone battery level when eating	NU	C
	Y = bef/aft to indicate before eating or after			screen_on/off	Number of screen on/off events	NU	C
				rog	radius of gyration during eating time window	NU	C
				app_X	whether X app was used or not	CA(2)	C
					X = facebook, instagram, whatsapp, etc.		
				mood, stress	mood and stress while eating	CA(5)	C
				acc_A_B	Derived using accelerometer sensor	NU	A
					B = bef/aft to indicate before eating or after		
					A = Used indicate the X,Y, or Z axis		

Table 2: Gender and BMI Inference accuracy from the random forest classifiers (RF) when using different feature groups.

	CH-Dataset		MX-Dataset	
Feature Groups	**Gender**	**BMI**	**Gender**	**BMI**
Baseline	50.00%	50.00%	50.00%	50.00%
A	65.13%	67.41%	66.73%	65.79%
C	72.51%	70.49%	68.91%	67.46%
C+D	74.39%	72.72%	74.39%	73.64%
C+A	**77.38%**	**74.75%**	**77.21%**	**76.39%**
C+A+D	91.39%	89.12%	80.63%	81.29%

Table 3: Feature Importance (FI) for the top-five features using RF for sensitive inferences with C+A feature group. GQS and MSL corresponds to google quick search and microsoft launcher, respectively.

CH-Dataset				MX-Dataset			
Gender		BMI		Gender		BMI	
Feature	FI	Feature	FI	Feature	FI	Feature	FI
time_in_mins	0.109	time_in_mins	0.119	app_GQS	0.086	stress	0.070
steps_sd_aft	0.104	steps_sd_bef	0.094	whatelse	0.053	feeling	0.053
steps_sd_bef	0.089	time_since_meal	0.092	app_MSL	0.048	rog	0.041
steps_mean_aft	0.087	steps_sd_bef	0.091	rog	0.036	acc_Y_bef	0.032
steps_tot_aft	0.087	steps_mean_aft	0.087	acc_Z_bef	0.109	acc_Y_aft	0.031

5 INFERRING SENSITIVE ATTRIBUTES USING MOBILE FOOD DIARIES (RQ1)

In this section, we examine the feasibility of inferring sensitive attributes using the two low-dimensional datasets. We used support vector machines, neural networks, and random forest classifiers for this task. Due to space limitations, we only report results from random forest classifiers that were marginally higher than neural networks. For this experiment, we used Random Forest Classifiers (RF) with an ntree values of range 200-500 for different feature groups. We used 10-fold cross validation during training, and when preparing the dataset, we made sure that the classes are balanced by up-sampling the minority class. It should be noted that we followed a leave-k-participants-out strategy for all the experiments, where training, validation, and testing sets did not include data from the same user. We ended up with datasets with sizes 4200 in the CH-Dataset and 1000 in the MX-Dataset (corresponding to single eating events) for the experiment.

Results of this experiment are summarized in Table 2. In the CH-Dataset, when using sensor and self-reported contextual information alone (C), the classifiers achieved an accuracy of 72.51% using RF for gender inference. When we included BMI to contextual data (C+D), the accuracies were increased to 74.39%. Accuracy was even higher when using C+A feature group. However, when additional demographic information (BMI category) was also used to form the feature group C+A+D, gender inference accuracy increased to 91.39% with RF. Similar results were attained for gender inference in MX-Dataset as well. Moreover, in the BMI category inference task, we used gender as the feature in the D feature group. Results for BMI inference showed reasonably high accuracies in the range 74%-76% for both datasets, for C+A feature group. C+A+D feature group showed accuracies of 89.12% for the CH-Dataset and 81.29% for the MX-Dataset, again showing how knowing one sensitive attribute makes it easier to infer another sensitive attribute. Furthermore, since we are specifically interested in demonstrating the effects of smartphone sensing and self-reported data, when presenting accuracy values for sensitive inference and application inferences in later sections, we only present the accuracies obtained with the contextual and physical activity feature (C+A) for both sensitive and application inferences. Moreover, Table 3 summarizes the top-five features based on feature importance in RF classifier, for each inference done with C+A feature group in Table 2.

6 PROTECTING SENSITIVE ATTRIBUTES IN MOBILE FOOD DIARIES (RQ2)

This section will be divided into methodology and results. The methodology and results sections contain two subsections each. First, in order to facilitate the process of transforming dataset features such that sensitive attributes are protected, we train a Multi-Task Neural Network (MT-NN) [26] (Step 1). Step 2 describes the procedure to use an Autoencoder (AE) [46] together with the trained MT-NN to transform features using a modified loss function such that using the output data from the AE, sensitive inferences cannot longer achieve high accuracies, while still enabling high accuracies for application inferences.

6.1 Methodology

6.1.1 Step 1: Multi-task Neural Networks for Sensitive and Application Inferences. Most applications and third party services that use sparse, low-dimensional datasets such as the one studied here, use such data for application inferences. To show the conundrum between utility of application inferences and risks of sensitive inferences, we train a MT-NN, and show that the model is able to perform an application inference (e.g. meal vs. snack, sweet vs. non-sweet or dairy vs. non-dairy, etc.), and a sensitive inference (men vs. women or high-BMI vs. low-BMI) on the same dataset. We use examples of application inferences that mobile health apps could target, to illustrate the possibility of such joint inferences. Example of such a joint inference is using a MT-NN to infer meal vs. snack and men vs. women in the CH-Dataset. Similarly, for each dataset, we considered six joint inference tasks (using two sensitive inferences and three application inferences), hence leading to a total of 12 inferences.

The MT-NN consisted of five layers, where the input layer had dense neurons equal to the number of input features. Intermediate layers had 32-64, 32-64 and 16-32 dense neurons, respectively depending on the inference task, whereas the two outputs corresponded to binary values representing the two inference tasks. Dropout was used for regularization in intermediate layers, relu was the activation function of intermediate layers, sigmoid activation was used for outputs, binary cross entropy was used to calculate loss for both inference tasks, and 10-fold cross validation was used. Even though the results hold for both C+A and C+A+D feature groups, we provide results only for the C+A feature group due to space limitations, and because that feature group represents a use-case where app servers have no sensitive information about users.

6.1.2 Step 2: An Autoencoder Based Architecture to limit Sensitive Inferences. We propose how deep learning techniques can be adjusted to suit a low-dimensional dataset, such that further privacy risks are reduced. Initially, we trained and tested the MT-NN as described in Step 1 using binary cross entropy loss function for both sensitive inference and application inference. Then, we created an AE with an equal number of dense neurons in the input/output layers (also equal to the number of features in the dataset); with 12,10,8,10,12 dense-neurons in each intermediate layer, elu activations for intermediate layers, and sigmoid activations for the output layer. The AE + MT-NN based architecture is shown in Figure 2.

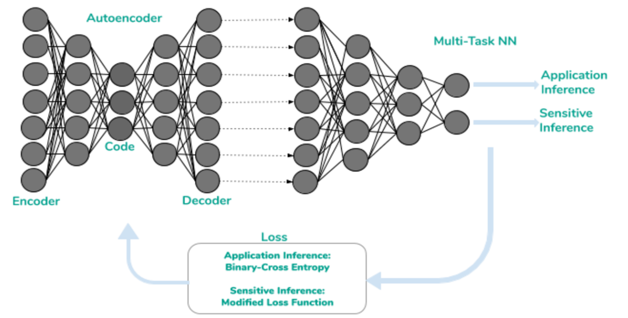

Figure 2: AE and MT-NN based architecture for privacy preserving feature transformation. Output of the AE is directly mapped to the input of MT-NN. AE's loss function is based on the losses of sensitive inference and application inference.

We locked the weights of the MT-NN so that its weights do not get tuned during the training process of the AE, and then trained the AE using the training dataset.

$$L_{sen} = |\alpha - F_{sen}(B_i)| \tag{1}$$

$$L_{app} = -(F_{app}(B_i) \times log(p) + (1 - F_{app}(B_i)) \times log(1-p)) \tag{2}$$

$$F_{ae} = \arg\min_{B_i}(L_{sen} - L_{app}) \tag{3}$$

If we define our dataset as X_n, the two functions for sensitive and application inferences can be defined as $F_{sen}(.)$ and $F_{app}(.)$. The objective is to find a feature transformation function for AE, denoted by $F_{ae}(.)$, where the resultant dataset from the autoencoder is $X_n^* = F_{ae}(X_n)$ such that $F_{sen}(X_n^*)$ accuracy is not high, hence preserving sensitive attributes about users, and $F_{app}(X_n^*)$ is high (closer to 100%), providing high inference accuracies for application inferences. In the training phase of the AE, for a given data point B_i, the output of the MT-NN for the sensitive inference would be $F_{sen}(B_i)$, and the application inference output would be $F_{app}(B_i)$ whereas the two losses are indicated by Equations 1 and 2, respectively. The objective of the autoencoder is represented by Equation 3 which combines the losses from the two inferences in the MT-NN, and aims at minimizing the loss for the training dataset. Finally, p is the probability of the outcome.

To make sure that AE learns its parameters to create a dataset that provides higher accuracies for application inference and lower accuracies for sensitive inferences, we used a modified loss function as in Equation 1 for gender/BMI (we use the value α=0.5 because it is desired accuracy for the binary classification task to make sure that it has a lower accuracy [53]), and traditional binary cross entropy (given in Equation 2) for application inference. Hence, the loss for the AE was derived from the two output losses of the MT-NN as given in Equation 3, whereas no matter how high the loss for gender/BMI classification is, it is not conveyed as it is to the AE due to the modified objective. This allows the AE to tune its weights such that resultant dataset after the feature transformation care less about the accuracy of sensitive inferences, and the features are transformed to ensure reasonable accuracies for application inferences. After the training process, we obtain the data with transformed features using the AE. The trained AE is the component than can be used in the points marked as 2 or 3 in the diagram of Figure 1.

As the final step, using the trained AE, we obtained a final dataset that is Privacy Preserved. We trained the RFs and NNs for sensitive and application inferences for both datasets using the final dataset. The intuition here is to check whether the modified dataset can provide good accuracies for application inferences, and lower the accuracies for sensitive inferences, even if a new model is trained.

6.2 Results

6.2.1 Step 1: Multi-task Neural Networks can jointly infer sensitive attributes and eating events . Results from this experiment are shown under the column *MT-NN Before AE* in Table 4 and Table 5 for the CH-Dataset and MX-Dataset, respectively. In the CH-Dataset, the MT-NN achieved a meal vs. snack inference accuracy of 86%, sweet vs non-sweet inference accuracy of 83%, and dairy vs. non-dairy inference accuracy of 78%. These results are similar to results obtained using the RF (RF Before AE). Moreover, C+A feature groups provide significantly high accuracies for gender/BMI inference which highlights the need for privacy-preserving solutions for low-dimensional and sparse data from mobile food journals. Similar results hold for the MX-Dataset where application inference accuracies using both RF and MT-NN were in the range 80%-85% and sensitive inference accuracies were in the range 72%-79% before using the AE based feature transformation.

6.2.2 Step 2: Our architecture limits sensitive inferences while providing utility for eating-related inferences. After training the AE to transform dataset features so that sensitive inferences are made difficult following the procedure given above, we measure both the application inference and sensitive inference accuracies for the transformed dataset using the newly trained RFs and MT-NNs. Table 4 and Table 5 show the results for the CH-Dataset and MX-Dataset respectively, using a comparison between accuracy results before and after the use of AE for MT-NN and RF for three inference pairs in both datasets. Application inference accuracies have been kept reasonably high for all three inference pairs in both datasets (the CH-Dataset: above 81% for MT-NN and 85% for RF in meal vs. snack and similar results hold for other two application inferences as well; the MX-Dataset: above 78% for meal vs. snack and similar results hold for other two inferences). At the same time, in the CH-Dataset, we were able to reduce the gender inference accuracy from 67% to 51% for MT-NN and from 77% to 48% for RF, and a similar trend can be seen for the BMI-category. A similar pattern in results can be seen for other two application inferences in the CH-Dataset, and for sensitive inferences in the MX-Dataset too. Hence the output from this procedure is still low-dimensional (similar to the original dataset), but also privacy preserving because the sensitive attributes can not be directly inferred with high accuracies from the resultant data even if a model is newly trained.

6.2.3 Generalization of our technique. In the results, we showed that our technique generalizes well to two datasets from mobile food diaries with passive sensing from two different countries. For both datasets, we attempted two sensitive inference tasks paired with three application inferences. Hence, we believe the above combination of datsets, sensitive inferences, and application inferences reasonably show the generalization potential of our technique. Moreover, it should also be noted that we were able to obtain similar results for other application inferences such as fruit vs. no-fruit and cereal vs. non-cereal too for both datasets, when used with both sensitive inferences gender and BMI category. However, the results are not included in the paper due to space limitations.

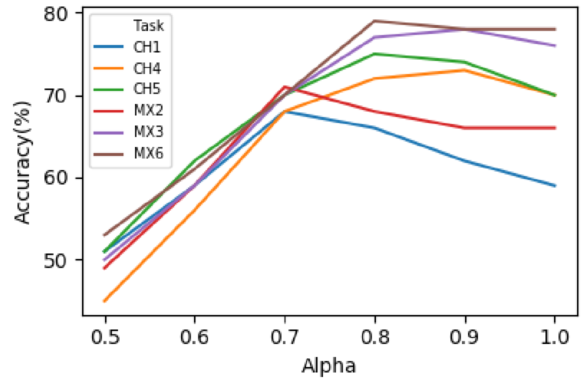

Figure 3: MT-NN After AE Inference Accuracy for Sensitive Inferences in 6 Different Tasks.

6.3 Inference Results for Different α Values in the Loss Function

In the results presented in prior sections, in Equation 1, we used α=0.5 to make sure the accuracy for sensitive inference is minimum in a binary classification task. However, since this is a tunable parameter, this value could vary from 0.5. Figure 3 shows results from the MT-NN after using the AE for six example tasks (these tasks are randomly chosen, we carried out experiments for all 12 tasks, and results are similar) from both datasets. Results show that by increasing the value of α, the accuracy of the sensitive inferences increases. Importantly, for all these inferences, application inference accuracies were kept above application inference accuracies gained with α=0.5 for all tasks. In addition, for all 12 tasks, for α<0.5, we obtained sensitive inference accuracies in the range 35%-55%, and application inference accuracies of the range 75%-87% using MT-NN. Hence, these results show that the α value in Equation 1 can be used as a tunable parameter in order to control the sensitive inference accuracy.

7 DISCUSSION

Using Feature Transformation Techniques on High Dimensional vs. Low Dimensional Data.

If we just consider the dimensionality of raw data traces, the higher the number and diversity of features in the data, the higher the potential amount of information available in the dataset, thus increasing the ability of discriminating sensitive attributes. On the other hand, low-dimensional or low-resolution datasets are already

Table 4: CH-Dataset: Accuracy for Application Inferences vs. Gender Inference and Application Inferences vs. BMI Category Inference using MT-NN and RF, before and after feature transformation using the AE. Results use C+A feature group

	Application Inference and Gender Inference						Application Inference and BMI Category Inference				
Task	Classification	MT-NN Before AE	MT-NN After AE	RF Before AE	RF After AE	Task	Classification	MT-NN Before AE	MT-NN After AE	RF Before AE	RF After AE
CH1	Meal vs. Snack	86%	81%	86%	85%	CH2	Meal vs. Snack	85%	84%	86%	82%
	Men vs. Women	67%	51%	77%	48%		High BMI vs. Low BMI	71%	48%	75%	53%
CH3	Sweet vs. Non-Sweet	83%	79%	82%	81%	CH4	Sweet vs. Non-Sweet	82%	79%	82%	80%
	Men vs. Women	69%	53%	77%	48%		High BMI vs. Low BMI	73%	45%	75%	52%
CH5	Dairy vs. Non-Dairy	78%	78%	73%	71%	CH6	Dairy vs. Non-Dairy	77%	76%	73%	72%
	Men vs. Women	76%	51%	77%	57%		High BMI vs. Low BMI	78%	54%	75%	44%

Table 5: MX-Dataset: Accuracy for Application Inferences vs. Gender Inference and Application Inferences vs. BMI Category Inference using MT-NN and RF, before and after feature transformation using the AE. Results use C+A feature group

	Application Inference and Gender Inference						Application Inference and BMI Category Inference				
Task	Classification	MT-NN Before AE	MT-NN After AE	RF Before AE	RF After AE	Task	Classification	MT-NN Before AE	MT-NN After AE	RF Before AE	RF After AE
MX1	Meal vs. Snack	81%	78%	83%	79%	MX2	Meal vs. Snack	82%	79%	83%	79%
	Men vs. Women	77%	53%	76%	51%		High BMI vs. Low BMI	72%	49%	77%	54%
MX3	Fatty vs. Non-Fatty	80%	78%	82%	79%	MX4	Fatty vs. Non-Fatty	81%	80%	82%	81%
	Men vs. Women	79%	52%	76%	59%		High BMI vs. Low BMI	78%	51%	77%	60%
MX5	Meat vs. Non-Meat	84%	81%	85%	82%	MX6	Meat vs. Non-Meat	82%	78%	85%	79%
	Men vs. Women	79%	53%	76%	56%		High BMI vs. Low BMI	79%	53%	77%	59%

processed in some way, reducing the information embedded in them. For example, the step count of a person is derived by processing high-resolution accelerometer and gyroscope data where many features (x,y,z axis of accelerometer and gyroscope, time) are combined to derive one single value i.e. the step count in a particular time window. Because step counts are low-resolution, it is comparatively difficult to engineer more features by processing them with different techniques. Therefore, from our findings, we advocate the idea that preserving sensitive attributes from high-dimensional or high-resolution datasets might have some limitation if novel discriminative features can still be generated. On the other hand, preserving sensitive attributes from low-dimensional or low-resolution data might mitigate the privacy risk discussed here to a larger extent. Researchers and developers who use mobile sensing datasets should be aware of these findings, specially when they store or share data with other parties.

Data Before and After Feature Transformation. The feature transformation process proposed here makes significant changes to dataset features after transformation. One such change is the conversion of categorical variables to numerical variables. For example, during an experiment, the dataset had two values each for the categorical variables "with_family", "with_friends" and "with_date" before the transformation, and after the transformation resulted in numerical values. This is because the feature transformation happens to each data row separately, and not to each column separately, unaware of the categorical nature of the dataset. Hence, the dataset after feature transformation would be uninterpretable unless the party using the transformed data had prior knowledge of the feature transformation process. This naturally protects the dataset from privacy risks from third parties who may gain access to the transformed data. For example, if a transformed dataset was shared with a third party by the data owner together with instructions regarding useful application inferences, it would be difficult for the third party to interpret data for other purposes. As another example, if the data was stored after feature transformation (i.e., in processed form) by the data owner, even if the data fell in the hands of a third party through hacking or a data breach, since the data was only interpretable for the original data owners, the dataset would become of less use for the third party. In other words, the technique we propose would create uninterpretable datasets for sharing and storage, increasing the likelihood that datasets are used only for required purposes, and not for anything else.

Dataset Diversity. A limitation of our study is the relative homogeneity of the participants who volunteered in the CH and MX datasets. The dataset used is from university students of two countries, hence, even though the participants are diverse in terms of eating routines, ethnicities, and behaviors, they are homogeneous in terms of age and occupation. While the results show evidence of sensitive inference using food diary entries, and that a feature transformation technique can preserve privacy, we believe that conducting a larger scale experiment more countries with people having different behavioral habits, ages, professions would shed more light into the results we present here. We hypothesize that even though using more diverse user populations might demonstrate varieties of eating behaviors, the technique we have proposed might still be useful.

Personalization, Privacy, and Utility. As researchers, we usually strive to enhance utility of applications and algorithms, and often use personalisation as a tool to increase utility. While this is important, an increasing body of work has also emphasized the importance of privacy preservation and the use of less sensitive data [16, 20, 33, 55, 58, 59]. Personalization and privacy preservation are at the two opposite ends of the spectrum because personalisation has typically required more personal data to provide high utility, while privacy preservation aims at providing reasonable utility from the application, while preserving privacy of users from known risks. The trade-off between these goals are also reflected among people who value different aspects while using mobile health applications, and online applications in general. Hence, it should be understood that while some users might prefer to distribute their personal information and health related information for personalized services, there are other users who have concerns regarding application developers, and also regarding how their personal data would be used if they provided such information. As seen from the results in Section 6.2, application inference utility slightly drops when privacy is preserved (after feature transformation). While we understand that personalisation of algorithms and services is an important research direction, we endorse the idea that app users, app developers, and data owners should be aware of the risks they might face when sharing and storing personal information from foreseen and unforeseen circumstances. We believe that designing ubicomp technology for joint privacy and utility, and not only for personalisation, is important for the advancement of the field in a progressive and ethical manner. Recent literature further discusses why new privacy preservation techniques are needed by pointing out that simple anonymization techniques are no longer enough to preserve user privacy [16].

Future Research Directions. We identify several topics for future research. As highlighted previously in this section, it is worth examining how eating behaviors or eating routines in general vary across different countries and diverse user groups. Moreover, it is also important to analyze whether there are application inferences that would not allow to transform features such that application inference accuracy is high, and sensitive inference accuracy remains low. Such situations might occur in cases where both application inference related tasks and sensitive inference tasks highly depend on the same features. Hence, it is worth experimenting novel privacy preserving techniques for such scenarios. In addition, our work builds upon prior work that proposes the idea that mobile app users can distinguish between trusted and non-trusted applications. However, the validity of this idea should be examined with a user study in the future.

8 CONCLUSION

In this paper, we examined how to provide both privacy and utility in mobile food diary applications that generate low-dimensional data consisting of sparse mobile sensor data and self-reports of eating behavior. Using two datasets involving 160 people from two countries, we first demonstrated that behavioral features around eating events can be used to infer sensitive attributes like gender and BMI-category. After demonstrating the scale of the potential privacy risk, we show how deep learning techniques based on autoencoders and multi-task neural networks can be leveraged to process dataset features such that application inferences achieve accuracies of around 75%-90%, while sensitive inference accuracies drop to around 50%. We show that the technique generalizes well on two datasets, and two sensitive inferences and three application inferences for each dataset. We emphasized how prior work on applying sensitive inference preservation techniques on high-dimensional and high-resolution data might not be useful if more low-dimensional, sparse datasets could still be generated from such data. We also highlighted the need to think about both application inference utility and sensitive attribute protection in addition to personalisation when creating, storing, and sharing even low-dimensional datasets derived from mobile sensing based food diaries. We believe that thinking along this line would help create more privacy-preserving mobile health applications, and also would increase awareness in the ubicomp community regarding privacy risks in other mobile health and well-being applications.

ACKNOWLEDGMENTS

This work was funded by the European Union's Horizon 2020 WeNet project, under grant agreement 823783. D. Gatica-Perez acknowledges the former support of EPFL Integrative Food and Nutrition Center, through the Understanding Eating Routines in Context project.

REFERENCES

[1] 2016. *APPFAIL: Threats to Consumers in Mobile Apps.* Retrieved Nov 14, 2020 from https://fil.forbrukerradet.no/wp-content/uploads/2016/03/Appfail-Report-2016.pdf
[2] 2018. *The Guardian view on big data and insurance: knowing too much.* Retrieved Feb 13, 2020 from https://www.theguardian.com/commentisfree/2018/sep/27/the-guardian-view-on-big-data-and-insurance-knowing-too-much
[3] 2018. *Health Insurers Are Vacuuming Up Details About You — And It Could Raise Your Rates.* Retrieved Feb 13, 2020 from https://www.npr.org/sections/health-shots/2018/07/17/629441555/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates?t=1581596827223
[4] 2019. *FitBit Privacy Policy.* Retrieved Nov 11, 2019 from https://www.fitbit.com/eu/legal/privacy-policy
[5] 2019. *Lose It!* Retrieved Nov 06, 2019 from https://www.loseit.com/
[6] 2019. *MyFitnessPal.* Retrieved Nov 06, 2019 from https://www.myfitnesspal.com/
[7] 2019. *OUT OF CONTROL: How consumers are exploited by the online advertising industry.* Retrieved Nov 11, 2020 from https://www.forbrukerradet.no/undersokelse/no-undersokelsekategori/report-out-of-control/
[8] 2020. *Google Fit - Coaching you to a healthier and more active life.* Retrieved February 12, 2020 from https://www.google.com/fit/
[9] 2020. *HCI Guidelines for Gender Equity and Inclusivity.* Retrieved October 15, 2020 from https://www.morgan-klaus.com/gender-guidelines.html
[10] 2020. *A more personal Health app. For a more informed you.* Retrieved February 12, 2020 from https://www.apple.com/ios/health/
[11] 2020. *S Health Terms of Use.* Retrieved Feb 13, 2020 from https://account.samsung.com/membership/etc/specialTC.do?fileName=shealth.html
[12] 2020. *Samsung Health App.* Retrieved February 12, 2020 from https://www.samsung.com/us/support/owners/app/samsung-health
[13] Ibrahim Mousa Al-Zuabi, Assef Jafar, and Kadan Aljoumaa. 2019. Predicting customer's gender and age depending on mobile phone data. *Journal of Big Data* 6, 1 (19 Feb 2019), 18.
[14] O. Amft and G. Troster. 2009. On-Body Sensing Solutions for Automatic Dietary Monitoring. *IEEE Pervasive Computing* 8, 2 (April 2009), 62–70.
[15] 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.
[16] B. Baron and M. Musolesi. 2020. Interpretable Machine Learning for Privacy-Preserving Pervasive Systems. *IEEE Pervasive Computing* (2020), 1–10.
[17] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. 2017. EarBit: Using Wearable Sensors to Detect Eating Episodes in Unconstrained Environments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 37 (Sept. 2017), 20 pages.
[18] France Bellisle. 2014. Meals and snacking, diet quality and energy balance. *Physiology & Behavior* 134 (2014), 38 – 43. Eating Patterns, Diet Quality and Energy Balance.
[19] Jennifer Berry. 2019. *Is dairy good or bad for your health?* Retrieved Jan 22, 2020 from https://www.medicalnewstoday.com/articles/326269.php
[20] Alex Beutel, Jilin Chen, Zhe Zhao, and Ed H. Chi. 2017. Data Decisions and Theoretical Implications when Adversarially Learning Fair Representations. *CoRR*

abs/1707.00075 (2017). arXiv:1707.00075 http://arxiv.org/abs/1707.00075

[21] 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.

[22] Joseph Bonneau and Sören Preibusch. 2010. The Privacy Jungle:On the Market for Data Protection in Social Networks. In *Economics of Information Security and Privacy*, Tyler Moore, David Pym, and Christos Ioannidis (Eds.). Springer US, Boston, MA, 121–167.

[23] George A. Bray and Barry M. Popkin. 2014. Dietary Sugar and Body Weight: Have We Reached a Crisis in the Epidemic of Obesity and Diabetes? *Diabetes Care* 37, 4 (2014), 950–956.

[24] Sanja Brdar, Dubravko Ćulibrk, Vladimir Crnojević, and Trg Dositeja Obradovića. [n.d.]. Demographic Attributes Prediction on the Real-World Mobile Data.

[25] Liangliang Cao, Mert Dikmen, Yun Fu, and Thomas S. Huang. 2008. Gender Recognition from Body. In *Proceedings of the 16th ACM International Conference on Multimedia* (Vancouver, British Columbia, Canada) *(MM '08)*. Association for Computing Machinery, New York, NY, USA, 725–728.

[26] Rich Caruana. 1997. Multitask Learning. *Machine Learning* 28, 1 (01 Jul 1997), 41–75.

[27] Didier Chapelot. 2010. The Role of Snacking in Energy Balance: a Biobehavioral Approach. *The Journal of Nutrition* 141, 1 (12 2010), 158–162.

[28] Ramnath K. Chellappa and Raymond G. Sin. 2002. Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. *Information Technology and Management* 6 (2002), 181–202.

[29] Gillian Cleary. 2018. *Mobile Privacy: What Do Your Apps Know About You?* Retrieved Nov 06, 2019 from https://www.symantec.com/blogs/threat-intelligence/mobile-privacy-apps

[30] Rob Copeland. 2019. *Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans*. Retrieved Jan 21, 2020 from https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790

[31] Cees de Graaf, F. Matthew Kramer, Herbert L. Meiselman, Larry L. Lesher, Carol Baker-Fulco, Edward S. Hirsch, and John Warber. 2005. Food acceptability in field studies with US army men and women: relationship with food intake and food choice after repeated exposures. *Appetite* 44, 1 (2005), 23 – 31.

[32] District, Hemamalini, Silambuselvi, and Ramya Devi. 2017. Relationship Between Eating Attitude and Body Mass Index Among Adolescents in Selected College , Kancheepuram. (2017).

[33] Yanai Elazar and Yoav Goldberg. 2018. Adversarial Removal of Demographic Attributes from Text Data. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Brussels, Belgium, 11–21.

[34] Alex Elliott-Green, Lirije Hyseni, Ffion Lloyd-Williams, Helen Bromley, and Simon Capewell. 2016. Sugar-sweetened beverages coverage in the British media: an analysis of public health advocacy versus pro-industry messaging. *BMJ Open* 6, 7 (2016).

[35] Lynnette R. Ferguson. 2010. Meat and cancer. *Meat Science* 84, 2 (2010), 308 – 313. Special Issue: 55th International Congress of Meat Science and Technology (55th ICoMST), 16-21 August 2009, Copenhagen, Denmark.

[36] Daniel Gatica-Perez, Joan-Isaac Biel, David Labbe, and Nathalie Martin. 2019. Discovering eating routines in context with a smartphone app. In *UbiComp/ISWC Adjunct*.

[37] Henner Gimpel, Christian Regal, and Marco Schmidt. 2015. myStress: Unobtrusive Smartphone-Based Stress Detection. In *ECIS*.

[38] C. M. Goldstein, J. G. Thomas, R. R. Wing, and D. S. Bond. 2017. Successful weight loss maintainers use health-tracking smartphone applications more than a nationally representative sample: comparison of the National Weight Control Registry to Pew Tracking for Health. *Obesity Science & Practice* 3, 2 (2017), 117–126.

[39] Robert P Heaney and Donald K Layman. 2008. Amount and type of protein influences bone health. *The American Journal of Clinical Nutrition* 87, 5 (05 2008), 1567S–1570S.

[40] Julie M Hess, Satya S Jonnalagadda, and Joanne L Slavin. 2016. What Is a Snack, Why Do We Snack, and How Can We Choose Better Snacks? A Review of the Definitions of Snacking, Motivations to Snack, Contributions to Dietary Intake, and Recommendations for Improvement. *Advances in nutrition* 7 3 (2016), 466–75.

[41] Lena Holzer. 2018. *Reporton third gender markeror no gender marker options*. Retrieved Nov 11, 2019 from https://www.ilga-europe.org/sites/default/files/non-binary_gender_registration_models_in_europe_0.pdf

[42] A. Jain and V. Kanhangad. 2016. Investigating gender recognition in smartphones using accelerometer and gyroscope sensor readings. In *2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT)*. 597–602.

[43] Fanyu Kong and Jindong Tan. 2012. DietCam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing* 8, 1 (2012), 147 – 163.

[44] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5802–5805.

[45] D. Kotz, C. A. Gunter, S. Kumar, and J. P. Weiner. 2016. Privacy and Security in Mobile Health: A Research Agenda. *Computer* 49, 6 (June 2016), 22–30.

[46] Mark A. Kramer. 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal* 37, 2 (1991), 233–243.

[47] N. D. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, and A. Campbell. 2011. Bewell: A smartphone application to monitor, model and promote wellbeing.. In *Proceedings of the Conference PervasiveHealth (PervasiveHealth '11)*. http://pac.cs.cornell.edu/pubs/PervasiveHealth_BeWell.pdf

[48] Rebecca M. Leech, Anthony Worsley, Anna Timperio, and Sarah A. McNaughton. 2015. Understanding meal patterns: definitions, methodology and impact on nutrient intake and diet quality. *Nutrition Research Reviews* 28, 1 (2015), 1–21.

[49] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu. 2008. Gait Components and Their Application to Gender Recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 38, 2 (March 2008), 145–155.

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

[51] Christos Louizos, Kevin Swersky, Yujia Li, Max Welling, and Richard Zemel. 2015. The Variational Fair Autoencoder. arXiv:stat.ML/1511.00830

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

[53] Mohammad Malekzadeh, Richard G. Clegg, Andrea Cavallaro, and Hamed Haddadi. 2018. Protecting Sensory Data Against Sensitive Inferences. In *Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems* (Porto, Portugal) *(W-P2DS'18)*. ACM, New York, NY, USA, Article 2, 6 pages.

[54] Mohammad Malekzadeh, Richard G. Clegg, Andrea Cavallaro, and Hamed Haddadi. 2019. Mobile Sensor Data Anonymization. In *Proceedings of the International Conference on Internet of Things Design and Implementation* (Montreal, Quebec, Canada) *(IoTDI '19)*. ACM, New York, NY, USA, 49–58.

[55] Mohammad Malekzadeh, Richard G. Clegg, and Hamed Haddadi. 2017. Replacement AutoEncoder: A Privacy-Preserving Algorithm for Sensory Data Analysis. *CoRR* abs/1710.06564 (2017). arXiv:1710.06564 http://arxiv.org/abs/1710.06564

[56] Vijini Mallawaarachchi, Lakmal Meegahapola, Roshan Madhushanka, Eranga Heshan, Dulani Meedeniya, and Sampath Jayarathna. 2020. Change Detection and Notification of Web Pages: A Survey. *ACM Comput. Surv.* 53, 1, Article 15 (Feb. 2020), 35 pages.

[57] Roberta Masella and Walter Malorni. 2017. Gender-related differences in dietary habits. *Clinical Management Issues* 11, 2 (2017).

[58] L. Meegahapola, N. Athaide, K. Jayarajah, S. Xiang, and A. Misra. 2019. Inferring Accurate Bus Trajectories from Noisy Estimated Arrival Time Records. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. 4517–4524.

[59] Lakmal Meegahapola, Thivya Kandappu, Kasthuri Jayarajah, Leman Akoglu, Shili Xiang, and Archan Misra. 2019. BuScope: Fusing Individual & Aggregated Mobility Behavior for "Live" Smart City Services. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services* (Seoul, Republic of Korea) *(MobiSys '19)*. Association for Computing Machinery, New York, NY, USA, 41–53.

[60] Lakmal Meegahapola, Salvador Ruiz-Correa, and Daniel Gatica-Perez. 2020. Alone or with Others? Understanding Eating Episodes of College Students with Mobile Sensing. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia*. Association for Computing Machinery, New York, NY, USA.

[61] Mark Mirtchouk, Christopher A. Merck, and Samantha Kleinberg. 2016. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. In *UbiComp*.

[62] Toan Nguyen, Aditi Roy, and Nasir D. Memon. 2018. Kid on The Phone! Toward Automatic Detection of Children on Mobile Devices. *CoRR* abs/1808.01680 (2018). arXiv:1808.01680 http://arxiv.org/abs/1808.01680

[63] University of Groningen. 2020. *Sensitive data and medical confidentiality*. Retrieved Jan 21, 2020 from https://www.futurelearn.com/courses/protecting-health-data/0/steps/39608

[64] Lukasz Olejnik, Tran Minh-Dung, and Claude Castelluccia. 2013. Selling Off Privacy at Auction. (Dec. 2013). https://hal.inria.fr/hal-00915249 working paper or preprint.

[65] Fabrice Pierre, Sylviane Taché, Claude R. Petit, Roelof Van der Meer, and Denis E. Corpet. 2003. Meat and cancer: haemoglobin and haemin in a low-calcium diet promote colorectal carcinogenesis at the aberrant crypt stage in rats. *Carcinogenesis* 24, 10 (10 2003), 1683–1690.

[66] Arya Pourzanjani, Tom Quisel, and Luca Foschini. 2016. Adherent Use of Digital Health Trackers Is Associated with Weight Loss. *PLOS ONE* 11, 4 (04 2016), 1–14.

[67] Mollie Powles. 2018. *Personalization Versus Privacy: Making Sense of the Privacy Paradox*. Retrieved Jan 21, 2020 from https://blog.hubspot.com/marketing/personalization-versus-privacy

[68] privacyrights. 2017. Mobile Health and Fitness Apps: What Are the Privacy Risks? https://pdfs.semanticscholar.org/c52c/67541657fd71022771edaed75148999b3c00.pdf

[69] Mashfiqui Rabbi, Min Hane Aung, Mi Zhang, and Tanzeem Choudhury. 2015. MyBehavior: Automatic Personalized Health Feedback from User Behaviors and Preferences Using Smartphones. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 707–718.

[70] 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* (Copenhagen, Denmark) *(UbiComp '10)*. ACM, New York, NY, USA, 281–290.

[71] Christopher Riederer, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, and Pablo Rodriguez. 2011. For Sale: Your Data: By: You. In *Proceedings of the 10th ACM Workshop on Hot Topics in Networks* (Cambridge, Massachusetts) *(HotNets-X)*. Article Article 13, 6 pages.

[72] Kathryn M Ross and Rena R Wing. 2016. Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: A randomized pilot study. *Obesity* 24 8 (2016), 1653–9.

[73] 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* (Anchorage, AK, USA) *(KDD '19)*. ACM, New York, NY, USA, 2260–2268.

[74] Krushnapriya Sahoo, Bishnupriya Sahoo, Ashok Choudhury, Nighat Sofi, Raman Kumar, and Ajeet Bhadoria. 2015. Childhood obesity: causes and consequences. *Journal of Family Medicine and Primary Care* 4 (04 2015), 187–92.

[75] A. Sano and R. W. Picard. 2013. Stress Recognition Using Wearable Sensors and Mobile Phones. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 671–676.

[76] Ana Santamaria, Isaac Vazquez, Dolores Caballero, and Concepcion Fernandez-Rodriguez. 2009. Eating habits and attitudes and their relationship with Body Mass Index (BMI). *European Journal of Psychiatry - EUR J PSYCHIAT* 23 (10 2009).

[77] 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.

[78] Sougata Sen, Vigneshwaran Subbaraju, Archan Misra, Rajesh Balan, and Youngki Lee. 2020. Annapurna: An automated smartwatch-based eating detection and food journaling system. *Pervasive and Mobile Computing* 68 (2020), 101259.

[79] 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.

[80] Caifeng Shan. 2012. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters* 33, 4 (2012), 431 – 437. Intelligent Multimedia Interactivity.

[81] Courtney C. Simpson and Suzanne E. Mazzeo. 2017. Calorie counting and fitness tracking technology: Associations with eating disorder symptomatology. *Eating Behaviors* 26 (2017), 89 – 92.

[82] Vaclav Smil. 2014. Eating meat: Constants and changes. *Global Food Security* 3, 2 (2014), 67 – 71.

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

[84] Edison Thomaz, Irfan A. Essa, and Gregory D. Abowd. 2015. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. *Proceedings of the ... ACM International Conference on Ubiquitous Computing . UbiComp* 2015 (2015), 1029–1040.

[85] Edison Thomaz, Aman Parnami, Irfan Essa, and Gregory D. Abowd. 2013. Feasibility of Identifying Eating Moments from First-person Images Leveraging Human Computation. In *Proceedings of the 4th International SenseCam & Pervasive Imaging Conference* (San Diego, California, USA) *(SenseCam '13)*. ACM, New York, NY, USA, 26–33.

[86] Eran Toch, Yuhuai Wang, and Lorrie Faith Cranor. 2011. Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction* 22 (2011), 203–220.

[87] Britta N. Torgrimson and Christopher T. Minson. 2005. Sex and gender: what is the difference? *Journal of Applied Physiology* 99, 3 (2005), 785–787.

[88] European Union. 2019. *Reporton third gender markeror no gender marker options*. Retrieved Nov 13, 2019 from https://eugdpr.org/

[89] Tim Van hamme, Giuseppe Garofalo, Enrique Argones Rúa, Davy Preuveneers, and Wouter Joosen. 2019. A Systematic Comparison of Age and Gender Prediction on IMU Sensor-Based Gait Traces. *Sensors* 19, 13 (2019). https://www.mdpi.com/1424-8220/19/13/2945

[90] Eugene Volokh. 2000. Personalization and Privacy. *Commun. ACM* 43, 8 (Aug. 2000), 84–88.

[91] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. ACM, New York, NY, USA, 3–14.

[92] Richard J. Wurtman and Judith J. Wurtman. 1995. Brain Serotonin, Carbohydrate-Craving, Obesity and Depression. *Obesity Research* 3, S4 (1995), 477S–480S.

[93] Najat Yahia, Alice Achkar, Abbass Abdallah, and Sandra Rizk. 2008. Eating habits and obesity among Lebanese university students. *Nutrition Journal* 7, 1 (2008), 32.

[94] Lydia Zepeda and David Deal. 2008. Think before you eat: photographic food diaries as intervention tools to change dietary decision making and attitudes. *International Journal of Consumer Studies* 32, 6 (2008), 692–698.

[95] Sean Zheng. 2015. *WHAT ARE GENDER STATISTICS*. Retrieved Nov 14, 2019 from https://unstats.un.org/unsd/genderstatmanual/What-are-gender-stats.ashx

[96] Erheng Zhong, Ben Tan, Kaixiang Mo, and Qiang Yang. 2013. User demographics prediction based on mobile data. *Pervasive and Mobile Computing* 9, 6 (2013), 823 – 837. Mobile Data Challenge.