Privacy Threats from Seemingly Innocuous Sensors

CUCS-002-20

Shirish Singh
Columbia University
New York, NY
shirish@cs.columbia.edu

Anthony Saieva
Columbia University
New York, NY
ant@cs.columbia.edu

Gail Kaiser
Columbia University
New York, NY
kaiser@cs.columbia.edu

ABSTRACT

Smartphones incorporate a plethora of diverse and powerful sensors that enhance user experience. Two such sensors are the accelerometer and gyroscope, which measure acceleration in all three spatial dimensions and rotation along the three axes of the smartphone, respectively. These sensors are often used by gaming and fitness apps. Unlike other sensors deemed to carry sensitive user data, such as GPS, camera and microphone, the accelerometer and gyroscope do not require user permission on Android to transmit data to apps. This paper presents our IRB-approved studies showing that the accelerometer and gyroscope gather sufficient data to quickly infer the user's gender. We started with 33 in-person participants, with 88% accuracy, and followed up with 259 on-line participants to show the effectiveness of our technique. Our unobtrusive *ShakyHands*[1] technique draws on these sensors to deduce additional demographic attributes that might be considered sensitive information, notably pregnancy. We have implemented *ShakyHands* for Android as an app, available from Google Play store, and as a Javascript browser web-app for Android and iOS smartphones. We show that even a low-skilled attacker, without expertise in signal processing or deep learning, can succeed at inferring demographic information such as gender and pregnancy. Our approach does not require tampering with the victim's device or specialized hardware; all our study participants used their own phones.

CCS CONCEPTS

• **Security and privacy** → **Mobile and wireless security**; **Privacy protections**; **Usability in security and privacy**; • **Human-centered computing** → *Smartphones.*

KEYWORDS

Privacy, Side channels, Mobile sensors, Demographic Inference

1 INTRODUCTION

Most mobile devices, including smartphones, tablets, and smartwatches, come equipped with built-in sensors such as accelerometer, gyroscope, and magnetometer. A study suggests that accelerometers are the most widely used sensor accessed by mobile apps [12]. Accelerometer and gyroscope sensors measure acceleration and rotation forces caused by the movements and vibrations of an object. When the smartphone goes from a standstill to any velocity, the accelerometer is designed to respond to the vibrations associated with such movements. Similarly, when the smartphone rotates about any of the smartphone axes, the gyroscope captures that change.

[1]In this paper, shaky hands does not refer to essential tremor disorder or any other disorder.

The axes of acceleration and rotation for a smartphone are shown in Figure 1.

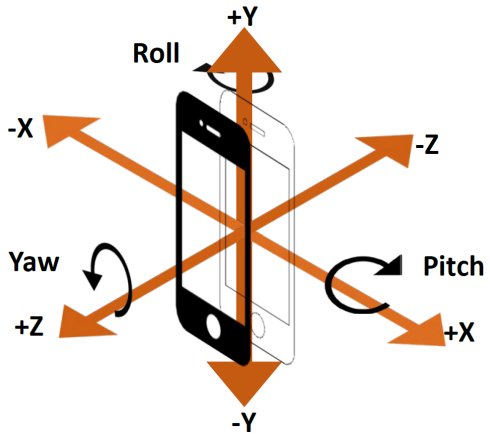

Figure 1: The x, y, and z-axis represent the horizontal axis along the face of the phone, the vertical axis along the face, and the axis along the perpendicular plane passing through the center of the phone, respectively. Figure from [42].

The privacy implications of inferences from smartphone sensors have previously been reviewed in [54]. For example, [83] showed how to infer the user's gender from their hip movements, gait, and activity patterns while walking for 5-10 minutes with the phone in their pocket. Instead, we observe different patterns in accelerometer and gyroscope data in male and female participants when they hold their phones. We collect accelerometer and gyroscope sensor data from the user's smartphone while the user is reading or browsing the internet, which enables detecting the user's gender within thirty seconds with reasonably high confidence. This difference in the holding pattern between the user groups stems from user's hand stability, which is attributed to physiologic tremor in hands. Tremor is an involuntary, rhythmic muscle contraction leading to shaking movements in one or more parts of the body. We leverage physiologic tremor in hands to infer the demographic information of smartphone users. Physiologic tremor occurs in all healthy individuals. It is rarely visible to the eye and typically involves a fine shaking of both of the hands and also the fingers. It is not considered a disease but is a normal human phenomenon that is the result of physical properties in the body (for example, rhythmical activities such as heart beat and muscle activation) [6]. Studies have investigated the impact of physiologic tremor on hand stability [33]. As computer scientists, we are not qualified to analyze the physiological differences in genders or pregnant women.

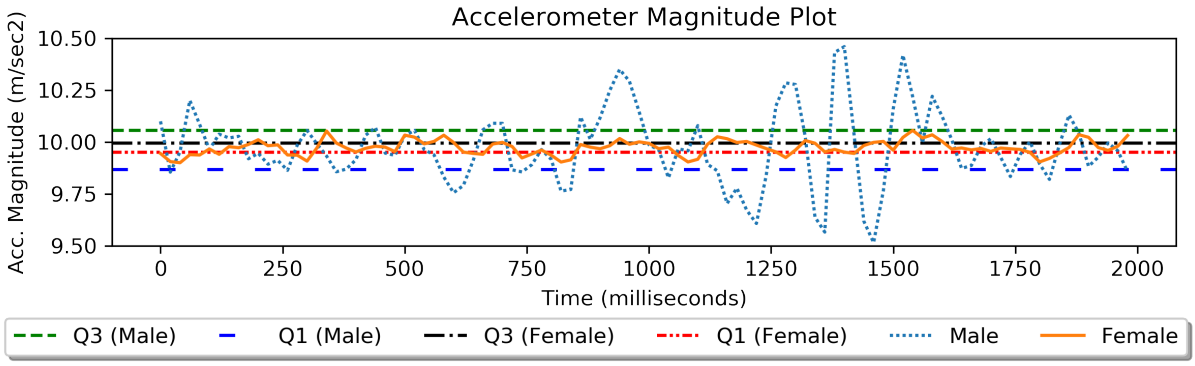

Figure 2: The graph shows two seconds of accelerometer magnitude time-series data of randomly selected male and female participants. Q1 and Q3 represent first and third quartile respectively. The quartile plots show lower spread in female's sensor readings, which implies lesser fluctuations or tremors while holding the phone.

Experts in ergonomics and neurology have, however, previously published about physiologic tremor and hand stability differences due to gender [34, 45] and pregnancy [5].

Since these holding patterns are subliminal in nature, it is difficult to replicate these patterns consciously. The difference becomes evident in the plot of accelerometer sensor readings corresponding to both genders. Figures 2 shows the differences in the accelerometer readings of randomly selected male and female participants from our study. The subtle differences in the phone holding characteristics of the two genders can be observed in the plot. The graph shows that the female participant's hand is more stable as compared to that of the male participant. These variations in sensor readings contribute to differences in the phone holding pattern in the genders. In this paper we study how these patterns can be used to predict gender and pregnancy state of the smartphone user.

We conducted 33 in-person and 259 online experiments with smartphone users to demonstrate the effectiveness of our proposed approach. The experimental results show that our technique can achieve accuracy up to 88% on gender and 86% on pregnancy prediction tasks.

1.1 Contributions

We evaluate ShakyHands in terms of both run-time performance and accuracy. In this paper, we answer three research questions:

- **RQ1**: Can we derive demographic information of the user by looking at only accelerometer and gyroscope data streams?
- **RQ2**: If yes, then which demographic attributes can we extract with high accuracy?
- **RQ3**: Is this a realistic threat? In particular, how much interaction time is needed before making an inference?

The primary contributions of this paper are four-fold:

- We demonstrate that smartphone sensor readings which, at the time of writing, do not require user permission and are available to arbitrary apps, can be used for near-real-time inference of demographic information about the user.
- In particular, we show that holding the phone for as little as 5 seconds can leak private information about the user, such as gender and pregnancy.
- We provide evidence that instituting permission requirements for apps to access these sensors may be ineffective.

To our best knowledge, this is the first work on predicting end users' demographics in near real-time (In the order of seconds) which does not require cooperation from the user.

If our paper is accepted, we will provide the Android app, JS based web-app, open-source implementation of *ShakyHands*, model training scripts, and anonymized dataset. The rest of the paper is organized as follows: Section 2 provides background for our research and threat model followed by Section 3 which discusses related research efforts to infer demographic information. Section 4 presents data acquisition methodology and study design. Section 5 provides detailed description of the proposed algorithm for demographic inference tasks. Section 6 presents evaluation scheme and experimental results. Section 7 discusses observations from our study. Section 8 describes the limitations of our approach. Section 9 describes our privacy scoring scheme and mitigation strategies for sensor based attacks. Finally, conclusion is presented in Section 10 followed by future work.

2 BACKGROUND

As smartphones become a ubiquitous part of society, our data becomes intertwined with smartphone tech [30]. Such data has become a valuable commodity in today's marketplace, with even user's demographic information holding significant value. In response, users have taken many measures to protect their privacy.

In a 2012 study [20], researchers found that 20% of the Facebook users did not reveal their gender, and another study in 2014 [63] showed that in the popular online game World of Warcraft 23% of the male users used female avatars and 7% of the female used male avatars. While there are many reasons for users to give fake information about gender, with privacy concerns being one of them, users maintain a right to know and control how much information about them is shared on the net. Nevertheless, unrestricted access to related user information like a list of installed apps allows service providers to identify users who entered fake information disregarding the user's wish for anonymity. Even something as seemingly benign as a person's gender can quickly become sensitive, e.g. targeted advertising, political or otherwise [14, 26, 43, 50].

As such, the Cambridge Analytica scandal has demonstrated how digital technology, such as weaponized social media, can efficiently micro-target voters [16].

According to a 2014 study [56], gender significantly impacts the ways in which political candidates communicate to voters and political communication experts have paid increasing attention to gender differences in political advertising. Holman et al. [43] studied the impact of political advertisements on genders, and found that candidates of either gender can use these ads to affect women's votes. Therefore, it seems reasonable to treat gender as sensitive.

User education remains a major problem with respect to smartphone related privacy concerns. Researchers observed that technically skilled and financially independent users risked potential privacy intrusions despite their awareness of potential risks [15]. Since the general public uses smartphones, many users do not fully realize the amount of information revealed by granting access to various smartphone hardware nor the level of permissions most apps retain. Furthermore, when app-specific knowledge is used in conjunction with orthogonal data sets available for purchase, it is nearly impossible to grasp the breadth of user information data consumers can infer.

Data collection related to smartphone usage is not limited to apps and service providers as might be initially thought, but also to general web activity. Javascript APIs [81] allow traditional browser activity to access smartphone sensor information creating another opportunity for privacy invasion. Researchers found that of the top 100,000 sites-as ranked by Amazon-owned analytics company Alexa - 3,695 incorporate scripts that tap into one or more of these accessible mobile sensors. That includes plenty of big names, including Wayfair, Priceline.com, and Kayak [69]. Another study of the top 2200 free apps collected from 28 categories on Google Play store, reported that 719 apps used sensor-related APIs [13]. Out of these 719 apps, 610 (84.8%) apps used accelerometer, and 107 (14.9%) apps used gyroscope sensors.

2.1 Threat Model and Existing Protections

In recent years, data collection and sale, also called data brokerage, has become a viable profit model for numerous companies and individuals [31, 57, 65]. Data consumers leverage this information to inform investment decisions, target advertising choices, and manipulate user behavior all at the expense of user privacy. The primary defense mechanisms protecting users against privacy invasions are the privacy policies enforced by the App Distribution System (ADS), e.g. Google Play Store [39] and Apple's App Store [10]. Such policies require that if an app is going to record sensitive data from the user (per the policy, this includes things like gender, personal preferences, location information, and health habits), the user must be explicitly informed about the data collection [40, 46]. Gender information is particularly valuable for targeted advertising, political or otherwise [14, 43]. While most advertising is mundane, political advertising advantages represent significant risk to election integrity as demonstrated by social media based propaganda campaigns in the 2016 US presidential election [35]. Furthermore a candidates ability to appeal to specific gender demographics can win or lose an election for a hopeful politician [23, 86]. As such, any invasion of privacy that reveals the gender of users and creates opportunity for the attacker to sway the user's political opinions on a large scale presents an opportunity to influence elections which would have far reaching political and societal impacts.

In a 2016 experiment, researchers found that 74% of people who joined a fictitious social network skipped reading the privacy policy. Furthermore, those who opened the terms and conditions must not have read them very carefully-because all of them agreed to give up their firstborn child to the social network [9]. In another study [64], researchers estimated that it would take an average individual 154 hours to skim the privacy policies for the approximately 1,462 websites they encountered each year. The study concluded that in terms of wages and lost time, that would amount to $2,226 per person. Moreover, it is essential to note that the privacy policies allow app developers to record most sensor data without user permission. In particular, the privacy policies permit the developer to record the accelerometer and gyroscope data without the user ever knowing. Additionally, ADS's generally provide basic defenses, such as Google Play Protect [8], against malware and spyware to protect users from the most malicious and invasive attacks. The smartphone's OS re-enforces the privacy policies by providing additional app-specific access control mechanisms for particularly sensitive hardware like the microphone and camera as well as sensitive user data like contacts and photos. No such access control mechanisms are provided for the gyroscope and accelerometer on Android [3] (and were apparently not provided on iOS until version 13 [2]).

We assume that the attacker wants to covertly collect marketable user information without the user knowing this data is being collected, but is still limited by the ADS' privacy policies and malware inspection as well as the phone's permission restrictions. The attacker circumvents these controls to obtain the marketable user demographic information by leveraging the data leakage from the built-in motion sensors of a smartphone. Any environment that lets the attacker run code with access to the phone's gyroscope and accelerometer for 30 seconds or longer and allows the attacker to send the collected data across the network functions as an attack vector. We show that even a low-skilled attacker, without expertise in signal processing, feature extraction or deep learning, can succeed at these classification tasks. Because deep learning algorithms and frameworks have a steep learning curve (many hyper-parameters and network architecture options) and require an advanced skill-set (such as defining the layers, activation functions, regularization, etc.) to train a model, the attacker uses off-the-shelf signal processing libraries and machine learning algorithms to extract relevant features and train models for predictions, respectively. For instance, Random Forest can be used without turning the hyper-parameters. On the other hand, a neural network requires the definition of the network architecture, hidden layers, activation function, regularization, etc. to ensure that it can produce any meaningful results [41, 73]. We demonstrate our proof of concept by collecting data while users read a piece of text, but other possible scenarios include collecting data while users wait for a gaming app to load, or while a user is typing in a messaging app. These heavyweight attack vectors like a maliciously developed app or website present obvious yet realistic threats [88]. However, more subtle approaches present a more severe threat. Most apps and websites run unmonitored 3rd-party ads, and since the attack requires so little overhead and leverages a unprotected resource, ads accessing sensor data

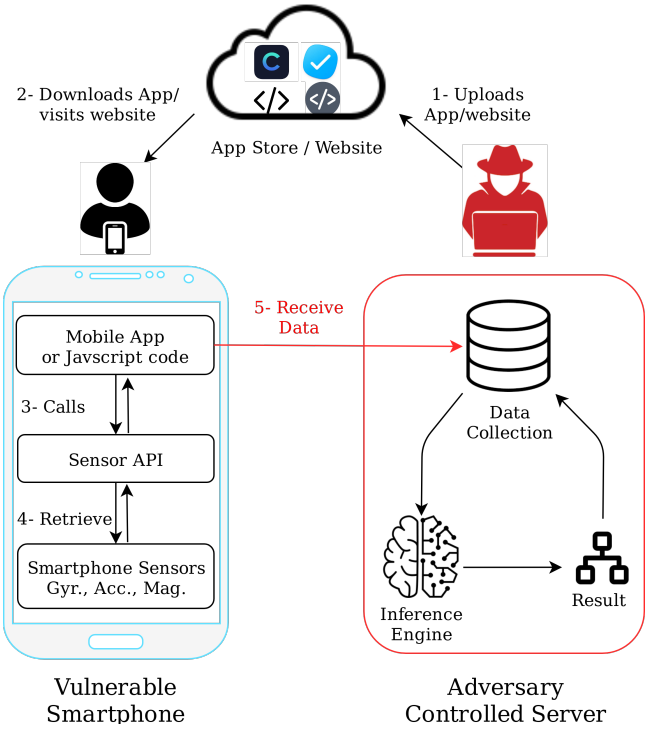

Figure 3: The adversary uploads an app or hosts a website. The user opens the adversarial app or visits an adversarial website on the phone's browser. The installed app or web-app accesses the accelerometer and gyroscope sensors through an API call and sends the data to the adversary. The adversary can then infer the demographics of the user in real-time. The app implementation could alternatively be self-contained and able to survive intermittent network connectivity, by running the interference engine locally and later sending only the results to the adversary.

via the Generic Sensor API [81] function as an omnipresent and omnipotent attack vector. Since most apps and (obviously) web pages require network connections to run, the malicious code in an ad will be able to connect to a remote server under the control of the attacker. Since most ad distribution services provide highly targeted deployments, an attacker has almost complete control over where and how the attack is deployed. The threat model is shown in figure 3. In our proof of concept, we only examine accelerometer and gyroscope sensor data collected via an Android and web app for demographics inference.

3 RELATED WORK

Researchers have studied the privacy implications of the accelerometer sensor [54]. Prior studies have been conducted to infer the behavior, location, health parameters, and demographics of the smartphone user by leveraging the accelerometer sensor alone. Most of the existing approaches for sensor-based demographic inference in mobile phones are based on gait analysis, installed apps, browser history, and call logs. Other approaches include the processing of visual or audio signals. In this section, we first discuss the related work based on smartphone data, then we look at inference techniques using sensors.

3.1 App/Browser Based Approaches

There are three popular techniques that can be applied on a smartphone to infer demographic information about the user. The first group of researchers used an application installed on the devices to infer the demographic information about the user. The second group investigates the browser history to infer the demographic attributes of the user. The third group looks at the call logs over long periods of time to make inferences.

3.1.1 *Installed App.*

Suranga et al. presented how third-party apps can leverage information about installed applications on the smartphone to predict the gender of the user [76]. Their work shows that the app installation patterns between male and female users can be used to train the linear SVM classifier for the prediction task. They achieved an accuracy of around 70%. In a study, researchers used a list of installed applications on the device, used at least once within a period of one month, to train the logistic regression model for the task of predicting gender, age, race, marital status, children count, and income [62]. They reported an accuracy of 82.3% for gender detection. These techniques of gender inference are effective to an extent. They cannot determine if the user of the device has been changed temporarily, which can happen often in a household [77].

3.1.2 *Browser History.*

The user's web browsing history has been studied extensively for demographic inference. Previous studies show that there is correlation between users' browsing behavior and their demographic attributes. Phuong et al. presented browser history-based gender detection techniques by leveraging the websites visited and the time of each visit over a period of one month [72]. They treat the problem as a binary classification problem and apply the SVM algorithm for the task. Their technique achieved a macro-average F1 score of 0.805. Hu et al. [44] used data on page clicks from a major website to predict the age and gender of the users through a Bayesian framework. They achieved 79.7% on gender and 60.3% on age in terms of Macro F1 score over a dataset collected over one week. Goel et al., paired web history over one year of 250K users (57% female and 43% male) with user-level demographics [38]. They reported accuracy of 80% and 76% for age and gender, respectively. In a similar study, Kalimeri et al. studied the relationship between demographics and multi-modal digital data from web browsing behavior and smartphone usage collected for one month [49]. They reported accuracy of 90% for gender and 71% for age.

3.1.3 *Call Logs.*

Felbo et al. used anonymized call detail records over fifteen weeks to predict the age and gender of the users [36]. They report an accuracy of 63.1% for age and 79.7% for gender prediction. In another study, Ying et al. proposed a multi-level classification model to predict gender, age, marital status, and job using call logs, location, and environmental features collected over a year [87]. They used MDC data set [52, 55] consisting of 200 participants and reported an accuracy of 85.47% and 77.77% for gender and age, respectively. Dong, et al. used more than 7 million users' call and SMS data,

collected over a period of one year, to predict the gender and age of the user [28]. Their dataset comprised of 45% female and 55% male participants. They recorded the performance to be 80% and 73% for gender and age detection, respectively.

The above-mentioned methods of demographic inference require system permissions and historic data from other applications for predictions. Each of these methods is intrusive since they access the user's private information of browser history, call logs, and applications installed on their device. Furthermore, these techniques require substantial data from the device over a long period of time. Hence, they are computationally ineffective.

3.2 Sensor Based Approaches

In the past, there have been studies to infer the demographics of the user through sensor data. We summarize work on two popular demographic attributes studied extensively: age and gender.

3.2.1 Age.

Researchers have investigated the relationship between the smoothness of walking and the age of the participants by measuring the acceleration patterns at the head and pelvis [66]. Researchers studied the variation in step length, step time, and velocity of the participants and observed that the magnitude of accelerations at the head and pelvis were generally smaller in older subjects. Davarci et al. studied the critical observation that the characteristics of children and adults differ in hand-holding and touching the smartphones [24]. In the study, participants were asked to use an app and perform simulated screen taps to gather data and train their model. They report an accuracy of 92.5% over 100 adults and 100 children participants. Study of age is outside the scope of our study.

3.2.2 Gender.

SH Cho et al. studied the gender-based differences in gait [21]. Gait analysis data were obtained with the optoelectronic system and force plates. They observed that gait analysis data had significant gender differences attributed to anatomy and habits. In another work, Gary Weiss and Jeffrey Lockhart studied the relationship between walking pattern and soft-biometric traits, gender, height, and weight, based on the walking pattern of the participants [83]. The volunteers were asked to walk, with an Android phone in their pocket, for approximately 5-10 minutes. They obtain an accuracy of 71% over 70 participants. In another study, researchers presented a gender recognition technique based on behavioral biometrics [47]. Researchers investigate gender recognition using gait data acquired from the inbuilt accelerometer and gyroscope sensors of a smartphone. They used Multi-level Local Pattern, and Local Binary Pattern features to train SVM and aggregate bootstrapping classifiers to achieve an accuracy of 76.83% over 42 participants.

Other methods of side channel attack on accelerometer and gyroscope for private information has been studied in the past [11, 67, 89].

The techniques mentioned above, though efficient, require either sophisticated data collection mechanism and/or the data to be collected over long intervals of time. Hence, these techniques are not always practical. In the next section, we discuss some techniques which leverage the smartphone data to make demographic predictions. In this work, we present a novel real-time technique of demographic inference, *ShankyHands*, which is time efficient and unobtrusive.

We have also experimented with implementing a browser version of *ShankyHands*, which uses JavaScript API calls to get the sensor readings of the phone. This method of inference requires continuous internet connectivity as the JavaScript transfers the data to the adversarial server for making an inference. However, such inference can be made in real-time in the mobile app, and only the result needs to be sent to the adversary, thus, saving significant network bandwidth. In this paper, all evaluation is done for Android app and JavaScript-based web app.

4 STUDY DESIGN

In this section, we describe our process of collecting labeled accelerometer and gyroscope data, which is used for generating and evaluating classification models. Prior to data collection, we obtained approval from our institution's IRB (Institutional Review Board), since we were "experimenting" on human subjects. All subjects provided informed consent before participating in our study and were also asked to fill out a demographics questionnaire. We had used IRB approved deception while collecting the data. Participants were not briefed about the nature of the study or the purpose of data collection to avoid bias. The participants were asked to answer several demographic questions, some of which were used in our study and some of which were not (e.g. height). Some people started the study but did not complete it, these are not included as participants in our data. As a consequence, there were 292 participants for gender and pregnancy prediction tasks. The details associated with the collected data set are provided in Section 4.2 and Table 1. We conducted the study in two phases. In the first phase, we recruited 33 participants (17 male and 16 female) for an in-person study. These participants belonged to the undergraduate and graduate population of our university. In this phase, we collected age, gender, and hand preference of the participants. In the second phase, we recruited 127 on-line participants to collect data using the Android app. We deployed the web application to collect data from an additional 132 on-line users using iPhone and Android devices. In both the phases, participants were asked to use their own smartphones to complete the study and were given clear instructions on how to perform the task.

4.1 Data Collection Apps

In order to collect data, we developed an Android app and a web-app. We designed the apps to collect data from the sensors at a frequency of 50 Hertz (one sample from all sensors after every 20 milliseconds). In our study, a combination of accelerometer and gyroscope sensors demonstrated good results on the demographics inference tasks. We recruited 145 male and 147 female participants in our study. Every participant was asked to read a short passage on their Android smartphone while holding the phone in either one hand or both hands, according to their preference. We asked the participants to be comfortable in standing posture without using any arm support.

Both apps consist of a set of three tasks; a demographic survey to collect the ground truth, a reading task, and a quiz. The apps collect accelerometer and gyroscope sensor data to capture the subtle hand

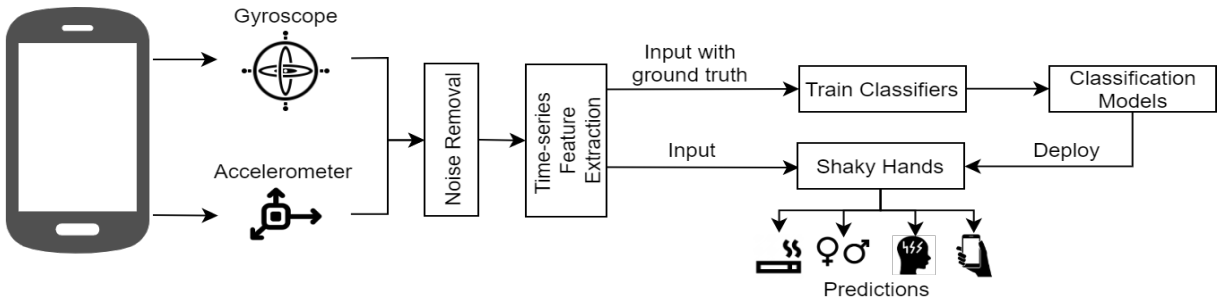

Figure 4: Overview of *ShakyHands*. The accelerometer and gyroscope sensor stream of the smartphone is used to train the classifiers. The models are then deployed to infer demographic information of the user in real-time.

movements of participants. The collected data allow us to infer the hand vibration and rotation along the three axes of the smartphone.

4.1.1 Demographic Survey. In the demographic survey, participants were asked about their demographic attributes, which include: age, gender, preferred phone holding hand, height, race, relationship status, education level, income, caffeine consumption, smoking habit, nicotine consumption, stress level, and pregnancy. The data provided by the participants and the demographic filters provided by online recruitment platform acted as ground truth for our classifiers. We were able to observe the apparent gender of the in-person participants.

4.1.2 Reading Task. After completing the demographic survey, participants were asked to read a short passage of 355 words while holding their smartphone in their preferred hand(s). Because people spend less than 72 seconds with an app at a time on an average [17], we designed our study to reflect a similar app use scenario. On average, male and female participants spent 86 and 81 seconds to read the passage, respectively.

4.1.3 Quiz. After the reading task, participants were asked three multiple-choice questions from the passage. The first question asked the participants to select the title of the passage. In the second question, participants were asked the summary of the passage, which was epitomized in the last sentence of the passage. The last question asked about the general idea of the passage. The first two questions were factual, and the last question was intuitive. We used the response to the first two questions of the quiz as a filter to weed out participants who did not read the passage carefully. The intuition is that if a participant had not read the passage carefully, they might not have read the instruction either.

4.2 Participant Recruitment

For data collection, we recruited adult participants who speak English either as a first or second language. We curated the data based on the response to the quiz and the availability of the accelerometer and gyroscope sensors on participants' phones. We would consider the data only if the participant answered first two factual questions correctly. Some participants' smartphones did not have a gyroscope senor; we disregard all such participants for our study. After curating, the results reported in this paper are based on data from 292 participants; 145 male and 147 female. Table 1 summarizes the characteristics of our dataset. Participants were spread out geographically, although most stayed in the United States or Europe. On average, male and female participants spent 86 and 81 seconds to read the passage, respectively. We advertised on Prolific [2], a well-known online crowd-sourcing service specifically designed for the scientific community [70], helping us to reach its growing number of users. Prolific boasts more naive, diverse, and less dishonest participants as compared to other online research platforms [71]. In addition, Prolific allows the researchers to get in touch with the study participants to verify the information if needed.

Parameter	Results
Age	Mean: 34 years
	Standard Deviation: 11
	Minimum: 18 years
	Maximum: 76 years
Gender	Male: 145 (50%)
	Female: 104 (35.27%)
	Pregnant: 43 (14.73%)
Smoking Habits & Nicotine Consumption	Smokers: 47
	Non-Smokers: 212
	Not Available: 33
Preferred Hand	Right Hand: 157
	Left-Hand: 80
	Both-Hands: 55
Stress Level	High Stress: 66
	Medium Stress: 98
	Low Stress: 95
	Not Available: 33

Table 1: Summary of demographics of study participants. We had 292 participants who participated in two separate data collection rounds.

5 *SHAKYHANDS* IMPLEMENTATION

In this section, we describe the implementation of *ShakyHands*. There are two steps to process the sensor data: Noise Removal and Feature Extraction. *ShakyHands* infers demographic information by leveraging patterns in the sensor readings extracted from the phone. We train two models 1) *Gender Inference*, and 2) *Pregnancy Detection*. Figure 4 shows the data pipeline used to process the raw sensor signals and train our classifiers.

[2]Prolific (www.prolific.co)

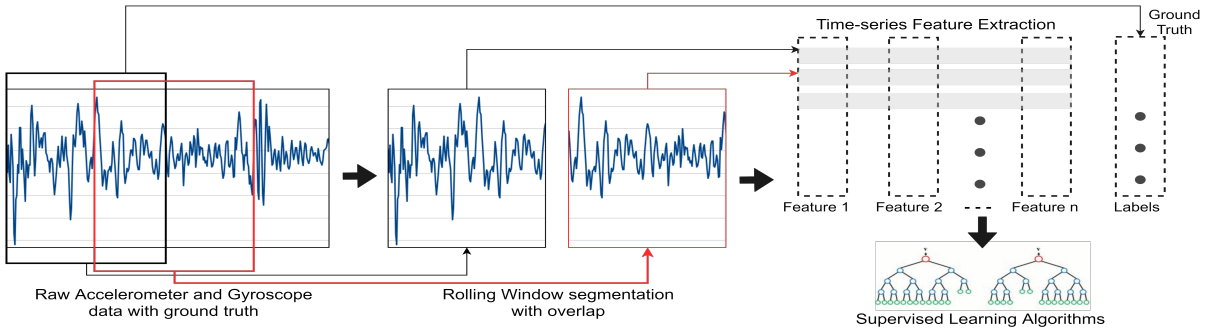

Figure 5: Supervised learning pipeline: The sensor data stream is segmented into windows of constant width and 50% overlap. Each window is then used to extract time-series features and train classifiers for predictions.

5.1 Noise Removal

The dataset acquired from smartphones will be subjected to various sensor noises. It is crucial to choose the right set of features from these noisy datasets that uniquely describes the behavior for the success of the data-driven algorithms. To remove data points that do not represent the reading task (Noise) we apply upper and lower threshold limits on accelerometer magnitude readings. Essentially, the optimal gravitational force should be 9.8 m/sec^2. The subtle movements of the participants' hand(s) lead to further fluctuation in this value. Hence, we trained our model to obtain an optimum range of threshold to remove such noise from the data. We observed that the accelerometer magnitude reading for our study task is between 8.0 m/sec^2 and 12.0 m/sec^2. These threshold limits only consider micro-fluctuations in values and do not overlap with any other activity that a user might perform on the smartphone. This threshold range has been observed in the data collected through the study. Hence, we use this limit to remove noise from the dataset.

5.2 Feature Extraction

We gather raw X, Y, and Z axes readings from each sensor. In addition to the raw data, we computed the magnitude of accelerometer and gyroscope from the raw sensor data. As mentioned in section 2.1, we assume that the attacker does not possess deep knowledge of signal processing or feature extraction, so they use an off-the-shelf signal processing library to extract relevant features. For each processed data value, we applied a feature extractor functions from the Tsfresh library [22] to extract all simple statistical time-series features which do not require additional parameters. These features are straight forward to use because they only rely on the time series window. Figure 5 shows how time-series data is segregated into windows and features are extracted from each window to construct the dataset. After extracting the time series features, we apply Sequential Forward Floating Selection (SFFS) algorithm [74] using the MLxtend library [75] to select the best features for each task. Our dataset consists of features comprising of 44 time-series metrics calculated from the two processed sensor data streams. For our study, we only use the accelerometer and gyroscope magnitude because we want to study the stability of the hand, independent of the orientation of the phone.

Let $X^{m \times n}$ be the multi-modal time series data comprised of m sensor streams and n attributes. We derive statistical features from the time series data over a rolling window of size d with an overlap of (ϕ). We use the feature extraction algorithm outlined in Algorithm 1 to derive the features from $X^{m \times n}$ to train our classifiers.

Algorithm 1: Time-Series Sensor Data Feature Extraction

Input : $X^{m \times n}$: n sensor streams obtained during m time instances, $t_1,\ t_2\ \ldots\ t_m$

Output: Extracted Features: $X^{r \times \lambda(n)}$: r is instances of processed data over window size W and $\lambda(n)$ is time-series features

Define : W (sliding window size), ϕ (overlap window), D (Sensor stream at a time instance), Acc_l (Lower threshold of accelerometer magnitude), Acc_h (Higher threshold of accelerometer magnitude)

Function `Noise_Removal`($X^{m \times n}$)**:**
- **foreach** $D \in X^{m \times n}$ **do**
  - **if** $D > Acc_h$ & $D \leq Acc_l$ **then**
    - Remove D from $X^{m \times n}$
- **return** $X^{\hat{m} \times n}$

Function `Feature_Extraction`($X^{\hat{m} \times n}$)**:**
- Segment $\hat{m}$ samples into r segments consisting of W samples each and ϕ overlap window
- **foreach** *segment* $s \in r$ **do**
  - compute $\lambda(n) = f_1,\ f_2\ \ldots\ f_j$
- **return** $X^{r \times \lambda(n)}$

If our paper is accepted, we will provide the Android app, web-app, open-source implementation of *ShakyHands*, model training scripts, and anonymized dataset. The web-app method of inference requires continuous internet connectivity as the JavaScript transfers the data to the adversarial server and offloads computation to server for making an inference. However, the skill level required to implement this technique is minimal as compared to developing an app, which requires registration on app stores and a steep learning curve. App can make inferences in real-time within app itself, and only the result needs to be sent to the adversary, thus, saving significant network bandwidth. In the next section, we discuss the mitigation strategies to obviate the privacy leak from *ShakyHands*.

6 EVALUATION

As part of our comprehensive evaluation of our classifiers, we used two different evaluation schemes: Stratified 10-fold cross-validation and Leave-One-Out Cross-Validation (LOOCV). Both schemes are similar in terms of the algorithm. In stratified k-fold cross-validation, the folds are formed such that they contain approximately the same proportions of labels as the original data-set. The model is trained on $k-1$ folds and tested on the k^{th} fold. LOOCV is a variant of k-fold cross-validation more suitable for small datasets. It uses a single observation from the original sample as the validation data, and the remaining observations as the training data. We selected these evaluation schemes because LOOCV is almost unbiased, as observed in [60], whereas 10-fold cross-validation has a lower variance [48]. However, we show only one or the other for each prediction task for space reasons. For Gender prediction, since we have more than hundred participants, we use both 10-fold cross validation and LOOCV. For pregnancy prediction task, we only use LOOCV. Moreover, our primary goal is to identify the features of one individual and hence, we have used LOOCV to classify the demographic features in both the tasks.

As mentioned in section 2.1, we assume that the attacker does not possess the expertise of deep learning algorithms and frameworks and uses off-the-shelf machine learning algorithms to train models for predictions. Since the Random Forests (RF) algorithm [18], is touted as one of the best "off-the-shelf" algorithms for classification available [25], we used Random Forest (RF) for both classification tasks. In addition to RF, we used K-Nearest Neighbour (KNN), SVM, Naive Bayes, Logistic Regression, and decision tree classifiers for all classification tasks. Because of space limitations we only report the performance of RF and KNN classifiers for evaluation to demonstrate the threat model. We ran our scripts on a Dell XPS 8930, with 9^{th} generation Intel Core i5-9600K 6-core processor and 32GB RAM, running Ubuntu 18.04 and Python 3.7.3.

6.1 Metrics

To evaluate the performance of the model, we use four metrics, namely, accuracy, precision, recall, and F1 score. Each one of them is defined below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

where TP represents true positive samples, TN represents true negatives, FP is the number of false-positive predictions, and FN is false-negative predictions.

6.2 ShakyHands

In this section, we present the results of models for two demographic detection tasks, namely, gender and pregnancy. We studied several other demographic attributes to demonstrate the effectiveness of *ShakyHands*. For space limitations, we only present the results of gender and pregnancy.

6.2.1 *Gender Detection.*

In the first phase, we recruited 33 participants (17 male and 16 female) for an in-person study. Participants belonged to the undergraduate and graduate population of our university. For this data set, we used LOOCV and achieved an accuracy of 88% and 85% using RF and KNN algorithms, respectively. Table 2 shows the results of each classifier. For the in-person study, our RF model achieved precision, recall, and F1 score of 0.93, 0.82, and 0.87, respectively. These results motivated us to perform a large scale study and observe if the same results hold true for larger populations across different demographics.

n = 33	**Predicted** Male	**Predicted** Female	
Actual Male	14 / 13	3 / 4	17
Actual Female	1 / 1	15 / 15	16

Table 2: Confusion matrix for gender classification task using LOOCV for in-person study. The values in red and blue represent the classification performed by random forest classifier and k-nearest neighbors algorithm, respectively.

In the second phase, for gender detection tasks we consider data from all 292 participants from in-person study and Prolific. First, we evaluate the performance of RF and KNN algorithms using LOOCV and 10 fold cross validation. We achieved 77% overall accuracy on 292 participants using LOOCV. We had recorded an accuracy of 79.31% for male and 74.83% for female participants. Then we applied stratified 10-fold CV to evaluate our technique. Our model achieved accuracy, precision, recall, and F1 score of 75.31%, 0.74, 0.79, and 0.76, respectively. The ROC curve of cross-validation is shown in Figure 6. The Area Under the Curve (AUC) is 0.75. Table 3 shows the results for both RF and KNN classifiers using 10-fold CV.

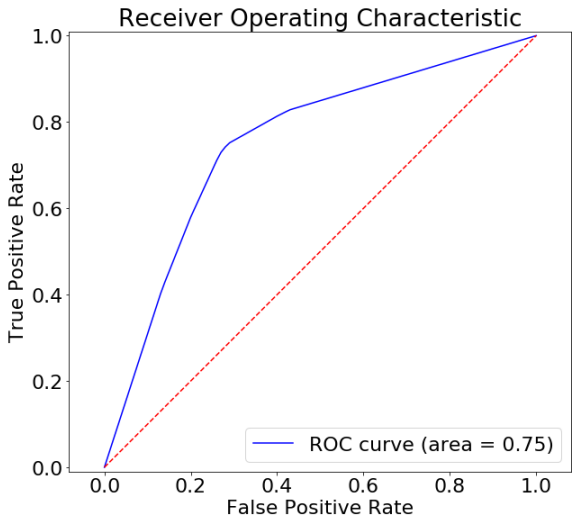

Figure 6: ROC curve of 10-fold cross validation.

n = 292	**Predicted** Male	**Predicted** Female	
Actual Male	115 / 117	32 / 30	145
Actual Female	37 / 58	110 / 89	147

Table 3: Confusion matrix for gender classification task using 10-fold Cross validation. The values in red and blue represent the classification performed by random forest classifier and k-nearest neighbors algorithm, respectively.

We also studied the minimum user interaction time required to make high confidence inference of gender. Figure 7 shows how the model accuracy increases over time. Time presented in x-axis represents the time spent by the participants on the app. The figure demonstrates how the model's accuracy varies depending on the time spent by the participants on the app. Our model predicted the claimed gender within 15 seconds of user holding the phone with an accuracy of 72%, within 30 seconds with an accuracy of 74%, and within 60 seconds with an accuracy of 76%. We note that these results are lower than for our in-person study. We suspect that some on-line participants did not answer the gender demographic question truthfully. This suspicion is coherent with the study [63] on online games. Various studies have suggested that participants in on-line crowd-sourcing websites do not necessarily represent their demographic information accurately [19, 27, 29, 79, 80, 84].

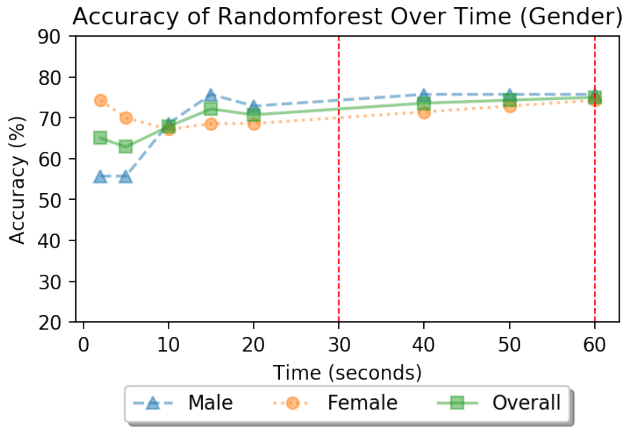

Figure 7: Gender classification accuracy increases over time.

We applied t-SNE dimentionality reduction [61] on the gender data-set to reduce the feature space to two features. To plot the graph, we used the points utilized by RF algorithm to classify the participants. Figure 8 shows clear distinction between genders.

6.2.2 Pregnancy Detection.

There have been numerous studies to detect pregnancies using smartphone [1, 78]. However, these techniques require external components to be added to the smartphone. In our study, we identify if a woman holding the phone is pregnant or not by using the motion and orientation sensors to measure the physiologic tremors. Increased physiologic tremor may occur in pregnancy caused by agents that cause an increase in adrenaline [5]. This increase in tremor is caused most often by increasing sympathetic nervous activity, which is a common phenomenon during pregnancy [68]. The

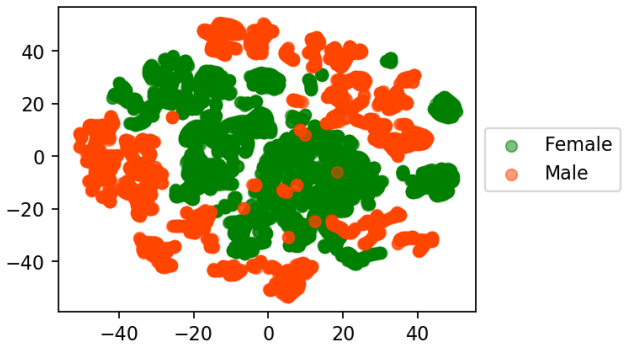

Figure 8: TSNE visualization shows distinction between male and female data points. Female data points are clustered in the center, and data points representing male participants lie along the edges.

sympathetic nerves are the system of nerves excited by adrenaline released from adrenal glands [4].

Our study had 43 pregnant participants, with ages ranging from 18 to 42 years, who were due to give birth in time ranging from 1 month to 9 months. We performed under-sampling of the non-pregnant females to construct a balanced dataset, consisting of 43 pregnant vs 43 randomly selected non-pregnant females, for supervised training and avoid the imbalanced classification problem [58]. All selected participants belonged to the same age group as pregnant women. Since we had less than 100 participants, we used LOOCV for evaluating our model. Table 4 shows the performance of our models. Figure 10 demonstrates the distinction between the data points representing both classes. Our model achieved accuracy, precision, recall, and F1 score of 86%, 0.92, 0.79, and 0.85, respectively using random forest classifier.

n = 86	**Predicted** Pregnant	**Predicted** Not Pregnant	
Actual Pregnant	34 / 32	9 / 11	43
Actual N. Pregnant	3 / 11	40 / 32	43

Table 4: Confusion matrix for pregnancy detection task using LOOCV. The values in red and blue represent the classification performed by random forest classifier and k-nearest neighbors algorithm, respectively.

Studying the minimum user interaction time required to make high confidence inference of pregnancy, we observed that it requires significantly less time than gender detection. Figure 9 shows how the model accuracy increases over time. Time presented in x-axis represents the time spent by participants on the app. The figure demonstrates how the model's accuracy varies depending on the time spent by the participants on the app. Our model predicted the pregnancy state within 5 seconds of user holding the phone with an accuracy of 83%. That figure remains constant throughout 60 secs, however, the accuracy of detecting non pregnant participants increases over time.

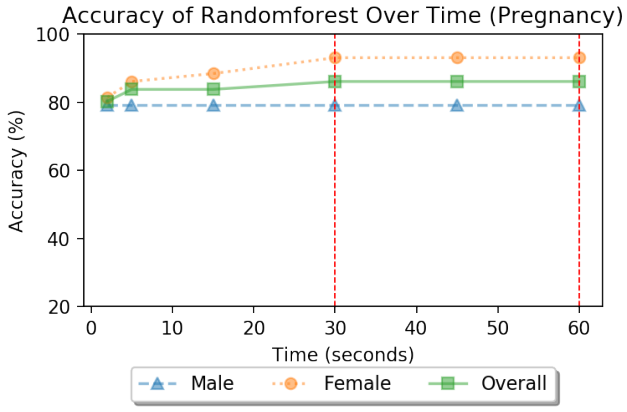

Figure 9: Pregnancy prediction accuracy increases over time. However, number of pregnant people detected remain constant throughout time.

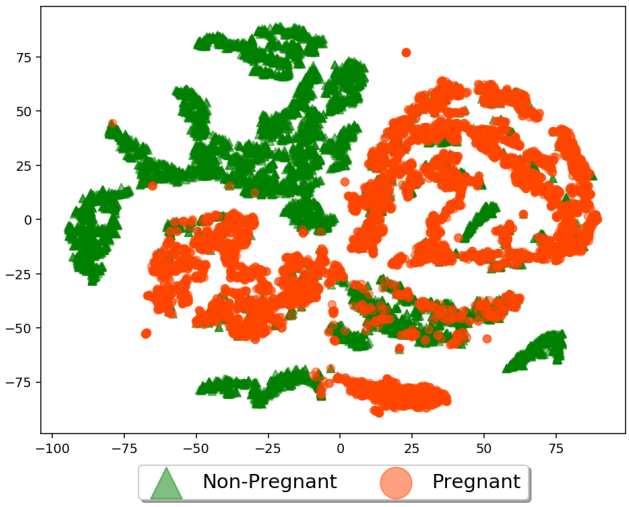

Figure 10: t-SNE visualization of data points of 43 pregnant and 43 non-pregnant females show class distinction.

7 DISCUSSION

In this section we first propose a scoring scheme to quantify the invasiveness of each side-channel attack approach presented in Section 3, then we discuss each research question.

7.1 Privacy Scoring Scheme

In Section 3 we reviewed related research that use different data sources to infer the user's demographic information. We propose a three-level scheme to quantify the intrusive nature of a data source.

(1) **High**: Users' private data, such as their browsing history, that can provide the content of items as well as their metadata. Accessing these sensitive data sources requires user permission. Even though the browser itself records this information, it does not share this information with web-apps.
(2) **Moderate**: Data sources that access semi-private information about the user, such as the metadata from call logs, emails, and messages, without accessing the direct content of individual items. To access moderate level sources, an app also has to acquire explicit user permission.
(3) **Low**: These data sources include sensors like the accelerometer, gyroscope, and magnetometer, which cannot directly provide users' private or semi-private information. These sensors can be easily accessed by any app on the device without user consent. However, as we show, they can be used to infer sensitive information about the user.

RQ1: Can we derive demographic information of the user by looking at only accelerometer and gyroscope data streams?

We demonstrate that using *ShakyHands*, it is possible to infer demographic information of the smartphone user in real-time by leveraging differences in physiologic tremor. Through supervised learning, our technique can achieve prediction accuracy of up to 77% about the user's gender and 86% accuracy for pregnancy prediction. We also show that our method of inference does not require any explicit permission from the user on Android devices. Table 5 summarizes the related work and the performance of each work. In comparison to existing works, *ShakyHands* performed better in the gender prediction task in two areas. First, *ShakyHands* achieved the highest accuracy in the sensor category. Second, *ShakyHands* used the minimum amount of time to make a high confidence inference. In comparison to works in other categories, our technique does not access any sensitive information.

We observe that when using an accelerometer alone for gender classification and pregnancy prediction tasks, models perform poor than the models leveraging both accelerometer and gyroscope sensors. Table 6 summarizes our findings regarding sensor contributions. Through this study we conclude that physiologic tremor not causes vibrations which are measured through the accelerometer, but also rotations which are measured through the gyroscope.

Classifier	Gender Classification	Pregnancy Prediction
Acc.	67.47% / 65.41%	62.79% / 54.65%
Acc.+Gry.	77.05% / 70.54%	86.05% / 74.41%

Table 6: Table shows the accuracy of our models. We studied the impact of each sensor on different tasks.

RQ1: We observe that it is indeed possible to infer demographic information from the accelerometer and gyroscope readings.

RQ2: If yes, then which demographic attributes can we extract with high accuracy?

In this paper we demonstrated that the gender and pregnancy prediction tasks can be performed with high accuracy of 77% and 86%, respectively. Other potential characteristics that influence physiologic tremors include smoking habits, stress level, alcohol consumption, and substance consumption [6]. We intend to study these features in the future work.

RQ2: *ShakyHands* can predict the gender and pregnancy of the user with high confidence.

RQ3: Is this a realistic threat? In particular, how much interaction time is needed before making an inference?

Type	Related Work	Summary	Participants	Reported Accuracy	Performance Period	Intrusion Level	Data / Code Availability
Browser	Phuong et. al [72]	Browser History	150K Users	Gen: 0.805 F1	1 mo	High	N/A
	Hu et al. [44]	Page Clicks	189K Users	Gen: 0.797 F1 Age: 0.603 F1	1 week	High	N/A
	Goel et. al, [38]	Browser History	250K Users	Gen:76% Age: 80%	1 year	High	N/A
	Kalimeri et. al [49]	Browser History	7633 Users	Gen: 90% Age: 71%	1 month	High	N/A
Call Logs	Felbo et. al [36]	Call Logs	150K Users	Gen: 79.7% Age: 63.1%	1 year	Moderate	N/A
	Ying et. al [87]	Call logs, location & env. features	200 Users	Gen: 85.47% Age: 77.77%	1 year	High	MDC Dataset
	Dong et. al [28]	Call Logs and SMS	7M Users	Gen: 80% Age: 73%	1 year	Moderate	N/A
Installed Apps	Seneviratne et. al [76]	Installed Apps	174 adults	Gen: 70%	Instant	Moderate	N/A
	Malmi et. al [62]	Installed Apps	3760 Users	Gen: 82.3%	1 month	Moderate	N/A
Sensor	Menz, et. al [66]	Gait Analysis	30 Young & 30 Older adults	N/A	N/A	Low	N/A
	Davarci, et. al [24]	Phone holding and touch/tap features	100 adults & 100 children	Age: 92.5%	N/A	Low	N/A
	Cho, et. al [21]	Barefoot gait analysis	98 adults	N/A	N/A	Low	N/A
	Weiss, et. al [83]	Hip movements	70 adults	Gen: 71%	5-10 mins	Low	Data
	Jain, et. al [47]	Gait analysis	42 adults	Gen: 76.83%	N/A	Low	N/A
	ShakyHands	**Phone Holding**	**292 adults**	**Gen: 77% Preg.: 86%**	**<1 minute**	**Low**	**Apps & Dataset**

Table 5: Related studies on demographic inference which reported their results in terms of accuracy or F1 score allows for a direct comparison with our results. For each work, we provide the intrusion level as per Section 7.1. "Gen" refers to Gender and "Preg" refers to pregnancy. *ShakyHands* performs the best in the sensor category and has significantly lower computational overhead and intrusion level. 'N/A' is stated for papers that had omitted this information.

Our simulations with different phone holding time suggest that the accuracy of individual prediction tasks increases over time. We achieved optimal results within one minute of the phone holding for gender detection task. For pregnancy prediction task, we observed that as little as 5 seconds is enough to achieve an accuracy of 83%. We observe that the performance of gender prediction model increases over time and the model achieves 76% accuracy within a minute. We show that *ShakyHands* has significantly lower overhead in terms of time and intrusion level required to make an inference.

RQ3: *ShakyHands* can infer gender and pregnancy within as little as 5 seconds.

8 THREATS TO VALIDITY

In this study, we demonstrate how to predict four demographic characteristics of a smartphone user by mining accelerometer and gyroscope sensor data in a time-efficient manner. In this section, we delineate the limitations of our approach.

(1) *User Activity*: In this study, we consider that the participant is standing. In reality, users might perform other activities, such as walking, while using their smartphone. When walking with the smartphone, the sensors do not capture the micro-movements of the hand, and we lose the vital information required to predict the gender of the user.

(2) *Arm Support*: Our approach assumes that the user is using their smartphone in a position that does not require them to take arm support. We observed that when the participant takes arm support, the pattern in the sensor data diminishes, leading to a higher number of mis-classifications. This case of mis-classification is particularly true for male users since their data exhibits higher variations on sensor readings, which are lost while taking arm support. Though we have limited success when the user rests their hands, monitoring the sensor readings over time will eventually get the correct classification.

(3) *Substance Consumption*: Our classifier does not consider the impact of food items containing caffeine. Researchers had observed that participants had higher variations in the hand movement (irrespective of gender) when they regularly consume caffeine (e.g., coffee) [6]. This case might be handled by training a separate classifier on caffeine consumption and studying the dependence of caffeine on physiologic tremor. Similar to caffeine, nicotine consumption (e.g., cigarettes) also impact physiologic tremors [32] and, consequently, the performance of the model. Earlier studies have investigated the relationship between nicotine and tremors [59]. There may be other commonly used substances that affect the results, such as recreational drugs. Research is being done to understand the dependency of drugs and medication on physiologic tremors [6]. These and other substances might have affected some of our participants, but are outside the scope of our study.

We also did not consider the health conditions of the participants or other environmental factors like temperature, which might also influence the results. We leave that as a future research problem.

9 POTENTIAL MITIGATION STRATEGIES

We considered some fairly simple potential mitigation strategies and conclude that these are not adequate, a more sophisticated solution is needed, which we hope to address in future work.

9.1 Security through Policy

Unlike other sensitive sensors of the smartphone, accelerometer and gyroscope do not require user permission on Android to transmit data to an app [53]. A malicious app can exploit this loophole to its advantage and gather sensor streams that could be used to infer sensitive information about the user, as in our *ShakyHands*.

Permission mechanisms serve as the main measure to protect users' privacy and security in Android apps. Modern smartphone operating systems prompt users to regulate permissions using the ask-on-first-use (AOFU) policy. Much research has been done to dynamically regulate permissions depending on user preferences and contexts [85]. However, one common limitation of all the existing techniques is that they heavily rely on users' historical decisions on granting permissions. By relying on historical data, they ignore the fact that users are not experts on privacy protection. Felt et al. [37] show that Android's AOFU policy is ineffective, since only 17% of the study participants read the requested permissions when installing an app, and only 3% were able to demonstrate an understanding of those permissions. Other similar studies [7], [82], [51] show that install-time prompts fail because users either do not understand or pay attention to them. Hence, adding permissions on the accelerometer and gyroscope sensors will not necessarily be a secure solution.

9.2 Enhancing Privacy via Noise

An alternative to controlling the sensor access through permissions is to add noise to the sensor data such that there is no apparent correlation between the demographics and the sensor readings.

For example, most smartphones come with a built-in vibration mechanism, primarily to alert users for incoming notifications and calls. The design of the vibrator allows it to vibrate the phone at different amplitudes and frequencies. One could leverage these parameters of the vibrator to develop a background app that triggers small vibrations in the phone when it detects that the user is using the phone. The vibrations would be limited in amplitude and duration such that the user would not explicitly feel the vibrations when using the phone.

Though apparently feasible, there are inherent limitations to this approach. First, vibration is a battery-intensive operation: Continued use of vibrations will lead to decay in battery life over time. Second, vibration sensors vary, and as a result, some users may indeed notice the vibrations unless there separate versions for different phone models — and such an app may simply not work on some models.

10 CONCLUSION

In this paper, we demonstrated that while the traditional smartphone privacy mechanisms do not protect sensors like the accelerometer and gyroscope from data collection, they still leave malicious actors the opportunity to gather protected information about the user. We present a non-invasive transparent technique to infer user demographic information by mining data from these seeming innocuous smartphone sensors. A study was conducted on 292 real smartphone users to demonstrate the effectiveness of the proposed approach. The experimental results show that our technique can achieve prediction accuracy of more than 77% on the user's gender and 86% on pregnancy prediction. Furthermore, we also demonstrate that the time needed to perform such an inference attack is less than a minute.

Our findings suggest that accelerometer and gyroscope sensors can be combined to create a profile consisting of a person's personality traits. While these sensors are a powerful attack vector on their own, our evaluation shows that, in combination, they pose a significantly more severe threat. These findings are powered by different digital data sources, which allowed us to perform a comparative study on the predictive power of each sensor alone and in combination. Overall, deviations for performance between the two scenarios, using accelerometer alone vs. the combination of the two sensors, were found to be significant for all demographic inference tasks. If our paper is accepted, we will provide the Android app, JS based web-app, open-source implementation of *ShakyHands*, model training scripts, and anonymized dataset.

11 FUTURE WORK

While our work can be used for demographic inference, our primary goal is to learn as much as possible about the user. This mined knowledge has value across a number of fields like marketing and intelligent application design. Our objective is to identify user traits, including human conditions, by building predictive models from labeled sensor data using supervised learning. We have identified potential areas of future research that build on *ShakyHands*:

(1) *ShakyHands* focuses on reading and browsing tasks; in the future, we would like to extend to include gaming activities.
(2) *ShakyHands* might also be extended to predict substance consumption (such as nicotine, caffeine) and psychological factors (such as stress), since these also impact the intensity of physiologic tremors [6].
(3) We would also like to work on potential defense mechanisms to mitigate the privacy threat posed by *ShakyHands* attacks.
(4) *ShakyHands* predicts demographic attributes. Such traits are often referred to as "soft biometrics" because these traits are not sufficiently distinctive to uniquely identify an individual, but they can be used in conjunction with other information for identification. In our next work, we plan to investigate biometric authentication based on physiologic tremor.

ACKNOWLEDGMENTS

The Programming Systems Laboratory is supported in part by NSF CNS-1563555, CCF-1815494 and CNS-1842456.

REFERENCES

[1] [n. d.]. Could Your Smartphone One Day Tell You You're Pregnant? | News Releases | The Optical Society. https://www.osa.org/en-us/about_osa/newsroom/news_releases/2015/could_your_smartphone_one_day_tell_you_you_re_preg/. (Accessed on 02/14/2020).

[2] [n. d.]. How to Request Device Motion and Orientation Permission in iOS 13. https://medium.com/flawless-app-stories/how-to-request-device-motion-and-orientation-permission-in-ios-13-74fc9d6cd140. (Accessed on 03/17/2020).

[3] [n. d.]. Motion sensors Âă|Âă Android Developers. https://developer.android.com/guide/topics/sensors/sensors_motion.html. (Accessed on 03/17/2020).

[4] [n. d.]. Other possible causes of hand tremors - CNN.com. https://www.cnn.com/2010/HEALTH/expert.q.a/04/07/hand.tremors.brawley/index.html. (Accessed on 03/14/2020).

[5] [n. d.]. Parkinsonism & tremors in Pregnancy: By Dr. De Leon | defeatparkinsons. https://defeatparkinsons.com/2014/02/11/parkinsonism-tremors-in-pregnancy-by-dr-de-leon/. (Accessed on 03/14/2020).

[6] 2019. Tremor Fact Sheet | National Institute of Neurological Disorders and Stroke. https://www.ninds.nih.gov/disorders/patient-caregiver-education/fact-sheets/tremor-fact-sheet. (Accessed on 03/13/2020).

[7] Mansour Alsaleh, Noura Alomar, and Abdulrahman Alarifi. 2017. Smartphone users: Understanding how security mechanisms are perceived and new persuasive methods. *PLOS ONE* 12, 3 (03 2017), 1–35. https://doi.org/10.1371/journal.pone.0173284

[8] Android. [n. d.]. Android - Google Play Protect. https://www.android.com/play-protect/. (Accessed on 12/05/2019).

[9] Julia Angwin. 2016. Protecting Your Digital Privacy - Consumer Reports. https://www.consumerreports.org/privacy/protecting-your-digital-privacy-is-not-as-hard-as-you-might-think/. (Accessed on 03/14/2020).

[10] Apple. [n. d.]. App Store - Apple. https://www.apple.com/ios/app-store/. (Accessed on 12/05/2019).

[11] Zhongjie Ba, Tianhang Zheng, Xinyu Zhang, Zhan Qin, Baochun Li, Xue Liu, and Kui Ren. [n. d.]. Learning-based Practical Smartphone Eavesdropping with Built-in Accelerometer. ([n. d.]).

[12] Xiaolong Bai, Jie Yin, and Yu-Ping Wang. 2017. Sensor Guardian: prevent privacy inference on Android sensors. *EURASIP Journal on Information Security* 2017, 1 (2017), 10.

[13] Xiaolong Bai, Jie Yin, and Yu-Ping Wang. 2017. Sensor Guardian: prevent privacy inference on Android sensors. *EURASIP Journal on Information Security* 2017, 1 (08 Jun 2017), 10. https://doi.org/10.1186/s13635-017-0061-8

[14] Sy Banerjee and Ruby Dholakia. 2012. Location Based Mobile Advertisements and Gender Targeting. *Journal of Research in Interactive Marketing* 6 (08 2012). https://doi.org/10.1108/17505931211274679

[15] Susanne Barth, Menno D.T. de Jong, Marianne Junger, Pieter H. Hartel, and Janina C. Roppelt. 2019. Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources. *Telematics and Informatics* 41 (2019), 55 – 69. https://doi.org/10.1016/j.tele.2019.03.003

[16] H. Berghel. 2018. Malice Domestic: The Cambridge Analytica Dystopia. *Computer* 51, 05 (may 2018), 84–89. https://doi.org/10.1109/MC.2018.2381135

[17] Matthias Böhmer, Brent Hecht, Johannes Schöning, Antonio Krüger, and Gernot Bauer. 2011. Falling Asleep with Angry Birds, Facebook and Kindle: A Large Scale Study on Mobile Application Usage. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*. ACM, New York, NY, USA, 47–56. https://doi.org/10.1145/2037373.2037383

[18] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[19] Avner Caspi and Paul Gorsky. 2006. Online deception: Prevalence, motivation, and emotion. *CyberPsychology & Behavior* 9, 1 (2006), 54–59.

[20] Abdelberi Chaabane, Gergely Acs, Mohamed Ali Kaafar, et al. [n. d.]. You are what you like! information leakage through usersâĂŹ interests. Citeseer.

[21] SH Cho, JM Park, and OY Kwon. 2004. Gender differences in three dimensional gait analysis data from 98 healthy Korean adults. *Clinical biomechanics* 19, 2 (2004), 145–152.

[22] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W. Kempa-Liehr. 2018. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh âĂŞ A Python package). *Neurocomputing* 307 (2018), 72 – 77.

[23] Courtney Connley. 2019. Women's Equality Day and the impact women voters have on elections. *CNBC* (August 2019). https://www.cnbc.com/2019/08/26/womens-equality-day-and-the-impact-women-voters-have-on-elections.html (Accessed on 03/24/2020).

[24] Erhan Davarci, Betul Soysal, Imran Erguler, Sabri Orhun Aydin, Onur Dincer, and Emin Anarim. [n. d.]. Age group detection using smartphone motion sensors. In *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2201–2205.

[25] Rahul C Deo. 2015. Machine learning in medicine. *Circulation* 132, 20 (2015), 1920–1930.

[26] Deborah Dinzes, Michael D Cozzens, and George G Manross. 1994. The role of gender in âĂŸattack ads:âĂIJRevisiting negative political advertising. *Communication Research Reports* 11, 1 (1994), 67–75.

[27] Judith S Donath. 2002. Identity and deception in the virtual community. In *Communities in cyberspace*. Routledge, 37–68.

[28] Yuxiao Dong, Yang Yang, Jie Tang, Yang Yang, and Nitesh V. Chawla. 2014. Inferring User Demographics and Social Strategies in Mobile Social Networks. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. ACM, New York, NY, USA, 15–24. https://doi.org/10.1145/2623330.2623703

[29] Michelle Drouin, Daniel Miller, Shaun M.J. Wehle, and Elisa Hernandez. 2016. Why do people lie online? âĂIJBecause everyone lies on the internetâĂİ. *Computers in Human Behavior* 64 (2016), 134 – 142. https://doi.org/10.1016/j.chb.2016.06.052

[30] The Economist. 2015. Planet of the phones - Smartphones. https://www.economist.com/leaders/2015/02/26/planet-of-the-phones. (Accessed on 12/05/2019).

[31] Max Eddy. 2018. How Companies Turn Your Data Into Money | PCMag.com. https://www.pcmag.com/article/364152/how-companies-turn-your-data-into-money. (Accessed on 12/05/2019).

[32] Dag G Ellingsen, Rita Bast-Pettersen, Jon Efskind, Merete Gjølstad, Raymond Olsen, Yngvar Thomassen, and Paal Molander. 2006. Hand tremor related to smoking habits and the consumption of caffeine in male industrial workers. *Neurotoxicology* 27, 4 (2006), 525–533.

[33] Hiroshi Endo and Koichi Kawahara. 2010. Relationship between hand stability and the 10-Hz physiological tremor during various manual tasks. *Ergonomics* 53, 4 (04 2010), 491. Copyright - Copyright Taylor Francis Group Apr 2010; Last updated - 2012-02-10; CODEN - ERGOAX.

[34] Hiroshi Endo and Koichi Kawahara. 2011. Gender differences in hand stability of normal young people assessed at low force levels. *Ergonomics* 54, 3 (2011), 273–281.

[35] Robert Faris, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. Partisanship, Propaganda, and Disinformation: Online media and the 2016 US Presidential Election. *Berkman Klein Center Research Publication 2017-6* (August 2017). https://ssrn.com/abstract=3019414

[36] Bjarke Felbo, Pål Roe Sundsøy, Alex Pentland, Sune Lehmann, and Yves-Alexandre de Montjoye. 2015. Using Deep Learning to Predict Demographics from Mobile Phone Metadata. *ArXiv* abs/1511.06660 (2015).

[37] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. 2012. Android Permissions: User Attention, Comprehension, and Behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12)*. ACM, New York, NY, USA, Article 3, 14 pages. https://doi.org/10.1145/2335356.2335360

[38] Sharad Goel, Jake Hofman, and M. Sirer. 2012. Who Does What on the Web: A Large-Scale Study of Browsing Behavior. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4660

[39] Google. [n. d.]. Google Play. https://play.google.com/store?hl=en_US. (Accessed on 12/05/2019).

[40] Google. [n. d.]. User Data | Privacy, Security, and Deception - Developer Policy Center. https://play.google.com/intl/en-US/about/privacy-security-deception/user-data/#!?zippy_activeEl=personal-sensitive#personal-sensitive. (Accessed on 12/05/2019).

[41] Brett Grossfeld. 2020. Deep learning vs machine learning: a simple way to understand the difference | Zendesk Blog. https://www.zendesk.com/blog/machine-learning-and-deep-learning/. (Accessed on 03/14/2020).

[42] Chris Hemedinger. 2018. Using your smartphone accelerometer to build a safe driving profile. https://blogs.sas.com/content/sgf/2018/09/26/accelerometer-driving-profile/

[43] Monica C. Schneider Holman, Mirya R. and Kristin Pondel. 2015. Gender Targeting in Political Advertisements. *Political Research Quarterly* 4 (2015).

[44] Jian Hu, Hua-Jun Zeng, Hua Li, Cheng Niu, and Zheng Chen. 2007. Demographic prediction based on user's browsing behavior. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 151–160.

[45] Gerald A. Hudgens, Linda T. Fatkin, Patricia A. Billingsley, and Joseph Mazurczak. 1988. Hand Steadiness: Effects of Sex, Menstrual Phase, Oral Contraceptives, Practice, and Handgun Weight. *Human Factors* 30, 1 (1988), 51–60. https://doi.org/10.1177/001872088803000105 arXiv:https://doi.org/10.1177/001872088803000105 PMID: 3350527.

[46] Iubenda. [n. d.]. Privacy Policy for Android Apps. https://www.iubenda.com/en/help/11552-privacy-policy-for-android-apps. (Accessed on 12/05/2019).

[47] Ankita Jain and Vivek 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)*. IEEE, 597–602.

[48] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2014. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.

[49] Kyriaki Kalimeri, Mariano G. BeirÃş, Matteo Delfino, Robert Raleigh, and Ciro Cattuto. 2019. Predicting demographics, moral foundations, and human values from digital behaviours. *Computers in Human Behavior* 92 (2019), 428 – 445. https://doi.org/10.1016/j.chb.2018.11.024

[50] Cindy D Kam, Allison MN Archer, and John G Geer. 2017. Courting the womenâĂŹs vote: The emotional, cognitive, and persuasive effects of gender-based appeals in campaign advertisements. *Political Behavior* 39, 1 (2017), 51–75.

[51] Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. 2012. A Conundrum of Permissions: Installing Applications on an Android Smartphone. In *Financial Cryptography and Data Security*, Jim Blyth, Sven Dietrich, and L. Jean Camp (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 68–79.

[52] N. Kiukkonen, Blom J., O. Dousse, Daniel Gatica-Perez, and J. K. Laurila. 2010. Towards rich mobile phone datasets: Lausanne data collection campaign. In *Proc. ACM Int. Conf. on Pervasive Services (ICPS,',',''), Berlin.*

[53] Jacob Kröger. 2019. Unexpected Inferences from Sensor Data: A Hidden Privacy Threat in the Internet of Things. In *Internet of Things. Information Processing in an Increasingly Connected World*, Leon Strous and Vinton G. Cerf (Eds.). Springer International Publishing, Cham, 147–159.

[54] Jacob Leon Kröger, Philip Raschke, and Towhidur Rahman Bhuiyan. 2019. Privacy Implications of Accelerometer Data: A Review of Possible Inferences. In *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy (ICCSP '19)*. ACM, New York, NY, USA, 81–87.

[55] J. K. Laurila, Daniel Gatica-Perez, I. Aad, Blom J., Olivier Bornet, Trinh-Minh-Tri Do, O. Dousse, J. Eberle, and M. Miettinen. 2012. The Mobile Data Challenge: Big Data for Mobile Computing Research. (2012). http://infoscience.epfl.ch/record/192489

[56] Yu-Kang Lee. 2014. Gender stereotypes as a double-edged sword in political advertising: Persuasion effects of campaign theme and advertising style. *International Journal of Advertising* 33, 2 (2014), 203–234.

[57] Kalev Leetaru. 2018. The Data Brokers So Powerful Even Facebook Bought Their Data - But They Got Me Wildly Wrong. https://www.forbes.com/sites/kalevleetaru/2018/04/05/the-data-brokers-so-powerful-even-facebook-bought-their-data-but-they-got-me-wildly-wrong/#1ee257de3107. (Accessed on 12/05/2019).

[58] Camelia Lemnaru and Rodica Potolea. 2012. Imbalanced Classification Problems: Systematic Study, Issues and Best Practices. In *Enterprise Information Systems*, Runtong Zhang, Juliang Zhang, Zhenji Zhang, Joaquim Filipe, and José Cordeiro (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 35–50.

[59] Elan D Louis. 2007. Kinetic tremor: differences between smokers and non-smokers. *Neurotoxicology* 28, 3 (2007), 569–575.

[60] Alan C. Luntz and Viktor L. Brailovsky. 1969. On estimation of characters obtained in statistical procedure of recognition.

[61] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

[62] Eric Malmi and Ingmar Weber. 2016. You are what apps you use: Demographic prediction based on user's apps. In *Tenth International AAAI Conference on Web and Social Media*.

[63] Rosa Mikeal Martey, Jennifer Stromer-Galley, Jaime Banks, Jingsi Wu, and Mia Consalvo. 2014. The strategic female: gender-switching and player behavior in online games. *Information, Communication & Society* 17, 3 (2014), 286–300.

[64] Aleecia M McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *Isjlp* 4 (2008), 543.

[65] Steven Melendez and Alex Pasternack. 2019. The data brokers quietly buying and selling your personal information. https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information. (Accessed on 12/05/2019).

[66] Hylton B Menz, Stephen R Lord, and Richard C Fitzpatrick. 2003. Age-related differences in walking stability. *Age and ageing* 32, 2 (2003), 137–142.

[67] Yan Michalevsky, Dan Boneh, and Gabi Nakibly. 2014. Gyrophone: Recognizing speech from gyroscope signals. In *23rd {USENIX} Security Symposium ({USENIX} Security 14)*. 1053–1067.

[68] Taeko Mizuno, Koji Tamakoshi, and Keiko Tanabe. 2017. Anxiety during pregnancy and autonomic nervous system activity: A longitudinal observational and cross-sectional study. *Journal of psychosomatic research* 99 (2017), 105–111.

[69] Lily Hay Newman. 2018. Mobile Websites Can Tap Into Your Phone's Sensors Without Asking | WIRED. https://www.wired.com/story/mobile-websites-can-tap-into-your-phones-sensors-without-asking/. (Accessed on 12/05/2019).

[70] Stefan Palan and Christian Schitter. 2018. Prolific.acâĂŤA subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (2018), 22 – 27. https://doi.org/10.1016/j.jbef.2017.12.004

[71] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* 70 (2017), 153 – 163. https://doi.org/10.1016/j.jesp.2017.01.006

[72] Do Viet Phuong and Tu Minh Phuong. 2014. Gender Prediction Using Browsing History. In *Knowledge and Systems Engineering*, Van Nam Huynh, Thierry Denoeux, Dang Hung Tran, Anh Cuong Le, and Son Bao Pham (Eds.). Springer International Publishing, Cham, 271–283.

[73] Piotr Plonski. 2019. Random ForestsÂő vs Neural Networks: Which is Better, and When? https://www.kdnuggets.com/2019/06/random-forest-vs-neural-network.html. (Accessed on 03/14/2020).

[74] Pavel Pudil, Jana Novovičová, and Josef Kittler. 1994. Floating search methods in feature selection. *Pattern recognition letters* 15, 11 (1994), 1119–1125.

[75] Sebastian Raschka. 2018. MLxtend: Providing machine learning and data science utilities and extensions to PythonâĂŹs scientific computing stack. *The Journal of Open Source Software* 3, 24 (2018).

[76] Suranga Seneviratne, Aruna Seneviratne, Prasant Mohapatra, and Anirban Mahanti. 2015. Your installed apps reveal your gender and more! *ACM SIGMOBILE Mobile Computing and Communications Review* 18, 3 (2015), 55–61.

[77] Laura Silver, Aaron Smith, Courtney Johnson, Kyle Taylor, Jingjing Jiang, Monica Anderson, and Lee Rainie. 2019. Mobile Connectivity in Emerging Economies. https://www.pewresearch.org/internet/2019/03/07/mobile-connectivity-in-emerging-economies/. (Accessed on 12/01/2019).

[78] Krystal Steinmetz. 2016. Now Your Smartphone Can Tell You If YouâĂŹre Pregnant | Money Talks News. https://www.moneytalksnews.com/now-your-smartphone-can-tell-you-youre-pregnant/. (Accessed on 02/14/2020).

[79] Sherry Turkle. 1995. *Life on the Screen: Identity in the Age of the Internet.* Simon Schuster Trade.

[80] Sherry Turkle. 1999. Cyberspace and Identity. *Contemporary Sociology* 28, 6 (1999), 643–648. http://www.jstor.org/stable/2655534

[81] W3C. 2019. Generic Sensor API. https://www.w3.org/TR/generic-sensor/. (Accessed on 12/01/2019).

[82] Xuetao Wei, Lorenzo Gomez, Iulian Neamtiu, and Michalis Faloutsos. 2012. Permission Evolution in the Android Ecosystem. In *Proceedings of the 28th Annual Computer Security Applications Conference (ACSAC '12)*. ACM, New York, NY, USA, 31–40. https://doi.org/10.1145/2420950.2420956

[83] Gary M Weiss and Jeffrey W Lockhart. 2011. Identifying user traits by mining smart phone accelerometer data. In *Proceedings of the fifth international workshop on knowledge discovery from sensor data*. ACM, 61–69.

[84] Monica T Whitty. 2008. Revealing the âĂŸrealâĂŹme, searching for the âĂŸactualâĂŹyou: Presentations of self on an internet dating site. *Computers in Human Behavior* 24, 4 (2008), 1707–1723.

[85] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David A. Wagner, and Konstantin Beznosov. 2017. The Feasibility of Dynamically Granted Permissions: Aligning Mobile Privacy with User Preferences. *2017 IEEE Symposium on Security and Privacy (SP)* (2017), 1077–1093.

[86] Venessa Williams. 2016. Black women vow to be a powerful voting force again this year. *The Washington Post* (January 2016). https://www.washingtonpost.com/politics/black-women-vow-to-be-a-powerful-voting-force-again-this-year/2016/01/10/f0c290fc-b324-11e5-a842-0feb51d1d124_story.html (Accessed on 03/24/2020).

[87] Josh Jia-Ching Ying, Yao-Jen Chang, Chi-Min Huang, and Vincent S. Tseng. 2012. Demographic Prediction Based on User's Mobile Behaviors.

[88] Jiexin Zhang, Alastair R. Beresford, and Ian Sheret. 2019. SensorID: Sensor Calibration Fingerprinting for Smartphones. In *Proceedings of the 40th IEEE Symposium on Security and Privacy (SP)*. IEEE.

[89] Li Zhang, Parth H Pathak, Muchen Wu, Yixin Zhao, and Prasant Mohapatra. 2015. Accelword: Energy efficient hotword detection through accelerometer. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 301–315.