Struggling, But Not Sufficiently So?: Exploring Unintentional Malingering in Adult ADHD

ANONYMOUS AUTHOR(S)
SUBMISSION ID: 123-A56-BU3

As adult ADHD diagnosis relies heavily on self-report, accurate diagnosis of adult ADHD is easily sabotaged by biased self-perceptions of symptoms. One key consequence of this that we introduce is *unintentional malingering*, which is when people overstate their symptoms and risk being falsely diagnosed. In this paper, we explore how we may detect unintentional malingering using technology, and provide insights into why the phenomenon may occur. We recruit 40 participants and through a thorough diagnostic procedure, identify 17 out of the 40 participants that may classify as unintentional malingerers. Then, we use mobile sensing to collect participants' behaviors and build a model that predicts the unintentional malingerers. We conduct further analysis using psychological scale data to interpret our model. Through regression and correlation analysis, we speculate that unintentional malingering may be related to higher stress levels and lower self-efficacy and general well-being. We discuss implications for the HCI community.

CCS Concepts: • **Applied computing** → *Psychology*; **Health care information systems**; • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: Adult ADHD; Measurement

ACM Reference Format:
Anonymous Author(s). 2022. Struggling, But Not Sufficiently So?: Exploring Unintentional Malingering in Adult ADHD. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

There is an increasing trend of ADHD diagnosis among adults worldwide, but there are controversies over this trend. A literature review on ADHD medication misuse suggests that the prevalence rate of misuse of stimulant medications is approximately 5% to 35% in college students [13]. There are several complexities of adult ADHD diagnosis that contribute to the phenomenon. First, the diagnostic procedure for adult ADHD relies heavily on retrospective reports. Criteria for diagnosis, according to the DSM-V criteria for adult ADHD, is the persistence of symptoms for at least *six months* "to the degree that negatively impacts social and academic/occupational activities" [4]. Thus, the validity of assessment - even in the context of formal diagnosis - could easily be compromised by the poor recall of past experiences. Second, there are other psychological conditions that may account for symptoms that resemble those of adult ADHD. For example, according to the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) symptoms of ADHD such as impaired concentration and feeling fidgety are diagnostic criteria for depression and anxiety [33].

There are issues specific to ASRS and *adult* ADHD - especially in the context of college students - that arise from the subjective nature of self-report surveys. First, college students (or any high-achieving group of the sort) may overestimate the level of their inattention and/or hyperactivity-impulsivity. The tendency to make judgments about oneself relative to prominent acquaintances rather than in absolute terms (i.e., the *reference-group effect*) could easily cause homogeneous groups such as college cohorts to be overly critical of themselves in terms of their level of focus or self-discipline. In fact, the increase in ADHD and academic accommodations were granted disproportionately to those who are "white, from upper-class socioeconomic backgrounds, and from private schools" [34]. Also, some college students facing academic concerns may find it more compelling to attribute their difficulties to an external factor such as adult ADHD than to admit to their struggles. In other words, some students may have a subconscious need to assume the sick role (i.e., *self-handicapping*) [5, 20] in the hopes to find an external factor to attribute their life struggles. While this could be a passing phase for some, others may go so far as to exaggerate their symptoms in the hopes of receiving help from stimulant medications or academic accommodations. Moreover, ADHD symptoms are non-specific, and other comorbid conditions such as depression and anxiety disorders could often account for the symptoms [38]. Chronically stressful situations such as performance pressures in competitive groups could also easily cause a temporary increase in certain behavioral symptoms of ADHD [31]. We refer to such groups of people who tend to overstate their symptoms as *unintentional malingerers*.

Unfortunately, there is no way to systematically detect unintentional malingering during ADHD diagnosis yet. ADHD is a highly individualized condition, with each person expressing symptoms in different ways. There is no conclusive physical test to diagnose ADHD, and diagnosis is typically given in response to a series of behavioral observations and self-reports. However, many psychiatrists lack specific expertise in making a conclusive diagnoses for adult ADHD; as many as 66% of 400 primary care physicians (PCPs) in a study[31] admitted they have inadequate knowledge and training to diagnose ADHD. Also, evaluators often fail to detect such biases due to the tendency to fall into the advocate role of the clients [34]. Finally, the diagnosis of adult ADHD is heavily reliant on self-reported symptoms. To this day, the American Academy of Pediatrics (AAP) guidelines stipulates conducting a clinical interview and administering the adult ADHD self-report scale (ASRS) as the two essential parts of adult ADHD diagnosis. However, self-reports are easily subject to bias and imprecision. These ongoing debates about the diagnosis and treatment of ADHD, along with the increasing rates of psychostimulant prescription suggest that the disorder may have been overdiagnosed and that ways to better detect unintentional malingering should be developed.

In this paper, we introduce the phenomenon of unintentional malingering in adult ADHD, propose how we may be able to better detect them using smartphone sensing, and provide insights into the psychological factors underlying unintentional malingering. We first conduct a motivational study to understand the group of people who may be inclined to over-report and over-perceive ADHD symptoms. Then, we administer thorough diagnostic procedures grounded in clinical work and identify 14 out of 37 participants who unintentionally malinger ADHD symptoms. We collect passive mobile sensing data and develop an early-stage prediction model that classifies the false positive group (i.e. people who unintentionally malinger) from the rest. Instead of building a black-box classification model grounded on self-report data, we collect a range of psychological scale data and interview data to comprehend how each sensor data feature relates to the behavioral, affective, and cognitive characteristics of the unintentional malingering group. Through regression and correlation analysis, we speculate that unintentional malingering may be related to higher stress levels and lower self-efficacy and general well-being. The likelihood of unintentional malingering is positively associated with self-reported traits such as (physical) aggression and anger, while negatively associated with those such as initiative, competence, optimism, and flourishing. Lastly, we discuss the implications of our study for the HCI community in terms of mental health support, self-report bias, and interpretability.

We explore the important yet underexplored issue of unintentional malingering via passive long-term tracking enabled by mobile sensors. Using mobile sensing for ADHD evaluation yields some unique advantages. First, as a tool that can easily be deployed in the wild, mobile sensing could complement diagnostic procedures by confirming the persistence of ADHD symptoms. Current diagnostic procedures such as clinical interviews and the ASRS rely on subjective interpretations of symptoms and behaviors. While people often lose objectivity and lack insight into their behaviors, smartphones are able to track behaviors in a consistent and ubiquitous manner. Also, smartphones can passively and objectively track behaviors for extended periods so the influence of cognitive bias or intentions to malinger would be minimized. Any fluctuations or irregularities in behavioral patterns can also be identified and used to assist in diagnosis. Moreover, smartphone sensing is affordable, unobtrusive, and scalable. Anyone with access to a smartphone can benefit from a smartphone app with a detection model built with the expertise of a few clinicians.

2 MOTIVATIONAL STUDY

This research is motivated by the intuition that the diagnostic procedure of adult ADHD is highly subjective which causes unintentional malingering to occur. To validate our motivation, we conducted an Institutional Review Boards (IRB)-approved study to capture how bias may influence the perception of ADHD-related characteristics.

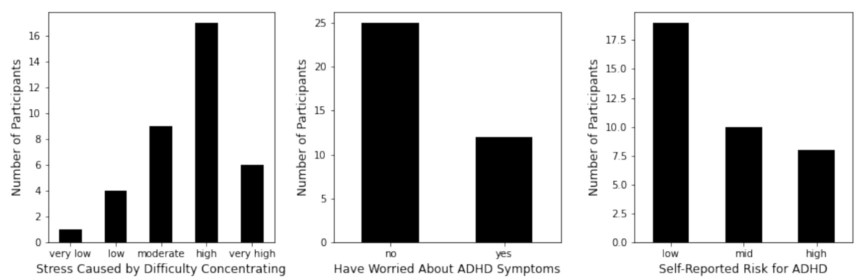
Participants. We recruited undergraduate students from a local university because prior literature suggests college students as the group that is most influenced by the misdiagnosis of adult ADHD. The students are often left under extreme academic pressures and feelings of constant stress, depression, and anxiety due to the intrinsic and extrinsic motivation to excel academically. Also influenced by the cultural contexts of the student body, students at the university experiencing academic difficulties often expose themselves to self-criticism[27]. Thus, we expected that this group demonstrates bias in their self-assessment (*i.e.,* overestimate their levels of ADHD symptoms) of ADHD traits. A total of 40 students (F=19, M=21, X=0) were recruited. The average age of the participants was 21.5 (SD=1.94), ranging from 19 to 26. As the school has limited ethnic diversity, all recruited participants were of one race. Participants were recruited via a school community website widely used among the students. We also asked each of the participants to recruit three informants for this study, resulting in 111 additional participants in the study.

Procedure. This motivational study captures the discrepancy between self and informant reports of the participant group's ADHD-related traits. Through this study, we test our hypothesis that college students may have biased perceptions of themselves. We created a survey consisting of the ASRS, psychological scales such as the Big-Five Personality Inventory (BFI), and three general questions measuring perceptions regarding adult ADHD. We administered the survey to our participants and asked them to gather responses on the survey from their three closest observers. We specifically requested responses from a long-time friend/family, a friend with interactions from multiple contexts, and a close friend with recent and frequent interactions to gather a comprehensive reference. [1]

[1]While informants' responses may also be biased, they can provide valuable insight into understanding the influence of subjectivity in ADHD self-report measures. It is common practice to reference another source of observation in ADHD diagnosis. For ADHD in children, clinicians use parent and instructor reports to make appropriate judgments on diagnosis and treatment. Such practice is sometimes adopted in diagnosing adult ADHD through assessment tools such as the Adult ADHD Symptom Rating Scale – Observer Version (ASRS-O) and Conners' Adult ADHD Rating Scales–Observer (CAARS-O) likewise. Moreover, it is a well-established theory in social psychology that there exists an asymmetry between the understanding of self and in others[9, 23, 36, 37, 49, 51], that "others sometimes know us better than we know ourselves"[50]. Specifically, observers often make better judgments of traits with higher observability and high evaluativeness. The behavioral symptoms of adult ADHD, as evident in the ASRS[26], could be presumed to be high in observability and evaluativeness, especially in high-achieving groups. Hence, it is reasonable to conjecture that observers could offer keener insights into the ADHD-related behaviors of our participant group.

Measurement. Based on prior psychological and medical research on adult ADHD, we selected 30 scales highly related to ADHD-like behavior and characteristics. We first conducted a pilot study of the survey with 48 students who do not participate in the actual study. The questionnaires were provided in both English and the local language. Scales without a validated local-language version were translated and agreed upon by at least two experts. Also, ambiguous or misleading terms were modified in the process. Then, by analyzing the scales that are most correlated with the ASRS, we obtained five scales that best reflect adult ADHD to our best knowledge. The scales included the ASRS, the BFI to measure conscientiousness, the General Procrastination Scale (GPS-9), the Temperament and Character Inventory (TCI) to measure persistence, and the short-form Buss-Perry Aggression Questionnaire to measure various forms of aggression.

Results. The results revealed that the participant group may indeed overestimate the level of their ADHD-related traits. First, the mean group ratings from participants on traits that are positively correlated to ADHD were higher than those ratings from their informants. Specifically, the group mean of participant ratings on procrastination, ADHD symptoms, and verbal aggression were 29.18 (out of 45), 11.27 (out of 24), and 8.46 (out of 25), respectively, while the average rating of the informant group on the same scales were 21.63, 9.20, and 6.74. The differences between group mean of the three scales were statistically significant (t=-6.715, p=< .001***; t=-3.150, p=.003**; t=-4.799, p=< .001***). On the other hand, participant ratings on traits negatively correlated to ADHD were lower than the informant ratings. Participants rated an average of 5.11 out of 7 on conscientiousness and 3.29 out of 5 on persistence, while informants rated an average of 5.99 and 3.74 on the same scales. The differences between these ratings were also statistically significant (t=6.599, p=< .001***; t=3.676, p=.001**).



(a) Stress from attention issues (b) Concerned about ADHD (c) Self-reported risk for ADHD

Fig. 1. Bar plots of self-reported survey responses to attention and ADHD related questions from a total of 37 participants.

While the prevalence rate of adult ADHD is around 3-9% worldwide [48], 42.5% (17 out of 40) of recruited participants were screened for adult ADHD on the ASRS. Meanwhile, only 0.88% (1 out of 117) of the informant ratings screened the participants as possibly having adult ADHD. The ASRS is a widely used self-report screening scale for adult ADHD developed by the World Health Organization (WHO) [26] with a high specificity of 99.5%. In other words, the participants, despite their reputation for hard work and self-management, rated themselves considerably low not only in comparison to the informant group but also in absolute terms. Also, our participant group expressed chronic stress regarding their focus or distraction levels. As shown in Figure 1, 86% of the participants responded that they have moderate to severe concerns regarding the matter. 32% answered that they worried they might have adult ADHD. Though no one sought a formal diagnosis from psychiatrists, almost half (49%) of the participants showed moderate (27%) to high (22%) certainty of having adult ADHD.

The large discrepancy between self and informant perspectives on the participant group's ADHD-related traits provides a compelling case to confirm our hypothesis that the inherent issue of subjectivity in self-reports could contaminate the outcome of ADHD diagnosis.

3 RELATED WORK

Existing Clinical Tools to Assist Adult ADHD Diagnosis. Recent studies in clinical psychology and psychiatry have explored the potential of objective measures such as existing neuropsychological tests to complement the diagnostic procedure of adult ADHD. Various neuropsychological assessments such as the Conner's Continuous Performance Test (CPT)[14] are believed to potentially provide an objective view of the symptoms of ADHD, yet controversy about its efficacy remains [1, 22, 40]. Moreover, as these tools do not yield conclusive results in diagnosing adult ADHD, some clinicians employ a test battery of multiple neuropsychological tests. Such practice requires hours of professional guidance, which is costly and time-consuming. As there is a limited number of clinicians with expertise in diagnosing adult ADHD many do not have access to such opportunities; even for those who do, the results from these tests may not help identify or understand bias or subjectivity in the diagnostic procedure.

Many recent studies in the clinical psychology and psychiatry domains were also dedicated to evaluating the potential of using objective measures such as symptom validity tests (SVTs) and other neuropsychological tests to detect ADHD malingering. Most of these tools are tested and developed in lab settings by comparing the performances of those properly diagnosed with adult ADHD against neurotypicals instructed to feign symptoms [19, 24, 44]. However, performance on such tests can easily be compromised by situational factors such as a belief about the diagnostic saliency of the given task. People instructed to feign ADHD with limited knowledge on the deficiency were able to malinger ADHD behaviors [21, 52]. As such, the results from these studies may still be influenced by cognitive bias and have limited ecological validity. Also, the tests are developed to test cognitive abilities such as sustained attention so the interpretation of the results from these tests concerning adult ADHD is only indirect and hypothetical. The findings are inconsistent [21, 52] and there is still debate over the usefulness of these tools in diagnosing adult ADHD [6].

Technology Solutions to Assist Adult ADHD Diagnosis. As it is a well-known issue that bias can interfere with a mental health diagnosis, many efforts have been made to leverage technology as an objective and non-biased method of assessment [11]. The most common method of computationally supported ADHD diagnosis is computerized continuous performance tests. Snappy App [53] uses a single CPT to measure ADHD symptoms and uses the UPPS-P Impulsive Behavior Scale [15] to obtain ground truth for ADHD. MATH-CPT[42] and QbCheck also incorporate CPT to detect ADHD in children[18] and adults[8]. Virtual Reality (VR) systems in which virtual classrooms are provided to monitor behaviors have also been developed [41]. Researchers have also explored the combination of games and electroencephalogram (EEG) for children who often find it difficult to follow through CPTs [2, 3]. Mock et al. [32] used touch interactions while solving multiple-choice math problems to predict ADHD risk. However, the study was conducted with fourth-grade students, and self-report questionnaire results were used as ground truth without further validation. Spachos et al. [46] developed WHAAM to monitor ADHD symptoms, though the app mainly focuses on providing a network for the caregivers of children with ADHD to easily track ADHD behaviors and adequate interventions. Some works have used complex technology such as brain activity analysis and eye movement analysis [17]. Such technology is costly, requires active user input, and has questionable accuracy. Machine learning models built using wearable sensing data have shown high accuracy [35, 39], but most researchers tested the technology only with children. Also, the machine learning approaches have left some issues with reproducibility, which calls for safer procedures and computations that are suitable for clinical practices [11].

While all of these technological approaches were relatively successful in assisting the prediction of adult ADHD, we put the focus of our research on a different angle. While bias can interfere with adult ADHD diagnosis in multiple ways, we focus on an aspect where people in homogeneous groups such as college students would over-report their symptoms as per the reference-group effect. We also try to understand the inner workings of our machine learning model by integrating psychological data so that our approach is appropriate for clinical purposes. Further, we focus on the use of smartphones rather than other sophisticated technologies as we believe that smartphones are yet the most accessible and ubiquitous.

4 ANALYZING UNINTENTIONAL MALINGERING

Motivation. As confirmed in our motivational study, certain groups of people are more inclined to have negatively biased perceptions of themselves regarding ADHD symptoms. However, the ADHD screening tool is not very effective in detecting such groups. There is no standard measure to detect such unintentional malingerers and it is not viable that all clinicians become equipped with experience and specialized expertise in adult ADHD diagnosis. While there are many possible ways to explore this issue, we chose to leverage the smartphone sensing method for the detection and understanding of bias in adult ADHD. The reasons are as follows:

(1) **Ubiquity.** Mental health assessments require comprehensive observations and reports of clients' behaviors, yet ADHD evaluations rely much on self-report of internal feelings. Many people carry their phones with them at all times so behaviors tracked from smartphone sensors would be useful diagnostic information.
(2) **Longitudinal Tracking.** Criteria for adult ADHD diagnosis is the persistence of symptoms for at least six months. This means that the reliability of the assessment could be called into question if the person being assessed cannot give accurate accounts of past experiences. For instance, someone experiencing a recent phase of inattention and restlessness caused by depressive feelings or anxiety could reinterpret memories in hindsight and falsely believe that they have had ADHD symptoms throughout their life.
(3) **Passive Sensing.** Unlike neuropsychological tests, smartphone sensing does not require any active user input. Sensors can passively capture behaviors that are correlated to adult ADHD symptoms, which would minimize the influence of cognitive bias.
(4) **Scalability.** The number of people that can be diagnosed with ADHD by PCPs with expertise in adult ADHD is low. Also, opportunities to receive rigorous evaluation are limited, costly, and time-consuming. A model developed using sensing data can be deployed so that anyone possessing a smartphone can benefit from the technology.

Procedure. In this main study, we aim to identify who and why some participants may have overestimated their ADHD behaviors by evaluating smartphone sensing data and self-reports. We conduct this main study through three phases of data collection and analysis steps to explore how sensing and scale data can complement adult ADHD diagnosis and understanding. In Phase 1, we conduct a thorough clinical diagnosis to determine the unintentional malingerers among our participants. Then, in Phase 2, we build a passive sensing app, collect data for an extended period, and build a prediction model that classifies the unintentional malingerers from the rest. Finally, in Phase 3, we analyze the data using psychological scale data to understand the unintentional malingering group.

The study was conducted for 22 days. The ASRS was administered on the first day of the study, along with the 54 psychological scales needed for Phase 3. Each participant attended the clinical diagnostic procedure in a scheduled time slot during the second week of the study. The sensing data collection process in Phase 2 was run every day throughout the 22 days.

The same 40 undergraduate students (F=19, M=21, X=0) that participated in the motivational study participated in this study, 37 of whom (F=19, M=18, X=0) participated in all phases of our main study. Students who were planning to take at least 9 credits and are currently not engaged in a full-time occupation were eligible to participate. In addition, recruitment was limited to Android phone users with an OS version 10.0 installed. Participants were offered up to USD 88.5, with a bonus $27 for full participation throughout the study.

4.1 Phase 1: Identifying Unintentional Malingering

Defining Unintentional Malingering. Studies that investigate malingering of ADHD typically recruit neurotypical participants and instruct a subgroup of them to display behaviors of malingerers [19, 24, 44]. However, as we intend to study malingerers who *unintentionally* report exaggerated symptoms of ADHD, we have to take a different approach in recruiting our participants. As shown in Figure 2, we want to determine the true positive (TP), false positive (FP), and true negative (TN) groups from self-reports of ADHD symptoms in respect to the clinical diagnosis results, whereby the FP group represents the unintentional malingerers. Each group represents the following:

- *The TP Group*: individuals who self-screened as having adult ADHD, and were formally diagnosed likewise
- *The TN Group*: individuals who self-screened negative, and were also determined as not having adult ADHD
- *The FP Group*: individuals who self-screened as having adult ADHD, but do not actually have adult ADHD (*i.e.,* the unintentional malingerers)
- *The FN Group*: individuals who self-screened negative, but were formally diagnosed with adult ADHD

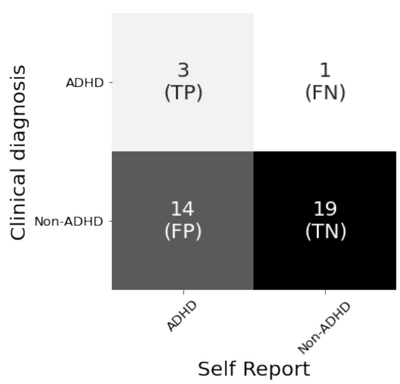


Fig. 2. Confusion matrix of self report and clinical diagnosis

Clinical Diagnosis. To obtain the ground truth of adult ADHD diagnosis, participants attended a clinical interview performed by a trained clinical psychologist and an experienced psychiatrist. The psychologist conducted validated clinical interviews using the Structured Clinical Interview for DSM-5 (SCID-5) to thoroughly examine the participants' possibility of having ADHD. The SCID-5 is the most recent version of SCID, a widely used semi-structured interview guide for determining whether an individual meets the criteria for DSM disorders [4]. Based on the results of the interview, the psychiatrist made a final diagnosis along with a further classification of the subtype of ADHD - inattentive, hyperactive, or combined. Additionally, clinical assessment of psychological disorders such as depression, bipolar disorder, anxiety, and post-traumatic stress disorder (PTSD) was conducted using the native language version of Mini-International Neuropsychiatric Interview (M.I.N.I.) [30]. The M.I.N.I. is a short structured diagnostic interview for DSM-5 and ICD-10 psychiatric disorders. The additional interviews ensured the absence (or presence) of comorbid diseases or conditions that may resemble symptoms of adult ADHD, such as frequent mood swings or lack of focus.

The clinician and psychiatrist both had a profound understanding of the general characteristics of the university student population. The psychiatrist especially had years of experience in the diagnosis and treatment of students from the university that express concerns about having adult ADHD. However, the clinician and psychiatrist were blind to the research hypothesis and were asked to carry out their usual procedures in diagnosing ADHD and comorbid disorders. Participants were informed that they were not obligated to respond to the questions during the interview if they felt uncomfortable. In addition, they were notified that the diagnostic results would not be disclosed to them nor be used for any purpose other than research. Participants attended the interview via phone call due to COVID-19. While there are benefits to conducting in-person assessments rather than telephone assessments, the protocol for adult ADHD diagnosis relies much on the behavioral accounts learned through the structured interview questions.

Results. We administered the ASRS screener three times throughout our study to confirm that the result from the motivational study (*i.e.,* 17 out of 40 participants were screened through the ASRS) is not coincidental. We evaluated the test-retest reliability of the three responses of ASRS self-ratings using intra-class correlation coefficient (ICC) estimates and their 95% confidence intervals (CI) calculated based on a single rating, absolute-agreement, 2-way mixed-effects model [28]. The results indicated good reliability with an ICC of 0.75 (CI = .59-.86) [28].

Results from the clinical interview revealed that only four participants (3 out of the 17 participants that self-screened positive using the ASRS, and one false negative) were diagnosed as having adult ADHD. Two participants belonged to the *hyperactive-impulsive* type, and the other two were each classified as the *inattentive* and *combination* types. As seen in Figure 2, we identified 14 FPs, 3 TPs, 1 FN, and 19 TNs from the results of self-screening and interviews. The evaluators commented that several participants reported overstated worries about their inattentiveness and impulsivity. The pieces of evidence given by some of the participants as inattentive or hyperactive/impulsive behaviors include *not strictly adhering to schedules at times* and *not being able to focus when listening to online lectures for self-improvement.*

30-50% of the participants that belong to the FP group often felt overwhelmed by the tasks at hand, much more so than in the TN group. 43% of the FP group often fidget or squirm, possibly indicating states of boredom and/or anxiety. Also, a portion of them reported feelings of chronic fatigue and emptiness. On the other hand, the FP group seemed to display little to no impulsivity or hyperactivity, unlike the TP group.

Most of the examples provided by the FP group in giving positive answers on the SCID-5 were evaluated as false pieces of evidence. As the interviewers commented, some of our participants took even small signs of inattention or impulsivity very seriously. For example, answers given to question H5 in the SCID-5 (*i.e.,* often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace) include: *"I have trouble finishing all my work when many assignments are given at once"*, *"I put off work during the first semester of college. But I do tend to finish all my assignments"*, and *"I try to do extra work for self-improvement, but often fail to focus"*.

Some participants in the FP group also reported that they often procrastinate. It is when individuals fail to manage feelings of overwhelming pressure that they begin to procrastinate. The FP group possibly experiences a lot of stress due to reasons such as academic pressure, low self-efficacy, perfectionism, and procrastinating on some of their tasks. Examples of such responses were *"I have trouble sticking to my long-term plans. I tend to avoid them"*, and *"I want to procrastinate my graduation thesis. It is too overwhelming"*. Meanwhile, some responses such as *"I have lost any purpose in life. I am not depressed, but I do feel lethargic"*, *"I feel empty, and I don't find my work to be meaningful"* and *"I often use* *my smartphone while I'm doing my assignments or when I'm attending my lectures"* suggest feelings of emptiness and boredom.

The school psychiatrist who supervised the diagnostic procedures in our study had extensive knowledge and experience in adult ADHD and the student population. However, in general, clinicians are unaware of the demographics of the client which means diagnoses are influenced by unintentional malingering. In fact, students in higher education are much more frequently diagnosed with adult ADHD than the general adult population, even in countries where there is a stigma against psychiatric illnesses and little or no accommodations for such [29].

4.2 Phase 2: Predicting Unintentional Malingering

This phase consists of a daily tracking procedure that leverages smartphone sensing, and a machine learning model evaluation procedure to predict the unintentional malingerers. To obtain behavioral patterns of the participants, we collect Bluetooth, GPS, call, activity recognition, screen event, notifications, and app usage data using our passive sensing app over 22 days. We then extract a superset of features that cover a range of behavioral patterns. To build a prediction modeling using sensing features to classify unintentional malingerers, we first select features based on the variance of the feature set. We further reduce the dimensionality of the features by applying principal component analysis (PCA) to the features. We employ the principal components (PCs) obtained from PCA to build prediction models using five classifiers: logistic regression (LR), random forest (RF), naive Bayes (NB), k-nearest neighbors (KNN), and support vector machines (SVMs). We validate and evaluate our model by performing a 5-fold Group K-fold cross-validation (CV).

Data Collection. We acquire phone usage and sensor data by utilizing our sensing app. The sensing app collects five different types of phone usage data (i.e., app usage, Bluetooth, call, notification, and screen event) and two types of sensing data (i.e., motion, location). The notification, motion, and screen data are stored upon every event (*e.g.*, receiving or clicking on a notification, unlocking the screen, changing in motion from walking to running). The light tracker is designed to detect and save light intensity at 5-minute intervals. The app passively collects sensing data in an unobtrusive manner once a user installs and starts our app, requiring no further user interaction afterward. While running in the background, the app status is displayed on the status bar at all times. The app restarts automatically when the phone is rebooted. The app uploads the attained data to our data collection server every six hours; the data upload happens at the next cycle when the device has a sufficient battery or is connected to WiFi.

Privacy Considerations & Data Management. As important as the quality of acquired data was participants' privacy since our sensing app inevitably collects sensitive data. At the beginning of the study, participants were informed thoroughly about the details on what kind of sensor data will be collected for which analysis and signed a consent form. We also developed a chatbot so that concerns regarding privacy issues could be addressed directly to us at all times. We also ran a blog to keep all the participants updated on any information or questions raised regarding privacy and security risks.

We kept the participant data anonymized throughout the study by allocating random IDs to each participant and keeping the map separate from all other data. The sensing data was stored on secure servers and access was limited to the researchers.

Compliance and Data Quality. Most participants remained participative throughout the study, resulting in high-quality data from 37 out of 40 participants. Two participants dropped out during the study due to personal reasons. We omitted one participant in the analysis phase due to the poor quality of sensor data associated with device-specific problems in the hardware sensors. To achieve high data quality, we maintained a script for monitoring purposes and examined the sensor data on the server daily. A message was sent via a chatbot to participants in cases of incompliance,

Table 1. List of sensor data collected by *Dopameter*

Sensor	**Data Collected**	**Feature**
App Usage	first/last timestamp, total time foreground/visible, app package name	app usage by category
Bluetooth	timestamp, device type, device class signal strength, device name	freq. of interactions with encountered devices, ratio of close devices[1] to total devices
Call	timestamp, type (incoming, outgoing), duration, phone number in hash	freq. of incoming/outgoing phone calls
Location	timestamp, latitude, longitude, altitude	total distance traveled, coverage area[2], no. of unique places visited, time spent at home
Motion	timestamp, transition (enter/exit), type (walking, vehicle, etc)	freq. of each motion, duration of each motion
Notifications	timestamp, app package name, type (received, dismissed, clicked, etc)	notifications by app category
Screen Event	timestamp, type (on, off, phone unlocked)	duration of screen on/phone unlock sessions, sessions with durations ≤ 3/5 seconds

[1] RSSI signal ≤ -60dBm
[2] computed based on [10]

such as turning off the GPS or Bluetooth sensors. As a result, data from all 37 remaining participants demonstrated high data quality, with negligible levels of missing data. In total, we collected 4,021,192 usage logs from 37 participants over 22 days. We filled any missing data through interpolation; a data point missing from an afternoon on Sunday of a participant was filled in with the average of data points from other Sunday afternoons of that participant.

Feature Extraction. We were not able to build a theory-driven model for this topic as little prior work has investigated the inner workings of unintentional malingerers' psychology. Instead, we try to obtain a superset of features we could extract from all eight sensors to explore a range of behavioral patterns. We draw on features from previous research on mobile sensing to extract behavioral features from smartphone sensing data. As presented in Table 1, we extract any major features from each sensor such as the frequency and duration of app usage, incoming and outgoing calls, motion (e.g., walking, vehicle ride), notification dismissal or click, or screen event (*i.e.,* screen on and phone unlock). To clarify, screen-on events reflect the act of pressing the power button, while phone-unlock events indicate the actual use of the phone by swiping the lock screen upon turning the screen on.

To analyze app usage data in more detail, we crawl category information for each app from Google Play. Of the 17 categories crawled, we remove 5 categories with too many zeros, modify a few wrongly categorized packages, and obtain 12 categories of apps - education, shopping, communication, browser, email, photography, organization, travel, fitness/health, weather, game, and finance.

To infer the mobility patterns of the participants, we extract the total distance traveled, the radius of gyration, and the number of places visited from GPS data. To extract the features, we first compute the movement trajectories (*i.e.,* sequence of places visited) within given time intervals. Then, we leverage a time-based clustering algorithm to determine significant places and relevant trajectories. A distance threshold of 30 meters and a time threshold of 5 minutes are employed. Using an algorithm from the literature [10], we merge identical geographic locations assigned with different coordinates or IDs.

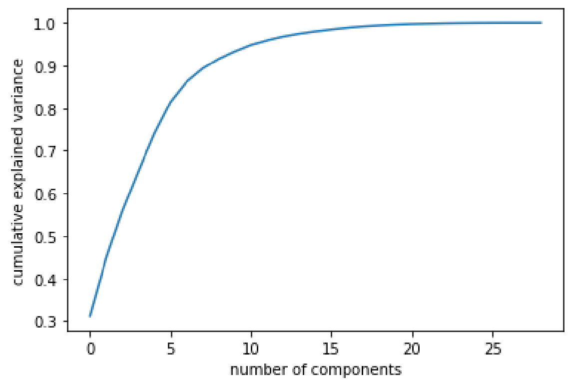


Fig. 3. The cumulative variability in the data explained by the top n principal components.

We utilize Bluetooth sensor data to compute the frequency of encounters with known devices and the ratio of devices within proximity to estimate the diversity and frequency of social interactions. We consider a device is *within proximity* when the RSSI signal strength is less than or equal to -60dBm, as prior work states that two phones within 2 to 4-meter distances apart from each other cover a Received Signal Strength Indicator (RSSI) value of -60dBm to -50dBm under ideal indoor circumstances.

We also extract the frequency of screen-on sessions with a duration of 3 seconds or less, and phone unlock sessions that last 5 seconds or less. This is to capture behaviors of frequently checking the phone without further use, such as habitually picking up the phone for checking the time or notifications only.

Upon extracting the basic features, we compute the average, median, maximum, and standard deviations of each set of sensing data across 22 days. In order to analyze behavioral characteristics within specific time periods, we also compute features by dividing time across the day into daily epochs (*i.e.*, *night* 12am-8am, *day* 8am-4pm, *evening* 4pm-12am). We also make variations specific to each sensor and obtain a superset of 403 features.

Data Analysis. In this study, we use behavioral features extracted from passive sensing data to predict unintentional malingering of adult ADHD symptoms. We consider this as a binary classification problem of differentiating the unintentional malingerers from the rest. The unintentional malingering group can be identified as the FP group; that is, those who consider themselves as having ADHD but are not diagnosed with ADHD upon formal diagnosis. The rest of the participants include those diagnosed as having ADHD (*i.e.,* the true positive and false negative groups) and those who correctly identified themselves as not having ADHD (*i.e.,* the true negative group).

To prevent overfitting our model, we select a more representative subset of the features. We obtain 29 features with higher variance, using the Analysis of Variance (ANOVA) F-statistic. To further reduce the feature dimensionality, we perform PCA on the matrix of 29 features and the classification results of each participant for unintentional malingering. We assess the prediction performance using a different set of PCs that explain 74%, 81%, 86%, 89%, and 92% of the cumulative explained variance ratio. The cumulative explained variance is illustrated in Figure 3. We observe that six PCs that explain approximately 81% of the variance of our matrix are most suitable for our analysis.

Using the 6 PCs selected, we train our model to predict the unintentional malingerers. We select five different algorithms - logistic regression (LR), support vector machine (SVM), k-nearest neighbors (KNN), decision tree (DT), and random forest (RF) - to train and evaluate our model. We first perform a Group 5-fold cross-validation (CV) and

Table 2. Classification results of different classifiers using PCA data

Classifier	Acc.	F-score	Sen.	Spec.	AUC
LR	0.92 (0.03)	0.89 (0.05)	0.87 (0.08)	0.96 (0.04)	0.98 (0.02)
SVM	0.83 (0.08)	0.8 (0.11)	0.87 (0.13)	0.81 (0.09)	0.94 (0.05)
KNN	0.87 (0.04)	0.84 (0.04)	0.87 (0.08)	0.87 (0.08)	0.92 (0.05)
RF	0.92 (0.03)	0.86 (0.06)	0.83 (0.11)	0.96 (0.04)	0.92 (0.05)
NB	0.84 (0.08)	0.74 (0.12)	0.67 (0.15)	0.95 (0.05)	0.88 (0.06)

Note. Darkly shaded row indicates the classifier with the best performance. Acc.: accuracy, Sen.: sensitivity, Spec.: specificity. LR: Logistic Regression, SVM: Support Vector Machine, KNN: K-Nearest Neighbor, RF: Random Forest, NB: Naive Bayes. Standard errors are given in parentheses.

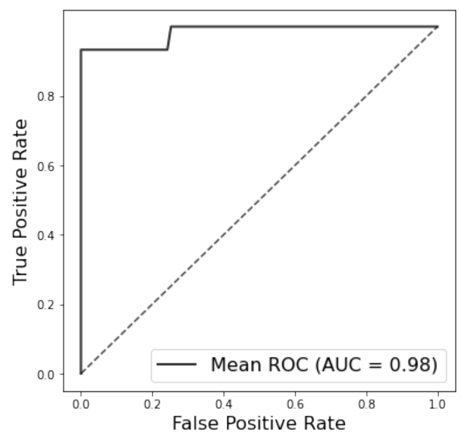


Fig. 4. Mean receiver operating characteristic (ROC) curves across 5 folds using different data

a grid search to find the best hyper-parameters for each model and prevent overfitting. Both the inner and outer CV schemes are 5-fold. The outer CV process yields 7 participants for the test set and the rest for training. To evaluate the performance of our models, we compute the following five metrics for each model: accuracy, F-score, sensitivity, specificity, and area under the ROC curve (AUC). We empirically determine the prediction threshold (*i.e.,* cut-off) to a value that prioritizes the F-score, as the metric prioritizes both precision and recall.

Results. As presented in Table 2, the LR classifier achieved the highest scores across all evaluation metrics. The mean metrics across the five folds were accuracy=0.92, F-score=0.89, sensitivity=0.87, specificity=0.96, and AUC=0.98. Figure 4 illustrates the mean ROC curve of the LR classifier performance across five folds. The model demonstrated strong predictive power with high total classification accuracy and AUC. In other words, the prediction model that we built using sensing features is effective in differentiating the behaviors of unintentional malingerers from those of the rest of the participants.

4.3 Phase 3: Understanding Unintentional Malingering

In this section, we aim to understand our model - which sensing features contributed in what ways to our model, and what the sensing features represent in terms of the psychology of the unintentional malingerers - through a three-step analysis.

Method. We first conduct regression analysis to explore the relationship between sensing features and unintentional malingering. Since the likelihood or extent of unintentional malingering cannot be expressed on a continuous scale, we are not able to perform Pearson's correlation analysis on the data. Instead, we choose to assess the coefficients from logistic regression analysis for evaluation. We first choose sensing features that contributed much to our model; that is we select sensing features that are representative of the PCs by identifying those with loadings of 0.65 or higher. We obtain 14 features that meet the condition. We remove one unnecessary feature and obtain 13 sensing features.[2] Then, we train each of the 13 sensing features to classify participants into the FP group (*i.e.,* unintentional malingerers) and the rest (*i.e.,* TP, TN, and FN groups) using sensing features. We leverage a nested 5-fold cross-validation (CV) and a grid search to find the best hyper-parameters for each model and prevent over-fitting. We use coefficients from the logistic regression models to understand how each of the 13 sensing features is associated with our model.

Then, we conduct correlation analysis to assess the meaning of the sensing features in terms of psychological characteristics. We experiment with the association between the 13 sensing features used for regression analysis and responses to the group of psychological scales. As correlation coefficients of 0.3 or greater are regarded as significant by convention, we employ 0.3 as the cutoff to determine the cases of meaningful associations. We found 35 such scales out of 54 scales that we administered. There were cases where a single survey had multiple instances of meaningful associations with distinct sensing features. Thus, we determine whether the directions of association between the survey data and the likelihood of unintentional malingering are consistent by taking both the coefficients from logistic regression analysis and correlation analysis into account. We found that one scale, which measures levels of extravagance, shows inconsistent associations. Moreover, 11 surveys had only one instance of significant correlation with the sensing features. We eliminate such 12 scales with insufficient explanatory power from further analysis. All other scales yielded coherent results, implying that the model we built using sensing features, in fact, yields valuable insights into the prediction of unintentional malingering.

Finally, we conduct t-tests to compare the characteristics of unintentional malingerers against those of specific groups, such as the TN or TP groups. The results from the regression and correlation analysis offer some insight into how various psychological traits relate to the likelihood of unintentional malingering. However, classification of the unintentional malingering group during the regression analysis was performed against a heterogeneous group of participants (*i.e.,* the TP, TN, and FN groups). While the results from the analysis are useful in understanding our model, we additionally perform Student's t-test on the survey data to better understand what the psychological surveys tell us about the unintentional malingering group in relation to the other groups. First, we compare the unintentional malingering group against the groups that were diagnosed with ADHD. Of the 21 selected surveys, 7 scales yielded high effect sizes.[3]

[2] To prevent an over-representation of similar features, we use variance inflation factor (VIF) values to remove correlated features. The VIF values are used to diagnose multicollinearity among features (*i.e.,* the extent of correlation between one feature and other features in a model). In general, VIF values greater than 10 are considered highly correlated; we remove features until the largest VIF value is smaller than 10 and obtain our final set of sensing features to use for analysis in this section.

[3] In this part of our analysis, we used Cohen's d effect sizes - instead of p-values to determine the features with statistically significant group mean differences between the FP and diagnosed groups. The effect sizes measure the magnitude of the difference between groups independent of the sample sizes, not just the probability of the observed differences. Because the sample size of the (TP+FN) is small, the effect sizes can give a much more accurate view of the actual differences between our groups. An effect size of .20 indicates a small difference, .50 a medium difference, and .80 a large difference. In

Table 3. Results of logistic regression analysis

No.	Feature	Coef.	F-score
1	No. of unique far Bluetooth devices during the evening	-0.935	0.75
2	No. of phone unlock events with durations ≤ 5 sec in a day	-0.937	0.75
3	Mean duration of walking motion activity in a day	-1.130	0.75
4	Mean duration of walking motion activity during the daytime	-1.316	0.75
5	Mean duration of vehicle motion activity during the daytime	-1.016	0.75
6	Mean no. of outgoing calls in a day	-1.425	0.75
7	Time spent at home during the daytime	0.820	0.75
8	Mean no. of phone unlock events during the daytime	-1.115	0.67
9	Mean no. of vehicle motion activity in a day	-1.369	0.67
10	No. of places visited during the evening	-0.694	0.67
11	Time spent at home during the evening	0.842	0.67
12	Mean duration of communication app use in foreground in a day	-0.355	0.6
13	Sd. of no. of outgoing calls in a day	-0.517	0.6

Table 4. Results from correlation analysis on sensor features' association with survey data

No.	Traits that predict unintentional malingering	Associated sensing feature no.
1	Higher likelihood of bipolar disorder	1 (+), 9 (+)
2	Higher aggression	1 (-), 3 (-), 4 (-), 10 (-), 11 (+)
3	Higher anger	1 (-), 3 (-), 4 (-), 7 (+), 11 (+)
4	Higher physical aggression	1 (-), 4 (-), 7 (+), 11 (+)
5	Higher boredom proneness	3 (-), 5 (-), 6 (-), 9 (-), 12 (-)
6	Higher procrastination	3 (-), 6 (-), 10 (-)
7	Lower promotion-focus	6 (+), 10 (+), 11 (-), 12 (+)
8	Lower persistence	6 (+), 9 (+), 10 (+), 11 (-), 12 (+), 13 (+)
9	Lower initiative	6 (+), 12 (+)
10	Lower competence	6 (+), 9 (+), 10 (+), 11 (-), 12 (+), 13 (+)
11	Lower achievement-striving	6 (+), 7 (-), 13 (+)
12	Lower industriousness	3 (+), 4 (+), 6 (+), 7 (-), 8 (+), 9 (+), 10 (+), 11 (-), 12 (+), 13 (+)
13	Lower self-directedness	3 (+), 4 (+), 10 (+), 12 (+)
14	Lower optimism	3 (+), 4 (+), 5 (+), 9 (+), 10 (+), 12 (+)
15	Lower impulse control	3 (+), 4 (+), 10 (+), 12 (+)
16	Higher rebelliousness	1 (-), 11 (+)
17	Lower agreeableness	7 (-), 10 (+), 11 (-)
18	Lower conscientiousness	6 (+), 9 (+), 13 (+)
19	Lower flourishing	3 (+), 4 (+), 10 (+), 12 (+)
20	Lower gratitude	4 (+), 6 (+), 7 (-), 10 (+)
21	Lower self-monitoring	6 (+), 9 (+), 12 (+), 13 (+)

Note. (+) and (-) indicate positive and negative coefficients from individual feature logistic regression models, respectively. The sensing feature numbers can be found in Table 3.

Results. The coefficients from the logistic regression analysis on each model are as listed in Table 3. All models showed adequate predictive power for a single feature prediction model, with F-scores of 0.6 or greater. This suggests

this case, we set an absolute effect size of .80 or larger as our cut-off. We use the 21 selected surveys from the correlation analysis in Phase 3 as those surveys provide a solid explanation for our prediction model.

Table 5. *t*-test between unintentional malingering vs. ADHD diagnosed groups using survey data

Scale	**Unintentional malingerers** Mean (SD)	**ADHD diagnosed** Mean (SD)	***t***	***d***
Optimism	3.24 (0.56)	4.02 (0.4)	2.578	1.550
Rebelliousness	2.41 (0.35)	2.9 (0.52)	2.231	1.342
Bipolar disorder	10.21 (4.26)	14.75 (3.86)	1.909	1.148
Anger	6.57 (1.95)	9 (3.37)	1.876	1.128
Flourishing	20.5 (5.73)	26.25 (5.74)	1.768	1.063
Self-directedness	3.24 (0.34)	3.55 (0.11)	1.748	1.051
Self-monitoring	34.79 (7.33)	40.75 (7.18)	1.441	0.866

Note. Cohen's *d* effect size = .20 small, .50 medium, and .80 large.

that the coefficients from the models can justifiably be used to interpret the associations between the sensing features and unintentional malingering. Specifically, the coefficients signify the relationship each sensing feature has with the likelihood of unintentional malingering; that is, a positive coefficient implies a higher likelihood of unintentional malingering for higher magnitudes of the corresponding sensing feature and vice versa.

The results of the correlation analysis - *i.e.,* the implications of the remaining 21 survey features for unintentional malingering and their directions of correlation with sensing features - are as demonstrated in Table 4. Using data from psychological scales, we find that higher levels of self-reported likelihood of bipolar disorder, (physical) aggression, anger, boredom proneness, procrastination, and rebelliousness are useful in predicting a higher likelihood of unintentional malingering. On the other hand, a higher likelihood of being classified as the unintentional malingering group seems to be associated with lower levels of self-assessed promotion-focus, persistence, initiative, competence, achievement-striving, industriousness, self-directedness, optimism, impulse control, agreeableness, conscientiousness, flourishing, gratitude, and self-monitoring. The table also shows how each psychological characteristic is related to the sensing features. For example, higher self-reported levels of aggression are negatively associated with the number of unique Bluetooth devices in far distances during the evening, the daily mean duration of walking motion activity, the mean duration of walking motion activity during the daytime, and the number of places visited during the evening. The trait was positively associated with time spent at home during the evening epoch.

The results from the *t*-test on FP and the diagnosed groups (*i.e.,* TP+FN groups) are summarized in Table 5. The differences between the groups on each feature in the table were statistically significant. The FP group showed significantly lower ratings on scales that measure well-being (*i.e.,* optimism, self-directedness, and flourishing), impulsivity (*i.e.,* rebelliousness, symptoms of bipolar disorder, and anger), and self-monitoring in comparison to the ADHD group. The results presented in Table 6 are those with statistically significant group mean differences between the FP and TN groups. 12 of the 21 surveys revealed significant results. The effect sizes of all results were at least 0.76, indicating strong effects. The unintentional malingering group rated themselves higher in levels of boredom proneness and procrastination. On the other hand, ratings on well-being (*i.e.,* gratitude, satisfaction, optimism), drive (*i.e.,* initiative, persistence, competence, promotion focus, self-directedness, industriousness), and impulse control were lower.

The key takeaways of our analysis were that the unintentional malingering group expresses lower levels of optimism and self-directedness in comparison to both those diagnosed with ADHD and the true negative group. Such results are consistent with the interview results, as participants of the unintentional malingering group showed a tendency to express higher levels of academic stress and feelings of emptiness than other groups. The *t*-test results, in particular, tell

Table 6. *t*-test between unintentional malingering vs. true negative groups using survey data

Scale	**Unintentional malingerers** Mean (SD)	**True negatives** Mean (SD)	***t***	***d***
Boredom proneness	22.93 (4.32)	4.32 (17.68)	-3.354	-1.219
Gratitude	32.64 (5.58)	5.58 (37.11)	3.158	1.148
Procrastination	33.43 (4.75)	4.75 (25.63)	-3.119	-1.133
Initiative	2.46 (0.52)	0.52 (3.11)	2.929	1.065
Persistence	3.04 (0.36)	0.36 (3.51)	2.889	1.050
Impulse control	3.01 (0.41)	0.41 (3.51)	2.788	1.013
Optimism	3.24 (0.56)	0.56 (3.79)	2.675	0.972
Competence	3.45 (0.38)	0.38 (3.83)	2.532	0.920
Regulatory focus promotion	3.37 (0.53)	0.53 (3.82)	2.480	0.901
Flourishing	20.5 (5.73)	5.73 (24.89)	2.327	0.846
Self-directedness	3.24 (0.34)	0.34 (3.54)	2.289	0.832
Industriousness	3.09 (0.48)	0.48 (3.52)	2.111	0.767

Note. Cohen's *d* effect size = .20 small, .50 medium, and .80 large.

us in detail how and why some of the psychological scale features from the correlation analysis was associated with the likelihood of unintentional malingering. For instance, lower levels of self-reported competence and flourishing were associated with a higher likelihood of classification into the unintentional malingering group; results from our *t*-test indicate that this may be because the unintentional malingerers displayed lower levels of competence and flourishing in comparison to the ADHD and true negative groups, respectively. In other words, the struggles that the unintentional malingerers as college students are real - that having ADHD is the cause of those struggles was probably not.

5 DISCUSSION

The Unintentional Malingering Group. The unintentional malingering group that we introduce in this paper is new but is not small; the unintentional malingering group took up a significant 37.8% of our participant group. We believe that the group deserves more attention from HCI researchers. HCI researchers have put many efforts into researching neurodiverse populations [7, 16, 43]. Many have also researched ways to make increase accessibility for and to assist people with ADHD and related conditions [12, 25, 45, 47]. Yet, groups such as the unintentional malingerers that do struggle but do not identify as neurodivergent populations are left as blind spots of research. No one in the group was diagnosed with adult ADHD or any other mental health conditions, but the stress, the struggles, and the "subjective impairments" that the group experiences are very much real. The unintentional malingering group would certainly welcome and find solidarity from research that support and accommodate their mental health.

Implications for Self-report Data and Bias. Instead of directly screening for adult ADHD like most research in the domain would, we discovered the unique problem of unintentional malingering in adult ADHD diagnosis. We had initially tried to predict adult ADHD using mobile sensing in our research. However, while interviewing college students to develop our research, we noticed that the number of students who report serious cases of adult ADHD symptoms was much larger than expected. While the struggles they were having were very real, it was quite unlikely that all of them actually have adult ADHD. We consulted an expert in the field who has examined multiple college students for adult ADHD: most students that visit the school clinic due to attention or discipline problems incline to believe that they have ADHD but are usually diagnosed with mood disorders such as depression, bipolar, chronic stress, or anxiety, or they are sometimes diagnosed with obsessive-compulsive personality disorder. What struck us is that these students could certainly be diagnosed with adult ADHD by some clinicians as the diagnostic procedure relies heavily on self-reports.

As did much prior research developed by the HCI community, we saw that technology could contribute in a unique way to the issue of bias and malingering in the adult ADHD diagnostic process. Our mind cannot be measured and mental health evaluations will necessarily rely on (in many cases, self-reported) behavioral proxies: the influence of bias and subjectivity in such assessments is inevitable and the strengths of technology such as objectivity can certainly provide alleviate the issue.

Implications for Clinical Prediction and Interpretability. We believe the interpretation of machine learning models to be a critical part of HCI research concerning clinical topics. In Phase 3 of our main study, we tried to leverage mobile sensing, psychological scale, and diagnostic interview data to understand our prediction model. We needed to estimate the interpretability of our models, as the diagnosis of adult ADHD or unintentional malingering both concern the realm of psychiatry. Research in the area seeks an understanding of clinical phenomena and theory-driven approaches to increase accountability and transparency while reducing unforeseen side effects. Thus, through such analysis, we hoped to gain a more comprehensive understanding of the implications our model has for the psychology of unintentional malingerers and gain some clinical validity. While this is an early work, exploratory study such as this can also provide insights into the clinical research community. Instead of only creating a tool or developing a technological infrastructure, we attempted to also gain a theoretical understanding of a unique population regarding an important clinical issue.

Implications for Research Involving Mental Health Diagnosis. There were a few diagnostic considerations we had to take into account in designing our research due to the unique characteristics of adult ADHD diagnosis:

First, the TP group had to consist of people that are not on medication. As stimulant medications are highly effective in reducing ADHD symptoms, those on medication may not display the true characteristics of ADHD patients before diagnosis. However, medications are the most recommended treatment for ADHD symptoms, and most patients diagnosed with adult ADHD take stimulant medications daily. As such, we recruited people who have not been diagnosed with adult ADHD previously and went through the formal diagnostic procedures with each participant.

Second, when determining the criteria for the FP group, we had to consider the nature of mental health diagnosis. While the best way to determine the FP group would have been by 1) conducting a formal diagnostic procedure with a clinician untrained for adult ADHD diagnosis or diagnosis of the student group and 2) observing how diagnosis from the novice clinician is overturned by an expert clinician. However, such a procedure would be highly unethical; it is already established that uncertainties in diagnosing psychiatric disorders are fundamentally inevitable. While it is possible to ask for second opinions, second opinions are conventionally served as references that primary clinicians can look into rather than conclusive judgments that should overturn the initial diagnosis. Therefore, we determined our FP group by taking self-screening survey results and comparing those to the ground truths (*i.e.,* clinical interview results by an expert).

Finally, ground truth for ADHD diagnosis had to be obtained from clinicians with sufficient experience and expertise in detecting comorbid conditions and malingering of ADHD. The diagnostic procedure in our study was administered by two experts in the field; the clinical psychologist received training in adult ADHD diagnosis, and the school psychiatrist had years of experience in diagnosing and treating students at the university the participants are matriculated at. As one of the two psychiatrists at the university, the clinician most likely has enough understanding of the university culture and insights into assessing the psychological difficulties that the students express.

The many unique issues of ADHD diagnosis that we discovered during our research process demonstrate the importance of a thorough literature review in fields outside of the HCI field so that the research would also have more impact in communities outside the HCI community. The steps we took in our research to consult with a psychiatrist and a psychologist to ground truth our subjects properly were important in achieving clinical plausibility.

Limitations and Future Work. Despite promising results, the findings of this early work are subject to three major limitations. First, the sample size is relatively small due to procedural difficulties in determining the unintentional malingerers and the cost involved in administering clinical diagnosis. The number of participants diagnosed with adult ADHD is particularly small. As such, we could not build models that directly classify unintentional malingerers from true positives. Second, all participants of our study are recruited from a university with a distinct academic culture so the extent of generalizability of this study is yet to be determined. Third, our experiment was conducted under the COVID-19 situation. Government policies restricted movement and contact with other people during the time of our study. Data collected from sensors such as GPS, Bluetooth, and motion, may have been affected by the circumstances. Due to these limitations, our results cannot be generalized to the entire relevant subgroup of our target population. These limitations could be overcome by validating our results in further research with larger and more diverse participant populations. Future work may also benefit from conducting experiments for an extended period as the guideline for ADHD diagnosis is the persistence of symptoms for six months or longer.

6 CONCLUSION

Adult ADHD diagnosis relies heavily on self-report, which means overstated ADHD symptoms can significantly influence diagnostic outcomes. While some clinicians would have an understanding of clients' backgrounds and how the level of reported symptoms may be biased, many are not experienced or trained to make such judgments. To overcome this issue, we explored how objective data collected via smartphone sensors longitudinally can assist the detection of unintentional malingering in adult ADHD. We achieve this by 1) identifying unintentional malingerers (*i.e.,* people who overestimate the severity of their ADHD symptoms) through the official diagnostic procedure of self-screen survey and clinical interview, 2) making a classification model that predicts the behaviors of unintentional malingerers, and 3) providing insights into the behavioral and cognitive characteristics of unintentional malingerers analyzed using mobile sensing and psychological scale data. We discover that a model built using smartphone sensors can successfully predict unintentional malingerers and that the cause of unintentional malingering may be lower self-efficacy, perhaps due to their struggles at academic work.

REFERENCES

[1] Claire Advokat, Leslie Martino, B. D. Hill, and William Gouvier. 2007. Continuous Performance Test (CPT) of college students with ADHD, psychiatric disorders, cognitive deficits, or no diagnosis. *Journal of Attention Disorders* (2007). https://doi.org/10.1177/1087054706292106

[2] Alaa Eddin Alchalabi, Shervin Shirmohammadi, Amer Nour Eddin, and Mohamed Elsharnouby. 2018. FOCUS: Detecting ADHD Patients by an EEG-Based Serious Game. *IEEE Transactions on Instrumentation and Measurement* 67, 7 (July 2018), 1512–1520. https://doi.org/10.1109/TIM.2018.2838158

[3] Alaa Eddin Alchalcabi, Amer Nour Eddin, and Shervin Shirmohammadi. 2017. More attention, less deficit: Wearable EEG-based serious game for focus improvement. In *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*. IEEE, Perth, Australia, 1–8. https://doi.org/10.1109/SeGAH.2017.7939288

[4] American Psychiatric Association and American Psychiatric Association (Eds.). 2013. *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed ed.). American Psychiatric Association, Washington, D.C.

[5] American Psychiatric Association. 1994. *Diagnostic and statistical manual of mental disorders (4th ed.).* American Psychiatric Association, Washington, D.C.

[6] Russell A. Barkley. 2019. Neuropsychological Testing is Not Useful in the Diagnosis of ADHD: Stop It (or Prove It)! *The ADHD Report* 27, 2 (April 2019), 1–8. https://doi.org/10.1521/adhd.2019.27.2.1

[7] Laura Benton, Asimina Vasalou, Rilla Khaled, Hilary Johnson, and Daniel Gooch. 2014. Diversity for design: a framework for involving neurodiverse children in the technology design process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 3747–3756. https://doi.org/10.1145/2556288.2557244

[8] Denise Bijlenga, Fredrik Ulberstad, Lisa B. Thorell, Hanna Christiansen, Oliver Hirsch, and J.J. Sandra Kooij. 2019. Objective assessment of attention-deficit/hyperactivity disorder in older adults compared with controls using the QbTest. *International Journal of Geriatric Psychiatry* 34, 10 (Oct. 2019), 1526–1533. https://doi.org/10.1002/gps.5163

[9] Philip S. Brenner and John DeLamater. 2016. Lies, Damned Lies, and Survey Self-Reports? Identity as a Cause of Measurement Bias. *Social Psychology Quarterly* 79, 4 (2016), 333–354. https://doi.org/10.1177/0190272516628298

[10] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*. ACM Press, Osaka, Japan, 1293–1304. https://doi.org/10.1145/2750858.2805845

[11] Franceli L. Cibrian, Gillian R. Hayes, and Kimberley D. Lakes. 2021. *Research advances in ADHD and technology*. Springer, Cham, Switzerland. OCLC: 1232008543.

[12] Franceli L. Cibrian, Kimberley D. Lakes, Arya Tavakoulnia, Kayla Guzman, Sabrina Schuck, and Gillian R. Hayes. 2020. Supporting Self-Regulation of Children with ADHD Using Wearables: Tensions and Design Challenges. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376837

[13] David B. Clemow and Daniel J. Walker. 2014. The potential for misuse and abuse of medications in ADHD: A review. *Postgraduate Medicine* 126, 5 (2014), 64–81. https://doi.org/10.3810/pgm.2014.09.2801

[14] C K Conners, M H S Staff, Vincent Connelly, Sonya Campbell, M MacLean, and J Barnes. 2000. Conners' Continuous Performance Test II (CPT II V. 5). *Multi-Health Systems Inc* (2000).

[15] Melissa A Cyders, Andrew K Littlefield, Scott Coffey, and Kenny A Karyadi. 2014. Examination of a short english version of the upps-p impulsive behavior scale. *Addictive Behaviors* 39, 9 (sep 2014), 1372–1376. https://doi.org/10.1016/j.addbeh.2014.02.013

[16] Nicholas Sheep Dalton. 2013. Neurodiversity & HCI. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. ACM Press, Paris, France, 2295. https://doi.org/10.1145/2468356.2468752

[17] Senuri De Silva, Sanuwani Dayarathna, Gangani Ariyarathne, Dulani Meedeniya, Sampath Jayarathna, Anne M P Michalek, and Gavindya Jayawardena. 2019. A rule-based system for adhd identification using eye movement data. In *2019 Moratuwa Engineering Research Conference (MERCon)*. IEEE, Moratuwa, Sri Lanka, 538–543. https://doi.org/10.1109/MERCon.2019.8818865

[18] Theresa S. Emser, Blair A. Johnston, J. Douglas Steele, Sandra Kooij, Lisa Thorell, and Hanna Christiansen. 2018. Assessing ADHD symptoms in children and adults: evaluating the role of objective measures. *Behavioral and Brain Functions* 14, 1 (Dec. 2018), 11. https://doi.org/10.1186/s12993-018-0143-x

[19] Thomas W. Frazier, Allison R. Frazier, Robyn M. Busch, Melissa A. Kerwood, and Heath A. Demaree. 2008. Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology* 23 (2008). Issue 5. https://doi.org/10.1016/j.acn.2008.04.001

[20] Allyson G. Harrison. 2006. Adults Faking ADHD: You Must Be Kidding! *The ADHD Report* 14, 4 (2006), 1–7. https://doi.org/10.1521/adhd.2006.14.4.1

[21] Allyson G. Harrison, Melanie J. Edwards, and Kevin C.H. Parker. 2007. Identifying students faking ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology* 22, 5 (2007), 577–588. https://doi.org/10.1016/j.acn.2007.03.008

[22] Ramzi Hasson and Jodene Goldenring Fine. 2012. Gender differences among children with ADHD on continuous performance tests: A meta-analytic review. https://doi.org/10.1177/1087054711427398

[23] Erik G. Helzer and David Dunning. 2012. Why and when peer prediction is superior to self-Prediction: The weight given to future aspiration versus past achievement. *Journal of Personality and Social Psychology* 103, 1 (2012), 38–53. https://doi.org/10.1037/a0028124

[24] Lindsey J. Jasinski, Jordan P. Harp, David T.R. Berry, Anne L. Shandera-Ochsner, Lisa H. Mason, and John D. Ranseen. 2011. Using symptom validity tests to detect malingered ADHD in college students. *Clinical Neuropsychologist* 25, 8 (2011), 1415–1428. https://doi.org/10.1080/13854046.2011.630024

[25] Xinlong Jiang, Yiqiang Chen, Wuliang Huang, Teng Zhang, Chenlong Gao, Yunbing Xing, and Yi Zheng. 2020. WeDA: Designing and Evaluating A Scale-driven Wearable Diagnostic Assessment System for Children with ADHD. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376374

[26] Ronald C. Kessler, Lenard Adler, Minnie Ames, Olga Demler, Steve Faraone, Eva Hiripi, Mary J. Howes, Robert Jin, Kristina Secnik, Thomas Spencer, T. Bedirhan Ustun, and Ellen E. Walters. 2005. The World Health Organization adult ADHD self-report scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine* 35, 2 (2005), 245–256. https://doi.org/10.1017/S0033291704002892

[27] Shinobu Kitayama, Hazel Rose Markus, Hisaya Matsumoto, and Vinai Norasakkunkit. 1997. Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of Personality and Social Psychology* 72, 6 (1997), 1245–1267. https://doi.org/10.1037/0022-3514.72.6.1245

[28] Terry K. Koo and Mae Y. Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15, 2 (June 2016), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

[29] Soo Jin Kwon, Yoonjung Kim, and Yeunhee Kwak. 2018. Difficulties faced by university students with self-reported symptoms of attention-deficit hyperactivity disorder: A qualitative study. *Child and Adolescent Psychiatry and Mental Health* 12, 1 (2018), 1–8. https://doi.org/10.1186/s13034-018-0218-3

[30] Y Lecrubier, Dv Sheehan, E Weiller, P Amorim, I Bonora, K Harnett Sheehan, J Janavs, and Gc Dunbar. 1997. The Mini International Neuropsychiatric Interview (Mini). A short diagnostic structured interview: reliability and validity according to the CIDI. *European Psychiatry* 12, 5 (1997), 224–231. https://doi.org/10.1016/S0924-9338(97)83296-8

[31] Paul Marshall, James Hoelzle, and Molly Nikolas. 2019. Diagnosing Attention-Deficit/Hyperactivity Disorder (ADHD) in young adults: a qualitative review of the utility of assessment measures and recommendations for improving the diagnostic process. *Clinical Neuropsychologist* 0, 0 (2019), 1–34. https://doi.org/10.1080/13854046.2019.1696409

[32] Philipp Mock, Maike Tibus, Ann-Christine Ehlis, Harald Baayen, and Peter Gerjets. 2018. Predicting adhd risk from touch interaction data. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. ACM, Boulder CO USA, 446–454. https://doi.org/10.1145/3242969.3242986

[33] Mandi W. Musso, Benjamin D. Hill, Alyse A. Barker, Russell D. Pella, and Wm Drew Gouvier. 2016. Utility of the Personality Assessment Inventory for Detecting Malingered ADHD in College Students. *Journal of Attention Disorders* 20, 9 (2016), 763–774. https://doi.org/10.1177/1087054714548031

[34] Jason M. Nelson, Benjamin Lovett, and Will Lindstrom. 2015. Assessing, Documenting, and Accommodating ADHD in College Students. *The ADHD Report* 23, 6 (2015), 7–11. https://doi.org/10.1521/adhd.2015.23.6.7

[35] Niamh O'Mahony, Blanca Florentino-Liano, Juan J. Carballo, Enrique Baca-García, and Antonio Artés Rodríguez. 2014. Objective diagnosis of ADHD using IMUs. *Medical Engineering & Physics* 36, 7 (July 2014), 922–926. https://doi.org/10.1016/j.medengphy.2014.02.023

[36] Emily Pronin, Justin Kruger, Kenneth Savtisky, and Lee Ross. 2001. You don't know me, but I know you: The illusion of asymmetric insight. *Journal of Personality and Social Psychology* 81, 4 (2001), 639–656. https://doi.org/10.1037/0022-3514.81.4.639

[37] Emily Pronin, Daniel Y. Lin, and Lee Ross. 2002. The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin* 28, 3 (2002), 369–381. https://doi.org/10.1177/0146167202286008

[38] Javier Quintero, Isabel Morales, Rosa Vera, Pilar Zuluaga, and Alberto Fernández. 2019. The impact of adult adhd in the quality of life profile. *Journal of Attention Disorders* 23, 9 (July 2019), 1007–1016. https://doi.org/10.1177/1087054717733046

[39] Mariachiara Ricci, Monica Terribili, Franco Giannini, Vito Errico, Antonio Pallotti, Cinzia Galasso, Laura Tomasello, Silvia Sias, and Giovanni Saggio. 2019. Wearable-based electronics to objectively support diagnosis of motor impairments in school-aged children. *Journal of Biomechanics* 83 (Jan. 2019), 243–252. https://doi.org/10.1016/j.jbiomech.2018.12.005

[40] Cynthia A. Riccio and Cecil R. Reynolds. 2001. Continuous performance tests are sensitive to ADHD in adults but lack specificity: A review and critique for differential diagnosis. https://doi.org/10.1111/j.1749-6632.2001.tb05776.x

[41] A.A. Rizzo, J.G. Buckwalter, T. Bowerly, C. Van Der Zaag, L. Humphrey, U. Neumann, C. Chua, C. Kyriakakis, A. Van Rooyen, and D. Sisemore. 2000. The Virtual Classroom: A Virtual Reality Environment for the Assessment and Rehabilitation of Attention Deficits. *CyberPsychology & Behavior* 3, 3 (June 2000), 483–499. https://doi.org/10.1089/10949310050078940

[42] Celestino Rodríguez, Débora Areces, Trinidad García, Marisol Cueli, and Paloma González-Castro. 2018. Comparison between two continuous performance tests for identifying ADHD: Traditional vs. virtual reality. *International Journal of Clinical and Health Psychology* 18, 3 (Sept. 2018), 254–263. https://doi.org/10.1016/j.ijchp.2018.06.003

[43] Moushumi Sharmin, Md Monsur Hossain, Abir Saha, Maitraye Das, Margot Maxwell, and Shameem Ahmed. 2018. From Research to Practice: Informing the Design of Autism Support Smart Technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–16. https://doi.org/10.1145/3173574.3173676

[44] Myriam J. Sollman, John D. Ranseen, and David T.R. Berry. 2010. Detection of feigned adhd in college students. *Psychological Assessment* 22, 2 (2010), 325–335. https://doi.org/10.1037/a0018857

[45] Tobias Sonne, Jörg Müller, Paul Marshall, Carsten Obel, and Kaj Grønbæk. 2016. Changing Family Practices with Assistive Technology: MOBERO Improves Morning and Bedtime Routines for Children with ADHD. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 152–164. https://doi.org/10.1145/2858036.2858157

[46] Dimitris Spachos, Antonella Chifari, Giuseppe Chiazzese, Gianluca Merlo, Gavin Doherty, and Panagiotis Bamidis. 2014. WHAAM: A mobile application for ubiquitous monitoring of ADHD behaviors. In *2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014)*. IEEE, Thessaloniki, Greece, 305–309. https://doi.org/10.1109/IMCTL.2014.7011153

[47] Katta Spiel, Eva Hornecker, Rua Mae Williams, and Judith Good. 2022. ADHD and Technology Research – Investigated by Neurodivergent Readers. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–21. https://doi.org/10.1145/3491102.3517592

[48] Lut Tamam, Gonca Karakus, and Nurgul Ozpoyraz. 2008. Comorbidity of adult attention-deficit hyperactivity disorder and bipolar disorder: prevalence and clinical correlates. *European Archives of Psychiatry and Clinical Neuroscience* 258, 7 (Oct. 2008), 385–393. https://doi.org/10.1007/s00406-008-0807-x

[49] Simine Vazire. 2010. Who Knows What About a Person? The Self-Other Knowledge Asymmetry (SOKA) Model. *Journal of Personality and Social Psychology* 98, 2 (2010), 281–300. https://doi.org/10.1037/a0017908

[50] Simine Vazire and Erika N. Carlson. 2011. Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science* 20, 2 (2011), 104–108. https://doi.org/10.1177/0963721411402478

[51] Simine Vazire and Matthias R. Mehl. 2008. Knowing Me, Knowing You: The Accuracy and Unique Predictive Validity of Self-Ratings and Other-Ratings of Daily Behavior. *Journal of Personality and Social Psychology* 95, 5 (2008), 1202–1216. https://doi.org/10.1037/a0013314

[52] Christina Wei and Julie A. Suhr. 2015. Examination of the role of expectancies on task performance in college students concerned about ADHD. *Applied Neuropsychology:Adult* 22, 3 (2015), 204–208. https://doi.org/10.1080/23279095.2014.902836

[53] Zoe Young, Michael P. Craven, Maddie Groom, and John Crowe. 2014. Snappy app: A mobile continuous performance test with physical activity measurement for assessing attention deficit hyperactivity disorder. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8512 LNCS, PART 3 (2014), 363–373. https://doi.org/10.1007/978-3-319-07227-2_35