Dopameter: Predicting Subtle Malingering of Adult ADHD Symptoms

JAEWON KIM, College of Liberal Studies, Seoul National University, Korea

YUNA JEONG, Department of Economics, Seoul National University, Korea

YURI KIM, College of Liberal Studies, Seoul National University, Korea

EUN YOUNG KIM, College of Medicine, Seoul National University, Korea

YOUNGKI LEE, Department of Computer Science and Engineering, Seoul National University, Korea

Extant approaches to assess adult ADHD are easily subject to malingering of symptoms, and solutions to detect subtle malingering (*i.e.,* subconsciously reporting exaggerated symptoms) are especially rare. In this paper, we propose the use of a scalable sensor-enabled framework designed for predicting subtle malingering of adult ADHD symptoms. In our 3-week-long user study, we gather 8 types of mobile sensing data such as GPS and Bluetooth from 37 participants,. We perform principal component analysis (PCA) on the sensing features that we extracted and build our model using the 6 principal components that we obtained. We were able to predict the 14 subtle malingerers that we identified at an AUC of 0.98 and F-score of 0.89. Then, we try to understand our model by assessing the psychological implications of the sensing features through regression and correlation analysis. We open-source our code for sensing application and data analysis and anonymized data. [1]

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:

JaeWon Kim, Yuna Jeong, Yuri Kim, Eun Young Kim, and Youngki Lee. 2022. Dopameter: Predicting Subtle Malingering of Adult ADHD Symptoms. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 23 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Adult ADHD is notorious for its non-specific and non-exclusive symptoms, making the accurate diagnosis of the disorder particularly challenging. Adult ADHD is frequently associated with comorbid psychiatric conditions such as sleep disorder, anxiety disorder, bipolar disorder, and depression [29, 51]. Chronically stressful situations such as performance pressures in competitive groups could often cause a temporary increase in certain behavioral symptoms of ADHD [42]. Also, some students facing academic concerns may find it compelling to attribute their difficulties to an external factor such as adult ADHD than to validate their 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 [49, 56]. Moreover, college students (or any members of competitive groups of the sort) may

[1] link to anonymized code

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. As these external incentives may encourage students to intentionally or unintentionally falsify their symptoms [47], the complete legitimacy of adult ADHD diagnosis remains questionable.

The term *malingering* refers to a range of behaviors, from outright feigning of symptoms driven consciously for external secondary gains to exaggeration of actual symptoms motivated by a subconscious need to assume the sick role (i.e. *self-handicapping*) [4, 21]. Recent literature suggests that the base rate of malingering in adult ADHD evaluations is between 18% to 48% [23, 30]. Also, a literature review on ADHD medication misuse and abuse suggests that the prevalence rate of misuse of stimulant medications is approximately 5% to 35% in college students [7]. Despite the significance of the matter, evaluators' practices on the assessment of ADHD malingering are suboptimal for a variety of reasons such as the tendency to fall into the advocate role of the clients [49]. Subtle malingering differs from intentional malingering, which is when people deliberately feign symptoms for external incentives such as stimulant medications or academic accommodations. We propose that the issue of subtle malingering (*i.e.,* subconscious exaggeration of symptoms) impedes the accurate diagnosis of adult ADHD.

Many recent studies in the clinical psychology and psychiatry domains were dedicated to evaluating the potentials of using objective measures such as symptom validity tests (SVTs) and other neuropsychological tests to detect ADHD malingering. The line of research, though promising, exhibits some fundamental limitations. For instance, existing studies have mostly focused on detecting the conscious malingering of symptoms by comparing the performances of individuals properly diagnosed with ADHD and neurotypicals instructed to feign symptoms [15, 26, 56]. Findings from these studies may lack ecological validity, and may not be generalized for the detection of the more subtle kind of malingering (i.e., subconscious feigning or exaggeration of symptoms). In addition, performance on cognitive tasks could easily be influenced by situational factors such as a belief about the diagnostic saliency of the given task. As such, people instructed to feign ADHD with limited knowledge on the deficiency were able to malinger ADHD behaviors [24, 67]. Results on the efficacy of using existing tools to detect feigned ADHD have expectedly been inconsistent, and point to the need for the use of several SVTs. However, such a procedure is less than ideal as the administration of SVTs is costly and time-consuming while the results remain highly speculative.

In this paper, we introduce a scalable sensor-enabled ADHD detection framework designed for assessing subtle malingering of ADHD. Extant approaches lead to a call for an in-the-wild method that can robustly detect and understand more subtle cases of malingering in an unobtrusive, affordable, and reliable manner. We propose that passive long-term tracking using mobile sensors (e.g. Bluetooth, GPS, motion) has the potential to serve as such assessment. Using mobile sensing data for ADHD evaluation yields some unique advantages. First, when used for long-term tracking purposes, mobile sensing could complement diagnostic procedures by confirming the persistence of ADHD symptoms. Second, we could use mobile sensing to passively and objectively track behaviors for an extended period to minimize the influence of bias or intentionality in ADHD assessment and nullify efforts to malinger the deficit. Further, a mobile sensor-based solution for ADHD evaluation could prove useful in promoting early detection and alleviating the problems of underdiagnosis, as it has the potential to screen individuals with high risks of adult ADHD without any direct user engagement.

We evaluate the feasibility of using a mobile sensing framework to predict subtle malingering through a three-phase procedure. First, we identify the subtle malingerers using results from ADHD self-screening and clinical diagnosis. Out of the 17 participants whose initial Adult ADHD Self-Report Scale (ASRS) results yielded positive results, only three of them were formally diagnosed, while 14 of them were false positives. Second, we develop a sensor-enabled framework to passively collect various sensing and phone usage data (*i.e.,* screen events, notifications, app usage, in and outgoing calls, Bluetooth connections, mobility patterns, and physical activities), extract and select sensing features, and apply machine learning models to predict exaggerated reports of ADHD symptoms. We evaluate our sensor-enabled framework by conducting an IRB-approved in-the-wild user study with 37 students over 22 days, and use the sensing data to build a prediction model that classifies subtle malingerers. Our model accurately classifies subtle malingerers with total classification accuracy=0.92, F-score=0.89, sensitivity=0.87, specificity=0.96, and AUC=0.98. Finally, we evaluate the implications of our model by analyzing the relationships between the key sensing features and subtle malingering, and the associations between sensing features and data collected using psychological scales.

Overall, our contributions can be summarized as follows:

(1) Our user study reveals that existing self-evaluation tools for adult ADHD such as the ASRS is insensitive to subtle malingering. We also identify the existence of subtle malingerers by leveraging results from clinical diagnosis.
(2) We show the accurate predictive power of our prediction model in assessing feigned adult ADHD. Our model showed greater predictive accuracy than any other existing tools for identifying malingered ADHD, with an Area Under the Curve (AUC) of 0.98.
(3) We analyze our model using clinician interview results along with self and informant responses to various psychological scales to infer the psychological characteristics of the malingerers. By leveraging data collected from our interview, mobile sensing, and psychological scales, we are able to unveil the behavioral, cognitive, and affective characteristics of ADHD malingerers as a whole.

2 RELATED WORK

2.1 Technology for Mental Health Evaluations

Mobile Sensing. Digital phenotyping and monitoring mental illnesses based on sensing technology have proven effective in many contexts [1, 45, 50, 62]. Much research is devoted to detecting mental health issues, such as bipolar disorder [2, 19], schizophrenia [64, 66], depression [53, 54], and stress [34, 70]. Some works have addressed mental health issues in the college student population in particular [11, 16, 46, 65]. However, little work has explored the use of mobile sensing methods to evaluate adult ADHD, especially the issue of malingering.

ADHD Evaluation. Prior work has developed technology to support ADHD assessment in ways other than mobile sensing [71]. For example, Mock et al. [44] 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. [57] developed WHAAM to monitor ADHD symptoms, but the app mainly focuses on providing a network for the caregivers of children with ADHD to easily track ADHD behaviors and provide adequate interventions. Snappy App [69] is another mobile app developed for ADHD detection, but it uses a single continuous performance test (CPT) to rate symptoms and uses the UPPS-P Impulsive Behavior Scale [10] as ground truth for ADHD. Some works have used complex technology such as brain activity analysis and eye movement analysis [12]. Still, the application of such systems is limited as the systems require active user input and costly equipment. The research studies also do not take motivations to malinger into consideration.

2.2 Detecting Feigned Adult ADHD

Symptom Validity Tests (SVTs). There has been considerable research in adopting objective measures - from classical neuropsychological tests and SVTs [33] to computer-aided methods [71] - to determine the presence of ADHD in ways that could minimize issues associated with traditional self-report based assessments. Yet, the results are not very strong. Though there is some variability in the results, most self-report questionnaires including the ASRS and various neuropsychological tests were not effective in detecting ADHD malingering. Several studies suggest that using several SVTs is the most effective but the predictive power of the approach is questionable; the Structured Inventory of Malingered Symptomatology (SIMS) [18] showed 100% specificity but low sensitivity (.36-.52), and using the criterion of failure of 2 or more SVTs had a sensitivity of .475 and a specificity of 1.00 [27]. Some studies have found no success with using the SVTs [24, 48, 60, 68].

There are also issues concerning the experimental design of the studies. In the existing studies, participants are split into groups, usually the control group, the group coached to feign ADHD, and the true ADHD group. Then, the malingerers are classified based on different cutoffs, on the basis that those instructed to malinger tend to overexaggerate and perform worse than the clinical group [18, 61]. However, such experimental setup has some fundamental flaws. Despite some success with the approach, a thorough review of prior work reveals that the profiles between individuals who malinger and those who do not are too similar to draw solid conclusions solely from SVTs [47]. Furthermore, a significant portion of malingerers today are people who actually experience ADHD symptoms to an extent, while consciously or subconsciously desiring to be diagnosed with ADHD [21, 47]. However, groups instructed to malinger ADHD typically consisted of individuals with no true intent or interest to malinger ADHD in reality. Thus, the ecological validity of the classification results or the generalizability of the results to subtle malingerers is questionable. In fact, a study revealed that detection of exaggeration was consistently poorer than detection of feigning [20].

Independently Developed Scales. Although SVTs show moderate performance overall, the tests are less than ideal for most primary care settings as they are very costly, time-consuming, and require an expert for administration and interpretation [28]. Thus, some researchers have taken a step further to create stand-alone tools developed explicitly for detecting feigned ADHD. For instance, Harrison et al. designed the Exaggeration Index (EI) to discriminate between those suspected or instructed to malinger ADHD and all other clinical groups [22]. However, the new measure demonstrated great specificity (.97) but poor sensitivity (.24). There were some scales with adequate predictive power. Courrégé et al. [9] also developed the ADHD Symptom Infrequency Scale (ASIS), which can identify feigned or exaggerated symptoms of ADHD with strong sensitivity (.79-.86) and specificity (.89). Ramachandran et al. conducted a thorough study to develop the Subtle ADHD Malingering Screener (SAMS) [52] with sensitivity of .90 and specificity of .80. Though the scales can offer detailed insight into the cognitive and affective characteristics of malingerers, the scales were developed using experimental designs rather than a real-world population. Also, people would easily be able to learn how to fabricate their answers in the long run.

3 PROCEDURE AND USER STUDY

This work is motivated by the intuition that capturing the influence of subtle malingering on self-assessment and using mobile sensors as objective measures for longitudinal tracking purposes can benefit the overall ADHD diagnosis process. We presume that it is difficult or almost impossible to feign symptoms of ADHD detected via mobile sensing for an extended period. Thus, the detection of intentional malingering is less of an issue in the case of mobile sensing. On the other hand, individuals who subconsciously exaggerate their symptoms are likely to actually experience ADHD symptoms to an extent. Hence, our concern is to detect and classify the behavioral characteristics of subtle malingerers using mobile sensors. Then we use data from surveys and interviews to additionally understand and explain the cognitive and affective traits of the individuals.

3.1 Study Procedures

Studies that investigate intentional malingering of ADHD typically recruit neurotypical participants and instruct a subgroup of them to display behaviors of malingerers [15, 26, 56]. However, as we intend to study subtle 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 subtle 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 diagnosed as not having adult ADHD
- *The FP Group*: individuals who self-screened as having adult ADHD, but were diagnosed as not having adult ADHD (*i.e.,* the subtle malingerers; the individuals who tend to exaggerate their symptoms)
- *The FN Group*: individuals who self-screened negative, but were formally diagnosed with adult ADHD
- *The TP & FN Groups*: individuals formally diagnosed with adult ADHD, regardless of self-screening results

To validate our motivation, we conduct an IRB-approved study on the participant group. Figure 1 illustrates the overall study procedure and methods. Our research employs a three-phase study process.

3.1.1 Phase 1 (Section 4). We first conduct a motivational study to test our hypotheses that 1) some adult participants tend to unintentionally report exaggerated symptoms of ADHD and that 2) existing ADHD assessment tools such as the ASRS are not sensitive to such behaviors of subtle malingering. To this day, conducting a clinical interview and administering ADHD behavior rating scales are the essential parts of adult ADHD diagnosis, as stipulated in the American Academy of Pediatrics (AAP) guidelines. Thus, we ask the participants to answer the ASRS, a self-report survey for adult ADHD, and then attend an interview with clinical experts for formal ADHD diagnosis. We first confirm, through the self-screening results, that exaggerated reports of ADHD symptoms are in fact common among our participants. Then, combining the results from the self-reports with those from formal diagnosis, we identify the subtle malingerers (i.e., people that unintentionally exaggerate their symptoms); that is, participants who subtly malinger are revealed as false positives, which are the participants that were screened via self-reports but were not actually diagnosed upon formal examinations.

3.1.2 Phase 2. This phase consists of a daily tracking procedure that leverages smartphone sensing, and a machine learning model evaluation procedure to predict the subtle 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 subtle 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) on 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).

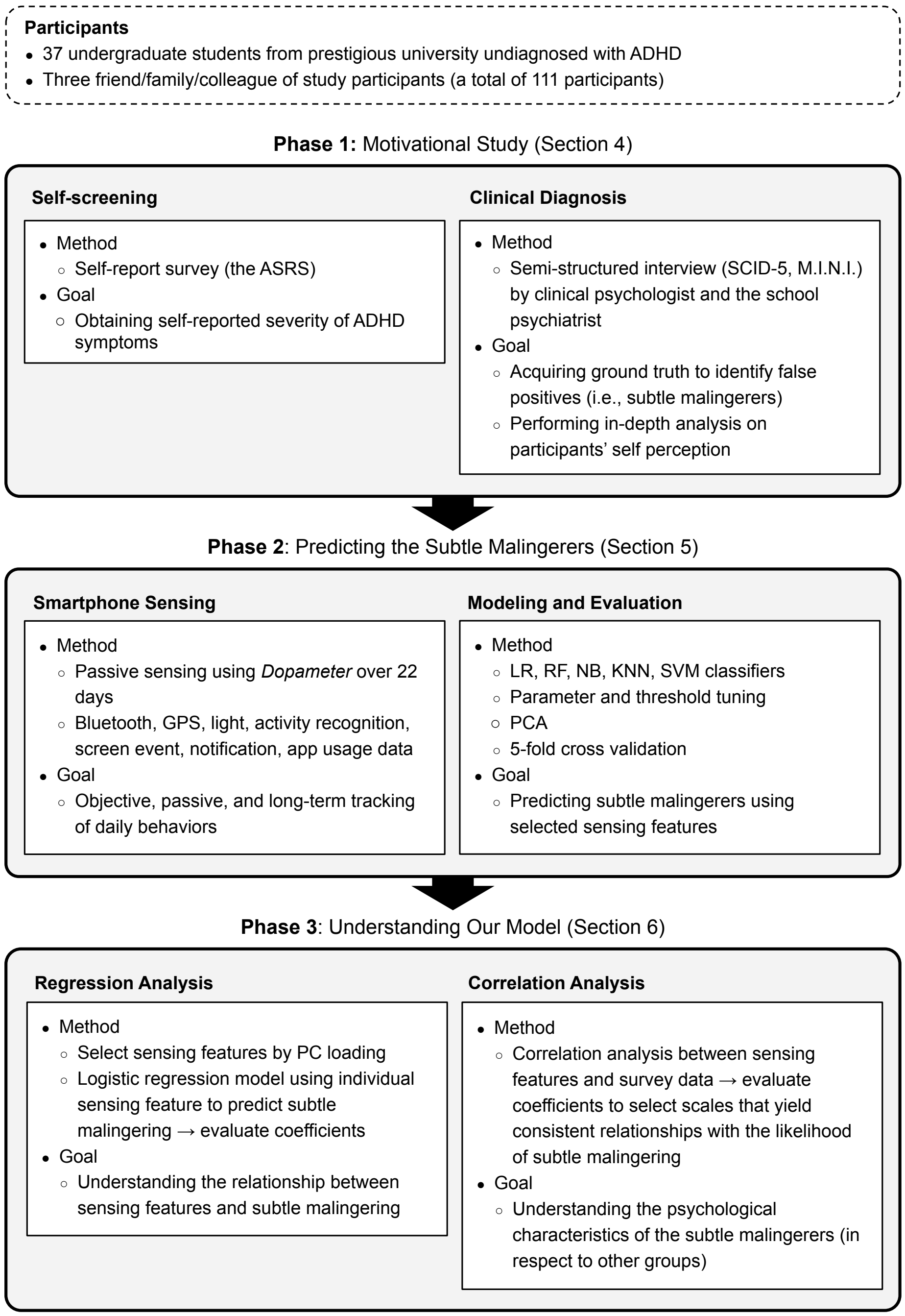

Fig. 1. Overall study procedure and the phase-specific methods, goals, etc.

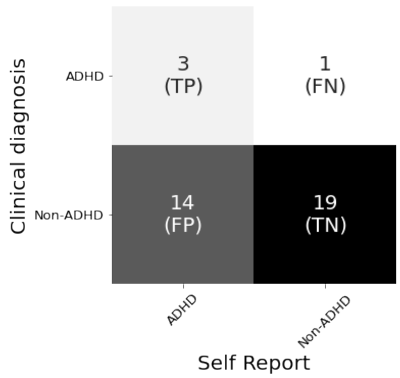

Fig. 2. Confusion matrix of self report and clinical diagnosis

3.1.3 Phase 3. We explore the implications our sensing features have for the prediction of subtle malingering in a two-step process. First, we perform logistic regression analysis between individual sensing features and the likelihood of subtle malingering. We build binary prediction models using individual sensing features to understand how each feature relates to subtle malingering. In performing this analysis, we analyze the factor loadings the sensing features have with each PC and select sensing features with high loadings. We infer the relationship between the sensing features and the likelihood of subtle malingering from the coefficient of each of the regression models, upon confirming the validity of the models using F-scores. Second, we perform association analysis on the selected sensing features and the survey data we collected. For each sensing feature, we select scales with correlation coefficients with sizes of 0.3 or greater. Then, considering the coefficients from individual sensing feature prediction models and the association between the sensing feature and survey, we validate whether each selected survey is related to subtle malingering coherently. We utilize surveys that display consistent associations with subtle malingering to understand our model and explore the possible psychological characteristics of the subtle malingerers.

3.2 User Study

Participants. 40 undergraduate students (F=19, M=21) participated in the user study, 37 of whom (F=19, M=18) participated in all phases of our study. The average age of the 37 participants was 21.5 (SD=1.94), ranging from 19 to 26. Participants were recruited via a school community website widely used among the students. The study was conducted for 22 days, which included the last week of summer vacation and the first two weeks of the Fall semester of 2020. 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 day period. 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.

Psychological Scales To use for analysis in Phase 3, we utilize 54 psychological scales reflecting a range of psychological traits and states that help promote understanding of the underlying cognitive and affective characteristics of the subtle malingering group. A group of scales such as the Patient Health Questionnaire depression scale (PHQ-8)

[36], General Anxiety Disorder 7-item scale (GAD-7) [58], Bipolar Spectrum Diagnostic Scale (BSDS) [63], and the Sleep Disorders Symptom Checklist-17 (SDS-CL-17) [25] screen the common comorbid mental health disorders of adult ADHD. Correlates of adult ADHD revealed in the literature are measured using scales such as the Risk Propensity Scale (RPS) [43], the Regulatory Focus Questionnaire (RFQ) [32], the General Procrastination Scale (GPS-9) [55], the 12-item short form of the Buss-Perry Aggression Questionnaire (BPAQ-SF) [5], the Boredom Proneness Scale Short Form (BPS-SR) [59], and the 30 Preliminary IPIP scales [17] measuring constructs similar to those in Cloninger's Temperament and Character Inventory (TCI) [8]. We also employ scales that measure general psychological characteristics, such as the Short 15-item Big Five Inventory (BFI-S) [39], the Gratitude Questionnaire (GQ-6) [37], the Revised Self-Monitoring Scale (RSMS) [41], the Satisfaction With Life Scale (SWLS) [13], and the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R) [14]. We gather responses to a total of 492 individual survey questions from each participant.

Diagnostic Considerations. There were a few difficulties we encountered due to some unique characteristics of adult ADHD diagnostic process.

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, the FP group had to be determined based on self-screened results from the ASRS. Due to the unobservable nature of the human mind, risks and uncertainties in diagnosing psychiatric disorders are fundamental and inevitable. Therefore, repeating and overturning a diagnosis, by its nature, is unconventional and should be an extremely careful process. We chose to classify the FP group by first administering the ASRS to obtain self-screened results and then employing diagnosis from the clinical interviews as the ground truth.

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. Neither one of the evaluators was aware of our research hypothesis.

Privacy Considerations & Data Management. As important as the quality of acquired data was participants' privacy since our sensing system inevitably collects sensitive data including mobile sensor data and psychological survey results. At the beginning of the study, participants were informed about details on what kind of sensor data will be collected and signed a consent form. 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. 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 on a daily basis. A message was sent via a chatbot to participants in cases of incompliance, 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.

4 PHASE 1: MOTIVATIONAL STUDY

We confirm our hypothesis by identifying the group of participants that report exaggerated symptoms of adult ADHD via formal diagnosis. We administer the Adult ADHD Self-Report Scale (ASRS) to the participants to obtain self-reported levels of ADHD symptoms and then perform formal diagnosis by trained experts. Our motivational study reveals 14 subtle malingerers, as only 3 out of the 17 self-screened participants were actually diagnosed with ADHD.

4.1 Self-Screening

4.1.1 The ASRS. The ASRS is a widely used self-report screening scale of adult ADHD developed by the WHO [31]. Though it is a widely accepted screening tool for formal diagnostic procedures, the assessment is easily subject to misjudgments as it relies heavily on subjective interpretations of the self. The ASRS includes 18 questions about the frequency of recent Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV) Criterion *A* symptoms of adult ADHD [3]. The ASRS screener consists of six Likert scale questions out of the 18 questions, and four or more marks above the set cutoffs indicate a need for further examinations. Compared to the unweighted 18-question ASRS, the unweighted ASRS screener showed better sensitivity (68.7% vs. 56.3%), specificity (99.5% vs. 98.3%), and total classification accuracy (97.9% vs. 96.2%). Also, the criteria for diagnosing ADHD in DSM-5 have been loosened in comparison to those in DSM-IV [49]. Since we conjectured that our participant group is likely to report exaggerated symptoms of ADHD, we use the ASRS screener based on DSV-IV Criteria to ensure more specific and accurate self-screening.

4.1.2 Self-screening Results. Participants were asked to respond to the ASRS twice across the 2-week interval of the study. Upon administration of the ASRS, almost half of the remaining participants (17 out of 37) were screened. Considering that the participants were recruited from a context where attention and impulse control is highly desirable, the results imply that many participants did in fact exaggerate their symptoms of ADHD. Also, despite the fact that the ASRS has a high specificity of 99.5%, the screening tool was not very effective in specifying clinical levels of adult ADHD in our participant group. We additionally administered the ASRS screener twice throughout our study to confirm that the results from the first time were 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 single rating, absolute-agreement, 2-way mixed-effects model [35]. The results indicated good reliability with an ICC of 0.75 (CI = .59-.86) [35].

4.2 Clinical Diagnosis

4.2.1 Diagnostic Procedure. 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 criteria for DSM disorders [3]. Based on the results of the interview, the psychiatrist made a final diagnosis along with 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.) [40]. 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 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.

4.2.2 Diagnostic Results. 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 excessive 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 tasks at hand, much more so than 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. (Note that the TP group in our study consists of three participants that showed clinical levels of hyperactivity-impulsivity.)

Most of the examples provided by the FP group in giving positive answers on the SCID-5 were evaluated as false evidences. 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 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"* suggests feelings of emptiness and boredom. Some participants in the FP group 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, or perfectionism, and procrastinating 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"*.

In our study, the diagnostic procedures were supervised by a school psychiatrist with a profound knowledge and expertise in not only adult ADHD diagnosis but also the student group. However, in most cases, clinicians are unaware of the demographics of the client. 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 [38]. Thus, the results from our motivational study provide a compelling case to confirm our hypothesis; *the inherent issue of subjectivity in using self-reported symptoms for ADHD diagnosis may contaminate the outcome of the ADHD diagnosis in certain populations.*

5 PHASE 2: PREDICTING THE SUBTLE MALINGERERS

5.1 Smartphone Sensing

In this paper, we introduce the use of smartphone sensing to predict subtle malingering in adult ADHD diagnosis. Though extant approaches are somewhat useful in assessing the validity of expressed symptoms, we propose that smartphone sensing is particularly useful in assessing adult ADHD for the following reasons.

Bias-free Assessment. Passive sensing data, coupled with the capability of smartphones to passively track behaviors for an extended period, could minimize the influence of bias or intentionality in adult ADHD diagnosis. Unlike the procedure of diagnosing ADHD in children, informant ratings are seldom used. Most diagnostic procedures, such as clinical interviews and the ASRS, rely on subjective interpretations of symptoms and behaviors. However, people often lose objectivity and lack insight into their behaviors. Existing assessment tools such as selected neuropsychological tests accounting for this issue lack merit, even with the problems of high cost and low diagnostic accuracy aside. Also, sensing data can be used to verify subjective reports of ADHD symptoms, as long-term tracking may nullify efforts to malinger the deficit. As such, smartphone sensing is particularly useful in providing an additional source of objectivity to the diagnostic process.

Diagnosing at Scale. To this day, conducting clinical interview and administering ADHD behavior rating scales are the essential parts of adult ADHD diagnosis, as stipulated in the American Academy of Pediatrics (AAP) guidelines. However, many psychiatrists lack specific expertise in diagnosing adult ADHD; as many as 66% of 400 primary care physicians (PCPs) in a study [42] admitted they have inadequate knowledge and training to diagnose ADHD. Hence, many individuals do not have access to diagnosis by PCPs with expertise in adult ADHD, including the knowledge on ADHD malingering and comorbid disorders. Our sensor-enabled framework could serve as an affordable, robust, and easily scalable ADHD detection system to assist evaluation and increase awareness of ADHD symptoms in the general population and PCPs likewise.

Long-term Passive Tracking. ADHD diagnosis requires a reasonably accurate and insightful retrospective self-assessment of one's behaviors over a lengthy time period. As such, longitudinal behavioral patterns captured using mobile sensors could complement treatment and diagnosis. According to the DSM-5 diagnostic criteria for adult ADHD, one must display persistence of symptoms for at least six months. Thus, the validity of assessment - even in the context of formal diagnosis - could easily be compromised by the poor recall of past experiences. Traditional daily tracking methods could be employed, but are impractical as they require active user input. Smartphones are embedded with multiple sensors that can capture a range of daily behaviors passively and are ubiquitous devices that people most frequently use. In that sense, there are few alternatives to the use of smartphones.

5.2 Sensing Features

5.2.1 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., call, app usage, notification, screen events, Bluetooth) and two types of sensing data (i.e., motion, location). The notification, motion, and screen data are stored upon every event (*e.g.*,

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 $\leq$ 3/5 seconds

[1] RSSI signal $\leq$ -60dBm
[2] computed based on [6]

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.

5.2.2 Feature Extraction. As little prior work has investigated the inner workings of subtle malingerers' psychology, we try to obtain a superset of features we could attract 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 call, motion (e.g., walking, vehicle ride), notification dismissal or click, or screen event (*i.e.,* screen on and phone unlock). Note that 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.

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:

(1) 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.

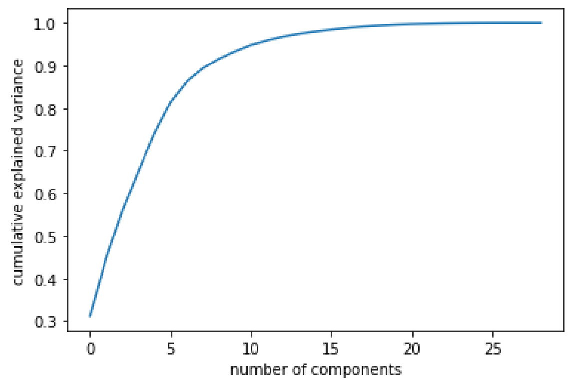

Fig. 3. The cumulative variability in the data explained by the top n principal components.

(2) To infer 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 [6], we merge identical geographic locations assigned with different coordinates or IDs.
(3) 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.
(4) 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.

5.3 Model and Prediction

5.3.1 Model Validation. In this study, we use behavioral features extracted from passive sensing data to predict subtle malingering of adult ADHD symptoms. We consider this as a binary classification problem of differentiating the subtle malingerers from the rest. The subtle 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 subtle 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.

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.

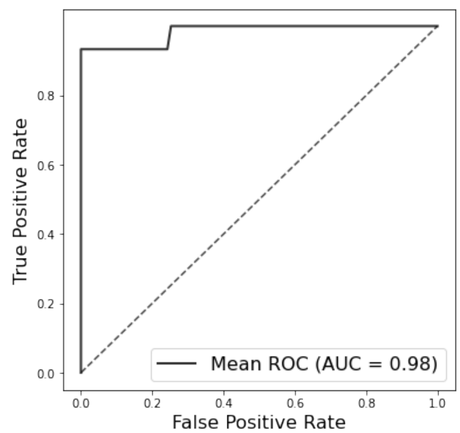

Fig. 4. Mean receiver operating characteristic (ROC) curves across 5 folds using different data

Using the 6 PCs selected, we train our model to predict the subtle 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 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.

5.3.2 *Classification 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. 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 classifying subtle malingerers from the rest of the participants. Figure 4 illustrates the mean ROC curve of the LR classifier performance across five folds.

6 PHASE 3: UNDERSTANDING OUR MODEL

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 self-assessed measurements on psychological scales - through a two-step analysis process. We first perform regression analysis using each sensing feature (rather than using all sensing features at once) to classify the subtle malingering group, then correlation analysis to understand the direction of association between the sensing features and the psychological scale data. Though mobile sensing offers some unique benefits such as the capability for ubiquitous and passive long-term tracking, sensing data only reflects behavioral patterns in an abstract manner. It is necessary that we estimate the interpretability of our models, as the diagnosis of adult ADHD or subtle 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 this analysis, we hope to gain a more comprehensive understanding of the implications our model has for the psychology of the subtle malingerers.

Table 3. Sensor features

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 $\leq$ 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

6.1 Regression Analysis

Here we aim to explore the relationship between sensing features and subtle malingering. Since the likelihood or extent of subtle 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. Then, 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

Table 4. Sensors associated with survey data

No.	Traits that predict subtle 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.

analysis in this section. We remove one feature and obtain 13 sensing features. Then, we train each of the 13 sensing features to classify participants into the FP group (*i.e.,* subtle 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 first use coefficients from the logistic regression models to understand how each of the 13 sensing features is associated with our model. The coefficients from 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 that the coefficients from the models can justifiably be used to interpret the associations between the sensing features and subtle malingering. Specifically, the coefficients signify the relationship each sensing feature has with the likelihood of subtle malingering; that is, a positive coefficient implies a higher likelihood of subtle malingering for higher magnitudes of the corresponding sensing feature and vice versa. In our analysis, coefficients of most features were negative with the exception of two features.

6.2 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 a 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 subtle 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 subtle malingering.

The implications of the remaining 21 survey features for subtle 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 subtle malingering. On the other hand, a higher likelihood of being classified as the subtle 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.

7 DISCUSSION

7.1 Understanding the Subtle Malingerers

The results from the regression and correlation analysis offer some insight into how various psychological traits relate to the likelihood of subtle malingering. However, classification of the subtle 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 analysis in Phase 3 are useful in understanding our model, we need to perform additional analysis to truly understand what the psychological surveys tell us about the subtle malingering group. Thus, we perform Student's t-test on the survey data between the subtle malingering group and other groups, such as the groups diagnosed with adult ADHD (*i.e.,* the true positive and false negative groups) and the true negative group (*i.e.,* those who correctly assessed themselves as not having adult ADHD).

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 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.

First, we compare the subtle malingering group against the groups that were diagnosed with ADHD. Of the 21 selected surveys, 7 scales yielded high effect sizes. 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.

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

Scale	**Subtle 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.

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

Scale	**Subtle 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.

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.

We also analyze the group mean difference between the TN group and the FP (*i.e.,* subtle malingering) groups. The results presented in Table 6 are those with statistically significant group mean differences. 12 of the 21 surveys revealed significant results. Effect sizes of all results were at least 0.76, indicating strong effects. The subtle 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.

Results from our analysis indicate that the subtle malingering group expresses lower levels of optimism and self-directedness in comparison to both those diagnosed with ADHD and the true negative group. The results are consistent with the interview results, as participants of the subtle malingering group showed a tendency to express higher levels of academic stress and feelings of emptiness than other groups. Moreover, the results tell us in detail the inner workings - in particular, how and why each psychological scale features from the correlation analysis in Phase 3 was associated with the likelihood of subtle malingering - of our analysis and features. For instance, lower levels of self-reported competence and flourishing were associated with a higher likelihood of classification into the subtle malingering group; results from our *t*-test indicate that this may be due to the fact that the subtle malingerers displayed lower levels of competence and flourishing in comparison to the ADHD and true negative groups, respectively.

7.2 Comparison to Conventional Screening Tools

This work presents a novel approach based on the smartphone sensing method to provide an alternative method for screening for subtle malingering of adult ADHD. Our study demonstrated that a prediction model using sensing features can outperform any other assessment methods developed to detect malingering. The best combinations of performance metrics of existing tools were sensitivity of .79-.86 and specificity of .89 [9], and sensitivity of .90 and specificity of .80 [52]. On the other hand, our model performed significantly better with sensitivity of .87 and specificity of .96. Such a sensor-enabled approach offers three additional benefits in assessing adult ADHD and the malingering of ADHD symptoms. First, smartphone sensing offers objective behavioral data. This is particularly useful as the diagnostic procedure for adult ADHD relies on subjective reports, which could easily be biased. College students in competitive environments, for instance, are prone to overexaggerate their symptoms as they have a skewed reference group to make objective judgments on ADHD symptoms. Second, mobile sensing is scalable. Access to qualified experts in diagnosing adult ADHD is limited. Models built with the diagnosis from experts could be useful for providing both self-screening and accurate diagnosis services at scale. Third, smartphones can passively collect data at all times for long periods of time. As the diagnosis of adult ADHD requires the persistence of symptoms for at least six months, the capabilities for long-term tracking is particularly useful.

7.3 Limitations and Future Work

Despite promising results, the findings of this study are subject to three major limitations. First, the sample size is relatively small due to procedural difficulties in determining the subtle 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 subtle malingerers from true positives. Second, all participants of our study are recruited from an Asian university with distinct academic culture. The exact 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 research could explicitly examine the feasibility of using behavioral sensing to screen intentional malingering of adult ADHD symptoms. In Phase 1, we administered the ASRS, a self-report assessment of adult ADHD symptoms formally acknowledged as a screening tool for clinical diagnosis. The scale had been validated extensively on a large population and has a sensitivity of 68.7%, specificity of 99.5%, and total classification accuracy of 97.9%. Despite the high specificity and accuracy, 17 out of the 37 participants in our study screened themselves as positive. To alleviate such problems, we could obtain reports from observers or use objective data such as smartphone sensing data. Various time windows could be employed to investigate the minimal duration of smartphone tracking required to identify intentional malingering. Moreover, future work could also assess sensing feature-based models that predict adult ADHD diagnosis, by obtaining data from a larger sample size with more participants diagnosed with ADHD. That way, a single sensor-enabled framework could robustly and objectively predict the diagnosis and malingering of adult ADHD.

8 CONCLUSION

We believe that this is the first work that explores the application of a smartphone sensing method to screen subtle malingering of adult ADHD. We proposed and rigorously evaluated the feasibility of using behavioral sensing data to predict adult ADHD subtle malingering. First, we conducted clinical diagnosis on 37 participants and identified 14 false positives, those who indicated significant levels of ADHD symptoms but were not diagnosed with ADHD, as the subtle malingerers. Then, we extract sensing features gathered from our 3-week-long user study and build a prediction model using the 6 PCs that we obtained. In comparison to existing tools for screening subtle malingering, our model demonstrates superior classification performance in terms of accuracy (.92), specificity (.96), and AUC (.98). We also assess the implications of the predictive performance of our model using a wide range of psychological scales. We envision that our sensing-based assessment could be used to establish a robust, cost-effective, and scalable solution to prevent overdiagnosis of adult ADHD.

REFERENCES

[1] Saeed Abdullah and Tanzeem Choudhury. 2018. Sensing Technologies for Monitoring Serious Mental Illnesses. *IEEE Multimedia* 25, 1 (2018), 61–75. https://doi.org/10.1109/MMUL.2018.011921236

[2] Saeed Abdullah, Mark Matthews, Ellen Frank, Gavin Doherty, Geri Gay, and Tanzeem Choudhury. 2016. Automatic detection of social rhythms in bipolar disorder. *Journal of the American Medical Informatics Association* 23, 3 (2016), 538–543. https://doi.org/10.1093/jamia/ocv200

[3] 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.

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

[5] Fred B Bryant and Bruce D Smith. 2001. Refining the architecture of aggression: a measurement model for the buss–perry aggression questionnaire. *Journal of Research in Personality* 35, 2 (jun 2001), 138–167. https://doi.org/10.1006/jrpe.2000.2302

[6] 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

[7] 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

[8] C Robert Cloninger (Ed.). 1994. *The temperament and character inventory (Tci): a guide to its development and use* (1st ed ed.). Center for Psychobiology of Personality, Washington University, St. Louis, Mo.

[9] Seth C. Courrégé, Reid L. Skeel, Abigail H. Feder, and Kaley S. Boress. 2019. The ADHD symptom infrequency scale (ASIS): A novel measure designed to detect adult ADHD simulators. *Psychological Assessment* 31, 7 (2019), 851–860. https://doi.org/10.1037/pas0000706

[10] 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

[11] Alex W. DaSilva, Jeremy F. Huckins, Rui Wang, Weichen Wang, Dylan D. Wagner, and Andrew T. Campbell. 2019. Correlates of stress in the college environment uncovered by the application of penalized generalized estimating equations to mobile sensing data. *Journal of Medical Internet Research* 21, 3 (2019), 1–9. https://doi.org/10.2196/12084

[12] 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

[13] ED Diener, Robert A Emmons, Randy J Larsen, and Sharon Griffin. 1985. The satisfaction with life scale. *Journal of personality assessment* 49, 1 (1985), 71–75.

[14] Greg Feldman, Adele Hayes, Sameet Kumar, Jeff Greeson, and Jean-Philippe Laurenceau. 2007. Mindfulness and emotion regulation: the development and initial validation of the cognitive and affective mindfulness scale-revised(Cams-r). *Journal of Psychopathology and Behavioral Assessment* 29, 3 (July 2007), 177–190. https://doi.org/10.1007/s10862-006-9035-8

[15] 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

[16] Martin Gjoreski, Hristijan Gjoreski, Mitja Lutrek, and Matja Gams. 2015. Automatic detection of perceived stress in campus students using smartphones. In *2015 International Conference on Intelligent Environments*. IEEE, Prague, Czech Republic, 132–135. https://doi.org/10.1109/IE.2015.27

[17] Lewis R. Goldberg, John A. Johnson, Herbert W. Eber, Robert Hogan, Michael C. Ashton, C. Robert Cloninger, and Harrison G. Gough. 2006. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality* 40, 1 (Feb. 2006), 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

[18] Alexandra F. Grant, John W. Lace, Carson L. Teague, Kimberly T. Lowell, Phillip D. Ruppert, Annie A. Garner, and Jeffrey D. Gfeller. 2020. Detecting feigned symptoms of depression, anxiety, and ADHD, in college students with the structured inventory of malingered symptomatology. *Applied Neuropsychology:Adult* 0, 0 (2020), 1–9. https://doi.org/10.1080/23279095.2020.1769097

[19] Agnes Grünerbl, Amir Muaremi, Venet Osmani, Gernot Bahle, Stefan Öhler, Gerhard Tröster, Oscar Mayora, Christian Haring, and Paul Lukowicz. 2015. Smartphone-based recognition of states and state changes in bipolar disorder patients. *IEEE Journal of Biomedical and Health Informatics* 19, 1 (2015), 140–148. https://doi.org/10.1109/JBHI.2014.2343154

[20] Jordan P. Harp, Lindsey J. Jasinski, Anne L. Shandera-Ochsner, Lisa H. Mason, and David T.R. Berry. 2011. Detection of Malingered ADHD Using the MMPI-2-RF. *Psychological Injury and Law* 4, 1 (2011), 32–43. https://doi.org/10.1007/s12207-011-9100-9

[21] 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

[22] Allyson G. Harrison and Irene T. Armstrong. 2016. Development of a symptom validity index to assist in identifying ADHD symptom exaggeration or feigning. *Clinical Neuropsychologist* 30, 2 (2016), 265–283. https://doi.org/10.1080/13854046.2016.1154188

[23] Allyson G. Harrison, A. Lynne Beal, and Irene T. Armstrong. 2021. Predictive value of performance validity testing and symptom validity testing in psychoeducational assessment. *Applied Neuropsychology: Adult* 0, 0 (2021), 1–15. https://doi.org/10.1080/23279095.2021.1943396 arXiv:https://doi.org/10.1080/23279095.2021.1943396 PMID: 34261385.

[24] 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

[25] Karen J. Klingman, Carla R. Jungquist, and Michael L. Perlis. 2017. Introducing the Sleep Disorders Symptom Checklist-25: A Primary Care Friendly and Comprehensive Screener for Sleep Disorders. *Sleep Medicine Research* 8, 1 (2017), 17–25. https://doi.org/10.17241/smr.2017.00010

[26] 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

[27] 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

[28] Lindsey J Jasinski and John D Ranseen. 2011. Malingered ADHD evaluations: A further complication for accommodations reviews. *The Bar Examiner* 79 (2011), 6–16.

[29] PETER S JENSEN, DAVID MARTIN, and DENNIS P CANTWELL. 1997. Comorbidity in ADHD: Implications for Research, Practice, and DSM-V. *Journal of the American Academy of Child and Adolescent Psychiatry* 36, 8 (1997), 1065–1079.

[30] Ellen E.H. Johnson and Julie Suhr. 2021. Self-reported functional impairment in college students: relationship to noncredible reporting, ADHD, psychological disorders, and other psychological factors. *Journal of Clinical and Experimental Neuropsychology* 00, 00 (2021), 1–13. https://doi.org/10.1080/13803395.2021.1935490

[31] 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

[32] Sena Kim, Dumi Pyo, Jisun Lee, Jungi Lee, Junho Min, Kanghyeon Shin, and Kyungil Kim. 2015. A Study of The Validity of Korean Versions of Regulatory Focus Scale. *Korean Journal of Social and Personality Psychology* 29, 3 (2015), 85–110.

[33] 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

[34] Saskia Koldijk, Mark A. Neerincx, and Wessel Kraaij. 2018. Detecting Work Stress in Offices by Combining Unobtrusive Sensors. *IEEE Transactions on Affective Computing* 9, 2 (2018), 227–239. https://doi.org/10.1109/TAFFC.2016.2610975

[35] 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

[36] Kurt Kroenke, Tara W. Strine, Robert L. Spitzer, Janet B.W. Williams, Joyce T. Berry, and Ali H. Mokdad. 2009. The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders* 114, 1-3 (2009), 163–173. https://doi.org/10.1016/j.jad.2008.06.026

[37] Sunjung Kwon, Kyoheon Kim, and Hongseock Lee. 2006. Validation of the Korean Version of Gratitude Questionnaire. *Korean Journal of Health Psychology* 11, 1 (2006), 177–190.

[38] 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

[39] Frieder R. Lang, Dennis John, Oliver Lüdtke, Jürgen Schupp, and Gert G. Wagner. 2011. Short assessment of the Big Five: robust across survey methods except telephone interviewing. *Behavior Research Methods* 43, 2 (June 2011), 548–567. https://doi.org/10.3758/s13428-011-0066-z

[40] 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

[41] Richard D. Lennox and Raymond N. Wolfe. 1984. Revision of the self-monitoring scale. *Journal of Personality and Social Psychology* 46, 6 (1984), 1349–1364. https://doi.org/10.1037/0022-3514.46.6.1349

[42] 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

[43] Ree M. Meertens and René Lion. 2008. Measuring an Individual's Tendency to Take Risks: The Risk Propensity Scale1. *Journal of Applied Social Psychology* 38, 6 (2008), 1506–1520. https://doi.org/10.1111/j.1559-1816.2008.00357.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1559-1816.2008.00357.x

[44] 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

[45] David C. Mohr, Mi Zhang, and Stephen M. Schueller. 2017. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology* 13, 1 (2017), 23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949

[46] Elizabeth L. Murnane, Tara G. Walker, Beck Tench, Stephen Voida, and Jaime Snyder. 2018. Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–27. https://doi.org/10.1145/3274396

[47] Mandi W. Musso and Wm Drew Gouvier. 2014. "Why Is This So Hard?" A Review of Detection of Malingered ADHD in College Students. *Journal of Attention Disorders* 18, 3 (2014), 186–201. https://doi.org/10.1177/1087054712441970

[48] 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

[49] 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

[50] Jukka Pekka Onnela and Scott L. Rauch. 2016. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology* 41, 7 (2016), 1691–1696. https://doi.org/10.1038/npp.2016.7

[51] 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

[52] Sujith Ramachandran, Erin R. Holmes, Meagen Rosenthal, Benjamin F. Banahan, John Young, and John P. Bentley. 2019. Development of the Subtle ADHD Malingering Screener. *Assessment* 26, 3 (2019), 524–534. https://doi.org/10.1177/1073191118773881

[53] Darius A Rohani, Maria Faurholt-Jepsen, Lars Vedel Kessing, and Jakob E Bardram. 2018. Correlations between objective behavioral features collected from mobile and wearable devices and depressive mood symptoms in patients with affective disorders: systematic review. *JMIR mHealth and uHealth* 6, 8 (Aug. 2018), e165. https://doi.org/10.2196/mhealth.9691

[54] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of Medical Internet Research* 17, 7 (July 2015), e175. https://doi.org/10.2196/jmir.4273

[55] Fuschia M. Sirois, Sisi Yang, and Wendelien van Eerde. 2019. Development and validation of the General Procrastination Scale (GPS-9): A short and reliable measure of trait procrastination. *Personality and Individual Differences* 146 (Aug. 2019), 26–33. https://doi.org/10.1016/j.paid.2019.03.039

[56] 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

[57] 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

[58] Robert L. Spitzer, Kurt Kroenke, Janet B. W. Williams, and Bernd Löwe. 2006. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine* 166, 10 (05 2006), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092 arXiv:https://jamanetwork.com/journals/jamainternalmedicine/articlepdf/410326/ioi60000.pdf

[59] Andriy A. Struk, Jonathan S.A. Carriere, J. Allan Cheyne, and James Danckert. 2017. A Short Boredom Proneness Scale: Development and Psychometric Properties. *Assessment* 24, 3 (2017), 346–359. https://doi.org/10.1177/1073191115609996

[60] Julie Suhr, Dustin Hammers, Katy Dobbins-Buckland, Eric Zimak, and Carrie Hughes. 2008. The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology* 23, 5 (2008), 521–530. https://doi.org/10.1016/j.acn.2008.05.003

[61] Brian K. Sullivan, Kim May, and Lynne Galbally. 2007. Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology* 14, 3 (2007), 189–207. https://doi.org/10.1080/09084280701509083

[62] J. Torous, J. P. Onnela, and M. Keshavan. 2017. New dimensions and new tools to realize the potential of RDoC: Digital phenotyping via smartphones and connected devices. *Translational Psychiatry* 7, 3 (2017), 2–4. https://doi.org/10.1038/tp.2017.25

[63] Heeryung Wang, Sangeok Kim, Shihyun Kang, Yeonho Joo, and Changyoon Kim. 2008. The Validation Study of the Korean Version of the Bipolar Spectrum Diagnostic Scale. *Journal of Korean Neuropsychiatric Association* 47, 6 (2008), 533–539.

[64] Rui Wang, Min S. H. Aung, Saeed Abdullah, Rachel Brian, Andrew T. Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, Michael Merrill, Emily A. Scherer, Vincent W. S. Tseng, and Dror Ben-Zeev. 2016. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Heidelberg Germany, 886–897. https://doi.org/10.1145/2971648.2971740

[65] Rui Wang, Weichen Wang, Alex DaSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26. https://doi.org/10.1145/3191775

[66] Rui Wang, Weichen Wang, Mikio Obuchi, Emily Scherer, Rachel Brian, Dror Ben-Zeev, Tanzeem Choudhury, John Kane, Martar Hauser, Megan Walsh, and Andrew Campbell. 2020. On predicting relapse in schizophrenia using mobile sensing in a randomized control trial. In *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, Austin, TX, USA, 1–8. https://doi.org/10.1109/PerCom45495.2020.9127365

[67] 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

[68] Kimberly D. Williamson, Hannah L. Combs, David T.R. Berry, Jordan P. Harp, Lisa H. Mason, and Maryanne Edmundson. 2014. Discriminating among ADHD alone, ADHD with a comorbid psychological disorder, and feigned ADHD in a college sample. *Clinical Neuropsychologist* 28, 7 (2014), 1182–1196. https://doi.org/10.1080/13854046.2014.956674

[69] 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

[70] Camellia Zakaria, Rajesh Balan, and Youngki Lee. 2019. StressMon. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–29. https://doi.org/10.1145/3359139

[71] Kamil Zyła. 2019. Attention deficit hyperactivity disorder detection-from psychological checklists to mobile solutions. *Studies in Logic, Grammar and Rhetoric* 60, 1 (2019), 85–100. https://doi.org/10.2478/slgr-2019-0047