



Published in final edited form as:

*Future Gener Comput Syst.* 2022 July ; 132: 266–281. doi:10.1016/j.future.2022.02.010.

## Exploratory Machine Learning Modeling of Adaptive and Maladaptive Personality Traits from Passively Sensed Behavior

Runze Yan<sup>a</sup>, Whitney R. Ringwald<sup>b</sup>, Julio Vega Hernandez<sup>c</sup>, Madeline Kehl<sup>b</sup>, Sang Won Bae<sup>d</sup>, Anind K. Dey<sup>e</sup>, Carissa Low<sup>c</sup>, Aidan G.C. Wright<sup>b</sup>, Afsaneh Doryab<sup>a</sup>

<sup>a</sup>School of Engineering and Applied Science, University of Virginia, United States

<sup>b</sup>Department of Psychology, University of Pittsburgh, United States

<sup>c</sup>Mobile Sensing + Health Institute, Center for Behavioral Health, Media, and Technology, University of Pittsburgh, United States

<sup>d</sup>School of Systems and Enterprises, Stevens Institute of Technology, United States

<sup>e</sup>Information School, University of Washington, United States

### Abstract

Continuous passive sensing of daily behavior from mobile devices has the potential to identify behavioral patterns associated with different aspects of human characteristics. This paper presents novel analytic approaches to extract and understand these behavioral patterns and their impact on predicting adaptive and maladaptive personality traits. Our machine learning analysis extends previous research by showing that both adaptive and maladaptive traits are associated with passively sensed behavior providing initial evidence for the utility of this type of data to study personality and its pathology. The analysis also suggests directions for future confirmatory studies

---

\* Corresponding author: ry4jr@virginia.edu (Runze Yan).

Runze Yan

Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing original draft, Review & editing, Visualization

Whitney R. Ringwald

Conceptualization, Methodology, Validation, Formal analysis, Formal analysis, Data Curation, Writing original draft, Review & editing

Julio Vega Hernandez

Data Curation

Madeline Kehl

Data Curation, Writing original draft

Sang Won Bae

Investigation

Anind K. Dey

Conceptualization

Carissa A. Low

Conceptualization, Investigation

Aidan G.C. Wright

Conceptualization, Methodology, Validation, Investigation, Writing original draft, Review & editing, Supervision, Funding acquisition

Afsaneh Doryab

Conceptualization, Methodology, Validation, Formal analysis, Writing original draft, Review & editing, Supervision, Funding acquisition

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

No conflict of interest existed in the study

into the underlying behavior patterns that link adaptive and maladaptive variants consistent with contemporary models of personality pathology.

## Keywords

Mobile and Wearable Sensing; Machine Learning; Data Mining; Behavior Modeling; Personality Prediction

---

## 1. Introduction

Personality refers to an individual's typical patterns of motivations, thoughts, feelings, and behavior that serve the functional roles of adaptively navigating the environment [18, 71]. As such, personality pathology (i.e., personality disorder) refers to typical patterns of motivations, thoughts, feelings, and behavior that are persistently and pervasively maladaptive, such that they result in dysfunctional patterns of day-to-day behavior [48]. These maladaptive patterns of behavior are thought to account for associations between personality pathology and poor psychosocial and physical health outcomes [19, 46, 69]. Clinical psychology and psychiatry are undergoing a paradigm shift from classifying personality pathology using a finite set of discrete categorical diagnoses to continuously distributed dimensional traits, and growing evidence suggests traits are more valid, reliable, and clinically useful [29]. Despite these recent advances, little is known about the specific, everyday, behavioral expressions of pathological traits that may contribute to poor psychosocial functioning and outcomes.

Most research on dimensional models has relied on cross-sectional, global self-report data, which are not well-suited to measure the dynamic moment-to-moment processes underlying personality and its pathology [24, 70]. Although the use of ecological momentary assessment (EMA) and intensive longitudinal studies to collect data from people in their naturalistic settings has remedied some of these issues, heavy reliance on self reports, which can be prone to bias and lack of engagement, can affect the reliability and validity of the collected data [58]. Passive sensing via a participant's smartphone or wearable devices has the potential to address these concerns through continuous and contextualized measures of behavior [40]. Smartphone sensing combined with machine learning has been used to predict personality traits from smartphone devices [27]. However, to our knowledge, no existing research has investigated the associations between passive sensing and maladaptive personality traits or compared adaptive and maladaptive traits by extracting the shared and unique behavioral patterns associated with each of them.

## 2. Background and Related Work

### 2.1. Adaptive vs. Maladaptive Personality Traits

Traditional psychiatric classifications, as reflected in the American Psychiatric Association's Diagnostic and Statistical Manual for Mental Disorders [2] and the World Health Organization's International Classification of Disease (ICD), have conceptualized personality pathology categorically, under diagnoses such as borderline personality disorder,

schizoid personality disorder, and narcissistic personality disorder. Each diagnosis describes a “type” that is based on meeting a certain number of criteria from a checklist of symptoms. However, categorical personality disorder diagnoses have critical weaknesses, including poor validity, low reliability, and limited clinical utility [57, 68]. Such issues have prompted a shift towards dimensional trait models of personality pathology that outperform categorical models psychometrically and may have greater clinical utility [34].

The dimensional approach is further supported by empirical and conceptual parallels between these maladaptive traits and the trait dimensions that have emerged from over a century of studying adaptive personality structure [38, 72]. Indeed, for each of the big five traits, personality’s most prominent model, maladaptive traits have been identified that can be understood to cover the same behavioral content, albeit often keyed towards the opposite direction. See Figure 1 for elaboration of these traits and their descriptions. For example, whereas adaptive personality research finds Extraversion, which manifests in sociability, assertiveness, positive emotions, and energy, the maladaptive variant emphasizes Detachment which manifests in social withdrawal, passivity, lack of positive emotions, and lethargy. Correspondence between these sets of traits suggests that instead of maladaptive personality having discrete principles and properties that differentiate it from normal range personality (i.e., being a separate category), the two operate along shared underlying dimensions or spectra. In other words, while there is a continuum of functioning from adaptive to maladaptive, the two share the same fundamental traits. In recent years, variations on trait-based models of personality disorder classification and diagnosis have been included in the DSM-5 (provisionally, as an alternative approach in need of more study [66]) and the ICD-11 (officially [44]). In fact, it has been argued that maladaptive trait models can serve as useful structures for organizing all of psychopathology [33].

## 2.2. Assessing Personality Traits from Self Reports

What is thought to link an individual’s personality to their level of psychosocial functioning and health are the dynamic and contextualized patterns of behavior they engage in. For example, a person who scores high on trait Detachment may be less motivated to socialize, so they may contact fewer people or go out in public less during the day than most people, which in turn contributes to poor relationships and low mood. Yet, most research on dimensional models has relied on cross-sectional, global self-report data, which are not well-suited to measure the dynamic moment-to-moment processes underlying personality and its pathology [24, 70]. Cross-sectional data is unable to capture the temporal aspect of maladaptive behaviors that unfold from moment-to-moment. To a large extent, issues with cross-sectional data can be remedied with intensive longitudinal study designs, such as ecological momentary assessment (EMA). EMA (sometimes referred to as experience sampling methodology or ambulatory assessment) can be used to assess people intensively and repeatedly in their naturalistic settings, thereby increasing ecological validity, reducing biases of retrospection, and capturing multiple points along an unfolding dynamic process. Each of these address the limitations of traditional global self-report scales. However, EMA typically uses self-report data which relies on the person being conscious of the behaviour and willing to report it—but many of the maladaptive patterns people engage in may be outside of awareness [11].

### 2.3. Assessing Personality Traits from Passively Sensed Behavior

Because self-report EMA study designs require participants to fill out multiple surveys per day or week, the burden to participants' in terms of time and effort limits the number of data points that can feasibly be collected throughout the day [58]. Passive sensing, or automatic data collection via a participant's smartphone or other device, is a methodological approach that has the potential to address these concerns [37]. It can provide continuous, contextualized, temporally sensitive, and direct measures of behavior (not reliant on participant report). Such an approach collects personality relevant information like frequency of texting and phone calls, average time spent in each location, and number of bluetooth devices scanned, which can be aggregated into higher order variables related to psychological constructs [21, 40, 64]. These nearly unbroken streams of high-resolution data on what people actually do during the day (not just what they are aware of) have the potential to offer new insights into maladaptive patterns linked to personality pathology.

One challenge associated with passive sensing, is that the type of data it produces, characterized by large numbers of variables, often overlapping, with unknown associations with behavioral outcomes, is not well suited to the statistical models traditionally applied in psychology research (i.e., mostly based on the general linear model). Such models often have a core assumption of relatively modest associations among predictors, can practically only handle a few variables, and assume a linear relationship between traits and behaviors that is often not met statistically or expected theoretically. In contrast machine learning has a number of tools available that can address these concerns. Decision-tree based regression algorithms such as XGBoost and random forest overcome the limitations of traditional analyses [9, 12]. Additionally the quantity of features derived from passive sensing cannot reasonably be included in traditional statistical analyses, especially since the number of features can often exceed the number of observations [55]. Decision trees can account for correlations among vast quantities of variables to identify predictive behaviors, and are also reported to be more stable than a typical step-wise approach used for variable selection in traditional analyses [39, 55]. Finally random forests are non-parametric and can be used to detect complex, non-linear associations [9, 39]. Machine learning can also be used to explore certain combinations of passive sensor features that are particularly predictive of certain trait levels. Characteristic combinations of daily behaviors are undetectable with covariance-based models in psychology or typical "black box" classification algorithms, but these combinations may reveal more nuanced patterns that contribute to a person's health and well-being.

Prior research provides initial evidence that passive sensing can provide meaningful information about individual differences in adaptive and maladaptive functioning. Studies have shown that attributes and events related to categorical DSM diagnoses (e.g., depression, psychotic episodes [6, 10, 47], and adaptive personality traits [1, 5, 14, 26, 27, 41, 54, 51] can be predicted from smartphone sensors from some degree, although these associations have been modest in size. Some machine learning methods have been used in previous research to predict personality including support vector classifiers, regularized regression models, and decision tree algorithms, but our study is the first to our knowledge to apply association mining to identify feature combinations linked to personality traits.

A few studies have investigated associations between heart rate measured by wearable devices and aspects of personality like emotional responses [22, 45, 17], but none have examined how personality traits relate to ambulatory heart rate. Given the transition towards conceptualizing personality pathology as dimensional variants of adaptive personality traits, linking smartphone sensor features to empirically-derived adaptive and maladaptive trait models has the potential to deepen our understanding of how personality shapes psychosocial functioning and important outcomes.

#### 2.4. Modeling behavior at the person-level and day-level

A key consideration for modeling behavior from any time-series data, including smartphone sensors, is what level of analysis to focus on. Psychological research typically uses *person-level* models that measures personality-related tendencies with a single index, most often the mean, that summarizes a person's behavior over time and across situations. Another approach is to include each repeated observation (e.g., daily samplings) in the analysis into an observation or *day-level* model. Day-level models capture behaviors that tend to characterize a given day for someone with a certain trait standing. Importantly, sets of behavior that tend to occur over time and across situations (person-level) are not necessarily expressed together on a given day (day-level). For example, more disinhibited people have more problems with concentration and engage in more impulsive behaviors overall, but they do not necessarily struggle with concentration more and act more impulsively on the same day. Thus, both levels of analysis offer distinct insights into adaptive and maladaptive personality. In addition to operationalizing behavior differently, each approach comes with different methodological trade offs. Person-level models will have far fewer observations than day-level models which can negatively impact the performance of machine learning models. Day-level models, however, do not account for differences in the number of observations for each person (e.g., number of days with complete data) which can bias the associations between traits and behaviors if they vary widely across individuals. There are multi-level, covariance-based models that can accommodate this nested data structure (i.e., observations within persons) to simultaneously preserve day-to-day variation and eliminate bias, but there are not currently comparable machine learning techniques. Thus, an important decision for studies using machine learning to predict personality pathology from passive sensing data is whether to model behavior at the person- or day-level.

#### 2.5. Current Study

The current study leverages a combination of passive sensing, machine learning, and data mining algorithms to identify patterns of behavior across the main spectra of adaptive and maladaptive personality traits. We also compared results from person-level and day-level models to help inform analytic decisions in future research. This study is strictly exploratory and we did not set out to test specific hypotheses; rather, we aimed to highlight key methodological issues for the field moving forward, generate new hypotheses about the everyday behavior patterns that maintain (or prevent) problems, and demonstrate the opportunities for these methods to enrich existing empirical models of personality pathology.

### 3. Methods

To explore the potential of passive sensing to reveal the dynamics of personality pathology, we designed and employed a series of analytical methods, including prediction and association rule mining. These methods allow us to 1) identify and rank behavioral features related to each personality dimension and 2) measure the level of the behavioral features associated with those dimensions. The following sections first provide an overview of our data processing pipeline followed by a description of the analytic methods.

#### 3.1. Data Processing

**3.1.1. Feature extraction**—To extract behavioral features from raw sensor data, we first divided each data stream into daily intervals (12:00am - 11:59pm). We aggregated the raw data and extracted different statistical measures such as minimum, maximum, mean, standard deviation as well as more complex behavioral features such as movement patterns and type, and duration of activities. We used RAPIDS, an open-source framework for extracting day-level behavioral features from mobile and wearable devices [62]. Details of the extracted features are documented in [20].

To create person-level features, we calculated the mean, minimum, maximum, and standard deviation of each individual's daily features. The test-retest reliability of the person-level features from week one to week two was acceptable (mean  $r = .53$ , median  $r = .58$ ). Full reliability results are in the supplementary materials.

**3.1.2. Handling missing values**—Datasets gathered from mobile sensor devices often suffer from a significant fraction of missing data due to issues such as poor communication, power depletion, and hardware failure. The missing rate across all 658 day-level features in our dataset was 58.2%.

For handling the missing values in day-level features, we took a two-step approach. The first step was removing participants and features with excessive missing data. If all sensor data from a participant on a certain day was missing, we deleted the participant's data for that day. We then removed features with a missing rate over 60%. This percentage was chosen based on the average proportion of missing values among all feature columns (57.2%). We also removed features with values outside of plausible ranges (e.g., negative value for distance travelled in a day). A total of 348 features were removed due to excessive missingness. The second step was imputing the remaining missing values. Because of the large blocks of missing data, we used multiple imputation techniques to increase the precision of missing data estimates [49]. Multiple imputation techniques can only work under the assumption that the data are missing at random (MAR) or the more stringent missing completely at random (MCAR). We believe that the MCAR assumption fits our data because sensor readings can fail randomly due to many different and non-systematic factors, including network error, faulty sensor, and phone battery issues [7]. We used scikit-learn with multiple imputation<sup>1</sup> and first implemented single imputation methods resulting in multiple complete datasets. Then, we trained the ML algorithms on each of those complete

---

<sup>1</sup> <https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html>

dataset to build models. Finally, we pooled the results by averaging the estimates from each model by computing the total variance over the repeated analyses, and estimating the average error. To avoid the data leakage in the leave-one-out cross-validation (LOOCV) stage, we imputed the training and testing set separately in each iteration of LOOCV as shown in Figure 2.

Only the non-missing day-level features were used to derive person-level features. Before calculating those features, we removed participants with less than three days of data ( $n = 14$ ). After calculating person-level descriptive statistics from each participant's data, there were 2632 features with a missing rate of 24.4%. After removing features with more than 40% missing values, 1664 person-level features remained. The percentage was chosen based on the average proportion of missing rates among all features (47.1%). Because there was considerably fewer missing values in the person-level features compared to the day-level, we used a single imputation method. Specifically, we imputed the missing features with a constant value smaller than the minimum value of each feature, i.e.,  $min - 0.01$ . This imputation method is appropriate for our data as it has a low effect on the center and scale of the observable values after the min-max normalization, i.e., the missing values become 0 after normalization and the respective features will be ignored by the machine learning algorithm [12].

### 3.2. Analysis Pipeline

To investigate passive sensing capabilities in predicting adaptive and maladaptive traits, we developed an analysis pipeline including machine learning, feature ranking, and association rule mining.

**3.2.1. Machine learning**—The structure of our machine learning pipeline is shown in Fig 2. The pipeline is built on a nested cross-validation, including an outer loop and an inner loop. The outer loop is a leave-one-participant-out cross-validation, which provides an estimate of the model's performance. The inner loop is a 5-fold Grid Search cross-validation used for parameter tuning. Both day-level and person-level machine learning models share the same general pipeline.

We used adaptive and maladaptive trait measures from the self-reports as ground truth for performance analysis of the machine learning algorithms. We compare the predictive performance of Lasso regression (Lasso), Random Forest (RF), XGBoost (XGB) and baseline models. Random Forest and XGBoost were chosen for modeling because of their ability to model potential nonlinear associations between behaviors and continuous personality trait values. Both methods use regression trees as base learners, but they have different strategies to prevent overfitting and performance improvement. RF uses Bagging (Bootstrap Aggregation) to reduce performance variance by first creating subsamples of the training set and then averaging the model performance resulted from each subsample [8, 52]. RF has a certain degree of randomness, which helps avoid overfitting and improves generalizability relative to a single tree model. XGB, on the other hand, uses an iterative learning strategy called Boosting that weighs the model outcomes based on the prediction results of the previous iteration [12]. The error is reduced in each iteration by giving correct

predictions lower weights and incorrect predictions higher weights. We used these two complementary models to reduce prediction errors and to build models that support a stable hypothesis generation process for further analyses. Both methods output feature importance as part of the model building process. In a nested cross-validation process, we selected the highly ranked features in each iteration and used them to identify the final set of behavioral features that distinguish personality traits. As a basis of comparison for the performance of the nonlinear, tree-based machine learning models, we evaluated the performance of a linear Lasso regression. Lasso is a modification of linear regression with penalty term to help in reducing overfitting. The Lasso regression, however, does not output the feature importance like RF and XGB. We aggregated the beta coefficients from Lasso regression to measure the linear correlation between input sensor features and personality within the nested cross-validation. We applied scikit-learn implementation<sup>2</sup> for Lasso and RF, and a public python implementation for XGB<sup>3</sup>. In addition, we built the baseline model to predict the mean target traits without taking any features into account using DummyRegressor class in scikit-learn<sup>4</sup>.

**3.2.2. Feature selection**—The large number of extracted features compared to the relatively small number of participants in our dataset required dimensionality reduction through feature selection. Our feature selection proceeded in two steps. First, we identified redundant features by calculating the Pearson correlation between all feature pairs. Correlation coefficient is an indicator of the linearity strength in the relationship between two features, and as such, a high correlation between two features indicates they will have nearly the same effect on the dependent variable. For our analyses, we considered correlations above .50 to indicate possible linear dependence. From each highly correlated pair of features, we kept the feature that correlated more strongly with the dependent variable (adaptive and maladaptive traits) and dropped the feature with lower correlation.

**3.2.3. Validation**—We applied leave-one-person-out cross-validation (LOOCV) to our dataset to evaluate the power of sensor features to predict personality traits [63]. LOOCV is a specific type of k-fold cross-validation, where the number of folds is equal to the number of participants in our dataset. In each iteration, data from n-1 participants is used for training, and the model is tested on the nth participant. LOOCV is a robust way to test the generalizability of a model that contains personal data. However, LOOCV is also the most computationally expensive, especially given that the data imputation and feature selection process is within the cycle of LOOCV [32].

**3.2.4. Parameter tuning**—We used Grid Search cross-validation to optimize the hyperparameters for machine learning models. We applied the Grid Search within each iteration of the LOOCV process. The training set of the LOOCV was split into the training set and validation set in the Grid Search CV. The Grid Search method tries every possible combination of hyperparameters from a given parameter space and selects the set of parameters that provides the best performance on the training set.

---

<sup>2</sup> <https://scikit-learn.org/stable/index.html>

<sup>3</sup> <https://xgboost.readthedocs.io/en/stable/index.html>

<sup>4</sup> <https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyRegressor.html>



**3.2.5. Feature ranking**—Each iteration in the LOOCV process assigns an importance value to each feature in the training process. A higher importance value for a feature indicates a larger contribution to the prediction process. We logged these values during training and aggregated the overall ranks of each feature at the end of the process. Sensor features were also ranked according to their frequency of being selected as well as their averaged importance across rounds in the LOOCV. For each training algorithm (RF or XGB), we considered features with frequency and importance greater than the median to be high ranked features selected by that algorithm. We then selected the highly-ranked features that were shared across both algorithms. Those features were then ranked according to the following formulas:

$$f(Imp_i) = \begin{cases} 1 & \text{if } Imp_i > 0; \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

$$RS = \sum_{i=1}^N \frac{(Imp_{RF_i} \sum_{i=1}^N f(Imp_{RF_i}) + Imp_{XGB_i} \sum_{i=1}^N f(Imp_{XGB_i}))}{N} \quad (2)$$

where  $Imp$  is the importance value of the sensor feature, and  $N$  is the number of iteration in the LOOCV, which equals the number of participants.

**3.2.6. Performance Measures**—To measure the performance of the models, we used the mean squared error (MSE) and the mean absolute error (MAE), both of which calculate the average distance between the predicted and observed values in the data [59]. MSE is the average of the squared difference between the predicted value and the observed value, MAE is the average of the absolute difference between the predicted value and the observed value. MAE is straightforward to understand and treats all prediction errors proportionately. Although MSE is less intuitive, it penalizes poor predictions by squaring the error. We therefore used both measures to give a more accurate picture of the prediction performance. To calculate the overall performance, we averaged the MSE and MAE results from the LOOCV process.

**3.2.7. Association Rules Mining**—To take another approach to identify behavioral patterns associated with personality traits, we applied Apriori, a well-known frequency-based method for analyzing transactional data and producing association rules that explain the frequency (support) and significance (confidence) of the observed patterns [28, 30]. We applied the Apriori method on both person-level and day-level features. In the association rules analysis, a collection of one or more items is called an itemset. If an itemset contains  $k$  items, it is called a  $k$ -itemset. Our analysis used the dataset containing participant's data and the frequently selected features in an iterative process to extract the 1 to  $k$ -itemsets. Since Apriori requires discrete data, we categorized the features and personality trait values into three ranges of low, moderate, and high. The sensor features were normalized into the 0 and 1 range and discretized into low (0 – 0.25), moderate (0.25 – 0.75), and high (0.75 – 1). Personality traits ranged from 0 to 4 and were discretized into low (0 – 1), moderate (1 – 3), and high (3 – 4). Our decision for this categorization was based on two empirical

observations. First, the computational cost was significantly reduced using three categories for each trait. Second, the personality traits at both ends (high or low values) were associated with behavioral features more significantly than the middle level, so we merged the 1-2 and 2-3 into the median level (a neutral option). The distribution of participants in each group is shown in Figure 4. Each feature with its corresponding levels was then entered as items into the frequent item-set mining step of the Apriori algorithm. This step generates itemsets, including combinations of features and the relative frequency they occur together for a person with a given trait standing or that tend to co-occur on the same day. For example, a 2-itemset can contain low location entropy and high frequent outgoing calls features that appear together 15 times (i.e., days) in the entire dataset. In the next step, the Apriori algorithm generated a set of association rules from those frequent itemsets. An association between X and Y ( $X \rightarrow Y$ ) exists if items in X and Y frequently appear together. Support is the proportion of data samples that contain both X and Y, while confidence is the proportion of samples containing X that also contain Y [25]. A rule must achieve a minimum level of support and confidence to be considered significant. Figure 3 shows our approach in using Apriori to extract frequent behavior patterns from highly ranked features.

## 4. Data collection

### 4.1. Participants and Procedures

Community members were recruited through posted flyers for a study of personality, daily stress, and social interactions. For inclusion, participants had to be between the ages of 18 and 40 and were not currently receiving treatment for psychosis or a psychotic disorder. Preliminary screening was used to recruit a gender-balanced sample and to ensure adequate representation of a range of personality pathology and interpersonal problems. The sample was also selected to balance individuals who had received recent mental health treatment within the past year with those who had not. Individuals were pre-screened using items from the Inventory of Interpersonal Problems - Personality Disorder Scales [42] and were recruited in an approximately 1-1-1 representation of low, moderate, and high levels of interpersonal difficulties within gender, treatment status, and the overall sample. For this study, we only analyzed a subset of participants that had smartphone sensor data.

The total sample size for our analyses was 128. Participants were mostly white (78%; 9% Asian; 9% Black/African-American), roughly balanced on gender (54% female), with an average age of 27.7 (SD = 6.6). Most participants had received mental health treatment in the past (23%) or were currently receiving treatment (37%).

Participation involved completing a battery of baseline assessments in an initial laboratory session followed by a 14-day EMA protocol including self-reports and passive sensor data collection. Participants received \$50 for the baseline session. Those who answered 90% or greater of the surveys during the EMA protocol earned an additional \$160. This amount was prorated by week for those who completed less than 90% of the surveys overall. Participation was also incentivized with random drawings for prizes, with chance of winning proportionally tied to rate of participation.

## 4.2. Self-report measures

Self-report measures of adaptive and maladaptive traits were completed during the initial laboratory session. During this session, participants were also provided instruction from a research assistant regarding the EMA procedures and the required smartphone applications were installed on their personal Android or iOS smartphone. Smartphone sensor data was collected using an application called AWARE described in the next section. Because the focus of this study is on examining the associations between dispositional personality traits and passive sensing features, we did not use the self-report EMA surveys.

**4.2.1. Personality traits**—Personality traits were assessed using the self-report International Personality Item Pool - NEO-120 [31], which is a 120-item inventory designed to map onto the widely used NEO-Personality Inventory-Revised [16] traits in an abbreviated and open-source format. For each item, participants rated the extent to which a characteristic applies to them (e.g., “I am someone who is outgoing”) on a scale from “Very Inaccurate” (0) to “Very Accurate” (4). The IPIP-NEO-120 was scored to provide a score for the five trait domains of Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness. Reliability for the trait scales was high (mean McDonald’s Omega = .88; range = .84 ~ .91).

**4.2.2. Maladaptive traits**—Maladaptive traits were assessed using the Comprehensive Assessment of Traits Relevant to Personality Disorder (CAT-PD; [53]). The CAT-PD is a 216-item self-report inventory, which asks participants to describe how they behave in general compared to others (e.g., “I get angry easily”). Participants rated each item on a scale from (0) “Very Untrue of Me” to (4) “Very True of Me.” The CAT-PD assesses 33 narrow maladaptive trait scales. Scores for these narrow trait scales were averaged to calculate the five, broad maladaptive trait domains that map on to the normative personality traits [72]). Specifically, these are Antagonism (maladaptive low Agreeableness), Detachment (maladaptive low Extraversion), Disinhibition (maladaptive low Conscientiousness), Negative Affectivity (maladaptive low Emotional Stability), and Psychoticism (maladaptive Openness). Reliability for the five maladaptive trait scales was good (mean McDonald’s Omega = .92; range = .90 ~ .95).

## 4.3. Descriptive statistics

Density distributions of self-reported personality trait levels are shown in Figure 5. Overall, participants reported levels of adaptive traits above the midpoint of the scale ( $M = 2.63$ ) and levels of maladaptive traits below the midpoint ( $M = 1.25$ ). These distributions show that there was considerable variance between people representing a wide range of functioning.

Consistent with the broader literature showing the correspondence between adaptive and maladaptive traits, the strongest bivariate correlations were between the expected constructs (Figure 6). Agreeableness, Extraversion, Conscientiousness, and Emotional Stability were strongly negatively correlated with their maladaptive counterparts of Antagonism, Detachment, Disinhibition, and Negative Affectivity respectively. Unlike the other trait pairs, Psychoticism and Openness were positively—not negatively—correlated. This result is

in line with previous research showing that the associations between these traits tend to be inconsistent and moderate in size [65, 67].

#### 4.4. Passive sensing

Data from participant's smartphone sensors was collected via the AWARE application [23]. AWARE runs in the background on both iOS and Android platforms and continuously collects data from phone channels. For this study, we used battery data, phone call logs, GPS, microphone sensors, and phone screen lock/unlock data. Collection of battery, call logs, and screen status sensors were event-based, such that a data sample or observation is created in the database when a change (event) is detected (e.g., a phone call is made). GPS and microphone sensors were sampled at regular intervals. Specifically, GPS sensors were sampled every 180 seconds and the microphone sensors recorded audio for one minute with three minutes of pause in between samplings. Participants were also provided with a Fitbit Blaze smartwatch to collect physiological measures related to heart rate, steps, and sleep via AWARE. Heart rate and movement data were sampled in one minute intervals. Sleep features were estimated from a combination of movement and heart-rate patterns. Seventy-three participants used an iOS equipped smartphone and 55 used an Android smartphone. On average, 12 days of sensor data were collected from every participant ( $SD = 1.5$ ). Out of 128 participants with passive sensing data, data from 128 participants was used for the day-level modeling and data from 114 was used for person-level modeling.

## 5. Results

### 5.1. Adaptive and Maladaptive Trait Prediction

Although the purpose of our study was not to develop models that optimized prediction, we indexed model performance to evaluate these methods for studying personality and to better understand the relationship between passive sensing features and personality traits. As shown in Figure 7, all of the models had modest predictive accuracy, but the RF and XGB models outperformed the Lasso regression models. This suggests that associations between personality and behavior may be nonlinear and may not be adequately captured by traditional linear regression models. The accuracy of the RF and XGB models for predicting each trait were very similar, with the RF models performing slightly better. We also found that the pairs of corresponding maladaptive and adaptive traits tended to have comparable levels of prediction errors. For example, Agreeableness and Antagonism had the lowest average prediction error ( $MAE=0.36$ ,  $MSE=0.24$ ), while Emotional Stability and Negative Affectivity had the highest average error ( $MAE=0.52$ ,  $MSE=0.42$ ). The overall patterns of performance for predicting different traits were comparable for the day-level and person-level models. The person-level models were somewhat less accurate in predicting traits, likely due to including fewer observations than the day-level models. We also compared each of the three methods (Lasso, RF, XGB) to a baseline model based on the expected value of the distribution (i.e., mean only model). As can be seen in the boxplot in Figure 8, the average performance of the Lasso, RF, and XGB models did not consistently improve upon the prediction of the baseline model. Thus, on the aggregate, all passive sensing features provided little increment in the prediction of these broad traits. Nevertheless, this does not preclude more circumscribed trait and feature associations.

## 5.2. Frequently selected behavioral features and their relationship to adaptive and maladaptive traits

A total of 52 high-ranking features were selected by the LOOCV machine learning and feature selection processes (out of 177, 14%) across the person- and day-level models. Ranking scores were calculated using Formula 2 (see Table 1). Recall that the high ranking scores indicate that the behavioral feature had both a high overall feature importance and was frequently selected during the iterative training process. More high-ranking features were selected by the person-level models ( $n = 27$ ) compared to the day-level models ( $n = 18$ ) and only 7 features were selected by both modeling approaches. Table 1 provides a description for each sensor feature, which model(s) it was selected by, and the abbreviations used to represent those features throughout the results section. Figure 9 shows feature ranking scores and correlations between sensor features and personality traits from both the person- and day-level models. For presentation clarity, we removed the features with ranking scores and/or correlations equal to 0 in the person- and day-level models. The y-axis lists the important and frequently selected features in pairs with the ranking score followed by the correlation of that feature for both day- and person-level features.

As noted above, most features were only highly ranked in either the person- or day-level models, not both. Different inferences can be made from results at each level; features that predict traits at the person-level summarize variation in behaviors that are expressed consistently across days and situations whereas features at the day-level indicate behaviors that are expressed consistently from within a given day. In the few instances in which the same feature was selected by the person- and day- level models, the rankings/Lasso regression coefficients were comparable. For ease of reading, we do not differentiate between features selected at the person- or day-level in describing the results below, but this information can be found in Table 2 and Figure 9.

There were some features that were associated with generally adaptive or maladaptive tendencies. Shorter durations of time with the phone unlocked (SR3/SR4) was associated with adaptive traits and longer unlock durations was associated with maladaptive traits. Shorter phone unlocks may indicate brief social communications whereas more extended time with the phone unlocked suggests different behavior such as gaming or internet browsing. Adaptive traits tended to be associated with the shortest duration of time spent in a significant location (LOC5), the number of scans of the most scanned WiFi device (WiFi1), more frequent detection of human voice in the surrounding environment (AUD3), and the number of phone unlocks (SR1)—however, the direction of associations with the latter varied by trait (e.g., Agreeableness and Openness were associated with fewer unlocks and Extraversion, Conscientiousness, and Emotional Stability with more unlocks). Maladaptive traits, on the other hand, tended to be associated with a higher heart rate (HR1-5), spending less time awake in bed (SL3), spending longer durations of time in REM sleep (SL7), and both higher and lower variation in time spent at significant locations (LOC4). Some features were predictive of specific adaptive/maladaptive trait pairs. Importantly, in nearly every case, the associations with each paired trait were in opposite directions in line with the idea that they manifest in opposite extremes of behavior in everyday life. Table 2 shows the common highly ranked sensor features shared by each

pair of adaptive and maladaptive traits. Agreeableness and Antagonism (the maladaptive counterpart of Agreeableness) were both associated with low variation in the duration of calls (CAL1, CAL2) and number of distinct contacts (CAL7). These results suggest that more cooperative and sympathetic people may have more regular social contact with more people whereas those who are more hostile and rude may have inconsistent social contact and with fewer people. High Extraversion was associated with longer calls (CAL3, CAL8), shorter phone unlocks (SR3; i.e., possibly using the phone more for social communication), and spending less time in bed consistent with sociability and energy characteristic of this trait, whereas high Detachment (the maladaptive counterpart of Extraversion) was associated with shorter calls and longer phone unlocks consistent with less social activity. Emotional Stability and Negative Affectivity (the maladaptive counterpart of Emotional Stability) were associated with call length, environmental noise (i.e., not human voice or silence), and distance travelled. These results suggest that more emotionally stable people make longer calls, tend to spend time in more quiet or social environments, and move around more throughout the day. In contrast, less emotionally stable people make shorter calls, spend time in noisy places, and are more sedentary. Against the expected opposite behavioral tendencies, both Emotional Stability and Negative Affectivity were associated with higher heart rate. Both Conscientiousness and Disinhibition were associated with patterns of sedentariness (ST2, ST3) and length of stay at significant locations (LOC5). Interestingly, both high Conscientiousness and high Disinhibition were associated with less time being sedentary but more conscientious people were sedentary for more consistent lengths of time whereas disinhibited people were sedentary for inconsistent lengths of time. High Conscientiousness was associated with shorter stays at significant locations and Disinhibition with longer stays. Finally, only variability in noise exposure (AUD1) was associated with Openness and Psychoticism. In line with evidence that these traits are not opposite poles of a trait continuum but are instead better thought of as maladaptive variants of one another, this result could indicate that more open people and those higher on Psychoticism both tended to spend time in a wide range of environments throughout the day.

### 5.3. Mining collective behavioral patterns associated with adaptive and maladaptive traits

Whereas the decision tree models and feature ranking showed the predictive power of various features, the Apriori models described associations between traits and characteristic combinations of sensor features that tended to co-occur between people or within people from day-to-day. Supporting the potential of this method for studying personality, we found that some levels of every trait could be predicted from feature combinations. Similar to the decision tree models, the mined patterns of features associated with different trait levels were almost entirely different in the person- and day-level models of the same trait. This suggests that trait-relevant behaviors that co-occur from person-to-person might be different than those that co-occur from day-to-day. For example, if aggregating behavior over multiple days, people low in Disinhibition tend to sleep for short periods of time (SL4), spend time in less noisy environments (AUD6), and are less sedentary (ST4) on average (person-level), but the co-occurrence of these tendencies on any given day (day-level) is not characteristic of low Disinhibition. Instead, a typical day for those low in Disinhibition is marked by consistent social contact (CAL1), low heart rate (HR1), and shorter stays at significant locations (LOC5). Contrary to expectations, there was very little overlap in combinations

of features associated with theoretically corresponding levels of adaptive and maladaptive traits (e.g., high Agreeableness and low Antagonism) in the person- or day-level models. These results raise the possibility that there are nuanced differences between adaptive and maladaptive variants that were not captured by the decision-tree models. The data, code and more results are publicly available at <https://github.com/HAI-lab-UVA/AAPECS>.

## 6. Discussion

There is growing evidence for conceptualizing personality pathology using dimensional traits in a unified model with adaptive personality, but there is a need to better understand the everyday manifestations of these traits that contribute to functioning and dysfunction. Towards this end, we applied multiple machine learning methods to sensor data collected from mobile and wearable devices and found that (1) adaptive and maladaptive traits can only modestly be predicted from smartphone sensor data though (2) there are behavior patterns that link adaptive and maladaptive variants consistent with contemporary models of personality pathology. This study adds to the literature showing smartphone sensor data captures personality-relevant patterns of behavior. We extended previous research by showing maladaptive traits are associated with several features derived from this type of data, providing initial evidence for the utility of passive sensing to study pathological personality. In addition to using regularized Lasso regression and decision tree models applied in previous work, we used association mining methods to show that specific combinations of features are associated with personality traits. We offer tentative interpretations of our results for hypothesis generating purposes, but emphasize that more work is needed to test these hypotheses.

### 6.1. Limited global trait prediction from sensors

In comparing the initial machine learning models (RF, XGB) to a baseline model using only the mean, we found little to no predictive power of the global traits from the sensors. This is consistent with prior work finding very modest prediction of broad traits from passive sensors (CITE) and is largely to be expected given the very broad and complex nature of these traits, which are highly distal constructs from the specific sensor features (see e.g., Mohr et al., 2017). Indeed, our goal was not to optimize prediction of traits, but rather to flesh out methods that would be useful for identifying associated patterns of features and behaviors in daily life. Thus, we see this finding as unsurprising and consistent with prior work, and should serve as a strong indication that fully capturing broad and complex constructs like higher-order personality traits using only passive sensing, at least with currently available tools, is unlikely to succeed. Nevertheless, we found that the methods we used did identify a number of interesting features and behaviors that are informative for understanding adaptive and maladaptive trait manifestation in daily life, and we consider these specific findings in the spirit of generating hypotheses and future studies.

### 6.2. Traits vary in how well they can be predicted from sensors

Our results indicate that some features are generally more (or less) associated with personality—only 52% of the features we extracted met our selection criteria. Very few features from certain sensors (e.g., WiFi, activity detection, Bluetooth) were selected by the

machine learning processes suggesting they may be less useful for studying personality. Screen time, heart rate, sleep, and location-based metrics, on the other hand, tended to exhibit important associations with adaptive and maladaptive personality across the different machine learning models. In particular, adaptive traits were most associated with social and contextualized activity features whereas maladaptive traits were generally predicted by duration of phone unlock episodes and physiological features (i.e., heart rate, sleep). Nearly all adaptive traits were associated with brief screen unlocks, time spent in environments with human voice, consistency of call lengths, and time spent at regularly-visited locations and around familiar WiFi devices. Maladaptive traits, on the other hand, were associated with extended screen time, higher heart rate throughout the day, amount of time spent awake in bed, and the consistency of time spent at different locations. We interpret longer durations of phone unlock episodes to possibly reflect engagement in recreational activities such as gaming or internet browsing and short unlock episodes to reflect texting or other brief, social communications. Taken together, the feature selection results and specific associations between features and traits suggest that engagement with other people and the environment may be especially related to adaptive functioning and typical physiological responses and heavy, recreational phone use may be especially important indicators of maladaptive functioning. We also showed that some traits are generally more (or less) predictable from passively sensed behavior. In the decision tree models, the most predictable trait pairs were Disinhibition/Conscientiousness and Antagonism/Agreeableness and the least predictable were Negative Affectivity/Emotional Stability. One reason certain traits were better predicted may be the nature of what can be detected by smartphone and FitBit sensors; namely, behavioral manifestations of traits rather than thoughts or emotions. Disinhibition and Conscientiousness reflect individual differences in impulse control and goal-directedness, which may be tracked well with features indexing behavioral (in)consistency (e.g., entropy/variability of calls or movement) and indicators of distractibility (e.g., phone unlocks). In contrast, Negative Affectivity and Emotional Stability encompass variation in typical emotional experiences that may not be expressed in overt, recognizable behaviors. In line with this possibility, human observers (like the passively sensed “observer”) struggle to detect Negative Affectivity/Emotional Stability in others and tend to be the least accurate in perceptions of these traits compared to other traits (i.e., lower inter-rater agreement and self-informant agreement; Connelly et al., [15]; Vazire et al., [61]). The high predictability of Antagonism and Agreeableness was less expected because these traits are generally defined by social intentions that are internally experienced rather than expressed by observable, detectable behaviors. Our results suggest that there may be behavioral manifestations of the tendency to be at odds with others (Antagonism) or to cooperate (Agreeableness) that may not have been considered in prior theorizing and research.

### 6.3. Evidence for behavioral indicators that span adaptive and maladaptive trait dimensions

Figure 10 shows the association rules between sensor features and personality traits. Our results bring a new perspective to the continuity between adaptive and pathological personality by identifying behavioral content that characterizes the full range of functioning in different trait dimensions. The corresponding adaptive and maladaptive traits were



strongly correlated with one another as expected, and the machine learning models selected several important features in common for most trait pairs. Overlap between features extracted across adaptive and maladaptive models suggest these behaviors are especially strong indicators of the underlying trait. These indicators, in turn, can help us understand what people do in everyday life that accounts for the link between personality and life outcomes. For example, several features were important indicators of the trait continuum from Disinhibition to Conscientiousness. These traits encompass individual differences in the motivation and ability to maintain focus on long-term goals and are strongly associated with academic and work success. It is thought that people who score high on Disinhibition do poorly in school and work because they struggle to complete tasks and tend to act impulsively whereas people who score high on Conscientiousness generally have better outcomes because they are able to focus on tasks and control their impulses. We found that people who scored high on Disinhibition tended to spend less time awake in bed, have shorter periods of being sedentary, and make calls of inconsistent durations, which may reflect difficulty staying still and a tendency to engage in more erratic social behavior. In contrast, people scoring high on Conscientiousness showed the opposite patterns, in line with more controlled and routine behavior. These results suggest that passive sensing may be used to detect consequential behavior patterns that potentially explain how this trait dimension helps (or hinders) functioning. Exceptions to the general pattern of adaptive/maladaptive trait pairs mapping onto opposite behavioral tendencies may be informative for understanding functioning as well. In the Apriori association mining models that identified configurations of behaviors that are characteristic of traits, there was almost no overlap in features between corresponding trait pairs. It is possible that these models uncovered more subtle differences in adaptive and maladaptive behaviors that could not possibly be found in general linear models typical in psychological research or even our decision tree analyses. Given the novelty of these findings, it will be important to replicate these analyses in other samples to evaluate their generalizability. Our results add new types of evidence to bear on important, clinical questions and provoke new hypotheses. For example, Negative Affectivity is a robust risk factor for cardiovascular disease (and Emotional Stability a protective factor), but it is unclear how exactly these traits influence cardiovascular health [35, 50, 56]. Because this trait dimension reflects how easily stressed and emotionally reactive a person tends to be, one hypothesis is that people who score high on Negative Affectivity overuse their stress response system, which over time takes a toll on cardiovascular health [3, 36]. There has been mixed support for this hypothesis, and most research investigating the association between stress responses and Negative Affectivity/Emotional Stability has been conducted in laboratory settings [13, 60]. Instead of sampling heart rate in a few experiments, we sampled heart rate continuously in everyday life, which may give a more representative picture of people's typical stress response patterns. Our results support the stress response hypothesis as we found that Negative Affectivity was strongly associated with a higher resting heart rate whereas Emotional Stability was associated with lower resting heart rate. Because the same behavior can relate to multiple traits for different reasons (e.g., longer sedentary durations may reflect the ability to focus for long periods of time or it may represent lack of energy and low mood), it can be difficult to infer why certain trait-relevant behaviors lead to life outcomes—but Apriori has the potential to disambiguate such associations. For instance, in these models we found that

people with high scores on Conscientiousness had more days with consistent call lengths, moderate sleep durations, and high levels of physical activity whereas people who scored high on Agreeableness also had more days with consistent call lengths but in combination with moderate levels of activity. Thus, although Agreeableness and Conscientiousness are both traits associated with more adaptive functioning, and both relate to consistent social behavior, they can be differentiated by co-occurring behaviors. Both behavior sets seem to reflect balanced activity, but perhaps being achievement-striving and disciplined is tied to well-regulated sleep and exercising more, whereas being more agreeable is primarily related to routine social interactions within a given day.

#### 6.4. Differences in day-level and person-level models

Our study also explored the methodological and conceptual implications of modeling behavior at the person- versus day-level. Each approach comes with different methodological benefits and drawbacks and results at each level provide insight into somewhat different questions about how personality relates to functioning. Day-level models indicate what behavior is consistently emitted from day-to-day for a person with that trait standing (day-level). We found that the day-level models performed slightly better than the person-level models in predicting personality traits according to global evaluation metrics (i.e., MAE, MSE). The day-level models may have been more accurate because they used considerably more observations than the person-level models or because they captured more personality-relevant information. At the same time, there were more person-level than day-level features that were predictive of personality in the decision tree analyses. This finding suggests that indices which account for consistent behavior over time may be generally more representative of traits, but there are also trait-relevant behaviors which are expressed in a highly consistent manner each day. Alternatively it may reflect the fact that more indices were auditioned in our models (e.g., for a given sensor we included mean, max, min, and standard deviation) because some only emerge aggregated over days and not at a single observation. Reinforcing the distinctiveness of these levels, there were few features that predicted personality in both the person- and day-level decision tree models and there were practically zero shared feature combinations for the same trait at different levels in the Apriori models. These discrepancies align conceptually, if not statistically, with the well-established observation in psychological research using multi-level modeling that patterns of associations can vary between and within-persons (e.g., [43, 4]). Our aim is not to argue in favor of either type of model, as it is clear that each provides unique information about personality, but we hope these results encourage researchers to consider the implications of each approach in developing their study designs, analyses, and interpretations.

#### 6.5. Limitations and future directions

The major advantage of using passive sensing and machine learning to study personality pathology is that they allow us to assess aspects of behavior that cannot be measured with any other methods and with potentially much greater precision. At the same time, because passive sensing is a relatively new technology, the psychometrics of the method have not been comprehensively evaluated yet. Before passive sensing research can be translated into theory or clinical applications, more work is needed to establish the reliability and validity of features and feature combinations. In terms of construct validity, some sensor features

are human-interpretable (e.g., more phone calls usually indicate more social activity), but others are more ambiguous and less readily interpreted. By examining associations with personality traits that have well-established psychometrics, our study provides some insight into how to interpret passive sensor features. Future research can build on this foundation by examining associations with contemporaneous participant reports; for example, to determine whether days with longer durations of being sedentary typically reflect high motivation and concentrated studying or lack of motivation and low mood, the passive sensor data could be correlated with participant reports of the activity they are engaging in.

Related to the unknown psychometrics of passive sensing methods, the technology is still in early stages of development and potentially prone to errors throughout the data collection process. There are currently no empirical criteria for determining valid data captured by sensor streams, meaning investigators have to apply ad hoc heuristics for assessing validity. Because of the exploratory nature of our study, we only removed observations with values outside of plausible ranges (e.g., negative value for distance travelled in a day). This liberal approach allowed us to maximize the amount of data analyzed, but it is possible that some of the data included errors (e.g., distance travelled that was calculated from inaccurately encoded raw GPS data). Alongside continued refinement of data collection technologies, it will be important to establish empirical criteria for determining valid data with more focused research efforts moving forward.

Another limitation of our study is that we were unable to identify combinations of features correlated with very high levels of personality pathology using the Apriori method. Because this approach is based on the frequency with which certain trait values appear in a dataset, there was less confidence and support for feature combinations related to high levels of maladaptive traits (or low levels of adaptive traits) that were less common in our sample. To maximize association mining methods to understand maladaptive personality, future research will need to collect data from samples selected for very high pathology (e.g., psychiatric patients with specific diagnoses).

A hallmark of personality measurement and literature is a keen focus on reliability and validity. We should not lose sight of these two guiding principles as we seek to incorporate passive-sensing into our toolkit. It is important to remember that reliability and validity are contextualized—that is, they are sample and purpose specific. For instance, an accelerometer may reliably encode a device's movement, but it may be an unreliable marker of a specific movement (e.g., footstep, traveling by automobile). Or, sensors may reliably capture whether a call comes in, but the number of calls that come in for an individual per day may not be a reliable feature of the individual because it varies widely day to day. Given the vast number of sensors and their combinations in higher-order features, and the interest among personality psychologists on different time-scales of assessment ranging from moments to decades, careful thought will need to be given to how reliability and validity should be defined and evaluated in passive-sensing.

## 7. Conclusion

As clinical psychology and psychiatry increasingly adopt dimensional models of classifying personality pathology, key questions remain about what the specific patterns of behavior people engage in that connect personality to poor psychosocial and physical health outcomes. Our study highlights the opportunities and challenges of using machine learning and passive sensor data to address these questions. We showed that maladaptive traits are unlikely to be strongly predictable from passive sensor features, at least as currently available. Despite this, a number of passive sensing features and feature combinations emerged as associated, and the results provoke new hypotheses about day-to-day processes underlying personality traits and open fresh directions of inquiry.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

Support for this research was provided by the National Institutes of Health (NIH) (R01 AA026879; L30 MH101760), the National Science Foundation (NSF-IIS-1816687), the University of Pittsburgh's Clinical and Translational Science Institute, which is funded by the NIH Clinical and Translational Science Award (CTSA) program (UL1 TR001857), and grants from the University of Pittsburgh Central Research Development Fund and a Steven D. Manners Faculty Development Award from the University of Pittsburgh University Center for Social and Urban Research.

## Biographies



Runze Yan is a Ph.D. student of Systems and Information Engineering at the University of Virginia. His research topics include mobile sensing, ubiquitous computing, and machine learning.



Whitney R. Ringwald is a Ph.D. student in the Clinical Psychology Program in the Department of Psychology at the University of Pittsburgh, and a member of the Personality Processes and Outcomes Laboratory. Her research focuses on understanding the link between adaptive and maladaptive personality functioning, with an emphasis on regulatory processes.



Julio Vega Hernandez is a postdoctoral associate at the Mobile Sensing + Health Institute. He is interested in personalized methodologies to monitor chronic conditions that affect daily human behavior using mobile and wearable data.



Madeline Kehl is currently an analyst at Arcadia.io, but until recently was the laboratory manager for the Personality Processes and Outcomes Laboratory. Her interests include using machine learning techniques, as well as advanced statistical models to better understand human personality and cognition.



Sang Won Bae is an assistant professor of Systems Engineering at Stevens Institute of Technology. Her research interests are human-computer interaction, mobile health, cognitive science, and ubiquitous computing.



Anind K. Dey is a Professor and Dean of the Information School and Adjunct Professor in the Department of Human-Centered Design and Engineering. Anind is renowned for his early work in context-aware computing, an important theme in modern computing, where computational processes are aware of the context in which they operate and can adapt appropriately to that context. His research is at the intersection of human-computer interaction, machine learning, and ubiquitous computing.



Carissa A. Low is Assistant Professor of Medicine (Hematology/Oncology), Psychology, and Biomedical Informatics at the University of Pittsburgh, Core Faculty of the Center for Behavioral Health and Smart Technology, Member of the University of Pittsburgh Medical Center Hillman Comprehensive Cancer Center, and Adjunct Faculty in the Human-Computer Interaction Institute at Carnegie Mellon University. Trained as a clinical health psychologist, her research leverages mobile technology for remote patient monitoring as well as delivery and personalization of behavioral interventions.



Aidan G.C. Wright is an Associate Professor in the Department of Psychology at the University of Pittsburgh, where he directs the Personality Processes and Outcomes Laboratory. Trained as a clinical psychologist, his work is at the intersection of personality and psychopathology, with a particular emphasis on studying each as ensembles of contextualized dynamic processes using ambulatory assessment methods.



Afsaneh Doryab is an Assistant Professor in the School of Engineering and Applied Science at the University of Virginia and the director of Human-AI Technology Lab. Her research is at the intersection of AI, HCI, and Ubiquitous Computing with a particular focus on computational modeling. She leverages passive sensing and machine learning to model human behavior and to build personalized context-aware systems.

## References

- [1]. Ai P, Liu Y, Zhao X, 2019. Big five personality traits predict daily spatial behavior: Evidence from smartphone data. *Personality and Individual Differences* 147, 285–291.
- [2]. American Psychiatric Association, A., Association, A.P., et al., 2013. *Diagnostic and statistical manual of mental disorders: Dsm-5*.
- [3]. Barlow DH, Ellard KK, Sauer-Zavala S, Bullis JR, Carl JR, 2014. The origins of neuroticism. *Perspectives on Psychological Science* 9, 481–496. [PubMed: 26186755]
- [4]. Beckmann N, Wood RE, 2017. Dynamic personality science. integrating between-person stability and within-person change. *Frontiers in psychology* 8, 1486. [PubMed: 28943855]

- [5]. Beierle F, Probst T, Allemand M, Zimmermann J, Pryss R, Neff P, Schlee W, Stieger S, Budimir S, 2020. Frequency and duration of daily smartphone usage in relation to personality traits. *Digital Psychology* 1, 20–28.
- [6]. Ben-Zeev D, Brian R, Wang R, Wang W, Campbell AT, Aung MS, Merrill M, Tseng VW, Choudhury T, Hauser M, et al. , 2017. Crosscheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. *Psychiatric rehabilitation journal* 40, 266. [PubMed: 28368138]
- [7]. Bennett DA, 2001. How can i deal with missing data in my study? *Australian and New Zealand journal of public health* 25, 464–469. [PubMed: 11688629]
- [8]. Biau G, Scornet E, 2016. A random forest guided tour. *Test* 25, 197–227.
- [9]. Breiman L, 2001. Random forests. *Machine learning* 45, 5–32.
- [10]. Canzian L, Musolesi M, 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*, pp. 1293–1304.
- [11]. Carlson EN, Vazire S, Oltmanns TF, 2013. Self-other knowledge asymmetries in personality pathology. *Journal of Personality* 81, 155–170. [PubMed: 22583054]
- [12]. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. , 2015. Xgboost: extreme gradient boosting. R package version 0.4-2 1.
- [13]. Chida Y, Hamer M, 2008. Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: a quantitative review of 30 years of investigations. *Psychological bulletin* 134, 829. [PubMed: 18954159]
- [14]. Chittaranjan G, Blom J, Gatica-Perez D, 2013. Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing* 17, 433–450.
- [15]. Connelly BS, Ones DS, 2010. An other perspective on personality: meta-analytic integration of observers' accuracy and predictive validity. *Psychological bulletin* 136, 1092. [PubMed: 21038940]
- [16]. Costa PT, McCrae RR, 1992. Normal personality assessment in clinical practice: The neo personality inventory. *Psychological assessment* 4, 5.
- [17]. Daly M, Delaney L, Doran PP, Harmon C, MacLachlan M, 2010. Naturalistic monitoring of the affect-heart rate relationship: a day reconstruction study. *Health Psychology* 29, 186. [PubMed: 20230092]
- [18]. DeYoung CG, 2015. Cybernetic big five theory. *Journal of research in personality* 56, 33–58.
- [19]. Dixon-Gordon KL, Whalen DJ, Layden BK, Chapman AL, 2015. A systematic review of personality disorders and health outcomes. *Canadian Psychology/Psychologie Canadienne* 56, 168. [PubMed: 26456998]
- [20]. Doryab A, Chikarsel P, Liu X, Dey AK, 2018. Extraction of behavioral features from smartphone and wearable data. *arXiv preprint arXiv:1812.10394* .
- [21]. Doryab A, Villalba DK, Chikarsal P, Dutcher JM, Tumminia M, Liu X, Cohen S, Creswell K, Mankoff J, Creswell JD, et al. , 2019. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data. *JMIR mHealth and uHealth* 7, e13209. [PubMed: 31342903]
- [22]. Ebner-Priemer UW, Kuo J, Schlotz W, Kleindienst N, Rosenthal MZ, Detterer L, Linehan MM, Bohus M, 2008. Distress and affective dysregulation in patients with borderline personality disorder: A psychophysiological ambulatory monitoring study. *The Journal of nervous and mental disease* 196, 314–320. [PubMed: 18414126]
- [23]. Ferreira D, Kostakos V, Dey AK, 2015. Aware: mobile context instrumentation framework. *Frontiers in ICT* 2, 6.
- [24]. Hamaker EL, Wichers M, 2017. No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science* 26, 10–15.
- [25]. Han J, Kamber M, Pei J, 2011. *Data mining concepts and techniques third edition*. The Morgan Kaufmann Series in Data Management Systems 5, 83–124.
- [26]. Harari GM, Müller SR, Stachl C, Wang R, Wang W, Bühner M, Rentfrow PJ, Campbell AT, Gosling SD, 2019. Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of personality and social psychology* .

- [27]. Harari GM, Vaid SS, Müller SR, Stachl C, Marrero Z, Schoedel R, Bühner M, Gosling SD, 2020. Personality sensing for theory development and assessment in the digital age. *European Journal of Personality* 34, 649–669.
- [28]. Hegland M, 2007. The apriori algorithm—a tutorial. *Mathematics and computation in imaging science and information processing* , 209–262.
- [29]. Hengartner MP, Zimmermann J, Wright AG, 2018. Personality pathology. .
- [30]. Ilayaraja M, Meyyappan T, 2013. Mining medical data to identify frequent diseases using apriori algorithm, in: 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, IEEE. pp. 194–199.
- [31]. Johnson JA, 2014. Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the ipip-neo-120. *Journal of Research in Personality* 51, 78–89.
- [32]. Kearns M, Ron D, 1999. Algorithmic stability and sanity-check bounds for leave-one-out cross-validation. *Neural computation* 11, 1427–1453. [PubMed: 10423502]
- [33]. Kotov R, Krueger RF, Watson D, Cicero DC, Conway CC, DeYoung CG, Eaton NR, Forbes MK, Hallquist MN, Latzman RD, et al. , 2021. The hierarchical taxonomy of psychopathology (hitop): A quantitative nosology based on consensus of evidence. *Annual review of clinical psychology* 17.
- [34]. Krueger RF, Eaton NR, 2010. Personality traits and the classification of mental disorders: Toward a more complete integration in dsm–5 and an empirical model of psychopathology. *Personality Disorders: Theory, Research, and Treatment* 1, 97.
- [35]. Kubzansky LD, Kawachi I, 2000. Going to the heart of the matter: do negative emotions cause coronary heart disease? *Journal of psychosomatic research* 48, 323–337. [PubMed: 10880655]
- [36]. Lahey BB, 2009. Public health significance of neuroticism. *American Psychologist* 64, 241.
- [37]. Lane ND, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell AT, 2010. A survey of mobile phone sensing. *IEEE Communications magazine* 48, 140–150.
- [38]. Markon KE, Krueger RF, Watson D, 2005. Delineating the structure of normal and abnormal personality: an integrative hierarchical approach. *Journal of personality and social psychology* 88, 139. [PubMed: 15631580]
- [39]. Matsuki K, Kuperman V, Van Dyke JA, 2016. The random forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading* 20, 20–33. [PubMed: 26770056]
- [40]. Mohr DC, Zhang M, Schueller SM, 2017. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annual review of clinical psychology* 13, 23–47.
- [41]. Mønsted B, Mollgaard A, Mathiesen J, 2018. Phone-based metric as a predictor for basic personality traits. *Journal of Research in Personality* 74, 16–22.
- [42]. Morse JQ, Pilkonis PA, 2007. Screening for personality disorders. *Journal of personality disorders* 21, 179–198. [PubMed: 17492920]
- [43]. Nezelek JB, Newman DB, Thrash TM, 2017. A daily diary study of relationships between feelings of gratitude and well-being. *The Journal of Positive Psychology* 12, 323–332.
- [44]. Organization, W.H., et al., 2018. International classification of diseases for mortality and morbidity statistics (11th revision).
- [45]. Pieper S, Brosschot JF, van der Leeden R, Thayer JF, 2007. Cardiac effects of momentary assessed worry episodes and stressful events. *Psychosomatic medicine* 69, 901–909. [PubMed: 17991822]
- [46]. Powers AD, Oltmanns TF, 2013. Borderline personality pathology and chronic health problems in later adulthood: The mediating role of obesity. *Personality Disorders: Theory, Research, and Treatment* 4, 152.
- [47]. Renn BN, Pratap A, Atkins DC, Mooney SD, Areán PA, 2018. Smartphone-based passive assessment of mobility in depression: Challenges and opportunities. *Mental health and physical activity* 14, 136–139. [PubMed: 30123324]
- [48]. Ringwald WR, Woods WC, Edershile EA, Sharpe BM, Wright AG, 2021. Psychopathology and personality functioning, in: *The Handbook of Personality Dynamics and Processes*. Elsevier, pp. 273–293.



- [49]. Royston P, White IR, et al. , 2011. Multiple imputation by chained equations (mice): implementation in stata. *J Stat Softw* 45, 1–20.
- [50]. Rugulies R, 2002. Depression as a predictor for coronary heart disease: a review and meta-analysis. *American journal of preventive medicine* 23, 51–61. [PubMed: 12093424]
- [51]. Schoedel R, Pargent F, Au Q, Völkel ST, Schuwerk T, Bühner M, Stachl C, 2020. To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day–night behaviour patterns. *European Journal of Personality* 34, 733–752.
- [52]. Shi T, Horvath S, 2006. Unsupervised learning with random forest predictors. *Journal of Computational and Graphical Statistics* 15, 118–138.
- [53]. Simms LJ, Goldberg LR, Roberts JE, Watson D, Welte J, Rotterman JH, 2011. Computerized adaptive assessment of personality disorder: Introducing the cat–pd project. *Journal of personality assessment* 93, 380–389. [PubMed: 22804677]
- [54]. Stachl C, Au Q, Schoedel R, Gosling SD, Harari GM, Buschek D, Völkel ST, Schuwerk T, Oldemeier M, Ullmann T, et al. , 2020. Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences* 117, 17680–17687.
- [55]. Strobl C, Malley J, Tutz G, 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods* 14, 323. [PubMed: 19968396]
- [56]. Thurston RC, Rewak M, Kubzansky LD, 2013. An anxious heart: anxiety and the onset of cardiovascular diseases. *Progress in cardiovascular diseases* 55, 524–537. [PubMed: 23621962]
- [57]. Trull TJ, Durrett CA, 2005. Categorical and dimensional models of personality disorder. *Annu. Rev. Clin. Psychol.* 1, 355–380. [PubMed: 17716092]
- [58]. Trull TJ, Ebner-Priemer U, 2013. Ambulatory assessment. *Annual review of clinical psychology* 9, 151–176.
- [59]. Tuli S, Tuli S, Tuli R, Gill SS, 2020. Predicting the growth and trend of covid-19 pandemic using machine learning and cloud computing. *Internet of Things* 11, 100222.
- [60]. Turner AI, Smyth N, Hall SJ, Torres SJ, Hussein M, Jayasinghe SU, Ball K, Clow AJ, 2020. Psychological stress reactivity and future health and disease outcomes: A systematic review of prospective evidence. *Psychoneuroendocrinology* 114, 104599. [PubMed: 32045797]
- [61]. Vazire S, 2010. Who knows what about a person? the self–other knowledge asymmetry (soka) model. *Journal of personality and social psychology* 98, 281. [PubMed: 20085401]
- [62]. Vega J, 2021. Rapids (reproducible analysis pipeline for data) streams. URL: <https://rapidspitt.readthedocs.io/en/latest/index.htm>.
- [63]. Vehtari A, Gelman A, Gabry J, 2017. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing* 27, 1413–1432.
- [64]. Wang W, Harari GM, Wang R, Müller SR, Mirjafari S, Masaba K, Campbell AT, 2018. Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1–21.
- [65]. Watson D, Clark LA, Chmielewski M, 2008. Structures of personality and their relevance to psychopathology: Ii. further articulation of a comprehensive unified trait structure. *Journal of personality* 76, 1545–1586. [PubMed: 19012658]
- [66]. Waugh MH, Hopwood CJ, Krueger RF, Morey LC, Pincus AL, Wright AG, 2017. Psychological assessment with the dsm–5 alternative model for personality disorders: Tradition and innovation. *Professional Psychology: Research and Practice* 48, 79.
- [67]. Widiger TA, Crego C, 2019. Hitop thought disorder, dsm-5 psychoticism, and five factor model openness. *Journal of Research in Personality* 80, 72–77.
- [68]. Widiger TA, Samuel DB, 2005. Diagnostic categories or dimensions? a question for the diagnostic and statistical manual of mental disorders–. *Journal of abnormal psychology* 114, 494. [PubMed: 16351373]
- [69]. Wright AG, Calabrese WR, Rudick MM, Yam WH, Zelazny K, Williams TF, Rotterman JH, Simms LJ, 2015. Stability of the dsm-5 section iii pathological personality traits and their longitudinal associations with psychosocial functioning in personality disordered individuals. *Journal of Abnormal Psychology* 124, 199. [PubMed: 25384070]

- [70]. Wright AG, Hopwood CJ, 2016. Advancing the assessment of dynamic psychological processes. *Assessment* 23, 399–403. [PubMed: 27313187]
- [71]. Wright AG, Pincus A, Hopwood C, 2020. Contemporary integrative interpersonal theory: Integrating structure, dynamics, temporal scale, and levels of analysis .
- [72]. Wright AG, Simms LJ, 2014. On the structure of personality disorder traits: conjoint analyses of the cat-pd, pid-5, and neo-pi-3 trait models. *Personality Disorders: Theory, Research, and Treatment* 5, 43.

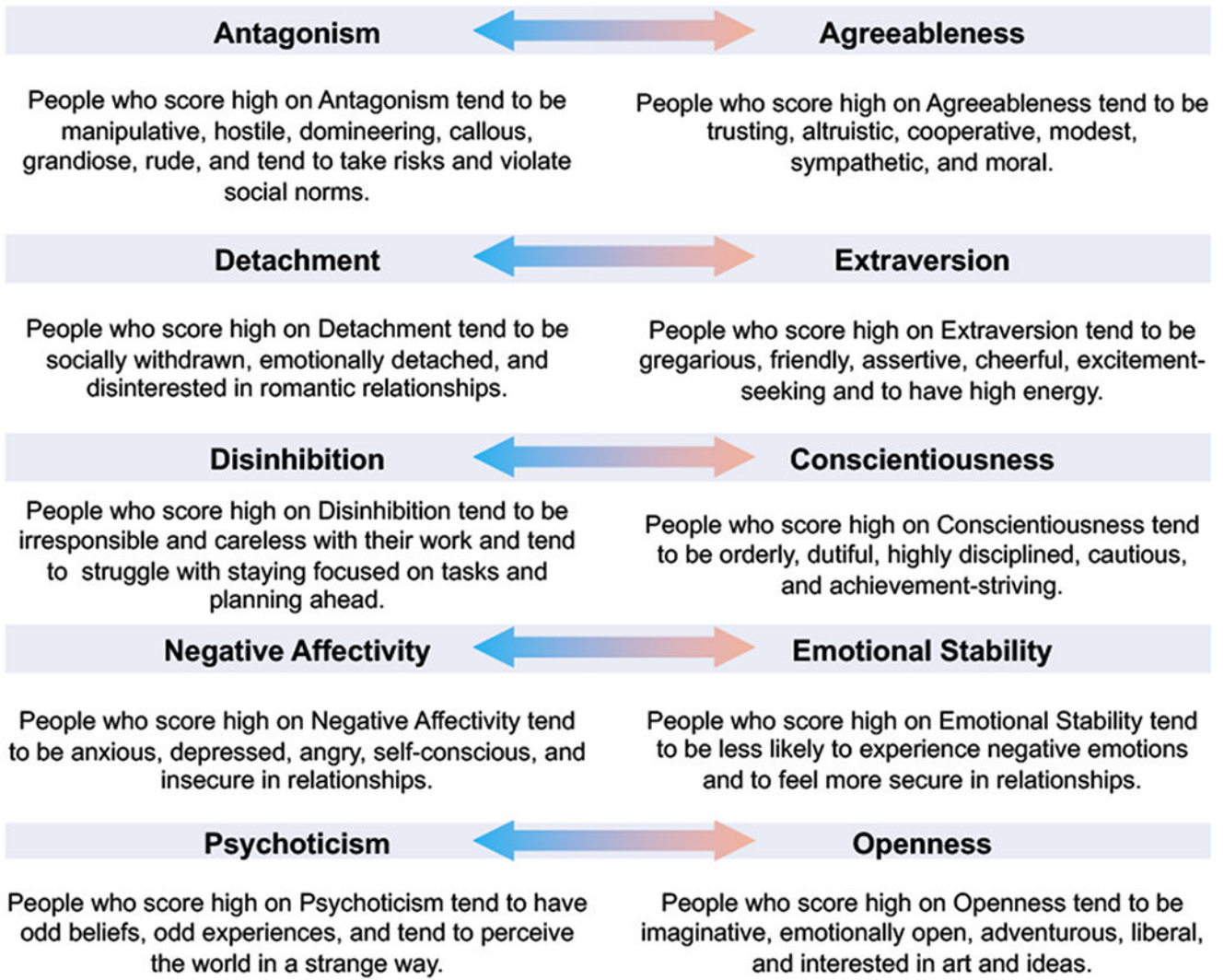
Machine learning analyses show that both adaptive and maladaptive traits can be predicted from passively sensed behavior using consumer devices (i.e., smartphones and smartwatches).

Association Rules Mining shows that specific combinations of features are associated with personality traits.

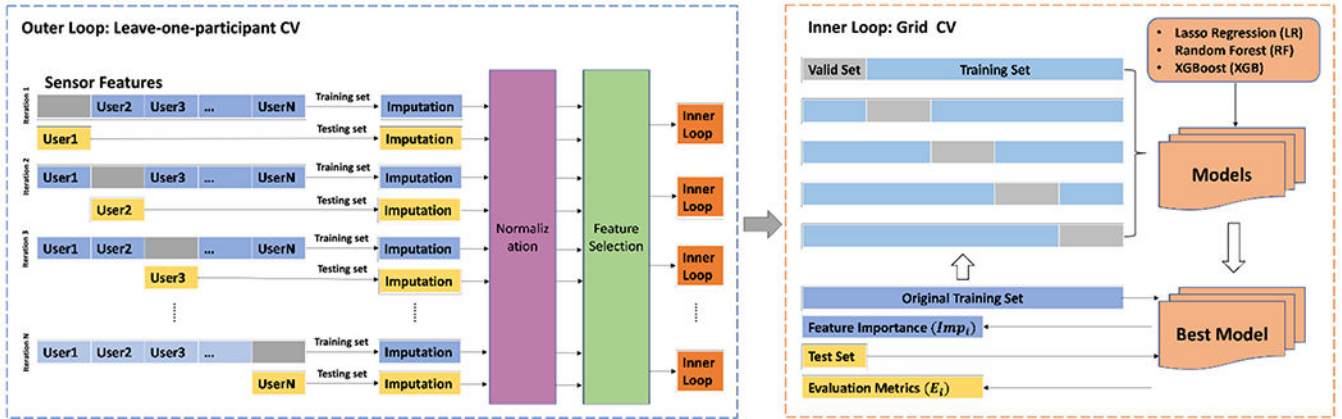
Behavior patterns that link adaptive and maladaptive variants consistent with contemporary models of personality pathology were identified.

Adaptive and maladaptive traits vary in how well they can be predicted from passively sensed behaviors.

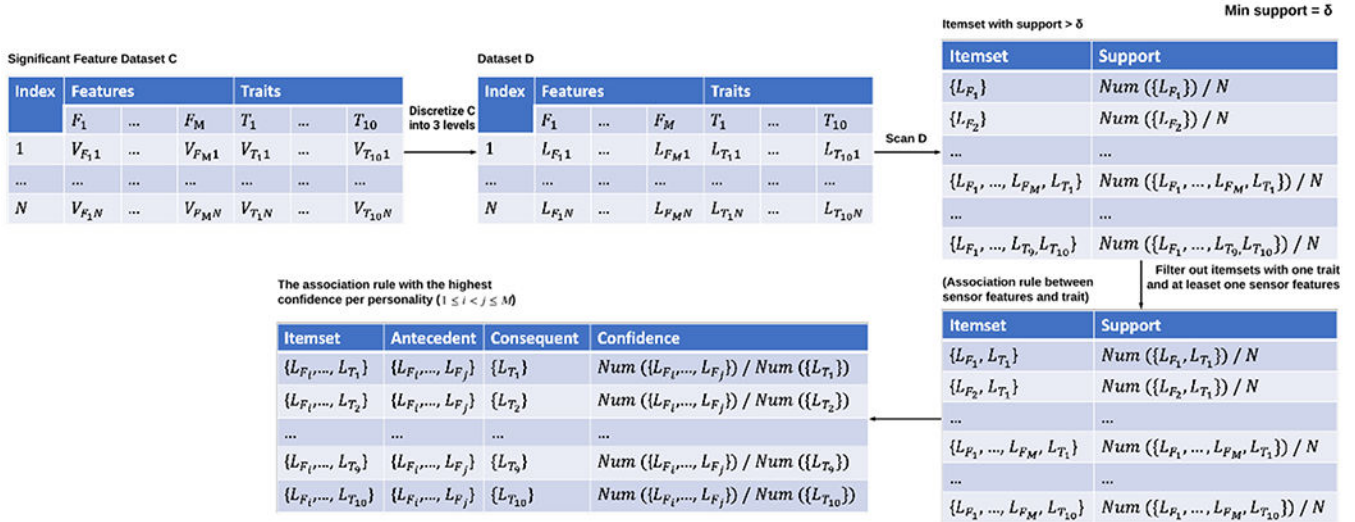
There is evidence for behavioral indicators that span adaptive and maladaptive trait dimensions.



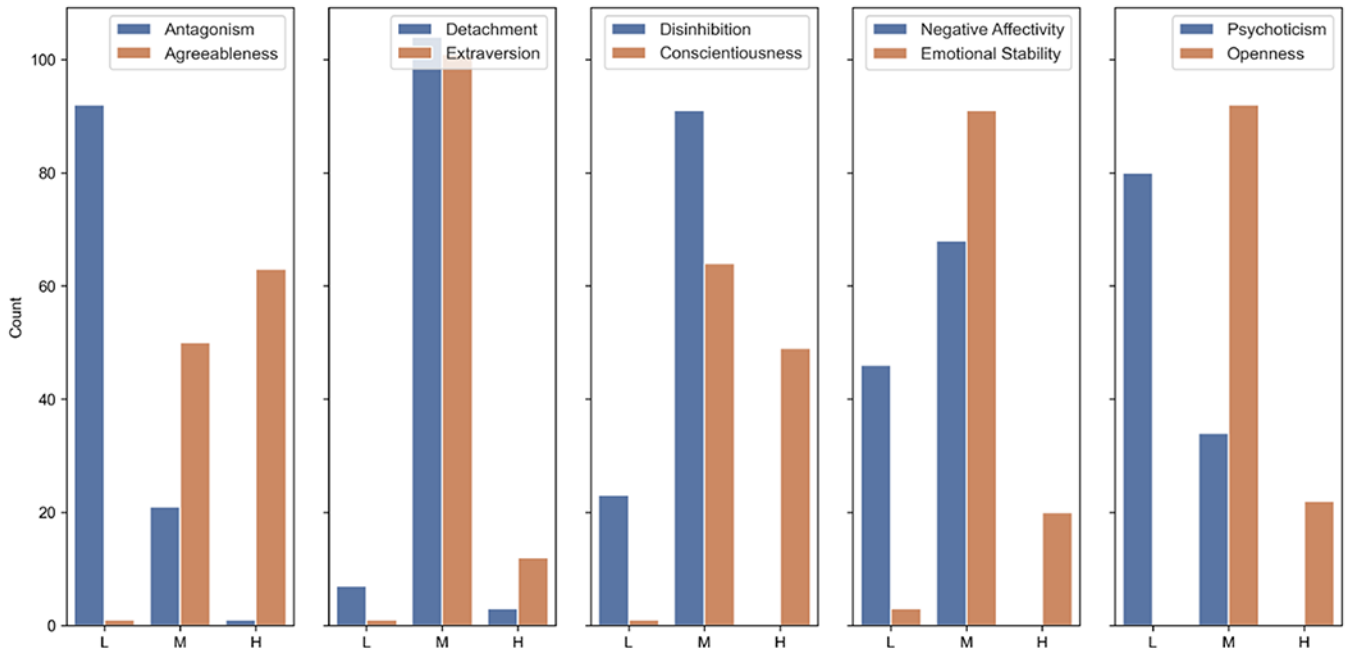
**Figure 1:**  
 Descriptions of corresponding adaptive and maladaptive traits.



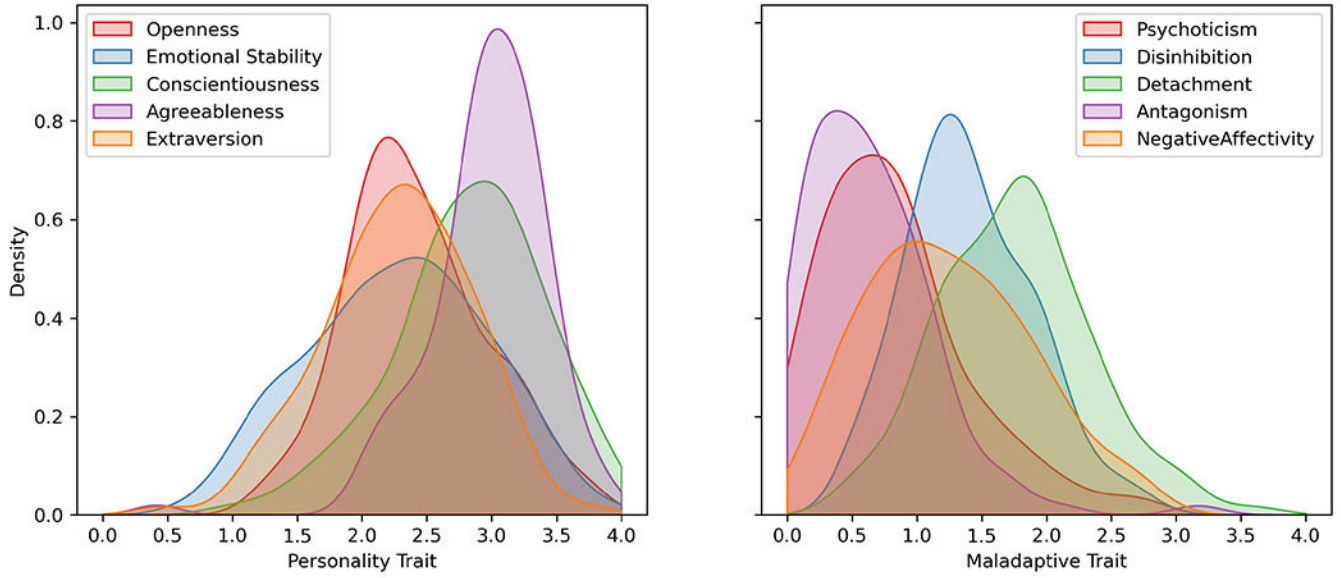
**Figure 2:** This figure shows the machine learning pipeline used to predict each personality trait. The pipeline includes nested cross-validation with two loops. The outer loop is leave-one-participant out cross-validation for evaluating models’ performance, and the inner loop is a grid search cross-validation used for tuning hyper-parameter. The figure provides the whole structure of the leave-one-participant-out cross-validation and a case of grid search cross-validation of an iteration  $i$  for the outer loop.



**Figure 3:** The process of mining association rules using the Apriori method.  $M$  is the number of highly ranked features we derived,  $V$  represents continuous values, and  $L$  represents the discrete levels. The input is the values of the day-level or person-level features selected by the machine learning pipeline, and the values of each participant’s personality traits. These values are then discretized into three levels: high (H), median (M) and low (L). The output is the association rule with the highest confidence per personality trait, indicating the most frequently occurring sensor feature combination with the personality trait of each level.



**Figure 4:** The number of participants in each group (low (L), moderate (M), and high (H) level) for each adaptive and maladaptive trait.



**Figure 5:** Density plot displays the distribution of the adaptive and maladaptive values of all participants. Each adaptive and maladaptive trait ranges from 0 to 4.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

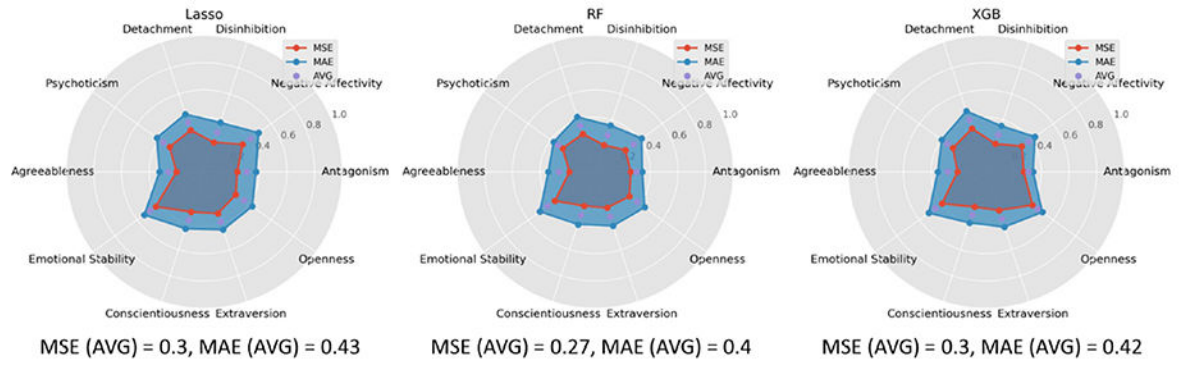


	Antagonism	Detachment	Disinhibition	Negative Affectivity	Psychoticism
Agreeableness	<b>-.71</b>	-.09	-.06	-.20	-.20
Extraversion	.07	<b>-.71</b>	-.27	-.46	-.12
Conscientiousness	-.27	-.12	<b>-.76</b>	-.35	-.37
Emotional Stability	-.13	-.35	-.49	<b>-.86</b>	-.42
Openness	.05	-.27	.15	.03	<b>.20</b>

**Figure 6:**

Bivariate correlations between maladaptive and adaptive traits. Values on the diagonal are correlations between corresponding adaptive and maladaptive variants of the same trait.

(a) Day-level



(b) Person-level

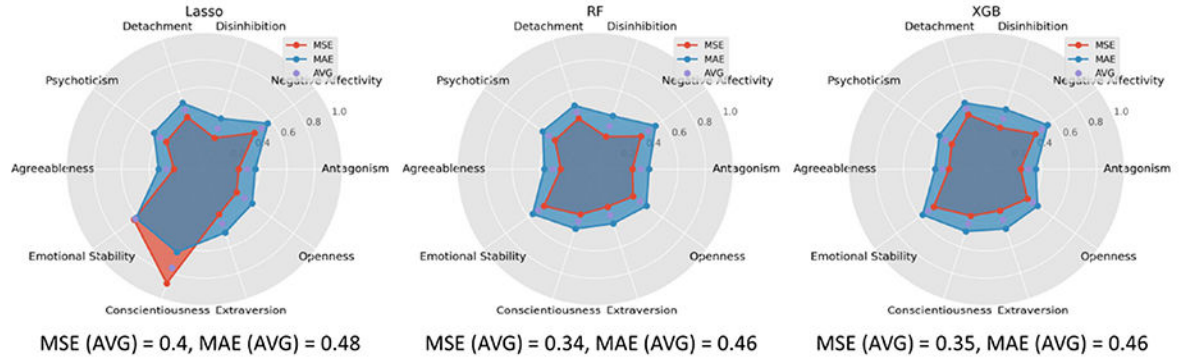
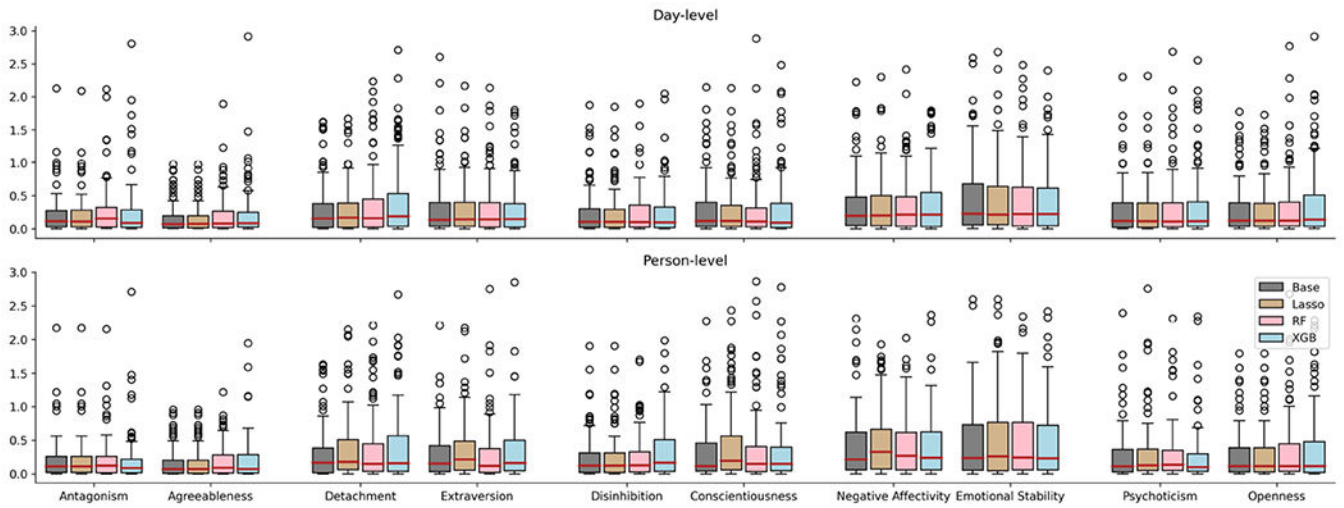
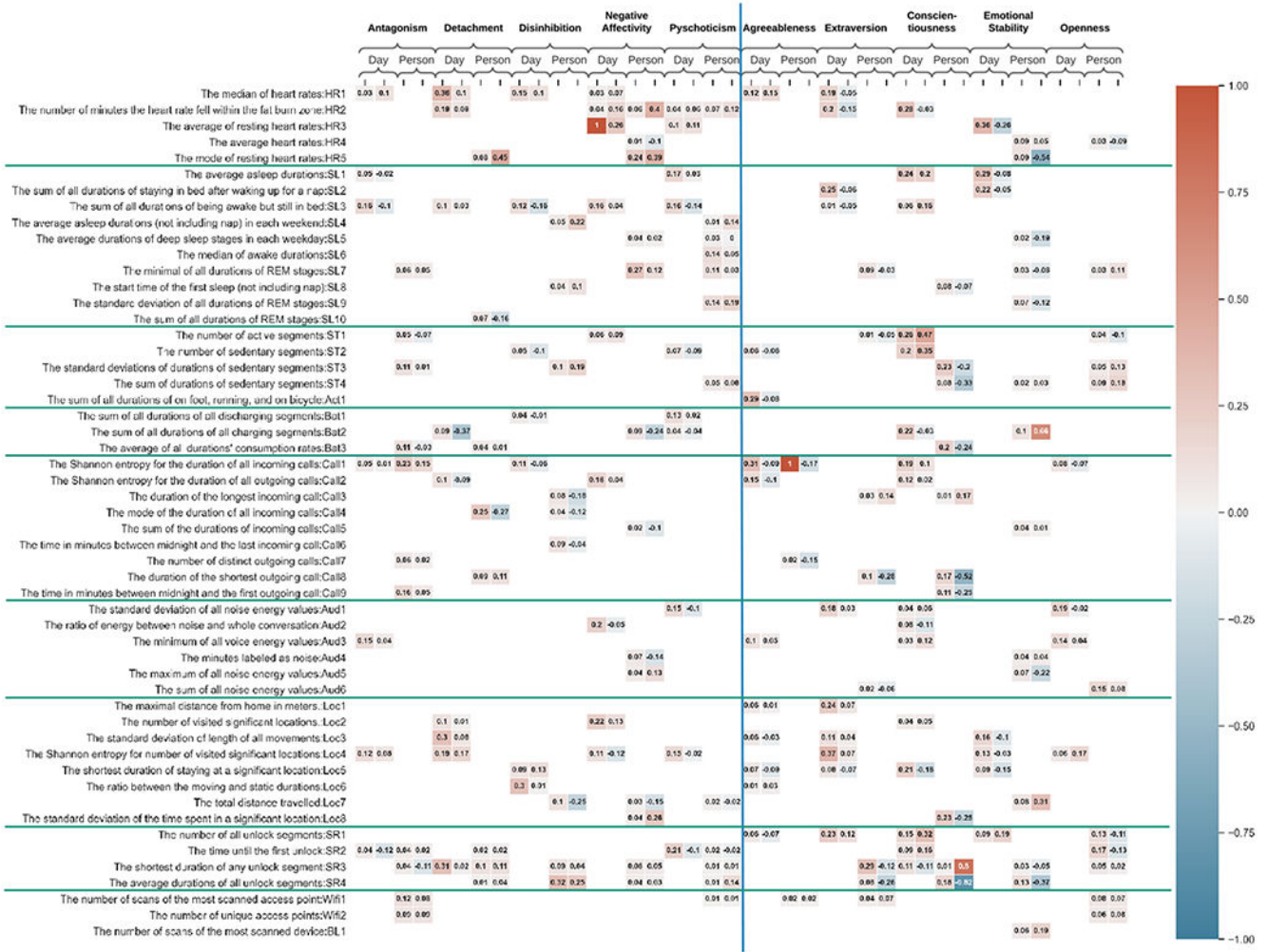


Figure 7:

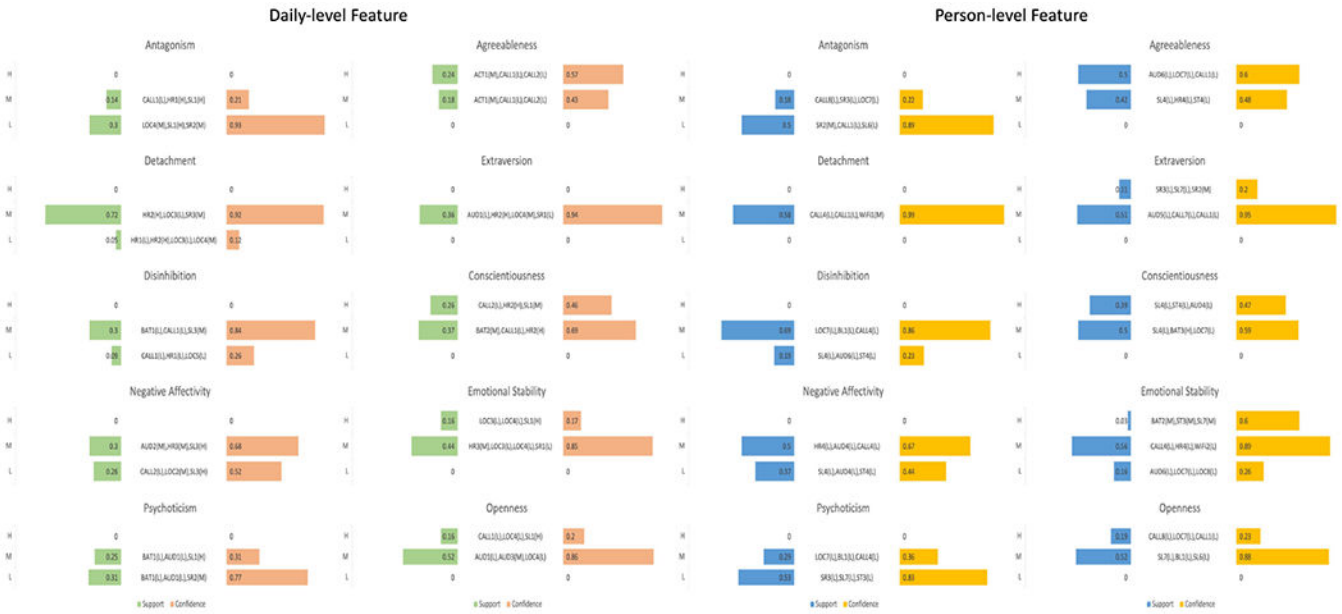
The radar plots visualize the MSE and MAE of three machine learning models (Lasso, RF, and XGB) when predicting each adaptive and maladaptive trait. The top row shows the results using day-level features, and the bottom row shows the results using the person-level features. AVG is the average of MSE and MAE per trait. MSE (AVG) is the average MSE across all traits, and MAE (AVG) is the average MAE across all traits.



**Figure 8:**  
The boxplot plots visualize the range of MSE of three machine learning models (Lasso, RF, and XGB) and the baseline when predicting each adaptive and maladaptive trait. The top row shows the results using day-level features, and the bottom row shows the results using the person-level features.



**Figure 9:** The heatmap shows the ranking for each feature assigned by the leave-one-person-out cross-validation process (Rank) and the beta coefficients (Beta) of Lasso regression with adaptive and maladaptive traits from the person-level (P) and day-level (D) models.



**Figure 10:** The figure shows the association rules between the combinations of multiple sensor features and personality traits. The used sensor features are highly ranked day-level features and person-level features shown in Table 1. Each sensor feature and trait has been discretized into three categories: low (L), moderate (M), and high (H). The combination of features with the highest confidence is shown in the figure for each level of adaptive and maladaptive trait. For the categories of traits with very low support values, the sensor features are not listed.

**Table 1:**

The column of the model shows whether the feature comes from the day-level model, participant model, or shared by the two models together (D: day-level model, P: participant model, C: day & person-level model). Nest Agg. stands for the statistical metrics used to aggregate day-level features to person-level features. NA represents no aggregation in the day-level model.

Sensor	Abbr.	Definition	Nest Agg.	Model
Heart Rate	HR1	The median of heart rates	NA	D
	HR2	The number of minutes the heart rate fell within the fat burn zone	Mean	C
	HR3	The average of resting heart rates	NA	D
	HR4	The average heart rates	Std	P
	HR5	The mode of resting heart rate	Mean	P
Sleep	SL1	The average asleep durations	NA	D
	SL2	The sum of all durations of staying in bed after waking up for a nap	NA	D
	SL3	The sum of all durations of being awake but still in bed	NA	D
	SL4	The average asleep durations (not including nap) in each weekend	Mean	P
	SL5	The average durations of deep sleep stages in each weekday	Std	P
	SL6	The median of awake durations	Max	P
	SL7	The minimal of all durations of REM stages	Min	P
	SL8	The start time of the first sleep (not including nap)	Mean	P
	SL9	The standard deviation of all durations of REM stages	Std	P
	SL10	The sum of all durations of REM stages	Min	P
Step	ST1	The number of active segments	Min	C
	ST2	The number of sedentary segments	NA	D
	ST3	The standard deviations of durations of sedentary segments	Mean	P
	ST4	The sum of durations of sedentary segments	Std	P
Activity	ACT1	The sum of all durations of on foot, running, and on bicycle activities	NA	D
Battery	Bat1	The sum of all durations of all discharging segments	NA	D
	Bat2	The sum of all durations of all charging segments	Mean	C
	Bat3	The average of all durations' consumption rates	Min	P
Call	Call1	The estimate of Shannon entropy for the duration of all incoming calls	Min	C
	Call2	The estimate of Shannon entropy for the duration of all outgoing calls	NA	D
	Call3	The duration of the longest incoming call	Mean	P
	Call4	The mode of the duration of all incoming calls	Min	P
	Call5	The sum of the durations of incoming calls	Mean	P
	Call6	The time in minutes between midnight and the last incoming call	Min	P
	Call7	The number of distinct contacts that are associated with outgoing calls	Max	P
	Call8	The duration of the shortest outgoing call	Min	P
	Call9	The time in minutes between midnight and the first outgoing call	Std	P
Audio	Aud1	The standard deviation of all noise energy values	NA	D

Sensor	Abbr.	Definition	Nest Agg.	Model
	Aud2	The ratio of energy between noise and whole conversation	NA	D
	Aud3	The minimum of all voice energy values	NA	D
	Aud4	The minutes labeled as noise	Std	P
	Aud5	The maximum of all noise energy values	Min	P
	Aud6	The sum of all noise energy values	Min	P
Location	Loc1	The maximal distance from home in meters.	NA	D
	Loc2	The number of visited significant locations.	NA	D
	Loc3	The standard deviation of length of all movements	NA	D
	Loc4	The estimate of Shannon entropy for the number of visited significant locations	NA	D
	Loc5	The shortest duration of staying at a significant location	NA	D
	Loc6	The ratio between the moving and static durations	NA	D
	Loc7	The total distance travelled	Min	P
	Loc8	The standard deviation of the time spent in a significant location	Min	P
Screen	SR1	The number of all unlock segments	Max	C
	SR2	The time until the first unlock	Max	C
	SR3	The shortest duration of any unlock segment	Min	C
	SR4	The average durations of all unlock segments	Min	P
WiFi	Wifi1	The number of scans of the most scanned access point	Std	P
	Wifi2	The number of unique access point	Std	P
Bluetooth	BL1	The number of scans of the most scanned device:BL1	Max	P

**Table 2:**

Common high ranked day-level features shared by the pair of adaptive and maladaptive traits.

ML Model	Antagonism / Agreeableness		Detachment / Extraversion		Disinhibition / Conscientiousness		Negative Affectivity / Emotional Stability		Psychoticism / Openness	
	Day-level	Person-level	Day-level	Person-level	Day-level	Person-level	Day-level	Person-level	Day-level	Person-level
Feature Abbr.	-	Call1	HR2	Call8	-	ST3	HR3	HR5	-	-
Beta	-	0.15 / -0.17	0.09 / -0.15	0.11 / -0.28	-	0.19 / -0.2	0.26 / -0.26	0.39 / -0.54	-	-
Frequency	-	111 / 111	120 / 102	108 / 108	-	114 / 114	108 / 91	66 / 89	-	-
	-	77 / 111	120 / 123	108 / 108	-	114 / 114	82 / 120	66 / 89	-	-
Importance	-	0.16 / 0.31	0.55 / 0.81	0.1 / 0.16	-	0.34 / 0.22	0.33 / 0.78	0.19 / 0.07	-	-
	-	0.16 / 1	0.4 / 0.31	0.04 / 0.22	-	0.1 / 0.04	0.25 / 0.87	0.16 / 0.04	-	-