Smartphone-derived Virtual Keyboard Dynamics Coupled with Accelerometer Data as a Window into Understanding Brain Health

Smartphone Keyboard and Accelerometer as Window into Brain Health

Emma Ning
University of Illinois Chicago
Chicago, Illinois, USA
yning7@uic.edu

Andrea Cladek
University of Illinois Chicago
Chicago, Illinois, USA
acladek2@uic.edu

Mindy K. Ross
University of Illinois Chicago
Chicago, Illinois, USA
mross27@uic.edu

Sarah Kabir
University of Illinois Chicago
Chicago, Illinois, USA
skabir4@uic.edu

Amruta Barve
University of Illinois Chicago
Chicago, Illinois, USA
amruta@uic.edu

Ellyn Kennelly
Wayne State University
Detroit, Michigan, USA
ellynk@wayne.edu

Faraz Hussain
University of Illinois Chicago
Chicago, Illinois, USA
farazh@uic.edu

Jennifer Duffecy
University of Illinois Chicago
Chicago, Illinois, USA
jduffecy@uic.edu

Scott L. Langenecker
University of Utah
Salt Lake City, Utah, USA
s.langenecker@hsc.utah.edu

Theresa Nguyen
University of Illinois Chicago
Chicago, Illinois, USA
tnguy271@uic.edu

Theja Tulabandhula
University of Illinois Chicago
Chicago, Illinois, USA
theja@uic.edu

John Zulueta
University of Illinois Chicago
Chicago, Illinois, USA
jzulueta@uic.edu

Olusola Ajilore
University of Illinois Chicago
Chicago, Illinois, USA
oajilore@uic.edu

Alexander P. Demos
University of Illinois Chicago
Chicago, Illinois, USA
ademos@uic.edu

Alex Leow
University of Illinois Chicago
Chicago, Illinois, USA
alexfeuillet@gmail.com

ABSTRACT

We examine the feasibility of using accelerometer data exclusively collected during typing on a custom smartphone keyboard to study whether typing dynamics are associated with daily variations in mood and cognition. As part of an ongoing digital mental health study involving mood disorders, we collected data from a well-characterized clinical sample (N = 85) and classified accelerometer data per typing session into orientation (upright vs. not) and motion (active vs. not). The mood disorder group showed lower cognitive performance despite mild symptoms (depression/mania). There were also diurnal pattern differences with respect to cognitive performance: individuals with higher cognitive performance typed faster and were less sensitive to time of day. They also exhibited more well-defined diurnal patterns in smartphone keyboard usage: they engaged with the keyboard more during the day and tapered their usage more at night compared to those with lower cognitive performance, suggesting a healthier usage of their phone.

CCS CONCEPTS

• **Applied computing** → **Health informatics**; • **Human-centered computing** → *Smartphones*.

KEYWORDS

Health - Clinical, Mobile Devices: Phones/Tablets, Empirical Study that tells us about people, Quantitative Methods

ACM Reference Format:
Emma Ning, Andrea Cladek, Mindy K. Ross, Sarah Kabir, Amruta Barve, Ellyn Kennelly, Faraz Hussain, Jennifer Duffecy, Scott L. Langenecker, Theresa Nguyen, Theja Tulabandhula, John Zulueta, Olusola Ajilore, Alexander P. Demos, and Alex Leow. 2023. Smartphone-derived Virtual Keyboard Dynamics Coupled with Accelerometer Data as a Window into Understanding Brain Health: Smartphone Keyboard and Accelerometer as Window into Brain Health. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3580906

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3580906

1 INTRODUCTION

Traditionally, healthcare professionals have used patient self-reports and in-person evaluations collected at clinical visits to assess behavior, affect, and cognition [44]. However, these assessments are limited by biases in patients' retrospective recall and lack of generalizability from lab or clinic to real world settings [5, 44]. The adoption of Ecological Momentary Assessment (EMA) offers a more ecologically valid and time-sensitive approach by repeatedly sampling data from patients in real time in their normal environments [44]. However, EMA data still relies on self-report, thus requiring active participation from users [32]. Smartphone-based mobile health applications can potentially overcome these limitations by leveraging a variety of passively collected data obtained via GPS, accelerometer, keyboard typing activity and other phone features. These data provide a glimpse into an individual's behavior and cognitive functioning in dynamic, naturalistic environments, and may be used to extract clinically relevant, ecologically valid biomarkers for mental and brain health. Smartphone keyboard and accelerometer data may be especially useful in monitoring behavioral and cognitive features of neuropsychiatric conditions such as Parkinson's disease and bipolar disorder, which show psychomotor and cognitive abnormalities [2]. Additionally, passive collection of smartphone sensor data allows for near-continuous monitoring of behaviors over time. Therefore, by leveraging keyboard typing dynamics and accelerometer data, we may identify distinctive patterns in participants' phone motion and orientation across naturalistic contexts.

Our team recently developed a customized, virtual smartphone keyboard that replaces the native iPhone and Android keyboard, allowing us to passively and unobtrusively collect keystroke metadata, such as inter-key delay, autocorrect, backspace usage, and accelerometer data during typing. So far, we have been able to link keyboard dynamics with mood in individuals with bipolar and depressive disorders [55]. We have also demonstrated keyboard dynamics fluctuations over the course of a day (i.e., diurnal pattern) as a function of participant age and mood [51]. Building on these results, we now extend this work by integrating accelerometer data with keyboard dynamics to investigate the interrelationship between activity patterns related to phone usage, cognitive performance, and time of day. By using a sensor fusion approach, we hypothesize that we will gain a more comprehensive understanding of brain functioning compared to using keyboard dynamics alone. Specifically, accelerometer data sampled during typing will provide information about phone orientation indicating an individual's typing position, such as typing upright or lying down, and whether they are in motion while typing.

1.1 Related work

1.1.1 *Using smartphone accelerometer data to gauge user activity.*

Accelerometers are often used to capture acceleration in their component axes to detect changes in motion and orientation. In this section, we review the use of wrist-worn and smartphone accelerometers to capture fluctuations in activity/movement.

Accelerometers are widely used in wrist-worn devices to track motor activity and sleep movements. Wrist-worn accelerometers can be used to measure tremors throughout the day in patients with Parkinson's disease, capturing the effects of treatment on tremor duration [50]. Additionally, accelerometers can track sleep disturbance and movement in individuals with post-traumatic stress disorder (PTSD) [49]. Individuals with PTSD exhibit more nocturnal activity and greater intraday variability in activity levels compared to controls, allowing for the differentiation between groups. In these studies, accelerometer data are acquired passively and unobtrusively from wrist-worn devices to quantify the amount of movement in clinical populations. Further, the variability in sleep quality and subsequent changes in intraday activity levels imply that frequent sampling of activity patterns throughout the day provides intra-subject variability important in diagnostics and treatment efficacy.

In addition to wrist-worn devices, accelerometers have been increasingly used in mobile computing. Researchers demonstrated that smartphone sensor data (including accelerometers) can predict bedtime, waketime, sleep duration, and sleep quality with reasonable accuracy and without requiring participants to wear wrist-worn sensors or place smartphones in close proximity [30]. Additionally, smartphone accelerometers can classify user activity (e.g., walking, running, in transit) and quantify the amount of movement during time intervals of interest [4, 36]. A recent study found that accelerometer data can be combined with contemporaneous mobile app data to uncover users' physical movement patterns specifically when interacting with their phones. Users moved their phones more when using messaging or navigation apps compared to browsing, and preferred portrait orientation unless streaming videos [41]. Accelerometer data can also provide the orientation of the phone because it always measures its relationship to gravity, allowing an inference of the person's posture (upright vs. laying down). Since multiple studies reported that patients with mood disorders experienced sleep disturbances [11, 24, 38], it is important to capture smartphone orientation during typing throughout the day. Thus, in addition to capturing an individual's general activity level, accelerometer data can be used to convey the nuances of an individual's interactions with their smartphone over various digital contexts. Further, prior studies have explored using multiple sensor streams (including accelerometer data) from wearable devices to characterize clinical populations such as individuals with bipolar disorders, depression, and schizophrenia (for review, see [43]). They found that over a total of 35 studies, 16 including accelerometer data, physical activity was associated with clinical symptom severity.

1.1.2 *Using smartphone keyboard dynamics to predict cognitive performance.*

Due to the pervasive adoption of smartphones, we can obtain big data per person to model intra-individual variability over time. Because of this sensitivity to both inter and intra-individual variability, prior studies have used keyboard dynamics as a password authentication mechanism and as proxies for emotion, stress, mood, and cognition.

People differ in their finger agility, thought processes, inherent rhythm, and touch angle and locations during interactions with a virtual keyboard [23]. Because of these factors, typing/touch behaviors and gestures are indicative of muscle behavior that can be used to differentiate between users [53]. With increasing computing capacity, traditional password authentication is vulnerable to attacks, and research has been focusing on incorporating biometric modalities such as typing dynamics to authenticate users [19, 45]. Overall, studies concluded that features extracted from typing dynamics were effective in classifying genuine users from impostors, and they also pointed out the huge intra-individual variability in keystroke dynamics affected by posture, user activity, and whether the keyboard allowed error-correcting mechanisms during typing [10]. Inevitably, some of the design factors of the keyboard added additional variability to keystroke dynamics. One way to reduce the variability is to restrict features, such as disabling one-handed keyboard or landscape mode to force users to fix hand postures [10]. The virtual keyboard used in the current study disables landscape mode and one-handed keyboard to minimize intra-individual variability.

Apart from behavioral biometrics, features derived from keystroke dynamics have been used to infer users' psychological state, a key element in designing more emotionally intelligent applications/systems. Prior studies found that timing features (e.g., press duration, keystroke latency) from keystroke dynamics successfully classified emotional states (e.g., frustration, anger, happiness, etc., [17]). In addition, combining keystroke timing and user traits predicted boredom and engagement during typing sessions [7], indicating that keystroke dynamics contain information about affective states and cognitive involvement.

Smartphone typing involves both psychomotor and visuospatial components. For an individual to form and communicate their thoughts in an efficient and coherent way, they need to coordinate the visual search for keys to achieve reasonable speed while monitoring for errors [22]. Since keystroke dynamics tap into these cognitive aspects, prior studies have used these features to detect sobriety from inebriation with the aim of preventing drunk driving [28]. Researchers showed that keystroke dynamics combined with reaction time tasks in a custom app estimated a person's blood alcohol level closely approximating what was produced by a breathalyzer [28]. A different study aiming at reducing binge-drinking episodes by providing spontaneous digital support used a sensor fusion approach and showed that high-risk drinking episodes were characterized by higher movement levels, more phone usage, slower typing speed, and more text deletions [3]. These studies show that altered cognitive states are detectable through keystroke dynamics.

Keystroke dynamics captured in the wild are also used to detect fine motor impairments typically associated with neurodegenerative disorders and mood disorders, such as Parkinson's disease and depression. Features derived from keystroke dynamics (such as key hold time and flight time) can successfully classify individuals' Parkinson's disease status (i.e., diagnosis groups assessed by clinicians [21, 48]). Apart from evidence showing that keystroke dynamics converge toward clinically validated assessment, they also predicted self-rated depression scores [29]. In addition, intensively sampled keystroke dynamics analyzed at lower levels (i.e., at different timepoints of a day) reveal detailed fluctuations in cognitive functioning. For example, our past study found that typing speed slowed with age and followed a diurnal pattern (changes with a period of 24 hours) that resembled the wakefulness pattern throughout the day [51], suggesting keyboard dynamics' role as a promising behavioral biomarker to measure cognitive performance. Furthermore, one study from our team used keyboard typing dynamics such as inter-key delay (IKD; time lapse between consecutive keys), typing session length (the number of keypresses within each typing session), and the amount of self-monitoring during typing (measured using the percentage of backspaces that occurred within each typing session) to predict participants' brain age [54]. Another study found that keystroke dynamics differences reflected participants' performance on an array of cognitive domains, including behavioral inhibition and processing speed in a multiple sclerosis sample [12].

In a mood-disordered clinical sample, using a direct measure of cognitive performance, we showed in a recent study that participants who typed slower took longer to complete a digital version of the Trail Making Test, Part B (dTMT-B), a validated measure for processing speed and executive functioning. In addition, typing speed was sensitive to fluctuations in cognitive performance: on days when participants typed slower, they also took longer to complete the dTMT-B relative to their own baselines [40]. Furthermore, this study showed that individuals who were more depressed experienced lower cognitive functioning compared to both others and relative to oneself. This approach, however, only uses one sensor stream (i.e., keystroke dynamics).

Very few studies to-date have employed a sensor fusion approach to model accelerometer data contemporaneously with keyboard dynamics in the context of inferring brain health. A recent study demonstrated the importance of temporally aligning data from heterogeneous sensor streams that are usually sampled at different frequencies (e.g., keystroke dynamics, motor function measured using accelerometer and gyroscope, physiological markers such as heart rate, sleep sensors, and EMA self-ratings) [13]. The authors used this sensor-fusion approach and reported the important features that classified participants into healthy controls, mild cognitive impairments, or early Alzheimer's disease. Notably, typing speed measured using an iPad typing task was one of the most informative features. Another study that combined accelerometer data with keystroke dynamics reported postural accelerations, keystroke timing (i.e., typing speed), and press duration to be useful predictors of Parkinson's disease [34]. Our current study expands on this sensor fusion approach by combining accelerometer data and keystroke dynamics to examine user-keyboard interaction patterns throughout the day. In addition, we assess whether this intra-individual pattern differs based on cognitive functioning and clinician assessments.

1.1.3 Using EMA tasks and accelerometer data to detect diurnal fluctuations.

Both accelerometer and EMA implemented in smartphones can frequently sample from users, affording researchers the opportunity to measure intraday variability in activity and cognitive functioning. Previously, we mentioned that wrist-worn accelerometers measured sleep quality and subsequent changes in activity patterns. To examine intraday fluctuations in alertness and cognition without requiring participants to wear wrist-worn devices, researchers designed smartphone applications that included toolkits composed of a few tasks, including the Psychomotor Vigilance Task, the Go/No-Go task, and asked participants to complete the toolkit assessments multiple times a day. Using response time measures, they reported that participants' cognitive speed deteriorated across the course of a day [1, 15]. These fluctuations that have a period of 24 hours are commonly referred to as diurnal fluctuations. Despite showing diurnal changes in cognitive performance, these studies did not use accelerometer data to passively capture movement and thus required active user participation. Additionally, they did not examine non-linear diurnal fluctuations in cognitive performance. Since people are generally most alert in the mid-mornings, less alert in the early afternoons, and then become alert again in the late afternoons and early evenings [42], testing for non-linear effects in cognitive performance would uncover more nuanced diurnal fluctuations above and beyond the linear effects.

In the current study, we combine frequently sampled accelerometer and keystroke dynamics data captured in the wild to study: 1) whether accelerometer and/or keystroke dynamics data present diurnal patterns; 2) whether these diurnal patterns follow a linear and/or a quadratic trend; 3) whether inter-individual differences explain differences in diurnal patterns.

2 METHOD

2.1 Participants

The study participants consisted of healthy controls (n = 26; women = 12) and individuals with mood disorder (n = 64; women = 44) between ages 25 and 50. At the first study visit, psychiatrists determined participants' eligibility: those with acute suicidal ideation, severe cognitive impairment, active alcohol and/or substance use disorders, major medical or neurological illness interfering with protocol adherence, contraindications to MRI, or that did not own smartphones were ineligible for the study. Participants were recruited using community outreach efforts such as listserv emails, social media posts, and clinical trial registries, and participant eligibility was determined after the completion of an initial screening session. Informed consent was obtained from all participants prior to the administering of the screening. Individuals in the mood disorder group were required to meet the DSM-5 criteria for major depressive disorder, bipolar disorder type I/II, persistent depressive disorder, or cyclothymia.

2.2 Data collection

Eligible study participants completed a virtual screening visit (Visit 1) and two in-person study visits (Visit 2 and Visit 3 to complete clinical symptom interviews) over the course of four to five weeks (see Figure 1 for the recruitment procedure).

2.2.1 Keyboard data.

During Visit 1, all eligible participants were instructed to download a custom-built smartphone keyboard, BiAffect (https://www.biaffect.com/), that replaced the standard iOS/Android keyboard on their smartphones. Participants completed an additional e-consent that described the goal of capturing keystroke dynamics and the extent of metadata the custom keyboard collected before enabling the keyboard. All eligible and consenting participants were instructed to use this keyboard exclusively for the duration of the study and participants' compliance was checked on a daily basis. Note that keyboard dynamics were recorded only in portrait mode and the custom keyboard does not support swiping.

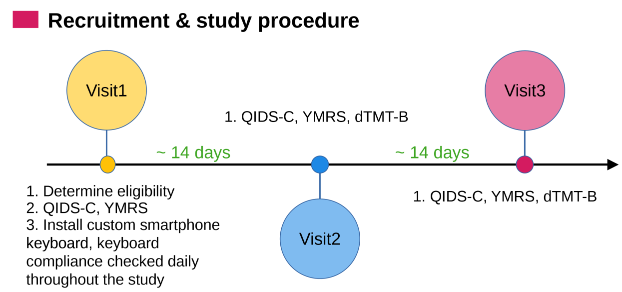

Figure 1: Recruitment and study procedure. QIDS-C: Quick Inventory of Depressive Symptomology; YMRS: Young Mania Rating Scale; dTMT-B: Digital Trail Making Task, part B.

A keyboard session begins as soon as the virtual keyboard is presented on-screen, when the user presses the first key, and terminates when the keyboard is no longer rendered or after 6 seconds of keyboard inactivity. Keyboard sessions are associated with unique identifiers with corresponding timestamps recorded in UTC (as well as the specific time zone the user is in). The keyboard records the timing of keypress events and their general category (i.e., alphanumeric, backspace, autocorrection, punctuation, suggestion, special characters, etc.). Note: that the actual text is not recorded. In addition, accelerometer data is recorded at a rate of 10 Hz along when a keyboard session is active.

2.2.2 Cognitive data.

At Visits 2 and 3, participants completed cognitive assessments using an adapted smartphone-based digital version of the Trail Making Test Part-B (dTMT-B), an executive functioning measure of visual attention, processing speed, and set-switching [8, 52].

2.2.3 Mood data.

As part of the research protocol, each study participant was evaluated by one of the three study psychiatrists (authors) at each visit using the Quick Inventory of Depressive Symptomatology (QIDS-C) and the Young Mania Rating Scale (YMRS).

2.3 Data processing

2.3.1 Keyboard data pre-processing.

Table 1 lists the relevant keyboard dynamics variables and their descriptions, as guided by recent published studies using keyboard dynamics. Keypress-level variables vary across keypresses for the duration of a session, whereas session-level variables are fixed for all keypresses within a session.

2.3.2 Accelerometer pre-processing.

Accelerometer data was used to determine: 1) Orientation (upright vs. non-upright): approximate phone orientation during typing and, 2) Motion: the percentage of active accelerometer samples in a typing session.

Before calculations were performed on accelerometer data to quantify orientation and motion, raw accelerometer data was first pre-processed. Each axis of the raw accelerometer was low-pass filtered (LPF) using a second-order Butterworth bidirectional filter (i.e., zero-phase filter) with a cutoff frequency at 4 Hz to remove

Table 1: Variables based on raw keyboard data, and both keypress-level and session-level metrics calculated from raw keyboard data.

Variables	Description
Transition type (Keypress-level)	The type of the current keypress and the type of the next keypress. For example, if the user is typing consecutive characters, then the transition type will be alphanumeric-alphanumeric; if the user is typing backspace then character, then the transition type will be backspace-alphanumeric, etc.
Distance from center (Keypress-level converted to session level)	A measure of how far the user's keypress was from the center of the key they pressed. We calculated the median distance from center for each typing session as measure of psychomotor performance.
Inter-key delay (see Figure 2) (IKD; Keypress-level converted to session level)	The time lapse between consecutive keys measured in milliseconds. Specifically, IKD is calculated in the same direction as transition type: next keypress's timestamp minus the current keypress's timestamp. IKD distribution was strongly positively skewed [35]. We calculated the median IKD for each typing session as measure of typing speed. Also, we calculated the variability of IKD using the Median Absolute Difference (MAD). See Figure 2 for the IKD distribution across sessions for one example user.
Handedness (Session-level)	The typing mode (i.e., whether the user was typing one-handed or two-handed) was determined for each session. To classify typing mode, IKD was correlated with the distance traveled from the previous key within each session and significant positive correlations were classified as two hand typing (> 95% accuracy in test dataset, see [51] for algorithm used).
Time of Day (TOD; Session-level)	Sessions that started between 6:00 and 11:59 were labeled as Morning, 12:00 and 17:59 as Afternoon, 18:00 pm and 23:59 as Evening, and 0:00 and 5:59 as Night.

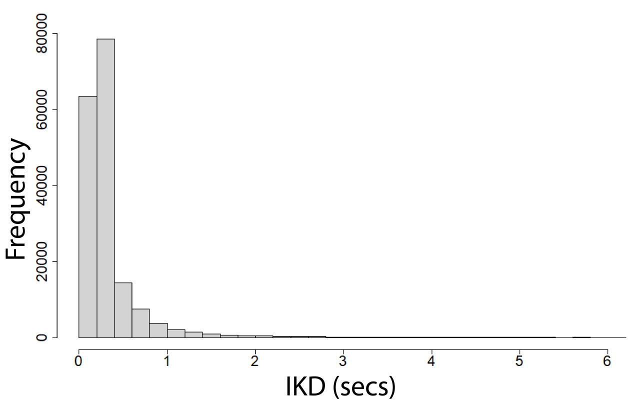

Figure 2: IKD distribution of one user across all keyboard sessions.

high frequency noise (Figure 3). 4 Hz was selected because daily human activities (including standing up, sitting down, stepping, and walking) were shown to have frequency components within the 0.3 to 3.5 Hz range [25, 47]. Then at each timepoint, a Euclidean norm (i.e., magnitude) was calculated using the low-pass filtered data to determine whether the phone was accelerating or decelerating with a total magnitude substantially different than gravity.

To determine the amount of motion for each session, we calculated the overall magnitude by combining the X, Y, Z accelerometer axes using $\sqrt{X^2 + Y^2 + Z^2}$. Note that a phone that is completely still would have an overall magnitude of 1, representing no deviation from gravity.

Next, each accelerometer sample was labeled as "not active" if its magnitude fell between 0.95 and 1.05, and "active" if outside this range. This range was determined using a ground truth dataset of typing during sitting. The accelerometer threshold was chosen to contain 95% of the data. We then calculated the percentage of active accelerometer samples within each typing session, which will be referred to as "motion", as the phone was in movement during a typing session.

To approximate phone orientation for each session from the accelerometer data, the median acceleration of the raw data for X and Z axes of that session were calculated. Based on these medians, each session orientation was labeled as either "not upright or "upright". Specifically, the session would be labeled as "upright" if its median acceleration in the Z axis was less than 0.1, and its median acceleration in the X axis is between -0.2 and 0.2, inclusive bounds. The cutoffs were obtained after pilot testing using a protractor to measure angles the phone was being tilted at. Z axis accounted for the phone screen orientation, and X axis accounted for the screen tilt (towards or away from the user, as typically seen in users that lay on their sides and type). The median acceleration was used instead of the mean because each axis' acceleration tended to be skewed, with peaks occurring at their respective fixed value (-1, 1, or 0) anticipated during normal typing (i.e., after cleaning out keypresses that involve excessive movements using the Euclidean norm).

2.4 Statistical analyses

The aim of BiAffect is to eventually replace infrequent clinician assessments with temporally dense samples collected from smartphones and wearables. Thus, by analyzing keyboard and accelerometer features as dependent variables, we are able to establish detailed

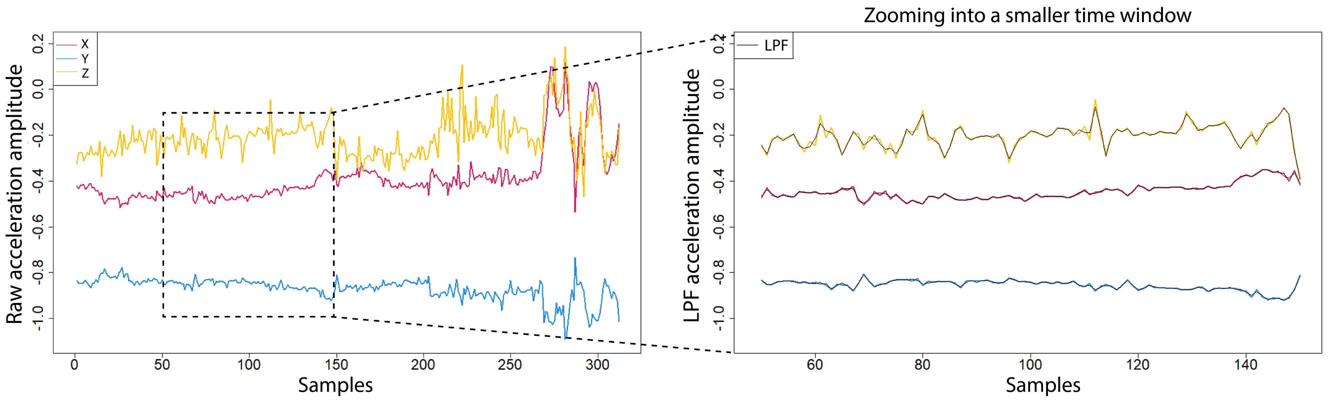

Figure 3: Raw (left panel) and low-pass filtered accelerometer data overlaying raw data (right panel) from a keyboard session of the same example user.

individual behavioral profiles over time. This personalized profile highlights not only the interconnections between different digital behaviors, but also the relationships between these behaviors and cognition, mood, and diurnal activities.

First, we report overall group differences using between-subjects t-test and chi-squares on important demographic characteristics of the healthy controls (n = 25) versus the mood disordered group (n = 60). Next, we report four mixed effects models with different dependent variables fitted using the glmmTMB package in R [9]. Out of the four models, two models predicted features based on keyboard dynamics, and the other two models predicted features based on accelerometer data: the first model predicted log number of keypresses per session, the second model predicted median IKD per session, the third model predicted the percentage of accelerometer samples that were in motion per session, and the fourth model predicted the probability of upright typing per session. All four models used alphanumeric to alphanumeric keypress transition types that took place in two-handed sessions, as determined using [51] with at least 20 keypresses to ensure the keyboard dynamics was representative of each individual. The decision to include only a subset of all typing sessions is because: 1) typing speed and accuracy differ considerably between one-handed vs. two-handed typing [14, 16, 18, 23], and our participants typed mostly two-handed (about 98% of all sessions were two-handed); 2) cognitive processes reflected through alphanumeric-alphanumeric keypress transitions are very different from other keypress transition types, e.g., alphanumeric-backspace. The former reflects psychomotor and visuomotor processes, whereas the latter reflects error-monitoring [14].

For all models, fixed effects from predictors QIDS, YMRS, age, log number of keypresses within session, session median distance from key center, dTMT-B completion time, median and MAD IKD, were all z-scored to improve the interpretability of intercepts, interaction terms, and resulting regression estimates such that all estimates would be on similar scales. All categorical predictors, including orientation (not upright vs upright), medication (medication vs. no medication), gender (women vs. men), were dummy-coded. TOD was calculated as clock time (corrected for time zone) as the data was unobtrusively collected and we do not know the time in which the individual began their day beyond their first typing session. Both orthogonal and non-orthogonal second-order polynomials for TOD were fitted. The former was used to isolate (i.e., semi partial) and interpret the significance of linear and quadratic effects of TOD. The latter was used to give interpretable regression estimates for the polynomial terms in the regression summary tables. Finally, because participants completed multiple visits, the clinical ratings (QIDS-C and YMRS) and dTMT-B scores were averaged for each participant.

The random effects structure for all four models included two random effects. The first random intercept was the same for all models: the individual was the cluster in which days (i.e., each day of using the keyboard) were nested, and nested within each day was the Time of Day (TOD) the typing session occurred. Then we allowed the random slope of linear and quadratic effects of TOD to vary within each day. In other words, the first random effect allowed each user to have their own unique TOD effect that varies across days. The second random effect included the individual as the cluster in which orientation, or motion, was nested. This allowed the random slope of orientation, or motion, to vary within the user. The random slope was changed based on what the model was predicting such that the random slope was not specified using the same variable as the outcome variable.

Figures were created by estimating the predicted values from the fixed effects of each model, specifically for the interaction between diurnal pattern (linear and quadratic TOD) and cognitive performance (dTMT-B). Each figure shows standard error bands (SE) generated from model predictions. Because this is a multi-level model with multiple within-subjects effects, significant differences cannot be easily inferred from the overlap of the SE bands as they do not remove between-subjects variation that the model accounted for as random effects. They are provided to give an estimate of overall within and between-subject variance as well as the amount of data at each TOD.

3 RESULTS

3.1 Overview

In Table 2, we report overall group differences on important demographic characteristics of the healthy controls (n = 25) versus the mood disordered groups (n = 60). First, to understand typing behavior and its relationship to cognitive performance, Table 3 reports the Gaussian mixed-effects models which examined cognitive performance interacted with time of day. Next, to understand the relationship between cognitive performance and accelerometer data over the course of the day, Table 4 shows the results of the Gaussian and logistic mixed-effects models which examined cognitive performance interacted with accelerometer data during typing over the course of a day. Each analysis contained 74,863 typing sessions within a total of 637 days.

Group affiliation (healthy control vs. mood disorder) was not entered into any of the mixed effects models because we entered depression and mania scores as well as cognitive performance scores, and these group differences replaced group affiliation (see Table 2). If we included group affiliation and these terms, it would have confounded our interpretation of the models. Note, for time of day effects, we report the slopes as non-orthogonal polynomials to increase interpretability of the estimates, and the *p*-values as orthogonal polynomials to ensure the independence of the terms.

3.2 Group differences

Individuals were recruited to ensure no age or gender differences exist between healthy controls and the mood disorder group, which was confirmed in Table 2. Unexpectedly, we found that the mood disorder group had significantly more typing sessions than the healthy controls, however, both groups typed using their two hands at roughly equal rates (determined using the algorithm in [51]). As expected, the mood disorder group was assessed to have significantly higher scores on depression (QIDS), mania (YMRS), and medication usage (any medication, not just psychotropic medications). However, it is important to note that the mood disorder group would be considered to have mild symptom severity given the low scores. Further, as expected, the mood disorder group showed significantly lower cognitive performance on the Digital Trail Making Test, Part B (dTMT-B) relative to the healthy controls, consistent

Table 2: Demographic information and differences between healthy controls and mood disorder group.

		Healthy Control (N = 25)	**Mood Disorder (N = 60)**	***p-value***
Age				
	Mean (SD)	31.4 (8.20)	34.0 (6.69)	0.159
	Median [Min, Max]	28.0 [24.8, 50.6]	32.1 [25.0, 50.1]	
Gender				
	Women	12 (48.0%)	32 (53.3%)	0.834
	Men	13 (52.0%)	28 (46.7%)	
Medication				
	Medication	5 (20.0%)	32 (53.3%)	**0.010**
	No medication	20 (80.0%)	28 (46.7%)	
YMRS (Mania Score)				
	Mean score (SD)	0.080 (0.40)	1.30 (2.37)	**<0.001**
	Median [Min, Max]	0 [0, 2.00]	0.167 [0, 14.0]	
QIDS (Depression Score)				
	Mean score (SD)	0.507 (0.746)	5.82 (4.35)	**<0.001**
	Median [Min, Max]	0 [0, 2.00]	5.00 [0, 18.0]	
Number of Typing Sessions				
	Mean (SD)	626 (668)	1080 (826)	**0.011**
	Median [Min, Max]	459 [46.0, 3110]	902 [2.0, 3880]	
Two-Handed Typing				
	Mean percentage (SD)	93.2 (2.53)	93.6 (2.28)	0.476
	Median percentage [Min, Max]	93.7 [84.8, 98.1]	94.0 [86.8, 100]	
dTMT-B (Digital Trail Making-B)				
	Median seconds (SD)	13.2 (4.80)	16.9 (7.45)	**0.007**
	Median seconds [Min, Max]	12.1 [8.00, 27.3]	14.7 [7.55, 45.8]	

with longstanding evidence demonstrating trait deficits in cognition in individuals with mood disorders relative to healthy controls, even in a euthymic state [37, 39].

3.3 Typing behavior

As seen in Table 3, we first examine naturalistic typing by looking at both the number of keypresses per session and the median inter-key delay (as an overall proxy for real-world processing speed) using Gaussian mixed effects models. We entered a series of fixed effect predictors that tested our main questions of interest, a) the relationship between typing dynamics and accelerometer data (i.e., motion and orientation), b) diurnal trajectories (time of day), and c) the relationship between diurnal trajectories and cognitive performance as measured by the dTMT-B (i.e., processing speed and executive function) into each model. We also simultaneously entered a series of control variables which included: medication status, depression and mania scores, age, gender, the degree of typing accuracy as measured by the distance of a keypress touch event to the center of each key recorded, and where appropriate other metrics of their typing dynamics (such as variability of typing speed).

3.3.1 Number of Keypresses per Session.

Table 3 contains the results of the number of keypresses in log, therefore we report the exponentiated betas for interpretability in the text. Overall, we found that men typed about 45 alphanumeric-to-alphanumeric keypresses per typing session, $b = 3.80$, $e^b = 44.7$, $p < .001$, and women tended to type more than men by about 1.1 keypresses per session, $b = .11$, $e^b = 1.1$, $p = .006$, when they were in the non-upright mode of typing with minimal movement. Further, with every percentage point increase in the active accelerometer samples per session, people typed about one keypress fewer. While small, the effect is significant, $b = -.04$, $e^b = .96$, $p < .001$. Interestingly, upright typing sessions on average were one keypress fewer relative to non-upright sessions. This effect was also small, $b = -.03$, $e^b = .97$, $p = .010$.

Table 3: Linear Mixed Effects Model of Number of Keypresses (in Log) and Median IKD.

		Number of Keypresses (in Log)		Median IKD	
Fixed Effect Predictors		*b*	*p*	*b*	*p*
	(Intercept) [Mean of men when non-upright/non-active]	3.80	**<0.001**	244.21	**<0.001**
	Women [vs Men]	0.11	**0.006**	-7.22	0.401
Predictors of Interest					
	Motion	-0.04	**<0.001**	-1.11	**0.034**
	Orientation [Upright typing session]	-0.03	**0.010**	-17.55	**<0.001**
	dTMT-B (Cognitive Performance)	-0.04	**0.022**	5.75	0.161
	Time of Day [1st degree]	-0.01	**<0.001**	-6.60	**<0.001**
	Time of Day [2nd degree]	-0.01	0.085	3.30	**<0.001**
	Time of Day [1st degree]*dTMT-B	-0.02	0.397	-0.89	0.051
	Time of Day [2nd degree]*dTMT-B	0.01	**0.029**	0.57	**0.040**
Control Variables					
	Medication [No medication]	-0.03	0.490	-0.07	0.993
	QIDS (Depression score)	0.03	0.215	1.06	0.840
	YMRS (Mania score)	-0.02	0.208	-2.79	0.387
	Age	0.08	**<0.001**	19.24	**<0.001**
	Median distance from center of key (Psychomotor functioning)	0.08	**<0.001**	-9.89	**<0.001**
	Median IKD	-0.17	**<0.001**	-	-
	MAD IKD	0.05	**<0.001**	34.57	**<0.001**
	Number of Keypresses (in Log)	-	-	-4.35	**<0.001**
Random Effects and Model Fit					
	Residual		0.34		865.43
	Intercept | User		0.03		1326.38
	Intercept | Day:User		0.01		99.57
	Time of Day [1st degree] Slope | Day:User		313.29		3693375
	Time of Day [2nd degree] Slope | Day:User		209.43		3251929
	Orientation [Upright typing session] | User		0		206.67
	Motion [% of accelerometer samples in motion] | User		0		17.76
	Marginal R^2 / Conditional R^2		0.063 / 0.166		0.518 / 0.826

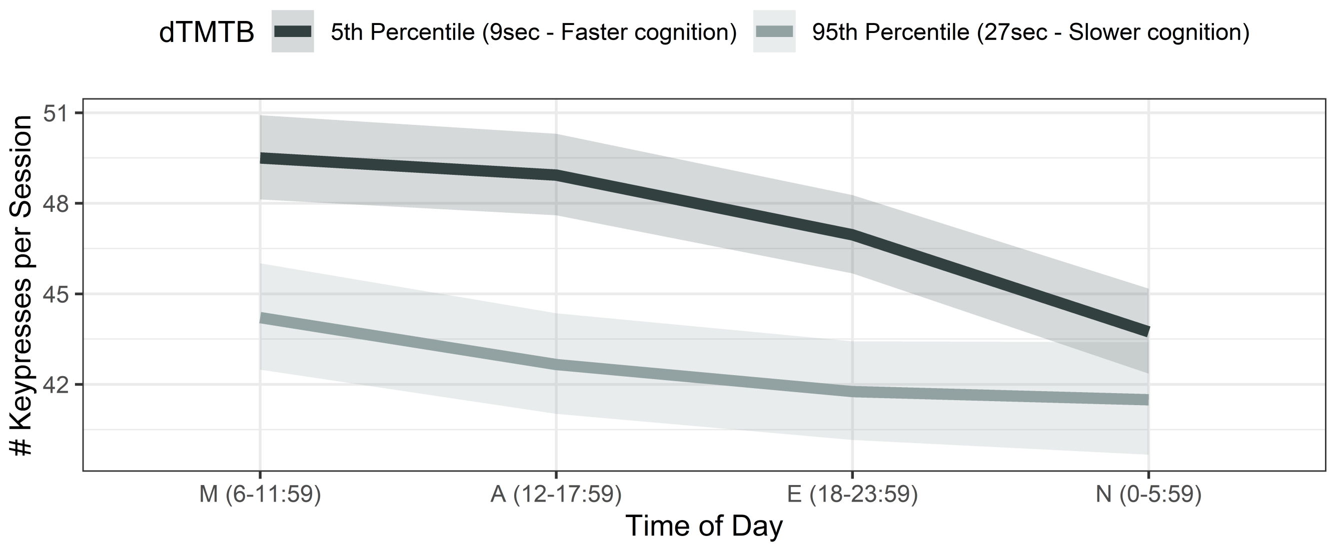

Figure 4: Predicted model results of number of keypresses per session for diurnal pattern by dTMT-B. Ribbons represent 1 SE. Note because analyses were done at the session-level, the 5th and 95th percentiles refer to the percentiles of all the sessions (N = 74,863) that had non-missing data for the variables used. M: Morning; A: Afternoon; E: Evening; N: Night. Numbers after TOD represent hour:minutes.

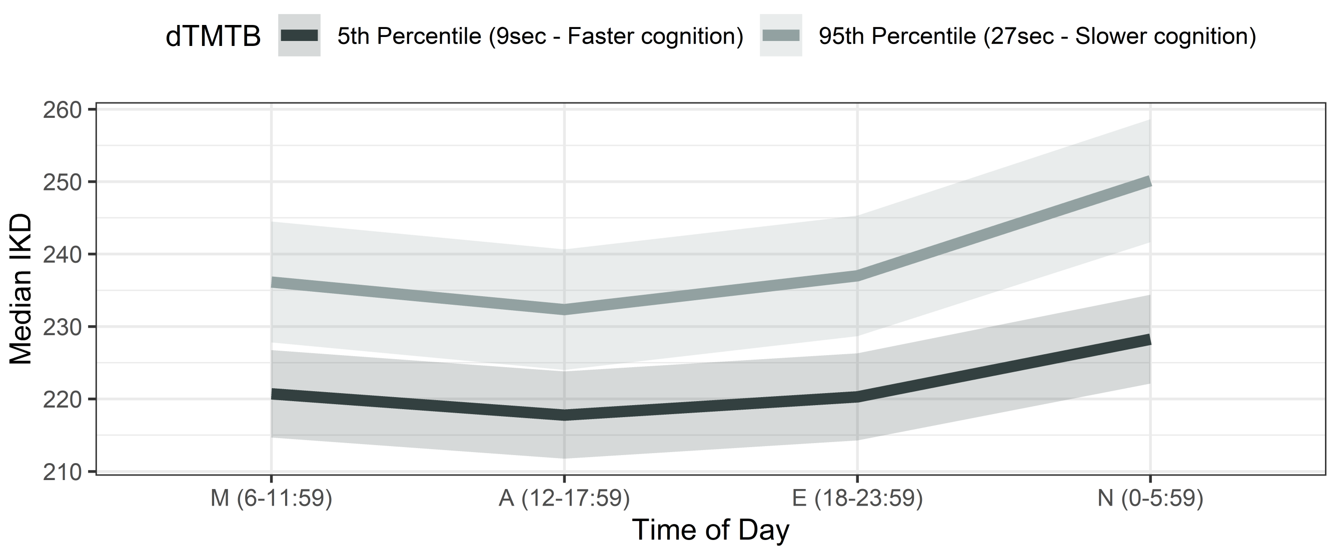

Figure 5: Predicted model results of median IKD for diurnal pattern by dTMT-B. Ribbons represent 1 SE. Note because analyses were done at the session-level, the 5th and 95th percentiles refer to the percentiles of all the sessions (N = 74,863) that had non-missing data for the variables used. M: Morning; A: Afternoon; E: Evening; N: Night. Numbers after TOD represent hour:minutes.

Next, we discuss the diurnal patterns from Table 3 for number of keypresses, which showed a significant interaction with cognitive performance, which we plot in Figure 4 in interpretable units (i.e., number of alphanumeric-to-alphanumeric keypresses). As seen in Figure 4, people in general entered fewer keypresses per session as the day progressed. The second order significant interaction with dTMT-B can also be seen in Figure 4, as shown by the difference in slopes over time of day between those with low and high cognitive performance. Those with higher cognitive performance, the 5th percentile (i.e., those taking about 9 seconds to complete dTMT-B on average) showed a more defined diurnal pattern in that they typed more in the mornings and afternoons, and decreased more over the course of the day. In short, those with higher cognitive performance showed a usage pattern consistent with a sleep-wake cycle. This stands in stark contrast to those with low cognitive performance, the 95th percentile (i.e., those taking about 27 seconds to complete the dTMT-B on average). Those participants showed an attenuated diurnal pattern less consistent with a distinctive sleep-wake cycle. The lines in Figure 4 represent the predicted number of keypresses from the first model in Table 3. We visualize the significant interaction between the quadratic effects of TOD and cognitive performance: those with faster cognition had a phone usage diurnal pattern that is concave down, whereas those with slower cognition had a concave up pattern (quadratic).

Last, we examine the control variables in Table 3 for number of keypresses and find no differences in number of keypresses for those who scored higher on QIDS and YMRS, or those on medication. We see those who pressed less closely to the center of the key within a typing session also typed more keypresses per session, about 1.05 more, $b = .08$, $e^b = 1.08$, $p < .001$. Typing sessions that were done quickly had significantly fewer keypresses, $b = -.17$, $e^b = .84$, $p < .001$, however those that were more variable in speed had significantly more, $b = .05$, $e^b = 1.05$, $p < .001$.

3.3.2 *Median IKD (Typing Speed).*

Table 3 contains the results of the median IKD per typing session in milliseconds. Overall, we found that in men, the inter-key delay was 244 ms between alphanumeric-to-alphanumeric keypresses, $b = 244.21$, $p < .001$, and women were non-significantly faster, $b = -7.22$, $p = .401$ than men when they were in the non-upright and non-active modes of typing. Further, when people were engaged in an upright typing session, they typed significantly faster by 17.55 ms, $b = -17.55$, $p < .001$. People also typed significantly faster when a higher percentage of the accelerometer samples per session were in motion, $b = -1.11$, $p = .034$.

Next, we discuss the diurnal patterns for median IKD, which showed significant linear and quadratic effects as well as significant interactions for both terms with cognitive performance, which we plotted in Figure 5. Here, we find that individuals tended to type the fastest in the afternoons and began slowing down towards the evening. Further, we can see that those slopes differed by cognitive performance. Those with higher cognitive performance, the 5th percentile in dTMT-B, showed both faster typing and less susceptibility to slowing over the course of the day (i.e., there was less slowing in typing speed in the evenings). On the other hand, those with lower cognitive performance, the 95th percentile in dTMT-B, exhibited more slowing in typing speed in the evenings and nights relative to those with higher cognitive performance. The lines in Figure 5 represent the predicted median typing speed. From Figure 5, we can visualize the significant interaction between the linear and quadratic effects of TOD and cognitive performance: overall typing speed got slower as the day progressed (linear), and although all individuals showed a concave up pattern for typing speed, those with lower cognitive performance typed even slower going from evenings to nights compared to those with higher cognitive performance (quadratic).

Next, we examine the control variables in Table 3 for median IKD and find no differences in number of keypresses for those who scored higher on QIDS and YMRS, or those on medication. Replicating a previous study [51], we do find an age effect which showed that as people age their typing speed slowed by about 20 ms for every 7 years (1 SD unit in age), $b = 19.24$, $p < .001$. Consistent with speed-accuracy trade-off, we see sessions with lower accuracy (larger distance from touch center to key center) had shorter IKDs by almost 10ms, $b = -9.89$, $p < .001$. Also, the more variable the IKDs were within a session, the slower the overall typing was for that session, $b = 34.57$, $p < .001$. Last, the more people typed per session, the faster they tended to type, $b = -4.35$, $p < .001$.

3.4 Accelerometer metrics per Typing Session

As seen in Table 4, we examine two accelerometer metrics, Motion and Orientation, using Gaussian and logistic mixed effects models, respectively. For the Motion model we predicted the percentage of accelerometer samples per session that were in motion, while for the Orientation model we predicted when people were typing in an upright vs. not-upright orientation. For each model, we entered a series of fixed effect predictors that tested our main questions of interest, a) diurnal trajectories (time of day), and b) the relationship between diurnal trajectories and cognitive performance as measured by the dTMT-B (i.e., processing speed and executive function). We also simultaneously entered the same control variables as Table 3 for typing behavior to make the models between dependent variables comparable. For the logistic mixed-effects model, we report odds ratios for interpretability and as effect sizes: > 1 OR means greater odds of an event happening (i.e., being upright) and < 1 OR means lower odds of an event happening (i.e., being upright).

3.4.1 *Motion.*

Table 4 contains the results of Motion, specifically the percentage of accelerometer samples per session that were in motion (i.e., active). Overall, we found that non-upright typing for men had around 6% of accelerometer samples per session in movement, $b = 5.70$, $p < .001$, which did not significantly differ from women $b = -0.41$, $p = .545$. Further, a session tagged as upright tended to involve more movements, $b = 5.73$, $p < .001$. There was no main effect for cognitive performance (dTMT-B) on the amount of motion per session.

Next, we discuss the diurnal patterns of motion, which showed both significant linear and quadratic interactions with cognitive performance, plotted in Figure 6 as the percentage of accelerometer samples per session that was in motion. As seen in Figure 6, we plotted the 5th and 95th percentile of cognitive performance relative to the time of day. As supported by the significant slope of time of day for both the linear and quadratic term, we find that afternoon typing sessions involved more motion during typing, but motion during typing gradually decreased as the day progressed into the evenings and nights. Further, this trajectory of the amount of motion per session differed by cognitive performance. Those with higher cognitive performance, the 5th percentile, showed both a steeper decrease in motion and a more distinctive diurnal pattern over the course of the day. On the other hand, those with lower

Table 4: Gaussian Mixed Effects Model of Motion and Logistic Mixed Effects Model of Orientation (Not upright vs. Upright).

		Motion		Orientation	
Fixed Effect Predictors		*b*	*p*	*Odds Ratios*	*p*
	(Intercept) [Mean of men when non-upright/non-active]	5.70	**<0.001**	19.49	**<0.001**
	Women [vs Men]	-0.41	0.545	1.08	0.815
Predictors of Interest					
	Motion	-	-	6.01	**<0.001**
	Orientation [Upright typing session]	5.73	**<0.001**	-	-
	dTMT-B (Cognitive Performance)	-0.15	0.628	0.99	0.951
	Time of Day [1st degree]	1.98	**<0.001**	2.52	**<0.001**
	Time of Day [2nd degree]	-1.31	**<0.001**	0.65	**<0.001**
	Time of Day [1st degree]*dTMT-B	-0.35	**<0.001**	0.92	**0.001**
	Time of Day [2nd degree]*dTMT-B	0.30	**0.021**	0.99	0.817
Control Variables					
	Medication [No medication]	0.85	0.203	1.00	0.990
	QIDS (Depression score)	0.25	0.519	0.73	0.097
	YMRS (Mania score)	-0.17	0.519	1.09	0.487
	Age	-0.66	**0.036**	1.66	**0.001**
	Median distance from center of key (Psychomotor functioning)	2.14	**<0.001**	0.78	**<0.001**
	Median IKD	0.14	0.183	0.37	**<0.001**
	MAD IKD	-0.47	**<0.001**	1.37	**<0.001**
	Number of Keypresses (in Log)	-0.66	**<0.001**	0.92	**<0.001**
Random Effects and Model Fit					
	Residual		152.44		3.29
	Intercept | User		4.99		1.56
	Intercept | Day:User		17.95		1.01
	Time of Day [1st degree] Slope | Day:User		748449.54		67878.36
	Time of Day [2nd degree] Slope | Day:User		870564.62		57383.35
	Orientation [Upright typing session] | User		16.65		-
	Motion [% of accelerometer samples in motion] | User		-		1.97
	Marginal R^2 / Conditional R^2		0.076 / 0.283		0.309 / 0.760

cognitive performance, the 95th percentile showed a more attenuated diurnal pattern over the course of the day. The lines in Figure 6 represent the predicted percentage of in-motion accelerometer samples per session. From Figure 6, we can visualize the significant interaction between the linear and quadratic effects of TOD and cognitive performance: the amount of motion decreased noticeably as the day went on for those with higher cognitive performance, but this decrease was not so large for those with lower cognitive performance (linear). In addition, for those with higher cognitive performance, the degree of the concave down pattern is more pronounced (quadratic).

Finally, we examine control variables. We find sessions to have less motion if the typing speed was more variable, b = -0.47, p < .001. Also, sessions that had more motion tended to be less accurate, b = 2.14, p < .001. In addition, the more keypresses people typed in a given session, the less motion was present during the session, b = -0.66, p < .001.

3.4.2 *Orientation.*

Table 4 contains the results of Orientation, specifically the odds of being upright while typing. There was no main effect for cognitive performance (dTMT-B) on the probability of being upright while typing.

Next, we discuss the diurnal patterns for upright typing, which showed a significant linear interaction with cognitive performance, plotted in Figure 7 as the probability of typing in an upright orientation. Here, we find that individuals were more likely to type in an upright orientation in the afternoons, but the overall probability of typing upright dropped as the day progressed. Further, the overall probability of being upright while typing differed by cognitive performance. Those with higher cognitive performance, the 5th percentile in dTMT-B, showed a more subdued decrease in upright typing as the day went on: they were less upright in the early mornings, more upright in the afternoons, and decreased in the evening, but this decrease was not as large as the 95th percentile in dTMT-B. Those with lower cognitive performance showed a slightly higher

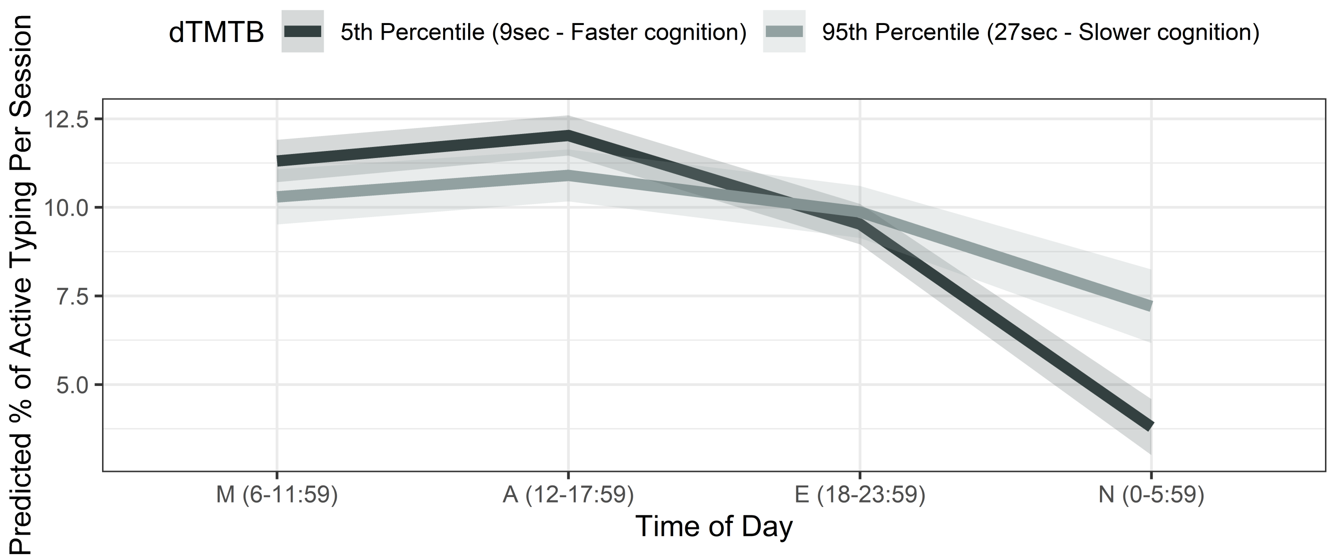

Figure 6: Predicted percentage of accelerometer samples per session that were in motion from diurnal pattern interacting with dTMT-B. Ribbons represent 1 SE. Note because analyses were done at the session-level, the 5th and 95th percentiles refer to the percentiles of all the sessions (N = 74,863) that had non-missing data for the variables used. M: Morning; A: Afternoon; E: Evening; N: Night. Numbers after TOD represent hour:minutes.

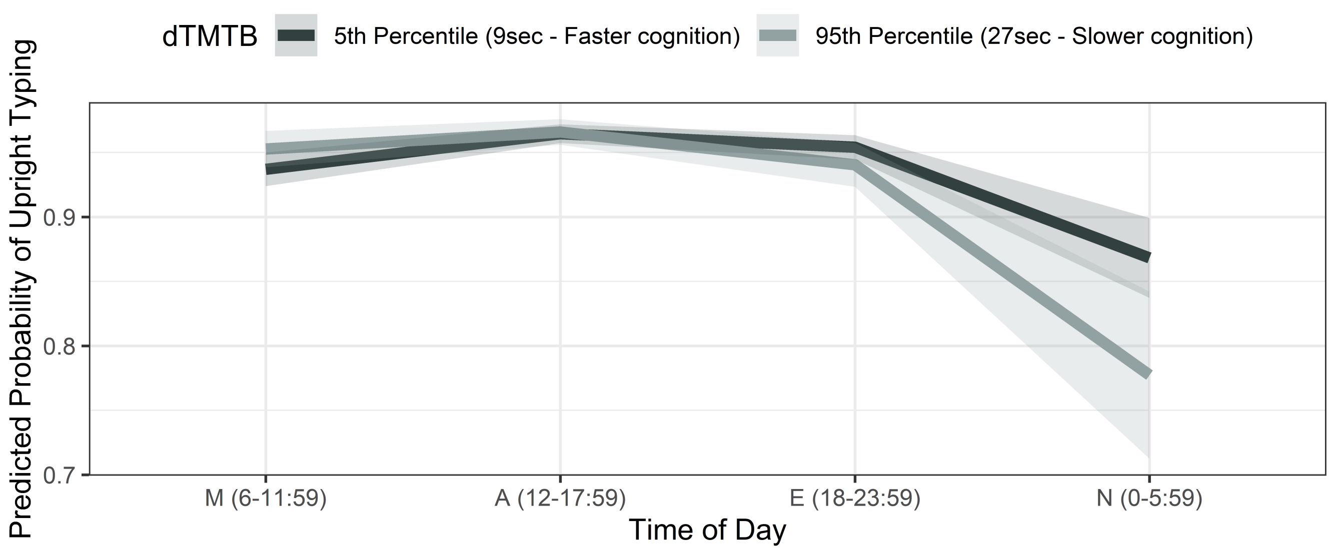

Figure 7: Predicted probability of Upright Typing from diurnal pattern interacting with dTMT-B. Ribbons represent 1 SE. Note because analyses were done at the session-level, the 5th and 95th percentiles refer to the percentiles of all the sessions (N = 74,863) that had non-missing data for the variables used. M: Morning; A: Afternoon; E: Evening; N: Night. Numbers after TOD represent hour:minutes.

probability to type upright in the early mornings, but we refrain from interpreting cognitive performance influencing upright typing at night because of the overlap in their respective error bars. The lines in Figure 7 represent the predicted probability of upright typing. From Figure 7, we can visualize the significant interaction between the linear effects of TOD and cognitive performance: those with lower cognitive performance decreased more in their probability of upright typing than those with higher cognitive performance as the day progressed (linear).

Finally, we examine control variables. We find that slower typing and lower accuracy were associated with lower odds in being upright while typing, OR = 0.37, p < .001 and OR = 0.78, p < .001, respectively, while more variable typing speed was associated with higher odds of being upright while typing, OR = 1.37, p < .001. In addition, the more keypresses people typed in a given session, the less likely they were to type in an upright orientation, OR = 0.92, p < .001. Last, we also find an age effect showing that older individuals were more likely to be typing in an upright orientation, OR = 1.66, p = .001.

4 DISCUSSION

Here we employed a sensor fusion approach, jointly modeling keyboard dynamics and accelerometer data obtained during typing, to investigate the inter-relationships between activity patterns related to phone usage, cognitive performance, and time of day. Unlike prior sensing studies that mostly utilized machine learning techniques to predict clinical diagnoses and extract relevant features [13, 20, 34, 48, 54], the current study modeled and statistically tested differences in granular temporal changes in cognition while simultaneously estimating effect sizes that can be used to determine clinical relevance in future studies. Consistent with literature demonstrating cognitive impairment in individuals with mood disorders, we found overall significantly lower performance in processing speed and executive functioning (as assessed using the Trail-Making Test, Part B) in the mood disorder group, even when their overall symptom severity of depression and mania was mild. This is consistent with prior findings that individuals with mood disorders may experience changes in cognition even in euthymic or non-acute states [27]. In addition, we expand on previous findings from EMA (such as the Psychomotor Vigilance Task) using passive and unobtrusive dependent measures and found that diurnal patterns interacted with differences in cognition. Specifically, inter-key delay, a proxy for processing speed, a core component of cognition [40], was not only shorter (i.e., typing faster) but also less sensitive to the impact of time of day (i.e., less slowing at night) in individuals with higher cognitive performance (i.e., those who performed better on the dTMT-B) [33]. This result is not surprising and likely related to the buildup of homeostatic sleep pressure (i.e., the increased need to sleep the longer we are awake) and converged with the established Psychomotor Vigilance Test, an active test measuring reaction times to cue presentations known to be sensitive to diurnal fluctuations [26]. Second, we found that participants with relatively higher performance exhibited more well-defined diurnal patterns as indexed by smartphone keyboard usage. Just like individuals with lower levels of depression who exhibited less attenuated positive mood intraday fluctuations [31], individuals with higher cognitive performance in the current study engaged with the keyboard more during the day and tapered their usage more at night, potentially indicative of self-regulatory behaviors and/or sleep hygiene habits related to phone usage in the evening.

Our current study has several implications for the use of real time passive and unobtrusive sensor data in measuring brain health. When employed at scale, smartphone technologies that incorporate these keyboard and accelerometer components can provide clinicians with additional insight into neurocognitive functioning that is more temporally granular and ecologically valid than standard clinical assessments. Further, moving beyond the "average patient" one-size-fits-all approach, recent work has shown that having granular and ecological measurements from smartphones predicted prospective clinician assessments individualistically [6, 46]. Therefore, our findings support the feasibility of precision-medicine-based approach for behavioral health using a digital platform. Precision medicine encompasses a wide range of innovative approaches that consider individual differences in genes, environments, and lifestyles. To this end, our data demonstrated how one can leverage rich data on a modern smartphone to probe brain function and behavior specific to each unique individual at a granular level.

Our current study has several limitations. As our mood disorder group had mild symptom severity over the course of the study (YMRS scores <= 8 count towards euthymia [8]), we were thus not able to longitudinally examine how mood fluctuations within a person differentially impacted keyboard dynamics and accelerometer data. Second, the current study used TOD instead of time since wake up to examine diurnal patterns, referencing everyone against the clock time (time zone corrected). Future studies can use time since wake up, such as integrating data from smart watches, to compare individuals against their own internal biological clock and investigate whether keystroke dynamics deteriorate differentially between subjects. Third, the current study averaged multiple cognitive scores and clinician assessments across different time points. Instead of taking the average, future work could better align the cognitive scores and clinician assessments temporally with corresponding keyboard sessions to characterize the relationship more precisely between activity patterns, phone usage, cognitive performance, and time of day.

5 CONCLUSION

We set out to answer if it is possible to leverage modern smartphone technology to passively and unobtrusively measure brain health in people's natural environments. To this end, we used data streams derived from a customized smartphone virtual keyboard and accelerometer data to examine how overall phone usage and keyboard dynamics exhibit diurnal patterns and how they are impacted by movement and orientation as well as baseline cognitive performance in a well-characterized sample of individuals across a mood disorder spectrum. We uncovered subtle yet clinically relevant differential diurnal patterns in individuals based on their cognitive performance, suggesting that these data streams can be employed to longitudinally assess and monitor changes in a person's neuropsychological functioning over not only days, but also weeks or even years. In sum, knowledge derived from our current line of research may be valuable in future studies of a wide range of neurological and psychiatric disorders thanks to the affordability, scalability, and ubiquity of modern smart and connected technologies.

REFERENCES

[1] Saeed Abdullah, Elizabeth L. Murnane, Mark Matthews, Matthew Kay, Julie A. Kientz, Geri Gay, and Tanzeem Choudhury. 2016. Cognitive rhythms: Unobtrusive and continuous sensing of alertness using a mobile phone. *UbiComp 2016* -

Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (2016), 178–189. https://doi.org/10.1145/2971648.2971712

[2] Hessa Alfalahi, Ahsan H. Khandoker, Nayeefa Chowdhury, Dimitrios Iakovakis, Sofia B. Dias, K. Ray Chaudhuri, and Leontios J. Hadjileontiadis. 2022. Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic review and meta-analysis. *Scientific Reports 2022 12:1* 12, 1 (may 2022), 1–24. https://doi.org/10.1038/s41598-022-11865-7

[3] Sangwon Bae, Tammy Chung, Denzil Ferreira, Anind K. Dey, and Brian Suffoletto. 2018. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive Behaviors* 83 (aug 2018), 42–47. https://doi.org/10.1016/J.ADDBEH.2017.11.039

[4] Luca Bedogni, Marco Di Felice, and Luciano Bononi. 2012. By train or by car? Detecting the user's motion type through smartphone sensors data. *IFIP Wireless Days* (2012). https://doi.org/10.1109/WD.2012.6402818

[5] Dror Ben-Zeev, Michael A. Young, and Joshua W. Madsen. 2009. Retrospective recall of affect in clinically depressed individuals and controls. *Cognition and Emotion* 23, 5 (2009), 1021–1040. https://doi.org/10.1080/02699930802607937

[6] Casey C. Bennett, Mindy K. Ross, Eu Gene Baek, Dohyeon Kim, and Alex D. Leow. 2022. Predicting clinically relevant changes in bipolar disorder outside the clinic walls based on pervasive technology interactions via smartphone typing dynamics. *Pervasive and Mobile Computing* 83 (jul 2022). https://doi.org/10.1016/J.PMCJ.2022.101598

[7] Robert Bixler and Sidney D'Mello. 2013. Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. *International Conference on Intelligent User Interfaces, Proceedings IUI* (mar 2013), 225–233. https://doi.org/10.1145/2449396.2449426

[8] C. Bourne, O. Aydemir, V. Balanzá-Martínez, E. Bora, S. Brissos, J. T.O. Cavanagh, L. Clark, Z. Cubukcuoglu, V. V. Dias, S. Dittmann, I. N. Ferrier, D. E. Fleck, S. Frangou, P. Gallagher, L. Jones, T. Kieseppä, A. Martínez-Aran, I. Melle, P. B. Moore, M. Mur, A. Pfennig, A. Raust, V. Senturk, C. Simonsen, D. J. Smith, D. S. Bio, M. G. Soeiro-de Souza, S. D.R. Stoddart, K. Sundet, A. Szöke, J. M. Thompson, C. Torrent, T. Zalla, N. Craddock, O. A. Andreassen, M. Leboyer, E. Vieta, M. Bauer, P. D. Worhunsky, C. Tzagarakis, R. D. Rogers, J. R. Geddes, and G. M. Goodwin. 2013. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: An individual patient data meta-analysis. *Acta Psychiatrica Scandinavica* 128, 3 (sep 2013), 149–162. https://doi.org/10.1111/acps.12133

[9] Mollie E. Brooks, Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Maechler, and Benjamin M. Bolker. 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* 9, 2 (2017), 378–400. https://doi.org/10.32614/RJ-2017-066

[10] Daniel Buschek, Alexander De Luca, and Florian Alt. 2015. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. *Conference on Human Factors in Computing Systems - Proceedings* 2015-April (apr 2015), 1393–1402. https://doi.org/10.1145/2702123.2702252

[11] G. B. Cassano, A. Benvenuti, M. Miniati, S. Calugi, M. Mula, L. Maggi, P. Rucci, A. Fagiolini, F. Perris, and E. Frank. 2009. The factor structure of lifetime depressive spectrum in patients with unipolar depression. *Journal of Affective Disorders* 115, 1-2 (may 2009), 87–99. https://doi.org/10.1016/J.JAD.2008.09.006

[12] Michelle H Chen, Alex Leow, Mindy K Ross, John Deluca, Nancy Chiaravalloti, Silvana L Costa, Helen M Genova, Erica Weber, Faraz Hussain, and Alexander P Demos. 2022. Associations between smartphone keystroke dynamics and cognition in MS. *Digital Health* 8 (2022), 1–12. https://doi.org/10.1177/20552076221143234

[13] Richard Chen, Luca Foschini, Lampros Kourtis, Alessio Signorini, Filip Jankovic, Melissa Pugh, Jie Shen, Roy Yaari, Vera Maljkovic, Marc Sunga, Nikki Marinsek, Han Hee Song, Hyun Joon Jung, Belle Tseng, and Andrew Trister. 2019. Developing measures of cognitive impairment in the real world from consumer-grade multimodal sensor streams. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (jul 2019), 2145–2155. https://doi.org/10.1145/3292500.3330690

[14] Vivek Dhakal, Anna Maria Feit, Per Ola Kristensson, and Antti Oulasvirta. 2018. Observations on typing from 136 million keystrokes. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (apr 2018). https://doi.org/10.1145/3173574.3174220

[15] Tilman Dingler, Albrecht Schmidt, and Tonja Machulla. 2017. Building Cognition-Aware Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (sep 2017), 1–15. https://doi.org/10.1145/3132025

[16] Rachel Eardley, Anne Roudaut, Steve Gill, and Stephen J. Thompson. 2018. Investigating how smartphone movement is affected by body posture. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (apr 2018). https://doi.org/10.1145/3173574.3173776

[17] Clayton Epp, Michael Lippold, and Regan L. Mandryk. 2011. Identifying emotional states using keystroke dynamics. *Conference on Human Factors in Computing Systems - Proceedings* (2011), 715–724. https://doi.org/10.1145/1978942.1979046

[18] Mayank Goel, Alex Jansen, Travis Mandel, Shwetak N. Patel, and Jacob O. Wobbrock. 2013. ContextType: Using hand posture information to improve mobile touch screen text entry. *Conference on Human Factors in Computing Systems - Proceedings* (2013), 2795–2798. https://doi.org/10.1145/2470654.2481386

[19] Safaa Hriez, Nadim Obeid, and Arafat Awajan. 2019. User authentication on smartphones using keystroke dynamics. *ACM International Conference Proceeding Series* (dec 2019). https://doi.org/10.1145/3368691.3368725

[20] Dimitrios Iakovakis, K. Ray Chaudhuri, Lisa Klingelhoefer, Sevasti Bostantjopoulou, Zoe Katsarou, Dhaval Trivedi, Heinz Reichmann, Stelios Hadjidimitriou, Vasileios Charisis, and Leontios J. Hadjileontiadis. 2020. Screening of Parkinsonian subtle fine-motor impairment from touchscreen typing via deep learning. *Scientific Reports 2020 10:1* 10, 1 (jul 2020), 1–13. https://doi.org/10.1038/s41598-020-69369-1

[21] Dimitrios Iakovakis, Stelios Hadjidimitriou, Vasileios Charisis, Sevasti Bostantjopoulou, Zoe Katsarou, Lisa Klingelhoefer, Heinz Reichmann, Sofia B. Dias, José A. Diniz, Dhaval Trivedi, K. Ray Chaudhuri, and Leontios J. Hadjileontiadis. 2018. Motor impairment estimates via touchscreen typing dynamics toward Parkinson's Disease detection from data harvested in-the-wild. *Frontiers in ICT* 5 (nov 2018), 28. https://doi.org/10.3389/FICT.2018.00028/BIBTEX

[22] Xinhui Jiang, Jussi P.P. Jokinen, Antti Oulasvirta, and Xiangshi Ren. 2022. Learning to type with mobile keyboards: Findings with a randomized keyboard. *Computers in Human Behavior* 126 (jan 2022). https://doi.org/10.1016/J.CHB.2021.106992

[23] Xinhui Jiang, Yang Li, Jussi P.P. Jokinen, Viet Ba Hirvola, Antti Oulasvirta, and Xiangshi Ren. 2020. How We Type: Eye and Finger Movement Strategies in Mobile Typing. *Conference on Human Factors in Computing Systems - Proceedings* (apr 2020). https://doi.org/10.1145/3313831.3376711

[24] Jennifer C. Kanady and Allison G. Harvey. 2015. Development and Validation of the Sleep Inertia Questionnaire (SIQ) and Assessment of Sleep Inertia in Analogue and Clinical Depression. *Cognitive therapy and research* 39, 5 (oct 2015), 601. https://doi.org/10.1007/S10608-015-9686-4

[25] Rinat Khusainov, Djamel Azzi, Ifeyinwa E. Achumba, and Sebastian D. Bersch. 2013. Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations. https://doi.org/10.3390/s131012852

[26] William D.S. Killgore. 2010. Effects of sleep deprivation on cognition. *Progress in Brain Research* 185, C (jan 2010), 105–129. https://doi.org/10.1016/B978-0-444-53702-7.00007-5

[27] Matthew M Kurtz and Raphael T Gerraty. 2009. A Meta-Analytic Investigation of Neurocognitive Deficits in Bipolar Illness: Profile and Effects of Clinical State. *Neuropsychology* 23, 5 (2009), 551–562. https://doi.org/10.1037/a0016277

[28] Alex Mariakakis, Sayna Parsi, Shwetak N. Patel, and Jacob O. Wobbrock. 2018. Drunk user interfaces: Determining blood alcohol level through everyday smartphone tasks. *Conference on Human Factors in Computing Systems - Proceedings* 2018-April (2018). https://doi.org/10.1145/3173574.3173808

[29] Rafail Evangelos Mastoras, Dimitrios Iakovakis, Stelios Hadjidimitriou, Vasileios Charisis, Seada Kassie, Taoufik Alsaadi, Ahsan Khandoker, and Leontios J. Hadjileontiadis. 2019. Touchscreen typing pattern analysis for remote detection of the depressive tendency. *Scientific Reports 2019 9:1* 9, 1 (sep 2019), 1–12. https://doi.org/10.1038/s41598-019-50002-9

[30] Jun Ki Min, Afsaneh Doryab, Jason Wiese, Shahriyar Amini, John Zimmerman, and Jason I. Hong. 2014. Toss 'N' turn: Smartphone as sleep and sleep quality detector. *Conference on Human Factors in Computing Systems - Proceedings* (2014), 477–486. https://doi.org/10.1145/2556288.2557220

[31] Greg Murray. 2007. Diurnal mood variation in depression: A signal of disturbed circadian function? *Journal of Affective Disorders* 102, 1-3 (sep 2007), 47–53. https://doi.org/10.1016/J.JAD.2006.12.001

[32] Jukka Pekka Onnela and Scott L. Rauch. 2016. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. https://doi.org/10.1038/npp.2016.7

[33] Venet Osmani, Alban Maxhuni, Agnes Grünerbl, Paul Lukowicz, Christian Haring, and Oscar Mayora. 2013. Monitoring activity of patients with bipolar disorder using smart phones. In *ACM International Conference Proceeding Series*. 85–92. https://doi.org/10.1145/2536853.2536882

[34] Alexandros Papadopoulos, Dimitrios Iakovakis, Lisa Klingelhoefer, Sevasti Bostantjopoulou, K. Ray Chaudhuri, Konstantinos Kyritsis, Stelios Hadjidimitriou, Vasileios Charisis, Leontios J. Hadjileontiadis, and Anastasios Delopoulos. 2020. Unobtrusive detection of Parkinson's disease from multi-modal and in-the-wild sensor data using deep learning techniques. *Scientific Reports 2020 10:1* 10, 1 (dec 2020), 1–13. https://doi.org/10.1038/s41598-020-78418-8

[35] Homa Rashidisabet, Olusola Ajilore, Alex Leow, and Alexander P. Demos. 2022. Revisiting power-law estimation with applications to real-world human typing dynamics. *Physica A: Statistical Mechanics and its Applications* 599 (aug 2022). https://doi.org/10.1016/j.physa.2022.127384

[36] Nishkam Ravi, Nikhil Dandekar, Preetham Mysore, and Michael L Littman. 2005. Activity recognition from accelerometer data. In *Proceedings of the National Conference on Artificial Intelligence*, Vol. 3. 1541–1546.

[37] Barbara Ravnkilde, Poul Videbech, Karin Clemmensen, Annette Egander, Niels Anton Rasmussen, and Raben Rosenberg. 2002. Cognitive deficits in major depression. *Scandinavian Journal of Psychology* 43, 3 (jul 2002), 239–251.

https://doi.org/10.1111/1467-9450.00292

[38] Philipp S. Ritter, Carolin Marx, Natalia Lewtschenko, Steffi Pfeiffer, Karolina Leopold, Michael Bauer, and Andrea Pfennig. 2012. The characteristics of sleep in patients with manifest bipolar disorder, subjects at high risk of developing the disease and healthy controls. *Journal of Neural Transmission* 119, 10 (oct 2012), 1173–1184. https://doi.org/10.1007/S00702-012-0883-Y/TABLES/7

[39] Lucy J. Robinson, Jill M. Thompson, Peter Gallagher, Utpal Goswami, Allan H. Young, I. Nicol Ferrier, and P. Brian Moore. 2006. A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. *Journal of Affective Disorders* 93, 1-3 (jul 2006), 105–115. https://doi.org/10.1016/j.jad.2006.02.016

[40] Mindy K. Ross, Alexander P. Demos, John Zulueta, Andrea Piscitello, Scott A. Langenecker, Melvin McInnis, Olusola Ajilore, Peter C. Nelson, Kelly A. Ryan, and Alex Leow. 2021. Naturalistic smartphone keyboard typing reflects processing speed and executive function. *Brain and Behavior* 11, 11 (nov 2021). https://doi.org/10.1002/brb3.2363

[41] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Kai Kunze, and Albrecht Schmidt. 2013. Upright or sideways? Analysis of smartphone postures in the wild. In *MobileHCI 2013 - Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 362–371. https://doi.org/10.1145/2493190.2493230

[42] Christina Schmidt, Fabienne Collette, Christian Cajochen, and Philippe Peigneux. 2007. A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology* 24, 7 (oct 2007), 755–789. https://doi.org/10.1080/02643290701754158

[43] Jussi Seppälä, Ilaria De Vita, Timo Jämsä, Jouko Miettunen, Matti Isohanni, Katya Rubinstein, Yoram Feldman, Eva Grasa, Iluminada Corripio, Jesus Berdun, Enrico D'Amico, and Maria Bulgheroni. 2019. Mobile phone and wearable sensor-based mhealth approaches for psychiatric disorders and symptoms: Systematic review. https://doi.org/10.2196/mental.9819

[44] Saul Shiffman, Arthur A. Stone, and Michael R. Hufford. 2008. Ecological momentary assessment. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

[45] Valeriu Daniel Stanciu, Riccardo Spolaor, Mauro Conti, and Cristiano Giuffrida. 2016. On the effectiveness of sensor-enhanced keystroke dynamics against statistical attacks. *CODASPY 2016 - Proceedings of the 6th ACM Conference on Data and Application Security and Privacy* (mar 2016), 105–112. https://doi.org/10.1145/2857705.2857748

[46] Jonathan P. Stange, John Zulueta, Scott A. Langenecker, Kelly A. Ryan, Andrea Piscitello, Jenna Duffecy, Melvin G. McInnis, Pete Nelson, Olusola Ajilore, and Alex Leow. 2018. Let your fingers do the talking: Passive typing instability predicts future mood outcomes. *Bipolar Disorders* 20, 3 (may 2018), 285–288. https://doi.org/10.1111/BDI.12637

[47] Ming Sun and J. O. Hill. 1993. A method for measuring mechanical work and work efficiency during human activities. *Journal of Biomechanics* 26, 3 (1993), 229–241. https://doi.org/10.1016/0021-9290(93)90361-H

[48] Shikha Tripathi, Teresa Arroyo-Gallego, and Luca Giancardo. 2022. Keystroke-Dynamics for Parkinson's Disease Signs Detection in An At-Home Uncontrolled Population: A New Benchmark and Method. *IEEE Transactions on Biomedical Engineering* (2022). https://doi.org/10.1109/TBME.2022.3187309

[49] Athanasios Tsanas, Elizabeth Woodward, and Anke Ehlers. 2020. Objective Characterization of Activity, Sleep, and Circadian Rhythm Patterns Using a Wrist-Worn Actigraphy Sensor: Insights Into Posttraumatic Stress Disorder. *JMIR mHealth and uHealth* 8, 4 (apr 2020). https://doi.org/10.2196/14306

[50] E. J.W. Van Someren. 1997. Actigraphic monitoring of movement and rest-activity rhythms in aging, Alzheimer's disease, and Parkinson's disease. , 394–398 pages.

[51] Claudia Vesel, Homa Rashidisabet, John Zulueta, Jonathan P. Stange, Jennifer Duffecy, Faraz Hussain, Andrea Piscitello, John Bark, Scott A. Langenecker, Shannon Young, Erin Mounts, Larsson Omberg, Peter C. Nelson, Raeanne C. Moore, Dave Koziol, Keith Bourne, Casey C. Bennett, Olusola Ajilore, Alexander P. Demos, and Alex Leow. 2020. Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: A BiAffect iOS study. *Journal of the American Medical Informatics Association* 27, 7 (jul 2020), 1007–1018. https://doi.org/10.1093/jamia/ocaa057

[52] Lakshmi N. Yatham, Ivan J. Torres, Gin S. Malhi, Sophia Frangou, David C. Glahn, Carrie E. Bearden, Katherine E. Burdick, Anabel Martínez-Arán, Sandra Dittmann, Joseph F. Goldberg, Aysegul Ozerdem, Omer Aydemir, and K. N.Roy Chengappa. 2010. The International Society for Bipolar Disorders-Battery for Assessment of Neurocognition (ISBD-BANC). *Bipolar Disorders* 12, 4 (jun 2010), 351–363. https://doi.org/10.1111/j.1399-5618.2010.00830.x

[53] Nan Zheng, Kun Bai, Hai Huang, and Haining Wang. 2014. You are how you touch: User verification on smartphones via tapping behaviors. *Proceedings - International Conference on Network Protocols, ICNP* (dec 2014), 221–232. https://doi.org/10.1109/ICNP.2014.43

[54] John Zulueta, Alexander Pantelis Demos, Claudia Vesel, Mindy Ross, Andrea Piscitello, Faraz Hussain, Scott A. Langenecker, Melvin McInnis, Peter Nelson, Kelly Ryan, Alex Leow, and Olusola Ajilore. 2021. The Effects of Bipolar Disorder Risk on a Mobile Phone Keystroke Dynamics Based Biomarker of Brain Age. *Frontiers in Psychiatry* 12 (dec 2021). https://doi.org/10.3389/fpsyt.2021.739022

[55] John Zulueta, Andrea Piscitello, Mladen Rasic, Rebecca Easter, Pallavi Babu, Scott A. Langenecker, Melvin McInnis, Olusola Ajilore, Peter C. Nelson, Kelly Ryan, and Alex Leow. 2018. Predicting mood disturbance severity with mobile phone keystroke metadata: A biaffect digital phenotyping study. *Journal of Medical Internet Research* 20, 7 (jul 2018). https://doi.org/10.2196/jmir.9775