Chapter 13
Passive Sensing of Affective and Cognitive Functioning in Mood Disorders by Analyzing Keystroke Kinematics and Speech Dynamics

Faraz Hussain, Jonathan P. Stange, Scott A. Langenecker, Melvin G. McInnis, John Zulueta, Andrea Piscitello, Mindy K. Ross, Alexander P. Demos, Claudia Vesel, Homa Rashidisabet, Bokai Cao, He Huang, Philip S. Yu, Peter Nelson, Olusola A. Ajilore, and Alex Leow

> Life is too sweet and too short to express our affection with just our thumbs. Touch is meant for more than a keyboard.
> — Kristin Armstrong, Olympic cyclist

F. Hussain · J. Zulueta · M. K. Ross · C. Vesel · H. Rashidisabet · O. A. Ajilore · A. Leow
Collaborative Neuroimaging Environment for Connectomics, University of Illinois, Chicago, USA
e-mail: farazh@uic.edu

J. Zulueta
e-mail: oajilore@uic.edu

M. K. Ross
e-mail: mross27@uic.edu

C. Vesel
e-mail: cvesel2@uic.edu

H. Rashidisabet
e-mail: hrashi4@uic.edu

O. A. Ajilore
e-mail: oajilore@uic.edu

A. Leow
e-mail: weihliao@uic.edu

J. P. Stange (✉)
Cognition and Affect Regulation Lab, University of Illinois, Chicago, USA
e-mail: jstange@uic.edu

S. A. Langenecker
University Neuropsychiatric Institute, University of Utah, Salt Lake City, UT, USA
e-mail: s.langenecker@hsc.utah.edu

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

C. Montag and H. Baumeister (eds.), *Digital Phenotyping and Mobile Sensing*,
Studies in Neuroscience, Psychology and Behavioral Economics,
https://doi.org/10.1007/978-3-030-98546-2_13

Abstract Mood disorders can be difficult to diagnose, evaluate, and treat. They involve affective and cognitive components, both of which need to be closely monitored over the course of the illness. Current methods like interviews and rating scales can be cumbersome, sometimes ineffective, and oftentimes infrequently administered. Even ecological momentary assessments, when used alone, are susceptible to many of the same limitations and still require active participation from the subject. Passive, continuous, frictionless, and ubiquitous means of recording and analyzing mood and cognition obviate the need for more frequent and lengthier doctor's visits, can help identify misdiagnoses, and would potentially serve as an early warning system to better manage medication adherence and prevent hospitalizations. Activity trackers and smartwatches have long provided exactly such a tool for evaluating physical fitness. What if smartphones, voice assistants, and eventually Internet of Things devices and ambient computing systems could similarly serve as fitness trackers for the brain, without imposing any additional burden on the user? In this chapter, we explore two such early approaches—an in-depth analytical technique based on examining meta-features of virtual keyboard usage and corresponding typing kinematics, and another method which analyzes the acoustic features of recorded speech—to passively and unobtrusively understand mood and cognition in people with bipolar disorder. We review innovative studies that have used these methods to build mathematical models and machine learning frameworks that can provide deep insights into users' mood and cognitive states. We then outline future research considerations and conclude with discussing the opportunities and challenges afforded by these modes of researching mood disorders and passive sensing approaches in general.

M. G. McInnis
Heinz C. Prechter Bipolar Research Program, University of Michigan, Ann Arbor, MI, USA
e-mail: mmcinnis@med.umich.edu

A. Piscitello
Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
e-mail: andrea1.piscitello@mail.polimi.it

B. Cao
Video Understanding Team, Applied Machine Learning Facebook, Menlo Park, CA, USA
e-mail: caobokai@fb.com

H. Huang · P. S. Yu
Department of Computer Science, University of Illinois, Chicago, USA
e-mail: hehuang@uic.edu

P. S. Yu
e-mail: psyu@cs.uic.edu

P. Nelson
College of Engineering, University of Illinois, Chicago, USA
e-mail: nelson@uic.edu

A. P. Demos
Department of Psychology, University of Illinois, Chicago, USA
e-mail: ademos@uic.edu

13.1 Introduction

Mood disorders take a sizable toll on the world's population, affecting more than 1 in 20 people annually and nearly 1 out of every 10 people over the course of their lifetime (Steel et al. 2014). Bipolar disorder, which alone accounts for at least 1% of years lived with disability globally (GBD 2017), is a mood disorder that causes patients to alternate between manic episodes of abnormally elevated mood and energy levels, and depressive episodes marked by diminished mood, interest, and energy (APA 2013). Compared to major depressive disorder (MDD), bipolar disorder can be harder to diagnose, and even when an accurate diagnosis is made, it is often delayed. The depressive episodes in both disorders share the same diagnostic criteria, and it is known that individuals suffering from bipolar disorder on average spend more time in the depressive phase than in mania. In particular, bipolar disorder type II, a subtype which is differentiated by attenuated levels of mania-like symptoms (termed hypomania) is difficult to diagnose by non-specialists as it can be challenging to distinguish from recurring unipolar depression. The presence of mood episodes with mixed features, i.e., those that exhibit characteristics of both mania and depression, can further complicate the process of diagnosis (Phillips and Kupfer 2013).

13.1.1 Current State of Diagnosis and Monitoring of Bipolar Disorder

Clinical approaches to diagnosing and monitoring bipolar disorder usually start with careful history-taking by the clinician (detailed interviews with patients and their family members as well as probing for a family history of the disorder), followed by the frequent use of self- and clinician-administered rating scales that assess for a history of possible mania or hypomania in patients with depression. Even with these tools at their disposal, it is often difficult for clinicians to ascertain whether any noted changes in mood, sleep, or energy are within normal ranges—or whether they are evidence of, say, a manic/hypomanic episode (Wolkenstein et al. 2011). Achieving inter-rater reliability between administered assessments and scales poses its own challenges.

After a correct diagnosis has been made, monitoring of symptoms commonly relies upon self-reports that may include mood charting and self-ratings or clinician-rated scales. These scales can only assess the severity of symptoms experienced by the patients and cannot actually screen for mania or hypomania; patients in manic states also may not be cognizant of their manic symptoms, casting doubt on the validity of some of these assessments (NCCMH 2018).

Ecological momentary assessments (EMA) have been used for supplementary monitoring in mood disorders with varying degrees of success (Ebner-Priemer and Trull 2009; Asselbergs et al. 2016; Kubiak and Smyth 2019). Asselbergs and colleagues reported that the clinical utility of self-report EMA is too often limited

by the heavy response burden that is imposed upon respondents—which can result in large dropout rates after an initial period of activity—and furthermore, that the predictive models constructed using unobtrusive EMA data were inferior to existing benchmark models.

In recent years, other techniques including neuroimaging (Phillips et al. 2008; Leow et al. 2013; Ajilore et al. 2015; Andreassen et al. 2018) and genomics (Hou et al. 2016; Ikeda et al. 2017) have also been used in attempts to discover biomarkers for bipolar disorder. Although they may not currently be feasible either for diagnosis or for monitoring on an individual level, in the near future we may begin finding immense value in these and related methods beyond their immediate research applications.

In addition to its affective components, bipolar disorder also influences cognitive ability (APA 2013). Among the most severely impaired domains of cognition are attention, working memory, and response inhibition (Bourne et al. 2013). These provide another avenue to further aid in distinguishing a possible diagnosis of bipolar disorder from other mood disorders and assessing its course and treatment.

13.1.2 Passive Sensing in Physical Health

Smartwatches, fitness trackers, and associated physical health and fitness apps in general have to a large extent enabled and encouraged users to self-manage chronic medical conditions and attempt to take better care of their physical health (Anderson et al. 2016; Canhoto and Arp 2017; Messner et al. 2019). The Apple Watch, for instance—which uses photoplethysmography to passively sense atrial fibrillation—and the associated Apple Heart Study (Turakhia 2018) have already been credited with saving several lives by alerting enrolled users to the onset of life-threatening conditions and directing them to seek immediate medical attention (Feng 2018; Perlow 2018).

13.1.3 What About Passive Sensing for Mental Health?

Portable sensors to track the health of the rest of the body have so far proven easier to develop than those that can track brain health. As yet, there are no portable functional magnetic resonance imaging (fMRI) scanners or brain-computer interfaces (BCI) that can be used to unobtrusively analyze brain functioning—although science fiction has proposed examples of each in the form of, respectively, cowboy hats that conduct brain scans to map wearers' cognition in television shows such as Westworld (Avunjian 2018) and biomechanical computer implants called neural lace in author Iain M. Banks' series *The Culture* (Banks 2002, 2010)—which science may in fact someday deliver instead in the shape of the startup Openwater's fMRI-replacing ski hats that are purportedly being designed to use infrared holography to scan oxygen utilization

by the wearer's brain (Jepsen 2017; Clifford 2017) and implantable electronic circuits capable of neural communication such as those being developed by Neuralink and others (Fu et al. 2016; Chung et al. 2018; Sanford 2018).

Until these nascent technologies reach maturity, there is a need for passive sensing tools that can bridge the divide and perhaps eliminate the need for more onerous means of sensing altogether. Smartphones are already ubiquitous enough and offer a wide array of sensors, which when used in concert with mHealth and digital phenotyping tools, offer a greater degree of precision medicine tools to users, researchers, and healthcare providers than ever before. Indeed, the very use of smartphones, and mobile social networking apps in particular, has been found to be associated with structural and functional changes in the brain (Montag et al. 2017); the corollary that smartphone usage patterns can be used to quantify the presence of established biomarkers has also been explored by Sariyska and colleagues (2018) in their preliminary study examining the feasibility of probing molecular genetic variables corresponding to individual differences in personality and linked social traits, in this case a variant of the promoter gene coding for the oxytocin receptor, and simultaneously surveying their real world behavior as reflected by the myriad different ways and purposes for which they used their phones over the course of the day.

The proliferation of touchscreen smartphones with software keyboards has, at least for the time being, tilted the balance of telecommunications in favor of typed rather than spoken messages (Shropshire 2015). Combined with the data provided by a phone's accelerometer, gyroscope, and screen pressure sensors, keystroke dynamics can be used to build mathematical models of a person's mood and cognition based only on how, and not what, they type.

Voice itself, of course, remains a valuable instrument for gaining insight into the speaker's mood state, and will only continue to become more so as the tide eventually turns toward speech-based interactions with both intelligent voice assistants and other human users of connected devices. Using similar statistical modeling and machine learning techniques, the acoustic features of speech are just as well-suited for analysis as typing kinematics (Cummings and Schuller 2019).

As more and more computing comes to be offloaded from personal devices to Internet of Things (IoT) devices and the cloud, and ambient computing becomes the norm, we expect that techniques like keystroke analysis will be supplanted by speech meta-feature analysis, facial emotional recognition (for more information on FER software, see Chap. 3 by Wilhelm and Geiger in this book), and altogether novel passive mood sensing tools. For the present time, being aware of the increasing ubiquity of algorithms and their influence on data analytics, digital architectures and digital societies (Dixon-Román 2016), as well as mindful of the absence of a codified analog for the Hippocratic Oath in the current practice of artificial intelligence in medicine as well as other applications (Balthazar et al. 2018), we nevertheless stand to learn a great deal from leveraging currently used input methods to derive models for sensing users' inner states.

13.2 Mobile Typing Kinematics

In the first known study of its kind, researchers from the University of Illinois at Chicago (UIC), the University of Michigan, the Politecnico di Milano, Tsinghua University and Sun Yat-sen University used passively obtained mobile keyboard usage metadata to predict changes in mood state with significant degrees of accuracy. The team recruited subjects from the Prechter Longitudinal Study of Bipolar Disorder at the University of Michigan as part of the BiAffect-PRIORI consortium for its pilot study based on an Android mobile keyboard and associated app. After winning the grand prize in the Mood Challenge supported by Apple and sponsored by the New Venture Fund of Robert Wood Johnson Foundation, UIC is currently conducting a full-scale study on the iOS platform using an app based on the open source ResearchKit mobile framework, enrolling both people with bipolar disorder as well as healthy controls from the general population.

The BiAffect study (https://www.biaffect.com/) involves the installation of a companion app containing a custom keyboard that is cosmetically similar to the stock system keyboard. The app includes mood surveys; self-rating scales; and active tasks such as a the go/no-go task and the trail-making test (part B) to measure reaction time, response inhibition, and set-shifting as part of executive functioning—all overlapping domains of cognition identified by Bourne and colleagues (2013) to be the most affected in bipolar disorder.

All data collected by the app and keyboard are first encrypted and then transmitted and stored on secure study servers; these were hosted at UIC for the Android pilot app, whereas study management services are being supported by Sage Bionetworks for the ongoing iOS study with the data being hosted on their Synapse platform. The Android pilot phase, which has concluded data collection, involved the keyboard, trail making test, Hamilton Depression Rating Scale (HDRS), Young Mania Rating Scale (YMRS), and slider-based daily self-rating scales for mood, energy, impulsiveness, and speed of thoughts; the main iOS study included each of these [with the notable substitution of the clinician-rated HDRS and YMRS with the self-reported Patient Health Questionnaire (PHQ) and the Altman Mania Rating Scale, respectively] as well as a daily self-rating scale querying ability to focus, and the aforementioned reaction time task. Metadata collected for keyboard usage include timestamps associated with each keystroke, residence time on each key, intervals between successive keystrokes, and accelerometer readings over the course of all active typing sessions. The actual character corresponding to any given keypress is not recorded, apart from noting whether it was a backspace, alphanumeric, or symbol key. In addition to backspace usage, instances of autocorrection and autosuggestion invocations are also logged.

Table 13.1 summarizes the literature that has been published thus far based on analyses of data collected during the pilot phase of the study, which included 40 participants—between 9 and 20 of whose data were used for any given one depending on the number of days of metadata logged, diagnosis of the participant, and other requirements; up to 1,374,547 keystrokes and 14,237,503 accelerometer readings

Table 13.1 A summary of analyses published by researchers using data from the BiAffect study

Author	Analytical technique	Predictors used	Main findings
Zulueta et al. (2018)	Linear mixed-effects models *(Preferred over ANOVA in settings where measurements are made on clusters of related statistical units due to advantages in dealing with missing values)*	Average inter-key delay, backspace ratio, autocorrect rate, circadian baseline similarity, average accelerometer displacement, average session length, and session count	Keystroke activity was predictive of depressive, and to a lesser extent, manic symptoms. Specifically, accelerometer displacement, average inter-key delay, session count, and autocorrect rate were positively correlated with the HDRS scores, whereas accelerometer displacement was positively correlated and backspace rate negatively correlated with YMRS scores
Stange et al. (2018)	Multilevel models to evaluate predictiveness of instability metrics computed using the root mean square successive difference *(Specific models for each level of multilevel data, thereby modeling the non-independence of observations due to cluster sampling)*	Instability of EMA affective ratings and daily typing speed	Greater instability of mood during baseline EMA was predictive of future depressive symptoms, while instability of energy predicted future manic but not depressive symptoms. Instability of typing speed predicted prospective depressive but not manic symptoms. Models built using data gathered during only 5–7 days were as reliable and predictive as those assessing instability over longer time periods
Cao et al. (2017)	Comparison of late fusion based DeepMood LSTM-type GRU ML architecture with a multi-view RNN machine layer, factorization machine layer, or conventional fully connected layer against early fusion approaches *(Recurrent connections between machine learning layers allow modeling of nonlinear time series that, after training on sufficient data, can solve problems with prolonged temporal dependencies, such as linguistic, semantic, and topic inference tasks.)*	Multiple representative views of the features of each typing session such as alphanumeric characters, special characters, and accelerometer values	Healthy people showed a wider range of variability in the time intervals between successive alphanumeric keypresses than people who were experiencing a mood disturbance. People in a manic state tend to hold down a keypress longer than people in a stable mood state, while depressed people pressed down on keys for shorter than average durations. The DMVM and DFM based architectures were the most predictive of depression scores, with prediction performances of 90.31% and 90.21%, respectively

(continued)

Table 13.1 (continued)

Author	Analytical technique	Predictors used	Main findings
Huang et al. (2018)	dpMood ML architecture based on early fusion, stacked CNNs to capture local typing dynamics and RNNs to capture temporal dynamics, and final predictions based on individual circadian calibrations *(More typically employed in computer vision models, CNNs outperform shallow architectures when predicting mental health related aspects from multiple data streams, while reaching at least comparable performance levels as predesignated architectures.)*	Metadata for alphanumeric characters, including duration of keypress, time since last keypress, distance from the center of the last pressed key along both axes, and corresponding accelerometer values during active sessions	The proposed dpMood architecture incorporating CNNs, RNNs, early fusion and time-based calibration taken together outperformed any individual approach alone or in combination with just a few others. The integrated analysis of local patterns and temporal dependencies allowed for the isolation of variations in keyboard usage at different times of the day and from day to day over the course of the week, and the personalized calibration was sensitive enough to be able to distinguish between healthy controls and subjects with type I and type II bipolar disorder
Vesel et al. (2020)	Growth curve mixed-effects (multilevel) models in R and lme4 using maximum likelihood fitting *(R version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria; lme4 version 1.1-21)*	Examined dependent variables of session-level typing speed, typing variability, typing accuracy, and session duration and their relationship to other session-level features and demographics	More severe depression relates to more variable typing speed ($P < 0.001$), shorter session duration ($P < 0.001$), and lower accuracy ($P < 0.05$). Additionally, typing speed and variability exhibit a diurnal pattern, being fastest and least variable at midday. Older users exhibit slower and more variable typing, as well as more pronounced slowing in the evening. The effects of aging and time of day did not impact the relationship of mood to typing variables
Ross et al. (2021)	Longitudinal mixed-effects models (with maximum likelihood estimator fitting) were used to analyze daily digital trail-making test, part B (TMT-B) performance as a function of typing and mood *(All analyses were conducted in R version 3.6.3; R Core Team 2020)*	Keypress metadata, paper and digital TMT-B completion times, and Hamilton Depression Rating Scale scores	Participants who typed slower took longer to complete dTMT-B, with this trend also being seen in individual fluctuations in typing speed and dTMT-B performance. Participants who were more depressed completed the dTMT-B slower than less depressed participants

(continued)

Table 13.1 (continued)

Author	Analytical technique	Predictors used	Main findings
Zulueta et al. (2021)	Two random forest regression models were trained using the caret and randomForest packages for R *(All statistical testing was performed in R version 4.0.0)*	Features derived from the smartphone kinematics were used to train random forest regression models to predict age	Smartphone kinematics were successfully used to predict chronological age. The absolute prediction error tended to be lower for participants with positive screens than those with negative screens, whereas the raw prediction error tended to be lower for participants with negative screens than those with positive screens

Abbreviations *EMA* ecological momentary assessment; *HDRS* Hamilton Depression Rating Scale; *YMRS* Young Mania Rating Scale; *LSTM* long short-term memory; *GRU* gated recurrent unit; *ML* machine learning; *DMVM* DeepMood multi-view machine; *DFM* DeepMood factorization machine; *DNN* DeepMood neural network; *CNN* convolutional neural network; *RNN* recurrent neural network

across 37,647 sessions were incorporated into some of the resulting models. Data collection for the main arm of the study is ongoing and has already resulted in over 8000 cumulative hours of active typing sessions culled from across hundreds of users.

Zulueta and colleagues (2018) built mixed-effects linear models to correlate keyboard activity metadata during the week preceding when each pair of mood rating scales was administered to the corresponding HDRS and YMRS scores. A representative sampling of these metadata over several weeks from one study participant is illustrated in Fig. 13.1, while Fig. 13.2 compares the scores predicted by these models against actual scores for both mood scales. Autocorrect rates were positively correlated with depression scores, probably because error-awareness becomes impaired when depressed (Fig. 13.3a). Backspace usage rate was found to be negatively correlated with higher mania scores, possibly because it is reflective of decreased self-monitoring and impaired response inhibition (Fig. 13.3b). Accelerometer activity was positively correlated with both depression and mania scores, possibly because study subjects were experiencing depression with mixed features or agitated/irritable depression. The trail making test, which consists of circles with alternating consecutive numbers and/or letters that respondents are directed to connect in the correct order, is a standard neuropsychological assessment that measures processing speed and task-switching, which are both good indicators of cognitive functioning; Fig. 13.4 shows how typing kinematics data were just as predictive as trail making test results at establishing cognitive ability.

Stange et al. (2018) took a different approach by constructing multilevel models based on instability metrics calculated for EMA ratings and daily typing speeds (Fig. 13.5) using the root mean square of the successive differences (rMSSD)—a time-domain measure that takes into account the magnitude, frequency, and temporal order of intra-user fluctuations (Ebner-Priemer et al. 2009). Greater instability in baseline mood EMA ratings was significantly predictive of elevated future symptoms of both depression (Fig. 13.6a) and mania, whereas instability in energy ratings

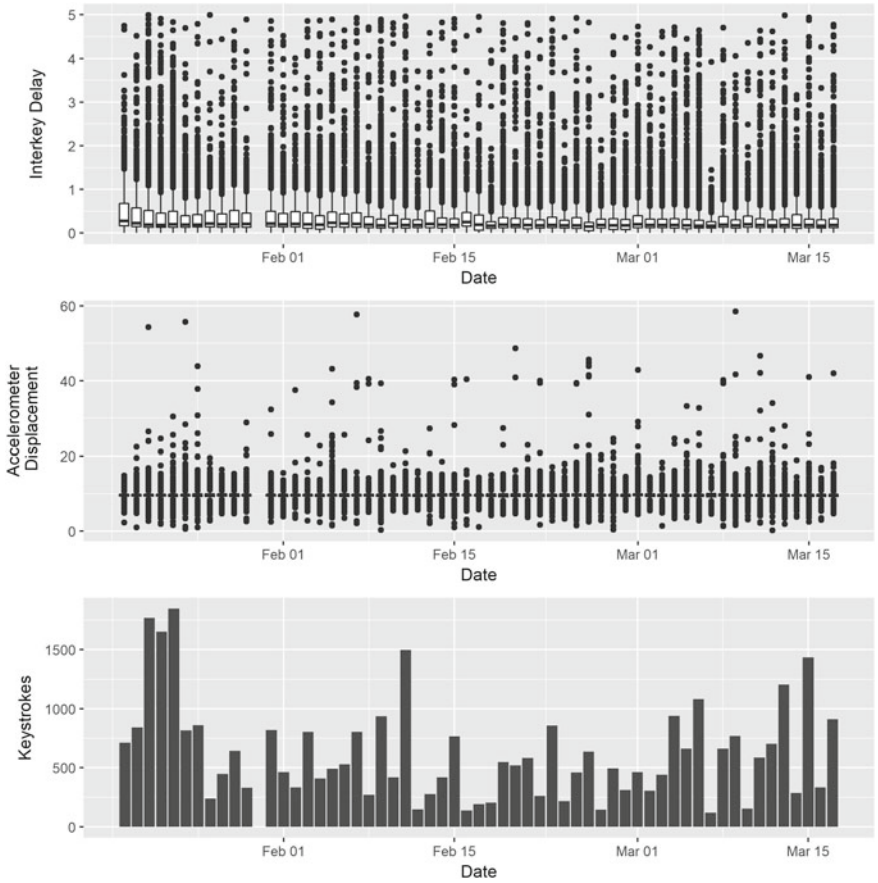


Fig. 13.1 An example of the deep personalized sensing possible with BiAffect showing the number of keystrokes, corresponding accelerometer readings, and the time between successive keypresses logged for an individual participant over the duration of the pilot study phase. Adapted from Zulueta et al. (2018)

was predictive of future mania but not depression; other affective EMA ratings were not found to be significantly predictive of either. Typing speed instability was predictive of elevated prospective symptoms of depression (Fig. 13.6b) but not of mania. Interestingly, as little as one week of data provided levels of predictiveness comparable to data collected over durations of time longer than 5–7 days, perhaps because this time period is a representative enough snapshot to capture day-to-day typing variability (Fig. 13.7). Turakhia and colleagues (2019) have subsequently gone on to demonstrate the feasibility of exploiting variability in similar irregular noncontinuous datastreams to identify, predict, and prevent potential serious episodes—atrial flutters and fibrillations in the case of their app- and wearable-based study on cardiac arrhythmia.

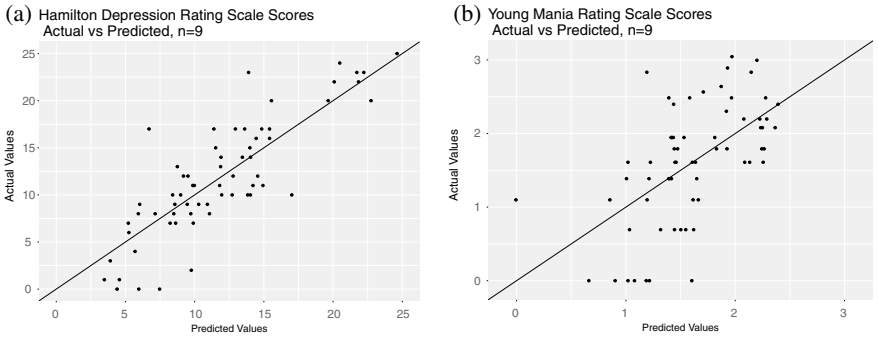


Fig. 13.2 Mixed effects modeling accounted for 63% of the variability of Hamilton Depression Rating Scale scores (Conditional $R^2 = 0.63$, Marginal $R^2 = 0.41$, $\chi^2_7 = 17.6$, $P = 0.014$). Ordinary least squares modeling accounted for 34% of the natural log of Young Mania Rating Scale scores (Multiple $R^2 = 0.34$, Adjusted $R^2 = 0.26$, $F_{7,56} = 4.1$, $P = 0.0011$) Adapted from Zulueta et al. (2018)

Cao and colleagues (2017) were among the first to model keystroke dynamics data using deep learning. Their method, DeepMood, consisted of comparing the predictive performance of a multi-view machine layer architecture (Fig. 13.8) to that of other late fusion approaches such as factorization and conventional fully connected layers as well as early fusion strategies like tree boosting systems, linear support vector machines, and logistic ridge regression models. For the uninitiated, a review on current applications of deep neural networks in the field of psychiatry by Durstewitz et al. (2019) may serve as a primer. DeepMood's early fusion approaches align each of the data views—alphanumeric characters, special characters, and accelerometer values—with their associated timestamps (Fig. 13.9), and then immediately concatenate the multi-view time series per session. However, this does not take into proper account unaligned features in certain views, such as special characters, that do not have corresponding data points from other views like acceleration or inter-key distance. This shortcoming is addressed by the late fusion approach, in which each of the multi-view series is first modeled separately by a recurrent neural network (RNN), and then fused in the next stage by analyzing first-, second-, and third-order interactions between each view's output vectors. Cao and colleagues established that their late fusion approach significantly outperformed early fusion in the ability to predict mood disturbances and their severity (Fig. 13.10), with the multi-view machines demonstrating the highest rate of accuracy at 90.31% followed by the factorization machines at 90.21%.

In a subsequent analysis, Huang et al. (2018) found that an early fusion approach integrating both convolutional and recurrent deep architectures and incorporating users' circadian rhythms allowed their model, dpMood, to attain even greater predictive performance as well as make more precise personalized mood predictions that took into fuller account an individual's biological clock and unique typing patterns. Their approach consisted of using convolutional neural networks (CNNs)

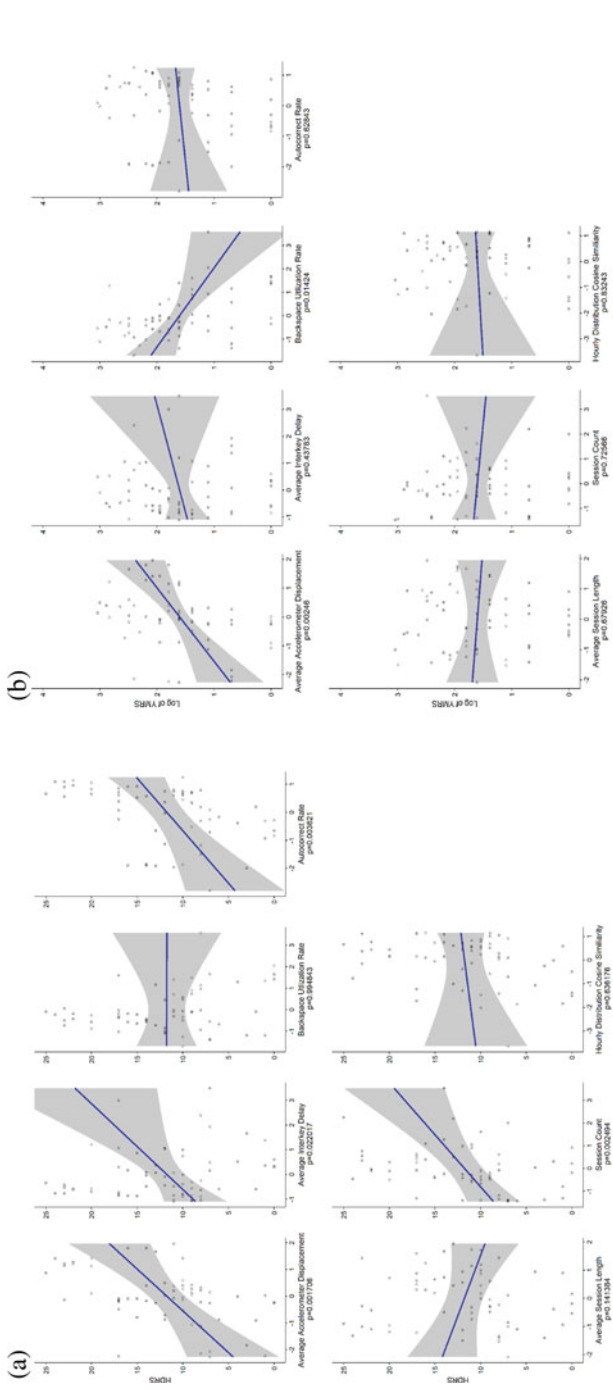


Fig. 13.3 Significant predictors for Hamilton Depression Rating Scale scores included accelerometer displacement ($P = 0.0017$), interkey delay ($P = 0.022$), autocorrect rate ($P = 0.0036$), and session count ($P = 0.0025$). Significant predictors for the natural log of the Young Mania Rating Scale scores include accelerometer displacement ($P = 0.003$) and backspace rate ($P = 0.014$). Adapted from Zulueta et al. (2018)

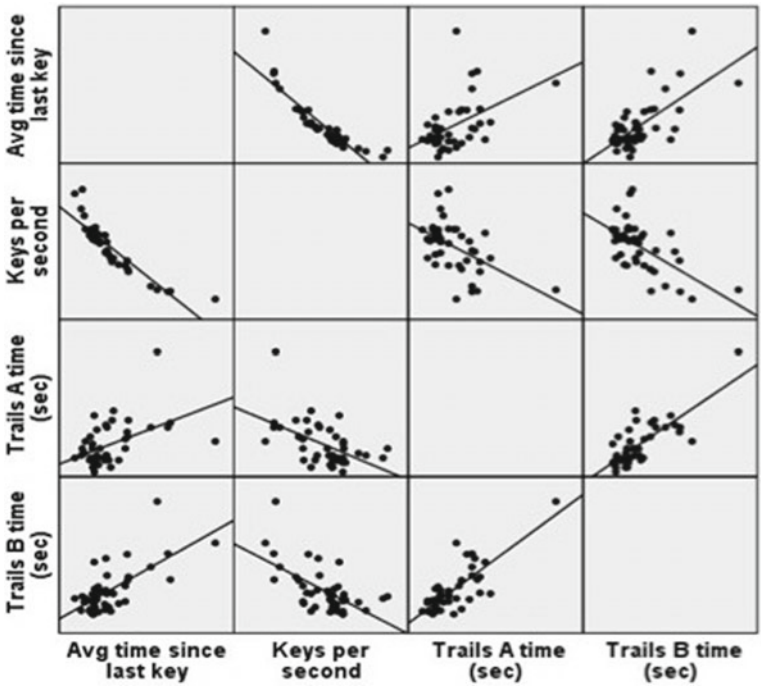


Fig. 13.4 Comparison of the predictiveness of keystroke data with that of trail making test results for assessing cognitive ability. Processing speed, as measured by trail taking test (part A) scores, was significantly correlated with average interkey delay (i.e., time since last key, $r = 0.5$, $p < 0.001$) and keys/second ($r = -0.54$, $p < 0.001$). Set shifting, as measured by trail taking test (part B) scores, was highly associated with average time since last key ($r = 0.68$, $p < 0.00001$) and keys/second ($r = -0.62$, $p < 0.00001$). Adapted from Zulueta et al. (2018)

that focused on temporal dynamics to analyze local features in typing kinematics over small periods of time, in conjunction with a special type of RNN called a gated recurrent unit (GRU) to model longer-term time-related dynamics (Fig. 13.11). GRUs address the vanishing gradient problem—the inherent inability of simpler RNNs to effectively learn those parameters that only cause very small changes in the neural network's output—and moreover have fewer parameters than comparable ameliorative approaches, allowing them to perform better on smaller datasets (Cho et al. 2014) such as the keystroke kinematics collected by BiAffect. This early fusion approach allowed for the alignment of features from multiple views to include additional information about temporal relationships between these data points that would otherwise be lost in late fusion models. In the final analysis, the proposed dpMood architecture with the best predictive performance and the lowest regression error rate was the one that made combined use of both CNNs and RNNs to learn local patterns as well as temporal dependencies, learned each user's individual circadian rhythm, and retained accelerometer values that had no contemporaneous alphanumeric keypresses by filling the unaligned alphanumeric features with zero values instead of dropping

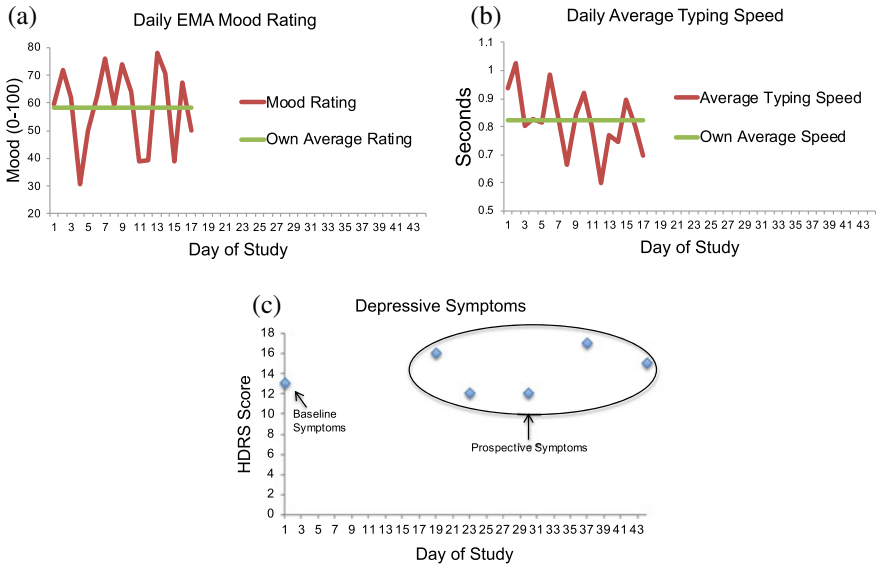


Fig. 13.5 An individual participant's a self-rated ecological momentary assessment scores, b passively collected daily typing speeds and c baseline and future course of depression symptom severity. Adapted from Stange et al. (2018) and reproduced with permission from the publisher

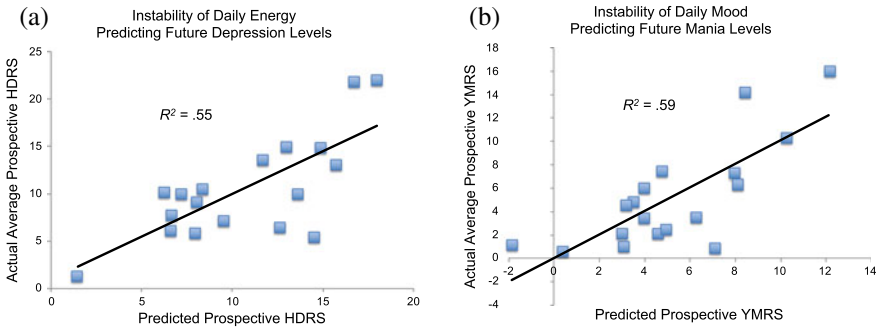


Fig. 13.6 Comparison of actual scores with those predicted by multilevel instability models for an individual participant's a Hamilton Depression Rating Scale and b Young Mania Rating Scale. Adapted from Stange et al. (2018) and reproduced with permission from the publisher

unaligned accelerometer values altogether. Accelerometric and time-based analyses elucidated both daily (Figs. 13.12 and 13.13) and hourly (Fig. 13.14) variations in keyboard use, with the notably smaller Z-axis accelerations that help pinpoint when a phone is being typed on from a supine position having been observed more predominantly in the evenings (Fig. 13.14c) and on weekends (Fig. 13.13d). Modeling individuals' circadian rhythms as a sine function with parameters automatically learned by gradient descent algorithms and backpropagation resulted in one of

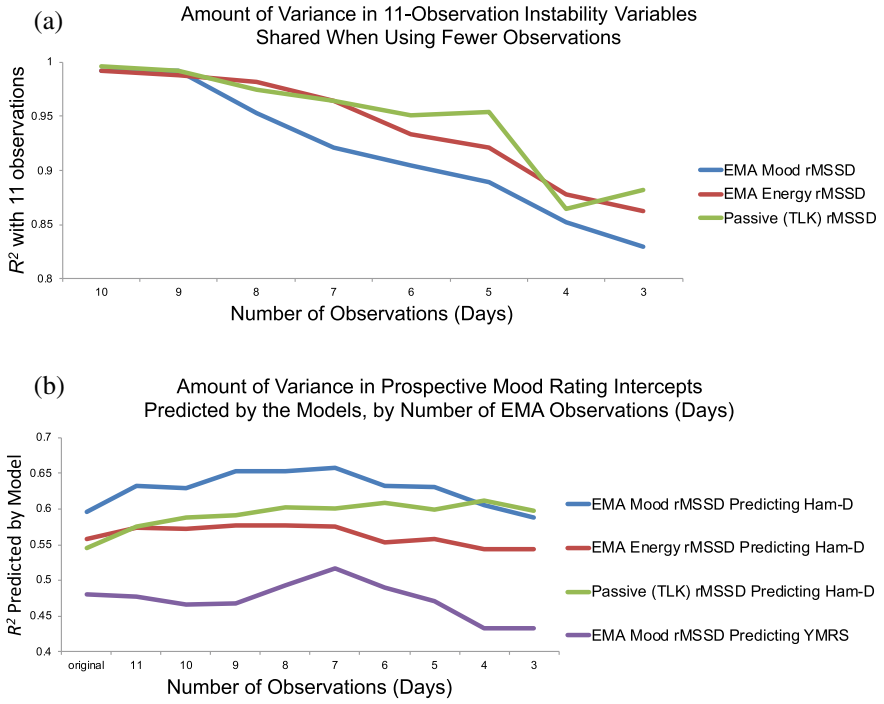


Fig. 13.7 **a** Reliability of active and passive assessments of instability depending on number of days of assessment. **b** Predictive utility of active and passive assessments of instability depending on number of days of assessment. Adapted from Stange et al. (2018) and reproduced with permission from the publisher

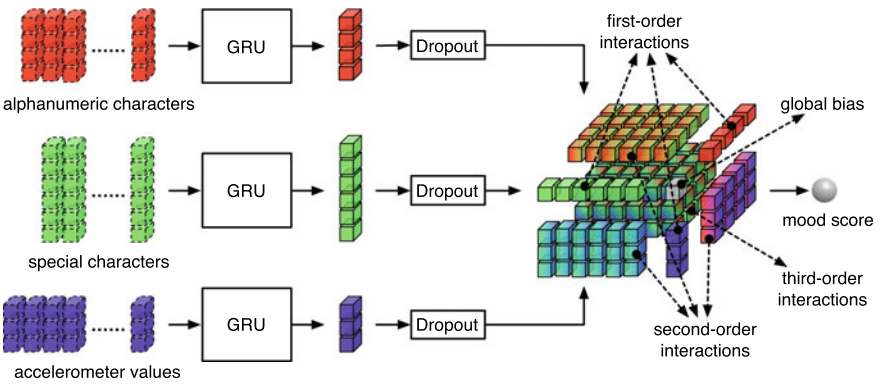


Fig. 13.8 DeepMood machine learning architecture with a multi-view machine layer for late data fusion. Adapted from Cao et al. (2017)

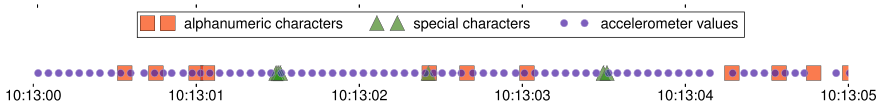
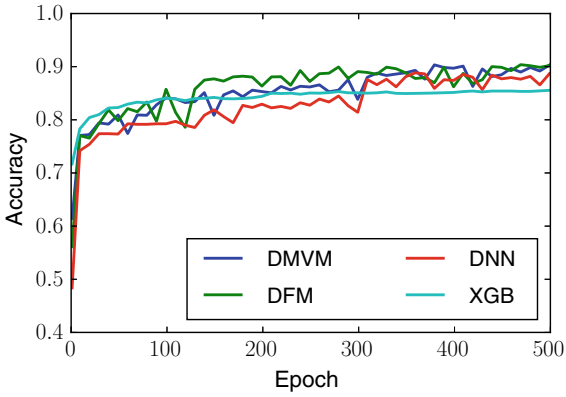


Fig. 13.9 A representative sample of the multi-view metadata collected in a time series. Adapted from Cao et al. (2017) and reproduced with permission from the publisher

Fig. 13.10 Comparison of the improvements in accuracy of different DeepMood architectural approaches over the course of successive training epochs. Adapted from Cao et al. (2017) and reproduced with permission from the publisher



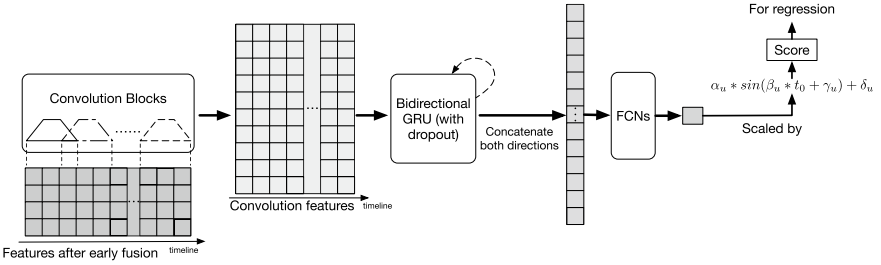


Fig. 13.11 dpMood machine learning architecture based on early data fusion stacked CNNs and GRUs, and time-based calibrations.Adapted from Huang et al. (2018) and reproduced with permission from the publisher

these parameters conspicuously clustering based on the subjects' diagnoses, permitting dpMood to successfully classify users as participants with bipolar I disorder, those with bipolar II disorder, or healthy controls (Fig. 13.15). These sophisticated techniques can combine to provide extraordinarily insightful mood-sensing tools to users and precision medicine practitioners alike.

Preliminary analysis of study participants' performance on the go/no-go task has indicated that reaction times vary both within and between individuals (Fig. 13.16a) as well as continue to change over time (Fig. 13.16b); variations in daily typing patterns in BiAffect users have been found to correlate with their performance on the

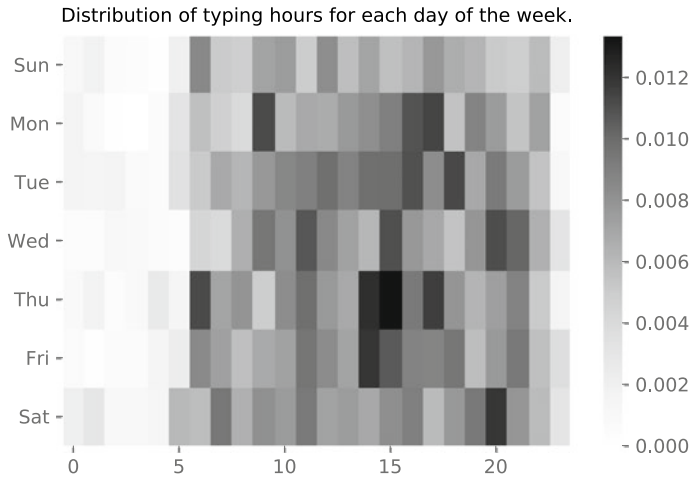


Fig. 13.12 Distribution of daily typing hours visualized as a 7 day × 24 h matrix. Adapted from Huang et al. (2018) and reproduced with permission from the publisher

go/no-go task, and concurrent analyses of both data streams are now under way to examine their interrelationships and interactions with mood and cognition as well.

Vesel and colleagues (2020) investigated the effects of mood, age, and diurnal patterns on intraindividual variability (IIV) in typing behaviors recorded in the iOS dataset, correlated against participants' responses to the PHQ. Interkey delay (IKD) was calculated as the time difference between 2 consecutive keypresses and analysis was restricted to only IKDs between character-to-character keypress events; typing speed of a session was operationally inferred using the median IKD of that session. Typing variability at the session level was quantified using the median absolute deviance of IKDs. Typing mode (the use of one or two hands when typing) was classified using a novel approach utilizing linear regression. Growth curve mixed-effects (multilevel) models were established using maximum likelihood fitting to examine dependent variables of session-level typing speed, typing variability, typing accuracy, and session duration and their relationship to other session-level features and demographics (Fig. 13.17).

It was established that typing speed exhibits slowing with age, while pausing between typing and variability in typing speed increase with age. The relationship between keystroke dynamics features and mood was supported by the significantly higher variability in IKDs observed with more severe depression, consistent with reported findings of higher IIV in task performance in mood disorders. Typing accuracy, as encoded using session-level autocorrect rates, was also found to decrease in more depressed individuals. Finally, sessions corresponding to elevated depressive symptoms were found to be shorter in duration, suggesting a decrease in smartphone keyboard use during more severe depression.

Ross et al. (2021) evaluated the efficacy of using smartphone typing dynamics along with mood scores in cognitive assessment as an adjunct to formal in-person

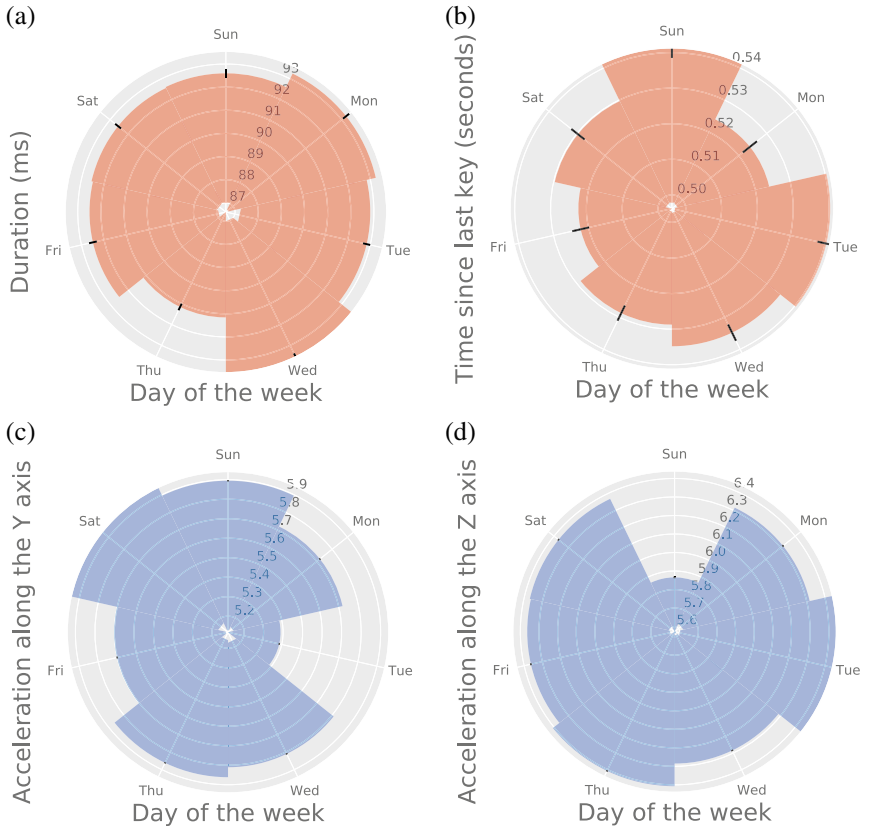


Fig. 13.13 Day-to-day fluctuations over the course of a week in a duration of a keypress, b time between successive keypresses, c acceleration along Y-axis, and d acceleration along Z-axis. Adapted from Huang et al. (2018) and reproduced with permission from the publisher

neuropsychological assessments through trail making tests. In addition to using the Android pilot app keyboard, participants were administered the pencil-and-paper version of the trail-making test, part B (pTMT-B) at the beginning and end of the study, as well as completed digital TMT-Bs (dTMT-B) throughout the study on their smartphones, and responded to the Hamilton Depression Rating Scale (HDRS) and Young Mania Rating Scale (YMRS) over the course of weekly phone interviews. For analysis, time windows were selected such that each consisted of one dTMT-B, one HDRS-17 score, and multiple keypresses, as shown in Fig. 13.18.

Intraclass correlations between the digital and paper-based forms of TMT-B were calculated to assess the consistency between both modalities. Comparison of the first dTMT-B to paper TMT-B showed adequate reliability. Longitudinal mixed-effects models were then used to analyze daily dTMT-B performance as a function of typing and mood. Participants who typed slower were observed to take longer to complete

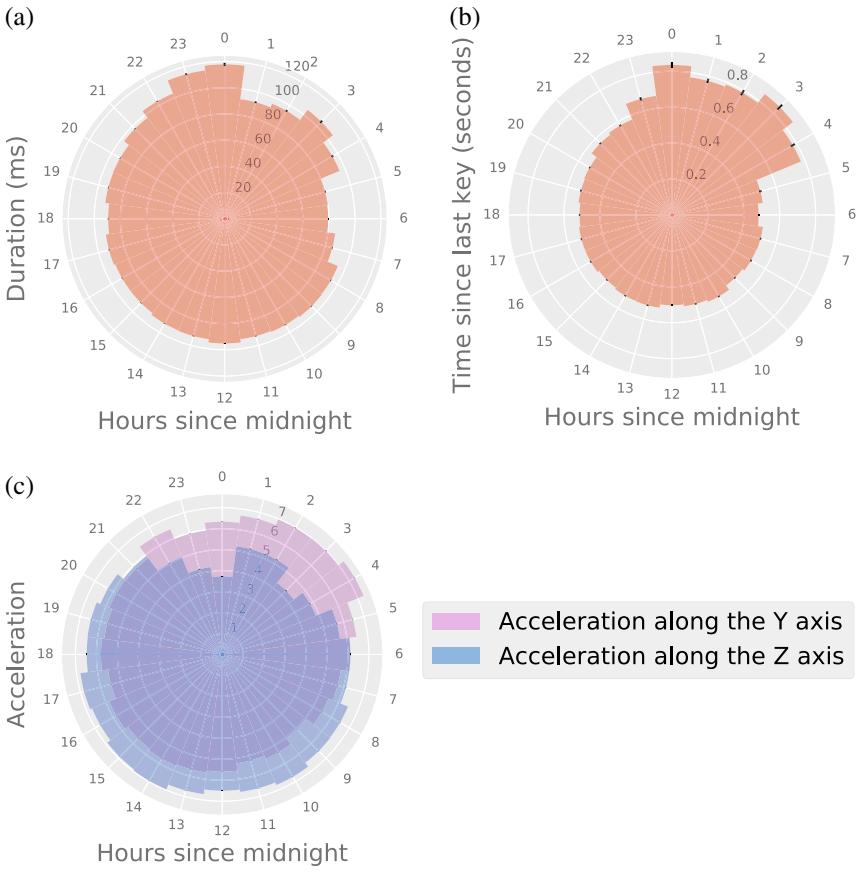


Fig. 13.14 Circadian rhythm mediated fluctuations in a duration of a keypress, b time between successive keypresses, and c acceleration along Y- and Z-axes. Adapted from Huang et al. (2018) and reproduced with permission from the publisher

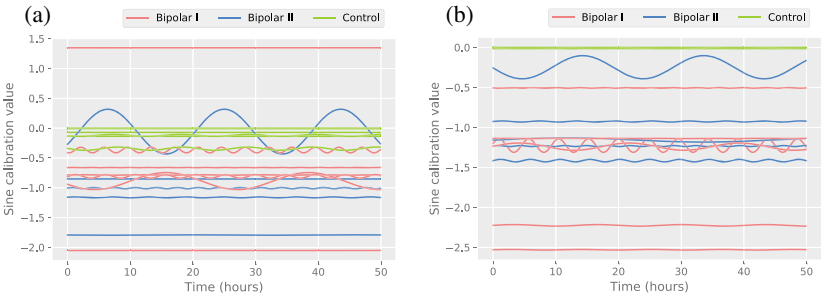


Fig. 13.15 Visualizations of each individuals' calibration sine functions for a Hamilton Depression Rating Scale scores and b Young Mania Rating Scale scores. Adapted from Huang et al. (2018) and reproduced with permission from the publisher

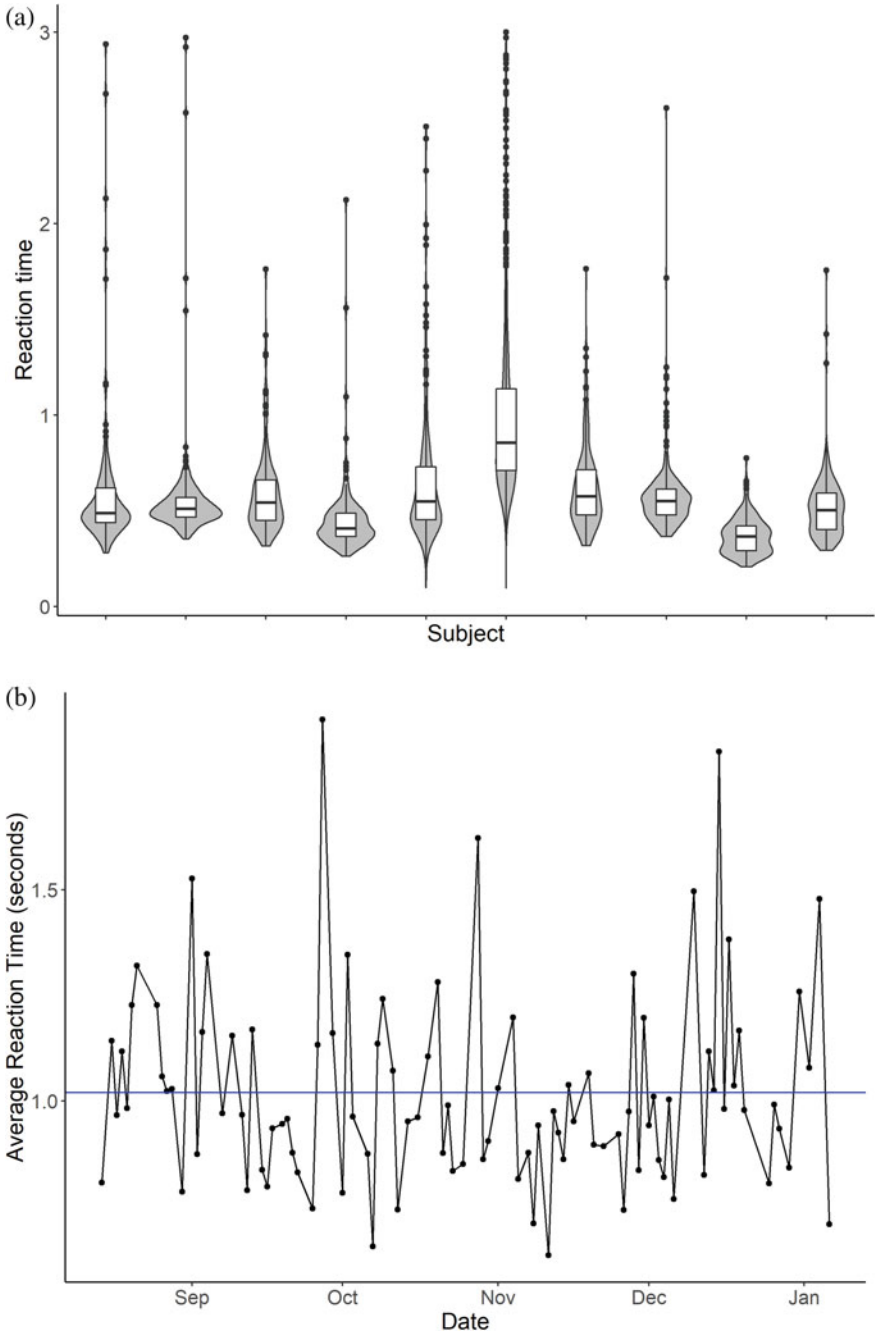


Fig. 13.16 A Go/no-go reaction time varies between and within individuals. b Average reaction time changes over the course of time

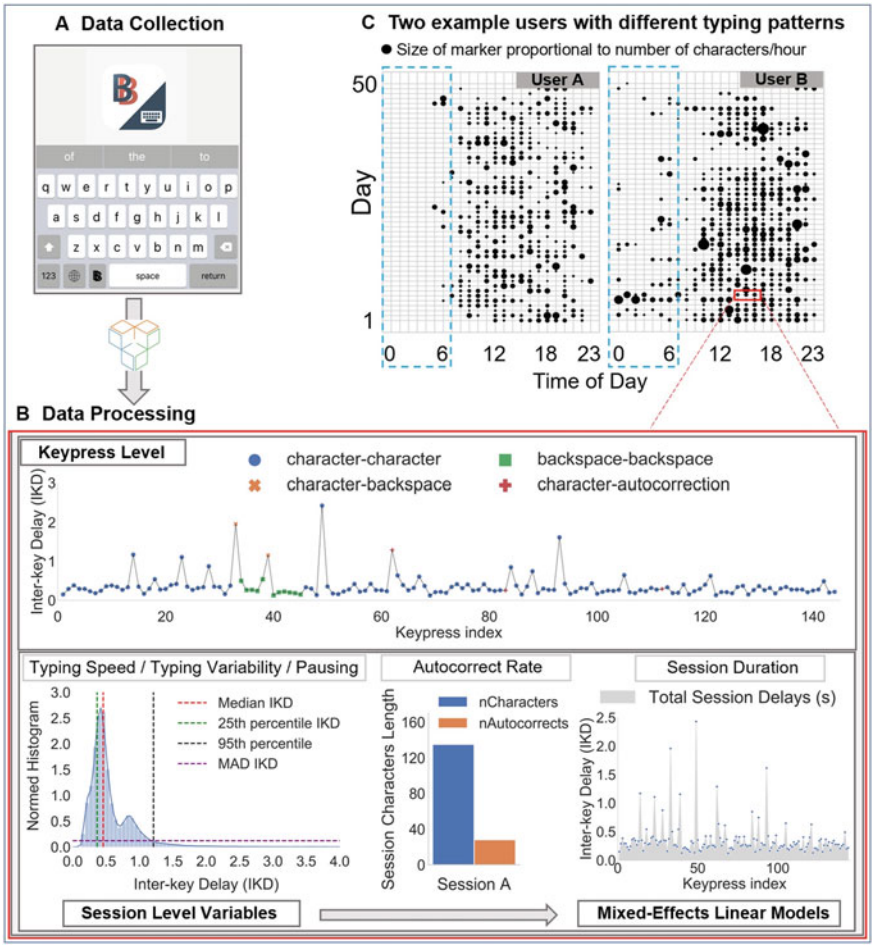


Fig. 13.17 Overview of BiAffect data collection and feature extraction process. **a** Keypress-level typing metadata are collected via the BiAffect keyboard and stored by Sage Bionetworks. **b** Interkey delays for keypress transitions from character to character are aggregated at a session level to compute median absolute deviance alongside typing accuracy and session duration. **c** An example for the hourly typing activity over multiple days from 2 active users is presented as an illustration of the potential patterns captured via continuous, unobtrusive collection. The blue dashed line highlights the different levels of activity at night, with user B exhibiting a more irregular activity pattern than user A. Size of the marker is proportional to the number of characters typed per hour

dTMT-B. This trend was also seen in individual fluctuations in typing speed and dTMT-B performance (Fig. 13.19). Moreover, participants who were more depressed completed the dTMT-B slower than less depressed participants (Fig. 13.20).

Depression severity was associated with the dTMT-B time at both the inter- and intrasubject level. Participants who were more depressed completed dTMT-B more slowly than participants who were not depressed. Typing speed was also associated

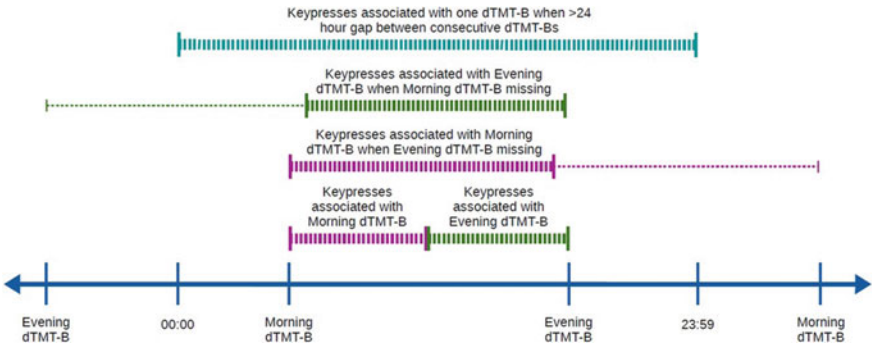


Fig. 13.18 Schematic outlining how keypresses were assigned to each digital trail making test part B (dTMT-B) to account for missing data

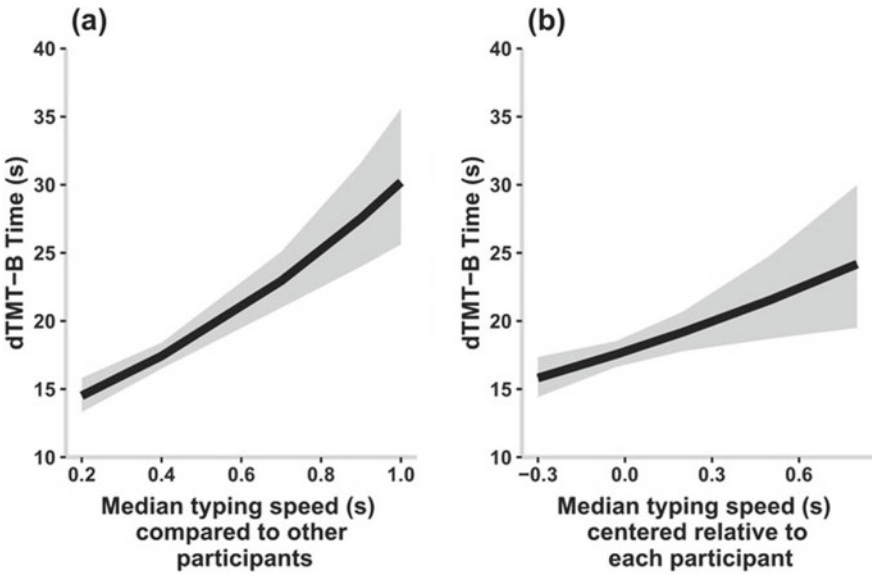


Fig. 13.19 Digital trail-making test part B completion time as a function of grand mean centered (**a**) and subject centered (**b**) typing speed with ribbons showing the 95th confidence interval

with the dTMT-B at both inter- and intrasubject levels. Faster typists completed the dTMT-B more quickly than slower typists. Participants' individual fluctuations in typing speed reflected their fluctuations in dTMT-B over the course of the study. A diagnosis of bipolar disorder was found to be a significant predictor of dTMT-B completion time, after controlling for depression score and typing speed.

Zulueta et al. (2021) analyzed participants' responses to the Mood Disorders Questionnaire (MDQ) and self-reported birth year against Features derived from the smartphone kinematics, which were used to train random forest regression models

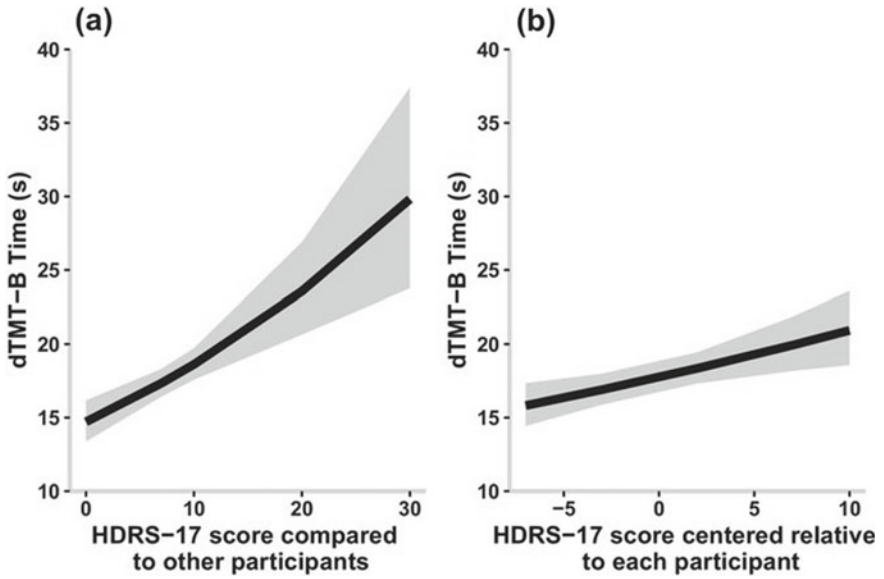


Fig. 13.20 Digital trail-making test part B completion time as a function of grand mean centered (**a**) and subject centered (**b**) Hamilton Depression Rating Scale score with ribbons showing the 95th confidence interval

to predict age. Data were split into training and validation sets (75:25). Two random forest regression models were trained using the caret and randomForest packages for R. The mtry value which minimized the Root Mean Square Error (RMSE) was selected as the value used in the final models. The models were constructed in a stepwise fashion with the first model including only typing related features, and the second model included all features from the first with the addition of gender and MDQ screening status. Each model's performance was assessed using the validation set to calculate RMSE, Breiman's pseudo R-squared, and median absolute error. Differences in model performance testing were assessed using paired Wilcoxon tests of their absolute errors. Feature importance was assessed using out-of-bag changes in Mean Square Error (MSE). Accumulated Local Effects plots (ALE Plots) were constructed for features which appeared important or interesting. Differences within model performance between participants based on MDQ screen status were assessed using Wilcoxon tests comparing raw prediction error scores and absolute prediction error scores.

Compared to participants with positive MDQ screens, participants with negative screens had a lower rate of reporting a diagnosis of bipolar disorder, a higher rate of reporting no history of bipolar disorder, and also provided no diagnosis history at a lower rate. The participants with negative screens tended to have lower MDQ scores comparted to those with positive screens and have a greater total number of keypresses. Plots A–D of Fig. 13.21 depict the ALE plots of four of the most important features: the median of mean interkey times, the mean session length, the

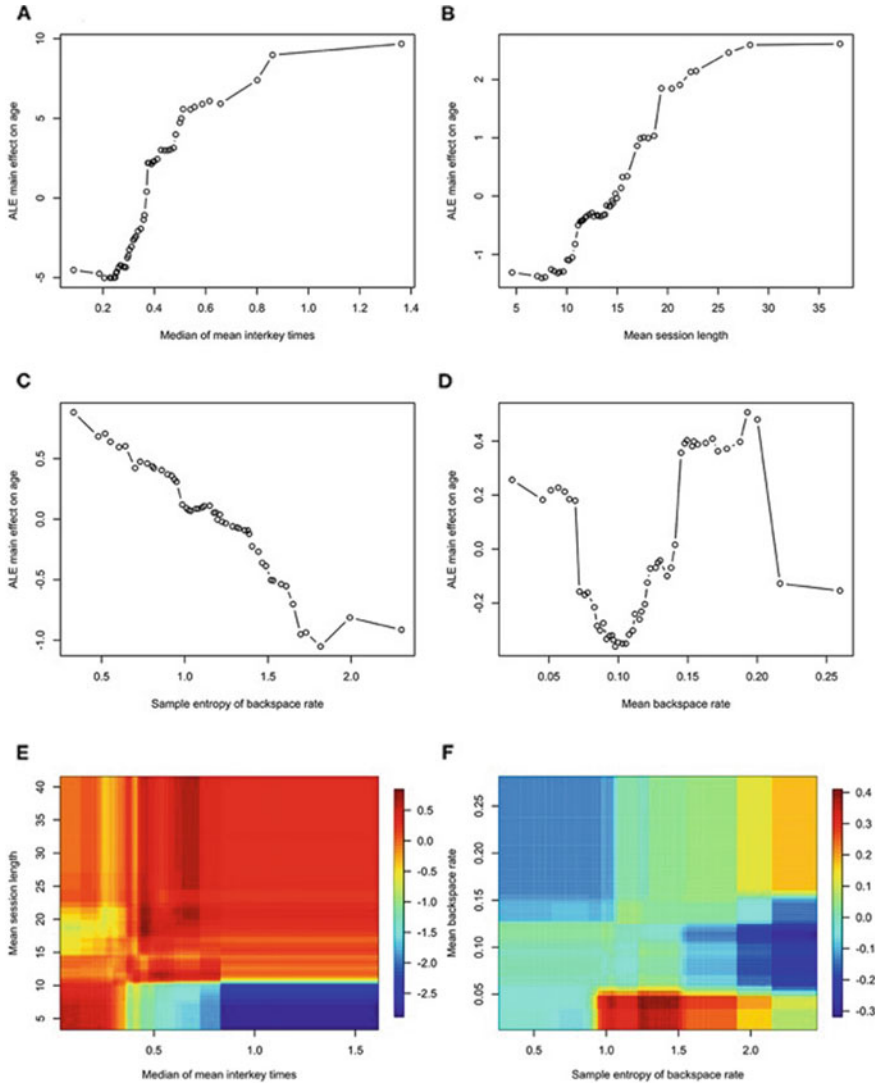


Fig. 13.21 Accumulated Local Effects plots for the second model. **a**–**b** Depict the effects of individual features on age prediction. **e**, **f** depict the interaction of the two indicated effects on age

sample entropy of the backspace rate, and the mean backspace rate. Many of the most important features are different summaries of the same essential feature (e.g., interkey time). Based on these plots, increased interkey time and session length are both generally associated with increased age; whereas, increased sample entropy of the backspace rate is associated with younger age, and the association between age and the mean backspace rate is not monotonic. Plots E and F of Fig. 13.21 depict the

interaction between the median of mean interkey times and the mean session length and between the mean backspace rate and the sample entropy of the backspace rate, respectively. In these plots, it is observed that the existence and directionality of linear trends between the predicted age and these features both depend on the range of a second associated feature, highlighting the complexity of the relationship between typing behaviors and predicted age.

The tendency to underestimate the chronological age of participants screening negative for bipolar disorder compared to those screening positive is consistent with the finding that bipolar disorder may be associated with brain changes that could reflect pathological aging. This interesting result could also reflect that those who screen negative for bipolar disorder and who engaged in the study were more likely to have higher premorbid functioning. This work demonstrates that age-related changes may be detected via a passive smartphone kinematics based digital biomarker.

13.3 Speech Dynamics

Research on keystroke kinematics was inspired by the work of colleagues at the University of Michigan's Heinz C. Prechter Bipolar Research Program on the Predicting Individual Outcomes for Rapid Intervention (PRIORI) project, which is based on analyzing voice patterns in participants enrolled in the longest longitudinal research study of bipolar disorder; BiAffect aims to infer mood from typing metadata just as PRIORI does from the acoustic meta-features of speech. Participants were enrolled in the PRIORI study for an average of 16 to 48 weeks and were provided a rooted Android smartphone with a preinstalled secure recording application that captured audio of the participant's end of every phone call. Study staff called participants weekly to administer HDRS and YMRS mood assessments; these calls were labeled separately from personal calls. The dataset has accumulated over 52,000 recorded calls totaling above 4,000 h of speech from 51 participants with bipolar disorder and 9 healthy controls.

Karam et al. (2014) used a support vector machine (SVM) classifier to perform participant-independent modeling of segment- and low-level features extracted by the openSMILE audio signal processing toolkit, and were able to separate euthymic speech from hypomanic and depressed speech using an average of 5 to 8 judiciously selected features. In a later study, Gideon et al. (2016) used a declipping algorithm to approximate the original audio signal, and performed noise-robust segmentation to improve inter-device audio recording comparability. Rhythm features were classified using multi-task SVM analysis, then transformed into call-level features, and finally Z-normalized either globally or individually by subject. Declipping and SVM classification was found to increase the performance of manic but not depressive predictiveness, whereas segmentation and normalization significantly increased both. Khorram et al. (2016) captured subject-specific mood variations using i-vectors, and utilized a speaker-dependent SVM to classify both these i-vectors as well as rhythm features. Fusion of the subject-specific model—using unlabeled personal calls—with

a population-general system enabled significantly improved predictive performance for depressive symptoms compared to the earlier approach used by Gideon and colleagues (2016). Khorram et al. (2018) went on to develop an 'in the wild' emotion dataset collating valence and activation annotations made by human raters drawing only upon the acoustic characteristics, and not the spoken content, of recordings from both personal and assessment calls.

Ongoing analyses, confounding challenges, and proposed solutions related to voice analysis have been outlined in a concise review by the PRIORI team (McInnis et al. 2017); their current focus is to isolate elements in the speech signal that are most strongly correlated with incipient disturbances in mood, enabling the development of on-device analytical systems without compromising limited mobile phone battery life.

13.4 Future Directions

The eventual goal of these projects is to be able to generate an early warning signal when changes in users' patterns of typing, speech, and behavior identify them to be at risk for an imminent manic or depressive episode. This would allow for just-in-time adaptive interventions that can circumvent or at least minimize the acuteness of the episode and any resulting cases of hospitalization, medication adjustment, or self-harm (Rabbi et al. 2019).

It has not escaped our attention that these passive sensing techniques can have applications in conditions other than bipolar disorder and indeed beyond just mood disorders; we have been investigating the use of a voice-enabled intelligent agent that are responsive to users' mood in order to provide emotionally aware education and guidance to patients with comorbid diabetes and depression (Ajilore 2018), as well as exploring the effectiveness of keystroke dynamics modeling in disparate conditions ranging from neurodegenerative processes such as Alzheimer's disease to cirrhotic sequelae such as hepatic encephalopathy.

The BiAffect keyboard has not only proven extremely adept at enabling digital phenotyping of its users' affective and cognitive states, but is also sensitive enough to their unique typing patterns that it can serve as an effective behavior-based biometric user identification and authentication tool. Sun et al. (2017) created DeepService, a multi-view multi-class deep learning method which is able to use data collected by the BiAffect keyboard to identify users with an accuracy rate of over 93% without using any cookies or account information. Until recently, the use of keystroke kinematics in hardware personal computer keyboards had been limited to similar continuous authentication applications, but physical keyboard sensing techniques are now expanding in scope to include identifying and measuring digital biomarkers as well (Samzelius 2016).

Mindful of the myriad potential concerns related to user privacy, data security and ethical implications inherent in the mass development and deployment of such applications, as well as in drawing conclusions based on findings generated using a

relatively small number of smartphone users from a handful of geographic regions (Lovatt and Holmes 2017; Martinez-Martin and Kreitmair 2018), and remaining particularly cognizant of the clinical imperative to only use those methods informed by established transtheoretical frameworks—the overarching lack of which may have led to the current replication crisis in psychology and the medical sciences (Muthukrishna and Henrich 2019)—the research teams investigating BiAffect data streams have endeavored to adopt a deliberately paced approach that harmonizes the latest developments in cognitive science, psychological theory, nosology, and treatment with state-of-the-art deep learning techniques and statistical methods. By paying close attention to safeguarding the individual privacy and protected health information of its users, and by adopting the most transparent possible model of sharing research techniques and findings in order to prioritize the use of digital phenotyping data for ethical medical applications, the BiAffect platform has been built on the twin paradigms of open source and open science as an invitation to collaborators from around the world to replicate, validate, amend or correct our hypotheses.

Perhaps one day we will all sport brain scanning ski caps that tell us how we feel, and install BCI implants to communicate wordlessly with our gadgets and with one another, while our IoT devices infer our emotions by analyzing our behavior at a distance; in the meantime, there is already no dearth of data streams readily available for passively mining users' mood, cognition, and much more with greater preservation of privacy and potential for predictiveness.

Acknowledgements We are grateful to the Robert Wood Johnson Foundation, the Prechter Bipolar Research Fund, Apple, Luminary Labs, and Sage Bionetworks, all of whom have helped enable much of the research discussed in this chapter. Jonathan P. Stange was supported by grant K23MH112769 from NIMH.

References

Ajilore O (2018) A voice-enabled diabetes self-management program that addresses mood—The DiaBetty experience. In: American Diabetes Association's 78th Scientific Sessions, Orlando, FL, USA

Ajilore O, Vizueta N, Walshaw P et al (2015) Connectome signatures of neurocognitive abnormalities in euthymic bipolar I disorder. J Psychiatr Res 68:37–44

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Publishing, Arlington, VA, USA

Anderson K, Burford O, Emmerton L (2016) Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS ONE 11(5):e0156164

Andreassen O, Houenou J, Duchesnay E et al (2018) 121. Biological insight from large-scale studies of bipolar disorder with multi-modal imaging and genomics. Biol Psychiatry 83(9):S49–S50

Asselbergs J, Ruwaard J, Ejdys M et al (2016) Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 18(3):e72

Avunjian N (2018) 'Westworld' cognition cowboy hats are a step up from a real science tool (inverse). USC Leonard Davis School of Gerontology. Retrieved from http://gero.usc.edu/2018/06/20/westworld-cognition-cowboy-hats-are-a-step-up-from-a-real-science-tool-inverse/

Balthazar P, Harri P, Prater A et al (2018) Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics. J Am College Radiol 15(3, Part B):580–586
Banks IM (2002) Look to windward. Simon and Schuster
Banks IM (2010) Surface detail. Orbit
Bourne C, Aydemir Ö, Balanzá-Martínez V et al (2013) Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. Acta Psychiatr Scand 128(3):149–162
Canhoto AI, Arp S (2017) Exploring the factors that support adoption and sustained use of health and fitness wearables. J Mark Manag 33(1–2):32–60
Cao B, Zheng L, Zhang C et al (2017) Deepmood: modeling mobile phone typing dynamics for mood detection. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 747–755
Cho K, van Merrienboer B, Gulcehre C et al (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. eprint arXiv:14061078:arXiv:1406.1078
Chung JE, Joo HR, Fan JL et al (2018) High-density, long-lasting, and multi-region electrophysiological recordings using polymer electrode arrays. bioRxiv:242693
Clifford C (2017) This former Google[X] exec is building a high-tech hat that she says will make telepathy possible in 8 years. This former Google[X] exec is building a high-tech hat that she says will make telepathy possible in 8 years. Retrieved from https://www.cnbc.com/2017/07/07/this-inventor-is-developing-technology-that-could-enable-telepathy.html
Cummings N, Schuller BW (2019) Advances in computational speech analysis for mobile sensing. In: Baumeister H, Montag C (eds) Mobile sensing and psychoinformatics. Springer, Berlin, pp x–x
Dixon-Román E (2016) Algo-Ritmo: more-than-human performative acts and the racializing assemblages of algorithmic architectures. cultural studies . Critical Methodol 16(5):482–490
Durstewitz D, Koppe G, Meyer-Lindenberg A (2019) Deep neural networks in psychiatry. Molecular Psychiatry
Ebner-Priemer UW, Eid M, Kleindienst N et al (2009) Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. J Abnorm Psychol 118(1):195–202
Ebner-Priemer UW, Trull TJ (2009) Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 21(4):463–475
Feng CH (2018) How a smartwatch literally saved this man's life and why he wants more people to wear one. South China Morning Post. Retrieved from https://www.scmp.com/lifestyle/health-wellness/article/2145681/how-apple-watch-literally-saved-mans-life-and-why-he-wants
Fu T-M, Hong G, Zhou T et al (2016) Stable long-term chronic brain mapping at the single-neuron level. Nat Methods 13:875
Gideon J, Provost EM, McInnis M (20–25 March 2016) (2016) Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 2359–2363
Global Burden of Disease Collaborative Network (2017) Global Burden of Disease Study 2016 (GBD 2016) Results. Institute for Health Metrics and Evaluation (IHME) Seattle, United States
Hou L, Bergen SE, Akula N et al (2016) Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. Hum Mol Genet 25(15):3383–3394
Huang H, Cao B, Yu PS et al (2018) dpMood: exploiting local and periodic typing dynamics for personalized mood prediction. In: Paper presented at the IEEE international conference on data mining
Ikeda M, Takahashi A, Kamatani Y et al (2017) A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. Mol Psychiatry 23:639
Jepsen ML, Open Water Internet Inc (2017) Optical imaging of diffuse medium. U.S. Patent No. 9,730,649

Karam ZN, Provost EM, Singh S et al (4–9 May 2014) (2014) Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 4858–4862
Khorram S, Gideon J, McInnis MG et al (2016) Recognition of depression in bipolar disorder: leveraging cohort and person-specific knowledge. In: INTERSPEECH
Khorram S, Jaiswal M, Gideon J et al (2018) The PRIORI emotion dataset: linking mood to emotion detected in-the-wild. ArXiv e-prints
Kubiak T, Smyth JM (2019) Connecting domains—ecological momentary assessment in a mobile sensing framework. In: Baumeister H, Montag C (eds) Mobile sensing and psychoinformatics. Springer, Berlin, pp x–x
Leow A, Ajilore O, Zhan L et al (2013) Impaired inter-hemispheric integration in bipolar disorder revealed with brain network analyses. Biol Psychiat 73(2):183–193
Lovatt M, Holmes J (2017) Digital phenotyping and sociological perspectives in a Brave New World. Addiction (abingdon, England) 112(7):1286–1289
Martinez-Martin N, Kreitmair K (2018) Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Mental Health 5(2):e32–e32
McInnis M, Gideon J, Mower Provost E (2017) Digital phenotyping in bipolar disorder. Eur Neuropsychopharmacol 27:S440
Messner E-M, Probst T, O´Rourke T et al (2019) mHealth applications: potentials, limitations, current quality and future directions. In: Baumeister H, Montag C (eds) Mobile sensing and psychoinformatics. Springer, Berlin, pp x–x
Montag C, Markowetz A, Blaszkiewicz K et al (2017) Facebook usage on smartphones and gray matter volume of the nucleus accumbens. Behav Brain Res 329:221–228
Muthukrishna M, Henrich J (2019) A problem in theory. Nat Human Behav
National Collaborating Centre for Mental Health (2018) Bipolar disorder: the NICE guideline on the assessment and management of bipolar disorder in adults, children and young people in primary and secondary care. In: British psychological society, pp 39–40
Perlow J (2018) How Apple Watch saved my life. ZDNet. Retrieved from https://www.zdnet.com/article/how-apple-watch-saved-my-life/
Phillips ML, Kupfer DJ (2013) Bipolar disorder diagnosis: challenges and future directions. Lancet 381(9878):1663–1671
Phillips ML, Ladouceur CD, Drevets WC (2008) A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. Mol Psychiatry 13:833
Rabbi M, Klasnja P, Choudhury T et al (2019) Optimizing mHealth interventions with a bandit. In: Baumeister H, Montag C (eds) Mobile sensing and psychoinformatics. Springer, Berlin, pp x–x
Ross MK, Demos AP, Zulueta J et al (2021) Naturalistic smartphone keyboard typing reflects processing speed and executive function. Brain Behav 11(11):e2363
Samzelius J, Neurametrix Inc (2016) System and method for continuous monitoring of central nervous system diseases. U.S. Patent No. 15,166,064
Sanford K (2018) Will this "Neural Lace" brain implant help us compete with AI? Retrieved from http://nautil.us/blog/-will-this-neural-lace-brain-implant-help-us-compete-with-ai
Sariyska R, Rathner E-M, Baumeister H et al (2018) Feasibility of linking molecular genetic markers to real-world social network size tracked on smartphones. Front Neurosci 12(945)
Shropshire C (2015) Americans prefer texting to talking, report says. Chicago Tribune. Retrieved from http://www.chicagotribune.com/business/ct-americans-texting-00327-biz-20150326-story.html
Stange JP, Zulueta J, Langenecker SA et al (2018) Let your fingers do the talking: passive typing instability predicts future mood outcomes. Bipolar Disord 20(3):285–288
Steel Z, Marnane C, Iranpour C et al (2014) The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013 43(2):476–493

Sun L, Wang Y, Cao B et al (2017) Sequential keystroke behavioral biometrics for mobile user identification via multi-view deep learning. In: Paper presented at the joint European conference on machine learning and knowledge discovery in databases, November 01, 2017

Turakhia MP (2018) Moving from big data to deep learning—the case of atrial fibrillation. JAMA Cardiol 3(5):371–372

Turakhia MP, Desai M, Hedlin H et al (2019) Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: the Apple heart study. Am Heart J 207:66–75

Vesel C, Rashidisabet H, Zulueta J et al (2020) Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: a BiAffect iOS study. J Am Med Inform Assoc 27(7):1007–1018

Wolkenstein L, Bruchmuller K, Schmid P et al (2011) Misdiagnosing bipolar disorder—do clinicians show heuristic biases? J Affect Disorders 130(3):405–412

Zulueta J, Demos AP, Vesel C et al (2021) The effects of bipolar disorder risk on a mobile phone keystroke dynamics based biomarker of brain age. Front Psychiatry 12(2284)

Zulueta J, Piscitello A, Rasic M et al (2018) Predicting mood disturbance severity with mobile phone keystroke metadata: a BiAffect digital phenotyping study. J Med Internet Res 20(7):e241