Current Practices in Mental Health Sensing

The ubiquity of smartphones and wearables makes it an attractive option to passively study human behavior. We explore the current practices of using passive sensing devices to assess mental health and wellbeing, including the limitations and future directions.

By Subigya Nepal, Weichen Wang, Bishal Sharma, and Prabesh Paudel

DOI: 10.1145/3481829

The rapid advancement of ubiquitous technologies is creating unprecedented opportunities to continuously assess human behavior. Devices such as smartphones, wearables, and fitness bands are equipped with a variety of embedded sensors, such as inertial and physiological sensors, offering the capability to passively obtain real-time, in-situ behavioral data. Such continuous monitoring of human behavior at a contextual and extremely granular level opens the way to study and make inferences about an individual's mental health and wellbeing in a more precise, objective, and personalized manner. In this article, we discuss current practices in mental health sensing and future directions.

MENTAL HEALTH AND PASSIVE SENSING

Over the last few years, research in mental health sensing has accelerated. While many works utilize passive forms of sensing, such as smartphones and wearables, there are also studies done with what people might consider to be a more intrusive form: videos, photos, and gaze trackers [1]. In this article, we mostly focus on passive forms of mental health sensing. Passive sensors are especially useful because they blend into the day-to-day life of users and unobtrusively collect objective data. In this regard, wearable devices have risen to the top as the go-to method for passively sensing mental health and wellbeing. While smartphones capture a range of objective data about human behavior (movement, activity, and sociability), wearables offer something truly valuable—physiological data. This includes heart rate, heart rate variability, skin temperature, respiration rate, and sleep among others. Wearable devices also consist of embedded sensors capturing similar behavior that smartphones do, and many times with more precision (e.g., sleep and steps). Some other forms of passive sensing that have been frequently utilized include Bluetooth trackers, environment or ambient sensors including temperature sensors, and smart shirts. The architecture of the passive sensing system is shown in Figure 1.

Mental health sensing technology is used in many studies to provide novel

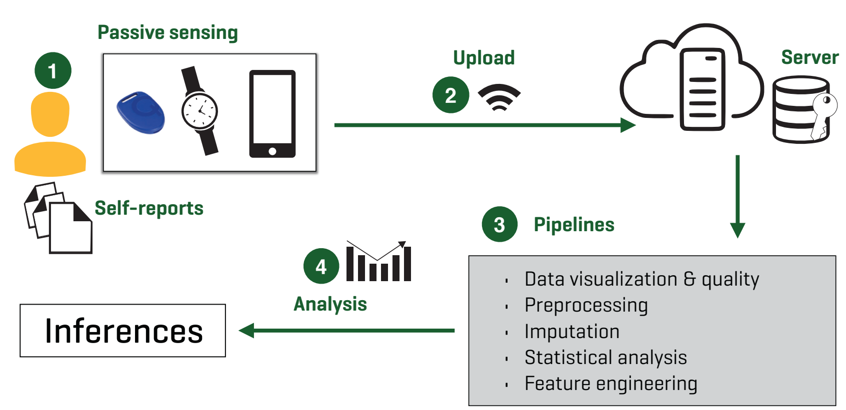
insights into depression, bipolar disorder, psychosis, anxiety, stress, and self-esteem. Self-reports are also used to obtain a quantitative assessment of individuals' mental health, although a few studies employ more objective approaches—salivary cortisol being one example. Cortisol is associated with the stress response systems of human biology. Cortisol secretion induces an increase in glucose level, blood pressure, and immune system suppression, all of which indicates an increased stress response in the body. Cortisol levels can be detected from blood serum as well as from saliva. Because it is non-invasive, salivary cortisol is frequently used as a biomarker of stress reaction.

In long-term longitudinal studies, participants are prompted with questions asking for responses to relevant self-reports throughout different periods of the day: morning, afternoon, and evening. This is usually termed as experience sampling, and in mental health sensing literature it is more commonly referred to as Ecological Momentary Assessment (EMA). This allows for a more fine-grained study of state anxiety, stress, and mood, enabling researchers to capture variation within the day as opposed to comparing behavior over weeks. Depression is studied both in the short- and long-term, mostly using the Patient Health Questionnaire (PHQ) or a modified shorter version of it (e.g., PHQ-8) [2]. The PHQ is a valid instrument for screening, diagnosing, monitoring, and measuring the severity of depression. There are several versions of PHQ such as PHQ-9, PHQ-8, and PHQ-4. The digit at the end indicates the number of items in the survey, for instance, PHQ-8 eliminates the ninth item on PHQ-9, which is used to assess suicide risk. Studies use different versions of PHQ based on the population under study, the frequency of the self-reports asked, and the major focus of the study. In terms of study population, different segments of populations that are on different spectrums with respect to symptom severity are studied. For example, that might mean the population under study might include those with clinical depression, e.g., veterans, or young adults, e.g., college students.

MOBILE SENSING PLATFORMS

Even though wearables are now an important part of mental health sensing, they are incomplete without a mobile sensing application in place that they can communicate with. Such smartphone-based applications are primarily developed for iOS and Android systems. The data the apps can collect depends on the platform they are built for. Android offers access to a wider range of data than iOS does. Android offers call and SMS logs, light amplitude, application usage log, opportunistic photo capture, and more. iOS is extremely strict on what an application can collect, not offering any of these data previously mentioned.

Figure 1. Overview of a passive sensing system. Data is continuously collected through different passive sensors, along with the ground truth in the form of self-reports. Different pipelines run on the server to visualize, pre-process, and maintain data quality. Inferences are made using the collected data at the end.



This limits what can be done with iOS-based sensing. In studies that use iOS sensing apps, it is common for researchers to ask participants to manually report some of their data. For example, researchers might request participants to report their application usage as a screenshot of the "Screen Time" details offered by iOS. With Android, such application usage logs can be automatically collected by the sensing app. Aside from the limitation in the data that can be collected, iOS sensing apps are also dependent on regular connection with an external server. For instance, it is not possible to schedule a notification locally for a later date if the app is not awake. As a result, we need to have a constant poll from the server that wakes the app (in the form of a push notification)—the app checks whether prior EMA was answered and schedules it for a later date in case it was not. Similar is the scenario with identifying phone usage; we must send constant push notification to tell whether the phone is locked or in use. The need to have a constant connection to the internet and a web server increases the number of movable parts to consider, reducing the effectiveness or usefulness of mobile sensing. Therefore, iOS is more limited in comparison to Android for mental health sensing. Common to both systems, researchers are always in an arms race of sorts to release an updated mobile sensing application. That is, with every new version of the operating system, developers must push new updates to deal with changes without losing any data in an ongoing study. This can be very challenging to coordinate in large-scale longitudinal studies. In addition, there may be limitations of the age of the device or the version of the operating system that serve as criteria for recruiting people into studies.

MODELING THE BEHAVIORAL DATA

With respect to modeling, most of the studies primarily explore detection of mental health issues as the main research problem. The analysis ranges anywhere from simple statistical association analysis with behavioral inferences to the use of different machine learning models to predict the score obtained from EMAs [3]. Specific to prediction, we see a greater number of studies that address the problem as a classification task rather than a regression-based analysis. This is mostly evident in proof-of-concept studies that are novel and may not have as much predictive capability to perform well in a regression task. Such studies usually employ threshold-based splitting to convert the scores obtained on a continuous survey scale to classes and then predict which class the participant falls in as per their score on the scale. Some surveys, such as PHQ-9, have recommended cut-off points to divide individuals into several groups based on symptom severity. For other studies that do not have a recommended cut-off point, usually, median or mean splitting is used to convert the regression task into a classification task. This might lead to there being either two classes (high or low in binary classification) or high, medium, or low class (multi-class classification). For example, in the case of PHQ-9, the high or low class are representative of two groups: one with a high depression score and the other with a low depression score. Transforming a regression task to a classification task is a way to decrease the complexity of the problem. Such transformations are even more valid and necessary for in-the-wild studies that have a lot of missing data and skewed survey responses.

Traditional machine learning approaches are still widely used for predictive modeling. Researchers generate several handcrafted features, such as circadian rhythms and routineness, based on early works that showed the relationship between such features and mental health. Use of a traditional machine learning approach also means that in most cases, averaging the data is commonplace—each feature obtained through passive sensing data is averaged out to a single representative value for every single day. Sometimes, the average of each feature is generated not only for the entire day, but also for specific periods of the day—for instance, average value of heart rate in the morning, during the afternoon, and during the night. By doing so, researchers hope to capture variations in the data throughout the day depending on whether it is participants' office time, family time, or resting period. However, this is slowly getting out of practice with the increasing use of deep learning in the field. Use of raw data directly with a deep model, or at least more granular data (i.e., hourly), is gaining momentum. By leveraging raw or extremely granular data, we may be able to perform more time sensitive predictions, rather than doing it over days or weeks. Also, aside from the traditional approaches, some researchers treat mobile sensing data as time series and utilize multivariate time series-based modeling techniques for prediction and analysis. Additionally, it is worth noting personalized modeling is also pursued within the field.

The field could benefit if both iOS and Android OS allowed users to provide access to their data for certain apps that are verified for research purposes.

CLOSING THE LOOP, INTERVENTIONS

There are not as many studies doing mental health sensing that explore interventions as well as explore the detection of it. By and large, there are more studies that focus on detecting mental health scores of participants. Generally, feedback is delivered to the participants who use smartphones in three ways: summary statistics, with the help of visual aids, and by issuing instructions or direct recommendations to the participants [4]. The way feedback is delivered varies across studies: either through the smartphone, in-app, or in-person. Bewell is an early example of one such sensing app that delivers feedback through the smartphone—by changing the wallpaper to notify the user about their social interaction, sleep, and physical activity [5]. We need further research into closing the loop and offering interven-

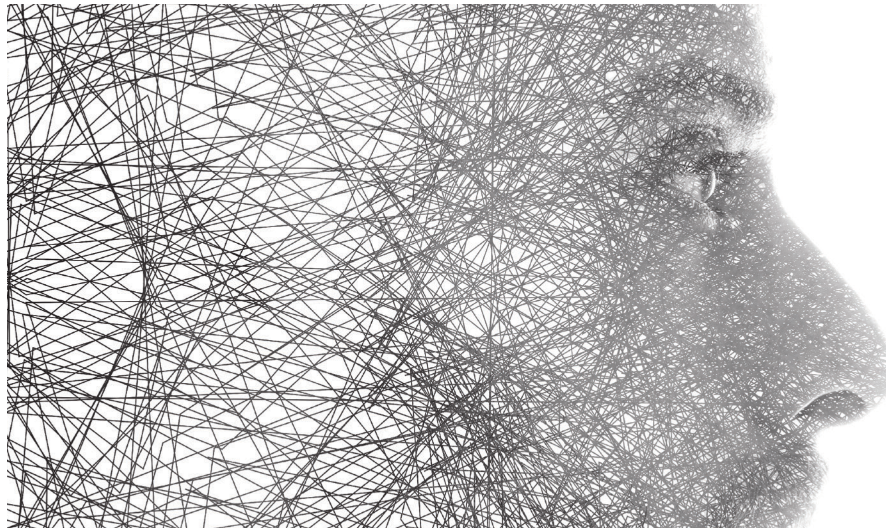


Image by LUMEZIA / Shutterstock.com

tions to users based on outcomes. As it stands now, most of the interventions that are currently utilized are a mere transition of traditional solutions to smartphones, relying heavily on the user's willingness to take those actions (e.g., cognitive behavioral therapy, CBT). We need more research into intervention approaches that exploit the smartphone's capability, maybe also delivering feedback in alternative means, such as using augmented reality or virtual reality. Personalized interventions that persuade users are especially important. We believe mobile apps offer the capability to offer real-time and in-situ psychological interventions that can be exploited fully only through multidisciplinary collaborations.

ETHICAL AND PRIVACY MEASURES

The current practices of maintaining privacy in data collection, involves separating personally identifiable information (PII) from participant's data. Participants are usually assigned two different numbers: one for identifying their sensing data and one for their demographics. The mapping for these is strictly protected so the enrollment team cannot identify the sensing data of a particular person, and the team involved in building the app and managing the data cannot know the identity of the person that the sensing data maps to. This makes it certain that the data of a participant cannot be tied back to their personal information they provided during enrollment. The collected data is kept in-house on local servers or in the cloud following encryption and security best-practices. Along with GPS data, Bluetooth, and Wi-Fi data can be used to identify location information of the participants. These are therefore restricted information and are not readily available to all the researchers in the team. They are kept in encrypted form and only accessible to team members upon valid and legitimate request. These are extremely important criteria that need to be put in place to receive Institutional Review Board approval for studies.

By leveraging raw or extremely granular data, we may be able to perform more time sensitive predictions, rather than doing it over days or weeks.

FUTURE DIRECTIONS

The rise of wearable technology and increasing applications of deep learning in the field point to a promising future of mental health sensing. The growth in the use of wearable devices for mental health sensing can in large part be attributed to the surge in demand of smart devices in healthcare and a boom in consumer fitness trends, which is directing more and more industry research into wearable technology. As more research and development is poured into wearable devices, new iterations of wearables are introducing newer forms of embedded sensors to measure physiology, offering increased battery life, and presenting a better way for users to interface with them. Such developments are increasing the possibility of what could be done in the future for mental health sensing.

Similarly, further research into the applications of deep learning in time-series modeling and with mobile sensing data is best placed to exacerbate more situated research in the field including just-in-time interventions, more accurate personalized models, and on-device private computations. The privacy of user data is going to be an important aspect of future research in the field. Regarding the increasing difficulty with collecting sensing data from mobile apps, there is a need for industry leaders to appreciate the impact of mobile phone sensing in mental health and other avenues. The field could benefit if both iOS and Android OS allowed users to provide access to their data for certain apps that are verified for research purposes. To this regard, Apple has introduced its own research app that allows studies to passively collect data which is otherwise unavailable to third-party apps.

With the increasing use of deep learning, there is a need to provide interpretability of sensing results for clinicians to interpret rather than black-box inferences, for example. In addition, researchers studying mental health through passive means should also try to utilize as many forms of sensors as possible, as long as it does not add on to the participant's burden by a lot. This would allow for a more holistic understanding of their behavior. Perhaps opportunistic sensing could be studied more as an approach allowing for holistic understanding while at the same time reducing burden on the participants. While most research focuses on directly predicting the outcomes, further investigation in analyzing or exploring the behavioral indicators of depression, anxiety, and

Monitoring of human behavior at a contextual and extremely granular level opens the way to study and make inferences about an individual's mental health.

other mental health related issues would offer a better understanding of them. An important direction worth considering for researchers is the repeatability and generalizability of the findings, which is missing right now. Currently, most studies are small and for short durations and focus on specific populations. There is a need to also repeat studies to see if findings are generalizable. Understanding how replicable the findings are along with their generalizability would be crucial in transforming the research into clinical practice.

Currently, there are not any generally acceptable behaviors associated with mental health topics even if those topics have been studied widely using passive sensing. This is because most studies use their own app to collect data and they report metrics independently, which makes it difficult to generalize the results or findings. Therefore, we either need an ImageNet equivalent dataset for mental health sensing or to agree to a certain set of standards to follow for studies (for instance, considering the sampling rates, the kind of features to collect or generate, the self-reports to use, or even just the availability of the data-set to public) so that they can be replicable and studied for generalizability. Because there is a complete lack of replication of study findings there is little confidence in the generalizability of study results so far—and, therefore, a lack of validity of inferences (e.g., being able to classify depressive symptoms). We need benchmarks here. When this is resolved and we can rely on, say, classification of time-varying mental health states, then we can drive just-in-time interventions to keep people healthy. This might start with behavioral inferences triggering CBT technology in the first instance, but ultimately developing new forms of personalized intervention beyond CBT not known today. Connecting a suite of validated machine learning pipelines for depression, bipolar disorder, psychosis, and anxiety to personalized intervention modules in an effective manner is the key challenge of our field.

All in all, we believe mental health sensing as a field is bound to benefit from the ongoing progress in all fronts, with the potential to increase positive outcomes for people and improve our understanding of mental health at the same time. However, much work is needed to make mental health sensing ubiquitous.

References

[1] Enrique, G. et al. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing* 51 (2018), 1–26.

[2] Kroenke, K. et al. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* 16, 9 (2001), 606–613.

[3] Seppälä, J. et al. Mobile Phone and wearable sensor-based mhealth approaches for psychiatric disorders and symptoms: Systematic review. *JMIR Mental Health* 6, 2 (2019).

[4] Aung, M. H. et al. Sensing behavioral symptoms of mental health and delivering personalized interventions using mobile technologies. *Depression and Anxiety* 34, 7 (2017) 603–609.

[5] Lane, N. D. et al. BeWell: Sensing sleep, physical activities and social interactions to promote wellbeing. *Mob. Netw. Appl.* 19, 3 (2014), 345–359.

Biographies

Subigya Nepal is a third-year Ph.D. student in the Department of Computer Science at Dartmouth College. He is a member of the Dartmouth Networking and Ubiquitous Systems (DartNets) lab, where he researches passive sensing and its applications in understanding and modeling human behavior.

Weichen Wang is a fifth-year Ph.D. student in the Department of Computer Science at Dartmouth College. He is also a member of the DartNets lab. His research focuses on ubiquitous computing, mobile sensing systems, mobile sensing applied machine learning and mobile health.

Bishal Sharma is a second-year Ph.D. student at Thayer School of Engineering, Dartmouth College. He is a member of the metabolic engineering group at Lynd Lab. His research interests are metabolic engineering, bioinformatics, machine learning and computational biology.

Prabesh Paudel is a senior undergraduate student at St. Olaf College in Northfield, MN majoring in mathematics and computer science. His research interests range from statistical analysis to using deep learning in sequential data.