The *Interactions* website (interactions.acm.org) hosts a stable of bloggers who share insights and observations on HCI, often challenging current practices. Each issue we'll publish selected posts from some of the leading and emerging voices in the field.

Machine Learning Applications

Reflections on Mental Health Assessment and Ethics

Anja Thieme, Microsoft Research Cambridge, **Danielle Belgrave,** Microsoft Research Cambridge, **Akane Sano,** Rice University, **Gavin Doherty,** Trinity College Dublin,

As part of the ACII 2019 conference in Cambridge, U.K., we ran a workshop on "Machine Learning for Affective Disorders" (ML4AD; http://mlformentalhealth.com/). The well-attended workshop had an extensive program, including an opening keynote by UC Irvine assistant professor of psychological science Stephen Schueller, presentations by authors of accepted workshop papers, and invited talks by established researchers in the field (http://mlformentalhealth.com/#speakers). Among the topics and application areas covered were: detection of depression from body movements; online suicide-risk prediction on Reddit; various approaches to assist stress recognition; a study of an impulse suppression task to help detect people suffering from ADHD; and strategies for generating better "well-being features" for end-to-end prediction of future well-being.

Discussions at the workshop touched on many common ML challenges regarding data processing, feature extraction, and the need for interpretable systems. Most conversations, however, centered on: 1) difficulties surrounding mental health assessment, and 2) ethical issues when developing or deploying ML applications. Here, we want to share a synthesis of these conversations and current questions that were raised by researchers working in this area.

MENTAL HEALTH ASSESSMENT

Workshop attendees described a range of assessment challenges, including: data labeling and establishing ground truth, definitions of mental health targets, and what measures were considered safe to administer to study participants or people who are perhaps self-managing their condition in everyday life. Two areas of debate received particular attention:

What healthcare need(s) to target and how to conceptualize mental health states or symptoms. In the types of ML tools or applications that are being developed, we noticed a predominant focus on the detection and diagnosis of mental health symptoms or states. This may partly be explained by the availability of data and clinically validated tools in this space, which inform how research targets are shaped. Currently, much of the existing ML work tries to match the data that is available about a person to a diagnosis category (e.g., depression). Here, attendees mentioned concerns that looking at a mental illness, like depression, as one broad category may not take into account the variability of depression symptoms and how the illness manifests, and could mean building models for monitoring depression that are less useful as a result. Further, they raised the question of whether mapping a person to a "relevant treatment" might present a more important ML task than diagnosis.

Related to this discussion, attendees raised some other key questions:

People may not want to be screened or diagnosed with a psychiatric condition due to the associated stigma.

- What is the health/medical problem that we are trying to address? Are we asking the right questions?
- How do we ensure that the (often complex) models and solutions we develop in computing science really meet a clinical need? What are the "right" use cases for ML?
- How can we define/select/develop good quality measures?

The value of objective versus subjective assessments. Do they need to compete with each other? Excitement about passively and continuously captured data about people's behaviors through sensors or content created online has shaped perceptions of ML approaches as providing "more objective" insights; especially when compared with other "more subjective" methods such as self-reports. It was pointed out that we cannot strictly define what is subjective or objective. Thus, instead of looking at these approaches in competition, perhaps a more promising route would be to look at interesting relations that surface through the combination of different data methods, and what each may say about the person. Rather than looking at ML insights, clinical expertise, and traditional health-assessment tools in competition, how can they complement each other? This leads us to ask how ML outputs can serve as a useful information resource to assist, and help empower, clinicians. When discussing examples such as mobile phone–based schizophrenia monitoring, it was apparent that providing clinicians with a wealth of automatically collected patient data was likely to be overwhelming and of little use unless the data was presented in ways that provided meaningful insights to clinicians and effectively complemented their work practices.

Workshop participants discussing assessment challenges.

Thus, key questions included:

• How can we empower clinicians through data tools?

• How can we help clinicians to appropriately trust data and related generated insights?

• How can the results of ML help make concrete actions/interventions for clinicians/patients?

ETHICAL CHALLENGES

Inevitably, when discussing the role of ML and possibilities of ML-enabled interventions for use as part of real-world mental health services, our conversations turned to ethical issues, specifically the following two themes:

(How) should we communicate ML-detected/diagnosed mental health disorders or risks? A key conversation topic was: if and how we should communicate to people that an ML application has diagnosed them with a mental health disorder or detected a risk. This was particularly a concern in contexts where people are perhaps unaware of a mental health problem and the processing of their data (e.g., from social media) for diagnostic purposes. On the one hand, being able to detect problems early can help raise awareness, validate the person's experience, encourage help seeking, and allow for better management of a condition. On the other hand, people may not want to be screened or diagnosed with a psychiatric condition due to the associated stigma and its implications on their personal or work life. For example, a diagnosis of a mental disorder can have severe consequences for professionals in the police force or firefighting. Thus, how do we balance both people's "right to be left alone" and their "right to be helped"?

Related questions were:

• How do we sensibly communicate the detection/diagnosis of a mental health problem or disorder?

• Should only passive data be collected and used for self-reflection and self-care of the person?

• How do we show risk factors to people in ways that are actionable (e.g., a diagnosis alone may not be helpful unless the person knows what they can do about it)?

• What kinds of interventions should not be developed or tested with people in the wild?

What are the broader implications of ML interventions, and how can we reduce the risks of misuse? It is hard to predict what unanticipated consequences a new ML intervention might have on a person, their life, or society at large. Partly this is due to the way in which we tend to study well-defined problems whose solutions may not transfer to other contexts outside of those for which they've been designed or trained. For example, in the context of developing an emotion recognizer based on a person's facial expressions, we discussed what the implications might be if someone was repurposing this technology, for example, to identify children who are not working enough at school or employees who appear less productive at work. Additional ethical concerns included: difficulties in preventing the (mis)use of developed tools with low clinical accuracy in clinical practice, and challenges related to user consent and data control.

Key questions included:

• How do we responsibly design and develop ML systems?

• How can we help reduce the risk of misuse for the technologies we develop?

• How do we rethink consent processes and support user control over their data?

We thank all organizers, keynote and invited speakers, paper authors, and attendees for their invaluable contributions to the workshop.

Anja Thieme is a senior researcher in the Healthcare Intelligence group at Microsoft Research, designing and studying mental health technologies.
→ anthie@microsoft.com

Danielle Belgrave is a principal researcher in the Healthcare Intelligence group at Microsoft Research. Her research focuses on ML for healthcare.
→ danielle.belgrave@microsoft.com

Akane Sano is an assistant professor at Rice University, developing technologies for detecting, predicting, and supporting mental health.
→ akane.sano@rice.edu

Gavin Doherty is an associate professor at Trinity College Dublin and co-founder of SilverCloud Health, developing engaging and effective mental health technology.
→ gavin.doherty@tcd.ie

DOI: 10.1145/3381342