Digital health technology data in biocomputing: Research efforts and considerations for expanding access (PSB2024)

Michelle Holko
International Computer Science Institute at Berkeley
Bethesda, MD 20817, USA
Email: michelle.holko@gmail.com

Chris Lunt
National Institutes of Health
Bethesda, MD 20892, USA
Email: chris.lunt@nih.gov

Jessilyn Dunn
Biomedical Engineering, Duke University
Durham, NC 27708, USA
Email: jessilyn.dunn@duke.edu

Data from digital health technologies (DHT), including wearable sensors like Apple Watch, Whoop, Oura Ring, and Fitbit, are increasingly being used in biomedical research. Research and development of DHT-related devices, platforms, and applications is happening rapidly and with significant private-sector involvement with new biotech companies and large tech companies (e.g. Google, Apple, Amazon, Uber) investing heavily in technologies to improve human health. Many academic institutions are building capabilities related to DHT research, often in cross-sector collaboration with technology companies and other organizations with the goal of generating clinically meaningful evidence to improve patient care, to identify users at an earlier stage of disease presentation, and to support health preservation and disease prevention. Large research consortia, cross-sector partnerships, and individual research labs are all represented in the current corpus of published studies. Some of the large research studies, like NIH's *All of Us* Research Program, make data sets from wearable sensors available to the research community, while the vast majority of data from wearable sensors and other DHTs are held by private sector organizations and are not readily available to the research community. As data are unlocked from the private sector and made available to the academic research community, there is an opportunity to develop innovative analytics and methods through expanded access. This is the second year for this Session which solicited research results leveraging digital health technologies, including wearable sensor data, describing novel analytical methods, and issues related to diversity, equity, inclusion (DEI) of the research, data, and the community of researchers working in this area. We particularly encouraged submissions describing opportunities for expanding and democratizing academic research using data from wearable sensors and related digital health technologies.

Keywords: digital health technologies; wearables; sensors; waveform data; time-series data; algorithms.

© 2023 The Authors. Open Access chapter published by World Scientific Publishing Company and distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License.

1. Background

Wearable devices and other digital health technologies (DHTs), such as smartwatches, fitness trackers, and smart rings, are becoming increasingly popular for tracking and monitoring a wide range of health and fitness metrics [1]. Figure 1 below reflects the growth of scientific publications with the word “wearable” with 5,713 papers published in 2022. A similar pattern is seen when searching for “digital health technology.” These devices can collect data on everything from heart rate and sleep patterns to activity levels and blood oxygen levels. In recent years, there has been a growing interest in using wearable devices and DHTs for health research. Wearable devices offer a number of advantages over traditional research methods, such as questionnaires and surveys. For example, wearable devices can collect data continuously and over long periods of time, providing researchers with a more complete picture of an individual's health and well-being. Additionally, wearable devices can be used to collect data in real-world settings, rather than in laboratory environments, which can provide more insights into how people behave in their everyday lives.

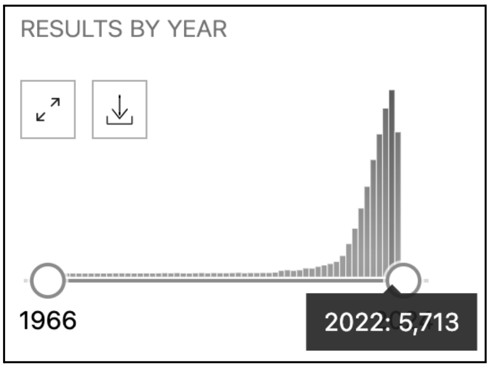

Fig. 1. Number of publications with “wearable” in PubMed from 1966-2023, highlighting exponential growth of this subject and 5,713 papers published in 2022.

Some specific examples of disease areas with active DHT and wearables research include:

- **Cardiovascular disease:** Apple watch devices have been used to study the relationship between heart rate and physical activity levels, and to develop algorithms to predict the risk of heart disease [2].
- **Respiratory disease:** Fitbit devices have been used to study the effects of different types of air pollution on lung function, and to develop algorithms to detect early signs of asthma attacks [3].

- **Metabolic disease:** Wearable devices tracking blood glucose and activity have been used to study gestational diabetes, and to develop algorithms to predict the risk of developing diabetes during pregnancy [4].
- **Mental health:** Wearable devices have been used to study the relationship between physical activity levels and mood, and to develop algorithms to predict symptom trajectory for bipolar disorder [5].

In addition to these specific examples, wearable devices and DHTs are also being used to study a wide range of other health conditions, such as cancer, infectious diseases, and chronic pain. But there are gaps in who has access to data and devices, who is performing the research and algorithm development, and therefore who the new technologies are poised to help improve health outcomes. Reviews of the current landscape of DHT research studies in the National Center for Biotechnology Information (NCBI)'s Clinical Trials database (clinicaltrials.gov), and of studies published by the top-20 funded private sector DHT companies, highlight several patterns and limitations:

1. **Small sample size:** Aside from a few large studies, most of the published clinical trials utilizing DHT have been relatively small, and are largely under-powered. "Nearly half the studies - 829, or 46.5% - had less than 100 enrollees. Only 8% had more than 1,000 [6]."
2. **Narrow Health Focus:** The majority of published DHT studies focus on cardiometabolic health and mental health/wellness, while relatively little published research examines critical healthcare burden diseases like stroke, chronic obstructive pulmonary disease (COPD), and diabetes [7].
3. **Narrow Population Focus:** Of studies published by the top 20 funded DHT private-sector companies, the majority (72%) include only healthy volunteers, rather than high-risk populations with comorbid conditions [8]. The breadth and diversity of the study population(s), including socioeconomic, healthcare status, and racial diversity, may be the most critical component of building AI-based DHT algorithms. This diversity is lacking in current published research, likely leading to biased results [9]. The "bring your own device" model has been used by many research studies, but this design may result in biased selection of participants, and therefore biased results [10].
4. **Limited Outcome Assessments:** Only 15% of published DHT studies measured clinical effectiveness, and only in relation to the patient outcomes and did not evaluate healthcare cost or access to care [11]. As healthcare cost and access are two of the most pressing needs in healthcare, it is important to expand research to examine these outcomes.
5. **Insufficient Reporting and Data Publishing:** Importantly, not only is reporting in clinicaltrials.gov not required for observational DHT trials, there is also no public database

for DHT data and algorithms. This complicates the ability to understand the full range of DHT "real world evidence" (RWE)-based research, and undermines research reproducibility and validation. The lack of a consensus DHT database also means that DHT data curation, feature (e.g., digital biomarker) discovery, and algorithm development is limited to those who have data, which is largely the private sector DHT companies. One attempt to develop standardized pipelines and data repositories for digital health data, the Digital Health Data Repository as part of the Digital Biomarker Discovery Pipeline [12], developed by co-organizer Jessilyn Dunn's lab, is still not fully funded.

6. **Bridging the Regulatory Gap and Moving to Clinical Implementation:** Despite tremendous progress in DHT research and development, there is still a lot of work to be done along the research → regulatory → clinical implementation continuum. The *All of Us* Research Program is uniquely situated within NIH to interact with FDA colleagues and assist in developing regulatory standards for this new and uncharted territory. The FDA also has a Center for Digital Health Excellence, and there is a Digital Health Consortium, housed within the Office of the National Coordinator, for senior leaders within the federal government to convene across the digital health continuum. The Digital Medicine Society is a professional organization that has been working across sectors with the community to support innovation and standardization, in part via the Digital Health Measurement Collaborative Community (DATAcc) [13] and the Digital Health Playbook [14]. For clinical implementation, HumanFirst has built the Atlas precision measures platform, a cloud-based platform with endpoints and measures being researched using DHTs across the industry to help pharma and clinicians decide on which devices and how they can be used in clinical research and healthcare [15].

The above limitations don't begin to address potential bias in algorithm development due to a limited pool of researchers interacting with these data. The purpose of this Session is to provide a forum for current research, address issues related to Diversity, Equity and Inclusion (DEI) in terms of the types of research and the researchers engaged, and ultimately to energize non-commercial research in the area. Our motivating question is how can this community work together to create more equitable research in the digital health tech space to benefit the research community and resulting impact?

2. Relevance to biocomputing

Computational biology approaches and algorithm development are critical enablers to the use of wearable devices and DHTs for biomedical research and health. Computational biologists are developing new methods for extracting meaningful insights from the large and complex datasets collected by these devices; algorithm developers are developing new algorithms to improve the accuracy and reliability of wearable devices and DHTs. The continuous or near-continuous data streams from DHTs are ripe for machine learning and artificial intelligence (ML/AI) research. Algorithms developed for detecting anomalies and other biomedically-related phenomena in wearable sensor data are increasingly being incorporated into research and moving into clinical practice and other health adjacent applications.

Despite the many advantages of using wearable devices and DHTs for health research, there are also a number of challenges that need to be addressed. One challenge is that the data collected by these devices can be noisy and complex with significant levels of missing data, making it difficult to extract meaningful insights. Another challenge is that the algorithms used to analyze this data need to be carefully validated to ensure that they are accurate and reliable. There are also many different devices, and the community doesn't yet have robust standards to compare between and among signals and data from different devices.

In this session, we bridge these gaps across sectors and domains to identify opportunities for researchers in the PSB community to contribute to the growing biomedical research leveraging wearables and DHT to understand and improve health. In prior years of PSB, there has been good representation of a variety of data types, including genomics, imaging and clinical data sets; there has been limited coverage of wearable sensors and digital health technologies research. Last year, PSB2023, we hosted the first year of this Session [16]. We wanted to continue to support this conversation and topic area as this field continues to grow and obstacles to academic research continue to need to be overcome. Many of the other conferences where DHT computational researchers are more focused on the clinical aspects and clinical trials, and not as much on the computational biology or biomedical research aspects of DHT data analysis and algorithm development.

The goal of this information sharing and discussion opportunity for participants and the community is to expand awareness and access to these data and tools, to enrich computational biology research, and bridge DEI gaps. The session includes a range of voices from academia, government, and private sector. It's important to represent private sector voices in this discussion since much of the research is currently happening in tech companies developing digital health devices. Creating a forum for dialogue across sectors is important for bridging gaps in awareness and understanding, and encouraging more researchers to participate in developing computational methods and analysis of data from digital health tech. The papers and discussion will focus on key challenges facing the field, and participants are encouraged to contribute ideas to potential solutions and initiate lasting collaborations with researchers and communities in this area. The session will also provide an opportunity to discuss as a community what is needed to truly enable cross-sector and expanded research for digital health technologies.

3. Session overview

The organizers will introduce the session, providing a background of the topic area, goals, and motivation for holding the session. There will then be a series of brief talks from the authors of the papers that were accepted for inclusion in the proceedings, a keynote by Vik Kheterpal from Care Evolution, ending with a panel discussion to include voices from academia, industry, and government including Q&A with attendees. The accepted papers/talks include causal data analysis of observational wearable device data, analysis of wearable silicone wristbands for chemical exposure, and digital biomarkers for detecting mild cognitive impairment. The talks are original research for publication, are widely varied, and the titles are listed below:

- Expanding the access of wearable silicone wristbands in community-engaged research through best practices in data analysis and integration
- Subject Harmonization of Digital Biomarkers: Improved Detection of Mild Cognitive Impairment from Language Markers
- Scalar-Function Causal Discovery for Generating Causal Hypotheses with Observational Wearable Device Data
- FedBrain: Federated Training of Graph Neural Networks for Connectome-based Brain Imaging Analysis (poster presentation only)

Following the original research talks, the keynote will be offered by Vik Kheterpal, the CEO and founder of Care Evolution. Vik is a nationally recognized expert in the area of healthcare informatics who has been focused on healthcare data exchange and interoperability for the past 11 years. He brings to the conversation the perspective of a serial entrepreneur working across IT, healthcare, and research sectors, and a go-to expert on real world data, healthcare IT, product design and usability, business, and leadership. After the keynote, attendees will be offered an opportunity to recommend DHT data collections and analysis methods that will help advance precision medicine research. This information will be shared with groups, such as the *All of Us* Research Program, that are collecting research data for the sake of advancing precision medicine.

The session will conclude with a panel discussion and audience Q&A. The panelists will feature speakers from industry, academia, and government; the session organizers will be joined by the keynote speaker and paper authors for a moderated discussion and Q&A from the participants. Session attendees are encouraged to participate in an interactive discussion on the current research, current challenges, and opportunities to expand access and use of DHT/wearables data in research.

References

1. https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/
2. AI detection of cardiac dysfunction from consumer watch ECG recordings. Nat Med 28, 2478–2479 (2022). https://doi.org/10.1038/s41591-022-02079-5
3. Tsang KCH, Pinnock H, Wilson AM, et alPredicting asthma attacks using connected mobile devices and machine learning: the AAMOS-00 observational study protocolBMJ Open 2022;12:e064166. doi: 10.1136/bmjopen-2022-064166
4. H. Y. Lu et al., "Digital Health and Machine Learning Technologies for Blood Glucose Monitoring and Management of Gestational Diabetes," in IEEE Reviews in Biomedical Engineering, doi: 10.1109/RBME.2023.3242261.
5. Bennett, C.C., Ross, M.K., Baek, E. et al. Smartphone accelerometer data as a proxy for clinical data in modeling of bipolar disorder symptom trajectory. npj Digit. Med. 5, 181 (2022). https://doi.org/10.1038/s41746-022-00741-3
6. https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/2725079
7. https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2018.05081
8. https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1377-7
9. https://preprints.jmir.org/preprint/29510/accepted
10. https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2018.05081
11. https://github.com/DigitalBiomarkerDiscoveryPipeline/Digital_Health_Data_Repository
12. https://datacc.dimesociety.org/
13. https://playbook.dimesociety.org/
14. https://pubmed.ncbi.nlm.nih.gov/33948242/
15. https://www.pnas.org/doi/full/10.1073/pnas.2119097118
16. Session Introduction: Digital health technology data in biocomputing: Research efforts and considerations for expanding access. M Holko, C Lunt, J Dunn - Pacific Symposium on Biocomputing, 2023