Navigating the Diabetes Jungle: AI for Daily Self-Management "In the Wild"

POOJA M., DESAI

Columbia University, pmd2137@cumc.columbia.edu

LENA, MAMYKINA

Columbia University, om2196@cumc.columbia.edu

The growth of AI tools within medicine has shown the immense potential for these technologies to support health in the wild. Within chronic-disease contexts, AI tools can support self-management through pattern identification, decision support, and personalized prediction. Here, we discuss findings from our in the wild deployment of GlucOracle, a diabetes self-management app that delivers personalized blood glucose forecasts to individuals and share insights to inform the future development of health technologies. Specifically, we highlight: (1) the need for AI systems that meet users' diverse expectations and adapt to their ever-evolving context, (2) challenges regarding the sporadic use of health tools in the wild and implications for developing unbiased models, and (3) users desire for explainable systems. We discuss opportunities for continued HCI work on these topics and hope to engage in collaborative discussion with the research community about designing effective health AI systems for the wild.

CCS CONCEPTS • Human-centered computing • HCI Design and Evaluation Methods • Field Studies

Additional Keywords and Phrases: diabetes, AI in the wild, self-management, explainability

ACM Reference Format:

Pooja M Desai, Lena Mamykina. 2021. Navigating the Diabetes Jungle: AI for Daily Self-Management "In the Wild. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 EA), May 8-13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 6 pages.*

1 INTRODUCTION

The rapid growth of AI technologies has profoundly touched every aspect of human life: from consumption habits, to social networks, to health. Emerging applications within the medical domain have shown the tremendous potential of AI technologies to identify health pattens, facilitate decision support, and leverage data to forecast future outcomes [2, 6, 12].

Within the health domain, the area of chronic disease self-management poses a unique challenge. Chronic diseases, such as diabetes, require individuals to rigorously manage various lifestyle factors including nutrition, physical activity and stress to maintain their health [13]. For individuals with diabetes, health-management challenges touch every dimension of their daily lives, impacting a myriad of their decisions in the wild. The ever-increasing volumes of data reflecting individuals' daily behaviors, coupled with emerging methods for computational data analysis, can help individuals with the task of self-management by recognizing important trends and patterns, recommending beneficial self-management behaviors and providing other forms of decision support. However, because such tools will need to integrate with individuals' daily lives, there remain questions about how to best design AI systems that can be easily adopted in the wild.

Our lab focuses on developing AI-based tools to support self-management of chronic diseases with a specific focus on type 2 diabetes (T2D). As part of this research, we have conducted several studies with participants, testing tools for T2D management in both small-scale pilot contexts and large-scale deployment "in the wild" studies. Our past work has highlighted several considerations that can inform the development of future AI health systems "in the wild":

1. When deployed in the wild, AI systems in health are inevitably faced with a great diversity of individuals' needs, preferences and expectations. AI systems should provide assistance and decision support that incorporates users' preferences and adapts to their evolving contexts and skills.
2. Individuals engage with health management technologies sporadically and inconsistently. This raises questions about how to best assess the effectiveness of interventions and how to handle data missingness in order to develop robust, unbiased models.
3. When presented with inferences and recommendations, users are eager to understand the inner decision logic of AI systems. This raises important ML and HCI questions about model explainability and how to best present explanations to users.

Below, we expand on these lessons and discuss implications for the future development AI systems to support health in the wild.

2 LESSONS FROM GLUCORACLE

In our previous research, we developed GlucOracle, a smartphone app to help individuals with T2D anticipate changes in their blood sugar. The app uses computational modeling and self-tracking data to generate personalized blood glucose forecasts in response to meals. Individuals use GlucOracle to log meal photos and short text descriptions, as well as blood glucose levels before and after meals to train their personalized model. Once the model is trained, users can log a meal they intend to eat and GlucOracle will present them with forecast for how that meal is anticipated to change their blood sugar, enabling them to make modifications to planned meals in real-time. GlucOracle's prediction model leverages data assimilation to generate personalized real-time predictions of an individuals' evolving blood glucose levels. In order to sufficiently train the model, users had to record 25 meals with pre- and post-meal blood glucose readings [1].

GlucOracle was evaluated in a 4-week "in the wild" pilot study with 10 participants with T2D who had varying levels of experience using technology for diabetes management (5 novice self-trackers, 5 experienced self-trackers). Users were invited to participate in qualitative interviews and share their perceptions and experiences with receiving personalized forecasts in the wild [3]. Specifically, we captured users' impressions of the usefulness of forecasts, and their attitudes toward receiving AI-generated forecasts to support nutrition management in everyday life.

Overall, GlucOracle users found the AI-generated forecasts compelling and informative. The study also opened many new questions about how to improve AI systems to support everyday decision-making and long-term engagement. Below we expand on insights from our deployment of GlucOracle and outline some perspectives and directions to consider for the future development of human-centered AI systems to support personal health.

2.1 Designing to support diversity and change in the wild

One finding that quickly emerged from our work with GlucOracle was that while users shared common challenges around nutrition and blood glucose management, they each had diverse needs, expectations and preferences for how to use the app.

First, users had varying levels of past experience with health technologies to manage blood glucose. Novice users found the process of tracking and receiving forecasts in and of itself to be eye-opening and rewarding. On the other hand, individuals who had more experience with both self-tracking and diabetes management had fewer open questions and were more likely to see continuous tracking as a burden, and simple forecasts as less informative. These individuals desired more detailed information about the dynamics of their blood glucose predictions, for example seeing more detailed depictions of changes in their blood glucose overtime, such as a blood glucose curve. Second, users had diverse approaches to incorporating forecasts into their daily meal practices. While some used forecasts to modify their planned meals, others noted challenges altering already prepared meals, especially when also cooking for family members. Instead, these users suggested flexible features for saving forecasted insights to consult beyond mealtime, such as during recipe planning and cooking.

These findings highlight a need for AI systems that capture the personal needs and preferences of users and evolve with them over time. For example, in addition to modeling physiological response to nutrition, as in GlucOracle, future AI systems for health management could attempt to model both individuals' physiology, as well as their preferences and usage patterns to provide assistance in the form appropriate for each user. Related work in chronic disease management also has shown the value of AI systems accounting for users' context and preferences when making specific suggestions (e.g. AI systems that recognize a user avoids meat, or AI systems used in family contexts that must account for the mutual preferences of several individuals) [11, 12]. Furthermore, it is crucial to remember that users' information needs evolve as they gain experience using digital tools for health management. Not only must in the wild AI systems accommodate needs of different users, but these systems must also adapt with each unique user as their needs, abilities and preferences change.

2.2 Mind the Data Gaps: Designing for Sporadic Use

Another important consideration that emerged from the GlucOracle study regarded users' perceptions around the long-term adoption of GlucOracle. Users reflected that tracking and receiving forecasts was most useful at the start of the study, while they learned what meals worked best for them. As users gained experience with the app and became more aware of how their blood glucose changed in response to common meals, forecasts became more predictable. Few participants saw themselves using GlucOracle with the same level of intensity in the long run. Instead, they saw it as a useful tool for occasional use to ensure they were on track, or to support ongoing discovery, for example, when they introduced new meals. They envisioned using the app for intense tracking at sporadic intervals, rather than continuously over long periods of time.

These findings have several implications for the design of future AI systems. First, they suggest the need for a more nuanced definition of engagement that distinguishes lapsing use due to *mastery* from lapsing use due to *abandonment*. This implies that in some contexts, “sustained engagement” may be an ill-fitting metric to accurately assess the usefulness of AI interventions [4, 8]. As we continue to develop AI tools for use in the wild, it is important to consider what other metrics we can use to assess the effectiveness of these interventions.

Second, sporadic data capture poses considerable challenges for developing accurate and unbiased algorithms that support health management [5]. Such tracking in “fits and spurts” can result in sparse and biased datasets that only capture individuals in a particular state of health (for example, users of GlucOracle only tracking their meals when their blood glucose was dysregulated) [10]. Current solutions include automated data collection through devices and app designs that simplify logging procedures (e.g., allowing users to duplicate previous entries, synchronizing data across multiple apps, etc.) in order to reduce tracking burden. However, challenges with comprehensive data capture persist and can make “in the wild” datasets suboptimal for algorithm training and development as they lack key data about individuals across a variety of health states.

This missingness makes it difficult to train models that can precisely detect unhealthy changes and offer personalized insights and suggestions to support users’ health. As we consider the development of AI to support health in the wild, it is important that we consider concerns about data quality and missingness in order to train accurate and unbiased models.

2.3 Unwrapping the Black Box: Making AI Insights Useful in Context

Finally, through our study we found that reflecting on forecasts made participants interested in understanding how forecasts were generated and opened new questions regarding model explainability. Overall, many participants found GlucOracle forecasts to be aligned with their own expectations, which increased their trust in the underlying AI engine. However, participants often wondered which aspects of meals contributed the most to the forecast. Furthermore, they wished to better understand both what influenced the predictions and the fuller extent of these predictions. For example, instead of a single numerical value, many wished to see the full blood glucose curve predicted by the model to gain a better understanding of how their blood glucose was predicted to change over time. These findings highlight new opportunities to incorporate insights from Explainable AI (XAI) research into AI systems that support health management in the wild.

However, there are many open questions about how to best design explainable systems. On the machine learning side, there are ongoing challenges with designing high performing models for which the decision logic can be easily explained (the so-called “performance-explainability tradeoff”) [7]. On the HCI side, there are questions about what model features are most necessary to explain, and how explanations should be presented to be useful in different interaction contexts [9].

With regards to Glucoracle, though the prediction model is highly explainable it is not immediately apparent what an a effective explanation would be for in-the-moment decision making context in which users engage with the app. Drawing from this work, we are interested in understanding how explanations (such as which elements of meals caused blood glucose forecasts to be higher or lower) can facilitate greater understanding, perceived usefulness, and willingness to act on AI insights in the context of routine heath management. We are also interested in how the format and style of useful explanations may differ between just-in-time settings where users want quick actionable feedback, and long-term recommendation settings where users want to reflect more deeply.

3 ONGOING WORK AND FUTURE DIRECTIONS

Our lab is continuing to explore these questions and engaging in studies to examine how AI-generated nutrition recommendations should be presented to users in the wild. Past work focusing on health recommendations has varied in the specificity of recommendations delivered by AI: ranging from surfacing a general trend, to recommending a type of behavior, to suggesting a very specific activity [2, 6, 12]. In the next steps of developing AI systems to support personal chronic disease management, it is vital to understand the kinds of AI-generated insights that can best support personal health, and find ways to present recommendations that are understandable, actionable, and trustworthy.

4 CONCLUSION

In this workshop we hope to engage with other HCI researchers and explore new ways to design AI technologies to better support personal health management. We believe the questions and insights from our studies will be a valuable contribution to this workshop and are eager to learn from fellow researchers. We look forward to a deep and engaging discussion about designing effective health AI systems for the wild.

ACKNOWLEDGMENTS

We thank the organizers of the CHI 2021 “Realizing AI in Healthcare: Challenges Appearing in the Wild” workshop for creating a forum for discussion about next steps for AI in health. Many thanks to the participants, translators, and members of the ARCH Lab who made this work possible. The original project was funded with support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) R01DK090372 and R56DK113189; and the National Library of Medicine (NLM) R01 LM012734-01.

REFERENCES

[1] David J Albers, Matthew Levine, Bruce Gluckman, Henry Ginsberg, George Hripcsak, and Lena Mamykina. 2017. Personalized glucose fore- casting for type 2 diabetes using data assimilation. PLoS computational biology 13, 4 (2017), e1005232.

[2] Frank Bentley, Konrad Tollmar, Peter Stephenson, Laura Levy, Brian Jones, Scott Robertson, Ed Price, Richard Catrambone, and Jeff Wilson. 2013. Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. ACM Transactions on Computer-Human Interaction 20, 5: 1–27. https://doi.org/10.1145/2503823

[3] Pooja M. Desai, Elliot G. Mitchell, Maria L. Hwang, Matthew E. Levine, David J. Albers, and Lena Mamykina. 2019. Personal Health Oracle: Explorations of Personalized Predictions in Diabetes Self- Management. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300600

[4] Daniel A. Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A. Munson. 2016. Beyond Abandonment to Next Steps: Understanding and Designing for Life after Personal Informatics Tool Use. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 1109–1113. DOI:https://doi.org/10.1145/2858036.2858045

[5] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A lived informatics model of personal informatics. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 731–742.

[6] Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin, and Steve Whittaker. 2017. What does all this data mean for my future mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being. Human–Computer Interaction 32, 5-6 (2017), 208–267.

[7] Andreas Holzinger. 2018. From machine learning to explainable AI. Proceedings of the 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), Košice, Slovakia, 2018, pp. 55-66, doi: 10.1109/DISA.2018.8490530.

[8] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to evaluate technologies for health behavior change in HCI research. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). Association for Computing Machinery, New York, NY, USA, 3063–3072. DOI:https://doi.org/10.1145/1978942.1979396

[9] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–15. DOI:https://doi.org/10.1145/3313831.3376590

[10] Jochen Meyer, Merlin Wasmann, Wilko Heuten, Abdallah El Ali, and Susanne C.J. Boll. 2017. Identification and Classification of Usage Patterns in Long-Term Activity Tracking. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, 667–678. DOI:https://doi.org/10.1145/3025453.3025690

[11] Shriti Raj, Kelsey Toporski, Ashley Garrity, Joyce M. Lee, and Mark W. Newman. 2019. "My blood sugar is higher on the weekends": Finding a Role for Context and Context-Awareness in the Design of Health Self-Management Technology. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 119, 1–13. DOI:https://doi.org/10.1145/3290605.3300349

[12] Darius A. Rohani, Andrea Quemada Lopategui, Nanna Tuxen, Maria Faurholt-Jepsen, Lars V. Kessing, and Jakob E. Bardram. 2020. MUBS: A Personalized Recommender System for Behavioral Activation in Mental Health. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376879

[13] Jaakko Tuomilehto, Jaana Lindström, Johan G Eriksson, Timo T Valle, Helena Hämäläinen, Pirjo Ilanne-Parikka, Sirkka Keinänen-Kiukaanniemi, Mauri Laakso, Anne Louheranta, Merja Rastas, et al. 2001. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. New England Journal of Medicine 344, 18 (2001), 1343–1350.