Counterfactual Explanations in Personal Informatics for Personalized Mental Health Management

Gyuwon Jung
KAIST
School of Computing
Daejeon, Republic of Korea
gwjung@kaist.ac.kr

Uichin Lee[*]
KAIST
School of Computing
Daejeon, Republic of Korea
uclee@kaist.ac.kr

ABSTRACT

Personal informatics has been widely used to support users' mental health management by identifying factors associated with health indicators from mobile data. However, it remains challenging for users to develop data-driven coping strategies for mental health issues, especially when targeting specific situations involving multiple factors. To address this, we suggest an analysis pipeline for investigating counterfactual scenarios using mobile data. We also show how the pipeline generates counterfactuals from an open dataset, illustrating the feasibility of this approach in providing practical guidelines for each unique situation. Moreover, we discuss several considerations for integrating the proposed pipeline into personal informatics systems.

CCS CONCEPTS

• **Applied computing → Health care information systems**; • **Human-centered computing → Mobile computing**.

KEYWORDS

Personal Informatics, Counterfactual Explanation, Mental Health, Human Behavior and Context

ACM Reference Format:
Gyuwon Jung and Uichin Lee. 2024. Counterfactual Explanations in Personal Informatics for Personalized Mental Health Management. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp Companion '24), October 5–9, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3675094.3678427

1 INTRODUCTION AND RELATED WORKS

Mobile devices such as smartphones and wearables have become pervasive in everyday life, creating opportunities to provide practical data-driven insights for health management. Systems like personal informatics enable users to collect data from their daily lives and use it to reflect on themselves or expand their self-knowledge [4, 13]. These systems are designed to address various health and well-being issues and are particularly useful for managing chronic diseases where long-term tracking and monitoring are necessary. Previous studies [11, 16] have also demonstrated the use of personal informatics to support users in the domains of mental health (e.g., stress) and affective computing (e.g., emotion and mood).

According to the stage-based model suggested by Li et al. [13], the final stage is 'action,' where users tailor their behaviors based on self-reflection on their data. To change their goals or take action in this stage, users first need to understand what affects their target outcomes from their data. Regarding this, previous personal informatics systems provided correlation analyses, such as the relationships between contextual factors and well-being indicators [1]. Additionally, they allowed users to conduct self-experimentation [10] or identify causal relationships by employing quasi-experimental approaches [8]. Findings from these analyses help users understand which factors are positively or negatively related to their mental health and plan their own strategies to improve their health status.

Yet, personal informatics systems suggested in previous studies typically conducted analyses on the entire dataset but focused less on individual instances. They could answer whether a specific factor is correlated (or causal) to the mental health indicator *in general*, but were limited in guiding what should be done in specific situations comprising multiple factors. For instance, systems may conclude that studying leads to higher stress levels. However, at the instance level, other contextual factors such as place, social settings, and time (e.g., studying with *friends* at a *cafe* in the *afternoon*) can relieve the stress caused by studying. This implies that different insights can be provided depending on whether the relationship between certain factors and mental health is viewed from the perspective of the entire dataset or from individual instances. In the latter case, personal informatics can suggest customized coping strategies targeting specific combinations of contextual factors, providing much more detailed and practical guidance. In the example above, systems can suggest '*studying with friends*' or '*studying at a cafe*' for lower stress levels instead of simply saying '*studying less*.'

To enable this analysis, we suggest employing counterfactual explanations [6], a well-known method in the explainable AI (XAI) field, in personal informatics. "Counterfactuals," which indicate scenarios contrary to established facts, allow people to consider alternative past events that might have led to different outcomes [2, 20]. For example, let us assume that students who usually do not study much (i.e., factual) receive poor grades on their recent exams. They might then think about events that did not actually happen such as "*What would have happened if I had spent more time studying or if I had attended all the classes?*" (i.e., counterfactual). The concept of counterfactuals provides useful insights into machine learning models, enabling researchers to examine how a prediction can change by modifying feature values. In the example above, changes in

[*] Corresponding author

UbiComp Companion '24, October 5–9, 2024, Melbourne, VIC, Australia
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1058-2/24/10
https://doi.org/10.1145/3675094.3678427

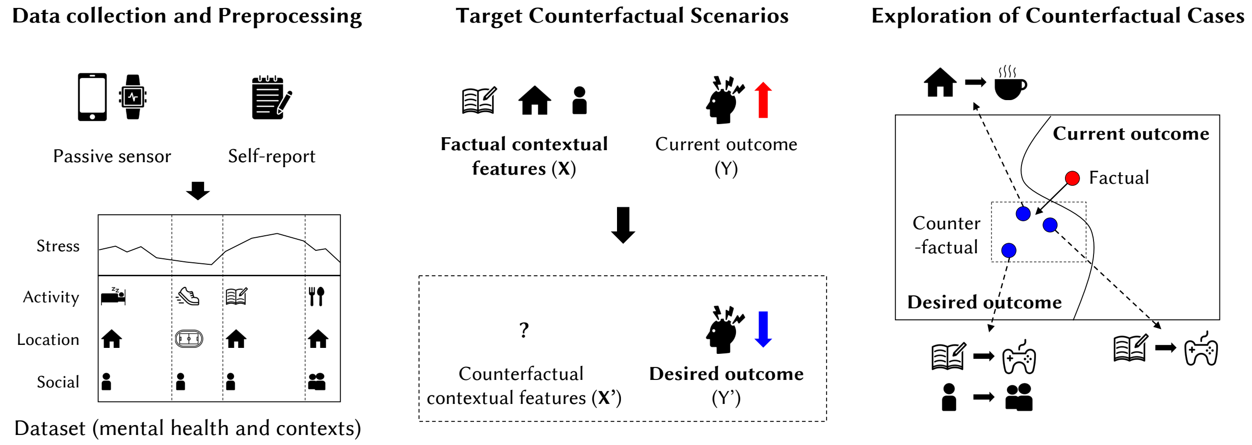
Data collection and Preprocessing
Passive sensor
Self-report
Stress
Activity
Location
Social
Dataset (mental health and contexts)
Target Counterfactual Scenarios
Factual contextual features (X)
Current outcome (Y)
?
Counterfactual contextual features (X')
Desired outcome (Y')
Exploration of Counterfactual Cases
Current outcome
Factual
Counter -factual
Desired outcome

Figure 1: The overview of the counterfactual explanation pipeline for mental health proposed in this study

features such as time spent studying or the attendance rate may alter the predicted grades on exams. Conversely, the model can suggest which features need to be altered to what extent in order to predict the desired outcome (e.g., study 15 hours more than in the recent case to get a good grade with a probability above 75%). Recent studies [21, 22] have applied counterfactual explanations to mental health prediction research to reveal significant indicators from machine learning models built on data from multiple participants. However, since our objective is to provide solutions to each individual through personal informatics, we chose to build a model and generate counterfactuals using data collected from specific individuals. Given that the same situations may affect each individual's stress differently, we expect that this approach would result in more personalized coping strategies for mental health.

In this study, we first propose a counterfactual analysis pipeline for personal informatics, outlining a process to explore which contextual features need to be changed to achieve desired mental health outcomes using everyday mobile data. In addition, we showcase how this pipeline generates counterfactual scenarios using an open dataset consisting of stress levels and multiple contextual factors (i.e., activity, place, social setting, and time). Based on this case study, we discuss design considerations for employing counterfactual explanations in personal informatics for mental health management.

2 COUNTERFACTUAL EXPLANATION PIPELINE FOR MENTAL HEALTH

We suggest an analysis pipeline for investigating counterfactual scenarios for managing mental health using mobile data. As illustrated in Figure 1, this pipeline consists of three steps: (1) data collection and preprocessing, (2) determining the target counterfactual scenario, and (3) exploration of counterfactual cases.

2.1 Data Collection and Preprocessing

The first step is collecting and preprocessing mobile data to construct a basic lifelogging dataset. As suggested in previous studies [17], mobile devices such as smartphones and wearables can automatically collect various types of sensor data, allowing us to extract features that represent human behaviors, contexts, and mental health states. As exemplified by Jung et al. [9], mobile data collected from passive sensors (e.g., GPS and accelerometer) and interactions with devices (e.g., touching or swiping the smartphone screen) can be preprocessed and converted into features such as significant places, physical activities, and mobile app usage behaviors. Moreover, mental health indicators such as perceived stress levels can also be inferred from data recorded through microphones during conversations [14] or collected via body-worn sensors such as photoplethysmography (PPG) and electrocardiography (ECG) [7].

However, some features, such as perceived stress levels, may be challenging to capture relying solely on automatically collected data. In such cases, the experience sampling method (ESM) with mobile devices [24] can be used, allowing individuals to self-report their in-situ experiences, including mental health status. ESM data is useful for gaining a more detailed understanding of the individual as it complements the information extracted from passive sensor data. Through this process, we can construct a dataset that describes the individual's situation and mental health status at specific moments.

2.2 Target Counterfactual Scenarios

In the second step, we determine the target counterfactual scenario by specifying the individual instance to be investigated and the desired outcome. As explained in the previous section, counterfactual scenarios provide insights into what the potential outcomes might have been if alternative choices or situations, which did not actually occur, had been made. By comparing the outcomes derived from the factual and counterfactual cases, it can be evaluated which choice would have been preferable.

Conversely, this concept can help identify the proper counterfactual cases that could change the outcome prediction to the desired one. In this step, we choose a target instance to be investigated, consisting of factual contextual features ($\mathbf{X}$) and the corresponding (predicted) mental health outcome (Y), from the dataset built in the previous step. Moreover, we decide on a desired health outcome (Y') different from Y, and subsequently identify the counterfactual feature set ($\mathbf{X}'$) that predicts the Y'.

2.3 Exploration of Counterfactual Cases

Based on the target instance ($\mathbf{X}$, Y) and the desired outcome (Y'), we explore the counterfactual feature set (**X'**). In this step, we train a machine learning model to predict the mental health outcome using the given features. Then, we explore the changes in feature values required to achieve the desired outcome using the model.

Previous studies have suggested several approaches to identifying counterfactuals following Wachter et al.'s pioneering work [25], such as Actionable Recourse [23], Diverse Counterfactual Explanations (DiCE) [18], and Multi-Objective Counterfactuals (MOC) [3]. We chose to employ the MOC due to its capability to provide diverse counterfactual sets while considering trade-offs between multiple objectives, making it practical for suggesting more alternatives to personal informatics users. Specifically, this approach searches for counterfactuals by optimizing four objectives, resulting in counterfactuals that (1) have predicted outcomes close to the desired one, (2) minimize the change in each feature, (3) minimize the number of features changed, and (4) are likely to be in the actual dataset. Dandl et al. [3] demonstrated the usefulness of the MOC by comparing it with other relevant approaches, showing that it offers more counterfactuals that resemble the training data more closely and require fewer feature changes.

Consequently, this process provides a counterfactual contextual feature set that would bring about the desired health outcome and describes the probability of predicting those outcomes. Also, it can produce multiple counterfactuals, allowing individuals to choose the most useful ones while considering their feasibility in practice.

3 CASE STUDY

3.1 Method

We showcase how the proposed pipeline generates counterfactual scenarios for mental health management from a lifelogging dataset. As a case study, we utilized an open dataset collected in a recent study [8] exploring causal relationships between contextual features and perceived stress levels. This dataset consists of ESM data from 24 participants collected over 6 weeks, including contextual features (i.e., activity, location, social setting, and time) and perceived stress levels on a 5-point Likert scale. Hereafter, we refer to combinations of these contextual features as 'contextual feature sets.'

To show the feasibility, we provide an in-depth analysis of one participant (ID: 24). We selected this participant from among several others to gather as many self-report stress labels as possible (726 samples) in order to create a model for counterfactual explanations. Out of a total of 726 samples, we identified 132 unique sets of contextual features, each consisting of different combinations of activity, location, social setting, and time. We used these contextual feature sets as target instances for exploring counterfactuals. Furthermore, for simplicity, we binarized the stress levels into high and low based on the mean stress level. This binarization may also make the explanations provided to users simpler and more intuitive (e.g., an explanation like "to have below-average stress" rather than "to have a stress level of 2"). The collected self-report stress data exhibited a bimodal distribution, so there was less potential for issues that could arise from dividing the data based on the mean value, which can occur with a unimodal distribution [5].

We trained a random forest model to predict the perceived stress levels based on the set of contextual features. Depending on the predicted outcome of the target instance, we explored counterfactual contextual features that would result in the opposite outcome via Multi-Objective Counterfactuals (MOC) [3]. If the predicted stress level was high, the pipeline provided sets of contextual features that would predict a low stress level, and vice versa. By doing so, individuals can alter their behaviors or environments by examining the suggested set of contextual features as potential stressors.

3.2 Results

Among the 132 unique sets of contextual features, 39 were predicted to result in high stress and 93 in low stress. For each of these sets, an average of 3.9 counterfactual cases were identified (SD: 4.1, Max: 22, Min: 1). Two-thirds of the unique contextual feature sets generated three or fewer counterfactuals, while three sets generated more than 20 counterfactuals. Also, 96.7% of the identified counterfactuals changed only 1 or 2 features from the factual case, indicating that changes were minimized when searching for counterfactuals.

As an example, we illustrate how the counterfactual cases were generated for the contextual feature set {studying, dormitory, alone, morning}, which occurred most frequently in the original dataset (65 times out of 726 samples). This set showed an average stress level of 3.92 (SD: 0.85) out of 5, and the trained model predicted high stress with a probability of 98.4% (Accuracy: 0.764, F1-Score: 0.766). During the counterfactual exploration process, the pipeline generated three counterfactual cases for this set of factual features. Table 1 shows details of the counterfactuals, including their contextual features and probabilities of being predicted as either high-stress or low-stress using the model. The results highlight that stress could be lowered if the activity is changed from 'studying' to 'resting,' 'leisure activity,' or 'eating.' In fact, these counterfactuals can act as coping strategies, helping individuals understand which contextual features should be changed to manage their stress better.

Table 1: The counterfactual cases for the contextual feature set {studying, dormitory, alone, morning} (F: Factual, CF: Counterfactual)

	Activity	Location	Social	Time	High stress	Low stress
F	Studying	Dormitory	Alone	Morning	**0.984**	0.016
CF	**Resting**	Dormitory	Alone	Morning	0.018	**0.982**
CF	**Leisure activity**	Dormitory	Alone	Morning	0.188	**0.812**
CF	**Eating**	Dormitory	Alone	Morning	0.384	**0.616**

Conversely, we could select cases predicted to have low stress and explore which feature changes would result in the undesired outcome (i.e., high stress). For example, we started with the contextual feature set {resting, dormitory, alone, afternoon}, which showed a low stress level (Mean: 2.67, SD: 1.41). For this set of features, the trained model predicted low stress with a probability of 90.8% (Accuracy: 0.770, F1-Score: 0.745), and four counterfactual cases were generated through the proposed pipeline, as illustrated in Table 2. This person may utilize these findings to avoid situations that could increase stress from the current situation. In this case, changing the activity from resting to studying, social activity, or class, or moving from the dormitory to the classroom, may cause the person to experience higher stress. The individual may then avoid these changes when they are in {resting, dormitory, alone, morning}. In addition, note that the counterfactuals can be generated by changing different feature types (i.e., activity or location), as shown in this example.

Table 2: The counterfactual cases for the contextual feature set {resting, dormitory, alone, afternoon} (F: Factual, CF: Counterfactual)

	Activity	Location	Social	Time	High stress	Low stress
F	Resting	Dormitory	Alone	Afternoon	0.092	**0.908**
CF	**Studying**	Dormitory	Alone	Afternoon	**0.998**	0.002
CF	**Social activity**	Dormitory	Alone	Afternoon	**0.954**	0.046
CF	**Class**	Dormitory	Alone	Afternoon	**0.738**	0.262
CF	Resting	**Classroom**	Alone	Afternoon	**0.700**	0.300

We also discovered cases where multiple features needed to be changed together to predict the desired outcome. Table 3 describes the counterfactual cases for the contextual feature set {class, classroom, alone, morning}, which was predicted to show high stress with a probability of 92.4% (Accuracy: 0.778, F1-Score: 0.790). In this case, the top three counterfactual cases with the highest probability of being predicted to have low stress required changes in two different feature types simultaneously (i.e., both activity and location). These results could occur because the MOC method produces counterfactuals that show different trade-offs between objectives. Individuals can observe these counterfactuals and choose one, considering the trade-off between the number of features to be changed (i.e., requiring more effort) and the closeness to the desired mental health outcome (i.e., achieving low stress).

Table 3: The counterfactual cases for the contextual feature set {class, classroom, alone, morning}, illustrating the top 5 out of 16 counterfactuals by the probability of low stress (F: Factual, CF: Counterfactual)

	Activity	Location	Social	Time	High stress	Low stress
F	Class	Classroom	Alone	Morning	**0.924**	0.076
CF	**Leisure activity**	**Private space**	Alone	Morning	0.022	**0.978**
CF	**Leisure activity**	**Private space**	Alone	Morning	0.174	**0.826**
CF	**Leisure activity**	**Private space**	Alone	Morning	0.180	**0.820**
CF	Class	**Home**	Alone	Morning	0.306	**0.694**
CF	Class	**Club room**	Alone	Morning	0.306	**0.694**
...	...	...	...	...	...	...

4 DISCUSSION AND CONCLUSION

This study proposes the application of counterfactual explanation techniques to mobile data to explore ways to manage mental health in everyday life. By following the steps of the suggested pipeline, individuals can understand which contextual features should be changed in specific situations to achieve the desired mental health status. As existing studies have revealed, users of personal informatics are highly interested not only in investigating which factors affect their health but also in developing appropriate coping behaviors [12]. We envision that this approach can be utilized in personal informatics to guide alternative ways to lower stress levels or improve emotional states.

The strength of this method lies particularly in providing answers tailored to an individual's situation, which involves a combination of multiple contextual features. Typical personal informatics only offer overall associations between each feature and health status (e.g., correlation or causality), limiting the understanding of whether the health status would improve by manipulating that feature within a specific situation. However, the suggested approach allows users to assume a unique contextual feature set that is likely to occur, predict stress levels based on it, and explore alternatives (i.e., counterfactuals) proactively. Given that counterfactual explanations enhance human interpretation of machine learning results, we could evaluate these counterfactuals through user studies. For instance, we could assess whether the counterfactuals effectively achieve the desired mental health outcomes (e.g., reducing stress levels) as well as evaluate their feasibility in real-world applications.

There are several considerations for employing counterfactual explanation techniques in personal informatics. First, we need to explore when and how to deliver these counterfactual scenarios to users better. They can be displayed when users reflect on their stress retrospectively or before they are exposed to situations where stress levels are expected to be high. The system may focus more on relatively frequent situations and provide coping strategies. In addition, we may enable users to set constraints in generating counterfactuals by changing features. For example, if users cannot help but continue the 'studying' activity in Table 1, the system may analyze counterfactual cases that keep the activity unchanged but alter other feature types. Furthermore, we can investigate how users apply these counterfactual findings in practice, particularly when there is a gap between their self-knowledge and the system's analysis [15] or those counterfactual cases are not applicable [19].

As a case study, we mainly employed ESM data that were manually reported. As outlined in the pipeline, it is also possible to use features extracted from passive sensor data, along with self-reported data. The data used in this study includes a limited number of contexts collected from users. However, by collecting a wider range of factors from daily life that could potentially influence or be causally related to stress, we can explore a variety of counterfactuals to support stress management. Moreover, we may utilize various counterfactual searching approaches for mobile data analysis and compare their outcomes.

We may extend this study to scenarios beyond mental health. For example, if individuals employ multiple strategies to lose weight (e.g., exercise, sleep, diet, water intake, and vitamin intake), the system may offer ways to improve the effectiveness of the current health management strategy. This approach would utilize continuous values for each feature, potentially generating more diverse counterfactual cases. In conclusion, by leveraging counterfactual explanation techniques, we may provide personalized, actionable insights that empower individuals to proactively manage their mental health and overall well-being.

ACKNOWLEDGMENTS

This research was supported by the Bio & Medical Technology Development Program and the Basic Science Research Program of the National Research Foundation (NRF), funded by the Korean government (MSIT) (No. 2021M3A9E4080780) and by the NIPA (National IT Industry Promotion Agency), funded by the Korean government (MSIT) (No. H1001-24-1002).

REFERENCES

[1] 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 (TOCHI)* 20, 5 (2013), 1–27. https://doi.org/10.1145/2503823

[2] Ruth M. J. Byrne. 2019. Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 6276–6282. https://doi.org/10.24963/ijcai.2019/876

[3] Susanne Dandl, Christoph Molnar, Martin Binder, and Bernd Bischl. 2020. Multi-Objective Counterfactual Explanations. In *International Conference on Parallel Problem Solving from Nature*. Springer, Springer International Publishing, Cham, 448–469. https://doi.org/10.1007/978-3-030-58112-1_31

[4] 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*. Association for Computing Machinery, New York, NY, USA, 731–742. https://doi.org/10.1145/2750858.2804250

[5] Eiko I Fried, Jessica K Flake, and Donald J Robinaugh. 2022. Revisiting the theoretical and methodological foundations of depression measurement. *Nature Reviews Psychology* 1, 6 (2022), 358–368.

[6] Riccardo Guidotti. 2022. Counterfactual Explanations and How to Find Them: Literature Review and Benchmarking. *Data Mining and Knowledge Discovery* (2022), 1–55. https://doi.org/10.1007/s10618-022-00831-6

[7] Karen Hovsepian, Mustafa Al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 493–504. https://doi.org/10.1145/2750858.2807526

[8] Gyuwon Jung, Sangjun Park, and Uichin Lee. 2024. DeepStress: Supporting Stressful Context Sensemaking in Personal Informatics Systems Using a Quasi-experimental Approach. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3613904.3642766

[9] Gyuwon Jung, Sangjun Park, Eun-Yeol Ma, Heeyoung Kim, and Uichin Lee. 2024. A Tutorial on Matching-based Causal Analysis of Human Behaviors Using Smartphone Sensor Data. *Comput. Surveys* 56, 9 (2024), 1–33. https://doi.org/10.1145/3648356

[10] Ravi Karkar, Jessica Schroeder, Daniel A Epstein, Laura R Pina, Jeffrey Scofield, James Fogarty, Julie A Kientz, Sean A Munson, Roger Vilardaga, and Jasmine Zia. 2017. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 6850–6863. https://doi.org/10.1145/3025453.3025480

[11] Rafal Kocielnik and Natalia Sidorova. 2015. Personalized Stress Management: Enabling Stress Monitoring With LifelogExplorer. *KI-Künstliche Intelligenz* 29 (2015), 115–122. https://doi.org/10.1007/s13218-015-0348-1

[12] Kwangyoung Lee, Hyewon Cho, Kobiljon Toshnazarov, Nematjon Narziev, So Young Rhim, Kyungsik Han, YoungTae Noh, and Hwajung Hong. 2020. Toward Future-Centric Personal Informatics: Expecting Stressful Events and Preparing Personalized Interventions in Stress Management. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376475

[13] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-Based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 557–566. https://doi.org/10.1145/1753326.1753409

[14] Hong Lu, Denise Frauendorfer, Mashfiqui Rabbi, Marianne Schmid Mast, Gokul T Chittaranjan, Andrew T Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. 2012. StressSense: Detecting Stress in Unconstrained Acoustic Environments Using Smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. Association for Computing Machinery, New York, NY, USA, 351–360. https://doi.org/10.1145/2370216.2370270

[15] Lena Mamykina, Elizabeth M Heitkemper, Arlene M Smaldone, Rita Kukafka, Heather J Cole-Lewis, Patricia G Davidson, Elizabeth D Mynatt, Andrea Cassells, Jonathan N Tobin, and George Hripcsak. 2017. Personal Discovery in Diabetes Self-Management: Discovering Cause and Effect Using Self-Monitoring Data. *Journal of Biomedical Informatics* 76 (2017), 1–8. https://doi.org/10.1016/j.jbi.2017.09.013

[16] Daniel McDuff, Amy Karlson, Ashish Kapoor, Asta Roseway, and Mary Czerwinski. 2012. AffectAura: An Intelligent System for Emotional Memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 849–858. https://doi.org/10.1145/2207676.2208525

[17] David C Mohr, Mi Zhang, and Stephen M Schueller. 2017. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annual Review of Clinical Psychology* 13 (2017), 23–47. https://doi.org/10.1146/annurev-clinpsy-032816-044949

[18] Ramaravind K Mothilal, Amit Sharma, and Chenhao Tan. 2020. Explaining Machine Learning Classifiers Through Diverse Counterfactual Explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, 607–617. https://doi.org/10.1145/3351095.3372850

[19] Rafael Poyiadzi, Kacper Sokol, Raul Santos-Rodriguez, Tijl De Bie, and Peter Flach. 2020. FACE: Feasible and Actionable Counterfactual Explanations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. Association for Computing Machinery, New York, NY, USA, 344–350. https://doi.org/10.1145/3375627.3375850

[20] Neal J Roese. 1997. Counterfactual Thinking. *Psychological Bulletin* 121, 1 (1997), 133–148. https://doi.org/10.1037/0033-2909.121.1.133

[21] Kei Shibuya, Zachary D King, Maryam Khalid, Han Yu, Yufei Shen, Khadija Zanna, Ryan L Brown, Marzieh Majd, Christopher P Fagunders, and Akane Sano. 2023. Predicting Stress and Providing Counterfactual Explanations: A Pilot Study on Caregivers. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 1–4. https://doi.org/10.1109/ACIIW59127.2023.10388130

[22] Ziang Tang, Zachary King, Alicia Choto Segovia, Han Yu, Gia Braddock, Asami Ito, Ryota Sakamoto, Motomu Shimaoka, and Akane Sano. 2023. Burnout Prediction and Analysis in Shift Workers: Counterfactual Explanation Approach. In *2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 1–4. https://doi.org/10.1109/BHI58575.2023.10313392

[23] Berk Ustun, Alexander Spangher, and Yang Liu. 2019. Actionable Recourse in Linear Classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, 10–19. https://doi.org/10.1145/3287560.3287566

[24] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *Comput. Surveys* 50, 6 (2017), 1–40. https://doi.org/10.1145/3123988

[25] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual Explanations Without Opening the Black Box: Automated Decisions and The GDPR. *Harvard Journal of Law & Technology* 31 (2017), 841. https://doi.org/10.2139/ssrn.3063289