Data-centric explanation methods for healthcare professionals

Dries Van Dievoort[0009−0008−3074−1404], Robin De Croon[0000−0002−1329−156X], Vero Vanden Abeele[0000−0002−3031−9579], and Katrien Verbert[0000−0001−6699−7710]

Augment, an imec research group at KU Leuven, Leuven (Arenberg), Celestijnenlaan 200a, Leuven, 3001, Belgium

Abstract. Generative AI offers potential in healthcare, especially through trustworthy and transparent virtual assistant solutions. Our D-STRESS research project, via its mental health use case, explores how to apply generative AI responsibly in high-stakes environments. By integrating explainable AI (XAI), multimodal longitudinal data, and human-centred design (HCD), we aim to support stress *recovery* by combining subjective self-reports, intervention reports, physiological signals, and contextual data to deliver a holistic understanding of stress patterns. Central to this effort is a clinician-facing dashboard co-designed with healthcare professionals. The dashboard presents interpretable and actionable insights using XAI techniques following a human-centred XAI (HCXAI) approach. It features a three-pane layout: patient data visualisation, analysis of model outputs, and a conversational agent for natural language interaction. This iterative, human-in-the-loop approach ensures task relevance and alignment with clinical workflows, promotes transparency, and fosters trust in AI-assisted decision-making.

Keywords: Stress recovery · explainable AI (XAI) · Human-centred design (HCD)

1 Introduction

Generative AI is transforming healthcare, with virtual assistant solutions emerging as a powerful tool to support clinical workflows, enhance patient engagement, and enable mental health interventions. Realising this potential in real-world settings requires advances in technical innovation, trustworthiness, explainability, and human-centred design (HCD).

Our D-STRESS research project addresses some of these challenges by enhancing existing and designing new methodologies to create explainable, ethical, and tailored virtual assistants powered by generative AI. D-STRESS seeks to deepen our understanding of stress recovery patterns and adaptive coping strategies by exploring daily-life stress patterns through a multimodal approach. This multimodal data includes subjective self-reports via a conversational agent (CA), physiological data, and contextual information, offering a holistic view of stress recovery and its influencing factors.

To support transparency and trust in clinical decision-making, we will explore a range of explanation methods (e.g., SHAP, LIME, counterfactuals, etc.), incorporating a data-centric approach combined with a human-centric perspective. These methods will be embedded into a user-friendly dashboard designed for healthcare professionals. We will evaluate the effectiveness and usability of these explanation tools in real-world settings through an iterative process involving at least 30 healthcare professionals (HPs). Our design builds on prior work (e.g., Fig. 1 & 2) to deliver clear, actionable insights to support risk stage management and stress recovery in clinical practice.

This short workshop paper is structured as follows: we begin with a high-level overview of related work, including our earlier work. We then describe our approach to embedding explanation methods into a clinician-oriented dashboard and conclude with a discussion of the potential impact and future directions of this research.



Fig. 1. Dashboard for coaches from the Personal Health Empower project, a collaboration with Idewe



Fig. 2. Illustration of earlier work. Data-centric explanation dashboard for nurses, published at ACM CHI [2]

2 Related work

2.1 Human-centred design

Human-centred design is a methodology that prioritises the needs, limitations, and preferences of end-users throughout the design process. In the context of mental healthcare, HCD is key to ensuring that AI systems are intuitive, usable, and meaningful for clinical professionals. In earlier work, we showed that interaction between users (AI novices) and AI research is crucial to generate valuable insights. The key findings were to prioritise explanation methods that emphasise the underlying data rather than solely focusing on the model's internal workings [12]. Given the high stakes of healthcare and the complexity of clinical decisions, collaboration with end-users throughout the development process is paramount. A deep understanding of clinicians' workflows, decision-making processes, and the clinical environment is crucial. In one example, researchers used iterative design techniques informed by real-world clinical contexts to develop an XAI system for diagnosing central nervous system tumours [10]. Finally, HCD is inherently iterative, requiring repeated testing, refinement, and feedback incorporation. In another paper, they detail how iterative prototyping and user

validation were used to develop an explainable clinical decision support system, demonstrating how continuous engagement helps the system adapt to changing clinical needs [7].

2.2 Explainable AI in healthcare

XAI enhances transparency by making model reasoning accessible, supporting ethical standards, regulatory compliance, and clinician trust. For example, [9,8] illustrate how natural language explanations improve understanding and assessment of AI outputs. Interpretability is vital for clinicians to trust AI recommendations, especially in critical scenarios like diagnosis and treatment planning [9]. Common XAI techniques such as SHAP, LIME, and counterfactuals can help clinicians understand AI-driven decisions, stimulating collaboration between humans and machines, as is shown by Tuan et al. [13]. Explainability also plays a role in identifying and mitigating bias. By illuminating how decisions are made, XAI contributes to equitable treatment outcomes across diverse patient populations. Hamida et al. [4] highlighted how fairness is intrinsically linked to transparency, emphasising the ethical imperative to avoid favouring specific groups or attributes. Moreover, in our recent work, we argue for involving domain experts to mitigate representation bias in AI systems [1]. Finally, integrating clinician feedback into AI systems ensures that explanations remain relevant and actionable. A human-in-the-loop approach fosters a feedback loop between design and application, improving both usability and trust [11].

2.3 Integration of human-centred design and explainability

Effectively integrating HCD and XAI requires a multidisciplinary approach that bridges AI development with clinical expertise and user experience design. This integration ensures that systems are not only technically robust but also aligned with the workflows, expectations, and values of end-users. For example, one study developed a user-centred XAI system for ECG diagnosis [6], involving both clinicians and patients to create explanations that were both understandable and actionable. Another example [7], as shown in section 2.1, describes a co-design approach to building an explainable clinical decision support system, emphasising the importance of clinician participation in ensuring technical accuracy and clinical relevance. Importantly, while user-centredness is crucial, it must be balanced with task-centred design to ensure systems effectively support core clinical objectives. Task relevance ensures that innovations are not only usable but also practical and impactful in real-world settings [5].

3 XAI dashboard for mental healthcare professionals

Using data-centric methods such as SHAP, LIME, counterfactuals, and others, our project co-develops innovative explanation procedures that integrate both global and local perspectives. Apart from the data-centric methods, research

shows that it is important to focus on the human, the HPs and the patients in our case. Striving for human-centred XAI (HCXAI) [3], HCXAI ensures that AI enhances human capabilities rather than replacing them by focusing on the actions that the explanations enable the user to take. All these different explanations are embedded into a user-friendly dashboard, enabling HPs to: examine patterns of stress recovery, assess the accuracy of risk stage predictions, and understand model behaviour within the patient's context.

The goal of this approach is to support transparency and build trust in clinical decision-making. Our current dashboard prototype consists of a three-pane layout and is illustrated in Fig. 3. The three panes consist of a single patient's data view, an analysis visualisation panel, and a chatbot interface panel. The patient's data is shown per measured variable, and each data plot consists of the latest measured value plus a line chart showing the evolution of this variable over a set period, two weeks in this case. The middle panel is the analysis view, where we can show any analysis visualisation plots, such as feature importance plots shown in Fig. 3, but also possible recommendations for further steps in the stress recovery process as illustrated in Fig 4. You can see the stress recovery stage assessment of the AI model and its certainty level. The idea of the recommendations is to base this on the stage achieved by the model, to tailor the recommendations to each stress stage's needs. The main idea of the chatbot interface is to provide a large language model like a chatbot, allowing the HP to talk/ask questions about the patient's profile in a natural language way.

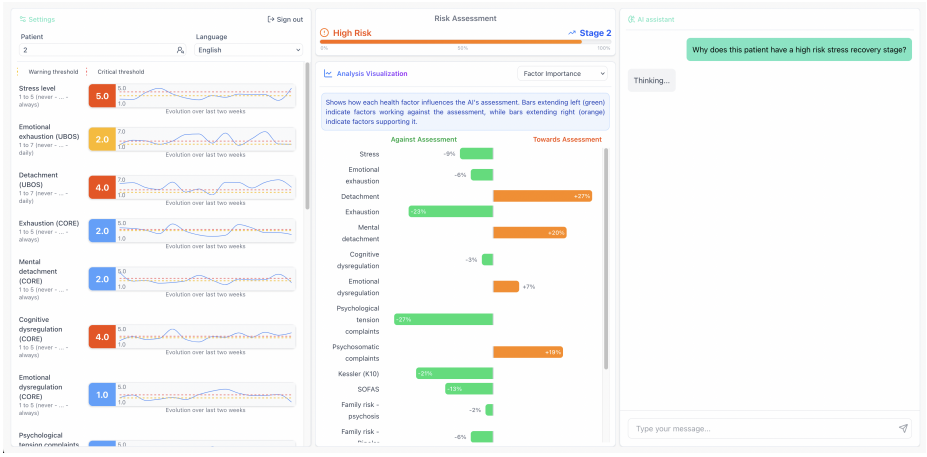


Fig. 3. D-STRESS dashboard

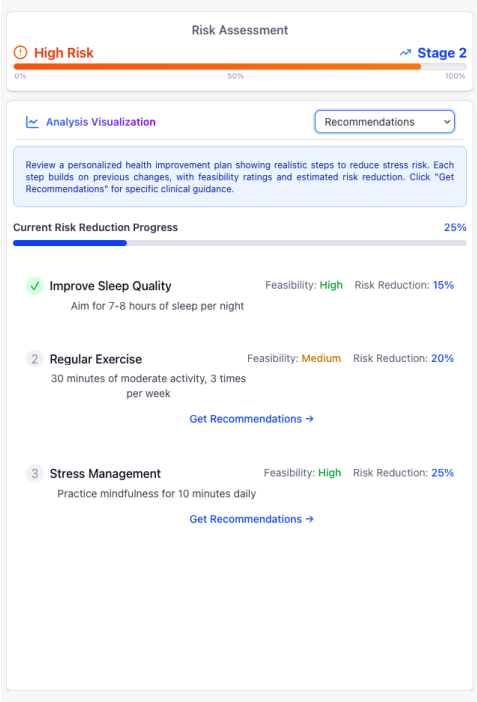


Fig. 4. Recommendations panel of the D-STRESS dashboard

4 Conclusion

Our work researches the design of clinician-facing dashboards that present actionable, interpretable insights, supporting both transparency and usability. The iterative involvement of healthcare professionals ensures that our tools align with real-world clinical needs, promoting adoption and long-term impact, while still holding task relevance close to our hearts.

Looking ahead, we will refine these tools based on user feedback, which we will gain from focus groups, co-design sessions, etc. This research endeavour contributes to the responsible integration of AI in healthcare, adding to the foundation for AI systems in clinical practice that are more ethical, transparent, and aligned with human tasks and values.

5 Acknowledgments

This research was supported by the Research Foundation–Flanders (FWO grant G0A4923N) and VLAIO (D-STRESS project HBC.2024.0681). Our workshop attendees' input, time and contribution are also highly valued.

References

1. Bhattacharya, A., Stumpf, S., Croon, R.D., Verbert, K., Ver-bert, K.: Explanatory debiasing: Involving domain experts in the data generation process to mitigate representation bias in ai systems (2025). https://doi.org/10.1145/3706598.3713497, https://doi.org/10.1145/3706598.3713497
2. Bhattacharya, A., Stumpf, S., Gosak, L., Stiglic, G., Verbert, K.: Exmos: Explanatory model steering through multifaceted explanations and data configurations. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. CHI '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3613904.3642106, https://doi.org/10.1145/3613904.3642106
3. Ehsan, U., Riedl, M.O.: Explainable ai reloaded: Challenging the xai status quo in the era of large language models (2024). https://doi.org/10.1145/3686169.3686185, https://doi.org/10.1145/3686169.3686185
4. Hamida, S.U., Chowdhury, M.J.M., Chakraborty, N.R., Biswas, K., Sami, S.K.: Exploring the landscape of explainable artificial intelligence (xai): A systematic review of techniques and applications. Big Data and Cognitive Computing 2024, Vol. 8, Page 149 **8**, 149 (10 2024). https://doi.org/10.3390/BDCC8110149, https://www.mdpi.com/2504-2289/8/11/149/htmhttps://www.mdpi.com/2504-2289/8/11/149
5. Holeman, I., Kane, D.: Human-centered design for global health equity. Information Technology for Development **26**, 477–505 (7 2020). https://doi.org/10.1080/02681102.2019.1667289, https://www.tandfonline.com/doi/abs/10.1080/02681102.2019.1667289
6. Nuamah, J.K.: Toward user-centered explainable displays for complex machine learning models in healthcare: A case study of heart disease prediction. Proceedings of the Human Factors and Ergonomics Society Annual Meeting (9 2024). https://doi.org/10.1177/10711813241264252/ASSET/2771DA0D-59BC-4DBA-846F-8F9152266DE2/ASSETS/IMAGES/LARGE/10.1177_10711813241264252-FIG5.JPG, https://journals.sagepub.com/doi/10.1177/10711813241264252
7. Panigutti, C., Beretta, A., Fadda, D., Giannotti, F., Pedreschi, D., Perotti, A., Rinzivillo, S.: Co-design of human-centered, explainable ai for clinical decision support. ACM Transactions on Interactive Intelligent Systems **13** (12 2023). https://doi.org/10.1145/3587271/ASSET/0FDF6383-F050-434C-925D-62A8BF4C8B38/ASSETS/GRAPHIC/TIIS-2022-MAR-005-F11.JPG, https://dl.acm.org/doi/10.1145/3587271
8. Patil, D.: Explainable artificial intelligence (xai): Enhancing transparency and trust in machine learning models (11 2024). https://doi.org/10.2139/SSRN.5057400, https://papers.ssrn.com/abstract=5057400
9. Praveen, S., Joshi, K.: Explainable artificial intelligence in health care: How xai improves user trust in high-risk decisions. Studies in Computational Intelligence **1072**, 89–99 (2023). https://doi.org/10.1007/978-3-031-18292-1_6, https://link.springer.com/chapter/10.1007/978-3-031-18292-1_6
10. Prince, E.W., Hankinson, T.C., Gorg, C.: The iterative design process of an explainable ai application for non-invasive diagnosis of cns tumors: A user-centered approach. ... IEEE Workshop on Visual Analytics in Healthcare. IEEE Workshop on Visual Analytics in Healthcare **2023**, 7–13 (2023). https://doi.org/10.1109/VAHC60858.2023.00008, https://pubmed.ncbi.nlm.nih.gov/38989292/

11. Sezgin, E.: Artificial intelligence in healthcare: Complementing, not replacing, doctors and healthcare providers. Digital Health **9**, 20552076231186520 (1 2023). https://doi.org/10.1177/20552076231186520, https://pmc.ncbi.nlm.nih.gov/articles/PMC10328041/
12. Szymanski, M., Abeele, V.V., Verbert, K.: Designing and evaluating explanations for a predictive health dashboard: A user-centred case study. Conference on Human Factors in Computing Systems - Proceedings (5 2024). https://doi.org/10.1145/3613905.3637140/SUPPL_FILE/3613905.3637140-TALK-VIDEO-SUBTITLES.VTT, https://dl.acm.org/doi/10.1145/3613905.3637140
13. Tuan, D.A.: Bridging the gap between black box ai and clinical practice: Advancing explainable ai for trust, ethics, and personalized healthcare diagnostics (9 2024). https://doi.org/10.20944/PREPRINTS202409.1974.V2, https://www.preprints.org/manuscript/202409.1974/v2