Mining Physical Activity and Dietary Intake Data for Guiding Healthier Behaviour

Xiaotong Yu*
xiaotong.yu@sydney.edu.au
University of Sydney
Sydney, NSW, Australia

Corinne Caillaud
corinne.caillaud@sydney.edu.au
University of Sydney
Sydney, NSW, Australia

Kalina Yacef
kalina.yacef@sydney.edu.au
University of Sydney
Sydney, NSW, Australia

Abstract

This research aims to design and create novel data mining approaches to analyze the combined impact of physical activity (PA) and diet intake on human health by identifying their healthy or unhealthy patterns. Our goal is twofold: (1) to identify and capture patterns, both healthy and unhealthy, in rich longitudinal PA data captured by wearable devices and dietary information collected through digital questionnaires, and (2) to generate actionable personalized insights via an interactive dashboard. The key challenges fall into: (1) **Data Integration:** How can we effectively integrate high-granularity PA data derived from wearable devices with self-reported dietary intake data to create a comprehensive representation of daily behaviour? (2) **Prediction and Interpretability:** Can we create a machine learning and data mining pipeline that not only predicts health outcomes accurately but also offers interpretable insights into how PA and dietary factors contribute to health outcomes? (3)**Interactive Dashboard Design:** How to design a user-centred interactive dashboard to visualize the trends and multi-factors, enabling users to explore and understand their data? (4) **Personalized Recommendation - Healthier Behaviour Pathways:** In what forms can we integrate data-driven insights into the interactive dashboard to provide meaningful health recommendation pathways for users to explore and continuously gain healthier behaviours? To tackle the challenges, we designed 3 studies, including a transformer-based feature extraction method integrated with machine learning models for health outcome prediction, an attention-based explainable transformer pipeline to interpret the temporal dynamics in PA data to health outcome prediction, an interactive dashboard embedded in the two models for personalized health recommendations.

ACM Reference Format:
Xiaotong Yu, Corinne Caillaud, and Kalina Yacef. 2025. Mining Physical Activity and Dietary Intake Data for Guiding Healthier Behaviour. In *Proceedings of (CHI '25 Workshop on Envisioning the Future of Interactive Health)*. ACM, New York, NY, USA, 4 pages.

1 Problem Statement

Digitalized health monitoring provides large amounts of data collected via wearable devices and digital questionnaires. Yet, users often find it challenging to understand and derive meaningful insights from these data. Therefore, discovering actionable insights from these multimodal health data remains a challenge, as the complexity of interpreting these data streams limits users' exploration on their healthy or unhealthy behaviors. As wearable devices become increasingly prevalent in the health domain, there are opportunities to analyze health indicators such as physical activity (PA), dietary intake, and sleep quality with high granularity. However, existing studies relied on aggregated metrics, such as daily average Moderate-to-Vigorous Physical Activity (MVPA) and weekly average sleeping time. These aggregations may overlook the temporal patterns in the rich wearable device data and the importance of timing. This interdisciplinary research aims to provide a novel, human-centred framework that integrates predictive modelling, explainability, and interactive dashboards to support personalized healthy behaviour recommendations. By bridging machine learning, explainable AI (XAI), and human-computer interaction (HCI), we aim to enhance user understanding and engagement with wearable device data and enable behaviour change through real-time, transparent, and personalized recommendations. Our approach moves beyond static analysis, creating a dynamic system that continuously refines data-driven insights based on user interaction and discovering health patterns. This research contributes to the growing need for intelligent, explainable, and user-driven digital health solutions that align with real-world health goals.

*All authors contributed equally to this research.

CHI '25 Workshop on Envisioning the Future of Interactive Health, Yokohama, Japan

2 Related Work

This section will focus on introducing the literature in related to how is the current interface design for healthy behaviour recommendation?

The interdisciplinary field of HCI, digital health, and explainable AI has been increasingly explored to enhance user engagement and support healthy behaviour change. Prior research has shown the need for integrating multimodal data sources [9] and explainable predictive models. [7] introduced MyBehavior, a mobile app that generates automated, contextualized, and actionable health recommendations based on user-tracked physical activity and diet, demonstrating the potential of mobile-based behaviour change interventions. [8] developed an interactive visualization tool leveraging parallel coordinates for sensor-based physical activity data analysis, enabling users to explore data patterns more intuitively. [5] proposed HealthPrism, a visual analytics system that integrates motion and contextual data to analyze children's physical and mental health profiles, highlighting the power of interactive, data-driven health monitoring.

In cardiovascular health, [1] designed Prevent Connect, a mobile health app that assesses behavioural risks and provides personalized recommendations for behaviour modification. Similarly, [2] proposed iCARE, a mobile health intervention system for cardiovascular event prevention, demonstrating the effectiveness of individualized, visualized health interventions. [4] applied a design thinking approach to co-develop an mHealth app for older adults, emphasizing the importance of participatory design in engaging end-users. [14] explored time-aware visualization in the HappyFit system, which supports self-monitoring and reflection on physical activity levels, highlighting the importance of aesthetic and intuitive health data presentation.

These studies pointed out the importance of personalized and explainable AI models in digital health. However, existing approaches often lack real-time adaptability and user-centred transparency. To fill this gap, we come up with a user-centered framework for an accuracy health outcomes prediction, at the same time, offer the interpretability of how the time-depend behaviours impact health outcomes. Another aspect is the co-design of an interactive dashboard to provide an interface for the user to explore and understand their own data and current behaviours. Furthermore, the interactive dashboard will embed real-time, explainable AI models and provide healthy behaviour insights and actionable guidance tailored to individual health trajectories instead of one-size-fits-all recommendations.

3 Research Questions

This research aims to address the following research questions:
(RQ1) Can we build a methodology to improve predictions of PA and Dietary Intake patterns on health outcomes?
(RQ2) Can we discover and explain meaningful patterns from wearable device data and digital questionnaires to better understand the time-dependent features that impact health outcomes?
(RQ3) How to create an interactive dashboard to support individuals in exploring and understanding their PA and dietary intake data, in order to adopt healthier behaviours?

Before tackling RQ3, we have already addressed RQ1 and RQ2, and the resulting solutions are integrated into the interactive dashboard to address RQ3.

4 Methodology

Our research consists of three interconnected studies:

4.1 Predictive Model for Health Outcomes

This study focuses on understanding the interplay between PA and dietary intake in predicting health outcomes. We introduce TimePAD (Time-Based Physical Activity and Dietary Intake), a three-stage predictive framework that integrates time-based PA features, dietary intake, and sociodemographic data to predict Healthy Weight Status (HWS) categories, specifically Healthy Fitness Zone (HFZ) and Need Improvement (NI). Unlike traditional methods that aggregate tracker data into daily or weekly estimates, TimePAD leverages a Transformer-based model for Multivariate Time Series analysis [15] to capture fine-grained temporal PA patterns. The model is fine-tuned using unsupervised learning techniques to extract relevant time-based PA features, which, when combined with dietary intake data, enhance HWS prediction. This approach identifies the key behavioural factors influencing HWS and provides new insights into the role of light PA (LPA), MVPA, sedentary (SED) behaviours, sleep duration, and dietary patterns. TimePAD was evaluated on a dataset of 206 adolescents, achieving an accuracy of 82.90% and an F1-score of 67.92%, demonstrating the importance of integrating temporal PA and dietary behaviours in predictive health modelling. This study only uses a Transformer Encoder to extract time-based PA features and further integrate them with dietary intake and sociodemographic features instead of using the whole black-box method for prediction. In this way, we potentially extend the feature extraction technique to a larger cohort, such as the UK Biobank dataset, to generate the first PA pre-train model to facilitate this interdisciplinary area.

4.2 Explainable Model for Time-Dependent Behaviours

Building on the findings of Study 1, this study introduces an explainable AI pipeline that enhances interpretability in predictive health modelling. The pipeline combines Transformer-based temporal modelling [15] with attention-driven interpretability [6] to uncover key time-dependent behavioural patterns from wearable device data. By identifying influential time steps in multivariate time series data, this approach bridges the gap between predictive accuracy and actionable insight generation. To validate the approach, we applied it to two datasets: a PA dataset analyzing movement patterns associated with HWS, and a sleep dataset investigating sleep behaviours linked to sleep health and disorders. Using heatmap-based visualizations, we highlighted critical temporal features driving model predictions, enabling a more intuitive understanding of health behaviours. The study contributes to personalized healthcare by facilitating interpretable, time-aware modelling of multivariate wearable device data. It can be expanded to other health domains, including stress monitoring and chronic disease management.

4.3 Co-designed Interactive Dashboard for Personalized Healthy Behaviour Recommendation

Study 3 builds on the predictive and explainability models developed in Studies 1 and 2, integrating them into a real-time interactive dashboard for personalized health monitoring. In the early stages of design, this study will employ a participatory co-design approach that involves health domain experts and end users to ensure usability and participation. The dashboard will embed the TimePAD predictive model and the Transformer-based interpretable pipeline to provide actionable health recommendations tailored to individual users.

The co-design process plans to begin with stakeholder engagement and requirement assessment, involving conducting structured interviews and focus group discussions with health domain experts and target users. These sessions aim to identify usability requirements, challenges in visualizing wearable health data, and user preferences for dashboard features. More importantly, how to intuitively present the multi-modal data, including data collected via wearable devices and digital questionnaires. Additionally, we will conduct a survey to validate findings and prioritize key functionalities. The following stage is iterative co-design workshops, participants will collaborate through scenario-based walkthroughs and think-aloud protocols to shape the functionality and usability of the dashboard. This iterative feedback loop ensures that core interface elements, including data visualizations and recommendation features, are user-friendly and effectively support adopting healthier behaviours. This dashboard design focuses on two critical functionalities: (1) Incorporate interactive visualizations, such as time-aware PA behavioural heatmaps and context-driven summaries, that facilitate intuitive data exploration. (2) Leverage the predictive capabilities from sections 4.1 and 4.2 to deliver real-time, personalized health guidance. Additionally, the scenario-based healthier behaviour pathway simulations allow users to visualize the outcomes of their behavioural changes, thereby enabling them to make informed actions on achieving healthier behaviours.

5 Evaluation

5.1 Evalute Predictive Model

Study 1 is evaluated from two aspects: (1) the effectiveness of the time-based PA feature extraction (2) the prediction performance. For the effectiveness of time-based PA features, we evaluate the extracted PA feature through unsupervised learning method. To mask the 15% of the PA data and calculate the mask prediction Mean Square Error (MSE). For the health outcome prediction performance, we compared the prediction accuracy and F1-score across 5 machine learning models, including Support Vector Classifier (SVC), k-nearest neigbour (KNN), eXtreme Gradient Boosting (XGBoost) and Adaptive Boosting (AdaBoost), with aggregated PA feature only, dietary intake features only, ARIMA-based PA feature extraction method. Our proposed method outperformed all the other compared method with a meaningful feature ranking.

5.2 Evaluate Explanable Model

Study 2 is evaluated from two aspects: (1) prediction performance against baseline model (2) exam the interpretability. For prediction performance, we compared with a baseline model GRU [3] on both accuracy and F1-score through 2 real-world datasets, the PA dataset [11–13] and the sleeping dataset [10]. Our method significantly outperforms the baseline model GRU. On the other hand, we first conduct statistical analysis on the contrast group, healthy versus unhealthy population, to evaluate which time steps significantly differentiate the two populations. Furthermore, we examine the model's interpretability by visualizing attention weights-based scores with heatmaps in order to highlight which time steps and corresponding features strongly impact health outcomes. Through these complementary quantitative and qualitative assessments, the evaluation demonstrates not only the pipeline's robustness and generalizability but also its capacity to offer meaningful insights into human behavioural patterns.

5.3 Evaluate Co-designed Interactive Dashboard

The final phase involves deploying the dashboard in real-world settings through a longitudinal study. Key evaluation metrics will include engagement rates, perceived usefulness, interpretability scores, and changes in health behaviours over time. Mixed-method evaluations will combine quantitative analytics with qualitative user interviews to measure the impact of the dashboard. Quantitative analytics will capture users' interaction with the dashboard, including frequency of use and adherence to the recommended healthy behaviour pathways. These analytics will be completed by pre- and post-intervention health outcome assessment to discover significant changes in behaviour or health indicators. Qualitative data will be collected through user interview and focus group discussion. We aim to find users' experience, especially for their perception of the dashboard interpretability and the effectiveness of the recommendation pathways. Combining quantitative analytics and qualitative user interviews will offer an overview of dashboard's performance and the impact on promoting healthier behavioural change. Furthermore, the evaluation will guide future refinement of the dashboard and ensure that the dashboard continues to meet users' needs.

6 Contribution

This research integrates predictive health modelling, explainable AI, and HCI-driven interaction for personalized health recommendations. The study introduces a novel framework that not only enhances prediction accuracy but also prioritizes prediction interpretability, making health insights more actionable. By leveraging multivariate time series analysis, the research extends beyond traditional static health assessments, enabling a more dynamic, time-sensitive understanding of individual health behaviours.

A key contribution of this work is integrating an interpretable prediction model within a co-designed dashboard. This setup enables users to interact directly with their complex health data, uncovering behavioural insights derived from the prediction model's interpretation. Rather than offering one-size-fits-all feedback, the dashboard provides real-time, adaptive, and personalized recommendations that dynamically adjust based on individual behaviour. Users can explore actionable recommendation pathways, simulate potential changes, and visualize expected outcomes through the interactive dashboard. This level of customization significantly enhances engagement and promotes sustainable behaviour change.

By combining predictive modelling, explainability, and user-centred design, this research represents a step forward in advancing transparent, adaptive, and personalized digital health interventions. The findings contribute to the broader HCI and Health interdisciplinary, providing for more interpretable, user-driven health analytics systems that enable individuals to make healthy behavioural changes in their well-being.

References

[1] Dahbia Agher, Karima Sedki, Sylvie Despres, Jean-Pierre Albinet, Marie-Christine Jaulent, and Rosy Tsopra. 2022. Encouraging behavior changes and preventing cardiovascular diseases using the prevent connect mobile health app: conception and evaluation of app quality. *Journal of medical Internet research* 24, 1 (2022), e25384.

[2] Yuling Chen, Meihua Ji, Ying Wu, Ying Deng, Fangqin Wu, and Yating Lu. 2021. Individualized mobile health interventions for cardiovascular event prevention in patients with coronary heart disease: study protocol for the iCARE randomized controlled trial. *BMC Cardiovascular Disorders* 21 (2021), 1–11.

[3] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

[4] Kim Daniels, Ryanne Lemmens, Els Knippenberg, Nastasia Marinus, Sharona Vonck, Jan Baerts, Jochen Bergs, Annemie Spooren, Dominique Hansen, and

Bruno Bonnechère. 2023. Promoting physical activity and a healthy active lifestyle in community-dwelling older adults: a design thinking approach for the development of a mobile health application. *Frontiers in public health* 11 (2023), 1280941.

[5] Zhihan Jiang, Handi Chen, Rui Zhou, Jing Deng, Xinchen Zhang, Running Zhao, Cong Xie, Yifang Wang, and Edith CH Ngai. 2023. HealthPrism: A Visual Analytics System for Exploring Children's Physical and Mental Health Profiles with Multimodal Data. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[6] Yao Qiang, Deng Pan, Chengyin Li, Xin Li, Rhongho Jang, and Dongxiao Zhu. 2022. Attcat: Explaining transformers via attentive class activation tokens. *Advances in neural information processing systems* 35 (2022), 5052–5064.

[7] Mashfiqui Rabbi, Angela Pfammatter, Mi Zhang, Bonnie Spring, Tanzeem Choudhury, et al. 2015. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. *JMIR mHealth and uHealth* 3, 2 (2015), e4160.

[8] Chuxuan Tong, Jinglan Zhang, Alok Chowdhury, and Stewart G Trost. 2019. An interactive visualization tool for sensor-based physical activity data analysis. In *Proceedings of the Australasian Computer Science Week Multiconference*. 1–4.

[9] Riaz Uddin, Eun-Young Lee, Shanchita R Khan, Mark S Tremblay, and Asaduzzaman Khan. 2020. Clustering of lifestyle risk factors for non-communicable diseases in 304,779 adolescents from 89 countries: A global perspective. *Preventive medicine* 131 (2020), 105955.

[10] K. Wang, J. Yang, A. Shetty, and J. Dunn. 2025. DREAMT: Dataset for Real-time sleep stage EstimAtion using Multisensor wearable Technology (version 2.0.0). PhysioNet. https://doi.org/10.13026/0vrv-nn81

[11] G. Wattelez, É. Paufique, P.-Y. Le Roux, A. Nedjar-Guerre, S. Ponidja, P. Zongo, C. Serra-Mallol, F. Wacalie, S. Frayon, and O. Galy. 2024. GENEActiv accelerometer files collected during the project entitled "Cultures et comportements alimentaires de la jeunesse dans les pays francophones du Pacifique au XXIème siècle: exemple de la Nouvelle-Calédonie" [Eng: "Eating cultures and behaviors of young people in French-speaking Pacific countries in the 21st century: the example of New Caledonia"] (anonymized version - first part) (1.0.0). Data set. https://doi.org/10.5281/zenodo.12615468 Zenodo.

[12] G. Wattelez, É. Paufique, P.-Y. Le Roux, A. Nedjar-Guerre, S. Ponidja, P. Zongo, C. Serra-Mallol, F. Wacalie, S. Frayon, and O. Galy. 2024. GENEActiv accelerometer files collected during the project entitled "Cultures et comportements alimentaires de la jeunesse dans les pays francophones du Pacifique au XXIème siècle: exemple de la Nouvelle-Calédonie" [Eng: "Eating cultures and behaviors of young people in French-speaking Pacific countries in the 21st century: the example of New Caledonia"] (anonymized version - second part) (1.0.0). Data set. https://doi.org/10.5281/zenodo.12638746 Zenodo.

[13] G. Wattelez, É. Paufique, P.-Y. Le Roux, A. Nedjar-Guerre, S. Ponidja, P. Zongo, C. Serra-Mallol, F. Wacalie, S. Frayon, and O. Galy. 2024. GENEActiv accelerometer files collected during the project entitled "Cultures et comportements alimentaires de la jeunesse dans les pays francophones du Pacifique au XXIème siècle: exemple de la Nouvelle-Calédonie" [Eng: "Eating cultures and behaviors of young people in French-speaking Pacific countries in the 21st century: the example of New Caledonia"] (anonymized version - third part) (1.0.0). Data set. https://doi.org/10.5281/zenodo.12682660 Zenodo.

[14] Soojeong Yoo, Lichen Xue, and Judy Kay. 2017. Happyfit: time-aware visualization for daily physical activity and virtual reality games. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. 391–394.

[15] George Zerveas, Srideepika Jayaraman, Dhaval Patel, Anuradha Bhamidipaty, and Carsten Eickhoff. 2021. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 2114–2124.