Multi-Modal AI for Remote Patient Monitoring in Cancer Care

Yansong Liu[1] **Ronnie Stafford**[1,2] **Pramit Khetrapal**[1,2] **Huriye Kocadag**[1]
Graça Carvalho[1,2,3] **Patricia de Winter**[1] **Maryam Imran**[1] **Amelia Snook**[1]
Adamos Hadjivasiliou[1] **D. Vijay Anand.**[1] **Weining Lin**[1]
John Kelly[1,2] **Yukun Zhou**[1] **Ivana Drobnjak**[1]

{yansong.liu.18, r.stafford.12, p.khetrapal, h.kocadag,
graca.carvalho, p.winter, maryam.imran.20, amelia.snook.20,
adamos.hadjivasiliou.17, v.dharmalingam, w.lin.16
j.d.kelly, yukun.zhou.19, i.drobnjak} @ucl.ac.uk

[1]University College London, UK, WC1E 6BT
[2]Ethera Health LTD, 125 Wood Street, London, UK, EC2V 7AW
[3]Centro Algoritmi, Universidade do Minho, Braga, Portugal

Abstract

For patients undergoing systemic cancer therapy, the time between clinic visits is full of uncertainties and risks of unmonitored side effects. To bridge this gap in care, we developed and prospectively trialed a multi-modal AI framework for remote patient monitoring (RPM). This system integrates multi-modal data from the HALO-X platform, such as demographics, wearable sensors, daily surveys, and clinical events. Our observational trial is one of the largest of its kind and has collected over 2.1 million data points (6,080 patient-days) of monitoring from 84 patients. We developed and adapted a multi-modal AI model to handle the asynchronous and incomplete nature of real-world RPM data, forecasting a continuous risk of future adverse events. The model achieved an accuracy of 83.9% (AUROC=0.70). Notably, the model identified previous treatments, wellness check-ins, and daily maximum heart rate as key predictive features. A case study demonstrated the model's ability to provide early warnings by outputting escalating risk profiles prior to the event. This work establishes the feasibility of multi-modal AI RPM for cancer care and offers a path toward more proactive patient support. https://github.com/LiuYYSS/EurIPS2025

1 Introduction

For individuals undergoing cancer systemic therapy (such as chemotherapy), the period between clinical appointments unfolds at home and is often fraught with uncertainty and the risk of unmonitored side effects or rapid health deterioration [1]. In the meantime, due to physical distance, clinicians have limited visibility into their patients' at-home daily well-being and the subtle onset of treatment-related side effects, creating a critical gap in cancer care [2].

RPM has the potential to address this problem by using multiple sources (such as wearable sensors, mobile apps, and other digital tools) to gather at-home health data [3]. To unlock the true potential of RPM, multi-modal AI was involved to fuse sources together to generate actionable decisions [4]. This approach is highly likely to fill the gap because it mirrors the multi-aspect reasoning process of clinicians at home environments.

Previous research in multi-modal AI RPM in cancer care handled the multi-modalities poorly. Some recent research was based on single modality (wearable only [5, 6], survey only [7, 3]). Some research had to reduce sampling frequency because certain modality was sampled slower than the others [8, 9]. There is a lack of standardised framework to address this.

This study presents our observational prospective trial using a framework that collects and uses multi-modal RPM data to forecast adverse clinical events. Specifically, we explored the asynchronous and non-random missingness problems through a token-based transformer, validated on the trial data. Ultimately, this work establishes the feasibility of multi-modal AI RPM for cancer care and offers a path toward more proactive patient support.

39th Conference on European Neural Information Processing Systems (EurIPS 2025) Workshop: Multimodal Representation Learning for Healthcare.

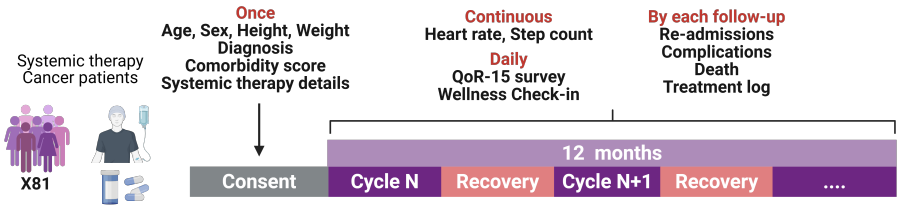

Figure 1: Overview of trial design. Multimodal RPM data collection up to 12 months

2 Methods

Multi-modal RPM data Collection: A prospective observational trial (IRAS 312296) was designed based on the HALO-X multi-modal data collection platform [10]. The trial aimed at testing the feasibility of the HALO-X platform in patients with advanced cancer undergoing systemic therapy. Patient recruitment started in March 2023 (inclusion & exclusion criteria see appendix A.1). Detailed collection process is illustrated in Figure 1. Multi-modal data was collected and combined via the platform: baseline characteristics, wearable data in 5-minute epochs, surveys, and clinical events. Surveys used were the 15-item Quality of Recovery questionnaire (QoR-15, 0-10 scale per item, 10=best) [11] and wellness check-in (an optional yes/no question to indicate overall health today). Treatment log contains information from treatment events such as treatment type (chemotherapy, hormone therapy, immunotherapy etc.), health-caused dose reduction or treatment delay (e.g. low blood platelets).

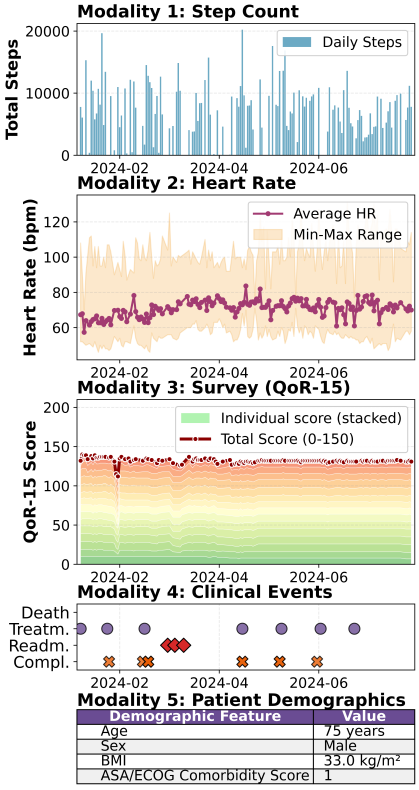

Figure 2: Representative patient timeline illustrating longitudinal asynchronous multi-modalities data with missingness. Raw sensor data were aggregated to daily values to avoid clustering in this visualisation. Treatnm.=Treatment; Readm.=Readmission; Compl.=Complications.

Multi-modal RPM Model Development: Data stream asynchronicity is a key consideration when building multi-modal AI for RPM data. The problem appears as: wearable sensors sampled every 5 minutes, survey responses arrived daily, and clinical events occurred irregularly across months of treatments. The way that traditional supervised learning model works either causes the least frequent feature to be re-used numerous times or binning the most frequent feature to a lower frequency (which significantly reduces temporal resolution). Additionally, patients with poor health are known to use monitoring services less [12–14]. This introduces missingness not at random (MNAR). MNAR in healthcare is notorious for prevalence and tricky to solve [15, 16]. To develop an AI model tailored to these data characteristics, we took the concept and adapted a transformer-architecture from the STraTS [17] with some key adjustments. First, tokenizer encodes each observation as an independent token, creating token sequences where each modality contributes tokens at its native sampling rate, avoiding the first problem mentioned. Second, because each observation is tokenized independently, missing observations can be simply skipped rather than forced to be imputed, avoiding the second problem mentioned. Third, we explicitly engineered missingness features (daily device wearing percentage and continuous absence duration) as additional tokens to enhance missingness learning. The model architecture and training/evaluation pipeline is illustrated in Figure 3.

The model was trained to forecast future adverse clinical events (general practitioner visit due to treatment-related reasons, accident&emergency department visit, re-admission, treatment delay/dose reduction, or death) within rolling 4-week windows. The output is a float value between 0 and 1, where 1 represents extremely high risk. The model takes all available historical monitoring data

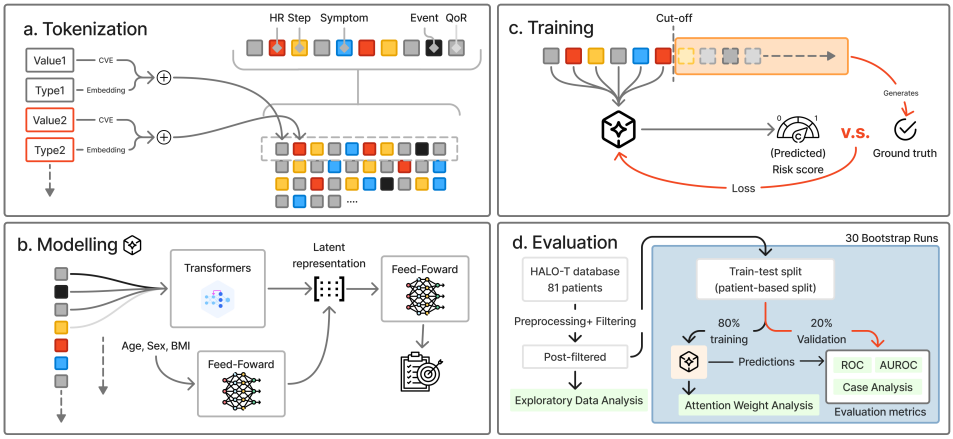

Figure 3: Model architecture and training pipeline. (a) Value and type of each observation were extracted and tokenized, CVE=Continuous Variable Encoder; (b) Transformer layers for temporal modelling. Static features were encoded via a separate feed-forward neural network; (c) Model uses data before cut-off to generate a binary classification forecast; (d) Evaluation paradigm.

(minimum 2 weeks) as input. Prior to modelling, filtering rules (Appendix table 1) were applied to remove patients with artifacts. Heart rate was clipped to 40-200 bpm, steps clipped to 0-600 per 5-minute epoch. Training employed sliding window re-sampling to generate multiple samples from each patient's longitudinal trajectory, with patient-level stratified splitting (80-20) to prevent data leakage. To reduce the stress on GPU memory and improve learning efficiency, raw sensor data were aggregated into representative daily values and max sequence length was kept at 1000 tokens. The hyperparameters were selected manually based on the original hyperparameters from STraTS with batch size set to 128, 80 epochs, and 5e-4 learning rate. During evaluation, bootstrap aggregation with 30 iterations was used to provide performance estimates with confidence intervals. Attention weight was extracted from the attention layer for feature importance analysis. Evaluation results from each sliding window over time were put together and connected to generate risk trajectory over time for case-analysis. A single NVIDIA RTX GPU was used, and took 2 hours to complete all bootstraps.

3 Results

Up to December 2024 (the date the dataset was extracted), the HALO-X platform collected over 2.1 million data points (top three sources: 1.2M in heart rate, 0.5M in step count, 0.07M in QoR-15) from 6,080 patient-days of monitoring data, marking one of the largest prospective trial datasets in multi-modal RPM in cancer systemic therapy to date. 50 patients (5,296 patient-days, 87%) were included in the final analysis after filtering (Appendix table 1). On average, each patient was monitored for 98 days (median 76 days, max 298 days, min 42 days). 30 features covering multiple modalities were extracted for training and prediction (Appendix table 2). Among 50 patients, 17 patients produced 66 adverse events during the monitoring period (treatment delay/dose reduction was the most frequent adverse event, accounting for 32 occurrences, 48%). Because only events within 4-week windows were considered as positive labels, among 9,313 sliding windows generated, only 1,265 (13.6%) were labelled positive. The dataset is moderately unbalanced with a 6.4:1 negative-to-positive ratio.

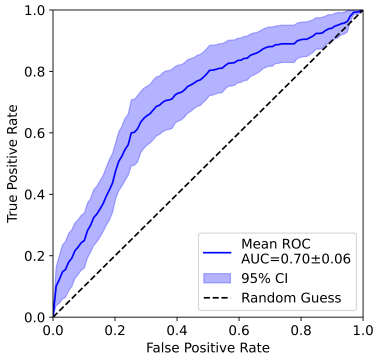

Figure 4: ROC curve of the model with 95% CI.

The model achieved 0.70 (95% CI: 0.64-0.76) area under ROC and 83.9% (95% CI: 81.5-86.3%) accuracy (figure 4). Ranking the extracted attention weights by importance (figure 5 a and b), features in the clinical event type are generally stronger than remote monitoring features, likely due to lower frequency, thus higher information density. The strongest feature in event type is previous

chemotherapy treatment, followed by A&E department visit and immunotherapy treatment. The top remote monitoring feature is wellness check-in, followed by good sleep and able to work (two items in QoR-15). The top wearable sensor feature is daily maximum heart rate. Percentage of daily wearing is also among the top features, indicating the model learned missingness information. In figure 5 c, we showcase a representative patient risk trajectory over time with clinical events annotated. The patient was selected from the testing set on which the model had never been trained. The predicted risk elevated ahead of the treatment delay event. Even though haemoglobin and platelet level were not among the features, the model was able to capture the subtle health deterioration via other data. After about a month of recovery, the patient received the next treatment with intended dosage, and the risk score dropped back to a low level. This case demonstrates the potential of the model in identifying high-risk periods and comparing recovery trajectories.

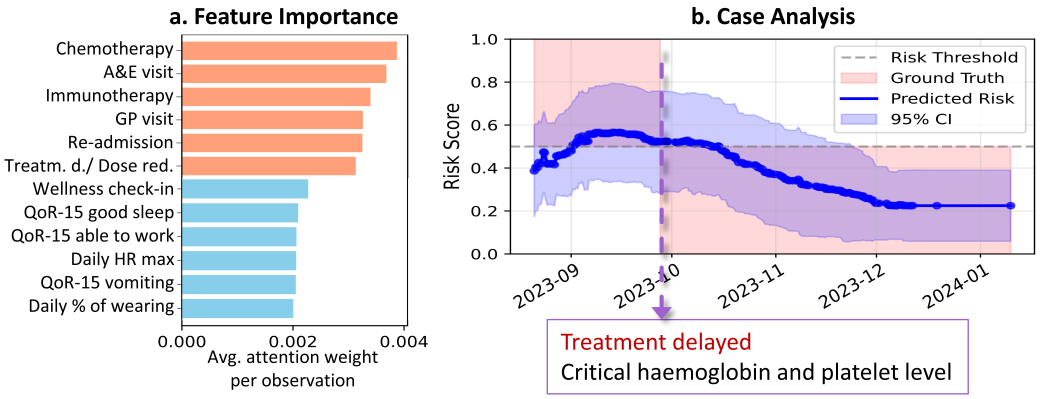

Figure 5: (a) Overview of the most important features, orange bars indicate event features which are less frequent but impactful. Blue bars are remote monitoring feature which happens more regularly and indicates recovery status; (b) Example patient trajectory over time with clinical events annotated. The risk score (y-axis) ranges from 0 to 1, where 1 indicates extremely high risk. Treatm. d./Dose Red.=Treatment Delay/Dose Reduction; A&E=Accident & Emergency department; GP=General Practitioner; QoR-15=Quality of Recovery 15 survey.

4 Discussion

This study shows that a multi-modal AI remote patient monitoring (RPM) in cancer therapy is feasible. We ran a prospective trial that collected >2.1 million data points across 6,000 patient-days, and developed and adapted an algorithm that addresses asynchronous data streams and non-random missingness (MNAR) common in real-world monitoring. The model's initial performance (AUC = 0.70, accuracy = 84%) is solid, but the key contribution is the practical and scientific insight: combining continuous wearable signals, daily surveys, and clinical event history produces a dynamic risk profile that better captures patient state between clinic visits.

Our exploratory analyses revealed which inputs drive prediction. Clinical event history holds strong predictive value. Among remote monitoring signals, our optional wellness check-ins, sleep quality, able to work, and daily maximum heart rate emerged as important indicators, showing the value of collecting daily well-being measurements. A case study shows a rising risk score before a treatment delay, illustrating how this multi-modal AI RPM can be used for early warning signs and could enable proactive clinical intervention. Overall, this work demonstrates feasibility, identifies high-value features for RPM in cancer care, and provides a roadmap for larger validation and clinical integration.

As an analysis of an ongoing study, this work has several limitations. The biggest being data size, and a series of other limitations caused by this. The analysis was based on longitudinal data from just 50 patients. With that said, given how new this field is and the difficulty of collecting prospective data in healthcare, the dataset is already one of the largest prospective trials in multi-modal RPM in cancer systemic therapy to date. Future work should focus on expanding the cohort, validating across multiple centres, and improving accuracy. This study serves as a crucial first step, providing the blueprint and initial evidence needed to build toward a future where multi-modal AI and remote monitoring become integral tools for delivering more personalized and proactive cancer care.

Acknowledgments and Disclosure of Funding

We would like to thank Ethera Health LTD for the support given to this work through the research collaboration, *Proof of Concept Study to Test a Framework for Integration of Data Streams to Explore the Relation between Measures of Molecular Genomic Markers and Performance in Patients Receiving Chemotherapy for Metastatic Cancer* and the use of the HALO-X RPM data collection platform.

References

[1] Claire M Oakley, Catherine Taylor, Emma Ream, and Alison Metcalfe. Patients making potentially life threatening delays in reporting neutropenic sepsis: Grounded theory study. *Journal of Clinical Oncology*, 35(15_suppl):e21560–e21560, 2017. doi: 10.1200/JCO.2017.35.15_suppl.e21560. Grounded theory study on patient delays in reporting life-threatening neutropenic sepsis during chemotherapy.

[2] Leah Chen, Christianna Bartel, Xinran Cai, Yu Cheng, Adam Perer, and Carissa A Low. Patient and provider perspectives on symptom monitoring during outpatient chemotherapy: interview study. *JMIR Formative Research*, 7:e46001, 2023. doi: 10.2196/46001. Interview study revealing gap in clinical visibility and monitoring between outpatient chemotherapy visits.

[3] Callie Rockey-Bartlett, Jennifer Morelli, Melissa Coffel, John Geracitano, Jennifer Elston Lafata, and Saif Khairat. Effect of remote patient monitoring on healthcare use among patients with cancer: A systematic review. *DIGITAL HEALTH*, 11:20552076251384220, 2025.

[4] Luis R Soenksen, Yu Ma, Cynthia Zeng, Léonard Boussioux, Kimberly Carballo, Liangyuan Zhang, Holly Lizotte, and Dimitris Bertsimas. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ digital medicine*, 5(1):149, 2022. doi: 10.1038/s41746-022-00689-4. Comprehensive framework for multimodal AI data fusion in healthcare.

[5] G Cay, YH Sada, M Dehghan Rouzi, et al. Harnessing physical activity monitoring and digital biomarkers of frailty from pendant based wearables to predict chemotherapy resilience in veterans with cancer. *Scientific Reports*, 14(1):5025, 2024. doi: 10.1038/s41598-024-53025-z.

[6] Carissa A Low, Christianna Bartel, Krina Durica, Heidi S Donovan, Roby Thomas, and Jennifer Fedor. Consumer wearable device measures of gait cadence and activity fragmentation as predictors of survival among patients undergoing chemotherapy. *JCO Clinical Cancer Informatics*, 9:e2500111, 2025.

[7] Elizabeth Kairis, Jennifer Fedor, Christianna Bartel, Krina C Durica, Heidi S Donovan, Roby Thomas, and Carissa A Low. Daily patient-reported symptoms as predictors of unplanned health care encounters during chemotherapy: Longitudinal observational study. *JCO oncology practice*, pages OP–25, 2025.

[8] Malte Jacobsen, Rahil Gholamipoor, Till A Dembek, Pauline Rottmann, Marlo Verket, Julia Brandts, Paul Jäger, Ben-Niklas Baermann, Mustafa Kondakci, Lutz Heinemann, et al. Wearable based monitoring and self-supervised contrastive learning detect clinical complications during treatment of hematologic malignancies. *Npj Digital Medicine*, 6(1):105, 2023.

[9] Jen-Hsuan Liu, Chih-Yuan Shih, Hsien-Liang Huang, Jen-Kuei Peng, Shao-Yi Cheng, Jaw-Shiun Tsai, and Feipei Lai. Evaluating the potential of machine learning and wearable devices in end-of-life care in predicting 7-day death events among patients with terminal cancer: cohort study. *Journal of Medical Internet Research*, 25:e47366, 2023.

[10] Ronnie Stafford, Graça Carvalho, Yansong Liu, Pramit Khetrapal, Huriye Kocadag, John Kelly, and Ivana Drobnjak. HALO-X: A remote patient monitoring platform for personalised cancer recovery: System design, deployment, and clinical insights. *Arxiv*, 2025.

[11] Peter A Stark, Paul S Myles, and Justin A Burke. Development and psychometric evaluation of a postoperative quality of recovery score: the qor-15. *Anesthesiology*, 118(6):1332–1340, 2013.

[12] S McClaine, J Fedor, C Bartel, L Chen, KC Durica, et al. Factors associated with participant engagement with daily symptom reporting, passive smartphone sensing, and wearable device data collection during chemotherapy: longitudinal observational study. *JMIR Cancer*, 10: e57347, 2024. doi: 10.2196/57347.

[13] Simon B Goldberg, Daniel M Bolt, and Richard J Davidson. Data missing not at random in mobile health research: Assessment of the problem and a case for sensitivity analyses. *Journal of Medical Internet Research*, 23(6):e26749, 2021.

[14] BL Ortiz, V Gupta, R Kumar, A Jalin, X Cao, C Ziegenbein, A Singhal, M Tewari, and SW Choi. Data preprocessing techniques for ai and machine learning readiness: Scoping review of wearable sensor data in cancer care. *JMIR mHealth and uHealth*, 12:e59587, 2024. doi: 10.2196/59587. PMID: 38626290.

[15] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2nd edition, 2002.

[16] Jonathan AC Sterne, Ian R White, John B Carlin, Michael Spratt, Patrick Royston, Michael G Kenward, Angela M Wood, and James R Carpenter. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*, 338, 2009.

[17] Sindhu Tipirneni and Chandan K Reddy. Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. *ACM Transactions on Knowledge Discovery from Data*, 16(6):1–17, 2022. doi: 10.1145/3516367.

A Technical Appendices and Supplementary Material

A.1 Inclusion and Exclusion Criteria

- Inclusion criteria:
  - Diagnosis of advanced or haematological cancer undergoing systemic therapy;
  - Planned for at least one cycle of standard-of-care systemic therapy;
  - Ability to provide informed consent;
  - Ambulatory without assistance or walking aids;
  - Possession of an Android or iOS smartphone with willingness to use the HALO-X monitoring smartphone application.
- Exclusion criteria:
  - Physical disabilities precluding daily walking;
  - Inability to provide informed consent;
  - Inability to operate the HALO-X monitoring smartphone application;
  - Medical or psychiatric conditions that would affect study completion, as determined by the investigator.

A.2 Data Filtering and Feature Overview

	Filtering reason	Num. Patient Removed
1	Patients that were not followed up for clinical events	4
2	Patients with empty record	3
3	Patients that were not allocated with a device	3
4	Patients with no treatment received in monitoring period	5
5	Patients providing any RPM data for <3 days	11
6	Patient involved in a car accident	1
7	Patients received stem cell therapy (distinctive pattern and limited population)	5
8	Patients who joined the trial too shortly before data extraction	3
	Population extracted	85
	Patients Affected	35
	Final population	50

Table 1: Patient filtering rules, number of patients affected by each rule, and total number of patients remaining after filtering.

Categories	Wearable	Survey	Previous Clinical Events	Demographics
	Daily max HR	QoR-15 total score	Chemotherapy	Age
	Daily total steps	QoR-15 individual score x15	Hormone therapy	Gender
	Daily % of wearing	Complication seriousness	Immunotherapy	BMI
	Duration of continuous absence	Wellness check-in	Mixed therapy	
			Re-admission	
			General practitioner visit	
			Accident & Emergency department visit	
			Dose reduction / Treatment delay due to health	

Table 2: Overview of features used in the model. BMI=Body Mass Index; HR=Heart Rate; QoR=Quality of Recovery.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract exactly matches the scope of the content.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Despite there not being a stand-alone limitation section due to page limits, the limitations of this work are well discussed in the last paragraph of the discussion.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA] .

   Justification: No theoretical result was proposed in this work.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The main result and contribution of this paper are the knowledge and insights we gained while developing the multi-modal AI remote patient monitoring framework and modelling the data. We have tried our best to document all necessary settings to reproduce the exact modelling framework under the page limit. For detailed implementation that can not be fitted into the paper, we have hosted a open source repository on GitHub. With the open sourced code, we have high hope others would be able to replicate the modelling framework we explored in this study.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Albeit the core contribution of this paper did not rely on the source code, the code is open sourced to help the understanding of the framework. The dataset can be access upon request.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All relevant details were fully specified in methodology section.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Error range and 95% confidence interval are reported in all applicable results. Bootstrapping method was used to ensure the representativeness of the results.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Specified in Methodology section.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conforms with NeurIPS Code of Ethics in every respect.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The work has potential to improve patient safety through early detection of adverse events, reduce healthcare burden via proactive monitoring, and empower patients through continuous engagement. As with any AI healthcare system, considerations include ensuring model fairness across diverse populations, integrating predictions appropriately with clinical judgment, protecting patient privacy in data collection, and ensuring equitable access for patients with varying levels of technical literacy. These broader considerations are discussed in the Discussion section.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA] .

    Justification: Not applicable because no model or data was released for open access in this work.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes] .

    Justification: All original creators or owners were properly cited in the article.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes] .

    Justification: Code is accompanied with readme files.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA] .

    Justification: Despite this trial involving asking patients to wear wearable sensors and use a mobile app, the data collection is mainly observational and does not in any form or way influence the original intended way of treatment or care. All patients gave written consent before being on-boarded to the trial. No crowdsourcing is conducted in this work. Hence, this question is not applicable to the trial involved in this research.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [Yes] .

    Justification: Tche data collection is mainly observational and does not in any form or way influence the original intended way of treatment or care. IRB approvals were obtained prior to data collection. And, the trial was registered on the integrated research application system (IRAS number withheld for blind review).

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The study did not use LLMs.